



- Expert Verified, Online, **Free**.



## **CERTIFICATION TEST**

- [CertificationTest.net](https://CertificationTest.net) - Cheap & Quality Resources With Best Support

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

### Existing Environment -

#### Identity Environment -

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

#### Data Environment -

Contoso has the following data environment:

The Sales division uses a Microsoft Power BI Premium capacity.

The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.

The Research department uses an on-premises, third-party data warehousing product.

Fabric is enabled for contoso.com.

An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.

A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

### Requirements -

#### Planned Changes -

Contoso plans to make the following changes:

Enable support for Fabric in the Power BI Premium capacity used by the Sales division.

Make all the data for the Sales division and the Research division available in Fabric.

For the Research division, create two Fabric workspaces named Productline1ws and Productline2ws.

In Productline1ws, create a lakehouse named Lakehouse1.

In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

#### Data Analytics Requirements -

Contoso identifies the following data analytics requirements:

All the workspaces for the Sales division and the Research division must support all Fabric experiences.

The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.

The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.

For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and

shortcuts by using SQL endpoints.

For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.

All the semantic models and reports for the Research division must use version control that supports branching.

#### Data Preparation Requirements -

Contoso identifies the following data preparation requirements:

The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.

All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

#### Semantic Model Requirements -

Contoso identifies the following requirements for implementing and managing semantic models:

The number of rows added to the Orders table during refreshes must be minimized.

The semantic models in the Research division workspaces must use Direct Lake mode.

#### General Requirements -

Contoso identifies the following high-level requirements that must be considered for all solutions:

Follow the principle of least privilege when applicable.

Minimize implementation and maintenance effort when possible.

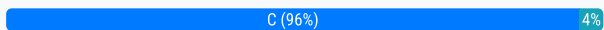
You need to ensure that Contoso can use version control to meet the data analytics requirements and the general requirements.

What should you do?

- A. Store at the semantic models and reports in Data Lake Gen2 storage.
- B. Modify the settings of the Research workspaces to use a GitHub repository.
- C. Modify the settings of the Research division workspaces to use an Azure Repos repository.
- D. Store all the semantic models and reports in Microsoft OneDrive.

**Suggested Answer: B**

Community vote distribution



**Nicofr** 9 months, 1 week ago

**Selected Answer: C**

"Currently, only Git in Azure Repos is supported." <https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration#considerations-and-limitations>

upvoted 16 times

**stilferx** 9 months, 1 week ago

**Selected Answer: C**

IMHO, the Azure Repos is the answer.

It is directly mentioned here:

<https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration>

Considerations and limitations

Currently, only Git in Azure Repos with the same tenant as the Fabric tenant is supported.

If the workspace and Git repo are in two different geographical regions, the tenant admin must enable cross-geo exports.

Azure DevOps on-prem isn't supported.

Sovereign clouds aren't supported.

upvoted 9 times

**Ce\_\_\_\_\_** 1 week, 5 days ago

**Selected Answer: C**

The versioning part of the question narrows the answer to B or C, while the specificity of the workspace: "Research Division", just as referenced in the case scenario makes the answer "C"

upvoted 1 times

**goldy29** 5 months ago

**Selected Answer: C**

You can use both Git and Azure as Repos for version control but if you read the question properly it clearly says that Research Division Workspace. Hence, C is the correct answer.



upvoted 1 times

  **rafaelfiss** 5 months ago

**Selected Answer: C**

Starting in September 2024, GitHub has also been supported for integration with Microsoft Fabric. At the time this question was created, the correct answer was only option C. Currently, both B and C are correct and support branching.



upvoted 1 times

  **Egocentric** 5 months, 2 weeks ago

**Selected Answer: C**

C is the correct answer



upvoted 1 times

  **NRezgui** 6 months, 1 week ago

**Selected Answer: C**

« Actuellement, seul Git dans Azure Repos est pris en charge. » <https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration#considerations-and-limitations>

upvoted 1 times

  **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

C is the answer.

<https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration#considerations-and-limitations>

upvoted 1 times

  **Naqib** 7 months, 3 weeks ago

<https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration?tabs=azure-devops>



Based on the Microsoft Learn new update dated 09/22/2024, both Git in Azure Repos(with the same tenant as the Fabric tenant) and GitHub is supported now. So the right answer for now are Azure Repos and GitHub.

upvoted 3 times

  **jass007\_k** 8 months, 1 week ago

Now Github is also supported. What would be the correct answer now? Github or Azure Repos?

upvoted 2 times

  **Egocentric** 8 months, 1 week ago

C is the answer Azure Repos

upvoted 1 times

  **36eeabb** 11 months ago

B. HitHub integration is now supported. (Preview)

<https://blog.fabric.microsoft.com/en-US/blog/announcing-github-integration-for-source-control-preview/>

upvoted 2 times

  **RalphLiang** 11 months, 3 weeks ago

Microsoft copilot recommend option C:

upvoted 2 times

  **ralavi** 1 year ago

I guess C is correct

upvoted 1 times



  **Jons123son** 1 year, 1 month ago

**Selected Answer: C**

C - "Currently, only Git in Azure Repos with the same tenant as the Fabric tenant is supported."

<https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration#considerations-and-limitations>

upvoted 2 times

  **gills** 1 year, 1 month ago

<https://learn.microsoft.com/en-us/fabric/cicd/git-integration/intro-to-git-integration>

Only Git in Azure Repos is supported at the moment.

upvoted 2 times

  **dev2dev** 1 year, 1 month ago

**Selected Answer: C**

Git in Azure Repos

upvoted 2 times

## HOTSPOT -

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

### Existing Environment -

#### Identity Environment -

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

#### Data Environment -

Contoso has the following data environment:

The Sales division uses a Microsoft Power BI Premium capacity.

The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.

The Research department uses an on-premises, third-party data warehousing product.

Fabric is enabled for contoso.com.

An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.

A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

### Requirements -

#### Planned Changes -

Contoso plans to make the following changes:

Enable support for Fabric in the Power BI Premium capacity used by the Sales division.

Make all the data for the Sales division and the Research division available in Fabric.

For the Research division, create two Fabric workspaces named Productline1ws and Productline2ws.

In Productline1ws, create a lakehouse named Lakehouse1.

In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

#### Data Analytics Requirements -

Contoso identifies the following data analytics requirements:

All the workspaces for the Sales division and the Research division must support all Fabric experiences.

The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.

The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name. For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.

For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.

All the semantic models and reports for the Research division must use version control that supports branching.

#### Data Preparation Requirements -

Contoso identifies the following data preparation requirements:

The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.

All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

#### Semantic Model Requirements -

Contoso identifies the following requirements for implementing and managing semantic models:

The number of rows added to the Orders table during refreshes must be minimized.

The semantic models in the Research division workspaces must use Direct Lake mode.

#### General Requirements -

Contoso identifies the following high-level requirements that must be considered for all solutions:

Follow the principle of least privilege when applicable.

Minimize implementation and maintenance effort when possible.

You need to recommend a solution to group the Research division workspaces.

What should you include in the recommendation? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

#### Answer Area

Grouping method:  ▼

Capacity

Domain

Tenant

Tool:  ▼

OneLake data hub

The Fabric Admin portal

The Microsoft Entra admin center

**Answer Area**

Grouping method:  ▼

Capacity

**Domain**

Tenant

Tool:  ▼

**OneLake data hub**

The Fabric Admin portal

The Microsoft Entra admin center

**Suggested Answer:**

**Momoanwar** Highly Voted 1 year, 4 months ago

Domain\Fabric Admin Portal  
upvoted 62 times

**XiltroX** Highly Voted 9 months, 1 week ago

<https://learn.microsoft.com/en-us/fabric/governance/domains#create-a-domain>  
From MS Learn

1. "In Fabric, a domain is a way to logically group together services in the organization"
  2. You have to use Fabric Admin portal to do the groupings.
- upvoted 16 times

**testtaker45** Most Recent 5 months, 2 weeks ago

The right answer is Doamin \ Onelake Data Hub.

The key wording is the Research Group will use 'Lakehouse Explorer'. This tool will connect to the Onelake Data Hub.

From an admin perspective, you use the Fabric Admin portal to create the Domain Grouping. I think that is the confusion. Hope that helps someone.

upvoted 2 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

Domain & The Fabric Admin portal.

upvoted 1 times

🗨️ 👤 **Egocentric** 8 months, 1 week ago

Domain is for grouping resources so Domain and Fabric Admin portal

upvoted 2 times

🗨️ 👤 **hjhjh123** 9 months, 2 weeks ago

its a capacity and fabric

: Domain Purpose: Domains are typically used for organizing and managing resources in a broader context and for access control rather than directly impacting performance and resource allocation.

Not Ideal for Resource Management: While domains can help with organizational structure, they do not offer the performance benefits and cost management that a dedicated capacity does.

upvoted 1 times

🗨️ 👤 **buhari** 1 year ago

ans: Domain and Fabric Admin portal ( to create the domain.)

Onelake hub is to view those domains and underlying artifacts.

upvoted 2 times

🗨️ 👤 **Ozyoluebube** 3 months ago

Fabric Admin Portal allows Fabric Admins to create and manage domains.

In the Domains tab of the Fabric Admin Portal, admins can:

Create and edit domains.

Assign domain admins and contributors.

Associate workspaces with domains.

Incorrect Options:

OneLake Data Hub – Used to view and filter data by domain, but not to create or manage domains.

Microsoft Entra Admin Portal – Manages user identity and permissions, not Fabric domains.

upvoted 1 times

🗨️ 👤 **Omer38** 1 year ago

grouping is done via Data Hub not Fabric Admin Tool

upvoted 3 times

🗨️ 👤 **Refined** 1 year ago

The right answer is Domain/Fabric Admin Portal.

You can create group with Domain, and Domain is managed at the Fabric Admin Portal level.

upvoted 2 times

🗨️ 👤 **2dc6125** 1 year, 1 month ago

<https://learn.microsoft.com/en-us/fabric/governance/domains#create-a-domain>

1- Domain is the way to logically group your resources.

2- Domain is located in the Admin portal.

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO,

\*\* Domain / One Lake Data Hub \*\*

Link: <https://learn.microsoft.com/en-us/fabric/governance/domains#create-a-domain>



1. Because of this: In Fabric, a domain is a way of logically grouping together all the data in an organization that is relevant to a particular area or field.

2. Because of this:

For instance, in the OneLake data hub, users can filter content by domain in order find content that is relevant to them. In addition, some tenant-level settings for managing and governing data can be delegated to the domain level, thus allowing domain-specific configuration of those settings.

upvoted 3 times

  **dev2dev** 1 year, 1 month ago

2nd is admin portal. You can explore/navigate domains and contents from the data hub.

upvoted 2 times

  **stilferx** 1 year, 1 month ago

But, if the question is how to do this, than answer may be Fabric Admin Portal, because of that: To assign workspaces to a domain or subdomain in the admin portal, you must be a Fabric admin or a domain admin.

So, there is an ambiguity in a question

upvoted 3 times

  **rmeng** 1 year, 2 months ago

Grouping Method: Domain. This method allows for logical grouping based on department names, which supports OneLake data hub filtering as required.

Tool: The Fabric Admin Portal provides the necessary tools for managing and organizing workspaces.

upvoted 1 times

  **sraakesh95** 1 year, 4 months ago

Tenant and MS Entra ID

The grouping method should likely be "Tenant," as this would allow for the organization of workspaces within the same Power BI service tenant, adhering to the principle of least privilege by keeping all resources under the same administrative boundary and simplifying management.

The tool to manage this could be "The Microsoft Entra admin center," formerly known as the Azure Active Directory admin center, which is used for managing various aspects of Microsoft services, including user and group management within a tenant. This would align with minimizing implementation and maintenance effort, as it is a centralized management tool.

upvoted 1 times

  **SamuComqi** 1 year, 4 months ago

- Domain

- The Fabric Admin Portal

<https://learn.microsoft.com/en-us/fabric/governance/domains#configure-domain-settings>

upvoted 3 times

  **David\_Webb** 1 year, 4 months ago



Group method: Domain

Tool: The Fabric Admin Portal

The tool here should mean the tool to implement the grouping solution. Thus, the answer should be the Fabric admin portal instead.

<https://learn.microsoft.com/en-us/fabric/governance/domains#create-a-domain>

upvoted 4 times

  **Bharat** 1 year, 4 months ago

Thank you for the link. I stand corrected.

upvoted 2 times



  **praisekim** 1 year, 4 months ago

I think theseon is right

One Lake Data Hub is used to filter data based on domain, but It doesn't create the domain or group data



When you want to make a domain, you should use Admin portal

upvoted 2 times

  **Bharat** 1 year, 4 months ago

One Lake Data Hub is the right answer according to the documentation link that you have provided.

upvoted 1 times

  **rmeng** 1 year, 4 months ago

Domain: Group the Research division workspaces based on their departmental context.

One Lake Data Hub: Use One Lake Data Hub to filter and organize the Research division workspaces effectively.

Explanation:

Domain allows you to group workspaces based on their purpose or business context. In this case, grouping by department (Research division) aligns with the requirement.

One Lake Data Hub provides the necessary filtering capabilities for organizing workspaces based on department names.

upvoted 1 times

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

### Existing Environment -

#### Identity Environment -

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

#### Data Environment -

Contoso has the following data environment:

The Sales division uses a Microsoft Power BI Premium capacity.

The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.

The Research department uses an on-premises, third-party data warehousing product.

Fabric is enabled for contoso.com.

An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.

A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

### Requirements -

#### Planned Changes -

Contoso plans to make the following changes:

Enable support for Fabric in the Power BI Premium capacity used by the Sales division.

Make all the data for the Sales division and the Research division available in Fabric.

For the Research division, create two Fabric workspaces named Productline1ws and Productline2ws.

In Productline1ws, create a lakehouse named Lakehouse1.

In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

#### Data Analytics Requirements -

Contoso identifies the following data analytics requirements:

All the workspaces for the Sales division and the Research division must support all Fabric experiences.

The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.

The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.

For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and

shortcuts by using SQL endpoints.

For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.

All the semantic models and reports for the Research division must use version control that supports branching.

#### Data Preparation Requirements -

Contoso identifies the following data preparation requirements:

The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.

All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

#### Semantic Model Requirements -

Contoso identifies the following requirements for implementing and managing semantic models:

The number of rows added to the Orders table during refreshes must be minimized.

The semantic models in the Research division workspaces must use Direct Lake mode.

#### General Requirements -

Contoso identifies the following high-level requirements that must be considered for all solutions:

Follow the principle of least privilege when applicable.

Minimize implementation and maintenance effort when possible.

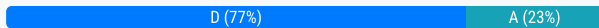
You need to refresh the Orders table of the Online Sales department. The solution must meet the semantic model requirements.

What should you include in the solution?

- A. an Azure Data Factory pipeline that executes a Stored procedure activity to retrieve the maximum value of the OrderID column in the destination lakehouse
- B. an Azure Data Factory pipeline that executes a Stored procedure activity to retrieve the minimum value of the OrderID column in the destination lakehouse
- C. an Azure Data Factory pipeline that executes a dataflow to retrieve the minimum value of the OrderID column in the destination lakehouse
- D. an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse

#### Suggested Answer: D

Community vote distribution



**theseon** 1 year, 4 months ago

**Selected Answer: D**

we need to retrieve the maximum OrderID in the destination table to minimize the number of rows added during refresh. this would be an incremental load. can be done with data flows

upvoted 27 times

**AsitTrivedi** 1 year, 3 months ago

<https://learn.microsoft.com/en-au/fabric/data-factory/tutorial-setup-incremental-refresh-with-dataflows-gen2>

upvoted 5 times

**sraakesh95** 1 year, 4 months ago

Totally agree on the max value to be retrieved on incremental load

upvoted 3 times

**Jons123son** 1 year, 1 month ago

**Selected Answer: D**

D - As other people pointed out, the exact same use case for retrieving the max OrderID is showcased in the documentation

<https://learn.microsoft.com/en-us/fabric/data-factory/tutorial-setup-incremental-refresh-with-dataflows-gen2#add-a-query-to-the-dataflow-to-filter-the-data-based-on-the-data-destination>

Thought at first that A would be correct because SP support least privilege and because how real incremental refresh is not yet supported in data flow gen 2 <https://ideas.fabric.microsoft.com/ideas/idea/?ideaid=4814b098-efff-ed11-a81c-6045bdb98602>

upvoted 9 times

🗄️ 👤 **Ce\_\_\_\_\_** Most Recent 1 week, 5 days ago

**Selected Answer: D**

While option A, can also be considered. Option D is the best choice given that it requires minimal development and maintenance  
upvoted 1 times

🗄️ 👤 **Egocentric** 5 months, 2 weeks ago

**Selected Answer: D**

key word minimize maintenance effort. answer is D  
upvoted 2 times

🗄️ 👤 **NRezgui** 6 months, 1 week ago

**Selected Answer: D**

an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse  
upvoted 1 times

🗄️ 👤 **NRezgui** 6 months, 1 week ago

**Selected Answer: D**

an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse  
upvoted 1 times

🗄️ 👤 **NRezgui** 6 months, 1 week ago

**Selected Answer: D**

an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse  
upvoted 1 times

🗄️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: D**

an Azure Data Factory pipeline that executes a dataflow to retrieve the maximum value of the OrderID column in the destination lakehouse

<https://learn.microsoft.com/en-au/fabric/data-factory/tutorial-setup-incremental-refresh-with-dataflows-gen2>

upvoted 1 times

🗄️ 👤 **Naqib** 7 months, 3 weeks ago

Both dataflow and SP should work is it? This question a bit confusing.  
upvoted 1 times

🗄️ 👤 **semauni** 8 months ago

**Selected Answer: D**

I'm also choosing D alongside the other answers. My reasoning is:

- 1) The showcased example of doing incremental refresh by dataflows (see the link below), which is almost an answer in itself because it tells you how Microsoft views the solution to this issue.
- 2) Maximum ID instead of minimum: see the same link for the specific use. But even without this knowledge you can read in the case study that new (higher) numbers represent newer orders, so for an incremental refresh it makes way more sense to retrieve the ID of the \*latest\* order placed than the ID of the first.
- 3) dataflow instead of stored procedure: because of the link, but it also makes sense from the "minimize implementation and maintenance effort" requirement: writing an incremental refresh SP is very, very complicated.

Link: <https://learn.microsoft.com/en-au/fabric/data-factory/tutorial-setup-incremental-refresh-with-dataflows-gen2>

upvoted 2 times

🗄️ 👤 **Egocentric** 8 months, 1 week ago

D is the answer also A can be correct  
upvoted 1 times

🗄️ 👤 **AzurePart** 9 months, 1 week ago

D

<https://learn.microsoft.com/en-au/fabric/data-factory/tutorial-setup-incremental-refresh-with-dataflows-gen2>

-> "You now have a query that returns the maximum OrderID in the lakehouse. This query is used to filter the data from the OData source. The next section adds a query to the dataflow to filter the data from the OData source based on the maximum OrderID in the lakehouse."

Don't ask why

The problem is Fabric, so find the answer in the document

Is this the first time you've seen a test in your life?

upvoted 1 times

🗨️ 👤 **LasAnsias** 9 months, 1 week ago

**Selected Answer: A**

Azure Data Factory "pipelines" is different from Azure Data Factory "Data Flows".

All the options are directing us to use Azure Data Factory "pipelines", so it should be using a stored procedure.

upvoted 2 times

🗨️ 👤 **semauni** 8 months ago

How does this piece of information impacts your answer? Because a pipeline is just the trigger for an activity to happen - which can either be a Stored Procedure activity or a dataflow activity. What limitation do pipelines have for dataflows in this regard?

upvoted 1 times

🗨️ 👤 **sepiida** 9 months, 1 week ago

**Selected Answer: A**

we need to retrieve the maximum OrderID in the destination table to minimize the number of rows added during refresh. This can be achieved with both the dataflow and a stored procedure.

It mentions that "All the semantic models and reports for the Research division must use version control that supports branching."

Dataflows are not supported in the git integration. Hence I choose A as the answer.

upvoted 1 times

🗨️ 👤 **Nefirs** 1 year, 2 months ago

only semantic model and reports must use version control. However, dataflows are not mentioned, therefore, irrelevant whether supported or not.

upvoted 2 times

🗨️ 👤 **nyoike** 10 months, 3 weeks ago

**Selected Answer: D**

I was initially leaning to A but got real confused when I read the choices again. Using FABRIC data factory (one would presume that what they would mean in a FABRIC exam), when you use a Stored Procedure activity, you only see Warehouses and other SQL sources and NOT Lakehouses. Using Azure Data Factory, one could add an Azure SQL DB linked service and connect to the SQL Endpoint of a Lakehouse and execute a stored procedure associated with that SQL Endpoint. Even for Fabric Pipelines, one could use an Azure SQL Database connection (instead of Lakehouse), connect to the SQL Endpoint of a Lakehouse and execute a stored procedure associated with that SQL Endpoint. This I believe is the most efficient way to do it. The issue I have with D is the fact that Dataflows require significant resources to spin up and execute. Good thing with it is that there is no ambiguity mentioned above and if you want to get the answer right, might not be the most efficient but without more verbosity in the choices, I painfully chose it.

upvoted 2 times

🗨️ 👤 **semauni** 8 months ago

And that's how you pass Microsoft exams, by (painfully) choosing the Microsoft way :) It's never 100% about what is 'the best' answer according to you or the community, it's about the answer that Microsoft will count as right.

In all seriousness: the general requirements do state that "implementation and maintenance efforts" should be minimized where possible. Writing a stored procedure to incrementally load a table gets very complex very fast.

upvoted 1 times

🗨️ 👤 **agente232** 1 year ago

I asked chatgpt and I've got this: Based on the requirements for the semantic model of the Online Sales department, the best solution to refresh the Orders table would be to include an Azure Data Factory pipeline that executes a Stored procedure activity to retrieve the maximum value of the OrderID column in the destination lakehouse. This approach ensures that only new orders are processed, maintaining the sequence and integrity of the OrderID values as per the system of origin. Therefore, the correct answer is:

A. an Azure Data Factory pipeline that executes a Stored procedure activity to retrieve the maximum value of the OrderID column in the destination lakehouse.

upvoted 1 times

🗨️ 👤 **ca63a55** 1 year ago

IMHO, it has to be done with dataflow (D) because the semantic model uses an Import mode so I think it doesn't support a store procedure (SQL)

upvoted 3 times

#### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

#### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

#### Overview -

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

#### Existing Environment -

##### Identity Environment -

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

##### Data Environment -

Contoso has the following data environment:

The Sales division uses a Microsoft Power BI Premium capacity.

The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.

The Research department uses an on-premises, third-party data warehousing product.

Fabric is enabled for contoso.com.

An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.

A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

#### Requirements -

##### Planned Changes -

Contoso plans to make the following changes:

Enable support for Fabric in the Power BI Premium capacity used by the Sales division.

Make all the data for the Sales division and the Research division available in Fabric.

For the Research division, create two Fabric workspaces named Productline1ws and Productline2ws.

In Productline1ws, create a lakehouse named Lakehouse1.

In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

##### Data Analytics Requirements -

Contoso identifies the following data analytics requirements:

All the workspaces for the Sales division and the Research division must support all Fabric experiences.

The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.

The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.

For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and

shortcuts by using SQL endpoints.

For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.

All the semantic models and reports for the Research division must use version control that supports branching.

#### Data Preparation Requirements -

Contoso identifies the following data preparation requirements:

The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.

All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

#### Semantic Model Requirements -

Contoso identifies the following requirements for implementing and managing semantic models:

The number of rows added to the Orders table during refreshes must be minimized.

The semantic models in the Research division workspaces must use Direct Lake mode.

#### General Requirements -

Contoso identifies the following high-level requirements that must be considered for all solutions:

Follow the principle of least privilege when applicable.

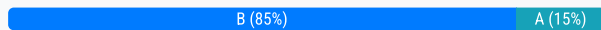
Minimize implementation and maintenance effort when possible.

Which syntax should you use in a notebook to access the Research division data for Productline1?

- A. `spark.read.format("delta").load("Tables/productline1/ResearchProduct")`
- B. `spark.sql("SELECT * FROM Lakehouse1.ResearchProduct ")`
- C. `external_table("Tables/ResearchProduct")`
- D. `external_table(ResearchProduct)`

**Suggested Answer: A**

Community vote distribution



**David\_Webb** Highly Voted 1 year, 4 months ago

**Selected Answer: B**

The correct answer is B.

The folder hierarchy of Tables in Lakehouse is incorrect for A.

upvoted 15 times

**a998450** 1 year, 3 months ago

Yes B is correct answer as per the requirement specified in case study - For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.

upvoted 6 times

**wispa2001** Highly Voted 9 months, 1 week ago

**Selected Answer: B**

```
df = spark.read.format("delta").load("Tables/MyShortcut")
```

```
display(df)
```

OR

```
df = spark.sql("SELECT * FROM MyLakehouse.MyShortcut LIMIT 1000")
```

```
display(df)
```

<https://learn.microsoft.com/en-us/fabric/onelake/onelake-shortcuts>

upvoted 10 times

**Ce\_\_\_\_\_** Most Recent 1 week, 5 days ago

**Selected Answer: B**

B is a valid answer

upvoted 1 times

**Atiwari95** 3 months ago

**Selected Answer: B**



The Correct Answer is B.

The hierarchy of table is correct in B option.

upvoted 1 times

🗨️ 👤 **NRezgui** 6 months, 1 week ago

**Selected Answer: B**

spark.sql("SELECT \* FROM Lakehouse1.ResearchProduct ")

upvoted 1 times

🗨️ 👤 **NRezgui** 6 months, 1 week ago

**Selected Answer: B**

spark.sql("SELECT \* FROM Lakehouse1.ResearchProduct ")

upvoted 1 times

🗨️ 👤 **MultiCloudIronMan** 6 months, 3 weeks ago

**Selected Answer: A**

Based on the case study details, the data for Productline1 is stored in an Azure Data Lake Storage Gen2 storage account named storage1 in the delta format. A shortcut to this storage, named ResearchProduct, is created in Lakehouse1 within the Productline1ws workspace.

Given this structure, the path "Tables/productline1/ResearchProduct" is justified because it references the shortcut created in Lakehouse1 that points to the data stored in storage1. This path aligns with the case study's description of the data environment and planned changes.

upvoted 2 times

🗨️ 👤 **Chandler9714** 3 months, 3 weeks ago

"In Lakehouse1, create a shortcut to storage1 named ResearchProduct."

The shortcut is to storage 1 not productline1. If the shortcut was specific to productline1 and placed in the folder, I think this would then be correct, but thats not the case here.

upvoted 1 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: B**

B-->spark.sql("SELECT \* FROM Lakehouse1.ResearchProduct ")

upvoted 1 times

🗨️ 👤 **jass007\_k** 8 months, 1 week ago

The correct answer is B)

Though A also looks correct, but the path mentioned is incorrect. The path should be Tables/ResearchProduct. We are directly creating a shortcut with the name ResearchProduct under Tables folder in Lakehouse1. There is no mention of the productline1 folder created.

upvoted 3 times

🗨️ 👤 **jass007\_k** 8 months, 1 week ago

Both seem to be correct option A and option B. I have tried both syntaxes with shortcut data. Also its mentioned that format of data is in delta so I will go with option A)

upvoted 2 times

🗨️ 👤 **Egocentric** 8 months, 1 week ago

A is for when you want to load data.

Answer is B, its only when you requesting specific data from specific table

upvoted 3 times

🗨️ 👤 **Egocentric** 10 months, 1 week ago

answer is A. in B there is no productline1

upvoted 2 times

🗨️ 👤 **nyoike** 10 months, 3 weeks ago

**Selected Answer: A**

With the recent introduction of schema-enabled Lakehouses, BOTH A and B are correct. That is assuming ResearchProduct table was created in a schema-enabled Lakehouse in the productline1 schema. I have tested A in a Fabric Spark notebook that is schema-enabled and it works.

upvoted 6 times

🗨️ 👤 **Miro\_dd** 12 months ago

**Selected Answer: B**

Hierarchy for answer A is not correct

upvoted 1 times

🗨️ 👤 **ziggy1117** 1 year ago

**Selected Answer: B**

<https://learn.microsoft.com/en-us/fabric/onelake/onelake-shortcuts>

```
df = spark.read.format("delta").load("Tables/MyShortcut")
```

OR

```
df = spark.sql("SELECT * FROM MyLakehouse.MyShortcut LIMIT 1000")
```

upvoted 1 times

🗨️ 👤 **ca63a55** 1 year, 1 month ago

If answer B is the correct, why Files folder doesn't appear between Lakehouse1 and ResearchProduct? It has to be like "lakehouse1.Files.ResearchProduct", hasn't it?

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO,

I would go with A, because of there is a clear statement ProductLine1.

Technically, A and B should work, but B doesn't have "ProductLine", what is confusing

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

My bad, the right answer is B.

There is a requirement: Research Division - all tables should be managed. It means, no subfolders should be.

<https://learn.microsoft.com/en-us/training/modules/use-apache-spark-work-files-lakehouse/5-spark-sql>

upvoted 5 times

**HOTSPOT -****Case study -**

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

**To start the case study -**

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

**Overview -**

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

**Existing Environment -****Fabric Environment -**

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

**Available Data -**

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

**Survey -****Question -****Response -**

For each survey submitted, the following occurs:

One row is added to the Survey table.

One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.

Customers can submit a survey after each purchase.

**User Problems -**

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic

models, but the logic does NOT always match across implementations.

#### Requirements -

##### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store

DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake

DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

A data store (type to be decided)

A custom semantic model -

A default semantic model -

##### Interactive reports -

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

##### Technical Requirements -

The data store must support the following:

Read access by using T-SQL or Python

Semi-structured and unstructured data

Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

List prices that are less than or equal to 50 are in the low pricing group.

List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.

List prices that are greater than 1,000 are in the high pricing group.

##### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

Fabric administrators will be the workspace administrators.

The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.

The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.

The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook

The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.

The date dimension must be available to all users of the data store.

The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

FabricAdmins: Fabric administrators

AnalyticsTeam: All the members of the analytics team  
DataAnalysts: The data analysts on the analytics team  
DataScientists: The data scientists on the analytics team  
DataEngineers: The data engineers on the analytics team  
AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

Enables a user to select a product to filter customer survey responses to only those who have purchased that product.

Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.

Shows data as soon as the data is updated in the data store.

Ensures that the report and the semantic model only contain data from the current and previous year.

Ensures that the report respects any table-level security specified in the source data store.

Minimizes the execution time of report queries.

You need to assign permissions for the data store in the AnalyticsPOC workspace. The solution must meet the security requirements.

Which additional permissions should you assign when you share the data store? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

#### Answer Area

DataEngineers: ☐ Build Reports on the default dataset  
☐ Read All Apache Spark  
☐ Read All SQL analytics endpoint data

DataAnalysts: ☐ Build Reports on the default dataset  
☐ Read All Apache Spark  
☐ Read All SQL analytics endpoint data

DataScientists: ☐ Build Reports on the default dataset  
☐ Read All Apache Spark  
☐ Read All SQL analytics endpoint data

#### Answer Area

Suggested Answer: DataEngineers: ☐ Build Reports on the default dataset  
☐ Read All Apache Spark  
☐ Read All SQL analytics endpoint data

DataAnalysts: ☐ Build Reports on the default dataset  
☐ Read All Apache Spark  
☐ Read All SQL analytics endpoint data

DataScientists: ☐ Build Reports on the default dataset  
☐ Read All Apache Spark  
☐ Read All SQL analytics endpoint data

 **Bharat**  1 year, 4 months ago

Here is my take on it:

Data Engineers: Read all Apache Spark - because they need to be able to work with Spark for Data curation.

Data Analysts: Build Reports on the default dataset - because they are report builders

Data Scientists: Read SQL Endpoints - They leverage curated data (by engineers) to do predictive analytics.

Let me know what you think.

upvoted 89 times

 **scorradi** 5 months, 3 weeks ago



My guest based in the question description:

Data analysts --> Build Reports on the default dataset.

Data Scientist -> Spark, because they need to use Notebooks as the question says.

Data Engineers -> Spark, because the need to write data.

upvoted 2 times

  **Training\_be2** 6 months, 2 weeks ago

Data analysts --> Build Reports on the default dataset

Data Scientist will access the data through a Notebook --> Read all Apache Spark

Data Engineers will read and write to the data store (lakehouse) --> Read all Apache Spark

SQL Analytics Endpoint is read-only mode. You can analyze Delta tables using T-SQL (instead of Spark SQL), save functions, or create views.

However, you cannot write data to a Delta table.

upvoted 3 times

  **7d97b62** 1 year ago

I think Data Analysts Building Reports on the default dataset is correct

upvoted 1 times

  **janineh** 1 year ago

it says the data scientist need to use spark - so spark for them in my opinion

upvoted 5 times

  **vissu\_settipally**  1 year, 4 months ago

Data Engineers = Read sql endpoints

Data Analyst = Build Reports



Data Scientist = They prefer Spark Always

upvoted 38 times

  **semauni** 8 months ago

agreed for the analyst and scientist, but the engineers also need spark access since they will be making notebooks

upvoted 1 times

  **SVCDIA** 4 months, 3 weeks ago


engineers will be making notebooks in DataEngPOC workspace but not in AnalyticsPOC workspace. So they don't need to create any in that

upvoted 1 times

  **zxc01**  2 months ago

my question is does Read all Apache Spark can write data in lakehouse? The data engineers must be able to read from and write to the data store. I checked many documents but they just said Read all Apache Spark can read data by spark notebook or the Lakehouse explorer.

upvoted 1 times

  **Rakesh16** 7 months, 2 weeks ago

Data Engineers-->Read all Apache spark

Data Analysts-->Build reports on the default dataset

Data Scientists-->Read all Apache spark

upvoted 12 times

  **Richdata23** 7 months, 3 weeks ago

DataEngineers: "Read All Apache Spark"

Explanation: To load and manage data in OneLake, data engineers need the ability to read and write data in Spark-compatible formats, which is ideal for handling semi-structured and unstructured data.

DataAnalysts: "Build Reports on the default dataset"

Explanation: Data analysts need to create Power BI reports using semantic models. This permission allows them to interact with the necessary datasets without granting access to perform data transformations.

DataScientists: "Read All Apache Spark"

Explanation: Data scientists will use Spark notebooks to access and analyze data. By granting "Read All Apache Spark," they can directly interact with the data in their Spark environment, which is optimized for their analysis needs.

upvoted 8 times

  **Naqib** 7 months, 3 weeks ago

DataEngineers: ReadAll Apache Spark  
DataAnalyst: Build Reports on the default dataset  
DataScientist: ReadAll Apache Spark

The DS and DE need to have access to Apache Spark otherwise they wont able to work with the data transformation/curation ect.  
upvoted 3 times

🗨️ 👤 **Ahmadpbi** 11 months, 1 week ago  
as the provided data (The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.), according to this It would not be possible for them to build report using the default dataset.  
upvoted 1 times

🗨️ 👤 **vish9** 1 year, 1 month ago  
Data Analysts – Build Reports  
Data Scientists – Read All Apache Spark.  
The confusion is about the Data Engineers: They must be able to read and write to the data store. The SQL end point is read only. They should not have build reports. Hence the remaining option is read all Apache spark. Hence in this question no one gets access to the SQL Endpoint.  
upvoted 10 times

🗨️ 👤 **2dc6125** 1 year, 1 month ago  
i think this is part of a solutio of the security requirement and data engineer cannot perform write action with any of these permission so i think  
Data Engineers: Read SQL Endpoints  
Data Analysts: Build Reports on the default dataset  
Data Scientists: Read all Apache Spark  
upvoted 7 times

🗨️ 👤 **stilferx** 1 year, 1 month ago  
IMHO,  
- DEs - Spark,  
- Analysts - Build Report,  
- DS - Spark (as well).

Why?

1. Here is the description of the options: <https://blog.fabric.microsoft.com/en-us/blog/data-warehouse-sharing/>

2. Here is the role description in the question itself:  
The data engineers must be able to !!!read from and write!!! to the data store. No access must be granted to datasets or reports.  
The data scientists must be able to read from the data store, but not write to it. They will access the data by using a !!! Spark notebook  
The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by !!! using the semantic models created by the analytics engineers.  
The date dimension must be available to all users of the data store.  
upvoted 14 times

🗨️ 👤 **rmeng** 1 year, 2 months ago  
DataEngineers: ReadAll Apache Spark  
DataAnalyst: Build Reports on the default dataset  
DataScientist: ReadAll Apache Spark  
upvoted 9 times

🗨️ 👤 **CertPeople** 1 year, 3 months ago  
"The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook" so for data scientists we have to give Read All Apache Spark  
upvoted 5 times

🗨️ 👤 **a\_51** 1 year, 3 months ago  
Feeding the question to ChatGPT it says given the option of A, B and C respectively:  
Data Engineers: Option B: Read All Apache Spark (Allows reading data from Apache Spark notebooks)  
Option C: Read All SQL analytics endpoint data (Allows reading data from SQL analytics endpoint)  
  
Data Scientists: Option B: Read All Apache Spark (Allows reading data from Apache Spark notebooks)  
Data Analysts: Option A: Build Reports on the default dataset (Allows creating reports)

upvoted 2 times

🗨️ 👤 **dp600** 1 year, 2 months ago  
stop using chatgpt as a source, it's not trustworthy  
upvoted 11 times

🗨️ 👤 **thuss** 1 year, 4 months ago  
My take: the SQL Analytics Endpoint is read-only, so it's perfect for the DataScientist. DataEngineers need read and write, so Spark. The Analysts need the report capabilities obviously.  
upvoted 3 times

🗨️ 👤 **momo1165** 1 year ago  
SQL Endpoints do not support Spark Notebooks  
upvoted 1 times

🗨️ 👤 **PazaBlandData** 1 year, 2 months ago  
All of these are read-only. It confuses me  
upvoted 4 times

🗨️ 👤 **David\_Webb** 1 year, 4 months ago  
DataEngineers: ReadAll Apache Spark  
DataAnalyst: Build Reports on the default dataset  
DataScientist: ReadAll Apache Spark

Data engineers will use Pyspark in notebooks to transform data from the Files folder in the Lakehouse.  
Data analysts will build reports and dashboards from the prepared dataset.  
Data scientists will use MLlib in notebooks to build models.  
The "Read All SQL analytics endpoint data" should be for the AnalyticsEngineers. The analytics team has four types of members.  
upvoted 10 times

🗨️ 👤 **metiii** 1 year, 3 months ago  
This is about access in AnalyticsPOC, Data Scientists don't need to access Apache Spark in this workspace they should only be able to read from SQL endpoint, they will create Spark notebooks in their own workspace and this question is not concerned about that workspace.  
upvoted 1 times

🗨️ 👤 **c8f5bdf** 1 year, 1 month ago  
actually it says that data scientists will use notebooks so they need ReadAll Apache Spark  
upvoted 1 times

🗨️ 👤 **wojciech\_wie** 1 year, 4 months ago  
Data Engineers = ReadAll Apache Spark  
Data Analyst = Build Reports  
Data Scientist = ReadAll Apache Spark  
upvoted 5 times

🗨️ 👤 **SamuComqi** 1 year, 4 months ago  
Data Engineers: Read all SQL Analytics Endpoint data (use SQL to explore and create/modify tables, views, stored procedures).  
Data Analysts: Build reports on the default dataset (using Power BI).  
Data Scientists: Read all Apache Spark (they will use Notebooks to analyze data and apply ML models).  
upvoted 4 times



## HOTSPOT -

## Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview -

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

## Existing Environment -

## Fabric Environment -

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

## Available Data -

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

## Survey -

## Question -

## Response -

For each survey submitted, the following occurs:

One row is added to the Survey table.

One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.

Customers can submit a survey after each purchase.

## User Problems -

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic

models, but the logic does NOT always match across implementations.

#### Requirements -

##### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store

DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake

DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

A data store (type to be decided)

A custom semantic model -

A default semantic model -

##### Interactive reports -

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

##### Technical Requirements -

The data store must support the following:

Read access by using T-SQL or Python

Semi-structured and unstructured data

Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

List prices that are less than or equal to 50 are in the low pricing group.

List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.

List prices that are greater than 1,000 are in the high pricing group.

##### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

Fabric administrators will be the workspace administrators.

The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.

The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.

The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook

The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.

The date dimension must be available to all users of the data store.

The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

FabricAdmins: Fabric administrators

AnalyticsTeam: All the members of the analytics team  
DataAnalysts: The data analysts on the analytics team  
DataScientists: The data scientists on the analytics team  
DataEngineers: The data engineers on the analytics team  
AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

Enables a user to select a product to filter customer survey responses to only those who have purchased that product.

Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.

Shows data as soon as the data is updated in the data store.

Ensures that the report and the semantic model only contain data from the current and previous year.

Ensures that the report respects any table-level security specified in the source data store.

Minimizes the execution time of report queries.

You need to create a DAX measure to calculate the average overall satisfaction score.

How should you complete the DAX code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

#### Answer Area

```
Rolling 12 Overall Satisfaction =  
VAR NumberOfMonths = 12  
VAR LastCurrentDate = MAX ( 'Date' [Date] )  
VAR Period = DATESINPERIOD ( 'Date' [Date], LastCurrentDate,  
    - NumberOfMonths, MONTH )  
VAR Result =  
    CALCULATE (  
        AVERAGE('Survey'[Response Value]),  
        AVERAGEA('Question'[Question Text]),  
        AVERAGEX(VALUE('Survey'[Customer Key]),  
            LastCurrentDate,  
            NumberOfMonths,  
            Period,  
            'Survey Question' [Question Title] = "Overall Satisfaction"  
    )  
RETURN  
    Result
```

#### Answer Area

```
Rolling 12 Overall Satisfaction =  
VAR NumberOfMonths = 12  
VAR LastCurrentDate = MAX ( 'Date' [Date] )  
VAR Period = DATESINPERIOD ( 'Date' [Date], LastCurrentDate,  
    - NumberOfMonths, MONTH )  
VAR Result =  
    CALCULATE (  
        AVERAGE('Survey'[Response Value]),  
        AVERAGEA('Question'[Question Text]),  
        AVERAGEX(VALUE('Survey'[Customer Key]),  
            LastCurrentDate,  
            NumberOfMonths,  
            Period,  
            'Survey Question' [Question Title] = "Overall Satisfaction"  
    )  
RETURN  
    Result
```

Suggested Answer:

I would say:

AVERAGE('Survey'[Response Value]) and Period



I do not understand why we would use the CustomerKey for Average Calculation.

upvoted 102 times

  **a\_51** Highly Voted 1 year, 3 months ago

Average('Survey'[Response Value]) and Period

upvoted 10 times

  **a\_51** 1 year, 3 months ago

Response Value is the only one that even likely could be aggregated. The question text and customer key do not make sense.

upvoted 6 times

  **rohitrc8521** Most Recent 4 months ago


Calm down guys, the correct answer is "Average" and "Period"

upvoted 2 times

  **Rakesh16** 7 months, 2 weeks ago

Average('Survey'[Response Value]) & Period is the answer

upvoted 3 times

  **jass007\_k** 8 months, 1 week ago

This will be the final answer

Rolling 12 Overall Satisfaction =

VAR NumberOfMonths = 12

VAR LastCurrentDate = MAX('Date'[Date])

VAR Period = DATESINPERIOD(

'Date'[Date],

LastCurrentDate,

-NumberOfMonths,

MONTH

)

VAR Result =

CALCULATE (

AVERAGE('Survey'[Response Value]),

Period,

'Survey Question'[Question Title] = "Overall Satisfaction"

)

RETURN

Result

The DATESINPERIOD function calculates a dynamic period starting from the most recent date (LastCurrentDate) and going back NumberOfMonths months. It provides a rolling 12-month window to calculate averages within this range.

upvoted 2 times

  **VAzureD** 9 months, 1 week ago

AVERAGE('Survey'[Response Value]),

Period.

AVERAGEA('Question'[Question Text]),

It doesn't make much sense to average a text field when you have to extract a value from it, although AVERAGEA can do it.

AVERAGEX(VALUES('Survey'[Customer Key]),

Syntax is wrong, discarded. AVERAGEX(<table>, <expression>)

Additionally VALUES leaves only the distinct values

LastCurrentDate,

In this case we would have only half a day. Has no sense.

NumberOfMonths,  
We would be filtering by 12. It doesn't make sense.

Period,  
In the variable we have the dates of the last 12 months.  
upvoted 7 times

🗳️ 👤 **282b85d** 9 months, 1 week ago

Rolling 12 Overall Satisfaction =  
VAR NumberOfMonths = 12  
VAR LastCurrentDate = MAX ( 'Date' [Date] )  
VAR Period = DATESINPERIOD ( 'Date' [Date], LastCurrentDate, -NumberOfMonths, MONTH )  
VAR Result =  
CALCULATE (  
AVERAGEX (  
VALUES('Survey'[Customer Key]),  
CALCULATE(  
AVERAGE('Survey'[Response Value])  
)  
),  
Period,  
'Survey Question'[Question Title] = "Overall Satisfaction"  
)  
RETURN  
Result  
\*\*There should be some missing part in this screenshot, but the correct answer would be the VALUES('Survey'[Customer Key]) with the missing part  
CALCULATE(AVERAGE('Survey'[Response Value]))  
upvoted 1 times

🗳️ 👤 **Darshan6232** 1 year ago

It should be :  
1. Average(Survey[Response Value]) as this one is the only value that can be aggregated.  
Customer Key Does not make any sense , as it just a key to identify a customer, this can never provide us the score.  
2. Period : Variable is defined as to select 1 year of date range. Can directly be passed in the Filter context of Calculate formula. Thanks!  
upvoted 2 times

🗳️ 👤 **2dc6125** 1 year, 1 month ago

1 = AVERAGE('Survey'[Response Value]). customer satisfaction can be determined by response value.  
2= Period. choose this just because it state overall no specific time or requires a date range.  
upvoted 1 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

IMHO,  
AVERAGE('Survey'[Response Value]) and Period

Why?  
The first one defines actually the metric, which is avg from survey(response value),  
the second one defines the date range, which is precalculated before as a variable Period.  
upvoted 2 times

🗳️ 👤 **rmeng** 1 year, 2 months ago

AVERAGE('Survey'[Response Value]) , Period  
upvoted 1 times

🗳️ 👤 **SamuComqi** 1 year, 4 months ago

For the first part, "AVERAGE('Survey'[Response Value])" because the second option uses a text column as argument, and the third option is not relevant in this context (no need to perform row calculations).  
For the second part, "Period": data is filtered to compute the average in the last 12 months (interval defined in the variable).  
upvoted 6 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

Average and Period

upvoted 5 times

  **Nicknamefordiscussions69** 1 year, 4 months ago

Response Value and Period

upvoted 4 times

**HOTSPOT -****Case study -**

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

**To start the case study -**

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

**Overview -**

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

**Existing Environment -****Fabric Environment -**

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

**Available Data -**

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

**Survey -****Question -****Response -**

For each survey submitted, the following occurs:

One row is added to the Survey table.

One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.

Customers can submit a survey after each purchase.

**User Problems -**

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic

models, but the logic does NOT always match across implementations.

#### Requirements -

##### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store

DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake

DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

A data store (type to be decided)

A custom semantic model -

A default semantic model -

##### Interactive reports -

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

##### Technical Requirements -

The data store must support the following:

Read access by using T-SQL or Python

Semi-structured and unstructured data

Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

List prices that are less than or equal to 50 are in the low pricing group.

List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.

List prices that are greater than 1,000 are in the high pricing group.

##### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

Fabric administrators will be the workspace administrators.

The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.

The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.

The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook

The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.

The date dimension must be available to all users of the data store.

The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

FabricAdmins: Fabric administrators



AnalyticsTeam: All the members of the analytics team  
DataAnalysts: The data analysts on the analytics team  
DataScientists: The data scientists on the analytics team  
DataEngineers: The data engineers on the analytics team  
AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

Enables a user to select a product to filter customer survey responses to only those who have purchased that product.

Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.

Shows data as soon as the data is updated in the data store.

Ensures that the report and the semantic model only contain data from the current and previous year.

Ensures that the report respects any table-level security specified in the source data store.

Minimizes the execution time of report queries.

You need to resolve the issue with the pricing group classification.

How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

#### Answer Area

CREATE  [dbo].[ProductsWithPricingGroup]

FUNCTION  
PROCEDURE  
TABLE  
VIEW

AS

SELECT ProductId,  
ProductName,  
ProductCategory,  
ListPrice,

CASE  
COALESCE  
IIF  
SET

WHEN ListPrice <= 50 THEN 'low'

WHEN(ListPrice > 50 AND ListPrice < 1000) THEN 'medium'  
WHEN(ListPrice >= 50 AND ListPrice < 1000) THEN 'medium'  
WHEN ListPrice BETWEEN 50 AND 1000) THEN 'medium'

WHEN ListPrice > 1000 THEN 'high'

END AS PricingGroup

FROM dbo.Products

## Answer Area

```
CREATE [dropdown] [dbo].[ProductsWithPricingGroup]
      FUNCTION
      PROCEDURE
      TABLE
      VIEW

AS

SELECT ProductId,
       ProductName,
       ProductCategory,
       ListPrice,

Suggested Answer: [dropdown]
                  CASE
                  COALESCE
                  IIF
                  SET

                  WHEN ListPrice <= 50 THEN 'low'

                  [dropdown]
                  WHEN(ListPrice > 50 AND ListPrice < 1000) THEN 'medium'
                  WHEN(ListPrice >= 50 AND ListPrice < 1000) THEN 'medium'
                  WHEN ListPrice BETWEEN 50 AND 1000) THEN 'medium'
                  WHEN ListPrice > 1000 THEN 'high'
                  END AS PricingGroup
FROM dbo.Products
```

**SamuComqi** Highly Voted 1 year, 4 months ago

\* VIEW: from an existing table.

\* CASE: correct syntax before the WHENs.

\* WHEN ListPrice BETWEEN 50 AND 1000 THEN 'medium': the other two options miss value 1000; on the other hand, the BETWEEN includes both 50 and 1000.

upvoted 63 times

**semauni** 8 months ago

I agree with your answer, but with some nuance:

- I think that you're right that it is a view, but not because it is created from an existing table. Nothing prevents you from creating a table from a table (CTAS statement). However, it should be a view because 'data should be visible as soon as the datastore is updated'. A table from a table is not automatically refreshed when its source table is updated, but a view on this source table will immediately show the new results.

- I agree that in the last option, there now is no correct option since none meet the requirements and BETWEEN is the least wrong. Either a bad question or a badly remembered question :)

upvoted 9 times

**semauni** 8 months ago

I'm however not 100% sure it shouldn't be table, because reports also need minimal query execution times. A 'SELECT \*' on a table is simple, a 'SELECT \*' on the view always executes the underlying query. Makes me wonder what 'as soon as the datastore' is updated mean: as soon as new data is available, or as soon as the Power BI source table has been updated?

upvoted 1 times

**DinoD** 3 months, 3 weeks ago

VIEW: hence create view as "Then SQL", plus its being used in a report, so its not meant to a static representation of the data at a single point in time.

CASE: correct syntax before the WHENs

WHEN: There is no correct answer, none of the solutions either exclude 50 and always include 1000. Look at the bottom of high and it says > 1000, so what catches 51 through 1000? Garbage question

upvoted 1 times

**2fe10ed** 1 year, 1 month ago



\* VIEW: from an existing table.

\* CASE: correct syntax before the WHENs.

\* WHEN(ListPrice > 50 AND ListPrice < 1000 THEN 'medium': The second option will generate 2 rows for any '50' answer; the third option has the same problem plus a miss spelled ' ) ' that will end up generating a syntax error.

The only answer that will not generate a cartesian product is the first one, although it will miss all the '1000' answers.

upvoted 7 times

  **dev2dev** 1 year, 1 month ago

wrong 3rd option. You get NULL PriceGroup when price is exactly 1000

upvoted 2 times

  **utsuha** 1 year, 1 month ago

The 3rd option is still the best--when doing a select with CASE WHEN, the logic is evaluated linearly. If the value is 50, the first case is true and the returned value is therefore "low". This can be tested with dummy data--just upload a csv file of numbers from 1 to 20 into a test warehouse and query in Fabric.

```
select number,  
case when number <= 5 then 'low'  
when number between 5 and 10 then 'med'  
when number > 10 then 'high'  
end as test  
from dbo.numbers  
order by number asc
```

As for the missing '(' or extra ')', I assume this is a typo.

upvoted 5 times

  **BennyBenz** 1 year, 2 months ago



I agree that 1000 must be included (in the medium), but if the BETWEEN includes 50, and low should include 50, and it states before that <= 50 is low, then 50 shouldn't be included... ?

upvoted 2 times

  **thuss**  1 year, 4 months ago

All options for the last one are wrong, which is irritating. Should be > 50 AND <= 1000.

upvoted 30 times

  **Tuki93** 1 year, 3 months ago

The range BETWEEN 50 AND 1000 is also a valid answer. However, if the value is 50, it will be returned from the previous 'when' clause, ListPrice <= 50 Then 'low', before reaching the second 'when' clause."

upvoted 13 times

  **Kau1977**  2 months, 1 week ago

As per the given syntax it should be Create View in the first option.

upvoted 1 times

  **ba2bdf** 3 months ago

1. None of the option in the question satisfying the requirement. Especially for ListPrice 1000


2. Option with Between having an additional ")" which will throw syntax error.

List prices that are less than or equal to 50 are in the low pricing group.

List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.

List prices that are greater than 1,000 are in the high pricing group.

upvoted 1 times

  **pirate84** 5 months ago

I took the test the last week. This question was included, with a important modification:

They changes the information,i of the first option for "WHEN": ListPrice > 50 AND ListPrice <= 1000) THEN 'medium'.

-View

-Case

-WHEN(ListPrice > 50 AND ListPrice <= 1000) THEN 'medium'.

upvoted 8 times

  **Rakesh16** 7 months, 2 weeks ago

View, CASE, WHEN ListPrice BETWEEN 50 AND 100 THEN 'medium' is the answer

upvoted 1 times

  **hjhjh123** 9 months, 2 weeks ago

view, case - no correct answer - my guess is they misseed the = on a  
upvoted 4 times

🗨️ **ca\_acc** 10 months, 3 weeks ago

\* VIEW

\* CASE

\* WHEN ListPrice Between 50 and 1000 Then Medium: The first when is includes 50, so even though the between should be from 51 to 1000, it is (in this case) irrelevant, because in this execution order the first WHEN wins. Therefore between should produce correct results.

upvoted 1 times

🗨️ **AhmadAllied** 1 year ago

The valid answer for 3rd is first one as the 1st case is When List Price <= 50 then 'low', so the next case should not be include equal to 50 in anyway, but option 2 has List Price >= 50 which is wrong and between function will also consider = 50 values so it is also wrong.

upvoted 1 times

🗨️ **kronoszerg** 1 year ago

View

To satisfy "Shows data as soon as the data is updated in the data store"

You cannot use CTAS Statement as it will not update when source data is updated.

Case

The usual SQL Case Statement

Last one is..... no correct answer

Either question is wrong or answer options are wrong.

BETWEEN is inclusive by the way

upvoted 1 times

🗨️ **calvintcy** 1 year ago

I took the test a few days ago. This question was included, but Microsoft has made a change to: WHEN(ListPrice > 50 AND ListPrice <= 1000) THEN 'medium'. Therefore, I chose this option.

upvoted 21 times

🗨️ **vahox** 1 year ago

View

CASE

WHEN(ListPrice >= 50 AND ListPrice < 1000) then as between is inclusive of 50 it will provide incorrect result

upvoted 1 times

🗨️ **Darshan6232** 1 year ago

It should be

1. Table : As it should be part of Default Semantic Model. With view it's not possible to achieve this.

2. Case Statement : As we have "When" and "Then"

3. Between 51\* and 1000 or it should be Listprice>50 and Listprice <=1000. Thanks!

upvoted 3 times

🗨️ **Priyanka007** 1 year ago

Within the warehouse, a user can add warehouse objects - tables or views to their default Power BI semantic model. I would go with VIEW as view Shows data as soon as the data is updated in the data store.<https://learn.microsoft.com/en-us/fabric/data-warehouse/default-power-bi-semantic-model>

upvoted 1 times

🗨️ **b6daab0** 1 year ago

Is the solution provided the true correct answer? In many cases it's not, which is misleading. It definitely should be a view and not a table because the Product table changes and thus the pricing group being updated. Where do the solutions come from??

upvoted 1 times

🗨️ **282b85d** 1 year, 1 month ago

CREATE VIEW [dbo].[ProductsWithPricingGroup]

AS

SELECT

ProductId,

ProductName,

```
ProductCategory,  
ListPrice,  
CASE  
WHEN ListPrice <= 50 THEN 'low'  
WHEN ListPrice > 50 AND ListPrice <= 1000 THEN 'medium'  
WHEN ListPrice > 1000 THEN 'high'  
END AS PricingGroup  
FROM dbo.Products;  
upvoted 3 times
```

  **David\_Webb** 1 year, 1 month ago

```
VIEW  
CASE  
WHEN ListPrice BETWEEN 50 AND 1000 THEN 'medium'
```

Look at the requirement:

List prices that are less than or equal to 50 -> low

List prices that are greater than 50 and less than or equal to 1,000 -> medium

List prices that are greater than 1,000 -> high

Only using option "BETWEEN 50 AND 1000" will be able to get the correct answer. If the ListPrice is 50, the case will resolve it as "low" in the first condition and return the value. Even though 50 is included in the next case "BETWEEN 50 AND 1000", it will not fall into this condition as it has already been resolved with the first condition.

upvoted 4 times

  **omsingh** 1 year, 1 month ago

```
select country,first_name, age , case when age <=25 then "child"  
when age between 25 and 28 then "adult"  
when age > 28 then "old"  
end as ages  
from customers;  
country first_name age ages  
USA John 31 old  
USA Robert 22 child  
UK David 22 child  
UK John 25 child  
UAE Betty 28 adult
```

Correct answer will be VIEW, CASE AND BETWEEN.

upvoted 2 times

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

### Existing Environment -

### Fabric Environment -

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

### Available Data -

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

### Survey -

### Question -

### Response -

For each survey submitted, the following occurs:

One row is added to the Survey table.

One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.

Customers can submit a survey after each purchase.

### User Problems -

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

## Requirements -

### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store

DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake

DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

A data store (type to be decided)

A custom semantic model -

A default semantic model -

### Interactive reports -

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

### Technical Requirements -

The data store must support the following:

Read access by using T-SQL or Python

Semi-structured and unstructured data

Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

List prices that are less than or equal to 50 are in the low pricing group.

List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.

List prices that are greater than 1,000 are in the high pricing group.

### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

Fabric administrators will be the workspace administrators.

The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.

The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.

The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook

The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.

The date dimension must be available to all users of the data store.

The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

FabricAdmins: Fabric administrators

AnalyticsTeam: All the members of the analytics team

DataAnalysts: The data analysts on the analytics team

DataScientists: The data scientists on the analytics team  
DataEngineers: The data engineers on the analytics team  
AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

Enables a user to select a product to filter customer survey responses to only those who have purchased that product.

Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.

Shows data as soon as the data is updated in the data store.

Ensures that the report and the semantic model only contain data from the current and previous year.

Ensures that the report respects any table-level security specified in the source data store.

Minimizes the execution time of report queries.

What should you recommend using to ingest the customer data into the data store in the AnalyticsPOC workspace?

- A. a stored procedure
- B. a pipeline that contains a KQL activity
- C. a Spark notebook
- D. a dataflow

**Suggested Answer: D**

Community vote distribution

D (97%)

🗳️ 👤 **VAzureD** Highly Voted 1 year, 2 months ago

**Selected Answer: D**

D. a dataflow

Es la mejor opción.

"Whenever possible, the data engineers will use low-code tools for data ingestion."

"Data will be loaded without transformation in one area of the AnalyticsPOC data store"

A. a stored procedure

No tiene sentido usar un procedimiento almacenado para hacer la carga.

B. a pipeline that contains a KQL activity

KQL es para datos en tiempo real

C. a Spark notebook

Podría valer, pero el texto pone,

"Whenever possible, the data engineers will use low-code tools for data ingestion."

upvoted 9 times

🗳️ 👤 **amar5555** 1 year, 2 months ago

i feel answer is pipeline with KQL and using kql we get real time data

upvoted 2 times

🗳️ 👤 **semauni** 8 months ago

We don't need real-time data, it is updated either hourly or daily based on the source. The instant refresh is only a reporting requirement, not an ingestion requirement. I'm \*assuming\* that the customer data would be updated daily.

upvoted 1 times

🗳️ 👤 **David\_Webb** Highly Voted 1 year, 4 months ago

**Selected Answer: D**

In the Interactive reports requirement, it stated, "Whenever possible, the data engineers will use low-code tools for data ingestion".

upvoted 9 times

🗳️ 👤 **Ce\_\_\_\_\_** Most Recent 1 week, 5 days ago

**Selected Answer: D**



D - Dataflow is Correct

upvoted 2 times

🗳️ 👤 **Amine\_spiegel94** 3 months, 4 weeks ago

**Selected Answer: D**

"Whenever possible, the data engineers will use low-code tools for data ingestion."

"Data will be loaded without transformation in one area of the AnalyticsPOC data store"

upvoted 1 times

🗳️ 👤 **7addd81** 4 months, 3 weeks ago

**Selected Answer: D**

D (Dataflow) because low-code tools are to be used. There are no additional requirements, so no need, to do real-time processing.

upvoted 1 times

🗳️ 👤 **Vulkany** 5 months, 2 weeks ago

**Selected Answer: C**

C) Spark supports real-time or near real-time ingestion of Delta Lake-compliant Parquet data.

It provides full control over data processing and transformation, ensuring data availability as soon as it's ingested.

While not "low-code," it's the most capable option for real-time processing and Delta Lake integration.

Why D) is not the option (in my opinion)

Dataflows rely on scheduled refresh, introducing a delay.

This violates the real-time data availability requirement, making it less suitable despite being low-code and user-friendly.

upvoted 2 times

🗳️ 👤 **NRezgui** 6 months, 1 week ago

**Selected Answer: D**

a dataflow

upvoted 2 times

🗳️ 👤 **Iagraoui** 7 months ago

**Selected Answer: D**

D, because they said to use low-code tools for data ingestion.

upvoted 1 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: D**

Dataflow is the answer

upvoted 1 times

🗳️ 👤 **jass007\_k** 8 months, 1 week ago

Its option D) Dataflow because dataflows are generally better in case of small and medium datasets and size of Product data is 200 MB

upvoted 1 times

🗳️ 👤 **TimoRii** 8 months, 2 weeks ago

Strange, no one had got the correct answer yet: it is Spark Notebook.

If you had 50 GB, then dataflow is fine, but premium Fabric won't manage to handle 500 GB with dataflow if you need to do any updates on the data.

And you should manage with less than premium/F64 capacity for these requirements. So the extra cost for Notebook programming will be saved many times over with lower capacity requirements.

upvoted 1 times

🗳️ 👤 **semauni** 8 months ago

But we're not loading the survey data, the question asks about the customer data. Which is 50 MB. If we were loading the survey data, I would agree that notebooks would be best ("when POSSIBLE").

upvoted 1 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: D**

IMHO,

D (dataflow), because of this:

"Whenever possible, the data engineers will use low-code tools for data ingestion."

upvoted 4 times

🗳️ 👤 **rmeng** 1 year, 2 months ago

**Selected Answer: D**

Whenever possible, the data engineers will use low-code tools for data ingestion.

upvoted 2 times

🗨️ 👤 **dp600** 1 year, 2 months ago

**Selected Answer: C**

"Shows data as soon as the data is updated in the data store"

KQL?

upvoted 1 times

🗨️ 👤 **dp600** 1 year, 2 months ago

nevermind, it specified as soon as it gets to the data store, that implies as soon as it's loaded in fabric

upvoted 3 times

🗨️ 👤 **amar5555** 1 year, 2 months ago

i feel answer is pipeline with KQL and using kql we get real time data

upvoted 1 times

🗨️ 👤 **clux** 1 year, 2 months ago

**Selected Answer: D**

correct

upvoted 1 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

**Selected Answer: D**

Dataflow is the best choice.

upvoted 2 times

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

### Existing Environment -

#### Fabric Environment -

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

#### Available Data -

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

#### Survey -

#### Question -

#### Response -

For each survey submitted, the following occurs:

One row is added to the Survey table.

One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score.

Customers can submit a survey after each purchase.

#### User Problems -

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

## Requirements -

### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store

DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake

DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

A data store (type to be decided)

A custom semantic model -

A default semantic model -

### Interactive reports -

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

### Technical Requirements -

The data store must support the following:

Read access by using T-SQL or Python

Semi-structured and unstructured data

Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

List prices that are less than or equal to 50 are in the low pricing group.

List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.

List prices that are greater than 1,000 are in the high pricing group.

### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

Fabric administrators will be the workspace administrators.

The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.

The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.

The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook

The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.

The date dimension must be available to all users of the data store.

The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

FabricAdmins: Fabric administrators

AnalyticsTeam: All the members of the analytics team

DataAnalysts: The data analysts on the analytics team

DataScientists: The data scientists on the analytics team  
DataEngineers: The data engineers on the analytics team  
AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

Enables a user to select a product to filter customer survey responses to only those who have purchased that product.

Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.

Shows data as soon as the data is updated in the data store.

Ensures that the report and the semantic model only contain data from the current and previous year.

Ensures that the report respects any table-level security specified in the source data store.

Minimizes the execution time of report queries.

Which type of data store should you recommend in the AnalyticsPOC workspace?

- A. a data lake
- B. a warehouse
- C. a lakehouse
- D. an external Hive metastore

**Suggested Answer: C**

Community vote distribution

C (100%)

 **stilferx** Highly Voted 1 year, 1 month ago

**Selected Answer: C**

IMHO,

my answer is C. Because of that: ""Semi-structured and unstructured data"" in the question  
upvoted 15 times

 **VAzureD** Highly Voted 1 year, 2 months ago

**Selected Answer: C**

C. to lakehouse

As a concept it exists in Fabric and is the most logical response for the given requirements,

Technical Requirements -

Semi-structured and unstructured data

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

A. a data lake

As an answer it could be valid, but it is not really a concept that we have in Fabric.

B. a warehouse

It is not valid, because there is unstructured data.

D. an external Hive metastore

Has no sense.

The Hive Metastore is a centralized repository that stores metadata related to the tables, partitions, columns, schemas, and other data structures used by Hive and Spark.

Contains information about the PHYSICAL location of the data, the schemas of the tables, and the relationships between them.

upvoted 11 times

 **NRezgui** Most Recent 6 months, 1 week ago

**Selected Answer: C**

a lakehouse

upvoted 1 times

 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

Lakehouse is the answer.

Warehouse not possible since data is structured and unstructured.

upvoted 1 times

🗲️ 👤 **rmeng** 1 year, 2 months ago

**Selected Answer: C**

Semi-structured and unstructured data

upvoted 1 times

🗲️ 👤 **GPerez73** 1 year, 2 months ago

**Selected Answer: C**

Semi-structured and unstructured data

upvoted 2 times

🗲️ 👤 **a\_51** 1 year, 3 months ago

**Selected Answer: C**

Lakehouse makes the most sense.

upvoted 1 times

🗲️ 👤 **TashaP** 1 year, 4 months ago

lakehouse 100%

upvoted 1 times

🗲️ 👤 **David\_Webb** 1 year, 4 months ago

In the technical requirement, it stated "Semi-structured and unstructured data" for the AnalyticsPOC data store. Thus, it must be a lakehouse.

upvoted 4 times

🗲️ 👤 **SamuComqi** 1 year, 4 months ago

**Selected Answer: C**

C. a lakehouse

The data store must handle semi-structured and unstructured data, therefore a Lakehouse should be the optimal solution supporting read access with T-SQL and Python.

upvoted 3 times

🗲️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: C**

Analytic=lakehouse

upvoted 1 times

🗲️ 👤 **Nicofr** 1 year, 4 months ago

**Selected Answer: C**

"Read access by using T-SQL or Python

Semi-structured and unstructured data"

upvoted 4 times

🗲️ 👤 **theseon** 1 year, 4 months ago

**Selected Answer: C**

C. a lakehouse

upvoted 2 times

You have a Fabric warehouse that contains a table named Staging.Sales. Staging.Sales contains the following columns.

Name	Data type	Nullable
ProductID	Integer	No
ProductName	Varchar(30)	No
SalesDate	Datetime2(6)	No
WholesalePrice	Decimal(18,2)	Yes
Amount	Decimal(18,2)	Yes

You need to write a T-SQL query that will return data for the year 2023 that displays ProductID and ProductName and has a summarized Amount that is higher than 10,000.

Which query should you use?

- SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount  
 FROM Staging.Sales  
 A. WHERE DATEPART(YEAR,SaleDate) = '2023'  
 GROUP BY ProductID, ProductName  
 HAVING SUM(Amount) > 10000  
  
 SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount  
 FROM Staging.Sales  
 B. GROUP BY ProductID, ProductName  
 HAVING DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000  
  
 SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount  
 C. FROM Staging.Sales  
 WHERE DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000  
  
 SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount  
 FROM Staging.Sales  
 D. WHERE DATEPART(YEAR,SaleDate) = '2023'  
 GROUP BY ProductID, ProductName  
 HAVING TotalAmount > 10000

**Suggested Answer: A**

Community vote distribution

A (100%)

 **TashaP** Highly Voted 1 year, 4 months ago

Anything without HAVING() + an aggregate is incorrect. HAVING was created for SQL to deal with filtering using an aggregate. Any option that references TotalAmount is incorrect because there is no nested statement in the syntax. Anything that uses HAVING() + DATEPART() is incorrect because you use a where clause for that. The answer is A.

upvoted 24 times

 **NRezgui** Most Recent 6 months, 1 week ago

**Selected Answer: A**

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR, SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000
```

upvoted 1 times

 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: A**

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR, SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000
```

upvoted 1 times

 **semauni** 8 months ago

**Selected Answer: A**

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR, SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000
```

Answer A is the only answer with valid syntax.

Answer B is excluded because the filter on year should be included in the WHERE-clause. It is the result of a function, but it is not an aggregate.

Answer C is excluded because the filter on SUM(Amount) should be included in the HAVING-clause, since this is an aggregate.

Answer D is excluded because the HAVING-clause does not accept aliases.

upvoted 2 times



  **BhavaniSubu** 11 months, 2 weeks ago

I am confused between A and B. But A is a correct Answer. Here is the explanation-

If we use non-aggregate function to Having clause it will throw below error-

Column 'SalesDate' is invalid in the HAVING clause because it is not contained in either an aggregate function or the GROUP BY clause.

upvoted 1 times

  **abrahamzetz** 2 months, 3 weeks ago

We want to filter the data to year 2023 only. To filter before an aggregation we use WHERE. So the DATEPART should be inside WHERE. We use HAVING to filter after aggregation, in this case the the summed amount higher than 10000. Therefore, B is incorrect because DATEPART should not be on HAVING. D is incorrect because HAVING cannot take an alias. So, A is the correct answer.

upvoted 1 times

  **sen\_su** 1 year ago

Actually both D & A are correct, aren't they?

"TotalAmount" is already available after the GROUP BY clause, or are there any reasons D is not correct?

upvoted 1 times

  **axe\_17** 2 months, 2 weeks ago

It would work in databricks (please correct me if I'm wrong) but not in T-SQL

upvoted 1 times

  **werka** 1 year ago

In D TotalAmount is an alias (HAVING TotalAmount > 10000) and only ORDER BY accepts aliases.

upvoted 7 times

  **see007** 1 year, 1 month ago

In some instances, you might want to exclude individual rows from groups (using a WHERE clause) before applying a condition to groups as a whole (using a HAVING clause).

A HAVING clause is like a WHERE clause, but applies only to groups as a whole (that is, to the rows in the result set representing groups), whereas the WHERE clause applies to individual rows. A query can contain both a WHERE clause and a HAVING clause. In that case:

The WHERE clause is applied first to the individual rows in the tables or table-valued objects in the Diagram pane. Only the rows that meet the conditions in the WHERE clause are grouped.

The HAVING clause is then applied to the rows in the result set. Only the groups that meet the HAVING conditions appear in the query output. You can apply a HAVING clause only to columns that also appear in the GROUP BY clause or in an aggregate function.

upvoted 2 times

  **RunicSkye** 1 year, 1 month ago

the table includes 'SalesDate' but every answer includes 'SaleDate' so they are all wrong!!! E. None of the above

upvoted 1 times

  **sen\_su** 1 year ago

The Table does not have to include SalesDate. It just return data for the year 2023, which can be filtered under the hood and not to be shown in SELECT.

upvoted 1 times

  **stilferx** 1 year, 1 month ago

IMHO, A

upvoted 1 times



🗳️ 👤 **JpLZa** 1 year, 1 month ago

I would say Option D.

Sum(Amount) AS TotalAmount is effectively TotalAmount > 10000

upvoted 3 times

🗳️ 👤 **AdventureChick** 6 months, 3 weeks ago

The HAVING clause does not allow aliases. This question is testing your knowledge of the order of operations (the order in which SQL statements are processed).

upvoted 1 times

🗳️ 👤 **hmoej** 3 months, 1 week ago

It does though..

upvoted 1 times

🗳️ 👤 **rmeng** 1 year, 1 month ago

**Selected Answer: A**

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
```

```
FROM Staging.Sales
```

```
WHERE DATEPART(YEAR, SaleDate) = '2023'
```

```
GROUP BY ProductID, ProductName
```

```
HAVING SUM(Amount) > 10000
```

upvoted 4 times

🗳️ 👤 **shivamgaurr** 1 year, 2 months ago

second one could be correct but saleDate is neither used in group by nor in any aggregate function. Hence A is the correct answer.

upvoted 1 times

🗳️ 👤 **earlqq** 1 year, 4 months ago

BTW spark sql allows us to refer to the count() alias

upvoted 2 times

🗳️ 👤 **David\_Webb** 1 year, 4 months ago

**Selected Answer: A**

The answer is A, no-brainer.

upvoted 2 times

🗳️ 👤 **SamuComqi** 1 year, 4 months ago

**Selected Answer: A**

```
A. SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
```

```
FROM Staging.Sales
```

```
WHERE DATEPART(YEAR, SaleDate) = '2023'
```

```
GROUP BY ProductID, ProductName
```

```
HAVING SUM(Amount) > 10000
```

Selected data is first filtered by Year, then grouped by ProductID and ProductName to compute the TotalAmount. Finally, only SUM(Amount) can be used after HAVING (not the alias).

upvoted 2 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: A**

Where to filter year data

Having to filter summerized data

Alias like TotalAmount not work in having

upvoted 2 times

🗳️ 👤 **objecto** 1 year, 4 months ago

**Selected Answer: A**

TotalAmount can not be used with HAVING. You must use SUM(Amount)

upvoted 4 times

## HOTSPOT -

You have a data warehouse that contains a table named Stage.Customers. Stage.Customers contains all the customer record updates from a customer relationship management (CRM) system. There can be multiple updates per customer.

You need to write a T-SQL query that will return the customer ID, name, postal code, and the last updated time of the most recent row for each customer ID.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

## Answer Area

```


WITH CUSTOMERBASE AS (
    SELECT [CustomerID]
    , [CustomerName]
    , [PostalCode]
    , [LastUpdated]
    , X =  OVER (PARTITION BY CustomerID
    ORDER BY LastUpdated DESC)
    FROM [LakehousePOC].[dbo].[CustomerChanges]
)
SELECT CustomerID, CustomerName, PostalCode, LastUpdated
FROM CUSTOMERBASE


Having Max(LastUpdated) = 1
WHERE LastUpdated = Max(LastUpdated)
WHERE X = 1

```

**Suggested Answer:** Box 1: ROW\_NUMBER()

Box 2: WHERE X = 1

 **R3D\_ENGINEER** Highly Voted 1 year, 4 months ago

The correct query is:

```

WITH CUSTOMERBASE AS (
    SELECT CustomerID, CustomerName, PostalCode, LastUpdated,
    ROW_NUMBER() OVER(PARTITION BY CustomerID ORDER BY LastUpdated DESC) as X
    FROM LakehousePOC.dbo.CustomerChanges
)

```

```

SELECT CustomerID, CustomerName, PostalCode, LastUpdated
FROM CUSTOMERBASE
WHERE X = 1

```

upvoted 37 times

 **semauni** 8 months ago

Thanks, this makes sense. I was already wondering where the X was coming from, it appears part of the syntax is missing in the question.

upvoted 2 times

 **AdventureChick** 6 months, 3 weeks ago

In the CTE, the X = is assigning the alias "X" to the value returned from the ROW\_NUMBER statement. It's the same as using "AS X" to assign an alias.

upvoted 1 times

 **varun\_r** Highly Voted 1 year, 3 months ago

Answer is RowNumber and X=1 -- No Brainer

upvoted 15 times

 **rohitrc8521** Most Recent 4 months ago

calm down guys, the given answer is correct!

upvoted 1 times

🗲️ 👤 **Rakesh16** 7 months, 2 weeks ago

ROW\_NUMBER() and where X=1

upvoted 2 times

🗲️ 👤 **jass007\_k** 8 months, 1 week ago

ROW\_NUMBER(), WHERE X =1

upvoted 2 times

🗲️ 👤 **7d97b62** 11 months, 2 weeks ago

The answer is correct

RowNumber and X=1

upvoted 1 times

🗲️ 👤 **Darshan6232** 1 year ago

Its straight forwarded. Provided answer is correct.

upvoted 1 times

🗲️ 👤 **stilferx** 1 year, 1 month ago

IMHO,

1. Row\_Number()

2. x = 1

It is a typical pattern. We are numerating by desc, and then filtering the first one (actually the last because of DESC order)

upvoted 3 times

🗲️ 👤 **TashaP** 1 year, 4 months ago

ROW\_NUMBER() + X = 1

upvoted 5 times

🗲️ 👤 **David\_Webb** 1 year, 4 months ago

First drop-down box: ROW\_NUMBER()

Second drop-down box: WHERE X = 1

As ORDER BY LastUpdated DESC was used, the first row will be the most recent row.

upvoted 5 times

🗲️ 👤 **Momoanwar** 1 year, 4 months ago

Row\_number

X=1

Row\_number give row position and start from 1

upvoted 5 times

HOTSPOT -

You have a Fabric tenant.

You plan to create a Fabric notebook that will use Spark DataFrames to generate Microsoft Power BI visuals.

You run the following code.

```
from powerbiclient import QuickVisualize, get_dataset_config, Report
```

```
PBI_visualize = QuickVisualize(get_dataset_config(df))
```

```
PBI_visualize
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

### Answer Area

Statements	Yes	No
The code embeds an existing Power BI report.	<input type="radio"/>	<input type="radio"/>
The code creates a Power BI report.	<input type="radio"/>	<input type="radio"/>
The code displays a summary of the DataFrame.	<input type="radio"/>	<input type="radio"/>

### Answer Area

	Statements	Yes	No
Suggested Answer:	The code embeds an existing Power BI report.	<input type="radio"/>	<input checked="" type="radio"/>
	The code creates a Power BI report.	<input checked="" type="radio"/>	<input type="radio"/>
	The code displays a summary of the DataFrame.	<input checked="" type="radio"/>	<input type="radio"/>

**Momoanwar** Highly Voted 1 year, 4 months ago

I think : No yes no

To display summary its describe() or sumarry()

upvoted 47 times

**scorradi** 5 months, 3 weeks ago

My guess: No, Yes, Yes. The "summary" word in this question is not referring to a function called summary. But summary is what QuickVisualization does for us.

upvoted 2 times

**bltemrah** 1 year ago

The code displays a summary of the DataFrame.

The last one should be Yes: The QuickVisualize function generates and displays a summary visualization of the DataFrame.

upvoted 11 times

**dev2dev** 1 year, 1 month ago

Correct! but display() function is MS fabric native function which gives more control on summarizing df including various visualizations

upvoted 2 times

**TashaP** Highly Voted 1 year, 4 months ago

Answer is No, Yes, Yes

The code is displaying a summary of the data frame using the visualizations. Point 3 is a technicality and I think to some extent depends on your interpretation of the question. I am going with Yes.

upvoted 20 times

**dp600** 1 year, 2 months ago

I understand your reasoning but i would go with No in 3, when I think in a summary of a dataframe I'm imagining something different.

upvoted 4 times

🗨️ 👤 **Mfathli** Most Recent 3 weeks, 4 days ago

My guess: No, Yes, Yes.

The third point depends on how you interpret summary in the statement, it's not a textual or statistical summary like `.describe()`, so the term "summary" is only applicable if regarded as a visual summary of the dataframe contents(which is what Quickvisualize does)

upvoted 1 times

🗨️ 👤 **hmoej** 3 months, 1 week ago

No, Yes, Yes

IMO the report itself is a summary of the dataframe.

It is a bad question though, you could interpret it either way.

upvoted 1 times

🗨️ 👤 **goldy29** 5 months ago

So, My final answer would be No, Yes, No

```
from powerbiclient import QuickVisualize, get_dataset_config
# Create a Power BI visualization from a Spark DataFrame
PBI_visualize = QuickVisualize(get_dataset_config(df))
# Display the visualization in the notebook
PBI_visualize
```

To embed a power bi report in notebook, you need to import these libraries and then authenticate

```
from powerbiclient import Report, models
from IPython.display import display
# Provide your Power BI report ID and workspace ID
workspace_id = "your_workspace_id"
report_id = "your_report_id"
# Create an embedded report object
report = Report(group_id=workspace_id, report_id=report_id)
# Display the report inside the notebook
display(report)
```

For summary, you can either use `describe()` or `summary()`

upvoted 1 times

🗨️ 👤 **Egocentric** 5 months, 2 weeks ago

No, Yes, No if it is a typo dataframe or yes if not dataframe

upvoted 1 times

🗨️ 👤 **Sowwy1** 7 months, 2 weeks ago

I think it's no, yes, no

upvoted 1 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

No-->yes-->yes

upvoted 1 times

🗨️ 👤 **semauni** 8 months ago

No, No, Yes

1) The code will display the contents of a dataframe, not an existing report.

2) The code does not create a new PBI report. Only after you click 'save' is a new report made.

3) The result of this code is a newly rendered block called 'Quick summary' with PBI visualisations. So if Microsoft calls it a summary, it is a summary.

Reference: <https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization>

upvoted 4 times

🗨️ 👤 **MultiCloudIronMan** 6 months, 3 weeks ago

No No Yes for the reason above

upvoted 1 times

🗨️ **semauni** 7 months, 3 weeks ago

Correction for 2: Yes. What I wrote is not wrong, however, Microsoft itself talks about 'creating a new report' and 'an embedded' report. So while I personally wouldn't consider it a new report, Microsoft does, and as long Microsoft grade you and not me, I would say Yes :)

Reference:<https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization#embed-a-power-bi-report-in-a-notebook>  
upvoted 1 times

🗨️ **jass007\_k** 8 months, 1 week ago

The provided answer is correct (No Yes Yes). Because with this code, we do get a Quick summary in the visualization. Please check the documentation and visualization image on this link: <https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization>  
upvoted 3 times

🗨️ **cafb698** 10 months ago

No, Yes, Yes  
upvoted 1 times

🗨️ **vka11** 1 year ago

The answer is No Yes Yes.  
See the documentation link, contains the exact code that is in the question  
<https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization>  
upvoted 10 times

🗨️ **vka11** 1 year ago

The doc clearly states that the last code line i.e. PBI\_visualize renders new report  
upvoted 3 times

🗨️ **Pib2** 1 year ago

Creates a PBI report? Not so sure, it shows a visualization like a PBI report, does not save it as such until you hit save.  
upvoted 2 times

🗨️ **282b85d** 1 year, 1 month ago

The code embeds an existing Power BI report:

No: Typically, embedding an existing Power BI report involves using Power BI REST API or embedding URLs, not Spark DataFrame operations within a notebook. The provided code snippet seems more focused on data manipulation and visual creation within the notebook itself, rather than embedding an external report.

The code creates a Power BI report:

Yes: When working with Spark DataFrames in a Fabric notebook, you can generate visuals that can be published to Power BI. This involves converting DataFrames to a format that Power BI can consume and using integration features provided by Fabric to create reports based on this data.

The code displays a summary of a DataFrame:

Yes: If the code snippet includes functions like `display(df)`, `df.show()`, or similar methods, it is intended to display a summary or the contents of the DataFrame. These functions are commonly used to provide a quick overview of the data within the notebook environment.

upvoted 5 times

🗨️ **laitoanthang** 1 year, 1 month ago

No yes no

The third one is a little bit confused though.

Anyway, the QuickVisualize is using the DataFrame, but it doesn't directly display a summary of the DataFrame itself

You'd need to use another method like `describe()` or `summary()`

upvoted 2 times

🗨️ **btemrah** 1 year ago

The code displays a summary of the DataFrame.

The last one should be Yes: The QuickVisualize function generates and displays a summary visualization of the DataFrame.

upvoted 1 times

🗨️ **stilferx** 1 year, 1 month ago


IMHO,

No -> Yes -> Yes.

Link is here: <https://learn.microsoft.com/en-us/power-bi/create-reports/jupyter-quick-report#create-and-render-a-quick-visualize-instance>

It is not embedded, it is an actual report. The 3rd one - is philosophical, but it summarizes the info from the data frame, so it makes sense to have "Yes" there

upvoted 8 times

  **stilferx** 1 year, 1 month ago

Probably, as colleagues said before, maybe the last one is No (it would be 100% yes for describe() )

upvoted 2 times

  **cmonte2** 1 year, 1 month ago

No, Yes and According to <https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization#displaydf-summary-view> the third should be No

upvoted 1 times

You are the administrator of a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:

Table1: A Delta table created by using a shortcut

Table2: An external table created by using Spark

Table3: A managed table -

You plan to connect to Lakehouse1 by using its SQL endpoint.

What will you be able to do after connecting to Lakehouse1?

- A. Read Table3.
- B. Update the data Table3.
- C. Read Table2.
- D. Update the data in Table1.

**Suggested Answer: D**

Community vote distribution

A (97%)

 **conie** Highly Voted 1 year, 4 months ago

**Selected Answer: A**

B & D is out as you can't update a table in lakehouse using SQL endpoint as this is read only. You will need to use spark or dataflows.

C is out because when you create external table using spark, you can see the table from the lakehouse but you can't see the table from SQL endpoint let alone read.

A is the answer, as I was able to see and read a managed table using SQL Endpoint  
upvoted 60 times

 **i\_have\_a\_name** 1 year, 1 month ago

It seems there is a way to have an external table created using Spark viewed using the Lakehouse SQL End point .

This is the spark code I used for testing:

```
df.write.format("delta").option("path","Tables/externalfolder").saveAsTable("testmanagedcsveditors")
```

When I run the catalog query :

spark.catalog.listTables(), it is returning as below :

```
Table(name='unmanagedcsveditors', catalog='spark_catalog', namespace=['training_lakehouse'], description=None, tableType='EXTERNAL', isTemporary=False)
```

Now if I go to the SQL end point, I can see a table named "externalfolder" .

I agree that the table name should be "unmanagedcsveditors", I can infact see 2 tables in the Lakehouse explorer - externalfolder and unmanagedcsveditors

upvoted 2 times

 **semauni** 8 months ago

Usually if it's this complicated, it's not the answer ;) even if you're right

upvoted 2 times

 **contactodonuno** Most Recent 3 months, 3 weeks ago

**Selected Answer: A**

The right answer is A. A managed table, it is stored within the Fabric storage and becomes immediately accessible through the SQL endpoint upon connection.

upvoted 1 times

 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: A**

Read Table3 (A managed table)

upvoted 1 times

 **jass007\_k** 8 months, 1 week ago



Correct option is A) as Lakehouse SQL endpoint is read only

B) and D) are incorrect as Lakehouse SQL endpoint is read only

Option C) is also incorrect as external tables created using spark in lakehouse are not visible to Lakehouse SQL endpoint

Refer this document for option c: <https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization>

upvoted 1 times

🗨️ **cafb698** 10 months, 1 week ago

The SQL analytics endpoint operates in read-only mode over lakehouse Delta tables. You can only read data from Delta tables using the SQL analytics endpoint. They can save functions, views, and set SQL object-level security.

Hence, B & D are out.

External Delta tables created with Spark code won't be visible to the SQL analytics endpoint. Use shortcuts in Table space to make external Delta tables visible to the SQL analytics endpoint.

Hence C is out.

Correct Answer: A

upvoted 1 times

🗨️ **Sara0724** 11 months ago

**Selected Answer: A**

The SQL analytics endpoint operates in read-only mode over lakehouse Delta tables.

External Delta tables created with Spark code won't be visible to the SQL analytics endpoint. <https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sql-analytics-endpoint>

upvoted 3 times

🗨️ **Kang\_24** 11 months, 1 week ago

**Selected Answer: A**

SQL Endpoint cannot update table --> B & D are outed.

Table 2 is a external table so it won't be displayed in SQL Endpoint view --> C is outed.

upvoted 2 times

🗨️ **Ahmadpbi** 11 months, 3 weeks ago

D is the correct answer.

In a lakehouse, the tables that can be edited using SQL endpoints are primarily Delta tables. These tables are specifically designed to be compatible with SQL analytics endpoints, allowing you to perform various SQL operations such as querying, creating views, and applying SQL security.

Other table formats like Parquet, CSV, and JSON are not directly editable through SQL endpoints; they need to be converted to Delta format first.

upvoted 1 times

🗨️ **b65ecca** 12 months ago

**Selected Answer: A**

External Delta tables created with Spark code won't be visible to the SQL analytics endpoint.

<https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sql-analytics-endpoint>

So C is out.

B and D are also out since with a SQL endpoint we can't update tables.

upvoted 2 times

🗨️ **Parth\_Mehta** 1 year ago

We can read table 1 and table 3 not table 2

and can't write any of them....

so ans will be A

upvoted 1 times

🗨️ **282b85d** 1 year, 1 month ago

Managed tables are fully controlled by the database system. These tables typically allow both read and write operations, including updates, through SQL endpoints.

After connecting to Lakehouse1 using its SQL endpoint, you will be able to:

- \*\*Read Table3\*\* (A)

- \*\*Update the data in Table3\*\* (B)

So, the correct options are:

**\*\*A. Read Table3.\*\***

**\*\*B. Update the data in Table3.\*\***

upvoted 2 times

🗨️ **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO,

the answer is A.

updates are prohibited at all. reading spark external table is a big no-no.

Link: <https://www.linkedin.com/pulse/use-shortcuts-instead-external-tables-reference-data-fabric-popovic/>

upvoted 4 times

🗨️ **rmeng** 1 year, 1 month ago

**Selected Answer: A**

C is not correct because the external tables are not accessible from the endpoint.

B & D is out of question cause as SQL endpoint is read only.

upvoted 2 times

🗨️ **Chrys941** 1 year, 2 months ago

**Selected Answer: A**

D is completely wrong is not typically feasible through a shortcut in SQL endpoint setup.

B Generally Supported but depends on the SQL endpoint permissions for such operations

A you should be able to read data from table3 since it is a managed table and such operations are standard through Sql Endpoints

upvoted 2 times

🗨️ **GPerez73** 1 year, 2 months ago

**Selected Answer: A**

A is correct. Tested!

upvoted 2 times

🗨️ **a\_51** 1 year, 3 months ago

**Selected Answer: A**

A is correct, D is not. SQL endpoint is read-only.

<https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sql-analytics-endpoint>

upvoted 4 times

🗨️ **a\_51** 1 year, 3 months ago

In that link it tells you as well to modify you have to switch

"To modify data in Lakehouse delta tables, you have to switch to lakehouse mode and use Apache Spark."

upvoted 1 times

🗨️ **mtroyano** 1 year, 3 months ago

**Selected Answer: A**

Options B and D are incorrect because an endpoint does not support DML operations. Option C is not correct because the external tables are not accessible from the endpoint. The correct answer is A, the managed tables can be read from the connection point.

upvoted 2 times

You have a Fabric tenant that contains a warehouse.

You use a dataflow to load a new dataset from OneLake to the warehouse.

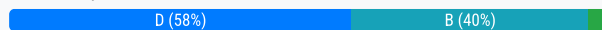
You need to add a PowerQuery step to identify the maximum values for the numeric columns.

Which function should you include in the step?

- A. Table.MaxN
- B. Table.Max
- C. Table.Range
- D. Table.Profile

**Suggested Answer: B**

Community vote distribution



**lengzhai** 1 year, 4 months ago

**Selected Answer: D**

D. Table.Profile

<https://learn.microsoft.com/en-us/powerquery-m/table-profile>

upvoted 38 times

**philippevandevyver** 11 months ago

Not correct, will not add another step.

upvoted 2 times

**amar5555** 1 year, 2 months ago

you will get summary in that max will include but it doesn't add as power query step so i think it is option b

upvoted 9 times

**PazaBlandData** 1 year, 2 months ago

Table.Max provides only a row with the max value per a single column. We have multiple of columns. It's d

upvoted 7 times

**woliveiras** 1 year ago

**Selected Answer: B**

B guys, "you need to add a step..."

upvoted 10 times

**woliveiras** 1 year ago

Sorry Guys, I changed my mind. It is the letter D.

upvoted 4 times

**Vulkany** 5 months, 2 weeks ago

**Selected Answer: D**

I initially thought that the right answer was B, but the question must be answered:

"You need to add a PowerQuery step to identify the maximum values of numeric columns."

Table.Max although it can find the maximum value, it is used for a single column and returns a single row based on one column. It doesn't work for several numeric columns at once.

upvoted 3 times

**pk07** 5 months, 3 weeks ago

**Selected Answer: D**

D. Table.Profile

upvoted 1 times

**pk07** 5 months, 3 weeks ago

**Selected Answer: B**

Table.Max:

The Table.Max function returns the maximum value for a specified column in a table.

It is the correct function to use when you need to identify the maximum value for numeric columns.

upvoted 1 times

  **sajjuh** 3 months, 2 weeks ago

We need to identify for multiple columns.

upvoted 1 times

  **pk07** 5 months, 3 weeks ago

**Selected Answer: B**

Correct Answer:

B. Table.Max

Explanation:

Table.Max is the appropriate function to use when you need to identify the maximum value in a specific column. You can apply this function to each numeric column in your dataset to find the maximum values.

Why Other Options Are Incorrect:

A. Table.MaxN: This function is used to get the top N rows, not the maximum value in a column.

C. Table.Range: This function is used to select a range of rows, not to find maximum values.

D. Table.Profile: While this function provides a summary of the data, including maximum values, it does not allow you to extract or use these maximum values directly in subsequent steps.

upvoted 3 times

  **51eef92** 6 months ago

**Selected Answer: D**

Chose D as it's looking for multiple numeric columns not a single one

upvoted 1 times

  **salini** 6 months ago

**Selected Answer: D**

D

Table.Profile:

This function provides a statistical summary of a table, including the maximum values for numeric columns, along with other statistics like minimum, average, and count.

It is specifically designed to give an overview of the dataset, making it ideal for identifying maximum values for numeric columns.

Table.MaxN:

Returns the top N rows based on a specified column but does not specifically identify the maximum value across numeric columns.

Table.Max:

Returns a single row with the maximum value for a specified column but does not provide maximum values for all numeric columns.

upvoted 1 times

  **MuralikumarCh** 7 months ago

**Selected Answer: B**

The correct answer is B. Table.Max.

Explanation:

Table.Max: This function is specifically designed to calculate the maximum value within a column of a table. It's the most suitable choice for identifying the maximum values for numeric columns in your PowerQuery step.

Table.MaxN: This function returns the top N rows of a table. It's not appropriate for calculating maximum values.

Table.Range: This function creates a new table containing a range of rows from an existing table. It's not relevant to finding maximum values.

Table.Profile: This function provides statistical information about a table, including measures of central tendency and dispersion. While it can be useful for understanding data, it's not the most direct way to find maximum values.

By using Table.Max in your PowerQuery step, you can efficiently identify the maximum values for the numeric columns in your dataset, enabling further analysis and decision-making.

upvoted 3 times

🗨️ **nappi1** 7 months, 1 week ago

it says columnS, so it is D  
upvoted 2 times

🗨️ **Sowwy1** 7 months, 2 weeks ago

I think it's D.  
upvoted 1 times

🗨️ **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: D**

Table.Profile  
upvoted 1 times

🗨️ **semauni** 8 months ago

**Selected Answer: D**

Option D, Table.Profile, returns all kind of aggregate values for \*all\* columns. Including the max values. It adds a step to PowerQuery; it is written in M, you're not watching it as part of the Power BI interface.

I was convinced it was B, but not that it is TABLE.Max, not List.Max. That is an important difference. This function does the following: "Returns the largest row in the table, given the comparisonCriteria." So you get one row of data, not the list of values you want.

Reference:

<https://learn.microsoft.com/en-us/powerquery-m/table-profile>

<https://learn.microsoft.com/en-us/powerquery-m/table-max>

upvoted 2 times

🗨️ **DataSturdy** 8 months, 4 weeks ago

**Selected Answer: D**

Table.Profile gives Min,Max,Average,StandardDeviation,Count,NullCount,DistinctCount for all columns  
Table.Profile("#last step name")

Table.Max gives only the maximum value for the specified column(entire row of the maximum value column)

Table.Max("#last step name","column name")

upvoted 2 times

🗨️ **nyoike** 10 months, 3 weeks ago

**Selected Answer: D**

I vote D too. The ask is for maximum value for numeric columnS. Note plural. Table.MaxN and Table.Max will return entire row(s) with max value for a specific column (singular) so they are out. Table.Profile returns the following stats for all columns where applicable:-

minimum

maximum

average

standard deviation

count

null count

distinct count

so that fits the bill.

upvoted 2 times

🗨️ **philippevandevyver** 11 months ago

**Selected Answer: B**

A & C Incorrect, other logic

D -> Will not add another power query step, will only display

B it is

upvoted 2 times

🗨️ **Pegooli** 11 months ago

**Selected Answer: D**

The reason why Table.Max is not the correct choice is that Table.Max returns the row in a table that contains the maximum value for a specified column, rather than providing the maximum values for all numeric columns.

upvoted 2 times

You have a Fabric tenant that contains a machine learning model registered in a Fabric workspace.

You need to use the model to generate predictions by using the PREDICT function in a Fabric notebook.

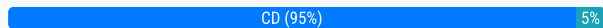
Which two languages can you use to perform model scoring? Each correct answer presents a complete solution.

NOTE: Each correct answer is worth one point.

- A. T-SQL
- B. DAX
- C. Spark SQL
- D. PySpark

**Suggested Answer:** CD

Community vote distribution



**mtroyano** Highly Voted 1 year, 3 months ago

**Selected Answer:** CD

Notebook only accepts the languages: PySpark, Spark, Spark SQL and SparkR  
upvoted 18 times

**DarioReymago** 1 year ago

yes but you can use "%sql" in a notebook  
upvoted 4 times

**dev2dev** 1 year ago

that indicates to use spark sql language.  
upvoted 6 times

**testtaker45** 5 months, 2 weeks ago

Thank you for that. I had the same question DarioReymago had.  
upvoted 1 times

**Vulkany** Most Recent 5 months, 2 weeks ago

**Selected Answer:** CD

Notebook only accepts the languages: PySpark, Spark, Spark SQL and SparkR  
upvoted 1 times

**NRezgui** 6 months, 1 week ago

**Selected Answer:** CD

PySpark, Spark  
upvoted 1 times

**shorymor** 7 months ago

**Selected Answer:** CD

Notebook only accepts the languages: PySpark, Spark, Spark SQL and SparkR  
upvoted 1 times

**Rakesh16** 7 months, 2 weeks ago

**Selected Answer:** CD

Spark SQL and PySpark  
upvoted 1 times

**jass007\_k** 8 months, 1 week ago

Its C and D  
upvoted 1 times

**gtc108** 1 year ago

<https://learn.microsoft.com/en-us/fabric/data-science/tutorial-data-science-batch-scoring>  
ou'll load the test dataset into a spark DataFrame and create an MLFlowTransformer object to generate batch predictions. You can then invoke the

PREDICT function using one of following three ways:

Transformer API from SynapseML

Spark SQL API

PySpark user-defined function (UDF)

upvoted 2 times

🗳️ 👤 **Gab13** 1 year ago

**Selected Answer: CD**

Notebook only accepts the languages: PySpark, Spark, Spark SQL and SparkR

upvoted 2 times

🗳️ 👤 **DarioReymago** 1 year ago

**Selected Answer: AD**

We can use Predict function with T-SQL:

<https://learn.microsoft.com/en-us/sql/t-sql/queries/predict-transact-sql?view=sql-server-ver16>

But I cannot find the Predict function in Spark SQL. Is different to run a predict process in PySpark

upvoted 1 times

🗳️ 👤 **semauni** 8 months ago

The question asks to do this from a notebook. Spark SQL can be accessed from the notebook using the %%sql-command, while T-SQL cannot.

Spark SQL, however, can also be used: <https://learn.microsoft.com/en-us/azure/synapse-analytics/machine-learning/tutorial-score-model-predict-spark-pool>

upvoted 1 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: CD**

IMHO,

the answer is C & D.

Link is here: <https://learn.microsoft.com/en-us/azure/synapse-analytics/machine-learning/tutorial-score-model-predict-spark-pool>

Here, ""You can call PREDICT three ways, using Spark SQL API, using User define function (UDF), and using Transformer API. "".

That means, UDF = PySpark in our case.

upvoted 1 times

🗳️ 👤 **rmeng** 1 year, 1 month ago

**Selected Answer: CD**

<https://learn.microsoft.com/en-us/azure/synapse-analytics/machine-learning/tutorial-score-model-predict-spark-pool>

upvoted 1 times

🗳️ 👤 **David\_Webb** 1 year, 4 months ago

**Selected Answer: CD**

Using Fabric notebook, thus must be C and D.

upvoted 1 times

🗳️ 👤 **BrandonPerks** 1 year, 4 months ago

The mention of Fabric Notebook, gives the hint to using Spark. Thus I went with CD

upvoted 1 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: CD**

C & d in notebook

upvoted 2 times

🗳️ 👤 **wojciech\_wie** 1 year, 4 months ago

CD - <https://learn.microsoft.com/en-us/azure/synapse-analytics/machine-learning/tutorial-score-model-predict-spark-pool>

upvoted 2 times

🗳️ 👤 **olavrab8** 1 year, 4 months ago

**Selected Answer: CD**

Answer CD Correct

T-SQL Cannot be used, nor DAX

upvoted 3 times



You are analyzing the data in a Fabric notebook.

You have a Spark DataFrame assigned to a variable named df.

You need to use the Chart view in the notebook to explore the data manually.

Which function should you run to make the data available in the Chart view?

- A. displayHTML
- B. show
- C. write
- D. display

**Suggested Answer:** D

Community vote distribution

D (100%)

  **stilferx**  1 year, 1 month ago

**Selected Answer:** D

IMHO,

the answer is D.

Link is here: <https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization>

Point is: You can use the display function on dataframes that created in PySpark and Scala on Spark DataFrames or Resilient Distributed Datasets (RDD) functions to produce the rich dataframe table view and chart view.

upvoted 12 times

  **JG1984** 1 year ago

Out of all the reviewers, I like the way you explain all the answers. Just wondering if you were able to ace the exam

upvoted 2 times

  **TashaP**  1 year, 4 months ago

D is the correct answer (the keyword is chart view)

A is another possibility for displaying data but not based on the requirements of this question, it is separate from chart view.

<https://learn.microsoft.com/en-us/fabric/data-engineering/notebook-visualization>


upvoted 5 times

  **NRezgui**  6 months, 1 week ago

**Selected Answer:** D


display

upvoted 1 times

  **Sowwy1** 7 months, 2 weeks ago

I think it's D.

upvoted 1 times

  **Rakesh16** 7 months, 2 weeks ago

**Selected Answer:** D

Display

upvoted 1 times

  **arrow040** 12 months ago

The .show function displays the contents of a Spark DataFrame in a tabular format within the notebook but does not enable interactive data exploration with the Chart view. To make the data available in the Chart view for interactive visualization, you should use the display function (Answer D)

upvoted 2 times

🗨️ 👤 **Miro\_dd** 12 months ago

**Selected Answer: D**

Display dataframe function should be correct  
upvoted 3 times

🗨️ 👤 **SilvanoRamalho** 1 year ago

**Selected Answer: D**

D. DISPLAY  
upvoted 1 times

🗨️ 👤 **Gabonia** 1 year ago

Show is the correct answer  
Thank you.  
upvoted 1 times

🗨️ 👤 **gshopper** 1 year ago

B: show is correct. plt.show()  
upvoted 2 times

🗨️ 👤 **fhlos** 1 year ago

**Selected Answer: D**

display(df)  
upvoted 1 times

🗨️ 👤 **Fmy** 1 year, 1 month ago

**Selected Answer: D**

Display(df)  
Allows you to view the data in chart  
upvoted 1 times

🗨️ 👤 **mtroyano** 1 year, 3 months ago

**Selected Answer: D**

Checked, the correct option is Display.  
upvoted 2 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: D**

Display allow to see chart and inspect statistiques  
upvoted 2 times

🗨️ 👤 **wojciech\_wie** 1 year, 4 months ago

D is correct  
upvoted 1 times

You have a Fabric tenant that contains a Microsoft Power BI report named Report1. Report1 includes a Python visual. Data displayed by the visual is grouped automatically and duplicate rows are NOT displayed. You need all rows to appear in the visual. What should you do?

- A. Reference the columns in the Python code by index.
- B. Modify the Sort Column By property for all columns.
- C. Add a unique field to each row.
- D. Modify the Summarize By property for all columns.

**Suggested Answer: A**

Community vote distribution



**TashaP** Highly Voted 1 year, 4 months ago

A - often the Microsoft learn pages give you the exact answer:  
Under tips on the link I posted:

In some cases, you might not want automatic grouping to occur, or you might want all rows to appear, including duplicates. In those cases, you can add an index field to your dataset that causes all rows to be considered unique and prevents grouping.

<https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals>

There are always multiple approaches in application but if you do the MS exam give the MS answer.  
upvoted 66 times

**SKN08** 1 year ago

sorry but that (A) sounds incorrect. This is from the MS documentation: "In those cases, you can add an index field to your dataset that causes all rows to be considered unique and prevents grouping." and it means create new column that has a unique value for each and every row of the dataset. \*C\* is correct  
upvoted 8 times

**[Removed]** 1 year, 2 months ago

Adding an index field to your dataset means that you add a column/field that's unique to each row, meaning that the correct answer is C, not A. "Reference columns by index" means writing `df.iloc[indexNo]` instead of `df["fieldName"]`, and that doesn't help anything.  
upvoted 28 times

**thuss** 1 year, 4 months ago

It even says "The default aggregation is Don't summarize" already. Helpful link, definitely correct answer.  
upvoted 3 times

**Fermd** Highly Voted 1 year, 4 months ago

**Selected Answer: D**

The right answer is D: By setting the "Summarize By" property to "None" for all columns, you disable automatic aggregation and ensure all rows, including duplicates, are displayed in the Python visual.  
upvoted 14 times

**semauni** 8 months ago

The default aggregation of a Python script is 'Don't summarize', but that still means that duplicate rows only appear once. It just means that, for instance, a number column is not added up, but if you have two rows with the same numbers, they won't both appear.  
upvoted 1 times

**NRezgui** Most Recent 6 months, 1 week ago

**Selected Answer: C**

Add a unique field to each row.  
upvoted 1 times

🗳️ 👤 **MultiCloudIronMan** 6 months, 3 weeks ago

**Selected Answer: D**

Modify the Summarize By property for all columns: By default, Power BI may summarize data in visuals, which can lead to grouping and aggregation of rows. Modifying the "Summarize By" property for all columns to "Do Not Summarize" will prevent this automatic grouping and ensure that all rows, including duplicates, are displayed in the visual.

upvoted 1 times

🗳️ 👤 **Sowwy1** 7 months, 2 weeks ago

I think it's C. Add a unique field to each row.

upvoted 2 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

Add a unique field to each row

upvoted 1 times

🗳️ 👤 **Naqib** 7 months, 3 weeks ago

A

Quoted from <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals>:

"In some cases, you might not want automatic grouping to occur, or you might want all rows to appear, including duplicates. In those cases, you can add an index field to your dataset that causes all rows to be considered unique and prevents grouping."

upvoted 4 times

🗳️ 👤 **semauni** 8 months ago

**Selected Answer: C**

"In some cases, you might not want automatic grouping to occur, or you might want all rows to appear, including duplicates. In those cases, you can add an index field to your dataset that causes all rows to be considered unique and prevents grouping."

I was leaning towards A because of the word 'index', and usually if a specific term is mentioned, Microsoft wants to hear you use that term. But option A talks about REFERENCING the index in the CODE. The quote above is not about how you write your code (like cgroven mentioned, e.g. `df["streetName"]` vs. `df[1]` for example), but it is about how you should modify your data structure and visual to handle the results. If you read the above statement with this distinction in mind, then it clearly reads C.

upvoted 7 times

🗳️ 👤 **semauni** 8 months ago

And about D: deduplication is not achieved by disabling grouping. The behaviour of Power BI is, unintuitively to the new user, to hide duplicate rows. So say you have two rows with exactly the same contents, also containing the same numbers for instance, no summarization just causes your visual not to aggregate them (sum, average, etc.). It will instead just show ONE row with the original number. You can't tell if this represents multiple rows or not unless you add a unique field.

upvoted 1 times

🗳️ 👤 **el00721** 8 months, 1 week ago

D. By setting the "Summarize By" property to "None" for all columns, you disable automatic aggregation and ensure all rows, including duplicates, are displayed in the Python visual.

upvoted 1 times

🗳️ 👤 **nasra2** 9 months, 1 week ago

A: <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals>

In some cases, you might not want automatic grouping to occur, or you might want all rows to appear, including duplicates. In those cases, you can add an index field to your dataset that causes all rows to be considered unique and prevents grouping.

upvoted 2 times

🗳️ 👤 **Training\_ND** 10 months ago

**Selected Answer: A**

MS ANSWER.

<https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals>

upvoted 3 times



🗳️ 👤 **cafb698** 10 months, 1 week ago

**Selected Answer: A**

As much as C is tempting, let's focus on the wording. "Add a unique field to each row". You simply say, "Add another column for unique id" or something like that.

In terms of Microsoft ways of asking, it's A.

upvoted 3 times

  **VishalTile** 10 months, 3 weeks ago

Power BI tends to automatically summarize data in visuals, which can cause rows to be grouped and duplicates to be hidden.

The "Summarize By" property controls how data is aggregated in a visual. By setting it to "Do Not Summarize," you prevent the automatic grouping of data, ensuring that all rows, including duplicates, are displayed.

None of the other options directly address the issue of rows being grouped and duplicates being hidden in a Power BI visual.

So, D is the right choice.

upvoted 1 times

  **fits08pistils** 10 months, 3 weeks ago

It seems not everyone can read and understand documentation. The following Python script is execute EVERY TIME you use a Python visual:

```
dataset = pandas.DataFrame(column1, column2, ...)
dataset = dataset.drop_duplicates()
```

So the ONLY way to prevent the duplicates from being dropped is to have a unique identifier.

Ergo, option C is the only possible answer here.

upvoted 2 times

  **SilvanoRamalho** 1 year ago

**Selected Answer: A**


<https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-visuals>

--- Respostas "A"... da MS...

Em alguns casos, talvez você não queira que o agrupamento automático ocorra ou queira que todas as linhas apareçam, inclusive as duplicadas.

Nesses casos, você pode adicionar um campo de índice ao seu conjunto de dados que faz com que todas as linhas sejam consideradas exclusivas e evita o agrupamento.

upvoted 2 times

  **ziggy1117** 1 year ago

**Selected Answer: C**

ANSWER IS C:

In some cases, you might not want automatic grouping to occur, or you might want all rows to appear, including duplicates. In those cases, you can add an index field to your dataset that causes all rows to be considered unique and prevents grouping.

This means you need to add an index field column in which case each row will have a unique index value per row

upvoted 2 times

  **SilvanoRamalho** 1 year ago

**Selected Answer: C**

C- ADIONAR UMA COLUA INDICE

upvoted 2 times

## DRAG DROP -

You have a Fabric tenant that contains a semantic model. The model contains data about retail stores.

You need to write a DAX query that will be executed by using the XMLA endpoint. The query must return a table of stores that have opened since December 1, 2023.

How should you complete the DAX expression? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

## Values

## Answer Area

Values	Answer Area
DEFINE	
EVALUATE	VAR_SalesSince = DATE (2023, 12, 01)
FILTER	
SUMMARIZE	FILTER (
TABLE	(Store, Store[Name], Store[OpenDate]), Store[OpenDate] >= _SalesSince )

## Answer Area

**Suggested Answer:**

DEFINE	VAR_SalesSince = DATE (2023, 12, 01)
EVALUATE	FILTER (
TABLE	(Store, Store[Name], Store[OpenDate]), Store[OpenDate] >= _SalesSince )

**wojciech\_wie** Highly Voted 1 year, 4 months ago

- 1) DEFINE
  - 2) EVALUATE
  - 3) SUMMARIZE
- upvoted 73 times

**seaverick** Highly Voted 9 months, 2 weeks ago

I have tested it in dax.do changing fields and ans is DEFINE, EVALUATE and SUMMARIZE:

DEFINE

VAR \_SalesSince =  
DATE ( 2000, 12, 01 )

EVALUATE

FILTER (

SUMMARIZE ( 'Product', 'Product'[Product Name], 'Product'[Available Date] ),

'Product'[Available Date] >= \_SalesSince

)

upvoted 11 times

**sajjuh** Most Recent 3 months, 2 weeks ago

Table fn doesnt exist?

upvoted 1 times

**Rakesh16** 7 months, 2 weeks ago

Define, Evaluate & Summarize

upvoted 2 times

🗨️ 👤 **Harsha\_Guggilla** 9 months, 2 weeks ago

DEFINE

EVALUATE

SUMMARIZE

upvoted 2 times

🗨️ 👤 **user12345678** 10 months ago

Just to save people some time, I haven't seen a function called TABLE, so the last one is SUMMARIZE.

upvoted 3 times

🗨️ 👤 **6d1de25** 11 months, 3 weeks ago

The correct syntax is:

Define

Evaluate

Table

<https://learn.microsoft.com/en-us/dax/define-statement-dax>

<https://learn.microsoft.com/en-us/dax/evaluate-statement-dax>

upvoted 5 times

🗨️ 👤 **Jons123son** 1 year ago

Was in exam. Scored 95%

Chose

1) DEFINE

2) EVALUATE

3) SUMMARIZE

upvoted 5 times

🗨️ 👤 **282b85d** 1 year ago

DEFINE

VAR \_SalesSince = DATE(2023, 12, 01)

EVALUATE

FILTER(

SUMMARIZE(

Store, Store[Name], Store[OpenDate]

),

Store[OpenDate] >= \_SalesSince

)

upvoted 2 times

🗨️ 👤 **DarioReymago** 1 year ago

Define, Evaluate, Summarize.

Option Table does not exist

upvoted 1 times

🗨️ 👤 **ca63a55** 1 year ago

Why is it necessary to SUMMARIZE? the stores are supposed to have different names, wouldn't a FILTER be enough?

upvoted 1 times

🗨️ 👤 **mooom0000** 1 year, 1 month ago

<https://stackoverflow.com/questions/69047367/table-visual-is-unintuitively-aggregating-my-data>

Refer to this post.

It explains why the don't summarize option is not correct.

upvoted 1 times

🗨️ 👤 **adilene** 1 year, 2 months ago

1) DEFINE

2) EVALUATE



3) SUMMARIZE

upvoted 3 times

  **woliveiras** 1 year, 3 months ago

Define, Evaluate and Summarize. Please, update the correct answer.

upvoted 2 times

  **a\_51** 1 year, 3 months ago

The VAR\_SalesSince should have a space between them to be:

VAR \_SalesSince.

upvoted 5 times

  **Jeff\_Zhu** 1 year, 4 months ago

DEFINE

EVALUATE

SUMMARIZE

upvoted 4 times

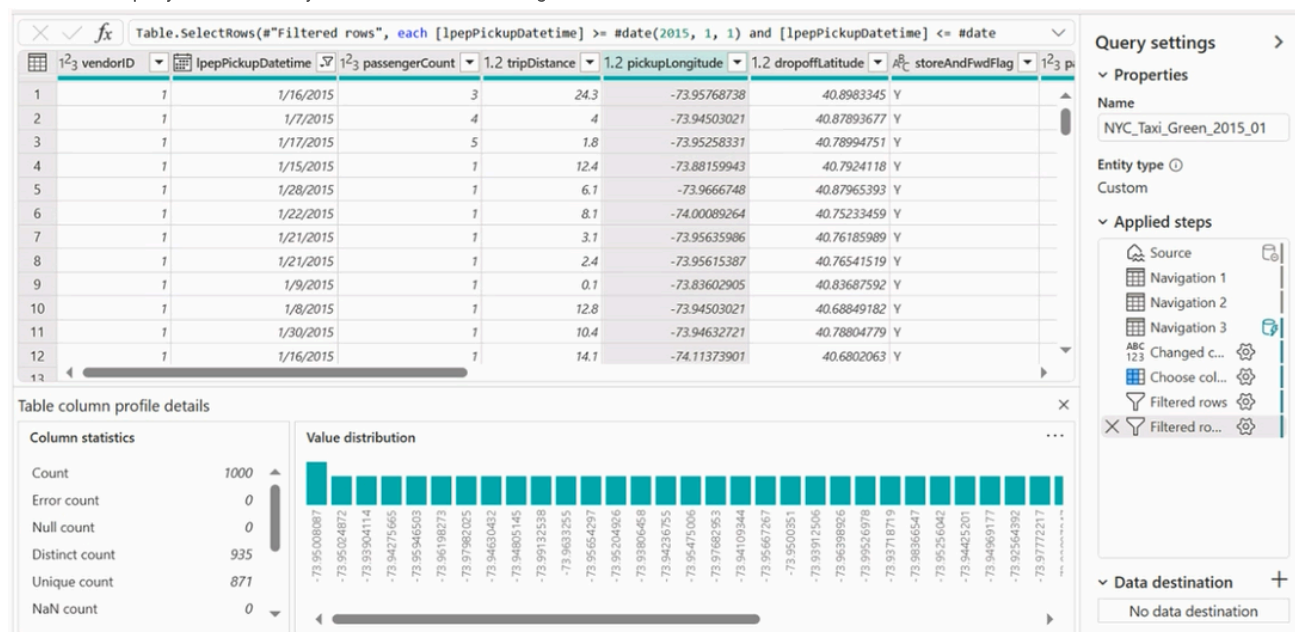
  **Momoanwar** 1 year, 4 months ago

Define - evaluate - summarize

upvoted 4 times



You have a Fabric workspace named Workspace1 that contains a dataflow named Dataflow1. Dataflow1 has a query that returns 2,000 rows. You view the query in Power Query as shown in the following exhibit.



What can you identify about the pickupLongitude column?

- A. The column has duplicate values.
- B. All the table rows are profiled.
- C. The column has missing values.
- D. There are 935 values that occur only once.

**Suggested Answer: A**

Community vote distribution

A (100%)

**nmosq** Highly Voted 1 year, 4 months ago

Selected Answer: A

Answer A.

B - Not all the rows are profiled in the sample (only 1000 of 2000)

C- From the column statistics, you don't have any missing values in the sample

D- The values that occur only once are 871 (unique count)

upvoted 32 times

**wojciech\_wie** 1 year, 4 months ago

correct

upvoted 3 times

**a\_51** Highly Voted 1 year, 3 months ago

Selected Answer: A

A

We see a count of 1000 (which is the limit by default) we do not know all the data is read, but we can see of the 1000 distinct is less and so we have duplicate values.

upvoted 5 times

**NRezgui** Most Recent 6 months, 1 week ago

Selected Answer: A

The column has duplicate values.

upvoted 1 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: A**

The column has duplicate values  
upvoted 1 times

🗨️ 👤 **Naqib** 7 months, 3 weeks ago

Answer A:

Distinct Value: This refers to all different values present in a dataset. When you retrieve distinct values from a column, you eliminate duplicate values so that each value is shown once. For example, if a column contains the values [1, 2, 2, 3, 3, 3], the distinct values would be [1, 2, 3].

Unique Value: This usually refers to values that appear only once in the dataset. Unlike distinct values, a unique value will only be considered if it has no duplicates at all. For example, if a column contains the values [1, 2, 2, 3, 3, 3], the unique values would be [1], since only 1 appears without repetition.

upvoted 2 times

🗨️ 👤 **b65ecca** 12 months ago

**Selected Answer: A**

Difficult to say if BCD are correct since we only see 1000 rows and not all columns. One thing is for sure though, there are columns that have values that occur more than once.

upvoted 2 times

🗨️ 👤 **gills** 1 year, 1 month ago

Answer is A

Distinct mean : count all the values as 1, even if there was more than one.

Unique mean : count only the value that are not repeated in the particular column

upvoted 3 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO, it is A, well explained below

upvoted 2 times

🗨️ 👤 **Nefirs** 1 year, 3 months ago

**Selected Answer: A**

A

my reasoning: every other answer option cannot be answered for sure since only 1000 values out of 2000 are profiled. -> B: only 1000 rows are profiled -> C: the column might have missings -> D: there might be more unique/distinct counts.

upvoted 2 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

**Selected Answer: A**

A is Best choice based on the picture.

upvoted 1 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: A**

Its A

Only one column selected here

No informations about missing values

Distinct count not mean exist only once

upvoted 4 times

You have a Fabric tenant named Tenant1 that contains a workspace named WS1. WS1 uses a capacity named C1 and contains a dataset named DS1.

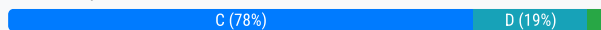
You need to ensure read-write access to DS1 is available by using XMLA endpoint.

What should be modified first?

- A. the DS1 settings
- B. the WS1 settings
- C. the C1 settings
- D. the Tenant1 settings

**Suggested Answer: C**

Community vote distribution



**a\_51** 1 year, 3 months ago

**Selected Answer: C**

As XMLA is set to Read-Only first, you must go to the capacity settings to enable read-write.

<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools#enable-xmla-read-write>

upvoted 26 times

**STH** 1 year, 3 months ago

**Selected Answer: C**

The question concerns changing from read-only to read-write (in the Capacity settings), not about enabling XMLA endpoints (in the Tenant settings), which, as per the query, are already set up.

upvoted 11 times

**NRezgui** 6 months, 1 week ago

**Selected Answer: C**

the C1 settings

upvoted 2 times

**Charley92** 7 months ago

**Selected Answer: C**

The XMLA endpoint in Microsoft Fabric allows read-write operations for datasets. To enable read-write access, you must first configure the capacity settings (C1 in this case) to allow XMLA read-write connectivity

upvoted 1 times

**Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

the C1 settings

upvoted 1 times

**jass007\_k** 8 months ago

Correct answer is option C)

R/W settings of XMLA endpoint is done under capacity settings by fabric admin. By default, read-only connectivity using the endpoint is enabled for the Semantic models.

Link for reference: <https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools#enable-xmla-read-write>

upvoted 2 times

**6d1de25** 11 months, 3 weeks ago

**Selected Answer: C**

Correct answer is C

<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools#:~:text=To%20enable%20read,to%20the%20capacity.>

upvoted 1 times

**momo1165** 1 year ago

Enable XMLA read-write:

Select Settings > Admin portal.

In the Admin portal, select Capacity settings > Power BI Premium > capacity name.

Expand Workloads. In the XMLA Endpoint setting, select Read Write. The XMLA Endpoint setting applies to all workspaces and semantic models assigned to the capacity.

upvoted 1 times

🗨️ 👤 **Rezako** 1 year ago

**Selected Answer: C**

The right answer is C,

You use tenant setting to Allow XMLA endpoints and Analyze in Excel with on-premises datasets. It is read only access but with Capacity settings you can choose between off, read only and read-write

upvoted 1 times

🗨️ 👤 **Jons123son** 1 year ago

**Selected Answer: D**

The questions asked what should be modified FIRST. Therefore, I would go with D because C requires that D is activated. Although, I am not sure if D is not activated by default, thus making C the ONLY modification required.

I hate it when MS exam questions leave room for interpretation. Why don't they ask for two answers...

Weirdly enough Questions 21 covers basically the same topic and there people agree that Tenant settings are definitely required next to capacity settings.

upvoted 3 times

🗨️ 👤 **user12345678** 10 months ago

Those are other 2 settings in the Tenant settings are enabled by default. If it explicitly said XMLA Read/Write then I'd want to go with the Capacity settings like others have said as it seems the most applicable. I do get what you mean though, I was having the same doubts.

upvoted 1 times

🗨️ 👤 **282b85d** 1 year, 1 month ago

To ensure read-write access to a dataset (DS1) using the XMLA endpoint in your Fabric tenant, you need to modify the settings that control XMLA endpoint access. This typically involves enabling read-write capabilities for the XMLA endpoint at the capacity level. Therefore, you should modify:

C. the C1 settings

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: C**

IMHO,

the answer is C. As described here: <https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools#enable-xmla-read-write>

in the section "To enable read-write for a Premium capacity"

upvoted 1 times

🗨️ 👤 **VAzureD** 1 year, 2 months ago

**Selected Answer: D**

we need to first configure tenant settings:

Allow XMLA endpoints and Analyze in Excel with on-premises semantic models:

upvoted 2 times

🗨️ 👤 **manolet** 1 year, 1 month ago

Yes, but the question refers to enabling read-write to the capacity.

upvoted 2 times

🗨️ 👤 **emmanuelkech** 1 year, 2 months ago

**Selected Answer: B**

To ensure read-write access to the dataset DS1 via the XMLA endpoint, you should modify the settings at the workspace level (WS1). Specifically, enable read-write operations for the XMLA endpoint within the workspace configuration.

upvoted 1 times

🗨️ 👤 **manolet** 1 year, 1 month ago

No. You are wrong.

upvoted 2 times

🗨️ 👤 **thuss** 1 year, 3 months ago

Again, bad wording in the question. If it is specifically about read-write, it's C. If it's about having XMLA endpoints available in the first place, it's D. Judging by Microsoft's wordings and ways of referring to their own documentation, I guess they want us to go for C.

upvoted 4 times

🗨️ 👤 **metiii** 1 year, 3 months ago

**Selected Answer: D**

We need to configure several settings but in terms of hierarchy we need to first configure tenant settings:

Allow XMLA endpoints and Analyze in Excel with on-premises semantic models: Users in the organization can use Excel to view and interact with on-premises Power BI semantic models. This also allows connections to XMLA endpoints.

<https://learn.microsoft.com/en-us/fabric/admin/tenant-settings-index>

upvoted 2 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

C is the right answer. The read/write permission is set in capacity settings in Admin Portal.

upvoted 1 times

You have a Fabric tenant that contains a workspace named Workspace1. Workspace1 is assigned to a Fabric capacity.

You need to recommend a solution to provide users with the ability to create and publish custom Direct Lake semantic models by using external tools. The solution must follow the principle of least privilege.

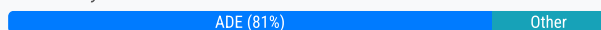
Which three actions in the Fabric Admin portal should you include in the recommendation? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. From the Tenant settings, set Allow XMLA Endpoints and Analyze in Excel with on-premises datasets to Enabled.
- B. From the Tenant settings, set Allow Azure Active Directory guest users to access Microsoft Fabric to Enabled.
- C. From the Tenant settings, select Users can edit data model in the Power BI service.
- D. From the Capacity settings, set XMLA Endpoint to Read Write.
- E. From the Tenant settings, set Users can create Fabric items to Enabled.
- F. From the Tenant settings, enable Publish to Web.

**Suggested Answer:** ACD

Community vote distribution



**sraakesh95** 1 year, 4 months ago

D - The XMLA endpoint needs to be enabled from the capacity settings as it is crucial for allowing external tools to only read, but also, publish and manage custom Direct Lake semantic models.

A - Again from Tenant settings, we need to enable XMLA endpoints ; This is to ensure that external tools interact with the data models as and when necessary within the tenant's workspace ; The analyze in excel is just complimentary to this setting in Fabric and is irrelevant

E - Sets appropriate permissions to users by allowing them to edit models and publish them as when required abiding by the principle of least privilege.

B&C have concerns with the principle of least privilege and security concerns while publishing to the Web  
upvoted 31 times

**282b85d** 1 year ago

**Selected Answer:** ADE

A. Enabling XMLA Endpoints is crucial because it allows external tools (like SQL Server Management Studio, Tabular Editor, or any other tool that can connect via XMLA) to connect to Power BI datasets. This is essential for users who need to create, modify, or manage semantic models using these external tools.

D. Setting the XMLA Endpoint to Read Write allows users not only to read the data from the Power BI service but also to write back to it. This is necessary for users to create and publish custom Direct Lake semantic models. Without write access, users would be unable to publish or update their models, which is a critical part of managing semantic models.

E. Enabling users to create Fabric items is essential because it allows users to create new Power BI datasets, dataflows, and other necessary components within the workspace.

upvoted 11 times

**NRezgui** 6 months, 1 week ago

**Selected Answer:** ADE

D - The XMLA endpoint needs to be enabled from the capacity settings as it is crucial for allowing external tools to only read, but also, publish and manage custom Direct Lake semantic models.

upvoted 2 times

**Rakesh16** 7 months, 2 weeks ago

**Selected Answer:** ADE

ADE is the answer

upvoted 2 times

**jass007\_k** 8 months ago

A- From tenant settings, setting allow XMLA endpoints and Analyze in excel with on-premises datasets to enabled

D- From capacity settings, set XMLA endpoint to ReadWrite

E- From Tenant settings, setUsers can create fabric items within workspace

upvoted 1 times

🗨️ 👤 **nyoike** 10 months, 3 weeks ago

**Selected Answer: ADE**

C is not correct for those voting for it. Below is the (verbatim) description for the "Users can edit data models in the Power BI service (preview)" Tenant setting in the Admin portal.

"Turn on this setting to allow users to edit data models in the service. This setting DOESN'T apply to DirectLake semantic models or editing a semantic model through an API or XMLA endpoint."

More information on that Tenant setting from the link below ....

<https://learn.microsoft.com/en-us/power-bi/transform-model/service-edit-data-models#enabling-data-model-editing-in-the-admin-portal>

upvoted 2 times

🗨️ 👤 **Ahmadpbi** 11 months, 1 week ago

C is wrong for 2 reasons, this setting is not in the tenant, it is in workspace settings. The second reason is that this feature is still in preview till date of my comment.

upvoted 1 times

🗨️ 👤 **Subhamssg13** 1 year ago

ADE

OPTION C should not be consider as :

Users can edit data models in the Power BI service (preview)

Enabled for the entire organization

Turn on this setting to allow users to edit data models in the service. This setting doesn't apply to DirectLake semantic models or editing a semantic model through an API or XMLA endpoint.

upvoted 2 times

🗨️ 👤 **rlo123** 1 year, 1 month ago

ADE

B. From the Tenant settings, set Allow Azure Active Directory guest users to access Microsoft Fabric to Enabled. This focuses on external user access, which might not be necessary for your scenario.

C. From the Tenant settings, select Users can edit data model in the Power BI service. This primarily affects in-browser editing of Power BI models and is less relevant to creating custom models with external tools.

F. From the Tenant settings, enable Publish to Web. This is unrelated to external tool development and is concerned with sharing Power BI reports publicly.

upvoted 2 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: ADE**

A -> D -> E

upvoted 2 times

🗨️ 👤 **hello2tomoki** 1 year, 2 months ago

**Selected Answer: ACD**

A. From the Tenant settings, set Allow XMLA Endpoints and Analyze in Excel with on-premises datasets to Enabled (<https://learn.microsoft.com/en-us/fabric/admin/service-admin-portal-integration>) . C. From the Tenant settings, select Users can edit data model in the Power BI service (<https://learn.microsoft.com/en-us/fabric/admin/service-admin-portal-data-model>) . D. From the Capacity settings, set XMLA Endpoint to Read Write (<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools>) .

upvoted 1 times

🗨️ 👤 **Priyanka007** 1 year ago

From the link you <https://learn.microsoft.com/en-us/fabric/admin/service-admin-portal-data-model> provided - It clearly says that "This setting doesn't apply to DirectLake datasets or editing a dataset through an API or XMLA endpoint." so it is not C.

upvoted 2 times

🗨️ 👤 **PazaBlandData** 1 year, 2 months ago

**Selected Answer: ADE**

AD for sure

<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-connect-tools#security>

upvoted 2 times

  **a61d298** 1 year, 2 months ago

A Allow XMLA Endpoints and Analyze in Excel with on-premises datasets:

From the Tenant settings, set Allow XMLA Endpoints and Analyze in Excel with on-premises datasets to Enabled. This allows users to interact with the dataset via XMLA endpoints and analyze data using Excel.

D. Set XMLA Endpoint to Read Write:

In the Capacity settings, configure the XMLA Endpoint to Read Write. This ensures that users have the necessary permissions to create and modify semantic models through external tools.

E Enable Users to Create Fabric Items:

From the Tenant settings, set Users can create Fabric items to Enabled. This grants users the ability to create custom semantic models within the Fabric workspace.

upvoted 2 times

  **kelvin3105** 1 year, 2 months ago

**Selected Answer: ADE**

sraakesh95 Highly Voted 1 month, 3 weeks ago

D - The XMLA endpoint needs to be enabled from the capacity settings as it is crucial for allowing external tools to only read, but also, publish and manage custom Direct Lake semantic models.

A - Again from Tenant settings, we need to enable XMLA endpoints ; This is to ensure that external tools interact with the data models as and when necessary within the tenant's workspace ; The analyze in excel is just complimentary to this setting in Fabric and is irrelevant

E - Sets appropriate permissions to users by allowing them to edit models and publish them as when required abiding by the principle of least privilege.

B&C have concerns with the principle of least privilege and security concerns while publishing to the Web

upvoted 8 times

JUST TO VOTE GUYS

upvoted 3 times

  **PCCCCC** 1 year, 2 months ago

**Selected Answer: ADE**

Not C : <https://learn.microsoft.com/en-us/power-bi/transform-model/service-edit-data-models#enabling-data-model-editing-in-the-admin-portal>

upvoted 1 times

  **a\_51** 1 year, 3 months ago

**Selected Answer: ADE**

ADE seems most appropriate

upvoted 1 times

  **azure\_bimonster** 1 year, 3 months ago

**Selected Answer: ADE**

Most probably ADE, I put E because users have the necessary permissions to create custom Direct Lake semantic models within the Fabric workspace.

upvoted 2 times



You are creating a semantic model in Microsoft Power BI Desktop.

You plan to make bulk changes to the model by using the Tabular Model Definition Language (TMDL) extension for Microsoft Visual Studio Code.

You need to save the semantic model to a file.

Which file format should you use?

- A. PBIP
- B. PBIX
- C. PBIT
- D. PBIDS

**Suggested Answer: B**

Community vote distribution

A (84%)

C (16%)

 **mtroyano** Highly Voted 1 year, 3 months ago

**Selected Answer: A**

The correct option is PBIP - <https://powerbiblogscdn.azureedge.net/wp-content/uploads/2024/02/tmdlPreviewFeature.png>

upvoted 23 times

 **Shiven** Highly Voted 1 year, 1 month ago

The correct file format for this purpose is:

A. PBIP

Explanation:

PBIP (Power BI Project): This format is specifically designed for managing Power BI projects, including semantic models, in a way that supports bulk editing and version control. It allows you to use tools like Visual Studio Code to edit the model's metadata directly.

Other formats:

PBIX (Power BI Desktop File): This is the standard file format for Power BI Desktop reports, but it is not designed for direct bulk editing using TMDL.

PBIT (Power BI Template File): This format is used for Power BI templates, which are useful for creating new reports based on a predefined structure, but it is not suitable for bulk editing using TMDL.

PBIDS (Power BI Data Source File): This format is used for defining data sources for Power BI reports, not for semantic models.


upvoted 21 times

 **NRezgui** Most Recent 6 months ago

**Selected Answer: A**

The correct option is PBIP

upvoted 1 times

 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: A**

PBIP is the answer

upvoted 1 times

 **KipngenhVinnie** 9 months, 2 weeks ago

To make bulk changes using the Tabular Model Definition Language (TMDL) extension is A. PBIP.

Explanation:

PBIP (Power BI Project) is a file format that supports the open-source TMDL format and is designed for integrating Power BI with external development environments like Visual Studio Code.

PBIX is the common Power BI report file format but is not intended for bulk edits through TMDL.

PBIT is a Power BI template file, used for creating new reports based on an existing structure but not for bulk editing in Visual Studio Code. PBIDS is for creating Power BI dataset connections and is unrelated to TMDL editing.

upvoted 2 times

🗨️ 👤 **woliveiras** 1 year ago

**Selected Answer: A**

PIBP is the correct one

upvoted 2 times

🗨️ 👤 **dev2dev** 1 year, 1 month ago

**Selected Answer: A**

PBIP for source code

upvoted 1 times

🗨️ 👤 **kkl03** 1 year, 1 month ago

**Selected Answer: A**

<https://powerbi.microsoft.com/en-us/blog/tmdl-in-power-bi-desktop-developer-mode-preview/>

upvoted 1 times

🗨️ 👤 **David\_Webb** 1 year, 1 month ago

**Selected Answer: A**

<https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-overview>

upvoted 2 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO,

I go with "A".

Here: <https://powerbi.microsoft.com/en-us/blog/tmdl-in-power-bi-desktop-developer-mode-preview/>

=====

Saving as a PBIP using TMDL is currently in preview. Before giving it a try, you must first enable this feature in Preview features: go to File > Options and settings > Options > Preview features and check the box next to "Store semantic model using TMDL format".

=====

upvoted 2 times

🗨️ 👤 **rlo123** 1 year, 1 month ago

C - PBIT

Here's why:

PBIT: The PBIT format is designed specifically for storing the semantic model definition in a way that is compatible with the Tabular Model Definition Language (TMDL) and readily editable in Visual Studio Code using the TMDL extension.

Let's clarify the other options:

PBIP, PBIX: These are standard Power BI Desktop file formats. They contain the semantic model but also include data, reports, visualizations, and other elements of a Power BI project.

PBIDS: This file format is related to Power BI datasets hosted in the Power BI service, not for local editing with TMDL.

upvoted 1 times

🗨️ 👤 **BennyBenz** 1 year, 1 month ago

But it is the semantic model we are focusing on here?

PBIX may be correct then?

<https://learn.microsoft.com/en-us/power-bi/create-reports/service-export-to-pbix#download-a-pbix-file-from-a-semantic-model>

upvoted 1 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

**Selected Answer: A**

A <https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-overview>

upvoted 3 times



🗨️ 👤 **sraakesh95** 1 year, 4 months ago

**Selected Answer: A**

<https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-overview>

Checkout the title in the above link: Programmatic generation and editing artifact definitions

upvoted 2 times

  **thuss** 1 year, 4 months ago

**Selected Answer: A**

Also think it's A: <https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-overview>

upvoted 1 times

  **Jeff\_Zhu** 1 year, 4 months ago

**Selected Answer: A**

The answer is A

The PBIP will create one file and two folders, PBIP.Dataset contains definition folder that is use to host the .tmdl files

upvoted 3 times

  **lengzhai** 1 year, 4 months ago

**Selected Answer: C**

<https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-templates>

upvoted 3 times

## HOTSPOT -

You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains three schemas named schemaA, schemaB, and schemaC.

You need to ensure that a user named User1 can truncate tables in schemaA only.

How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

GRANT   
CONNECT  
EXECUTE ON   
OBJECT::schemaA  
SCHEMA::schemaA TO User1;

**Answer Area**

Suggested Answer: GRANT   
CONNECT  
EXECUTE ON   
OBJECT::schemaA  
SCHEMA::schemaA TO User1;

**estrelle2008** Highly Voted 1 year, 4 months ago

ALTER/SCHEMA

This statement allows to alter (which includes truncating) tables within the specified schema. It ensures that the permission is restricted to schemaA and does not grant access to other schemas or objects.

upvoted 32 times

**KipngenohVinnie** Highly Voted 9 months, 2 weeks ago

GRANT ALTER ON SCHEMA::schemaA TO User1;

upvoted 6 times

**Rakesh16** Most Recent 7 months, 2 weeks ago

Alter, Schema\_SchemaA

upvoted 1 times

**py2356863** 1 year ago

ALTER/SCHEMA

<https://learn.microsoft.com/en-us/sql/t-sql/statements/truncate-table-transact-sql?view=sql-server-ver16#permissions>

<https://learn.microsoft.com/th-th/sql/t-sql/statements/alter-schema-transact-sql?view=fabric&preserve-view=true>

upvoted 1 times

**Valcon\_doo\_NoviSad** 1 year ago

truncating tables is not even supported in warehouse so this question is pointless, good job microsoft...

upvoted 5 times

**waltXc** 3 months, 1 week ago

TRUNCATE is now supported on Fabric's Data Warehouse:

<https://community.fabric.microsoft.com/t5/Fabric-Ideas/Support-TRUNCATE-TABLE-in-Fabric-Warehouse/idi-p/4510865>

upvoted 1 times

**stilferx** 1 year, 1 month ago

IMHO,

ALTER -> SCHEMA.

As well said below,

1) ALTER is the thing to have for truncating:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/truncate-table-transact-sql?view=sql-server-ver16&viewFallbackFrom=fabric>

2) SCHEMA:schema\_name to define a particular schema for tables

upvoted 1 times

🗨️ 👤 **VAzureD** 1 year, 2 months ago

ALTER/SCHEMA

<https://learn.microsoft.com/en-us/sql/t-sql/statements/grant-schema-permissions-transact-sql?view=sql-server-ver16>

GRANT permission [ ,...n ] ON SCHEMA :: schema\_name

TO database\_principal [ ,...n ]

[ WITH GRANT OPTION ]

[ AS granting\_principal ]

upvoted 4 times

🗨️ 👤 **cresclux** 1 year, 3 months ago

According to <https://learn.microsoft.com/en-us/sql/t-sql/statements/grant-object-permissions-transact-sql?view=sql-server-ver16>, answer seems to ALTER and OBJECT.

upvoted 2 times

🗨️ 👤 **Priyanka007** 1 year ago

It would have worked if the question said one table.. But it says "user named User1 can truncate TABLE"S" in schemaA only" so I would go with option C. SCHEMA.

upvoted 1 times

🗨️ 👤 **benja84** 1 year, 3 months ago

For schemas:

<https://learn.microsoft.com/en-us/sql/t-sql/statements/grant-schema-permissions-transact-sql?view=sql-server-ver16>

upvoted 2 times

🗨️ 👤 **AzureGeek79** 1 year, 4 months ago

ALTER is DDL not DML so the right answer is EXECUTE SCHEMA.

upvoted 2 times

🗨️ 👤 **dev2dev** 1 year, 1 month ago

TRUNCATE is DML

upvoted 2 times

🗨️ 👤 **dev2dev** 1 year, 1 month ago

Ignore my previous comment.

TRUNCATE is DDL so ALTER is what we need

upvoted 1 times

🗨️ 👤 **thomaski123** 1 year, 3 months ago

ALTER is required for Truncate <https://learn.microsoft.com/en-us/sql/t-sql/statements/truncate-table-transact-sql?view=sql-server-ver16&viewFallbackFrom=fabric>

upvoted 4 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

Correct

upvoted 2 times

You plan to deploy Microsoft Power BI items by using Fabric deployment pipelines. You have a deployment pipeline that contains three stages named Development, Test, and Production. A workspace is assigned to each stage.

You need to provide Power BI developers with access to the pipeline. The solution must meet the following requirements:

Ensure that the developers can deploy items to the workspaces for Development and Test.

Prevent the developers from deploying items to the workspace for Production.

Follow the principle of least privilege.

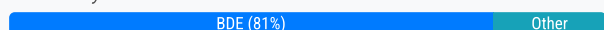
Which three levels of access should you assign to the developers? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Build permission to the production semantic models
- B. Admin access to the deployment pipeline
- C. Viewer access to the Development and Test workspaces
- D. Viewer access to the Production workspace
- E. Contributor access to the Development and Test workspaces
- F. Contributor access to the Production workspace

**Suggested Answer:** ADE

Community vote distribution



**sraakesh95** Highly Voted 1 year, 4 months ago

**Selected Answer:** BDE

As pointed out by XiltroX,

B - Admin access is provided to the developers for the developers to manage the deployment process across the various stages (in this case Dev and Test). This is a basic necessary.

D - To restrict the access on the Production workspace, provide an overriding Viewer access which lets the developers only view the Production environment and not make any changes.

E - This is to provide the developers with the permissions to develop, edit and update the Dev and Test pipelines.

upvoted 26 times

**elma\_qhor\_19** Most Recent 4 months ago

**Selected Answer:** BDE

B - Admin access is provided to the developers for the developers to manage the deployment process across the various stages (in this case Dev and Test). This is a basic necessary.

D - To restrict the access on the Production workspace, provide an overriding Viewer access which lets the developers only view the Production environment and not make any changes.

E - This is to provide the developers with the permissions to develop, edit and update the Dev and Test pipelines.

upvoted 1 times

**NRezgui** 6 months ago

**Selected Answer:** BDE

B - Admin access is provided to the developers for the developers to manage the deployment process across the various stages (in this case Dev and Test). This is a basic necessary.

D - To restrict the access on the Production workspace, provide an overriding Viewer access which lets the developers only view the Production environment and not make any changes.

E - This is to provide the developers with the permissions to develop, edit and update the Dev and Test pipelines.

upvoted 1 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: BDE**

B Admin access to the deployment pipeline

D Viewer access to the Production workspace

E Contributor access to the Development and Test workspaces

upvoted 1 times

🗳️ 👤 **KipngenohVinnie** 9 months, 2 weeks ago

To meet the requirements while following the principle of least privilege, you should assign the following levels of access to the developers:

B. Admin access to the deployment pipeline - This is needed so developers can manage the deployment pipeline itself, including moving items between stages.

E. Contributor access to the Development and Test workspaces - This allows developers to deploy and make changes to the items in the Development and Test workspaces.

D. Viewer access to the Production workspace - This ensures developers can view the Production workspace but cannot make any changes or deploy items to it.

upvoted 1 times

🗳️ 👤 **user12345678** 10 months ago

**Selected Answer: BDE**

B - There's only one role for pipelines so it's either having an Admin role or no access to a pipeline. You have to have an Admin role on a pipeline in order to do any deployments.

D - Pipeline permissions and workspace permissions are kind of linked but the way it works is if a user has viewer permissions on a workspace and Admin on the Pipeline - They won't be able to do a deployment to that workspace, but they can still see things inside of it (as is the case as a viewer in any workspace).

E - If you have the Contributor/Member/Admin role on a workspace AND Admin permissions on a pipeline, then you can deploy to that workspace.

upvoted 1 times

🗳️ 👤 **user12345678** 10 months ago

One common misconception is people are thinking (and I initially thought) that D is selected to limit the user. This isn't the case. The workspace roles are separate and if you don't give any role to the Prod workspace, then the user won't be able to access it at all. By giving viewer, they can at least see what's in the workspace.

upvoted 1 times

🗳️ 👤 **bad4b76** 10 months, 1 week ago

It does not say Developers need to manage deployments, just access items within the pipeline. So A, D, and E

upvoted 1 times

🗳️ 👤 **Ahmadpbi** 11 months, 2 weeks ago

A D E because: In Power BI, having Viewer access to a workspace does not automatically grant Build permission for a semantic model within that workspace. The Viewer role allows a user to view and interact with items in the workspace, but it does not include the ability to create new content or edit existing content.

To have Build permission, a user must be explicitly granted that permission, which allows them to build new content from the semantic model, access reports that use composite models on Power BI Pro workspaces, and pull data into Analyze in Excel, among other capabilities. This permission can be given by the workspace Admin or Member who has the authority to manage semantic model permissions

upvoted 1 times

🗳️ 👤 **semauni** 8 months ago

Agreed, but do they need build permissions?

upvoted 1 times

🗳️ 👤 **bmc15** 1 year ago

**Selected Answer: BDE**

B: It's admin or nothing for pipeline access.

D: Viewer, so they cannot create it in production.

E: You need Contributor on the workspaces you want to deploy and create.

upvoted 1 times

🗨️ 👤 **SilvanoRamalho** 1 year ago

Selected Answer: BDE

Respostas BDE

upvoted 1 times

🗨️ 👤 **Jons123son** 1 year ago

Was in exam. Scored 95%

Chose BDE

Answers switctied Contributor for Member role

upvoted 1 times

🗨️ 👤 **Gerald** 1 year ago

I agree with BDE

upvoted 1 times

🗨️ 👤 **282b85d** 1 year, 1 month ago

E. Contributor access to the Development and Test workspaces:

This allows developers to deploy and manage content in the Development and Test workspaces, meeting the requirement to allow deployments in these stages.

D. Viewer access to the Production workspace:

This provides developers with read-only access to the Production workspace, ensuring they can view content but cannot deploy or make changes, which aligns with the requirement to prevent deployments to Production.

B. Admin access to the deployment pipeline:

This allows developers to manage the deployment pipeline itself, including deploying items to the Development and Test stages but not to the Production stage. This ensures they can oversee the deployment process without overstepping into the Production environment.

upvoted 2 times

🗨️ 👤 **David\_Webb** 1 year, 1 month ago

Selected Answer: BDE

Admin access is needed in the development workspace for pipeline.

upvoted 2 times

🗨️ 👤 **72bd3bc** 1 year, 1 month ago

To deploy from one stage to another in the pipeline, you must be a pipeline admin, and either a member or an admin of the workspaces assigned to the stages involved. For example, a pipeline admin that isn't assigned a workspace role, can view the pipeline and share it with others. However, this user can't view the content of the workspace in the pipeline, or in the service, and can't perform deployments.

<https://learn.microsoft.com/en-us/fabric/cicd/deployment-pipelines/understand-the-deployment-process#permissions/?azure-portal=true>

From task: Prevent the developers from deploying to the production.

So B is not correct, we can't give pipeline admin access?

upvoted 1 times

🗨️ 👤 **72bd3bc** 1 year, 1 month ago

we can and we need viewer access to the Production, so that developers can't deploy at this stage. Without admin access they can't deploy from dev to test stages.

upvoted 2 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

Selected Answer: BDE

IMHO,

B->D->E

B. Admin access to the deployment pipeline - to be able to run

D. Viewer access to the Production workspace - to not be able to run Prod

E. Contributor access to the Development and Test workspaces - to be able to run Dev/Test

upvoted 2 times

🗨️ 👤 **72bd3bc** 1 year, 1 month ago



I see that contributors aren't able to deploy. Only members

Workspace contributor  
(and pipeline admin)  
Consume content  
Compare stages  
View semantic models  
Unassign a workspace from a stage

Workspace member  
(and pipeline admin)  
View workspace content  
Compare stages  
Deploy items (must be a member or admin of both source and target workspaces)  
Update semantic models  
Unassign a workspace from a stage  
Configure semantic model rules (you must be the semantic model owner)

<https://learn.microsoft.com/en-us/fabric/cicd/deployment-pipelines/understand-the-deployment-process#permissions-table>  
upvoted 3 times

  **Martin\_Nbg** 11 months, 1 week ago

So what does this mean referring to the auswärts? Do we have 3 correct answers at all? Obviously ist's Not Contributor to Test (Contributor to Dev would be enough because you only want to deploy FROM Dev but not TO Dev).  
upvoted 1 times

  **Mhmod48** 1 year, 1 month ago

I like your answers  
upvoted 1 times

  **PazaBlandData** 1 year, 2 months ago

**Selected Answer: BD**

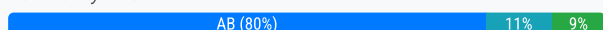
Member or Admin rights on the workspace level are required to deploy datasets so E doesn't match the requirements.  
<https://learn.microsoft.com/en-us/fabric/cicd/deployment-pipelines/understand-the-deployment-process#permissions-table>  
upvoted 2 times

You have a Fabric workspace that contains a DirectQuery semantic model. The model queries a data source that has 500 million rows. You have a Microsoft Power BI report named Report1 that uses the model. Report1 contains visuals on multiple pages. You need to reduce the query execution time for the visuals on all the pages. What are two features that you can use? Each correct answer presents a complete solution, NOTE: Each correct answer is worth one point.

- A. user-defined aggregations
- B. automatic aggregation
- C. query caching
- D. OneLake integration

**Suggested Answer: BD**

Community vote distribution



**sraakesh95** Highly Voted 1 year, 4 months ago

**Selected Answer: AB**

Agree with lengzhai's reference of the 2 links:

A - Custom aggregations enables PBI to not perform a Full Scan of the underlying datasets.

B - The AutoAggregations feature automatically creates aggregations on large datasets and based on query optimization determines the total number of rows that requires processing based on the generated query plan.

Incorrect to this question context:

C - Although caching helps improve performance on large datasets, it doesn't support DirectQuery (Important note in <https://learn.microsoft.com/en-us/power-bi/connect-data/power-bi-query-caching>) ; Also, it is a feature available in PBI Service that is automatic and needs no intervention from the user.

upvoted 19 times

**Momoanwar** Highly Voted 1 year, 4 months ago

**Selected Answer: AB**

D: onelake integration not for Direct Query

C: only at loading for first page

So AV

upvoted 11 times

**BrandonPerks** 1 year, 4 months ago

Agreed to AB.

Both UDA's and AA optimize direct query performance. One just requires more manual work and in depth knowledge data modelling and query optimization techniques (UDA), whereas the other makes simplifies this process through the use of ML algorithms (AA).

upvoted 3 times

**Momoanwar** 1 year, 4 months ago

I mean AB\*

upvoted 1 times

**NRezgui** Most Recent 6 months ago

**Selected Answer: AB**

A. user-defined aggregations

B. automatic aggregation

upvoted 1 times

**AshwiniVivek** 6 months, 2 weeks ago

**Selected Answer: AC**

User-defined aggregations (A):

User-defined aggregations allow you to create pre-aggregated tables that summarize data at a higher level, reducing the volume of data that needs to be queried for specific visuals. This can significantly improve performance for common queries by avoiding the need to scan the entire dataset.

Query caching (C):

Query caching stores the results of previous queries for reuse. When users interact with visuals that require similar data, the cached results can be returned more quickly than querying the source again, thereby reducing execution time for those visuals.

upvoted 2 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: AB**

A & B is the answer

upvoted 2 times

🗳️ 👤 **6d1de25** 11 months, 3 weeks ago

**Selected Answer: BD**

B&D are correct.

Direct Lakes are great for performance in the OneLake integration

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview>

upvoted 1 times

🗳️ 👤 **TimoRii** 8 months, 2 weeks ago

This was DirectQuery, not Direct Lake.

upvoted 1 times

🗳️ 👤 **282b85d** 1 year, 1 month ago

A&B

While query caching can be beneficial in certain scenarios, user-defined aggregations and automatic aggregations are typically more effective for improving query performance in Power BI reports with large datasets and complex queries. These methods reduce the volume of data processed in real-time queries, directly addressing the performance bottlenecks associated with querying large datasets.

upvoted 1 times

🗳️ 👤 **Murtaza\_007** 1 year, 1 month ago

CHATGPT saya AC

upvoted 1 times

🗳️ 👤 **Jane5** 3 months, 3 weeks ago

no, it says AB. I go for AB

upvoted 1 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: AB**

IMHO, A & B looks good

upvoted 1 times

🗳️ 👤 **lengzhai** 1 year, 4 months ago

**Selected Answer: AB**

Agree with A B

<https://learn.microsoft.com/en-us/power-bi/transform-model/aggregations-advanced>

<https://learn.microsoft.com/en-us/power-bi/enterprise/aggregations-auto>

upvoted 6 times

🗳️ 👤 **estrelle2008** 1 year, 4 months ago

**Selected Answer: AB**

Agreed AB.

Although query caching (C) will reduce query execution time too, you risk outdated cached results when working with real-time or dynamic data.

upvoted 2 times

🗳️ 👤 **Fermd** 1 year, 4 months ago

**Selected Answer: AC**

A. User-defined aggregations (UDAs) allow you to pre-aggregate specific calculations directly in the semantic model. This reduces the amount of data that needs to be retrieved from the source each time a visual requires the calculation, significantly improving query execution time.

C. Power BI Desktop enables query caching for DirectQuery models. This stores frequently used queries on the client machine, eliminating the need to re-send them to the source data for subsequent interactions.

upvoted 4 times

🗨️ 👤 **STH** 1 year, 3 months ago

Question is about Fabric workspace... not Power BI Desktop !

upvoted 2 times

🗨️ 👤 **Nicofr** 1 year, 4 months ago

**Selected Answer: BD**

<https://learn.microsoft.com/en-us/power-bi/enterprise/aggregations-auto>

<https://learn.microsoft.com/en-us/power-bi/enterprise/onelake-integration-overview>

upvoted 2 times

You have a Fabric tenant that contains 30 CSV files in OneLake. The files are updated daily.

You create a Microsoft Power BI semantic model named Model1 that uses the CSV files as a data source. You configure incremental refresh for Model1 and publish the model to a Premium capacity in the Fabric tenant.

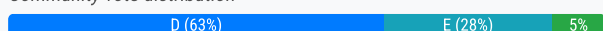
When you initiate a refresh of Model1, the refresh fails after running out of resources.

What is a possible cause of the failure?

- A. Query folding is occurring.
- B. Only refresh complete days is selected.
- C. XMLA Endpoint is set to Read Only.
- D. Query folding is NOT occurring.
- E. The delta type of the column used to partition the data has changed.

**Suggested Answer: D**

Community vote distribution



282b85d Highly Voted 1 year, 1 month ago

Selected Answer: D

D. Query folding is NOT occurring.

Query folding refers to the ability of Power Query to push data transformation logic back to the data source, which can perform the transformations more efficiently. When query folding does not occur, all the data is pulled into Power BI and transformations are applied locally, which can be resource-intensive and lead to running out of resources, especially with large datasets like your 30 CSV files.

E. The delta type of the column used to partition the data has changed: While this could cause issues with incremental refresh accuracy, it would not typically result in "running out of resources" during the refresh.

upvoted 21 times

Nefirs Highly Voted 1 year, 3 months ago

Selected Answer: E

i don't really know.

But A and D regarding query folding seem for me not to be relevant since query folding does not make sense anyway on CSV/Flatfile connections. In my understanding, query folding sends a query back to the source. And what kind of query would that be to a CSV source?

upvoted 15 times

Plb2 1 year ago

Don't think so too

<https://learn.microsoft.com/en-us/power-query/query-folding-examples#no-query-folding-example>

upvoted 4 times

mishravikas7 Most Recent 4 weeks, 1 day ago

Selected Answer: D

The key word is here running out of resources, which is why correct option is D as query folding not occurring causes power query to compute everything by itself.

Data type if changes, will cause error but not time out

upvoted 1 times

PradeepikaGadai 1 month ago

Selected Answer: D

query folding does not occur, Power BI must load all data into memory, apply filters after loading, and then perform transformations. This consumes a large amount of resources and may cause refresh failures, especially as the number of files or rows increases

upvoted 1 times

Ramy216 1 month, 1 week ago

Selected Answer: E

The answer is E.

Because CSV does not support Query Folding.

upvoted 1 times

🗳️ 👤 **kilowd** 3 months, 2 weeks ago

**Selected Answer: E**

Answer is E : <https://learn.microsoft.com/en-us/power-query/query-folding-examples#no-query-folding-example>

upvoted 1 times

🗳️ 👤 **kilowd** 3 months, 1 week ago

No, query folding does not occur when working with CSV files in OneLake from Power BI. Here's why:

1. Query Folding Requires a Supported Source

- Query folding happens when Power BI pushes transformations back to the data source (e.g., SQL Server, Snowflake).
- CSV files do not support query folding because they are flat files, not a database with an engine to process queries.

2. OneLake Stores Data but Doesn't Process Queries

- OneLake is a storage layer in Microsoft Fabric, not a query engine.
- When Power BI connects to a CSV in OneLake, it downloads the data and processes transformations in Power Query (M Engine)—not at the source.

3. Workaround for Performance

- If you need query folding, consider loading the CSV into a Lakehouse or Warehouse in Fabric and querying it using DirectQuery or SQL-based sources.

upvoted 1 times

🗳️ 👤 **sajjuh** 3 months, 2 weeks ago

**Selected Answer: E**

data type issue becoz PBI supports for lakehouse source

upvoted 1 times

🗳️ 👤 **Devoteam2025** 4 months ago

**Selected Answer: E**

The correct answer is "E." The answer is written incorrectly, they meant to say "data type" instead of "delta type," which is the reason for running out of resources.

upvoted 2 times

🗳️ 👤 **VLADCS** 5 months, 2 weeks ago

**Selected Answer: D**

Answer D.

If query folding is not supported, all transformations are performed in Power BI memory. This can cause excessive resource usage, especially for large data sources such as 30 CSV files. This is the most likely cause of the error.

Answer E is not correct. If the partitioning column type has changed, it will cause an update error, but not a resource overload. The error will be more related to data incompatibility.

upvoted 1 times

🗳️ 👤 **NRezgui** 6 months ago

**Selected Answer: E**

The delta type of the column used to partition the data has changed.

upvoted 2 times

🗳️ 👤 **pirate84** 5 months ago

I took my exam the last week, and the option "E", the text was: "The data type of the column used...."

For this, is "E" --> Correct

upvoted 1 times

🗳️ 👤 **rkandathil** 6 months, 3 weeks ago

**Selected Answer: D**

D. Query folding is NOT occurring

Query folding happens when Power BI sends this list (your transformations) to the database, so it does the heavy lifting for you. If folding doesn't happen, Power BI has to handle all the processing on its own, which is slower and less efficient.

upvoted 1 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: D**

Query folding is NOT occurring.

upvoted 1 times

🗨️ 👤 **jcu614** 8 months ago

D.

While query folding typically does not apply to CSV files, OneLake does provide some structure that can prevent loading all data into PowerBI Memory.

upvoted 1 times

🗨️ 👤 **jass007\_k** 8 months ago

Correct Option is D) If query folding is not occurring, it means that all 30 CSV files are likely being loaded into memory for processing, which can lead to resource exhaustion, especially if the files are large or numerous.

A. Query folding is occurring: If query folding were happening, it would typically improve performance and reduce resource usage, not cause a failure.

B. Only refresh complete days is selected: This option usually helps in managing the data being refreshed by only focusing on completed data, potentially alleviating resource issues rather than causing them.

C. XMLA Endpoint is set to Read Only: If the XMLA Endpoint were set to read-only, it wouldn't directly cause a refresh failure due to resource issues. It would restrict write operations but not necessarily impact resource allocation during a refresh.

E. The delta type of the column used to partition the data has changed: Changes in partitioning columns could cause refresh issues, but they wouldn't inherently lead to resource exhaustion unless they also disrupt query folding.

upvoted 2 times

🗨️ 👤 **Jons123son** 1 year ago

Was in exam. Scored 95%

Chose D.

Honestly, I was guessing. No clue. However, query folding does occur for CSV files stored IN OneLake. OneLake does the work. This different from semantic models created with files stored on a normal machine and what had been common knowledge for a Power BI user.

upvoted 8 times

🗨️ 👤 **haran939** 1 year, 1 month ago

**Selected Answer: D**

Answer D

upvoted 2 times

🗨️ 👤 **PiyushT** 1 year, 1 month ago

**Selected Answer: D**

For efficient data processing, Power BI aims to push as much of the filtering and calculations as possible to the source system (OneLake in this case). This is called query folding. When query folding fails, Power BI needs to pull all the raw data into the semantic model and perform operations there, increasing memory and processing strain.

E This could lead to refresh errors but is less likely to cause the specific behavior of running out of resources.

upvoted 5 times

You have a Fabric tenant that uses a Microsoft Power BI Premium capacity.  
You need to enable scale-out for a semantic model.  
What should you do first?

- A. At the semantic model level, set Large dataset storage format to Off.
- B. At the tenant level, set Create and use Metrics to Enabled.
- C. At the semantic model level, set Large dataset storage format to On.
- D. At the tenant level, set Data Activator to Enabled.

**Suggested Answer: C**

Community vote distribution

C (100%)

🗳️ 👤 **Nicofr** Highly Voted 👍 1 year, 4 months ago

**Selected Answer: C**

<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-scale-out-configure>  
upvoted 17 times

🗳️ 👤 **kaixin** Most Recent 🕒 5 months, 2 weeks ago

**Selected Answer: C**

<https://learn.microsoft.com/en-us/power-bi/enterprise/service-premium-scale-out>  
"The Scale-out queries for large semantic models setting for your tenant is enabled (default)."  
upvoted 1 times

🗳️ 👤 **NRezgui** 6 months ago

**Selected Answer: C**

. At the semantic model level, set Large dataset storage format to On.  
upvoted 1 times

🗳️ 👤 **Charley92** 7 months ago

**Selected Answer: C**

To enable scale-out for a semantic model in Power BI Premium, the Large dataset storage format must first be turned On. This is because the large dataset storage format allows models to take advantage of premium capacity features, including scale-out capabilities, which help distribute the load across multiple replicas for high-concurrency scenarios.  
upvoted 4 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

At the semantic model level, set Large dataset storage format to On  
upvoted 1 times

🗳️ 👤 **DarioReymago** 1 year ago

**Selected Answer: C**

Defini...defini...definitively is C  
upvoted 3 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: C**

Correct  
upvoted 1 times

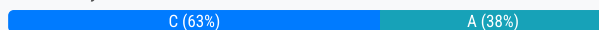


You have a Fabric tenant that contains a warehouse. The warehouse uses row-level security (RLS).  
 You create a Direct Lake semantic model that uses the Delta tables and RLS of the warehouse.  
 When users interact with a report built from the model, which mode will be used by the DAX queries?

- A. DirectQuery
- B. Dual
- C. Direct Lake
- D. Import

**Suggested Answer: D**

Community vote distribution



**wojciech\_wie** Highly Voted 1 year, 4 months ago

A. Direct Query "Row-level security only applies to queries on a Warehouse or SQL analytics endpoint in Fabric. Power BI queries on a warehouse in Direct Lake mode will fall back to Direct Query mode to abide by row-level security."

<https://learn.microsoft.com/en-us/fabric/data-warehouse/row-level-security>

upvoted 71 times

**FilipKr** 10 months, 3 weeks ago

Confirm:

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview>

"Queries using row-level security against tables in the warehouse (including the Lakehouse SQL analytics endpoint) will fall back to DirectQuery mode."

upvoted 8 times

**Estratech** Highly Voted 1 year, 2 months ago

**Selected Answer: C**

Direct lake . When users interact with a report built from a Direct Lake semantic model that uses Delta tables and RLS of the warehouse, the DAX queries will operate in Direct Lake mode. This mode is specifically designed for analyzing large data volumes in Power BI and is based on loading parquet-formatted files directly from a data lake without querying a Lakehouse or Warehouse endpoint. Unlike DirectQuery, there is no translation from DAX to other query languages, and it does not execute queries on other database systems. This results in performance similar to import mode, with the added benefit of picking up any changes at the data source as they occur<sup>12</sup>.

Direct Lake mode supports row-level security (RLS), ensuring that users only see the data they have permission to view. It combines the advantages of both DirectQuery and import modes while avoiding their disadvantages, making it an ideal choice for very large models and models with frequent updates at the data source<sup>2</sup>.

upvoted 26 times

**b8ef1c5** Most Recent 3 weeks, 5 days ago

**Selected Answer: A**

Direct Lake will fall back to DirectQuery when it can't load the data directly from a Delta table, such as when the data source ... uses RLS.

[https://learn.microsoft.com/en-us/fabric/fundamentals/direct-lake-](https://learn.microsoft.com/en-us/fabric/fundamentals/direct-lake-overview#:~:text=Direct%20Lake%20on%20SQL%20endpoints%20can%20fall%20back,when%20the%20Warehouse%20uses%20SQL%2Dbased%20Row%2Dlev)

[overview#:~:text=Direct%20Lake%20on%20SQL%20endpoints%20can%20fall%20back,when%20the%20Warehouse%20uses%20SQL%2Dbased%20Row%2Dlev](https://learn.microsoft.com/en-us/fabric/fundamentals/direct-lake-overview#:~:text=Direct%20Lake%20on%20SQL%20endpoints%20can%20fall%20back,when%20the%20Warehouse%20uses%20SQL%2Dbased%20Row%2Dlev)  
 upvoted 1 times

**Vbv\_8814** 1 month, 1 week ago

**Selected Answer: C**

Based on the table in latest documentation Semantic model with RLS supports Direct Lake mode

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview>

upvoted 1 times

**ba2bfdf** 3 months ago

**Selected Answer: A**

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview>

upvoted 1 times

🗄️ 👤 **4e5cf3d** 3 months ago

Selected Answer: A

Falls back to direct query  
upvoted 1 times

🗄️ 👤 **Amine\_spiegel94** 3 months, 2 weeks ago

Selected Answer: A

confirm  
upvoted 1 times

🗄️ 👤 **wudixh** 3 months, 2 weeks ago

Selected Answer: C

Answer from Microsoft Copilot: C.  
When users interact with a report built from a Direct Lake semantic model that uses Delta tables and RLS, the DAX queries will use Direct Lake mode (Option C) by default. This mode allows the semantic model to read Delta tables directly from OneLake.  
However, if there are certain conditions like memory pressure or unsupported features at the data source, the model might fall back to DirectQuery mode  
upvoted 1 times

🗄️ 👤 **aks2304** 3 months, 3 weeks ago

Selected Answer: A

The correct answer is A. DirectQuery.

When users interact with a report built from a Direct Lake semantic model that uses Delta tables and row-level security (RLS), the DAX queries are executed in DirectQuery mode. This is because RLS requires the queries to be evaluated at the source to enforce security rules dynamically, and Direct Lake does not currently support RLS natively. Therefore, the system falls back to DirectQuery mode to ensure RLS is applied correctly.  
upvoted 2 times

🗄️ 👤 **orallony** 6 months ago

Selected Answer: A

IT will fall back to direct query since the RLS is configure on the Data source and not in the semantic model  
upvoted 2 times

🗄️ 👤 **NRezgui** 6 months ago

Selected Answer: A

. DirectQuery  
upvoted 3 times

🗄️ 👤 **Red\_lotus85** 6 months, 1 week ago

Selected Answer: A

A! Read the limitation of direct lake  
upvoted 2 times

🗄️ 👤 **emboutchamani** 6 months, 1 week ago

Selected Answer: A

RLS is based on warehouse  
upvoted 2 times

🗄️ 👤 **MarkI** 6 months, 2 weeks ago

Selected Answer: A

A query always falls back when the semantic model queries a view in the SQL analytics endpoint, or a table in the SQL analytics endpoint that enforces row-level security (RLS).  
upvoted 3 times

🗄️ 👤 **Huepig** 6 months, 3 weeks ago

Selected Answer: A

RLS on the warehouse will fail back to direct query  
upvoted 2 times

🗄️ 👤 **Charley92** 7 months ago

Selected Answer: C

When a Direct Lake semantic model is created and used, the DAX queries interact directly with the data in OneLake without importing it into memory (as in Import mode) or querying it from a remote data source (as in DirectQuery mode).

The key features of Direct Lake mode include:

Direct access to Delta tables stored in OneLake.

Low latency querying without requiring data to be loaded into memory first.

Integration with features like Row-Level Security (RLS) when configured in the underlying warehouse.

upvoted 2 times

 **Shivam\_1122** 7 months ago

**Selected Answer: A**

Written in docs itself <https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview> it will fallback to DQ

upvoted 3 times

You have a Fabric tenant that contains a complex semantic model. The model is based on a star schema and contains many tables, including a fact table named Sales.

You need to create a diagram of the model. The diagram must contain only the Sales table and related tables.

What should you use from Microsoft Power BI Desktop?

- A. data categories
- B. Data view
- C. Model view
- D. DAX query view

**Suggested Answer: C**

Community vote distribution

C (100%)

🗨️ 👤 **DocE** 3 months ago

**Selected Answer: C**

See: <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-relationship-view>

upvoted 2 times

🗨️ 👤 **NRezgui** 6 months ago

**Selected Answer: C**

C. Model view

upvoted 2 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

Model view

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: C**

IMHO, "C"

Link: <https://learn.microsoft.com/en-us/power-bi/transform-model/desktop-relationship-view>

upvoted 2 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

**Selected Answer: C**

Well this was kinda Captain Obvious.

upvoted 4 times

🗨️ 👤 **SamuComqi** 1 year, 4 months ago

**Selected Answer: C**

C. Model view

In the Model view, it is possible to analyze the semantic model and create new layouts.

upvoted 2 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: C**

Model = model view

upvoted 1 times

You have a Fabric tenant that contains a semantic model. The model uses Direct Lake mode.

You suspect that some DAX queries load unnecessary columns into memory.

You need to identify the frequently used columns that are loaded into memory.

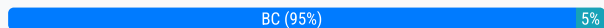
What are two ways to achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct answer is worth one point.

- A. Use the Analyze in Excel feature.
- B. Use the Vertipaq Analyzer tool.
- C. Query the \$System.DISCOVER\_STORAGE\_TABLE\_COLUMN\_SEGMENTS dynamic management view (DMV).
- D. Query the DISCOVER\_MEMORYGRANT dynamic management view (DMV).

**Suggested Answer: AC**

Community vote distribution



🗳️ **282b85d** Highly Voted 1 year, 1 month ago

**Selected Answer: BC**

Methods to Identify Frequently Used Columns:

B. Use the Vertipaq Analyzer tool.

Vertipaq Analyzer: This tool helps analyze the internal structure of your Power BI model. It provides detailed information about the storage and memory usage of your model, including which columns are frequently accessed and loaded into memory. This can help you identify unnecessary columns that are consuming resources.

Steps:

Export your Power BI model to a .pbix file.

Open the .pbix file in Power BI Desktop.

Use the Vertipaq Analyzer tool to analyze the model and review the column usage statistics.

C. Query the \$System.DISCOVER\_STORAGE\_TABLE\_COLUMN\_SEGMENTS dynamic management view (DMV).

DMVs: Dynamic Management Views (DMVs) provide detailed information about the operations of your Power BI models. Specifically, the \$System.DISCOVER\_STORAGE\_TABLE\_COLUMN\_SEGMENTS DMV can give you insights into the storage and usage patterns of individual columns within your model.

upvoted 13 times

🗳️ **Momoanwar** Highly Voted 1 year, 4 months ago

**Selected Answer: BC**

I think BC.

A is only to read data and D only memory allocations

upvoted 8 times

🗳️ **NRezgui** Most Recent 6 months ago

**Selected Answer: BC**

Use the Vertipaq Analyzer tool.

C. Query the \$System.DISCOVER\_STORAGE\_TABLE\_COLUMN\_SEGMENTS dynamic management view (DMV).

upvoted 1 times

🗳️ **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: BC**

B & C is the answer

upvoted 1 times

🗳️ **6d1de25** 11 months, 3 weeks ago

**Selected Answer: AB**

Answer is A&B

upvoted 1 times

🗨️ 👤 **FSCH\_111** 1 year ago

**Selected Answer: BC**

Other Options: WRONG

A. Use the Analyze in Excel feature: This feature allows for interaction with the model data in Excel but does not provide detailed insights into column-level memory usage.

D. Query the DISCOVER\_MEMORYGRANT DMV: This DMV provides information about memory grants for queries but does not provide detailed information about the columns loaded into memory.

upvoted 5 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: BC**

IMHO,

B & C

Because:

1. The DISCOVER\_STORAGE\_TABLE\_COLUMN\_SEGMENTS schema rowset returns information about the column segments used for storing data for in-memory tables.<336>

2. Very often there could be a few columns that are not required in your Power BI model, but they take up a lot of space. This is easy to find with Vertipaq Analyzer.

Links: [https://learn.microsoft.com/en-us/openspecs/sql\\_server\\_protocols/ms-sas/948d5135-5bf4-4cf7-82c5-3a38746c2fb8](https://learn.microsoft.com/en-us/openspecs/sql_server_protocols/ms-sas/948d5135-5bf4-4cf7-82c5-3a38746c2fb8)

<https://www.fourmoo.com/2020/11/11/how-to-use-vertipaq-analyzer-with-dax-studio-for-power-bi-model-analysis/>

upvoted 4 times

🗨️ 👤 **VAzureD** 1 year, 2 months ago

**Selected Answer: BC**

B and C

A. It's more about data exploration and visualization.

D. Provides information about memory grants for queries

upvoted 2 times

🗨️ 👤 **CLVASQUEZ** 1 year, 3 months ago

**Selected Answer: BC**

B and C is the right answer.

upvoted 2 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

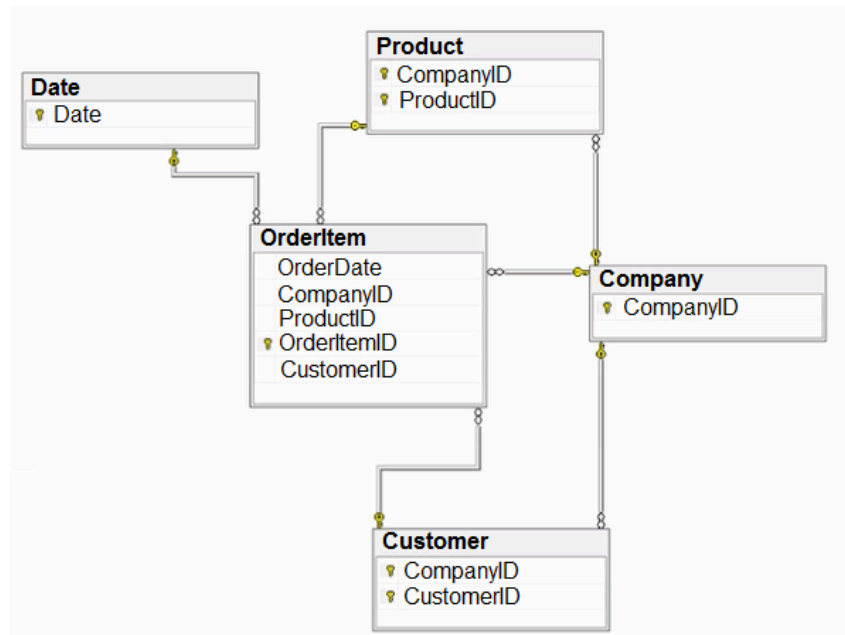
**Selected Answer: BC**

B and C is the right answer.

upvoted 3 times

## HOTSPOT -

You have the source data model shown in the following exhibit.



The primary keys of the tables are indicated by a key symbol beside the columns involved in each key.

You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer.

What should you include in the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

### Answer Area

The relationship between OrderItem and Product must be based on:

- ☐ The ProductID column
- ☐ Both the CompanyID and the ProductID columns
- ☐ A new key that combines the CompanyID and ProductID columns

The Company entity must be:

- ☐ Omitted
- ☐ Denormalized into the Product entity only
- ☐ Denormalized into the Customer and Product entities

### Answer Area

Suggested Answer:

The relationship between OrderItem and Product must be based on:

- ☐ The ProductID column
- ☒ Both the CompanyID and the ProductID columns
- ☐ A new key that combines the CompanyID and ProductID columns

The Company entity must be:

- ☐ Omitted
- ☐ Denormalized into the Product entity only
- ☒ Denormalized into the Customer and Product entities

**thuss** Highly Voted 1 year, 4 months ago

I think what they are trying to get at is that you denormalize the company information into both tables so that you do not need the CompanyID anymore, which would then make the relationship between OrderItem and Product only based on ProductID. Hence I think A and C.

But honestly that whole model is just weird.

upvoted 35 times

**d47320d** 1 year, 2 months ago

Agree up to the denormalization part.




But that does not imply that the relationship between OrderItem and Product should only be based on ProductID. The Product table has two

columns as a primary key, ProductID and CompanyID.

After denormalization, a two columns join should be performed in order to establish the proper relationships and proper analysis.

So correct answers are B, C i.e. the given one.

upvoted 13 times



  **zerone72**  1 year, 2 months ago

the question says : "You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer."

Therefore, the analysis of order item by date, product and customer doesn't need the company table (it can be omitted) !

The answer is A and A

upvoted 25 times

  **scorradi** 5 months, 3 weeks ago

B - The union between the tables must take into account the company ID and the product ID. It is not possible to have only the Product ID because the company ID is part of the composite key.

C - Regarding the entity, it is not possible to remove it. It is possible to remove the table only after denormalization. Some people make a relationship between the entity and the table itself, in this case it would be answer A, but I do not think that is the case here.

upvoted 3 times

  **pveiga**  1 week, 5 days ago

B and C

upvoted 1 times

  **PradeepikaGadai** 1 month, 2 weeks ago

I think answer for this is 1: Both the CompanyID and the ProductID columns because In the data model, Product has a composite primary key: (CompanyID, ProductID)

So, the relationship from OrderItem to Product must be based on both columns to uniquely identify each product.

and 2: Denormalized into the Customer and Product entities because both Customer and Product reference CompanyID, and we're designing a dimensional model (which favors denormalization for simplicity and performance), it is best to embed company data into the Customer and Product dimensions. This removes the need to model Company as a separate dimension and avoids extra joins.



so the answers are : Both the CompanyID and the ProductID columns , Denormalized into the Customer and Product entities

upvoted 1 times

  **zxc01** 1 month, 4 weeks ago

the question is "You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer." It is new dimensional data mode and we just focus on date, product, and customer. for both production and customer dimension, we can just keep distinct value of their ID, such as production ID and Customer ID, use it to build relationship with fact table. We don't need keep CompanyID. The answer should be A and A.

upvoted 1 times

  **5099e2d** 2 months, 2 weeks ago

To create a dimensional data model that will enable the analysis of order items by date, product, and company, you should consider the following approach:

The relationship between OrderItem and Product:

A: The ProductID column

The Company entity must be:

C: Denormalized into both Customer and Product entities

This approach ensures that the data model is optimized for analysis by including the necessary relationships and denormalizing the Company entity to provide comprehensive insights.

upvoted 1 times


  **MYPE** 4 months, 3 weeks ago

For the 1st part, I would say "C. A new key that combines the CompanyID and ProductID columns" because we are in the model view and there is no possibility to create a relationship based on 2 columns (that can be done in Power Query with a "merge" operation but NOT in the Model view).

For the 2nd part, C. "Denormalized into the Customer and Product entities"

Answer = C and C.

upvoted 2 times

  **Sowwy1** 7 months, 1 week ago



I think it's B and C

upvoted 1 times

🗨️ 👤 **Ous01** 7 months, 1 week ago

If company table represents cars such as Toyota and Honda while the product table has parts like a hood, the analysis is NOT interested in the car brand. The analysis should find out how many hoods the company sold, so I believe we should use the ProductID in the relation and the Company table can be omitted. I am leaning towards A & A

upvoted 2 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

A-->The product ID column

A-->Omitted

upvoted 6 times

🗨️ 👤 **Pegooli** 11 months, 1 week ago

Company doesn't have any more attribute. just company id and it's already on customer and product entities so it can be omitted.

upvoted 3 times

🗨️ 👤 **6d1de25** 11 months, 3 weeks ago

I think Company table should be completely omitted as this is extra data that is not mentioned in the requirement (You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer).

So the answer should be The Product ID Column and Omitted.

upvoted 3 times

🗨️ 👤 **c2834e0** 1 year ago

By denormalizing the relationships will become many to many no ?

upvoted 1 times

🗨️ 👤 **c2834e0** 1 year ago

forget what i said

upvoted 1 times

🗨️ 👤 **vspeter** 1 year ago

Given the Primary key for Product is CompanyID and ProductID, how do you model two (2) "active" one-to-many relationships between Product (dimension) and Orderitem (fact) objects?

upvoted 3 times

🗨️ 👤 **IK247** 1 year ago

C and C

The company ID and Product ID combined together will make a unique Identifier. You have to denormalise to reduce the number of joins in the model.

upvoted 6 times

🗨️ 👤 **Lion007** 11 months, 2 weeks ago

Correct. C and C

C. A new key that combines the CompanyID and ProductID columns

In a dimensional data model, especially in a star schema, it is important to create relationships that ensure uniqueness and properly link the fact table to the dimension tables.

C. Denormalized into the Customer and Product entities

The Company table should be denormalized into both the Customer and Product entities. This denormalization simplifies the model by embedding company information directly into the related tables, reducing the need for additional joins and improving query performance.

upvoted 2 times

🗨️ 👤 **282b85d** 1 year, 1 month ago

Relationship based on CompanyID and ProductID:

Why: In many real-world scenarios, products might be uniquely identified only when combined with the company context. Thus, ensuring the correct identification and linkage between OrderItem and Product necessitates using both columns. Create a composite key or use both columns to form the relationship between OrderItem and Product.

Denormalizing the Company entity:

Why: Denormalizing Company data into Customer and Product helps to flatten the structure, making it easier to query and reducing the complexity of joins.

How: Add the relevant Company attributes directly into the Customer and Product tables. For instance, each Customer and Product entry will carry information about the Company they are associated with.

upvoted 2 times

  **2dc6125** 1 year, 1 month ago

The requirement for Analysis "date, product, and customer" so we only need productId and nothing about the company so Omitted. I'm interested on why many answers go for including company in the model? would love to hear about that.

upvoted 5 times

  **66d0cf7** 1 year, 1 month ago

Because It seems more then one product will have the same ProductID if you omit the company, otherwise the current key of Product would not be ProductID and CompanyID

upvoted 2 times

You have a Fabric tenant that contains a semantic model named Model1. Model1 uses Import mode. Model1 contains a table named Orders. Orders has 100 million rows and the following fields.

Name	Data type	Description
OrderID	Integer	Column imported from the source
OrderDateTime	Date/Time	Column imported from the source
Quantity	Integer	Column imported from the source
Price	Decimal	Column imported from the source
TotalSalesAmount	Decimal	Calculated column that multiplies Quantity and Price
TotalQuantity	Integer	Measure

You need to reduce the memory used by Model1 and the time it takes to refresh the model.

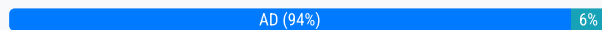
Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Split OrderDateTime into separate date and time columns.
- B. Replace TotalQuantity with a calculated column.
- C. Convert Quantity into the Text data type.
- D. Replace TotalSalesAmount with a measure.

**Suggested Answer: BD**

Community vote distribution



**Momoanwar** Highly Voted 1 year, 4 months ago

A : Best practice

D : measure better than column

upvoted 28 times

**fabric1** Highly Voted 1 year, 4 months ago

I was under the impression that A should be correct due to the fact that separate date and time column achieve higher columnar redundancy and allow better data compression. Whereas solution B, an additional calculated column, would inflate the memory usage. Measure are not stored in memory and would therefore be favorable concerning the stated objective of reducing memory used and minimizing refresh times.

upvoted 9 times

**NRezgui** Most Recent 6 months ago

**Selected Answer: AD**

A & D is the answer

upvoted 1 times

**Rakesh16** 7 months, 2 weeks ago

**Selected Answer: AD**

A & D is the answer

upvoted 1 times

**6d1de25** 11 months, 3 weeks ago

**Selected Answer: BD**

B&D are correct.

upvoted 1 times

**woliveiras** 1 year ago

**Selected Answer: AD**

A and D

upvoted 1 times

**282b85d** 1 year, 1 month ago

**Selected Answer: AD**

The correct answer is A and D

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

Selected Answer: AD

IMHO, A & D

upvoted 1 times

🗨️ 👤 **Nefirs** 1 year, 2 months ago

Selected Answer: AD

It's A D

upvoted 2 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

Selected Answer: AD

Its AD

upvoted 7 times

🗨️ 👤 **Nicofr** 1 year, 4 months ago

Selected Answer: AD

A should compress the memory size

D should reduce the memory usage but

upvoted 3 times

You have a Fabric tenant that contains a semantic model.

You need to prevent report creators from populating visuals by using implicit measures.

What are two tools that you can use to achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct answer is worth one point.

- A. Microsoft Power BI Desktop
- B. Tabular Editor
- C. Microsoft SQL Server Management Studio (SSMS)
- D. DAX Studio

**Suggested Answer: AC**

Community vote distribution

AB (100%)

🗳️ 👤 **XiltroX** Highly Voted 1 year, 4 months ago

**Selected Answer: AB**

Answer is A and B. You can prevent user-defined implicit measures by turning them off in PBI Desktop or Tabular Editor.

upvoted 15 times

🗳️ 👤 **a\_51** Highly Voted 1 year, 3 months ago

**Selected Answer: AB**

A,B turn enable discourage implicit measure for the semantic model.

<https://learn.microsoft.com/en-us/power-bi/transform-model/model-explorer#discourage-implicit-measures>

[https://docs.tabulareditor.com/api/TabularEditor.TOMWrapper.Model.html#TabularEditor\\_TOMWrapper\\_Model\\_DiscourageImplicitMeasures](https://docs.tabulareditor.com/api/TabularEditor.TOMWrapper.Model.html#TabularEditor_TOMWrapper_Model_DiscourageImplicitMeasures)

upvoted 8 times

🗳️ 👤 **goldy29** Most Recent 4 months, 4 weeks ago

**Selected Answer: AB**

Only Power BI Desktop and Tabular Editor allow for schema-level changes to a Power BI dataset, so you can add, modify, or delete tables, columns, and relationships in the data model.

upvoted 1 times

🗳️ 👤 **NRezgui** 6 months ago

**Selected Answer: AB**

A. Microsoft Power BI Desktop

B. Tabular Editor

upvoted 1 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: AB**

A & B is the answer

upvoted 1 times

🗳️ 👤 **woliveiras** 1 year ago

**Selected Answer: AB**

Just tabular editor and PowerBI can prevent user defined implicit measures

upvoted 2 times

🗳️ 👤 **282b85d** 1 year, 1 month ago

**Selected Answer: AB**

A. Microsoft Power BI Desktop: Power BI Desktop allows you to control and manage how measures are used within your reports. By carefully defining and using explicit measures within your data model, you can ensure that report creators use only these predefined measures instead of creating implicit measures automatically.

B. Tabular Editor: Tabular Editor is a powerful tool for managing and editing Power BI and Analysis Services tabular models. It allows you to enforce

best practices, such as disabling implicit measures, by modifying the model's properties and ensuring that only explicit measures are available for use in reports.

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO, A & B.

SSMS - big NO, Dax Studio - only for DAX programming.

upvoted 3 times

🗨️ 👤 **estrelle2008** 1 year, 4 months ago

**Selected Answer: AB**

Microsoft Power BI Desktop and Tabular Editor are your go-to for creating explicit measures.

upvoted 4 times

🗨️ 👤 **estrelle2008** 1 year, 4 months ago

and you can hide or disable implicit measures within the model, which prevents report builders from using them.

upvoted 2 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: AB**

Ssms have nothing to do here and dax studio not alter model

upvoted 3 times

🗨️ 👤 **IshtarSQL** 1 year, 4 months ago

**Selected Answer: AB**

AB: To prevent report creators from populating visuals using implicit measures in a Power BI semantic model within a Fabric tenant, you can utilize the following tools:

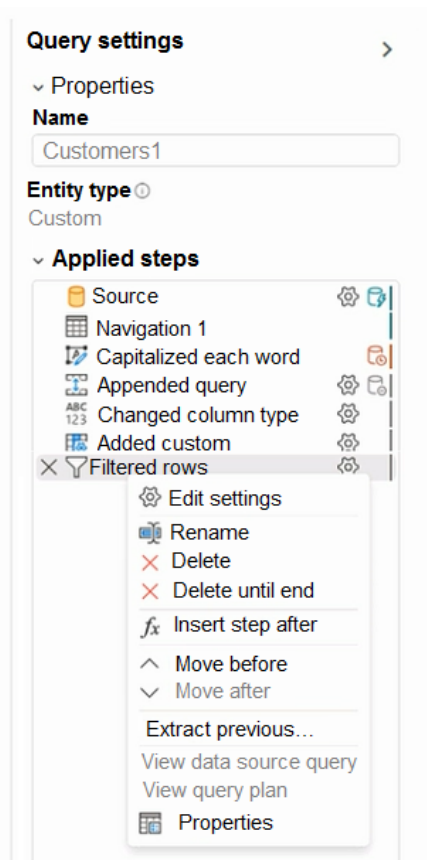
1. Tabular Editor:
2. Power BI Desktop (Data Model View):

upvoted 3 times

## HOTSPOT -

You have a Fabric tenant that contains two lakehouses.

You are building a dataflow that will combine data from the lakehouses. The applied steps from one of the queries in the dataflow is shown in the following exhibit.



Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

### Answer Area

[Answer choice] of the transformation steps in the query will fold.

▼  
 All  
 None  
 Some

The Added custom step will be performed in [answer choice].

▼  
 each lakehouse's query engine  
 the Microsoft Power Query engine  
 the source lakehouse query engine

### Answer Area

[Answer choice] of the transformation steps in the query will fold.

▼  
 All  
 None  
 Some

Suggested Answer:

The Added custom step will be performed in [answer choice].

▼  
 each lakehouse's query engine  
 the Microsoft Power Query engine  
 the source lakehouse query engine

Momoanwar Highly Voted 1 year, 4 months ago

1: Some. blue icon fold and red don't fold

2: on power query. all process after the step that don't fold are on power query

upvoted 27 times

SamuComqi Highly Voted 1 year, 4 months ago

The red icon suggests where query folding is interrupted, therefore only SOME steps are folded. All the other steps after the red icon will not fold and will be executed by the POWER QUERY engine.

upvoted 7 times

Spam\_Account Most Recent 5 months, 3 weeks ago

Isn't query folding when the query steps are pushed back into the engine, in order to optimise queries? This isn't happening at all, since the red icon occurs immediately after the navigation step. So the entire query WILL necessarily be pulled into power query, so no query folding will occur.

upvoted 2 times

Rakesh16 7 months, 2 weeks ago

'Some' and 'The Microsoft Power Query Engine'

upvoted 1 times

a\_51 1 year, 3 months ago

Correct as shown. If the query folded the work is pushed to the data source. the icons help to tell you if folding is occurring or not.

<https://learn.microsoft.com/en-us/power-query/step-folding-indicators#step-diagnostics-indicators>

upvoted 5 times

XiltroX 1 year, 4 months ago

The answers are correct. Some steps will be performed in Query Folding and query folding is only available in MS Power Query engine.

upvoted 2 times

NewB87 1 year, 4 months ago

Reference in MS Learn:

<https://learn.microsoft.com/en-us/power-query/step-folding-indicators>

upvoted 3 times



You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Table1.

You are creating a new data pipeline.

You plan to copy external data to Table1. The schema of the external data changes regularly.

You need the copy operation to meet the following requirements:

Replace Table1 with the schema of the external data.

Replace all the data in Table1 with the rows in the external data.

You add a Copy data activity to the pipeline.

What should you do for the Copy data activity?

- A. From the Source tab, add additional columns.
- B. From the Destination tab, set Table action to Overwrite.
- C. From the Settings tab, select Enable staging.
- D. From the Source tab, select Enable partition discovery.
- E. From the Source tab, select Recursively.

**Suggested Answer: B**

Community vote distribution

B (100%)

🗳️ 👤 **Momoanwar** Highly Voted 1 year, 4 months ago

**Selected Answer: B**

Replace all with overwrite

upvoted 8 times

🗳️ 👤 **Egocentric** Most Recent 5 months, 2 weeks ago

**Selected Answer: B**

Destination tab set table action to overwrite

upvoted 1 times

🗳️ 👤 **NRezgui** 6 months ago

**Selected Answer: B**

From the Destination tab, set Table action to Overwrite.

upvoted 1 times

🗳️ 👤 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: B**

From the Destination tab, set Table action to Overwrite.

upvoted 1 times

🗳️ 👤 **2dc6125** 1 year, 1 month ago

B Since the plan is to drop the destination schema and replace it with data source schema.

upvoted 2 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: B**

IMHO, B and only B.

Not D, because it is about partitions, not A, C, D - because of completely different purpose. A may be, generally, but it is not what a question about.

upvoted 1 times

🗳️ 👤 **VAzureD** 1 year, 2 months ago

**Selected Answer: B**

Overwrite

<https://learn.microsoft.com/en-us/fabric/data-factory/copy-data-activity#configure-your-source-under-the-source-tab>

Expand Advanced for more advanced settings.

upvoted 1 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

**Selected Answer: B**

B, the tool tip for Overwrite tells you:

"Enable overwrite table option to overwrite the existing data and schema in the table using the new values"

upvoted 1 times

🗨️ 👤 **IshtarSQL** 1 year, 4 months ago

**Selected Answer: B**

Correct: Enable "Truncate table" option: This option truncates the target table before copying data, ensuring that all existing data is replaced with the new data from the external source.

upvoted 2 times

🗨️ 👤 **Nefirs** 1 year, 2 months ago

this answer option is not available. Additionally, truncate a table does only delete the data but does not touch schema.

upvoted 3 times

You have a Fabric tenant that contains a lakehouse.

You plan to query sales data files by using the SQL endpoint. The files will be in an Amazon Simple Storage Service (Amazon S3) storage bucket. You need to recommend which file format to use and where to create a shortcut.

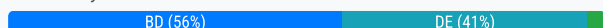
Which two actions should you include in the recommendation? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Create a shortcut in the Files section.
- B. Use the Parquet format
- C. Use the CSV format.
- D. Create a shortcut in the Tables section.
- E. Use the delta format.

**Suggested Answer:** CD

Community vote distribution



**Felgas** Highly Voted 1 year, 4 months ago

Answer is DE.

To be able to use the SQL Endpoint you need to create the shortcut in the Tables section. The file also needs to be in the delta format to be recognised as a managed table. If you try to add a parquet file to the tables section, it will not be recognised as a table object and you won't be able to query it.

upvoted 34 times

**d47320d** 1 year, 2 months ago

A parquet file shortcut is recognized as a managed table. Delta format not supported by S3. The answer cannot be E.

upvoted 11 times

**kilowd** 3 months, 2 weeks ago

E: Only the tables in Delta format are available in The SQL analytics endpoint. Parquet, csv and other formats cannot be queried using SQL analytics endpoint.

upvoted 1 times

**d47320d** Highly Voted 1 year, 2 months ago

B,D are the correct answers, since the Parquet file format within S3, and the pointing to it via a lakehouse Table shortcut, allows the shortcut to be recognized by Fabric as a Delta table and hence provide access to the SQL endpoint.

Explanation:

As per the provided links: "If the target of the shortcut contains data in the Delta\Parquet format, the lakehouse automatically synchronizes the metadata and recognizes the folder as a table".

It is stated that S3 contains sales data files, that is files not tables. Besides, delta tables are not supported in S3. So the file format needs to be either CSV or Parquet, leading us to choose the latter for all the reasons stated in other comments.

So creating a shortcut in Table section that points to S3 Parquet files, will allow Fabric to recognise it as a Delta table, which in turn enables the SQL endpoint.

upvoted 22 times

**uvan** 6 months, 1 week ago

In the question, it is not mentioned as a Parquet file. so we may not forcefully use the Parquet format.

upvoted 1 times

**Vbv\_8814** Most Recent 1 month, 1 week ago

**Selected Answer: BD**

S3 does not support delta format but supports parquet + shortcut in tables section can enable sql endpoint access .

Csv files cannot be queried by sql endpoint if I am not mistaken and data in file section is not available for sql endpoint query.

upvoted 2 times

🗄️ 👤 **zxc01** 2 months ago

Selected Answer: DE

This is another bad question. only delta format file can build tables and query by SQL endpoint. Other file formats, such as parquet and csv will put to unidentified folder and missing from SQL endpoint. However, why we keep delta type files inside S3?

upvoted 1 times

🗄️ 👤 **Amine\_spiegel94** 3 months, 2 weeks ago

Selected Answer: DE

SQL Endpoint you need to create the shortcut in the Tables section.

upvoted 1 times

🗄️ 👤 **Egocentric** 5 months, 2 weeks ago

Selected Answer: BD

B.use parquet format

D.create a shortcut in the table section

upvoted 1 times

🗄️ 👤 **NRezgui** 6 months ago

Selected Answer: DE

D. Create a shortcut in the Tables section.

E. Use the delta format

upvoted 1 times

🗄️ 👤 **Mhunity** 6 months, 4 weeks ago

Selected Answer: BD

B OR E (Use the Parquet format or Use the delta format)

D Create a shortcut in the Tables section

upvoted 1 times

🗄️ 👤 **Rakesh16** 7 months, 2 weeks ago

Selected Answer: BD

B & D is the answer

upvoted 1 times

🗄️ 👤 **moodi86** 9 months, 4 weeks ago

Delta uis supported:

<https://blog.fabric.microsoft.com/en-us/blog/public-preview-of-onelake-shortcuts-to-s3-compatible-data-sources?ft=All>

""If your data is already in the Delta Lake format, create your shortcut in the Tables section of your lakehouse. This will allow your table shortcut to benefit from metadata synchronization across Fabric engines, letting you use this structured data where tables are used in Fabric.""

upvoted 3 times

🗄️ 👤 **Pegooli** 11 months ago

Selected Answer: BD

chat GBT also agree with BD

upvoted 2 times

🗄️ 👤 **Ziggler** 11 months, 2 weeks ago

Selected Answer: CD

CSV is best if the data is not that large

upvoted 2 times

🗄️ 👤 **Ziggler** 11 months, 2 weeks ago

There is no mention of how big the data is , using CSV will be better if the data is small

upvoted 1 times

🗄️ 👤 **Pegooli** 11 months, 2 weeks ago

ChatGPT is saying amazon S3 supports both Delta and Parquet files

upvoted 2 times

🗄️ 👤 **6d1de25** 11 months, 3 weeks ago

Selected Answer: BD

B- Because Parquet format is supported in Amazon S3

<https://learn.microsoft.com/en-us/azure/data-factory/format-parquet>

D- Because you need to use the Shortcuts as managed portion of the lakehouse.

upvoted 3 times

  **OLGIS** 1 year ago

I guess MS documentation says Delta rather than Parquet? <https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sql-analytics-endpoint>

upvoted 1 times

  **woliveiras** 1 year ago

**Selected Answer: DE**

D and E. Tested. You can turn delta files available in your AWS Service and create a shortcut [here](#).

upvoted 6 times

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a subfolder named Subfolder1 that contains CSV files.

You need to convert the CSV files into the delta format that has V-Order optimization enabled.

What should you do from Lakehouse explorer?

- A. Use the Load to Tables feature.
- B. Create a new shortcut in the Files section.
- C. Create a new shortcut in the Tables section.
- D. Use the Optimize feature.

**Suggested Answer: A**

Community vote distribution

A (90%)

10%

🗳️ 👤 **Sanji931** Highly Voted 1 year, 4 months ago

**Selected Answer: A**

With "Load to tables" : tables are always loaded using the Delta Lake table format with V-Order optimization enabled.

<https://learn.microsoft.com/en-us/fabric/data-engineering/load-to-tables#load-to-table-capabilities-overview>

upvoted 11 times

🗳️ 👤 **Rakesh16** Most Recent 7 months, 2 weeks ago

**Selected Answer: A**

Use the Load to Tables feature.

upvoted 1 times

🗳️ 👤 **cafb698** 10 months, 1 week ago

**Selected Answer: A**

A, since optimize alone won't convert the CSV to Delta Format

upvoted 1 times

🗳️ 👤 **cafb698** 10 months, 1 week ago

**Selected Answer: D**

Optimize feature alone does not convert CSV files to Delta format. It is used for optimizing Delta tables after they have been created.

To fully address the requirement of converting CSV files into Delta format with V-Order optimization, the correct approach involves:

Loading CSV files into a Delta table: This is typically done using features that allow for the import and conversion of data into Delta format.

Applying V-Order optimization: Once the data is in Delta format, you use the Optimize feature to apply V-Order.

Hence, D

upvoted 1 times

🗳️ 👤 **Training\_ND** 10 months, 2 weeks ago

**Selected Answer: D**

The Optimize feature in a Lakehouse environment is specifically designed to optimize data, including converting it into more efficient formats like Delta and applying optimizations like V-Order. This feature is what you would use to convert CSV files into the Delta format and enable V-Order optimization, which reorganizes the data to improve query performance.

upvoted 1 times

🗳️ 👤 **byyleoo** 1 year ago

**Selected Answer: A**

A is correct

upvoted 2 times

🗳️ 👤 **282b85d** 1 year, 1 month ago

**Selected Answer: A**

The "Load to Tables" feature in Lakehouse explorer allows you to import data from various file formats (such as CSV) into a table within the lakehouse. During this process, you can specify the file format for the table, and by choosing the delta format, you can enable optimizations like V-Order.

upvoted 2 times

🗨️ **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO, "A" is right.

Because: The Lakehouse in Microsoft Fabric provides a feature to efficiently load common file types to an optimized Delta table ready for analytics. The Load to Table feature allows users to load a single file or a folder of files to a table.

<https://learn.microsoft.com/en-us/fabric/data-engineering/load-to-tables>

upvoted 2 times

🗨️ **prabhjot** 1 year, 2 months ago

I think Optimize ( Ans D) - <https://learn.microsoft.com/en-us/fabric/data-engineering/delta-optimization-and-v-order?tabs=sparksql>

upvoted 2 times

🗨️ **XiltroX** 1 year, 4 months ago

**Selected Answer: A**

A is the correct answer.

upvoted 4 times

🗨️ **Momoanwar** 1 year, 4 months ago

**Selected Answer: A**

Correct

upvoted 3 times

🗨️ **Nicofr** 1 year, 4 months ago

**Selected Answer: A**

"Load to Tables" functionality, the Optimize check mark is set by default.

upvoted 2 times

🗨️ **fabric1** 1 year, 4 months ago

**Selected Answer: A**

Hi IshtarSQL,

the "Optimize" feature is only applicable on already existing tables and cannot convert CSV files as far as I know. When loading a CSV file as table using the "Load to Tables" functionality, the Optimize check mark is set by default. Therefore, A should be correct.

Cheers fabric1

upvoted 3 times

🗨️ **IshtarSQL** 1 year, 4 months ago

**Selected Answer: D**

The "New Optimize" feature in the Lakehouse Explorer to convert CSV files into Delta format with V-Order optimization enabled.

The "New Optimize" feature allows you to optimize your data in Delta Lake format, including enabling V-Order optimization. V-Order optimization improves query performance by organizing data according to the values of frequently queried columns.

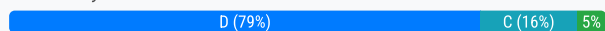
upvoted 1 times

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains an unpartitioned table named Table1. You plan to copy data to Table1 and partition the table based on a date column in the source data. You create a Copy activity to copy the data to Table1. You need to specify the partition column in the Destination settings of the Copy activity. What should you do first?

- A. From the Destination tab, set Mode to Append.
- B. From the Destination tab, select the partition column.
- C. From the Source tab, select Enable partition discovery.
- D. From the Destination tabs, set Mode to Overwrite.

**Suggested Answer: D**

Community vote distribution



**SamuComqi** Highly Voted 1 year, 4 months ago

**Selected Answer: D**

D. From the Destination tabs, set Mode to Overwrite.

When setting up the Copy Activity, you need to choose the Overwrite mode to make the partition option appear (not visible in Append mode).  
upvoted 15 times

**VAzureD** Highly Voted 1 year, 2 months ago

**Selected Answer: D**

D. From the Destination tabs, set Mode to Overwrite.

<https://learn.microsoft.com/en-us/fabric/data-factory/tutorial-lakehouse-partition#load-data-to-lakehouse-using-partition-columns>

Expand Advanced, in Table action, select Overwrite, and then select Enable partition, under Partition columns, select Add column, and choose the column you want to use as the partition column. You can choose to use a single column or multiple columns as the partition column.

upvoted 7 times

**Pegooli** Most Recent 3 weeks, 4 days ago

**Selected Answer: A**

In Microsoft Fabric, partitioning a table during a Copy activity is only supported in "Append" mode, not "Overwrite" mode.

upvoted 2 times

**praticewizards** 2 months ago

**Selected Answer: A**

To partition data in a destination table, you need to ensure that new data is appended. Setting the mode to "Append" is the first step because: Partitioning requires Append mode.

Overwrite mode would replace the entire table, negating any partitioning structure you aim to achieve. Once Append mode is set, you can specify the partition column in the Destination settings.

Why not the other options?

B. From the Destination tab, select the partition column: You cannot select the partition column until the mode is set to Append first.

C. From the Source tab, select Enable partition discovery: This option is useful for discovering partitions in source data, but it is unrelated to configuring partitioning in the destination table.

D. From the Destination tab, set Mode to Overwrite: Overwrite mode deletes existing data in the table, which contradicts the goal of appending and partitioning the data.

upvoted 2 times

**emboutchamani** 6 months, 1 week ago

**Selected Answer: A**

What should you do first ?

Before setting partition you have to set mode to append after you can select partition column



upvoted 1 times

🗨️ **emboutchamani** 6 months, 1 week ago

correction: D

upvoted 1 times

🗨️ **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: D**

From the Destination tabs, set Mode to Overwrite.

upvoted 1 times

🗨️ **Brainny** 9 months, 3 weeks ago

C. The "first" thing to do is to Enable Partition Discovery from the Source tab, this will enable the destination tab identify the partition column, then the destination tab can be set to overwrite.

<https://learn.microsoft.com/en-us/fabric/data-factory/connector-lakehouse-copy-activity>

upvoted 1 times

🗨️ **a2675f7** 8 months, 1 week ago

only when the source table is partitioned first... Enable partition discovery: For files that are partitioned, specify whether to parse the partitions from the file path and add them as extra source columns, in this case source is unpartitioned

upvoted 1 times

🗨️ **Pegooli** 11 months ago

**Selected Answer: C**

Enabling partition discovery on the Source tab allows the Copy activity to recognize the partition structure of the incoming data. This step is essential because it ensures that the data is copied into the destination table (Table1) with the correct partitions based on the specified date column.

upvoted 1 times

🗨️ **dev2dev** 1 year ago

**Selected Answer: D**

those who still think B is the answer, emphasis on "What should you do first?"

upvoted 1 times

🗨️ **Data\_Works** 1 year ago

**Selected Answer: D**

D. From the Destination tabs, set Mode to Overwrite.

in the question, it is specified that "we are going to specify the partition column". Yes, we will do that, BUT before that we have to set the "Table action" option to "Overwrite".

say, if you first set the partition column & after that you select the "overwrite" mode, then the selected partition column will be cleared & you will have to select that again.

And "overwrite" is necessary because the table is not currently partitioned. we need to re-write the existing data leveraging partition.

upvoted 1 times

🗨️ **282b85d** 1 year, 1 month ago

**Selected Answer: C**

Enable Partition Discovery: This option ensures that the Copy activity can identify the partition column and apply it correctly to the destination table. By enabling partition discovery, you make sure that the source data's partitioning information is considered during the copy process.

Why Not:

A. From the Destination tab, set Mode to Append: This setting controls how data is added to the existing table (whether new data is appended or existing data is overwritten). It does not directly address partitioning setup.

B. From the Destination tab, select the partition column: While this is necessary to specify the partition column, it is logically the next step after enabling partition discovery. The system needs to recognize partitions before you can configure them.

D. From the Destination tabs, set Mode to Overwrite: Similar to the Append mode, this setting determines how data is handled during the copy process but does not enable partitioning.

upvoted 3 times

🗨️ **David\_Webb** 1 year, 1 month ago

**Selected Answer: D**

Partition option need to switch to Overwrite mode first.

upvoted 1 times

🗨️ **stilferx** 1 year, 1 month ago

**Selected Answer: D**

IMHO, the answer is "D".

Because:

Expand Advanced, in Table action, select Overwrite, and then select Enable partition, under Partition columns, select Add column, and choose the column you want to use as the partition column. You can choose to use a single column or multiple columns as the partition column.

in <https://learn.microsoft.com/en-us/fabric/data-factory/tutorial-lakehouse-partition#load-data-to-lakehouse-using-partition-columns>  
upvoted 4 times

🗨️ **emmanuelkech** 1 year, 2 months ago

**Selected Answer: B**

From the Destination tab, select the partition column. This ensures that the data is partitioned based on the specified column in the destination table  
upvoted 3 times

🗨️ **estrelle2008** 1 year, 4 months ago

**Selected Answer: D**

A makes no sense, it leaves data that is already in table1 unpartitioned

B makes no sense, it is the statement that you are asked to decide for the step before that.

C MIGHT assign date column in source data as requested partition column, but only and if source data is complete and unambiguous

D re-creates table1, ensuring that the partitioning is applied consistently.

Hoping that the data that is already in table1 is also in this source data, D should be recommended answer, I guess?

upvoted 6 times

🗨️ **lengzhai** 1 year, 4 months ago

**Selected Answer: D**

D is correct. Partition is available when Overwrite is checked

upvoted 4 times

🗨️ **Momoanwar** 1 year, 4 months ago

**Selected Answer: C**

I think its C. For partition in pipeline we have to specify column in source

upvoted 3 times

🗨️ **Bharat** 1 year, 4 months ago

The table is unpartitioned to begin with.

upvoted 1 times

## HOTSPOT -

You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains a fact table named FactSales that has one billion rows.

You run the following T-SQL statement.

```
CREATE TABLE test.FactSales AS CLONE OF Dbo.FactSales;
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

## Answer Area

Statements	Yes	No
A replica of <code>dbo.Sales</code> is created in the test schema by copying the metadata only.	<input type="radio"/>	<input type="radio"/>
Additional schema changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input type="radio"/>
Additional schema changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input type="radio"/>

## Answer Area

Statements	Yes	No
<b>Suggested Answer:</b> A replica of <code>dbo.Sales</code> is created in the test schema by copying the metadata only.	<input checked="" type="radio"/>	<input type="radio"/>
Additional schema changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input checked="" type="radio"/>
Additional schema changes to <code>dbo.FactSales</code> will also apply to <code>test.FactSales</code> .	<input type="radio"/>	<input checked="" type="radio"/>

**thuss** Highly Voted 1 year, 3 months ago

Just took the exam. In the actual answers the typo in A is gone, and c actually reads "Additional data changes".

upvoted 18 times

**Pegooli** 11 months, 1 week ago

for additional data changes will apply, did you select "YES"?

upvoted 1 times

**Martin\_Nbg** 8 months, 3 weeks ago

The answer would be No. From the documentation: Table clones help to create historical reports that reflect the state of data as it existed as of a specific point-in-time in the past.

upvoted 1 times

**a\_51** 1 year, 3 months ago

That was the same for my exam. There were many other new questions, but I found at the end when I was reviewing there is the Microsoft Learn documentation that can be opened in the exam and really wish I had known that earlier as you are permitted to use it. I passed the exam, but sharing that other is there as there was some questions for new syntax I had not used, but once I looked up I had guessed right for them. One was a SQL statement to use a function called GREATEST and another was a question about merging schema in python. There were many questions of putting things in the right order to have something work proper. One was related to reviewing DAX performance and another related to order of pipeline for establishing a Bronze, Silver, Gold set of data.

upvoted 17 times

**SamuComqi** Highly Voted 1 year, 4 months ago

Y - N - N

The AS CLONE AS creates a replica of the original table by copying the metadata (no data).

The two copies are independent therefore any changes will not be inherited.

Source: <https://learn.microsoft.com/en-us/fabric/data-warehouse/clone-table>

upvoted 11 times

**Rakesh16** Most Recent 7 months, 2 weeks ago

Yes,No,No

upvoted 1 times

🗨️ 👤 **dev2dev** 1 year ago

actual typo is in the table name its going to create. its not dbo.Sales for sure. 1st option is yes only if there is no typo

upvoted 2 times

🗨️ 👤 **dev2dev** 1 year ago

If there are no typos in the choices first choice is NO because we are not closing dbo.Sales but dbo.FactSales

upvoted 3 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO, Y -> N -> N

Because of:

Creates a new table as a zero-copy clone of another table in Warehouse in Microsoft Fabric. Only the metadata of the table is copied. The underlying data of the table, stored as parquet files, is not copied.

in <https://learn.microsoft.com/en-us/sql/t-sql/statements/create-table-as-clone-of-transact-sql?view=fabric&preserve-view=true>

and

Any changes made through DML or DDL on the source of the clone table are not reflected in the clone table.

Similarly, any changes made through DDL or DML on the table clone are not reflected on the source of the clone table.

in <https://learn.microsoft.com/en-us/fabric/data-warehouse/clone-table#separate-and-independent>

upvoted 3 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

This is such a weird question. in the first statement, there is a typo and the question contains dbo.Sales when in fact, the original table is dbo.FactSales. I guess there was a typo so I'll let that pass and the answer is that YES, a replica is created of the original table.

The 2nd and 3rd questions are exact replicas of each other so both of their answers are NO. Any changes done after the cloning will not directly affect the cloned table.

upvoted 2 times

🗨️ 👤 **estrelle2008** 1 year, 4 months ago

YNN is correct now.

Second and third statements are equal "Additional schema changes to dbo.FactSales will also apply to test.FactSales"

What happens with this statement:

- A new table called test.FactSales is created.
- The structure (columns, data types, constraints) of test.FactSales will be identical to that of Dbo.FactSales at the time of creation.
- The data in test.FactSales will be a snapshot of the data in Dbo.FactSales at the moment of table creation.

upvoted 2 times

Note: This section contains one or more sets of questions with the same scenario and problem. Each question presents a unique solution to the problem. You must determine whether the solution meets the stated goals. More than one solution in the set might solve the problem. It is also possible that none of the solutions in the set solve the problem.

After you answer a question in this section, you will NOT be able to return. As a result, these questions do not appear on the Review Screen.

Your network contains an on-premises Active Directory Domain Services (AD DS) domain named contoso.com that syncs with a Microsoft Entra tenant by using Microsoft Entra Connect.

You have a Fabric tenant that contains a semantic model.

You enable dynamic row-level security (RLS) for the model and deploy the model to the Fabric service.

You query a measure that includes the USERNAME() function, and the query returns a blank result.

You need to ensure that the measure returns the user principal name (UPN) of a user.

Solution: You add user objects to the list of synced objects in Microsoft Entra Connect.

Does this meet the goal?

A. Yes

B. No

**Suggested Answer: B**

 **Mhunity**  6 months, 4 weeks ago

**Selected Answer: A**

Key Points for Evaluation:

The USERNAME() function in Power BI returns the UPN of the logged-in user, which corresponds to the user's identity in Microsoft Entra ID (Azure AD).

If the query returns a blank result, it indicates:

Either the UPN is not properly synced between AD DS and Microsoft Entra ID.

Or the user object is missing from the sync configuration.

By adding the user objects to the list of synced objects in Microsoft Entra Connect:

The UPNs for the users will be synced from AD DS to Microsoft Entra ID.

This ensures that the USERNAME() function can retrieve the UPN correctly.

Does the solution meet the goal?

Yes.

Adding the user objects to the sync configuration in Microsoft Entra Connect ensures that the UPNs are available in Microsoft Entra ID, allowing the USERNAME() function to return the correct result.



upvoted 8 times

 **MultiCloudIronMan**  7 months ago

**Selected Answer: B**

To ensure that the measure returns the user principal name (UPN) of a user, you need to make sure that the USERNAME() function is correctly configured to retrieve the UPN. This typically involves ensuring that the UPN is correctly mapped and available in the data model.

upvoted 7 times

 **VojtechSima**  2 months, 4 weeks ago

**Selected Answer: B**

I think it should be No.

You should use USERPRINCIPALNAME(), if you want to make sure, you get always UPN.

upvoted 4 times

🗨️ 👤 **PradeepikaGadai** 1 month, 2 weeks ago

Absolutely right!  
upvoted 1 times

🗨️ 👤 **joedxb** 2 months, 1 week ago

absolutely right!  
upvoted 2 times

🗨️ 👤 **Stants** 7 months ago

**Selected Answer: A**

The answer is A: Yes

Here's why:

The USERNAME() function requirement:

USERNAME() returns the UPN of the current user

UPN is needed for dynamic RLS to work properly

UPN must be available in Microsoft Entra ID

Microsoft Entra Connect sync:

By default, not all user attributes might be synced

User objects need to be explicitly included in sync configuration

UPN is a critical attribute for user identification

Why this solution works:

Adding user objects to sync ensures UPN attributes are synchronized

Ensures proper user identity flow from AD DS to Microsoft Entra ID

Enables USERNAME() function to retrieve the correct UPN

Completes the identity chain needed for dynamic RLS

upvoted 3 times

🗨️ 👤 **e810eb0** 7 months, 2 weeks ago

Here should be Yes  
upvoted 1 times

🗨️ 👤 **Ous01** 7 months, 2 weeks ago

Could you please let me know where and how you found the answer?  
upvoted 2 times

You have source data in a folder on a local computer.

You need to create a solution that will use Fabric to populate a data store. The solution must meet the following requirements:

Support the use of dataflows to load and append data to the data store.

Ensure that Delta tables are V-Order optimized and compacted automatically.

Which type of data store should you use?

- A. a lakehouse
- B. an Azure SQL database
- C. a warehouse
- D. a KQL database

**Suggested Answer: C**

Community vote distribution

A (61%)

C (39%)

 **bvmony2294** Highly Voted 1 year ago

I have recently taken the exam and this question was asked. Its a multiple choice question. we need to select 2 options. so as per the comments both lakehouse and warehouse are support delta tables and v-order optimization. Its A,C  
upvoted 21 times

 **testtaker45** 5 months, 1 week ago

Makes a ton of sense. Thanks for Sharing!  
upvoted 3 times

 **bigdave987** Highly Voted 1 year, 1 month ago

**Selected Answer: C**

The answer is correct - C Warehouse

The key to this question is "Ensure that Delta tables are V-Order optimized".

V-Order optimization isn't guaranteed in Lakehouse, and there are times when you need to run OPTIMIZE to ensure the tables are V-Order Optimized. This link here shows the answer:

<https://learn.microsoft.com/en-us/fabric/data-warehouse/ingest-data#best-practices>

The Note says "Regardless of how you ingest data into warehouses, the parquet files produced by the data ingestion task will be optimized using V-Order write optimization... Unlike in Fabric Data Engineering, V-Order is a global setting in Synapse Data Warehouse that cannot be disabled."


upvoted 10 times

 **Amine\_spiegel94** Most Recent 3 months, 2 weeks ago

**Selected Answer: A**

both lakehouse and warehouse are support delta tables and v-order optimization.

upvoted 2 times

 **4b35503** 6 months, 2 weeks ago

**Selected Answer: A**

All Fabric compute engines (spark notebooks, pipelines, Dataflow Gen2) automatically create Delta tables that are vorder'd.

upvoted 1 times

 **Rakesh16** 7 months, 2 weeks ago

**Selected Answer: C**

a warehouse

upvoted 1 times

 **Training\_ND** 10 months, 2 weeks ago

**Selected Answer: C**

In Microsoft Fabric, a warehouse is a specialized data store optimized for analytics and query performance. It uses V-Order write optimization, which is specifically designed to enhance read performance for parquet files across various compute engines such as Power BI, SQL, and Spark. This feature is automatically applied in Synapse Data Warehouse and cannot be disabled, ensuring that data stored in the warehouse is always optimized.

upvoted 1 times

🗨️ 👤 **6d1de25** 11 months, 3 weeks ago

Selected Answer: C

Both A & C are correct

upvoted 1 times

🗨️ 👤 **buhari** 1 year ago

C - Warehouse as for warehouse the optimized V-order is automatically enabled but for Lakehouse is setting that you need to enable or disable.

upvoted 1 times

🗨️ 👤 **282b85d** 1 year, 1 month ago

Selected Answer: A

Lakehouses can automatically handle the optimization and compaction of Delta tables, including V-Order optimization, which arranges data in an optimal order to improve query performance.

Why Not the Other Options?

Azure SQL Database (Option B): Azure SQL Database is a relational database service that does not natively support Delta tables or V-Order optimization. It is more suited for traditional OLTP workloads.

Warehouse (Option C): While warehouses are excellent for structured data and support dataflows, they may not provide the same level of native support for Delta tables and automatic V-Order optimization as a lakehouse.

KQL Database (Option D): KQL (Kusto Query Language) databases are optimized for log and telemetry data, primarily used with Azure Data Explorer. They are not designed to support Delta tables or the specific optimizations required for large-scale transactional data processing.

upvoted 3 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

Selected Answer: A

IMHO, "A"

because:

The Lakehouse and the Delta Lake table format are central to Microsoft Fabric, assuring that tables are optimized for analytics is a key requirement.

...

in

<https://learn.microsoft.com/en-us/fabric/data-engineering/delta-optimization-and-v-order?tabs=sparksql>

upvoted 1 times

🗨️ 👤 **PiotrO** 1 year, 2 months ago

Selected Answer: C

lakehouse doesn't automatically store data in delta tables while warehouse does.

upvoted 3 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

The whole new platform is focused on the lakehouse and optimization for it, so answer A.

upvoted 2 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

Selected Answer: A

A - The only logical answer here. B is an Azure SQL database and is an Azure product but doesn't have V-Order. C is just a generic warehouse and once again, doesn't necessarily contain any V-Order feature. D is a database that uses KQL and is irrelevant of the question.

upvoted 8 times

🗨️ 👤 **Gecig** 1 year ago

<https://learn.microsoft.com/en-us/fabric/data-warehouse/ingest-data#best-practices>

The Note says "Regardless of how you ingest data into warehouses, the parquet files produced by the data ingestion task will be optimized using V-Order write optimization... Unlike in Fabric Data Engineering, V-Order is a global setting in Synapse Data Warehouse that cannot be disabled."

Therefore the answer is warehouse

upvoted 5 times

🗨️ 👤 **mns0173** 9 months, 3 weeks ago

Read your quote one more time. It doesn't mean that the answer is a warehouse

upvoted 1 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

Selected Answer: A

Delta table... = Lakehouse



upvoted 4 times

  **IshtarSQL** 1 year, 4 months ago

**Selected Answer: A**

To meet the requirements of supporting dataflows to load and append data to the data store while ensuring that Delta tables are V-Order optimized and compacted automatically, you should use a lakehouse in Fabric as your solution.

upvoted 4 times

## HOTSPOT -

You have a Fabric tenant that contains a lakehouse.

You are using a Fabric notebook to save a large DataFrame by using the following code. `df.write.partitionBy("year", "month", "day").mode("overwrite").parquet("Files/SalesOrder")`

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

### Answer Area

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	<input type="radio"/>	<input type="radio"/>
The resulting file partitions can be read in parallel across multiple nodes.	<input type="radio"/>	<input type="radio"/>
The resulting file partitions will use file compression.	<input type="radio"/>	<input type="radio"/>

### Answer Area

Statements	Yes	No
<b>Suggested Answer:</b> The results will form a hierarchy of folders for each partition key.	<input checked="" type="radio"/>	<input type="radio"/>
The resulting file partitions can be read in parallel across multiple nodes.	<input checked="" type="radio"/>	<input type="radio"/>
The resulting file partitions will use file compression.	<input type="radio"/>	<input checked="" type="radio"/>

**Momoanwar** Highly Voted 1 year, 4 months ago

I think yes yes yes.

Parquet= compression

upvoted 37 times

**Blue\_MSBI** 1 year, 4 months ago

I think the same

upvoted 7 times

**Training\_ND** 10 months, 1 week ago

Yup, agreed

upvoted 1 times

**vimalan** 6 months ago

agreed

Yes, the resulting file partitions will use file compression. When you save a DataFrame in Parquet format using the `df.write.partitionBy(...).mode("overwrite").parquet(...)` method, Parquet files are automatically compressed by default

upvoted 2 times

**goldy29** 4 months, 4 weeks ago

The `.parquet()` format supports compression, but compression is not automatically applied unless specified (e.g., `df.write.option("compression", "snappy")`).

By default, Parquet often uses Snappy compression, but this depends on the implementation.

upvoted 2 times

**estrelle2008** Highly Voted 1 year, 4 months ago

I think so too: YYY

code snippet according to Learn

upvoted 8 times

**estrelle2008** 1 year, 4 months ago

additional: Parquet files are compressed by default, and you don't need to take any additional actions to enable compression. When writing Parquet files, you can specify the desired compression codec (if needed) to further optimize storage and performance

upvoted 5 times

🗨️ 👤 **Rakesh16** Most Recent 7 months, 2 weeks ago

yes,yes,yes

upvoted 1 times

🗨️ 👤 **Mitchell12345** 9 months, 3 weeks ago

Technically the results will not form a hierarchy of folders for EACH partition key right? Because for the day partition key, files are created. If it was phrased generically without the each-part, would've been better. Did someone have it on the exam?

upvoted 1 times

🗨️ 👤 **AdventureChick** 6 months, 2 weeks ago

Yes, there is a level for each "entity" in the partition key. Even if you only had 1 file in a day, that file would be found under a "day" folder.

upvoted 1 times

🗨️ 👤 **dev2dev** 1 year ago

Yes Yes Yes

The compression is optional parameter which uses 'snappy' compression by default. Unless we specify none, compression happens.

upvoted 3 times

🗨️ 👤 **282b85d** 1 year, 1 month ago

Y-Y-N

The results will form a hierarchy of folders for each partition key:

Yes: When using partitionBy in Spark, the data is organized into a hierarchical directory structure based on the specified partition keys. Therefore, you will see directories like year=YYYY/month=MM/day=DD within the specified output path.

The resulting file partitions can be read in parallel across multiple nodes:

Yes: Parquet files are designed for efficient querying and support parallel processing. Spark can read these partitions in parallel, enabling distributed query execution across multiple nodes.

The resulting file partitions will use file compression:

No: While Parquet format supports compression, it is not enabled by default in the code snippet provided. Compression needs to be explicitly specified if required. For example, you could use .option("compression", "snappy") to enable Snappy compression.

upvoted 3 times

🗨️ 👤 **vernillen** 1 year ago

Snappy is the default compression type of parquet, though. So it will create chunks of your file by default and compress them. If you want to have it all in one file, that's when you have to overwrite your default compression. So I disagree with the 'No'

upvoted 6 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO, fully agree with colleagues below - Y -> Y -> Y

upvoted 2 times

🗨️ 👤 **Nefirs** 1 year, 2 months ago

i think YYY as well

upvoted 3 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

I think it should be NYY as there is no mention in the code to form a hierarchy. Please correct me if I'm wrong.

upvoted 1 times

🗨️ 👤 **4fd861f** 1 year, 3 months ago

partitionBy will create it

upvoted 4 times

You have a Fabric workspace named Workspace1 that contains a data flow named Dataflow1 contains a query that returns the data shown in the following exhibit.

Table.TransformColumnTypes(#"Promoted headers", {{"VendorID", Int64.Type}, {"2015-01-01", Int64.Type}, {"2015-01-02", Int64.Type}, {"2015-01-03", Int64.Type}, {"2015-01-04", Int64.Type}, {"2015-01-05", Int64.Type}, {"2015-01-06", Int64.Type}, {"2015-01-07", Int64.Type}, {"2015-01-08", Int64.Type}})

VendorID	2015-01-01	2015-01-02	2015-01-03	2015-01-04	2015-01-05	2015-01-06	2015-01-07	2015-01-08
1	1	16	15	0	9	12	9	9
2	2	20	15	7	17	20	6	6

Column profiling based on top 1,000 rows

Query settings: Name: Generated-NYC-Taxi-Gre... Entity type: Custom Applied steps: Source, Promoted..., Changed c... Data destination: No data destination

Buttons: Step, Publish

You need to transform the data columns into attribute-value pairs, where columns become rows.

You select the VendorID column.

Which transformation should you select from the context menu of the VendorID column?

- A. Group by
- B. Unpivot columns
- C. Unpivot other columns
- D. Split column
- E. Remove other columns

#### Suggested Answer: C

Community vote distribution

C (100%)

**Momoanwar** Highly Voted 1 year, 4 months ago

Selected Answer: C

Correct

upvoted 11 times

**Rakesh16** Most Recent 7 months, 2 weeks ago

Selected Answer: C

Unpivot other columns

upvoted 1 times

**stilferx** 1 year, 1 month ago

Selected Answer: C

IMHO, "C",

Because of: Unpivot other columns

This command unpivots unselected columns. Use this command in a query when not all columns are known. New columns added during a refresh operation are also unpivoted.

description is here: <https://support.microsoft.com/en-us/office/unpivot-columns-power-query-0f7bad4b-9ea1-49c1-9d95-f588221c7098#:~:text=Select%20the%20columns%20you%20do,Transform%20%3E%20Unpivot%20Only%20Selected%20Columns.>



upvoted 3 times

**Nefirs** 1 year, 2 months ago

Selected Answer: C

clearly C

upvoted 1 times

  **XiltroX** 1 year, 4 months ago

C is the correct answer

upvoted 3 times

You have a Fabric tenant that contains a data pipeline.  
You need to ensure that the pipeline runs every four hours on Mondays and Fridays.  
To what should you set Repeat for the schedule?

- A. Daily
- B. By the minute
- C. Weekly
- D. Hourly

**Suggested Answer: D**

Community vote distribution

C (97%)

🗳️ 👤 **Sanji931** Highly Voted 👍 1 year, 4 months ago

**Selected Answer: C**

Answer C : Weekly.

The only way to do this is to set the schedule to "Weekly", set the days on Monday and Friday and add manually 6 Time of 4 hour intervals.

upvoted 17 times

🗳️ 👤 **Rakesh16** Most Recent ⌚ 7 months, 2 weeks ago

**Selected Answer: C**

Weekly

upvoted 1 times

🗳️ 👤 **Bonga1** 9 months, 4 weeks ago

**Selected Answer: C**

Weekly

upvoted 1 times

🗳️ 👤 **282b85d** 1 year, 1 month ago

**Selected Answer: C**

Schedule Type: Weekly

Days: Monday, Friday

Interval: Every 4 hours

upvoted 3 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: C**

IMHO, "C" is good

upvoted 1 times

🗳️ 👤 **Nefirs** 1 year, 2 months ago

**Selected Answer: C**

C - individual days can be selected only in Weekly schedule, and then add additional Times manually for each interval within a day.

upvoted 2 times

🗳️ 👤 **dvenkatesh** 1 year, 2 months ago

I would pick D.

To ensure that the pipeline runs every four hours on Mondays and Fridays, you should set the Repeat for the schedule to "4 hours".

upvoted 2 times

🗳️ 👤 **XiltroX** 1 year, 4 months ago

**Selected Answer: C**

This is a no-brainer as the only way to make this work is to set the schedule as weekly and specify the pipeline to run on Mondays and Wednesdays.



upvoted 4 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: C**

Weekly



upvoted 3 times

  **Nicofr** 1 year, 4 months ago

**Selected Answer: C**

Weekly allow to choose the days and the time

upvoted 4 times

  **fabric1** 1 year, 4 months ago

**Selected Answer: C**

Hi IshtarSQL,

the selection of individual week days is only possible in the "Weekly" schedule option. Therefore, C is correct here.

upvoted 3 times

  **IshtarSQL** 1 year, 4 months ago

**Selected Answer: A**

To ensure that the pipeline runs every four hours on Mondays and Fridays, you should set the "Repeat" frequency for the schedule to "Daily" and set the "Interval" to 4 hours. Then, you can specify the days of the week when the pipeline should run by selecting only Mondays and Fridays.

upvoted 1 times

You have a Fabric tenant that contains a warehouse.

Several times a day, the performance of all warehouse queries degrades. You suspect that Fabric is throttling the compute used by the warehouse. What should you use to identify whether throttling is occurring?

- A. the Capacity settings
- B. the Monitoring hub
- C. dynamic management views (DMVs)
- D. the Microsoft Fabric Capacity Metrics app

**Suggested Answer: D**

Community vote distribution

D (100%)

🗳️ 👤 **XiltroX** Highly Voted 1 year, 4 months ago

**Selected Answer: D**

From MS Learn:

The Microsoft Capacity Metrics app, also known as the metrics app, serves as a monitoring tool within the Microsoft Power BI service. It offers functionalities to track and analyze the resource utilization

upvoted 13 times

🗳️ 👤 **Rakesh16** Most Recent 7 months, 2 weeks ago

**Selected Answer: D**

the Microsoft Fabric Capacity Metrics app

upvoted 1 times

🗳️ 👤 **AndreaRosho1** 1 year, 1 month ago

**Selected Answer: D**

<https://learn.microsoft.com/en-us/fabric/data-warehouse/compute-capacity-smoothing-throttling>

"Just like most Warehouse operations, dynamic management views (DMVs) are also classified as background and covered by the "Background Rejection" policy. As a result, DMVs cannot be queried when capacity is throttled. Even though DMVs are not available, capacity admins can go to Microsoft Fabric Capacity Metrics app to understand the root cause."

upvoted 3 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: D**

IMHO, "D"

Because:

Track rejected operations

The Microsoft Fabric Capacity Metrics app drilldown allows admins to see operations that were rejected during a throttling event.

Here: <https://learn.microsoft.com/en-us/fabric/enterprise/throttling#track-rejected-operations>

upvoted 1 times

🗳️ 👤 **prabhjot** 1 year, 2 months ago

yes Ans D - <https://learn.microsoft.com/en-us/fabric/enterprise/metrics-app>

upvoted 1 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: D**

Microsoft Fabric Capacity Metrics app for throttling

upvoted 2 times



## HOTSPOT -

You have a Fabric workspace that uses the default Spark starter pool and runtime version 1.2.

You plan to read a CSV file named Sales\_raw.csv in a lakehouse, select columns, and save the data as a Delta table to the managed area of the lakehouse. Sales\_raw.csv contains 12 columns.

You have the following code.

```
from pyspark.sql.functions import year

(spark
 .read
 .format("csv")
 .option("header", 'true')
 .load("Files/sales_raw.csv")
 .select('SalesOrderNumber', 'OrderDate', 'CustomerName',
 'UnitPrice')
 .withColumn("Year", year("OrderDate"))
 .write
 .partitionBy('Year')
 .saveAsTable("sales")
 )
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

### Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv.	<input type="radio"/>	<input type="radio"/>
Removing the partition will reduce the execution time of the query.	<input type="radio"/>	<input type="radio"/>
Adding inferSchema= 'true' to the options will increase the execution time of the query.	<input type="radio"/>	<input type="radio"/>

### Answer Area

Statements	Yes	No
<b>Suggested Answer:</b> The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv.	<input checked="" type="radio"/>	<input type="radio"/>
Removing the partition will reduce the execution time of the query.	<input type="radio"/>	<input checked="" type="radio"/>
Adding inferSchema= 'true' to the options will increase the execution time of the query.	<input checked="" type="radio"/>	<input type="radio"/>

metiii **Highly Voted** 1 year, 3 months ago

- No, this is called filter pushdown / predicate pushdown / column pruning. This config is available when reading from a columnar type like parquet, I didn't find anything related to csv, I know that you can pushdown a predicate on csv to make it only read some rows in that case it works but it probably doesn't work for selecting columns so spark will read the entire file then filters the columns.
  - Yes partitioning creates some overhead since Spark needs to create more files
  - Yes inferSchema forces spark to read the file twice once for schema and once for data
- upvoted 40 times

282b85d **Highly Voted** 1 year, 1 month ago

N-N-Y

- No: The Spark engine will initially read all columns from the CSV file because the .select() transformation is applied after the data has been read into memory. Therefore, all 12 columns from Sales\_raw.csv are read before the selection of specific columns is applied.
- No: Removing the partition might not necessarily reduce the execution time. While there might be some overhead in writing data to partitions, the overall impact on read performance, especially for large datasets, is usually beneficial. The query execution time for saving might be higher due to partitioning, but the read performance improvement usually outweighs this cost.
- Yes: Adding inferSchema = 'true' will increase the execution time of the query because Spark will need to read through the entire dataset to determine the data types of each column. This extra pass over the data adds to the initial read time.

upvoted 21 times

🗨️ 👤 **Huepig** 6 months, 3 weeks ago

This is not correct. Partitioning is not recommended because of both read and write overheads. Most often than not, partitioning results in "many small file" problem and data skew. Also, partitioning will result in reshuffle which causes longer write durations.

In this case, unless the csv is so massive that each "year" partition is greater 128mb as parquet, the partition will increase both read and write durations.

To add to this

upvoted 1 times

🗨️ 👤 **Ielima** 1 year ago

This is correct

upvoted 1 times

🗨️ 👤 **testtaker45** Most Recent 5 months, 1 week ago

N, Y, Y

LoL, such a controversial question.

1. The key word is read. Load will read all of them, it will only select the specified. When Spark does a .select(), it will read everything into mem first.
2. Removing the partition would reduce compute requirements, making the query execute faster, but would likely increase the in memory footprint as ultimately more is loaded into memory, not just a part.
- 3.infer\_schema will try to guess the type, like int, date, etc, but it requires compute. This will slow down the query.

I hope this helps someone!

upvoted 1 times

🗨️ 👤 **Rakesh16** 7 months, 2 weeks ago

No,No,yes

upvoted 1 times

🗨️ 👤 **Pegooli** 11 months, 2 weeks ago

I'm going to Y-N-N

upvoted 1 times

🗨️ 👤 **calvintcy** 1 year ago

I took the test today. This question was included, but the option 'Removing the partition will reduce the execution time of the query' has been replaced by 'Will the Year column replace the OrderDate column?'. My answer was No.

upvoted 11 times

🗨️ 👤 **radamantes** 5 months, 3 weeks ago

Did you approve? Most of question were in this site? Can you share?

upvoted 1 times

🗨️ 👤 **mnc\_1997** 1 year, 1 month ago

just tried it, it only writes the columns that were selected.

Answer: YNY

upvoted 2 times

🗨️ 👤 **mnc\_1997** 1 year, 1 month ago

also, what spark does is perform a lazy evaluation approach, it does not read each method(read,load,option) into memory. The actual reading happens when an action is performed such as(display,show,write). Spark will create a plan on how to execute the entire query and will optimize this plan for efficient execution.

upvoted 2 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO,

1. N
2. N - arguable
3. Y

1 No - because it is CSV. It will be read in full (in contrast to parquet)

2 No - well, maybe 0.5% slower due to creating a new files. But actually - no

3 Yes - because inferring schema - it is additional process

upvoted 6 times

🗨️ 👤 **vish9** 1 year, 1 month ago

No, CSV will be read in full and then filtered.

No: Using the partition by clause in Spark's Delta format can impact write performance in several ways:

Increased Write Throughput: Partitioning your data can potentially increase write throughput by distributing the write workload across multiple partitions. This parallelism allows Spark to write data to different partitions concurrently, improving overall write performance, especially when dealing with large datasets.

Y. Infer schema will slow the performance

upvoted 5 times

🗨️ 👤 **dp600** 1 year, 2 months ago

I would go with NYY.

It's a CSV it is a row format, I don't think you can separate it by columns before reading the entire content.

Partitioning takes extra work, so it may slow down the process.

InferSchema requires an extra scan of the document or I think so, so maybe, I will go with yes.

upvoted 2 times

🗨️ 👤 **DilumD** 1 year, 2 months ago

1. Yes: Reason: Select columns: The code selects specific columns from the DataFrame using the select method. The selected columns are "SalesOrderNumber", "OrderDate", "CustomerName", and "UnitPrice".

2. Yes: Reason: removing the partitionBy will simplify the process. Partitioning data involves some overhead in organizing the data into separate folders/files based on the partitioning column.

3. No: Reason: Potentially Slower: Enabling inferSchema generally results in a slightly slower initial read operation. This is because Spark needs to do an additional scan of a portion of your data to analyze and determine data types before loading it.

upvoted 1 times

🗨️ 👤 **wellingtonluis** 1 year, 3 months ago

After read all file, engine will select just some. But, initially it runs the entire file.

upvoted 5 times

🗨️ 👤 **XiltroX** 1 year, 4 months ago

The answer is probably YNY

1. Those are exactly the columns that are being read. So Yes

2. Removing the PartitionBy line would not result in any performance changes. So NO

3. Adding inferSchema as True WILL result in extra time in execution as it will make the engine go over the data twice (one to read data and the other time to read Schema). So YES.

upvoted 11 times

🗨️ 👤 **estelle2008** 1 year, 4 months ago

full of typos, this one.

Anyhow, my guess:

YNN

inferSchema=true helps automatically determine column data types, but it needs a extra pass over the data, which comes with a slight query performance cost. So last statement = No

upvoted 2 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

Its read not red.

This question is ambiguous would say : no no yes.

For the point 1 : with case sensitivity sales\_raw is not Sales\_raw

upvoted 5 times

You have a Fabric tenant that contains a warehouse.

A user discovers that a report that usually takes two minutes to render has been running for 45 minutes and has still not rendered.

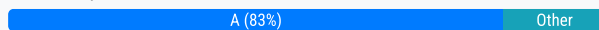
You need to identify what is preventing the report query from completing.

Which dynamic management view (DMV) should you use?

- A. sys.dm\_exec\_requests
- B. sys.dm\_exec\_sessions
- C. sys.dm\_exec\_connections
- D. sys.dm\_pdw\_exec\_requests

**Suggested Answer: D**

Community vote distribution



**Sanji931** Highly Voted 1 year, 4 months ago

**Selected Answer: A**

Answer is A.

<https://learn.microsoft.com/en-us/fabric/data-warehouse/monitor-using-dmv>

upvoted 14 times

**thuss** 1 year, 4 months ago

Yeah they finally got rid of the PDW bit.

upvoted 2 times

**a\_51** 1 year, 3 months ago

If you look at docs it is still out there for a parallel datawarehouse, but considering this is an exam about fabric, then the link mentioned for fabric monitor would make the answer A I agree.

upvoted 1 times

**MultiCloudIronMan** Most Recent 6 months, 3 weeks ago

**Selected Answer: D**

Given that Fabric uses a Massively Parallel Processing (MPP) architecture similar to Azure Synapse Analytics, sys.dm\_pdw\_exec\_requests (Option D) could indeed be relevant for monitoring requests in this environment.

upvoted 1 times

**Rakesh16** 7 months, 2 weeks ago

**Selected Answer: A**

sys.dm\_exec\_requests

upvoted 1 times

**Pegooli** 11 months ago

**Selected Answer: D**

sys.dm\_pdw\_exec\_requests: This DMV provides information about the status of requests (queries) executed in SQL Data Warehouse (now known as Azure Synapse Analytics) environments. It includes details about query execution, such as start time, end time, status, and any error messages. This DMV is particularly useful for diagnosing long-running or stuck queries.

upvoted 2 times

**Pegooli** 8 months ago

correcting my above comments is about Synapse Analytics. The Fabric data warehouse is using sys.dm\_exec\_requests command similar to Azure SQL DB

upvoted 1 times

**rlo123** 1 year, 1 month ago

D. A or D are really close, but since it is Azure Synapse Analytics D is probably the best answer. Why sys.dm\_pdw\_exec\_requests is better for this Fabric scenario:

Fabric's MPP architecture: Fabric warehouses utilize a distributed architecture where queries are broken down and processed across multiple

compute nodes. This is why `sys.dm_pdw_exec_requests` is the ideal tool.

MPP-specific insights: This DMV gives you visibility into how the query is being executed across the nodes, which can reveal bottlenecks or performance issues that wouldn't be apparent in `sys.dm_exec_requests`.

Targeted troubleshooting: With the MPP-specific data from `sys.dm_pdw_exec_requests`, you can pinpoint the exact steps or nodes causing the slow performance, leading to a more efficient resolution.

upvoted 3 times

  **rlol123** 1 year, 1 month ago

`sys.dm_exec_requests` is the correct answer after a lot more review and searching, can't delete my original comment, but here is why....In the context of Microsoft Fabric, `sys.dm_exec_requests` would be the appropriate DMV to use for monitoring running queries, as it is specifically designed for the SQL engine within Fabric.

`sys.dm_pdw_exec_requests` is a DMV designed for Azure Synapse Analytics dedicated SQL pools (formerly SQL Data Warehouse). While Fabric's warehouse engine is based on SQL Server technology, it's not directly compatible with the DMVs specific to Azure Synapse Analytics.

upvoted 3 times

  **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO, "A",

it may be found here: <https://learn.microsoft.com/en-us/fabric/data-warehouse/monitor-using-dmv#identify-and-kill-a-long-running-query>

upvoted 1 times

  **Estratech** 1 year, 2 months ago

**Selected Answer: A**

You can use the `sys.dm_exec_requests` dynamic management view (DMV) to identify what is preventing the report query from completing 1.

upvoted 1 times

  **dvenkatesh** 1 year, 2 months ago

Answer would be D, I guess...

Here the key word might be warehouse...

The `sys.dm_exec_requests` dynamic management view is used in SQL Server to provide information about each request that is executing within SQL Server. On the other hand, the `sys.dm_pdw_exec_requests` dynamic management view is used in Azure SQL Data Warehouse to provide similar information about each request executing within the data warehouse environment. Both views offer insights into the current status and details of queries running in their respective environments.

upvoted 3 times

  **azure\_bimonster** 1 year, 3 months ago

**Selected Answer: A**

A would be correct

upvoted 2 times

  **XiltroX** 1 year, 4 months ago

**Selected Answer: C**

Answer is C

<https://learn.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-exec-connections-transact-sql?view=sql-server-ver16>

upvoted 1 times

## DRAG DROP -

You are creating a data flow in Fabric to ingest data from an Azure SQL database by using a T-SQL statement.

You need to ensure that any foldable Power Query transformation steps are processed by the Microsoft SQL Server engine.

How should you complete the code? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

## Values

## Answer Area

EnableFolding	let	Source = Sql.Databases(
NativeQuery		"server.database.windows.net"
Optimize		),
Record		Database = Source{[Name = "db"]}[Data],
StopFolding		Query = [ ] . [ ] (
Table		Database,
Value		" SELECT * FROM customer WHERE country IN ('USA', 'UK')",
		null,
		[ ] = true]
		)
		in
		Query

## Answer Area

**Suggested Answer:**

```

let
    Source = Sql.Databases(
        "server.database.windows.net"
    ),
    Database = Source{[Name = "db"]}[Data],
    Query = Value . NativeQuery (
        Database,
        " SELECT * FROM customer WHERE country IN ('USA', 'UK')",
        null,
        [ EnableFolding = true]
    )
in
    Query
  
```

**SamuComqi** Highly Voted 1 year, 4 months ago

- \* Value
- \* NativeQuery
- \* EnableFolding

Source: <https://learn.microsoft.com/en-us/power-query/native-query-folding>  
upvoted 29 times

**Rakesh16** Most Recent 7 months, 2 weeks ago

Value.NativeQuery & EnableFolding  
upvoted 1 times

**282b85d** 1 year, 1 month ago

- Value.NativeQuery: This function is used to execute a native SQL query directly against the data source. This ensures that the query is processed on the server-side, which can significantly improve performance by leveraging the database engine's capabilities.
  - EnableFolding: By setting EnableFolding to true, you are instructing Power Query to allow query folding for any subsequent transformations that can be pushed down to the SQL Server engine. This means that Power Query will try to push as much of the query logic as possible to the database engine.
- upvoted 3 times

**stilferx** 1 year, 1 month ago

IMHO,  
Value -> NativeQuery -> EnableFolding is good :)



Because of here:

Value.NativeQuery(Source, "SELECT DepartmentID, Name FROM HumanResources.Department WHERE GroupName = 'Research and Development' ", null, [EnableFolding = true])

Link:



<https://learn.microsoft.com/en-us/power-query/native-query-folding#use-valuenativequery-function>

upvoted 3 times

  **XiltroX** 1 year, 4 months ago

Answer is correct

upvoted 3 times

  **Momoanwar** 1 year, 4 months ago

Correct

upvoted 2 times

## DRAG DROP -

You have a Fabric tenant that contains a lakehouse named Lakehouse1.

Readings from 100 IoT devices are appended to a Delta table in Lakehouse1. Each set of readings is approximately 25 KB. Approximately 10 GB of data is received daily.

All the table and SparkSession settings are set to the default.

You discover that queries are slow to execute. In addition, the lakehouse storage contains data and log files that are no longer used.

You need to remove the files that are no longer used and combine small files into larger files with a target size of 1 GB per file.

What should you do? To answer, drag the appropriate actions to the correct requirements. Each action may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

**Actions****Answer Area**

Set the autoCompact  
table setting.

Set the optimizeWrite  
table setting.

Run the VACUUM  
command on a schedule.

Set the autoCompact  
SparkSession setting.

Run the OPTIMIZE  
command on a schedule.

Set the parallelDelete  
SparkSession setting.

Remove the files:

Combine the files:


**Answer Area****Suggested Answer:**

Remove the files:

Run the VACUUM  
command on a schedule.

Combine the files:

Run the OPTIMIZE  
command on a schedule.

 **SamuComqi** Highly Voted 1 year, 4 months ago

VACUUM: to remove old files no longer referenced.

OPTIMIZE: to create fewer files with a larger size.

Sources:

\* <https://learn.microsoft.com/en-us/fabric/data-engineering/delta-optimization-and-v-order?tabs=sparksql>

\* VACUUM: <https://docs.delta.io/latest/delta-utility.html#delta-vacuum>

\* OPTIMIZE: <https://docs.delta.io/latest/optimizations-oss.html>

upvoted 23 times

 **Rakesh16** Most Recent 7 months, 2 weeks ago

Remove the files-->Run the VACUUM command on a schedule

Combine the files-->Run the OPTIMIZE command on a schedule

upvoted 2 times

 **Pegooli** 11 months, 2 weeks ago



answer is correct :)

upvoted 2 times

  **282b85d** 1 year, 1 month ago

- Remove the files that are no longer used:

Run the VACUUM command on a schedule: The VACUUM command cleans up old files and log files that are no longer needed by the Delta table, helping to free up storage and potentially improve performance by reducing the number of files the query engine needs to consider.

- Combine small files into larger files:

Run the OPTIMIZE command on a schedule: The OPTIMIZE command compacts small files into larger ones, improving read performance by reducing the overhead associated with opening many small files. This can be particularly useful when you have a large number of small files due to frequent appends of small data sets.

upvoted 1 times

  **stilferx** 1 year, 1 month ago

IMHO,



Vacuum & Optimize are good for optimizing Delta Lake :)

upvoted 1 times

  **Valcon\_doo\_NoviSad** 1 year, 3 months ago

I agree that it is VACUUM and OPTIMIZE, but I would say Set the optimizeWrite table setting (B) and not Run the OPTIMIZE command on a schedule (E).

upvoted 2 times

  **thuss** 1 year, 3 months ago

Isn't optimizeWrite set by default though? However that would only optimize the data as it is written, not over time.

upvoted 1 times

  **Momoanwar** 1 year, 4 months ago

Correct :

OPTIMIZE Improves query performance by optimizing file sizes. See Compact data files with optimize on Delta Lake.

VACUUM Reduces storage costs by deleting data files no longer referenced by the table. See Remove unused data files with vacuum.

upvoted 3 times

You need to create a data loading pattern for a Type 1 slowly changing dimension (SCD).

Which two actions should you include in the process? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Update rows when the non-key attributes have changed.
- B. Insert new rows when the natural key exists in the dimension table, and the non-key attribute values have changed.
- C. Update the effective end date of rows when the non-key attribute values have changed.
- D. Insert new records when the natural key is a new value in the table.

**Suggested Answer: AC**

Community vote distribution

AD (100%)

**fabric1** Highly Voted 1 year, 4 months ago

**Selected Answer: AD**

Type 1 SCD does not preserve history, therefore no end dates for table entries exists. A and D are correct.

upvoted 22 times

**CamiU** Most Recent 6 days, 7 hours ago

**Selected Answer: AD**

The wording here is confusing. Instead of "updating" it is about "replacing".

upvoted 1 times

**Rakesh16** 7 months, 2 weeks ago

**Selected Answer: AD**

A & D is the answer

upvoted 1 times

**Joly** 9 months, 4 weeks ago

A and D

upvoted 1 times

**Ahmadpbi** 11 months, 2 weeks ago

AD because C is representing type2 SCD

upvoted 1 times

**woliveiras** 1 year ago

**Selected Answer: AD**

Come on! A and D, for sure. Basic question about data warehouse (SCD)

upvoted 1 times

**282b85d** 1 year, 1 month ago

**Selected Answer: AD**

A. Update rows when the non-key attributes have changed: In a Type 1 SCD, when a change is detected in any of the non-key attributes of an existing row, the current row is updated with the new values. This type of SCD does not keep any historical data; it simply overwrites the old data with the new data.

D. Insert new records when the natural key is a new value in the table: When a new record (with a new natural key) is encountered that does not already exist in the dimension table, it is inserted as a new row. This is necessary to ensure that all new entities are captured in the dimension.

upvoted 2 times

**2dc6125** 1 year, 1 month ago

I would go with A and C: Type1 use case to make changes or correction without involving histo<https://www.examttopics.com/exams/microsoft/dp-600/view/13/#ric> data. D in my opinion is type 2 inserting new record with natural key for versioning.

upvoted 1 times

**stilferx** 1 year, 1 month ago

**Selected Answer: AD**

IMHO, A & D is good.

Taking into account this: A Type 1 SCD always reflects the latest values, and when changes in source data are detected, the dimension table data is overwritten.

Link: <https://learn.microsoft.com/en-us/training/modules/populate-slowly-changing-dimensions-azure-synapse-analytics-pipelines/3-choose-between-dimension-types>

upvoted 2 times

🗨️ 👤 **Nefirs** 1 year, 2 months ago

**Selected Answer: AD**

AD - other options do not concern scd type 1 (overwrite) but scd type 2 (insert new row)

upvoted 1 times

🗨️ 👤 **sraakesh95** 1 year, 4 months ago

**Selected Answer: AD**

AD makes sense given the type of SCD as no history is maintained and any updates on the columns are taken care of by replacing the values then and there.

But, still a bit skeptical about C, since that is a part of SCD too in case we have certain audit columns that we would have to update. If the end\_date is an audit column to update the date and time of last entry into a record, then, it would be required to be updated too.

upvoted 4 times

🗨️ 👤 **thuss** 1 year, 4 months ago

**Selected Answer: AD**

AD = SCD1, BC = SCD2

upvoted 3 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: AD**

No history for scd1

upvoted 1 times

## HOTSPOT -

You have a Fabric workspace named Workspace1 and an Azure Data Lake Storage Gen2 account named storage1. Workspace1 contains a lakehouse named Lakehouse1.

You need to create a shortcut to storage1 in Lakehouse1.

Which connection and endpoint should you specify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

Connection:  ▼  
abfs  
abfss  
https

Endpoint:  ▼  
blob  
dfs  
file

**Suggested Answer:** Box 1: abfss -

Box 2: dfs

🗲️ 👤 **wellingtonluis** Highly Voted 1 year, 3 months ago

HTTPS, DFS

<https://learn.microsoft.com/en-us/fabric/onelake/create-adls-shortcut>

upvoted 43 times

🗲️ 👤 **testtaker45** 5 months, 1 week ago

Someone give WellingtonLuis a medal for this answer. Spot on.

So if we are working in apache spark, we would use the abfss protocol.

As we are creating the shortcut in Fabric the protocol is https, and the endpoint is dfs.core.windows.net. I had to go deep to figure this out, and your documentation was very helpful.

upvoted 5 times

🗲️ 👤 **Felgas** Highly Voted 1 year, 4 months ago

It's asking to create a shortcut in the lakehouse. To do that, the URL should be <https://adls.dfs.core.windows.net/file>.

To access the shortcut in Fabric, you use the abfss path

upvoted 12 times

🗲️ 👤 **mishravikas7** Most Recent 1 month ago

from spark it's abfss and for creating shortcut it's https, and endpoint is dfs

upvoted 1 times

🗲️ 👤 **Ryan2025** 2 months, 2 weeks ago

abfss and dfs

<abfss://rawdata@storage1.dfs.core.windows.net/sales/2024>

upvoted 1 times

🗲️ 👤 **Rakesh16** 7 months, 2 weeks ago

Abfss and dfs

upvoted 4 times

🗲️ 👤 **ca63a55** 1 year ago

I have just passed the exam, instead of 'connection' the question said: 'protocol'

upvoted 5 times

🗲️ 👤 **DarioReymago** 1 year ago

HTTPS, DFS

<https://learn.microsoft.com/en-us/fabric/onelake/onelake-shortcuts#access>

upvoted 4 times

🗨️ 👤 **ManuelG00** 1 year ago

Connection: 2. abfss

Endpoint: 2. dfs

upvoted 2 times

🗨️ 👤 **282b85d** 1 year, 1 month ago

- Connection (abfss): The abfss (Azure Blob File System Secure) protocol is used for secure connections to Azure Data Lake Storage Gen2. This protocol ensures that the connection is encrypted, providing a secure method to access the storage.

- Endpoint (dfs): The dfs (Data Lake Storage) endpoint is used to connect to the hierarchical namespace of Azure Data Lake Storage Gen2, which is optimized for big data analytics workloads. It allows for file and directory-based operations, making it suitable for data lake scenarios.

upvoted 7 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

https, dfs

If you look at the tooltip in the connection settings it has the following about the endpoint, where the example shows https:

"The URL of the ADLSG2 endpoint to connect to. To avoid invalid credential errors, be aware to use the '.dfs' rather than '.blob' endpoint, ensure you are assigned a blob-specific role, and have the networking access set appropriately."

upvoted 3 times

🗨️ 👤 **Valcon\_doo\_NoviSad** 1 year, 3 months ago

Just tried it and when copying https://storage.dfs.core.windows.net/ it worked so definitely http, dfs

upvoted 4 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

Correct path exemple :

abfss://Dev@onelake.dfs.fabric.microsoft.com/lakehouse1...

upvoted 4 times

🗨️ 👤 **IshtarSQL** 1 year, 4 months ago

abfs, dfs

upvoted 3 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

Access Azure storage

Once you have properly configured credentials to access your Azure storage container, you can interact with resources in the storage account using URIs. Databricks recommends using the abfss driver for greater security.

Python

Copy

```
spark.read.load("abfss://<container-name>@<storage-account-name>.dfs.core.windows.net/<path-to-data>")
```

<https://learn.microsoft.com/en-us/azure/databricks/connect/storage/azure-storage>

upvoted 1 times

You are analyzing customer purchases in a Fabric notebook by using PySpark.

You have the following DataFrames:

transactions: Contains five columns named transaction\_id, customer\_id, product\_id, amount, and date and has 10 million rows, with each row representing a transaction. customers: Contains customer details in 1,000 rows and three columns named customer\_id, name, and country.

You need to join the DataFrames on the customer\_id column. The solution must minimize data shuffling.

You write the following code.

```
from pyspark.sql import functions as F
```

```
results =
```

Which code should you run to populate the results DataFrame?

- A. `transactions.join(F.broadcast(customers), transactions.customer_id == customers.customer_id)`
- B. `transactions.join(customers, transactions.customer_id == customers.customer_id).distinct()`
- C. `transactions.join(customers, transactions.customer_id == customers.customer_id)`
- D. `transactions.crossJoin(customers).where(transactions.customer_id == customers.customer_id)`

**Suggested Answer: A**

Community vote distribution

A (100%)

 **Momoanwar** Highly Voted 10 months, 2 weeks ago

**Selected Answer: A**

In Apache Spark, broadcasting refers to an optimization technique for join operations. When you join two DataFrames or RDDs and one of them is significantly smaller than the other, Spark can "broadcast" the smaller table to all nodes in the cluster. This approach avoids the need for network shuffles for each row of the larger table, significantly reducing the execution time of the join operation.

upvoted 31 times

 **sraakesh95** Highly Voted 10 months ago

**Selected Answer: A**

A - Broadcasting generates a copy of the data across all the nodes in the Spark cluster. Therefore, during a join operation, it won't require any I/Os from other nodes, thereby, reducing the shuffling requirement.

upvoted 7 times

 **282b85d** Most Recent 7 months ago

**Selected Answer: A**

Broadcasting: The `F.broadcast(customers)` function is used to broadcast the smaller DataFrame (customers). This ensures that the smaller DataFrame is replicated across all nodes, and each node can perform the join locally with its partition of the larger DataFrame (transactions). This significantly reduces the data movement (shuffling) required during the join operation.

upvoted 1 times

 **stilferx** 7 months, 3 weeks ago

**Selected Answer: A**

IMHO, "A" is correct!

Broadcast joining copies the smaller table to each worker in Spark, which may significantly improve performance by reducing shuffling

upvoted 3 times

 **SamuComqi** 10 months, 2 weeks ago

**Selected Answer: A**

A. `transactions.join(F.broadcast(customers), transactions.customer_id == customers.customer_id)`

Optimized method to perform a join between a very large table and a smaller one.

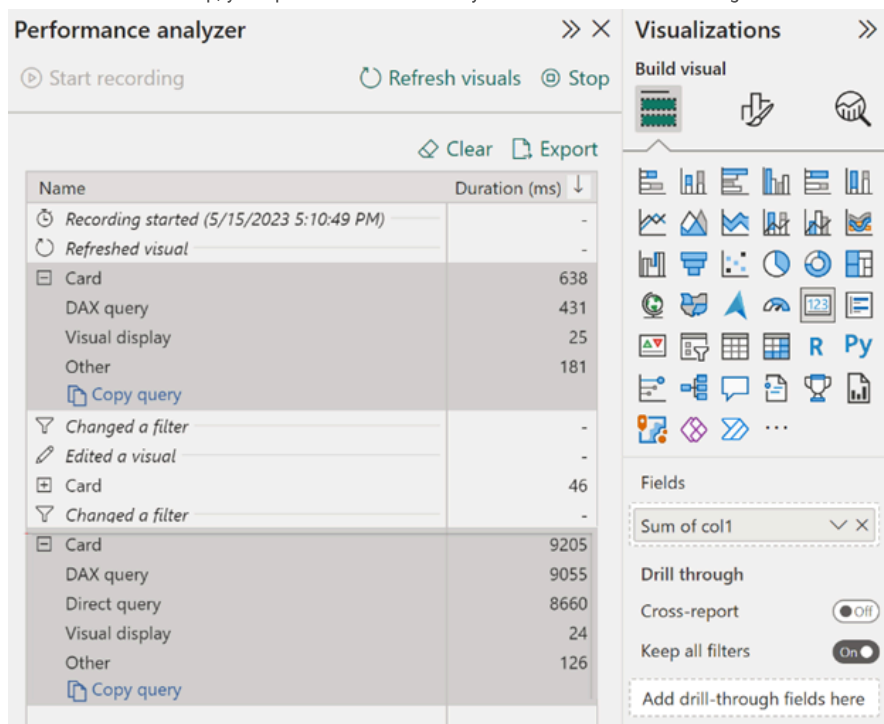
Source: <https://sparkbyexamples.com/spark/broadcast-join-in-spark/>

upvoted 2 times

## HOTSPOT -

You have a Microsoft Power BI report and a semantic model that uses Direct Lake mode.

From Power BI Desktop, you open Performance analyzer as shown in the following exhibit.



Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

### Answer Area

The Direct Lake fallback behavior is set to

▼

Automatic

DirectLakeOnly

DirectQueryOnly

The query for the table visual is executed by using

▼

the composite model

Direct Lake

Direct Query

### Answer Area

Suggested Answer:

The Direct Lake fallback behavior is set to

▼

Automatic

DirectLakeOnly

DirectQueryOnly

The query for the table visual is executed by using

▼

the composite model

Direct Lake

Direct Query

**lordcarlosv** Highly Voted 1 year, 4 months ago

The answer is Automatic and Direct Lake, actually the picture comes from

<https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-analyze-qv>

In this article you can see there are table1 and view1, performance analyser shows:

- First card is linked to Table1 so direct lake is used
- Second card is linked to View1 so it does direct query

As the model can use direct lake and direct query you can conclude that the fallback behavior is automatic.

For direct lake behavior you can read this: <https://powerbi.microsoft.com/en-us/blog/leveraging-pure-direct-lake-mode-for-maximum-query-performance>

upvoted 32 times



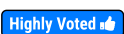
  **SKM1964** 6 months, 1 week ago

In the article mentioned, it is written

"If the semantic model falls back to DirectQuery mode to process the visual's DAX query, you see a Direct query performance metric, as shown in the following image:".

So in my opinion, it should be Automatic and DirectQuery

upvoted 4 times

  **wellingtonluis**  1 year, 3 months ago

There is no table visual in this image.

upvoted 8 times

  **Felix\_G** 1 year, 1 month ago

The table visual is on the bottom of the link, it says "processed in Direct Lake Mode." on top of the table.

upvoted 1 times

  **PradeepikaGadai**  1 month ago

The Direct Lake fallback behavior is set to:

Answer: Automatic

Since the semantic model uses Direct Lake mode, but the query was executed using Direct Query, this implies a fallback occurred.

If fallback occurred and wasn't blocked, it means the fallback behavior must be set to Automatic (allows fallback to DirectQuery when Direct Lake is not possible).

The query for the table visual is executed by using:

Answer: Direct Query

This is explicitly shown in the Performance Analyzer under the "Direct query" row, which indicates that the table visual ran using Direct Query, not Direct Lake.

upvoted 1 times

  **hecke** 2 months, 2 weeks ago

Since the right half of the screenshot is missing I suggest visiting here (as already mentioned below):

<https://learn.microsoft.com/en-us/fabric/fundamentals/direct-lake-analyze-query-processing>

There you can see that the first performance measurement was done with a table and the second one with a view. Hence the 2nd question is referring to the first performance measurement.

When DirectQuery is not mentioned the model uses DirectLake - since one measure is using DirectLake and one DirectQuery the mode is set to automatic.

upvoted 1 times

  **gtc108** 1 year ago

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-analyze-query-processing>

Direct Lake Mode, Direct Query

upvoted 2 times

  **282b85d** 1 year, 1 month ago

1. The Direct Lake fallback behavior is set to: DirectQueryOnly

The Performance analyzer shows that the query type is "Direct query". This indicates that the fallback behavior is set to only use DirectQuery and not Direct Lake. If the behavior were set to "DirectLakeOnly", the query would fail if Direct Lake could not be used. "Automatic" would use Direct Lake when possible and fall back to DirectQuery, but since it's specifically showing "Direct query", it suggests "DirectQueryOnly".

2. The query for the table visual is executed by using: Direct Query

The Performance analyzer directly mentions "Direct query" as the query type for the visual. This confirms that the data is being retrieved using DirectQuery mode, not Direct Lake or any composite model.

upvoted 1 times

  **AdventureChick** 6 months, 2 weeks ago



No, the 1st card shows that Direct Lake was used. DirectQueryOnly means that only Direct Query can be used.

Automatic mode - it uses Direct Lake where it can and fallbacks to Direct Query

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-manage#set-the-direct-lake-behavior-property>

upvoted 1 times

🗨️ 👤 **AndreaRosho1** 1 year, 1 month ago

Answer: Automatic - DirectLake.

the image is cut bad. if you follow this link: <https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-analyze-query-processing> you can see that in Data section of Power BI they are using two different tables

upvoted 7 times

🗨️ 👤 **282b85d** 1 year ago

In the link provided says: If the semantic model falls back to DirectQuery mode to process the visual's DAX query, you see a "Direct query" performance metric, as shown in the following image (the image of question). Is that correct?

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO, Automatic & DirectQuery

upvoted 2 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

sorry, Automatic & Direct Lake

upvoted 1 times

🗨️ 👤 **Valcon\_doo\_NoviSad** 1 year, 3 months ago

The question itself is uniquely confusing and ambiguous and at times simply wrong, but I would go with Automatic and DirectQuery bc DAX falls back to DirectQuery always (it is a known shortcoming --> all explained here: <https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-overview>)

upvoted 3 times

🗨️ 👤 **Valcon\_doo\_NoviSad** 1 year, 3 months ago

Additionally, this link (<https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-analyze-qp>) explicitly says the following: If the dataset falls back to DirectQuery mode to process the visual's DAX query, you see a Direct query performance metric, as shown in the following image --> and the image is the exact same as in the question.

upvoted 5 times

🗨️ 👤 **sraakesh95** 1 year, 4 months ago

Automatic, DirectLake

IMO, as the fallback method is not explicitly mentioned, by default any DirectLake query falls. back to DirectQuery, hence, it is automatic.

After the changes have been made, the query method has fallen back to DirectQuery, hence, I would choose DirectQuery.

Would be great to know further opinion into this.

upvoted 4 times

🗨️ 👤 **Momoanwar** 1 year, 4 months ago

I think automatic and direct lake : we have fallback direct query for card and no informations about table visual so its direct lake.

upvoted 3 times

🗨️ 👤 **a\_51** 1 year, 3 months ago

The final block of card shows Direct Query. I think the wording was wrong as there are no tables shown. So I suggest automatic and Direct query as noted by others.

upvoted 3 times

## HOTSPOT -

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Nyctaxi\_raw. Nyctaxi\_raw contains the following table:

Name	Data type
pickupDateTime	Timestamp
passengerCount	Integer
fareAmount	Double
paymentType	String
tipAmount	Double

You create a Fabric notebook and attach it to Lakehouse1.

You need to use PySpark code to transform the data. The solution must meet the following requirements:

Add a column named pickupDate that will contain only the date portion of pickupDateTime.

Filter the DataFrame to include only rows where fareAmount is a positive number that is less than 100.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

**Answer Area**

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")
```

```
df2= ( "pickupDate", df['tpepPickupDateTime']
```

df.columns  
df.select  
df.withColumn  
df.withColumnsRenamed

.alias('date')  
.cast('date')  
.cast('pickupDate')  
.getfield('date')

.filter("fareAmount > 0 AND fareAmount < 100")  
.filter(col("fareAmount").contains("1..100"))  
.when(df.fareAmount > 0 AND fareAmount < 100)  
.where(df.fareAmount.isin([1,100]))

**Answer Area**

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")
```

```
df2= ( "pickupDate", df['tpepPickupDateTime']
```

df.columns  
df.select  
df.withColumn  
df.withColumnsRenamed

Suggested Answer:

.alias('date')  
.cast('date')  
.cast('pickupDate')  
.getfield('date')

.filter("fareAmount > 0 AND fareAmount < 100")  
.filter(col("fareAmount").contains("1..100"))  
.when(df.fareAmount > 0 AND fareAmount < 100)  
.where(df.fareAmount.isin([1,100]))

 **XiltroX** Highly Voted 1 year, 4 months ago

The correct answers:

1. withColumn

2. cast('date')  
3. filter('fareAmount.....'  
upvoted 44 times

🗨️ 👤 **Momoanwar** Highly Voted 👍 1 year, 4 months ago

We need to add column not rename existing column. Here is the correct answer :

```
df.withColumn('pickupDate', df['pickupDateTime'].cast(DateType())) \
.filter("fareAmount > 0 AND fareAmount < 100")
```

upvoted 12 times

🗨️ 👤 **c119533** Most Recent 🕒 10 months, 3 weeks ago

Tested answer

```
df = spark.read.format("delta").load("Tables/factinternetsales")
df2 = df.withColumn("pickupdate", df['pickupDateTime'].cast('date')).filter("fareAmount > 0 AND fareAmount < 100")
df2.show()
```

upvoted 2 times

🗨️ 👤 **woliveiras** 1 year ago

The correct answers:

1. withColumn
2. cast('date')
3. .filter('fareAmount

Tested!!!

upvoted 8 times

🗨️ 👤 **72bd3bc** 1 year, 1 month ago

why the expression with where is not correct? Is it because it includes 100? not less than 100?

upvoted 1 times

🗨️ 👤 **PiyushT** 1 year, 1 month ago

.filter(fareamount >0 and <100) does not work, I tried the code myself. You have to use same condition but with "col" like this .filter(col(fareamount) > 0 and col(fareamount) < 100)

upvoted 2 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO,

withColumn -> cast(dateType()) -> filter("fareAmount > 0 AND fareAmount < 100")  
is correct.

As colleagues said, cast(dateType()), not cast('date'). Let's consider it a typo from Microsoft.

upvoted 6 times

🗨️ 👤 **earlqq** 1 year, 4 months ago

A bit incorrect.

```
.withcolumn
.cast(date)
.filter(fareamount >0 and <100)
```

upvoted 8 times

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Fabric tenant that contains a new semantic model in OneLake.

You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression:

```
df.explain()
```

Does this meet the goal?

A. Yes

B. No

**Suggested Answer: B**

Community vote distribution

B (100%)


  **282b85d**  7 months ago

**Selected Answer: B**

The df.explain() method in PySpark is used to print the logical and physical plans of a DataFrame, which helps in understanding how Spark plans to execute the query. It does not compute any statistical values like min, max, mean, or standard deviation.

\*\*To achieve the goal, you should use: df.describe().show()

upvoted 10 times

  **SamuComqi**  10 months, 2 weeks ago

**Selected Answer: B**

The correct syntax is df.describe().

Sources:

\* describe --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.describe.html>

\* explain --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.explain.html>

upvoted 5 times

  **SamuComqi** 10 months, 2 weeks ago

Also df.summary() is a valid solution.

Source --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html>

upvoted 1 times

  **stilferx**  7 months, 3 weeks ago

**Selected Answer: B**

IMHO, NOOO

explain() shows the execution plan...

upvoted 1 times

  **a\_51** 9 months, 1 week ago

**Selected Answer: B**

describe is how you get the information.

upvoted 2 times

  **Momoanwar** 10 months, 2 weeks ago

**Selected Answer: B**

No explain is for the execut plan

upvoted 4 times

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Fabric tenant that contains a new semantic model in OneLake.

You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression:

```
df.show()
```

Does this meet the goal?

A. Yes

B. No

**Suggested Answer: B**

Community vote distribution

B (100%)

🗳️ 👤 **282b85d** 7 months ago

**Selected Answer: B**

Correct methods: Use `df.describe().show()` for basic statistics and `df.agg()` with appropriate functions (min, max, mean, stddev) for detailed statistics.

upvoted 3 times

🗳️ 👤 **stilferx** 7 months, 3 weeks ago

**Selected Answer: B**

IMHO, NOOOOO

`df.show()` - shows the data in the dataframe

upvoted 2 times

🗳️ 👤 **a\_51** 9 months, 1 week ago

**Selected Answer: B**

Use describe

upvoted 2 times

🗳️ 👤 **XiltroX** 10 months, 1 week ago

**Selected Answer: B**

`df.summary()` is the only right answer.

upvoted 1 times

🗳️ 👤 **SamuComqi** 10 months, 2 weeks ago

**Selected Answer: B**

The correct syntax is `df.describe()`.

Sources:

\* describe --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.describe.html>

\* show --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.show.html>

upvoted 2 times

🗳️ 👤 **SamuComqi** 10 months, 2 weeks ago

Also `df.summary()` is a valid solution.

Source --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html>

upvoted 1 times

🗳️ 👤 **Momoanwar** 10 months, 2 weeks ago

**Selected Answer: B**

No show is to display data  
upvoted 1 times

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Fabric tenant that contains a new semantic model in OneLake.

You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression:

```
df.summary()
```

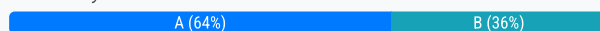
Does this meet the goal?

A. Yes

B. No

**Suggested Answer: A**

Community vote distribution



**stilferx** Highly Voted 1 year, 1 month ago

IMHO, A

Example:

```
df1 = spark.createDataFrame([(1, 10), (2, 10), (2, 15)], schema = ['fruit_id', 'amount'])
```

```
df1.summary()
```

```
summary fruit_id amount
count 3 3
mean 1.6666666666666667 11.666666666666666
stddev 0.5773502691896257 2.886751345948129
min 1 10
25% 1 10
50% 2 10
75% 2 15
max 2 15
upvoted 9 times
```

**Lotusss** Most Recent 2 months, 3 weeks ago

**Selected Answer: B**

df.summary() alone wont cut it.

df.summary().show() is correct. So anwser is B

upvoted 1 times

**b01d700** 4 months, 2 weeks ago

**Selected Answer: B**

The correct PySpark expression to calculate min, max, mean, and standard deviation for both numeric and STRING columns is:

```
df.describe()
```

upvoted 1 times

**slu239** 6 months ago

**Selected Answer: B**

Not meet the goal because it has to be df.summary().show()

upvoted 1 times

**2fe10ed** 6 months, 2 weeks ago

**Selected Answer: A**

<https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html#pyspark.sql.DataFrame.summary>

upvoted 1 times

🗋️ 👤 **Pegooli** 11 months ago

**Selected Answer: B**

Using `df.summary()` in PySpark will provide summary statistics, including min, max, mean, and standard deviation for all numeric columns. However, it will not provide these statistics for string columns since summary statistics like min, max, mean, and standard deviation are not applicable to string data.

upvoted 2 times

🗋️ 👤 **gover07** 10 months ago

so the questions doesn't make sense if you are asked to calculate things that aren't defined

upvoted 2 times

🗋️ 👤 **6d1de25** 11 months, 3 weeks ago

**Selected Answer: A**

Correct

upvoted 1 times

🗋️ 👤 **7d97b62** 11 months, 3 weeks ago

**Selected Answer: A**

In pandas, use `df.describe()` for summary statistics of numeric columns.

In PySpark, use `df.summary()` for summary statistics of both numeric and string columns in a distributed computing environment.

upvoted 3 times

🗋️ 👤 **282b85d** 1 year, 1 month ago

**Selected Answer: B**

while `df.summary()` does provide valuable information for numeric columns, it does not fully meet the goal of evaluating both string and numeric columns with the required statistical measures. Use `df.summary()` and `df.agg()` to cover numeric columns, and additional custom aggregations for string columns.

upvoted 4 times

🗋️ 👤 **XiltroX** 1 year, 4 months ago

`df.summary()` is the only option where you can get MIX, MAX and AVG

upvoted 1 times

🗋️ 👤 **SamuComqi** 1 year, 4 months ago

**Selected Answer: A**

Also `df.describe()` is a valid solution.

Sources:

\* summary --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.summary.html>

\* describe --> <https://spark.apache.org/docs/latest/api/python/reference/pyspark.sql/api/pyspark.sql.DataFrame.describe.html>

upvoted 4 times

🗋️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: A**

Correct

upvoted 2 times



Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer. When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table. You need to identify whether maintenance tasks were performed on Customer. Solution: You run the following Spark SQL statement:

DESCRIBE HISTORY customer -  
Does this meet the goal?

- A. Yes
- B. No

**Suggested Answer: A**

Community vote distribution

A (100%)

🗳️ 👤 **Martin\_Nbg** 8 months, 3 weeks ago

OPTIMIZE and VACUUM activities in the last 30 days are stored in the history. So, yes, you can see using this query whether maintenance took place in the last 30 days. <https://learn.microsoft.com/en-us/azure/databricks/delta/history>  
upvoted 3 times

🗳️ 👤 **282b85d** 1 year, 1 month ago

**Selected Answer: A**

Yes, running DESCRIBE HISTORY customer meets the goal of identifying whether maintenance tasks were performed on the Delta table.  
upvoted 3 times

🗳️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: A**

IMHO, A

Link: <https://learn.microsoft.com/en-us/azure/databricks/delta/history>  
upvoted 4 times

🗳️ 👤 **a\_51** 1 year, 3 months ago

**Selected Answer: A**

A correct.  
`display(spark.sql('describe history customer'))`  
upvoted 2 times

🗳️ 👤 **XiltroX** 1 year, 4 months ago

**Selected Answer: A**

Right answer. Describe HISTORY  
upvoted 2 times

🗳️ 👤 **Momoanwar** 1 year, 4 months ago

**Selected Answer: A**

Correct  
upvoted 2 times

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer. When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table. You need to identify whether maintenance tasks were performed on Customer. Solution: You run the following Spark SQL statement:

REFRESH TABLE customer -  
Does this meet the goal?

- A. Yes
- B. No

**Suggested Answer: B**

*Community vote distribution*

B (100%)

🗳️ 👤 **282b85d** Highly Voted 👍 7 months ago

**Selected Answer: B**

No, running REFRESH TABLE customer does not meet the goal of identifying whether maintenance tasks were performed on the Delta table. This Spark SQL command is used to refresh the metadata of a table. It ensures that the latest schema and data are available for queries but does not give any historical information about maintenance operations.  
upvoted 6 times

🗳️ 👤 **stilferx** Most Recent ⌚ 7 months, 3 weeks ago

**Selected Answer: B**

IMHO, NOOO  
upvoted 1 times

🗳️ 👤 **dp600** 8 months ago

**Selected Answer: B**

correct  
upvoted 1 times

🗳️ 👤 **Momoanwar** 10 months, 2 weeks ago

**Selected Answer: B**

Correct  
upvoted 1 times

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Customer. When you query Customer, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table. You need to identify whether maintenance tasks were performed on Customer.

Solution: You run the following Spark SQL statement:

EXPLAIN TABLE customer -

Does this meet the goal?

A. Yes

B. No

**Suggested Answer: B**

Community vote distribution

B (100%)

🗳️ 👤 **282b85d** Highly Voted 👍 7 months ago

**Selected Answer: B**

The EXPLAIN statement in Spark SQL is used to display the execution plan of a query. This plan shows how Spark will execute the query, including details about the operations and stages involved. While it is useful for understanding and optimizing query performance, it does not provide historical information about maintenance tasks like optimization, compaction, or vacuuming performed on the table.

upvoted 7 times

🗳️ 👤 **stilferx** Most Recent ⌚ 7 months, 3 weeks ago

**Selected Answer: B**

IMHO, NOOOO

upvoted 1 times

🗳️ 👤 **wellingtonluis** 9 months, 3 weeks ago

**Selected Answer: B**

Good luck !!!

upvoted 2 times

🗳️ 👤 **XiltroX** 10 months ago

**Selected Answer: B**

Correct answer is DESCRIBE HISTORY customer.

Good luck everyone.

upvoted 2 times

🗳️ 👤 **earlqq** 10 months, 1 week ago

Correct

upvoted 1 times

🗳️ 👤 **Momoanwar** 10 months, 2 weeks ago

**Selected Answer: B**

Given answer is correct. Explain is for query execution plan

upvoted 2 times

## Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview -

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

## Existing Environment -

## Fabric Environment -

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

## Available Data -

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

- Survey

- Question
- Response

For each survey submitted, the following occurs:

- One row is added to the Survey table.
- One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score. Customers can submit a survey after each purchase.

#### User Problems -

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

#### Requirements -

#### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

- AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
- DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
- DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

- A data store (type to be decided)
- A custom semantic model
- A default semantic model
- Interactive reports

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

#### Technical Requirements -

The data store must support the following:

- Read access by using T-SQL or Python
- Semi-structured and unstructured data
- Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

- List prices that are less than or equal to 50 are in the low pricing group.
- List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
- List prices that are greater than 1,000 are in the high pricing group.

#### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

- Fabric administrators will be the workspace administrators.
- The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
- The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
- The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
- The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
- The date dimension must be available to all users of the data store.
- The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

- FabricAdmins: Fabric administrators
- AnalyticsTeam: All the members of the analytics team
- DataAnalysts: The data analysts on the analytics team
- DataScientists: The data scientists on the analytics team
- DataEngineers: The data engineers on the analytics team
- AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

- Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
- Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.
- Shows data as soon as the data is updated in the data store.
- Ensures that the report and the semantic model only contain data from the current and previous year.
- Ensures that the report respects any table-level security specified in the source data store.
- Minimizes the execution time of report queries.

You need to recommend a solution to prepare the tenant for the PoC.

Which two actions should you recommend performing from the Fabric Admin portal? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Enable the Users can try Microsoft Fabric paid features option for the entire organization.
- B. Enable the Users can try Microsoft Fabric paid features option for specific security groups.
- C. Enable the Allow Azure Active Directory guest users to access Microsoft Fabric option for specific security groups.
- D. Enable the Users can create Fabric items option and exclude specific security groups.
- E. Enable the Users can create Fabric items option for specific security groups.

**Suggested Answer: BE**

Community vote distribution

BE (100%)



  **282b85d**  6 months, 3 weeks ago

**Selected Answer: BE**

Enable the Users can try Microsoft Fabric paid features option for specific security groups: This will allow specific security groups (like the AnalyticsTeam, DataAnalysts, DataScientists, DataEngineers, and AnalyticsEngineers) to access and use the paid features of Microsoft Fabric necessary for the PoC. This is important to ensure that only the relevant team members can utilize these advanced features while preventing unnecessary access for other users.

Enable the Users can create Fabric items option for specific security groups: This will allow only specific security groups to create Fabric items, ensuring that the creation of these items is controlled and managed by the appropriate team members. This helps maintain the principle of least privilege and ensures that only authorized personnel can create and manage Fabric items during the PoC.

upvoted 11 times



  **stilferx**  7 months, 3 weeks ago

**Selected Answer: BE**

IMHO, B & E looks good.

Here, we use only for "specific groups", but not "C" which is for guests.



upvoted 2 times

  **VAzureD** 8 months, 1 week ago

**Selected Answer: BE**

I agree with the answer. The POC is only for a certain group of users.

upvoted 2 times

  **4371883** 8 months, 1 week ago

**Selected Answer: BE**

This is a POC with internal users, B & E make sense on the principle of least privilege.

upvoted 2 times

## HOTSPOT

-

## Case study

-

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study

-

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview

-

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

## Existing Environment

-

## Fabric Environment

-

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

## Available Data

-

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB



The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

- Survey
- Question
- Response

For each survey submitted, the following occurs:

- One row is added to the Survey table.
- One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score. Customers can submit a survey after each purchase.

#### User Problems

-

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

#### Requirements

-

#### Planned Changes

-

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

- AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
- DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
- DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

- A data store (type to be decided)
- A custom semantic model
- A default semantic model
- Interactive reports

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data

engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

#### Technical Requirements

-

The data store must support the following:

- Read access by using T-SQL or Python
- Semi-structured and unstructured data
- Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

- List prices that are less than or equal to 50 are in the low pricing group.
- List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
- List prices that are greater than 1,000 are in the high pricing group.

#### Security Requirements

-

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

- Fabric administrators will be the workspace administrators.
- The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
- The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
- The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
- The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
- The date dimension must be available to all users of the data store.
- The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

- FabricAdmins: Fabric administrators
- AnalyticsTeam: All the members of the analytics team
- DataAnalysts: The data analysts on the analytics team
- DataScientists: The data scientists on the analytics team
- DataEngineers: The data engineers on the analytics team
- AnalyticsEngineers: The analytics engineers on the analytics team

## Report Requirements

-

The data analysts must create a customer satisfaction report that meets the following requirements:

- Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
- Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.
- Shows data as soon as the data is updated in the data store.
- Ensures that the report and the semantic model only contain data from the current and previous year.
- Ensures that the report respects any table-level security specified in the source data store.
- Minimizes the execution time of report queries.

You need to design a semantic model for the customer satisfaction report.

Which data source authentication method and mode should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

### Answer Area

Authentication method:   
Basic authentication  
Service principal authentication  
Single sign-on (SSO) authentication

Mode:   
Direct Lake  
DirectQuery  
Import

### Answer Area

Suggested Answer:

Authentication method:   
Basic authentication  
Service principal authentication  
Single sign-on (SSO) authentication


Mode:   
Direct Lake  
DirectQuery  
Import

 **dp600**  1 year, 2 months ago

Direct Lake also supports row-level security and object-level security so that users only see the data they have permission to see.

<https://learn.microsoft.com/es-es/power-bi/enterprise/directlake-overview>

upvoted 14 times

 **Allapanda** 9 months, 3 weeks ago

It's also saying "The analytics team has large volumes of data, some of which is semi-structured", as far as we know Direct Query can't handle semi-structured data, but DirectLake as well.

upvoted 3 times

 **gtc108**  1 year, 1 month ago

SSO and Direct Lake

upvoted 8 times

🗨️ 👤 **pirate84** Most Recent 5 months, 1 week ago

Single Sign-on (SSO) authentication

DirectQuery

I think this is correct 100%

upvoted 1 times

🗨️ 👤 **cafb698** 10 months ago

Where does it say the the Customer Survey Report needs RLS or am I missing something? I see that the report needs Table-level security which is in DirectQuery mode.

IMHO, Service Principle and DirectQuery are the correct answers.

upvoted 2 times

🗨️ 👤 **DarioReymago** 1 year ago

I'll select: SSO and Direct lake

<https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview#single-sign-on-ss-enabled-by-default>

upvoted 5 times

🗨️ 👤 **PaweuG** 1 year ago

I am hesitant... I wouldn't say that Direct Lake mode is super obvious here. Direct Lake does support RLS/OBS and you can configure it via web interface... BUT! From what I understood, if data source (e.g. database) has RLS/OBS configured, semantic model will fall back to Direct Query and obey RLS/OBS at data source.

Am I wrong where?

upvoted 5 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO, Service Principal & Direct Lake looks good.

And yes, Direct Lake supports RLS, proof is here:

Direct Lake also supports row-level security and object-level security so users only see the data they have permission to see.

Link: <https://learn.microsoft.com/en-us/fabric/get-started/direct-lake-overview>

upvoted 2 times

🗨️ 👤 **Fer079** 1 year, 1 month ago

According to the link you shared, it says "By default, Direct Lake models use single sign-on (SSO), so the effective permissions of the interactive user determine if the user is allowed or denied access to the data. If the Direct Lake model is configured to use a fixed identity, the effective permission of the fixed identity determines if users interacting with the semantic model can access the data."

Therefore, why do you think that the authentication method is Service principal (fixed identity) instead of SSO (interactive user)?.

upvoted 8 times

🗨️ 👤 **pppppppppie** 1 year, 2 months ago

it should be service principle and direct lake.

<https://learn.microsoft.com/en-us/power-bi/enterprise/directlake-fixed-identity>

upvoted 6 times

🗨️ 👤 **MultiCloudIronMan** 6 months, 3 weeks ago

Service Principle and Directlake

upvoted 1 times

🗨️ 👤 **neoverma** 1 year, 2 months ago

Direct lake wont support RLS. so it would be DQ

upvoted 2 times

🗨️ 👤 **dp600** 1 year, 2 months ago

Direct lake supports RLS

upvoted 8 times

## Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview -

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

## Existing Environment -

## Fabric Environment -

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

## Available Data -

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

- Survey

- Question
- Response

For each survey submitted, the following occurs:

- One row is added to the Survey table.
- One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score. Customers can submit a survey after each purchase.

#### User Problems -

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

#### Requirements -

#### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

- AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
- DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
- DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

- A data store (type to be decided)
- A custom semantic model
- A default semantic model
- Interactive reports

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

#### Technical Requirements -

The data store must support the following:

- Read access by using T-SQL or Python
- Semi-structured and unstructured data
- Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

- List prices that are less than or equal to 50 are in the low pricing group.
- List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
- List prices that are greater than 1,000 are in the high pricing group.

#### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

- Fabric administrators will be the workspace administrators.
- The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
- The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
- The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
- The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
- The date dimension must be available to all users of the data store.
- The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

- FabricAdmins: Fabric administrators
- AnalyticsTeam: All the members of the analytics team
- DataAnalysts: The data analysts on the analytics team
- DataScientists: The data scientists on the analytics team
- DataEngineers: The data engineers on the analytics team
- AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

- Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
- Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.
- Shows data as soon as the data is updated in the data store.
- Ensures that the report and the semantic model only contain data from the current and previous year.
- Ensures that the report respects any table-level security specified in the source data store.
- Minimizes the execution time of report queries.

You need to implement the date dimension in the data store. The solution must meet the technical requirements.

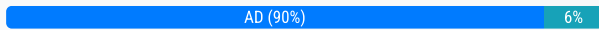
What are two ways to achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Populate the date dimension table by using a dataflow.
- B. Populate the date dimension table by using a Copy activity in a pipeline.
- C. Populate the date dimension view by using T-SQL.
- D. Populate the date dimension table by using a Stored procedure activity in a pipeline.

**Suggested Answer: CD**

Community vote distribution



**neoverma** Highly Voted 1 year, 2 months ago

**Selected Answer: AD**

as per the technical requirements -

"The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year."

No existing data source & The date dimension must always contain dates from 2010 through the end of the current year is the Key Point here

using the elimination method

View wont be appropriate and COPY activity cant be used since there is no data source for DATE table.

so the answer is A and D

upvoted 34 times

**DirectX** Highly Voted 1 year, 1 month ago

C,D.

I always create dimDate using a views (T-Sql select statement with recursive CTE).

A. is too complicated.

upvoted 9 times

**testtaker45** 5 months, 1 week ago

Thank you for your answer. I would say A, D. You would use T-SQL at the data frame layer in a Notebook, this wouldn't be at the Data Source Layer. That is the trick, to be affecting the Data Source.

upvoted 1 times

**os\_ca** 1 year ago

I am not able to create a view with CTE. Could you let us know how?

upvoted 3 times

**zxc01** Most Recent 1 month, 4 weeks ago

**Selected Answer: AB**

A is normal method. C is based on view but we cannot build date table in Lakehouse with view because there are no source data. D is not correct because we cannot build s SP which can load data to lakehouse because SQL endpoint is read only. We cannot add DML, such as insert, delete or update in SP.

B is complicated method, but it can be worked. We can build a SP in lakehouse to generate date dimension data and use copy activity to connect it with SQL endpoint(you cannot find lakehouse and need Azure Synapse Analyse as connection) as Source. And you can use Lakehouse as destination.

upvoted 1 times

**bc5468521** 6 months, 2 weeks ago

**Selected Answer: CD**

Traditional way to create date dimension

upvoted 1 times



🗨️ 👤 **c119533** 9 months, 2 weeks ago

It is specified: "The data store must support the following: Semi-structured and unstructured data".

So the data store is a lakehouse.

Stored procedures don't exist in a lakehouse.

I think A and C

upvoted 3 times

🗨️ 👤 **stv** 11 months, 1 week ago

C, D

it says POPULATE, not CREATE... you cannot create tables with SPs in a lakehouse, BUT you CAN populate already existing ones!

upvoted 1 times

🗨️ 👤 **b6daab0** 1 year ago

ChatGDP chose AB. It says "Copy activity in tools like Azure Data Factory (ADF) allows you to copy data from various sources to your data warehouse. For a date dimension, you can generate the required dates using a source query or script and use the Copy activity to load them into your dimension table." and "While using a stored procedure can also achieve this, it typically requires more complex management and might not be as straightforward or flexible as using dataflows or copy activities, which are specifically designed for ETL (Extract, Transform, Load) processes."

upvoted 1 times

🗨️ 👤 **b6daab0** 1 year ago

correction: ChatGDP says A and C

upvoted 1 times

🗨️ 👤 **b6daab0** 1 year ago

ChatGPT says the answer should be A and D :)

upvoted 1 times

🗨️ 👤 **DarioReymago** 1 year ago

**Selected Answer: AC**

I preferred A & C. B&D need more steps to be created

upvoted 2 times

🗨️ 👤 **6d1de25** 11 months, 3 weeks ago

I agree

upvoted 1 times

🗨️ 👤 **dev2dev** 1 year ago

**Selected Answer: CD**

C & D looks correct. A will be completed. C would need a new object to be created, rather we can simply use a one time script.

upvoted 1 times

🗨️ 👤 **David\_Webb** 1 year ago

**Selected Answer: AD**

A and D should be the right answer to go.

upvoted 1 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

**Selected Answer: AD**

IMHO, AD may be good

Agree, B is not good, C is just a weird way. But possible.

So, the question is kind of ambiguous, with no particular directions from the Microsoft site. Among ABD, I am choosing AD

upvoted 3 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

Sorry, among ACD, I am choosing AD. That's what I meant

upvoted 2 times

## Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview -

Litware, Inc. is a manufacturing company that has offices throughout North America. The analytics team at Litware contains data engineers, analytics engineers, data analysts, and data scientists.

## Existing Environment -

## Fabric Environment -

Litware has been using a Microsoft Power BI tenant for three years. Litware has NOT enabled any Fabric capacities and features.

## Available Data -

Litware has data that must be analyzed as shown in the following table.

Description	Original source	Total size
Customer data	Customer relationship management (CRM) system	50 MB
Product data	Customer relationship management (CRM) system	200 MB
Customer satisfaction surveys	SurveyMonkey	500 GB

The Product data contains a single table and the following columns.

Name	Data type
ProductID	Integer
ProductName	String
ProductCategory	String
ListPrice	Decimal

The customer satisfaction data contains the following tables:

- Survey

- Question
- Response

For each survey submitted, the following occurs:

- One row is added to the Survey table.
- One row is added to the Response table for each question in the survey.

The Question table contains the text of each survey question. The third question in each survey response is an overall satisfaction score. Customers can submit a survey after each purchase.

#### User Problems -

The analytics team has large volumes of data, some of which is semi-structured. The team wants to use Fabric to create a new data store.

Product data is often classified into three pricing groups: high, medium, and low. This logic is implemented in several databases and semantic models, but the logic does NOT always match across implementations.

#### Requirements -

#### Planned Changes -

Litware plans to enable Fabric features in the existing tenant. The analytics team will create a new data store as a proof of concept (PoC). The remaining Liware users will only get access to the Fabric features once the PoC is complete. The PoC will be completed by using a Fabric trial capacity

The following three workspaces will be created:

- AnalyticsPOC: Will contain the data store, semantic models, reports pipelines, dataflow, and notebooks used to populate the data store
- DataEngPOC: Will contain all the pipelines, dataflows, and notebooks used to populate OneLake
- DataSciPOC: Will contain all the notebooks and reports created by the data scientists

The following will be created in the AnalyticsPOC workspace:

- A data store (type to be decided)
- A custom semantic model
- A default semantic model
- Interactive reports

The data engineers will create data pipelines to load data to OneLake either hourly or daily depending on the data source. The analytics engineers will create processes to ingest, transform, and load the data to the data store in the AnalyticsPOC workspace daily. Whenever possible, the data engineers will use low-code tools for data ingestion. The choice of which data cleansing and transformation tools to use will be at the data engineers' discretion.

All the semantic models and reports in the Analytics POC workspace will use the data store as the sole data source.

#### Technical Requirements -

The data store must support the following:

- Read access by using T-SQL or Python
- Semi-structured and unstructured data
- Row-level security (RLS) for users executing T-SQL queries

Files loaded by the data engineers to OneLake will be stored in the Parquet format and will meet Delta Lake specifications.

Data will be loaded without transformation in one area of the AnalyticsPOC data store. The data will then be cleansed, merged, and transformed into a dimensional model

The data load process must ensure that the raw and cleansed data is updated completely before populating the dimensional model

The dimensional model must contain a date dimension. There is no existing data source for the date dimension. The Litware fiscal year matches the calendar year. The date dimension must always contain dates from 2010 through the end of the current year.

The product pricing group logic must be maintained by the analytics engineers in a single location. The pricing group data must be made available in the data store for T-SQL queries and in the default semantic model. The following logic must be used:

- List prices that are less than or equal to 50 are in the low pricing group.
- List prices that are greater than 50 and less than or equal to 1,000 are in the medium pricing group.
- List prices that are greater than 1,000 are in the high pricing group.

#### Security Requirements -

Only Fabric administrators and the analytics team must be able to see the Fabric items created as part of the PoC.

Litware identifies the following security requirements for the Fabric items in the AnalyticsPOC workspace:

- Fabric administrators will be the workspace administrators.
- The data engineers must be able to read from and write to the data store. No access must be granted to datasets or reports.
- The analytics engineers must be able to read from, write to, and create schemas in the data store. They also must be able to create and share semantic models with the data analysts and view and modify all reports in the workspace.
- The data scientists must be able to read from the data store, but not write to it. They will access the data by using a Spark notebook
- The data analysts must have read access to only the dimensional model objects in the data store. They also must have access to create Power BI reports by using the semantic models created by the analytics engineers.
- The date dimension must be available to all users of the data store.
- The principle of least privilege must be followed.

Both the default and custom semantic models must include only tables or views from the dimensional model in the data store. Litware already has the following Microsoft Entra security groups:

- FabricAdmins: Fabric administrators
- AnalyticsTeam: All the members of the analytics team
- DataAnalysts: The data analysts on the analytics team
- DataScientists: The data scientists on the analytics team
- DataEngineers: The data engineers on the analytics team
- AnalyticsEngineers: The analytics engineers on the analytics team

#### Report Requirements -

The data analysts must create a customer satisfaction report that meets the following requirements:

- Enables a user to select a product to filter customer survey responses to only those who have purchased that product.
- Displays the average overall satisfaction score of all the surveys submitted during the last 12 months up to a selected date.
- Shows data as soon as the data is updated in the data store.
- Ensures that the report and the semantic model only contain data from the current and previous year.
- Ensures that the report respects any table-level security specified in the source data store.
- Minimizes the execution time of report queries.

You need to ensure the data loading activities in the AnalyticsPOC workspace are executed in the appropriate sequence. The solution must meet

the technical requirements.

What should you do?

- A. Create a dataflow that has multiple steps and schedule the dataflow.
- B. Create and schedule a Spark notebook.
- C. Create and schedule a Spark job definition.
- D. Create a pipeline that has dependencies between activities and schedule the pipeline.

**Suggested Answer:** D

*Community vote distribution*



🗨️ 👤 **stilferx** 7 months, 2 weeks ago

**Selected Answer: D**

IMHO, D is correct.

I love Microsoft, ambiguity as is. But yes, generally, according to Microsoft BP, a pipeline in ADF with running all activities is the suggested solution. From it, it is easy to trigger different notebooks, for example  
upvoted 2 times

🗨️ 👤 **4371883** 8 months, 1 week ago

**Selected Answer: D**

D is correct. Pipeline can ensure the activities follow the required sequence.

<https://learn.microsoft.com/en-us/fabric/data-factory/activity-overview#data-transformation-activities>

upvoted 2 times

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

### Existing Environment -

### Identity Environment -

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

### Data Environment -

Contoso has the following data environment:

- The Sales division uses a Microsoft Power BI Premium capacity.
- The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.
- The Research department uses an on-premises, third-party data warehousing product.
- Fabric is enabled for contoso.com.
- An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
- A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

### Requirements -

## Planned Changes -

Contoso plans to make the following changes:

- Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
- Make all the data for the Sales division and the Research division available in Fabric.
- For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
- In Productline1ws, create a lakehouse named Lakehouse1.
- In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

## Data Analytics Requirements -

Contoso identifies the following data analytics requirements:

- All the workspaces for the Sales division and the Research division must support all Fabric experiences.
- The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
- The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
- For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.
- For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
- All the semantic models and reports for the Research division must use version control that supports branching.

## Data Preparation Requirements -

Contoso identifies the following data preparation requirements:

- The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
- All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

## Semantic Model Requirements -

Contoso identifies the following requirements for implementing and managing semantic models:

- The number of rows added to the Orders table during refreshes must be minimized.
- The semantic models in the Research division workspaces must use Direct Lake mode.

## General Requirements -

Contoso identifies the following high-level requirements that must be considered for all solutions:

- Follow the principle of least privilege when applicable.
- Minimize implementation and maintenance effort when possible.

You need to recommend which type of Fabric capacity SKU meets the data analytics requirements for the Research division.

What should you recommend?

- A. A
- B. EM
- C. P

D. F

**Suggested Answer:** D

Community vote distribution

D (100%)

🗳️ 👤 **4371883** Highly Voted 👍 8 months, 1 week ago

**Selected Answer:** D

Fabric capacity SKUs all start with F, they allow for Pay as you go by the minute. Reserved capacity is also available.

<https://azure.microsoft.com/en-us/pricing/details/microsoft-fabric/>

upvoted 10 times

🗳️ 👤 **282b85d** Most Recent 🕒 6 months, 3 weeks ago

**Selected Answer:** D

Dedicated, On-Demand Capacity with Per-Minute Billing

Support for All Fabric Experiences

Grouping for OneLake Data Hub Filtering

Read Access via SQL Endpoints and Lakehouse Explorer

upvoted 1 times

🗳️ 👤 **2dc6125** 7 months, 2 weeks ago

Fabric capacity SKU only start with F, so F is the only option.

upvoted 3 times

🗳️ 👤 **stilferx** 7 months, 2 weeks ago

**Selected Answer:** D

IMHO, D,

Requirement: The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.

Why: Microsoft Fabric Capacity (Pay-as-you-go or Reservation)

Get a shared pool of capacity that powers all capabilities in Microsoft Fabric, from data modeling and data warehousing to business intelligence and AI experiences (one minute minimum).

It is F license. Link: <https://azure.microsoft.com/en-us/pricing/details/microsoft-fabric/>

upvoted 2 times

🗳️ 👤 **593150a** 8 months, 1 week ago

D. F because : "The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing."

upvoted 3 times



## HOTSPOT

-

## Case study

-

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study

-

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview

-

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

## Existing Environment

-

## Identity Environment

-

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

## Data Environment

-

Contoso has the following data environment:

- The Sales division uses a Microsoft Power BI Premium capacity.
- The semantic model of the Online Sales department includes a fact table named Orders that uses Import made. In the system of origin, the OrderID value represents the sequence in which orders are created.
- The Research department uses an on-premises, third-party data warehousing product.

- Fabric is enabled for contoso.com.
- An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
- A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

#### Requirements

-

#### Planned Changes

-

Contoso plans to make the following changes:

- Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
- Make all the data for the Sales division and the Research division available in Fabric.
- For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
- In Productline1ws, create a lakehouse named Lakehouse1.
- In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

#### Data Analytics Requirements

-

Contoso identifies the following data analytics requirements:

- All the workspaces for the Sales division and the Research division must support all Fabric experiences.
- The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
- The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
- For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.
- For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
- All the semantic models and reports for the Research division must use version control that supports branching.

#### Data Preparation Requirements

-

Contoso identifies the following data preparation requirements:

- The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
- All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

#### Semantic Model Requirements

-

Contoso identifies the following requirements for implementing and managing semantic models:

- The number of rows added to the Orders table during refreshes must be minimized.
- The semantic models in the Research division workspaces must use Direct Lake mode.

#### General Requirements

-

Contoso identifies the following high-level requirements that must be considered for all solutions:

- Follow the principle of least privilege when applicable.
- Minimize implementation and maintenance effort when possible.

You need to migrate the Research division data for Productline1. The solution must meet the data preparation requirements.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

#### Answer Area

```
df=spark.read.format('csv').options(header="true",inferSchema="true").load("fs1@storage1.dfs.core.windows.net/files/productline1")
```

```
df.write.mode("overwrite").format(" ").save(" ")
```

csv  
delta  
parquet

productline1  
Tables/productline1  
Tables/research/productline1

#### Answer Area

##### Suggested Answer:

```
df=spark.read.format('csv').options(header="true",inferSchema="true").load("fs1@storage1.dfs.core.windows.net/files/productline1")
```

```
df.write.mode("overwrite").format(" ").save(" ")
```

csv  
**delta**  
parquet

productline1  
**Tables/productline1**  
Tables/research/productline1

**9878eb9** Highly Voted 1 year, 2 months ago

Correct.

- "delta"

- "Tables/productline1"

Requirements: Use managed tables.

If you use `saveAsTable()` you don't need to specify the path "Table/"

If you use `save()` you specify the full path

Ignore my previous comment!

upvoted 31 times

**MultiCloudIronMan** 6 months, 3 weeks ago

Something is missing in the folder - The case study mentions that Contoso plans to create a lakehouse named Lakehouse1 in the workspace Productline1ws. Additionally, a shortcut to storage1 named ResearchProduct will be created in Lakehouse1. The data for Productline1 is stored in storage1 in the delta format.

Given this information, the path "Tables/productline1/ResearchProduct" in the code snippet:

upvoted 1 times

**MultiCloudIronMan** 6 months, 3 weeks ago

I changed it should be Tables/productline1

upvoted 2 times

**VAzureD** Highly Voted 1 year, 2 months ago

I just checked it in Fabric,

```
df.write.mode("overwrite").format("delta").save("productline1")
```

It gives an error when executed, "productline1" would be valid if we use `saveAsTable()`.

```
df.write.mode("overwrite").format("delta").save("Tables/productline2")
```

It works, it does what we expect, it is the correct solution.

Create a managed table.

```
df.write.mode("overwrite").format("delta").save("Tables/research/productline3")
```



What it does is load a folder with our dataframe into the lakehouse.

upvoted 10 times

  **Pegooli**  11 months ago

delta and productline1

upvoted 1 times

  **Pegooli** 10 months, 2 weeks ago

sorry correcting myself Tables/productline1

upvoted 1 times

  **72bd3bc** 1 year, 1 month ago

can anyone explain why in the code we read csv file, not delta? if we have in the task:

An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.

Thank you.

upvoted 3 times

  **DarioReymago** 1 year ago

the option read csv in first line is not important. The question is in 2 line.

I in source the file could be csv. but it is not important for this exam

upvoted 1 times

  **stilferx** 1 year, 1 month ago

IMHO, delta and Tables/productline1

Data Preparation Requirements:


- The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
- All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

Answers:

- delta

- Tables/productline1 - because of the root folder in Tables, no subfolder

upvoted 3 times

  **9878eb9** 1 year, 2 months ago

Correct.

- "delta"

- "productline1"

Requirements: Use managed tables.

If you use saveAsTable() you don't need to specify the path "Table/"

If you use save() you specify the full path

upvoted 3 times

### Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

### To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

### Overview -

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

### Existing Environment -

### Identity Environment -

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

### Data Environment -

Contoso has the following data environment:

- The Sales division uses a Microsoft Power BI Premium capacity.
- The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.
- The Research department uses an on-premises, third-party data warehousing product.
- Fabric is enabled for contoso.com.
- An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
- A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

### Requirements -

## Planned Changes -

Contoso plans to make the following changes:

- Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
- Make all the data for the Sales division and the Research division available in Fabric.
- For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
- In Productline1ws, create a lakehouse named Lakehouse1.
- In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

## Data Analytics Requirements -

Contoso identifies the following data analytics requirements:

- All the workspaces for the Sales division and the Research division must support all Fabric experiences.
- The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
- The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
- For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.
- For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
- All the semantic models and reports for the Research division must use version control that supports branching.

## Data Preparation Requirements -

Contoso identifies the following data preparation requirements:

- The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
- All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

## Semantic Model Requirements -

Contoso identifies the following requirements for implementing and managing semantic models:

- The number of rows added to the Orders table during refreshes must be minimized.
- The semantic models in the Research division workspaces must use Direct Lake mode.

## General Requirements -

Contoso identifies the following high-level requirements that must be considered for all solutions:

- Follow the principle of least privilege when applicable.
- Minimize implementation and maintenance effort when possible.

What should you use to implement calculation groups for the Research division semantic models?

- A. Microsoft Power BI Desktop
- B. the Power BI service
- C. DAX Studio
- D. Tabular Editor

**Suggested Answer: B**

Community vote distribution

D (100%)



  **obii**  1 year, 2 months ago

The answer is D

[https://www.bing.com/videos/riverview/relatedvideo?](https://www.bing.com/videos/riverview/relatedvideo?q=how+to+implement+calculation+group+in+fabric&mid=DE305C502B1A6DA8AA03DE305C502B1A6DA8AA03&FORM=VIRE)

[q=how+to+implement+calculation+group+in+fabric&mid=DE305C502B1A6DA8AA03DE305C502B1A6DA8AA03&FORM=VIRE](https://www.bing.com/videos/riverview/relatedvideo?q=how+to+implement+calculation+group+in+fabric&mid=DE305C502B1A6DA8AA03DE305C502B1A6DA8AA03&FORM=VIRE)




upvoted 12 times

  **MM\_GG** 1 year, 2 months ago

seems the answer is really D

<https://powerbi.microsoft.com/en-us/blog/announcing-calculation-groups-for-direct-lake-datasets/>

upvoted 4 times

  **b65ecca**  12 months ago

**Selected Answer: D**

I was studying with practice assessment on Microsoft Learn's own page. I came across to a similar question:

"You have a Fabric workspace that contains a lakehouse named Lakehouse1.

Lakehouse1 requires additional time intelligence calculations added to its semantic model. The model has XMLA read/write permissions enabled.

You need to add a calculation group to the Lakehouse1 semantic model.

What should you use?"

The answer is Tabular Editor and explanation is as follows:

Only Tabular Editor 2/3 can currently add calculation groups to a lakehouse semantic model. <https://learn.microsoft.com/en-us/training/modules/create-calculation-groups/>


upvoted 7 times

  **zxc01**  1 month, 4 weeks ago

**Selected Answer: B**

I think this is old exam question. At that time, D(Tabular Editor) is best option. However, all of A, B and D can match this question. the requirement is "Minimize implementation and maintenance effort when possible.". I think B can match it better now. The reason is you don't need open model file in Power BI desktop or Tabular editor and publish your changes.



upvoted 1 times

  **c51777f** 4 months, 1 week ago

**Selected Answer: B**

You can create calculation groups in the Power BI Service. Why would I need to use a third party tool or power bi desktop to do it?

upvoted 1 times

  **SKM1964** 6 months, 2 weeks ago

**Selected Answer: A**

I am in trouble.

Today Power BI Desktop permits the creation of Calculation Group.

So A seems to be a valid answer

upvoted 5 times

  **maia01** 7 months, 1 week ago

Tabular Editor breaks the data model, you will be able to manage it again ONLY via Tabular Editor and not from Fabric workspace if you save the changes. This is an existing issue with it.

upvoted 1 times

  **stilferx** 1 year, 1 month ago

**Selected Answer: D**

IMHO, D (Tabular Editor)

Requirements:

The semantic models in the Research division workspaces must use Direct Lake mode.

Considering Contoso's requirement to minimize implementation and maintenance effort (general requirement), Tabular Editor offers a more efficient

way to define calculation groups compared to manually writing DAX code. Additionally, since calculation groups are part of the semantic model itself, they can be deployed and managed alongside the model, simplifying maintenance.

upvoted 4 times

🗨️ 👤 **2dc6125** 1 year, 1 month ago

The safe answer is D: C DAX work only with explicit measures, power BI generate dax calculation so I'm not sure but safely D

upvoted 1 times

🗨️ 👤 **zerone72** 1 year, 1 month ago

Since a few months ago you can create calculation group with a recent PowerBI Desktop version too.

<https://learn.microsoft.com/it-it/power-bi/transform-model/calculation-groups>

However, if I was to choose one only answer I'd go for Tabular Editor. You can create calculation group with Tabular editor since PowerBI desktop first version (years ago).

upvoted 3 times

🗨️ 👤 **4371883** 1 year, 2 months ago

**Selected Answer: D**

Should be multiple choice, D is probably the safest option here.

upvoted 1 times

🗨️ 👤 **belha** 1 year, 2 months ago

**Selected Answer: D**

ANSWER IS D

upvoted 2 times

🗨️ 👤 **neoverma** 1 year, 2 months ago

It should be A, using BI desktop... and if minimizing implementation was not mentioned then tabular editor.

upvoted 3 times

🗨️ 👤 **neoverma** 1 year, 2 months ago

Although it can be implemented using service as well. if someone can share the reasoning, that would be great!

<https://powerbi.microsoft.com/en-us/blog/model-explorer-and-calculation-groups-authoring-is-now-available-in-power-bi-service-including-direct-lake-semantic-models/>

upvoted 2 times

🗨️ 👤 **napoleonxiv** 1 year, 2 months ago

The solution should support version control which presumably rules that out

upvoted 1 times

🗨️ 👤 **gfors** 1 year, 2 months ago

So what is minimizing here? This can be done by Tabular editor, power bi desktop and now in the power bi service too according to that article.

Both Tabular editor and Power bi desktop must do some extra things to publish it. Well in this question the answers are very close

upvoted 4 times



## HOTSPOT

-

## Case study

-

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

## To start the case study

-

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

## Overview

-

Contoso, Ltd. is a US-based health supplements company. Contoso has two divisions named Sales and Research. The Sales division contains two departments named Online Sales and Retail Sales. The Research division assigns internally developed product lines to individual teams of researchers and analysts.

## Existing Environment

-

## Identity Environment

-

Contoso has a Microsoft Entra tenant named contoso.com. The tenant contains two groups named ResearchReviewersGroup1 and ResearchReviewersGroup2.

## Data Environment

-

Contoso has the following data environment:

- The Sales division uses a Microsoft Power BI Premium capacity.
- The semantic model of the Online Sales department includes a fact table named Orders that uses Import mode. In the system of origin, the OrderID value represents the sequence in which orders are created.
- The Research department uses an on-premises, third-party data warehousing product.

- Fabric is enabled for contoso.com.
- An Azure Data Lake Storage Gen2 storage account named storage1 contains Research division data for a product line named Productline1. The data is in the delta format.
- A Data Lake Storage Gen2 storage account named storage2 contains Research division data for a product line named Productline2. The data is in the CSV format.

#### Requirements

-

#### Planned Changes

-

Contoso plans to make the following changes:

- Enable support for Fabric in the Power BI Premium capacity used by the Sales division.
- Make all the data for the Sales division and the Research division available in Fabric.
- For the Research division, create two Fabric workspaces named Productline1ws and Productine2ws.
- In Productline1ws, create a lakehouse named Lakehouse1.
- In Lakehouse1, create a shortcut to storage1 named ResearchProduct.

#### Data Analytics Requirements

-

Contoso identifies the following data analytics requirements:

- All the workspaces for the Sales division and the Research division must support all Fabric experiences.
- The Research division workspaces must use a dedicated, on-demand capacity that has per-minute billing.
- The Research division workspaces must be grouped together logically to support OneLake data hub filtering based on the department name.
- For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.
- For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.
- All the semantic models and reports for the Research division must use version control that supports branching.

#### Data Preparation Requirements

-

Contoso identifies the following data preparation requirements:

- The Research division data for Productline1 must be retrieved from Lakehouse1 by using Fabric notebooks.
- All the Research division data in the lakehouses must be presented as managed tables in Lakehouse explorer.

#### Semantic Model Requirements

-

Contoso identifies the following requirements for implementing and managing semantic models:

- The number of rows added to the Orders table during refreshes must be minimized.
- The semantic models in the Research division workspaces must use Direct Lake mode.

#### General Requirements

-

Contoso identifies the following high-level requirements that must be considered for all solutions:

- Follow the principle of least privilege when applicable.
- Minimize implementation and maintenance effort when possible.

Which workspace role assignments should you recommend for ResearchReviewersGroup1 and ResearchReviewersGroup2? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

### Answer Area

ResearchReviewersGroup1:   
Contributor  
Member  
Viewer

ResearchReviewersGroup2:   
Contributor  
Member  
Viewer

### Answer Area

Suggested Answer:

ResearchReviewersGroup1:   
Contributor  
Member  
Viewer

ResearchReviewersGroup2:   
Contributor  
Member  
Viewer

**dp600** Highly Voted 1 year, 2 months ago

correct, you can see in this link: <https://learn.microsoft.com/en-us/fabric/get-started/roles-workspaces>  
upvoted 18 times

**stilferx** Highly Voted 1 year, 1 month ago

IMHO, correct, Viewer & Contributor

Requirements:

- For the Research division workspaces, the members of ResearchReviewersGroup1 must be able to read lakehouse and warehouse data and shortcuts by using SQL endpoints.
- For the Research division workspaces, the members of ResearchReviewersGroup2 must be able to read lakehouse data by using Lakehouse explorer.

From <https://learn.microsoft.com/en-us/fabric/get-started/roles-workspaces>

1) Viewer+: Read Lakehouse and Data warehouse data and shortcuts2 with T-SQL through TDS endpoint.

2) Contributor+: Read Lakehouse data through Lakehouse explorer.

upvoted 12 times

**OlivierFerrari** Most Recent 10 months, 3 weeks ago

1--> Viewer

2 --> Member

upvoted 1 times

**dp600** 1 year, 2 months ago

Viewer and contributor

Read using sql endpoint -> Min. Viewer

Read using lakehouse explorer -> Min. Contributor

upvoted 5 times

  **luiruipu** 1 year, 2 months ago

Answer is correct

upvoted 2 times

HOTSPOT

-

You have a Fabric tenant.

You need to configure OneLake security for users shown in the following table.

Name	Required access
User1	Read all the Spark data.
User2	Read all the SQL endpoint data.

The solution must follow the principle of least privilege.

Which permission should you assign to each user? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

### Answer Area

User1:


User2:

### Answer Area

Suggested Answer:

User1:

User2:

 **neoverma** Highly Voted 1 year, 2 months ago

Correct! ReadAll and readData

<https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sharing#sharing-and-permissions>

upvoted 17 times

 **AbhiShar** Most Recent 8 months, 3 weeks ago

the selections of ReadAll for User1 and Read for User2 ensure that each user gets the appropriate level of access according to their stated needs while adhering to the principle of least privilege. ReadData may imply limited access and could restrict User2 from reading all SQL endpoint data effectively.

upvoted 2 times

 **282b85d** 1 year, 1 month ago

- User1: Assign ReadAll. This is because User1 needs to read all Spark data, which would require broad access across multiple datasets and objects managed by Spark.

- User2: Assign ReadData. This permission is more likely to be specific to data endpoints, including SQL endpoints.

upvoted 3 times

🗨️ 👤 **stilferx** 1 year, 1 month ago

IMHO,

Req: user1 - spark -> \*\*ReadAll\*\*, user 2 - SQL endpoint -> \*\*ReadData\*\*.

Roles:

ReadData permission on SQL endpoint to access data without SQL policy.

ReadAll permission on the lakehouse to access all data using Apache Spark.

from <https://learn.microsoft.com/en-us/fabric/data-engineering/lakehouse-sharing#sharing-and-permissions>

upvoted 4 times

🗨️ 👤 **a998450** 1 year, 2 months ago

ReadAll and ReadData is Correct

<https://blog.fabric.microsoft.com/en/blog/lakehouse-sharing-and-access-permission-management?ft=All>

upvoted 3 times

You have an Azure Repos repository named Repo1 and a Fabric-enabled Microsoft Power BI Premium capacity. The capacity contains two workspaces named Workspace1 and Workspace2. Git integration is enabled at the workspace level.

You plan to use Microsoft Power BI Desktop and Workspace1 to make version-controlled changes to a semantic model stored in Repo1. The changes will be built and deployed to Workspace2 by using Azure Pipelines.

You need to ensure that report and semantic model definitions are saved as individual text files in a folder hierarchy. The solution must minimize development and maintenance effort.

In which file format should you save the changes?

- A. PBIP
- B. PBIDS
- C. PBIT
- D. PBIX

**Suggested Answer: A**

Community vote distribution

A (100%)

🗳️ 👤 **282b85d** Highly Voted 👍 7 months ago

**Selected Answer: A**

PBIP (Power BI Project):

-> PBIP format is designed to work with version control systems like Azure Repos. It breaks down Power BI artifacts into individual files that can be managed and versioned separately, facilitating better collaboration and change tracking.

-> Folder Hierarchy: It saves the project structure in a folder hierarchy, where each component of the Power BI project (like datasets, reports, data sources) is stored in separate files.

-> Text-Based: Being a text-based format, it integrates well with Git repositories and supports diff and merge operations.

upvoted 6 times

🗳️ 👤 **PaweuG** Most Recent 🕒 7 months, 1 week ago

**Selected Answer: A**

A - no-brainer

upvoted 2 times

🗳️ 👤 **stilferx** 7 months, 2 weeks ago

**Selected Answer: A**

IMHO, A) PBIP is correct.

Why:

Power BI Desktop introduces a new way to author, collaborate, and save your projects. When you save your work as a Power BI Project (PBIP), report and semantic model item definitions are saved as individual plain text files in a simple, intuitive folder structure.

Link: <https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-overview>

upvoted 1 times

🗳️ 👤 **luiruipu** 8 months, 4 weeks ago

A.PBIP Correct

upvoted 2 times

🗳️ 👤 **ES\_Capgemini\_landD** 8 months, 4 weeks ago

**Selected Answer: A**

[https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-](https://learn.microsoft.com/en-us/power-bi/developer/projects/projects-overview#:~:text=Power%20BI%20Desktop%20introduces%20a%20new%20way%20to%20author%2C%20collaborate%2C%20and%20save%20your%20projects)

overview#:~:text=Power%20BI%20Desktop%20introduces%20a%20new%20way%20to%20author%2C%20collaborate%2C%20and%20save%20your%20projects

upvoted 1 times

## DRAG DROP

You are implementing a medallion architecture in a single Fabric workspace.

You have a lakehouse that contains the Bronze and Silver layers and a warehouse that contains the Gold layer.

You create the items required to populate the layers as shown in the following table.

Layer	Data integration tool
Bronze	Pipelines with Copy activities
Silver	Dataflows
Gold	Stored procedures

You need to ensure that the layers are populated daily in sequential order such that Silver is populated only after Bronze is complete, and Gold is populated only after Silver is complete. The solution must minimize development effort and complexity.

What should you use to execute each set of items? To answer, drag the appropriate options to the correct items. Each option may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Triggers	Answer Area
A pipeline Copy activity	Orchestration pipeline: <input type="text"/>
A pipeline Dataflow activity	Bronze layer: <input type="text"/>
A pipeline Stored procedure activity	Silver layer: <input type="text"/>
A schedule	Gold layer: <input type="text"/>
A Spark job definition	
An Invoke pipeline activity	

**Suggested Answer:**

Answer Area	
Orchestration pipeline:	A schedule
Bronze layer:	A pipeline Copy activity
Silver layer:	A pipeline Dataflow activity
Gold layer:	A pipeline Stored procedure activity

**c8f5bdf** Highly Voted 1 year, 2 months ago

Bronze Layer answer is wrong. the Bronze Layer already has pipeline defined with it so you need an invoke pipeline activity to call them  
upvoted 28 times

**vernillen** Highly Voted 1 year ago

Should be;

Orchestration pipeline: Schedule

Bronze layer: An invoke pipeline activity (see table -> PIPELINE is already there)

Silver: Dataflow activity

Gold: Stored procedure activity

upvoted 13 times

**b8ef1c5** Most Recent 1 month, 1 week ago