Question #1 *Topic 1*

This is a case study. Case studies are not timed separately from other exam sections. You can use as much exam time as you would like to complete each case study. However, there might be additional case studies or other exam sections. Manage your time to ensure that you can complete all the exam sections in the time provided. Pay attention to the Exam Progress at the top of the screen so you have sufficient time to complete any exam sections that follow this case study.
To answer the case study questions, you will need to reference information that is provided in the case. Case studies and associated questions might contain exhibits or other resources that provide more information about the scenario described in the case. Information provided in an individual question does not apply to the other questions in the case study.
A Review Screen will appear at the end of this case study. From the Review Screen, you can review and change your answers before you move to the next exam section. After you leave this case study, you will NOT be able to return to it.

To start the case study -
To display the first question in this case study, select the “Next” button. To the left of the question, a menu provides links to information such as business requirements, the existing environment, and problem statements. Please read through all this information before answering any questions. When you are ready to answer a question, select the “Question” button to return to the question.

Overview -

Company Information -
Contoso, Inc. is a renewable energy provider that operates solar and wind farms across North America.

Existing Environment -

Azure Environment -
Contoso has a single Azure Databricks workspace named Workspace1 in the West US Azure region. Workspace1 is enabled for Unity Catalog. Workspace1 contains all-purpose clusters for both development and production workloads.
The company's Azure environment contains:
In the West US, Central US, and East US Azure regions, Azure event hubs that stream telemetry data and an Azure Data Lake Storage Gen2 account in each region for each hub
A single Azure SQL database in the West US region that hosts enterprise resource planning (ERP) data
An Azure Database for PostgreSQL server in the West US region that stores operational maintenance data

Data Environment -
Contoso ingests the following operational and business data:
Telemetry data: More than 40,000 IoT sensors across 28 sites emit JSON telemetry events every few seconds. Each site sends the events to the nearest event hub, which writes the data into the corresponding Data Lake Storage Gen2 account. These files frequently experience schema drift.
Maintenance logs: Maintenance systems generate historical repair logs, daily incremental updates, technician notes, and unstructured attachments that are stored in the Data Lake Storage Gen2 accounts.
Operational maintenance data: Structured operational maintenance data is stored on the Azure Database for PostgreSQL server.
External weather data: Hourly weather forecasts are retrieved from a REST API and written to the Data Lake Storage Gen2 accounts.
ERP data: Daily CSV extracts of 50 to 100 GB contain equipment metadata, work orders, and purchase order information.

Problem Statements -
The company’s existing analytics environment has several issues:

Ingestion -
Telemetry pipelines fall behind during peak loads.
Telemetry ingestion fails when schema drift occurs.
Streaming pipelines reprocess events after a pipeline restarts.

Compute -
Production and development workloads run on the same all-purpose clusters.
Production and development workloads do NOT support autoscaling or workload isolation.

Governance -
The ERP data is duplicated across systems and development teams.

Naming conventions are inconsistent across development teams, regions, and products.
Ownership of the IoT sensors changes over time, and analysts must track the full history of the ownership.
Occasionally, equipment manufacturers must correct data-entry mistakes in equipment names. Historical values are NOT required.

Pipeline operations -
Pipelines lack resiliency, alerting, and centralized scheduling.

Requirements -

Planned Changes -
Contoso plans to implement the following changes:
Implement scalable data pipeline orchestration.
Create a managed analytics catalog in Unity Catalog.
Implement a consistent approach to creating curated datasets.
Establish a centralized governance model across ingestion, cleansed, and curated layers.
Grant data engineers access to the ERP tables by using minimal development effort.
Adopt a compute strategy that isolates production workloads and supports autoscaling.
Adopt a slowly changing dimension (SCD) approach to address current data modeling issues.

Technical Requirements -
Contoso identifies the following environment and compute requirements:
Ensure that production ingestion workloads run on compute clusters that can scale automatically during telemetry spikes.
Provide fast and consistent performance for business intelligence (BI) workloads.
Prevent development activity from affecting production pipelines.
Production ingestion workloads must run as scheduled, non-interactive pipelines rather than on shared interactive development clusters.
Contoso identifies the following data ingestion and processing requirements:
Auto-scale ingestion pipelines to handle bursty workloads.
Handle schema drift for the maintenance and telemetry data.
Ingest file-based telemetry data by using minimal operational effort.
Store all the ingested data in a format that supports incremental processing.
Support the continuous ingestion of telemetry data from the event hubs by using exactly-once semantics.
Support the ingestion of the structured maintenance data from the Azure Database for PostgreSQL server.
Build a new telemetry pipeline that ingests raw events from the event hubs, cleanses the data, and publishes curated tables to Unity Catalog.
Ensure that the Apache Spark Structured Streaming pipelines reading from the event hubs write the data into a managed Delta table named telemetry.raw_events. The pipelines must support schema drift and resume processing after failures without reprocessing the data.
Contoso identifies the following data modeling and optimization requirements:
Build curated tables that standardize business logic.
Overwrite equipment metadata attributes, such as name, manufacturer, model, and commissioning date, when the attributes change. Historical values are NOT required.
Contoso identifies the following pipeline deployment and operation requirements:
Orchestrate multi-step ingestion and transformation workflows.
Define a clear execution order and dependencies.
Automatically retry failed steps and notify operators.
Schedule ingestion and transformation workloads consistently.

Governance Requirements -
Contoso identifies the following governance requirements:
Centralize the metadata catalog.
Provide isolated development areas that follow standard naming conventions.
Establish a consistent structure for organizing raw, cleansed, and curated data.
Provide a read-only mechanism to reference the ERP data through a foreign catalog.

Business Requirements -
Contoso identifies the following business requirements:
Improve ingestion reliability and reduce operational effort.
Standardize data definitions across development teams.
You need to configure compute for the ingestion of telemetry data. The solution must meet the data ingestion and processing requirements.
What should you do?

A. Move the ingestion pipelines to shared compute.

B. Enable Photon acceleration for a job compute cluster.

C. Increase an all-purpose cluster to a larger fixed node type.

D. Disable autoscaling for a job compute cluster.

Suggested Answer: *B*

Currently there are no comments in this discussion, be the first to comment!

Question #2 *Topic 1*

You have an Azure Databricks workspace.
You are creating a Lakeflow Spark Declarative Pipelines (SDP) pipeline that scales automatically.
You need to configure compute for the pipeline. The solution must minimize operational costs and effort.
What should you use?

A. the existing SQL warehouse

B. an all-purpose cluster that uses autoscaling

C. a job cluster that uses autoscaling

D. a single-node, all-purpose cluster

Suggested Answer: *C*

Community vote distribution

C (100%)

JackJung 2 weeks ago

Selected Answer: C

Delta Live Tables pipelines utilize cost-efficient, automated job clusters with built-in enhanced autoscaling to handle production pipeline workloads dynamically at the lowest possible cost.

upvoted 2 times

Question #3 *Topic 1*

DRAG DROP -

You have an Azure Databricks workspace that contains an all-purpose compute cluster named Cluster1. Cluser1 is used for interactive development.

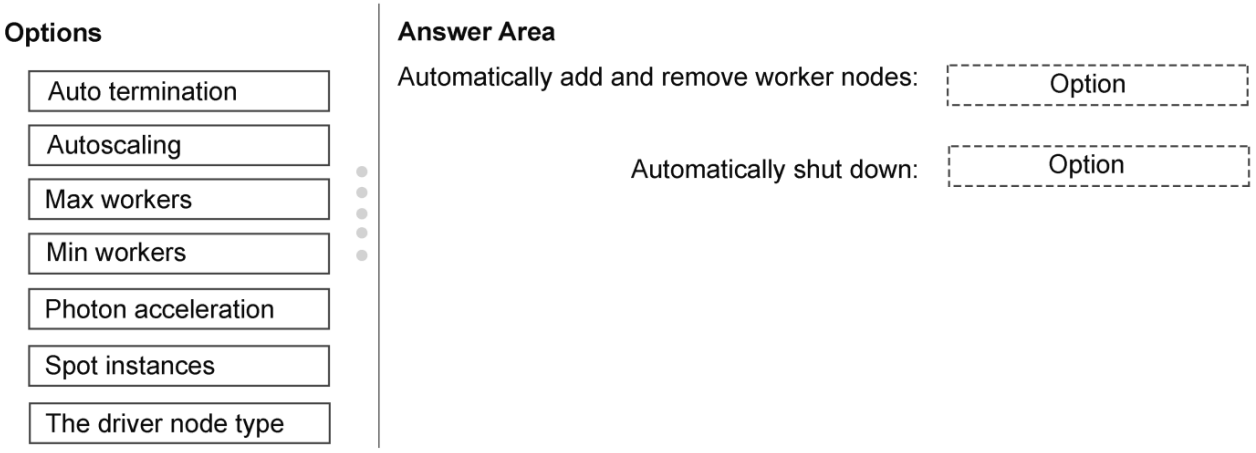
You need to configure Cluster1 to meet the following requirements:

Automatically add and remove worker nodes based on workload demand.

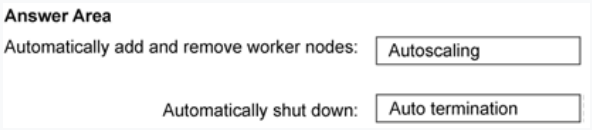
Automatically shut down when the cluster has been idle for a specific period.

What should you configure for each requirement? To answer, drag the appropriate options to the correct requirements. 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.



Suggested Answer:



JackJung 2 weeks ago

Correct answer

Autoscaling / Auto Termination

upvoted 1 times

Question #4 *Topic 1*

You have an Azure Databricks workspace that is attached to a Unity Catalog metastore named metastore1, metastore1 contains a catalog named catalog1.
You need to create a new schema named schema2 that meets the following requirements:

Is contained in catalog1 -
Uses abfss://container@storageaccount.dfs.core.windows.net/data as the managed location
Which SQL statement should you execute?

A. CREATE SCHEMA catalog1.schema2 -
LOCATION 'abfss://container@storageaccount.dfs.core.windows.net/data';

B. CREATE SCHEMA catalog1.schema2 -
MANAGED LOCATION 'abfss://container@storageaccount.dfs.core.windows.net/data';

C. CREATE CATALOG schema2 -
MANAGED LOCATION 'abfss://container@storageaccount.dfs.core.windows.net/data';

D. CREATE SCHEMA catalog1.schema2 -
WITH DBPROPERTIES (LOCATION-'abfss://container@storageaccount.dfs.core.windows.net/data');

Suggested Answer: *B*

Community vote distribution

B (100%)

JackJung 2 weeks ago

Selected Answer: B

The reference URL is https://learn.microsoft.com/en-us/azure/databricks/sql/language-manual/sql-ref-syntax-ddl-create-schema
upvoted 1 times

Question #5 *Topic 1*

You have an Azure Databricks workspace named Workspace1.
You create a compute cluster named Cluster1 that will be used to ingest data.
You need to install the required libraries on Cluster1. The solution must use Unity Catalog for access control.
What should you do?

A. Install the libraries by using pip3.

B. Create a custom dependency management script and run the script from a Databricks notebook.

C. Upload the libraries to Workspace1 and install the libraries on Cluster1.

D. Install the libraries on Cluster1 and manually restart the cluster.

Suggested Answer: *C*

Currently there are no comments in this discussion, be the first to comment!

Question #6 *Topic 1*

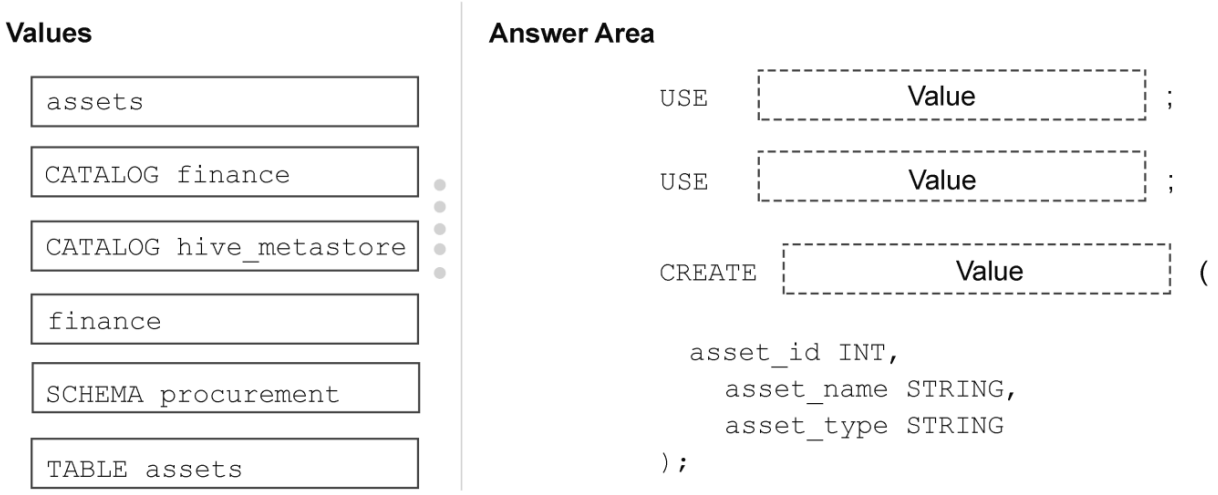
DRAG DROP -

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a catalog named finance, finance contains two schemas named default and procurement.

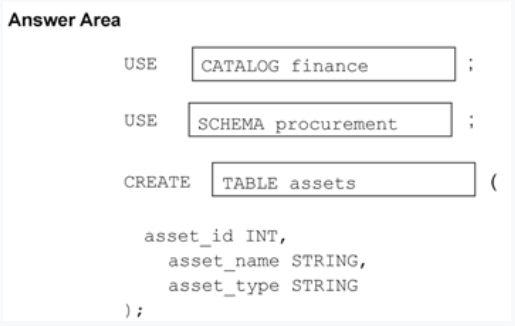
You need to create a table named assets in the procurement schema, assets must contain the following columns: asset_id asset_type asset_name

How should you complete the SQL statement? 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.



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #7 *Topic 1*

You have an Azure Databricks workspace that contains an all-purpose cluster named Cluster1.
You need to configure Cluster1 to meet the following requirements:
Scale up automatically when workloads increase
Scale down automatically when workloads decrease

Minimize costs -
Which two actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Disable Photon acceleration.

B. Enable autoscaling for Cluster1.

C. Apply a compute policy that enables users to manage the cluster settings.

D. Specify a fixed number of workers.

E. Configure Cluster1 to terminate after 30 minutes of inactivity.

Suggested Answer: *BE*

Currently there are no comments in this discussion, be the first to comment!

Question #8 *Topic 1*

You have an Azure Databricks workspace that is enabled for Unity Catalog.
You have an Apache Spark Structured Streaming job that writes data to a Delta table.
After the cluster restarts, the streaming job reprocesses previously ingested data.
You need to prevent the streaming job from reprocessing the data after the cluster restarts.
What should you do?

A. Configure a checkpoint location for the streaming query.

B. Increase the trigger interval of the streaming query.

C. Enable change data feed (CDF) for the target table.

D. Configure a watermark for the streaming query.

Suggested Answer: *A*

Currently there are no comments in this discussion, be the first to comment!

Question #9 *Topic 1*

HOTSPOT -

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a managed Delta table named Table1.

Table1 is written by batch jobs every hour and is queried frequently by filtering two columns named CustomerId and EventDate.

You expect Table1 to grow significantly over time.

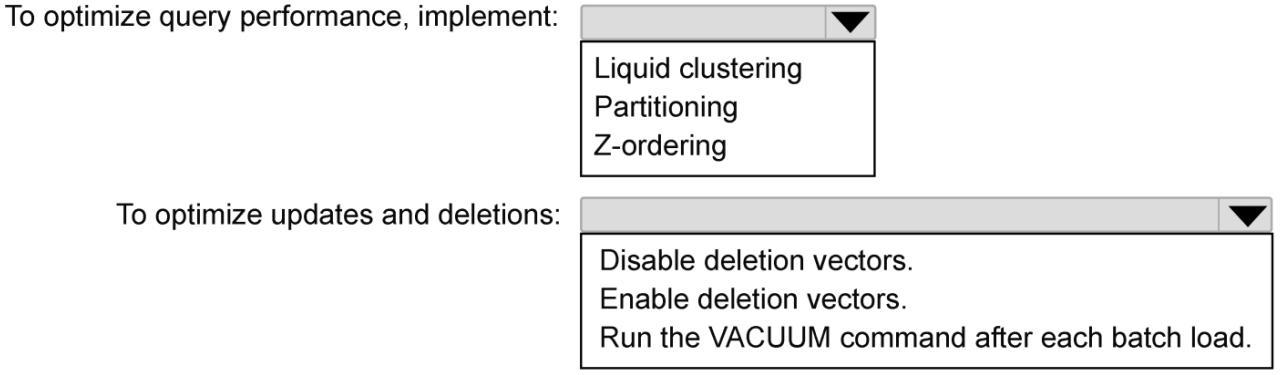
The rows in Table1 are frequently updated and deleted to support compliance requests.

You need to keep query performance consistent as Table1 grows. The solution must minimize update and deletion effort.

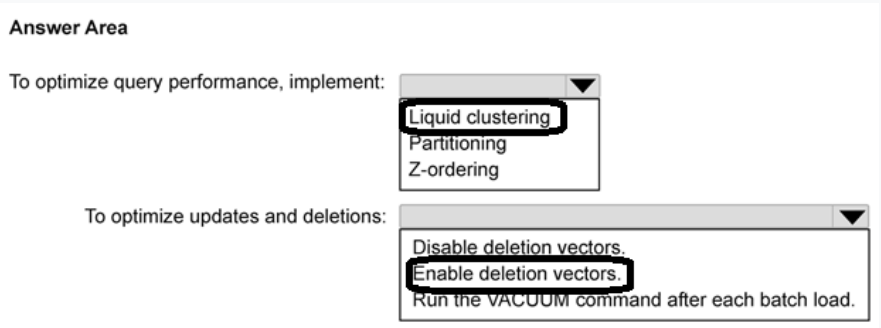
What should you include in the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #10 *Topic 1*

You have an Azure Databricks workspace that contains a cluster named Cluster1.
Performance monitoring shows that Cluster1 is consistently overprovisioned for its batch workload:
CPU utilization remains below 20 percent, including peak processing periods.
The workload is highly predictable and does not spike.
The current node type already meets the workload requirements.
You need to reduce compute costs without increasing job duration.
What should you do?

A. Enable Photon acceleration.

B. Configure Cluster1 to use a larger node type.

C. Decrease the autotermination timeout of Cluster1.

D. Disable autoscaling and reduce the number of worker nodes.

Suggested Answer: *D*

Currently there are no comments in this discussion, be the first to comment!

Question #11 *Topic 1*

You have an Azure Databricks workspace.
You are creating a Lakeflow Spark Declarative Pipelines (SDP) pipeline that scales automatically.
You need to configure compute for the pipeline. The solution must minimize operational costs and administrative effort.
What should you use?

A. serverless compute

B. a single-node, all-purpose cluster

C. an all-purpose cluster that uses autoscaling

D. an existing SQL warehouse

E. a job cluster that uses autoscaling

Suggested Answer: *A*

Currently there are no comments in this discussion, be the first to comment!

Question #12 *Topic 1*

You have an Azure Databricks solution that was developed by multiple engineers and deployed to development, test, and production environments.
You need to implement an application lifecycle management (ALM) process that meets the following requirements:
Deploys and provides source control for Databricks notebooks, workflows, and jobs in a consistent and repeatable way
Parameterizes the target deployment environment
What should you include in the solution?

A. the export and import of Databricks workspaces

B. Databricks resources managed by using the Databricks Terraform provider

C. Git folders and manual promotion

D. Declarative Automation Bundles in Git

Suggested Answer: *D*

Currently there are no comments in this discussion, be the first to comment!

Question #13 *Topic 1*

HOTSPOT -

You have an Azure Databricks workspace that contains a cluster named Cluster1.

Cluster1 has the following characteristics:

Configured as a single node cluster

Uses a general purpose virtual machine node type

The cluster runtime environment has the following configurations:

Uses Databricks Runtime for Machine Learning (Databricks Runtime ML)

Includes common machine learning libraries

Supports Python workloads -

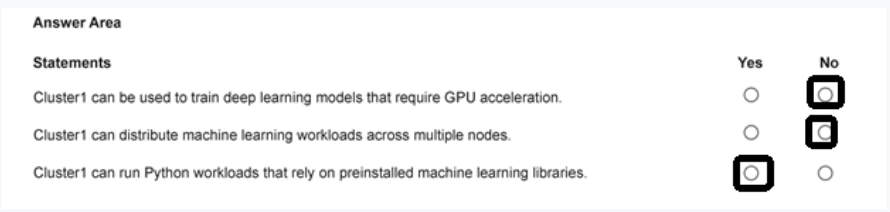
For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
Cluster1 can be used to train deep learning models that require GPU acceleration.	○	○
Cluster1 can distribute machine learning workloads across multiple nodes.	○	○
Cluster1 can run Python workloads that rely on preinstalled machine learning libraries.	○	○

Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #14 *Topic 1*

You have an Azure Databricks workspace that contains the objects shown in the following table.

Name	Type
Catalog1	Catalog
Schema1	Schema
Sales1	Table
Notebook1	Notebook
Space1	AI/BI Genie space

Users often use the following words to refer to a sale: transaction, event, order, and invoice.

You need to create a knowledge store. The solution must ensure that when the users use any of the words in Space1, Genie queries the Sales1 table. Any other Genie spaces must remain unaffected.

To which object should you add the instructions?

A. Sales1

B. Schema1

C. Space1

D. Catalog1

E. Notebook1

Suggested Answer: *C*

Currently there are no comments in this discussion, be the first to comment!

Question #1 *Topic 2*

This is a case study. Case studies are not timed separately from other exam sections. You can use as much exam time as you would like to complete each case study. However, there might be additional case studies or other exam sections. Manage your time to ensure that you can complete all the exam sections in the time provided. Pay attention to the Exam Progress at the top of the screen so you have sufficient time to complete any exam sections that follow this case study.
To answer the case study questions, you will need to reference information that is provided in the case. Case studies and associated questions might contain exhibits or other resources that provide more information about the scenario described in the case. Information provided in an individual question does not apply to the other questions in the case study.
A Review Screen will appear at the end of this case study. From the Review Screen, you can review and change your answers before you move to the next exam section. After you leave this case study, you will NOT be able to return to it.

To start the case study -
To display the first question in this case study, select the “Next” button. To the left of the question, a menu provides links to information such as business requirements, the existing environment, and problem statements. Please read through all this information before answering any questions. When you are ready to answer a question, select the “Question” button to return to the question.

Overview -

Company Information -
Contoso, Inc. is a renewable energy provider that operates solar and wind farms across North America.

Existing Environment -

Azure Environment -
Contoso has a single Azure Databricks workspace named Workspace1 in the West US Azure region. Workspace1 is enabled for Unity Catalog. Workspace1 contains all-purpose clusters for both development and production workloads.
The company's Azure environment contains:
In the West US, Central US, and East US Azure regions, Azure event hubs that stream telemetry data and an Azure Data Lake Storage Gen2 account in each region for each hub
A single Azure SQL database in the West US region that hosts enterprise resource planning (ERP) data
An Azure Database for PostgreSQL server in the West US region that stores operational maintenance data

Data Environment -
Contoso ingests the following operational and business data:
Telemetry data: More than 40,000 IoT sensors across 28 sites emit JSON telemetry events every few seconds. Each site sends the events to the nearest event hub, which writes the data into the corresponding Data Lake Storage Gen2 account. These files frequently experience schema drift.
Maintenance logs: Maintenance systems generate historical repair logs, daily incremental updates, technician notes, and unstructured attachments that are stored in the Data Lake Storage Gen2 accounts.
Operational maintenance data: Structured operational maintenance data is stored on the Azure Database for PostgreSQL server.
External weather data: Hourly weather forecasts are retrieved from a REST API and written to the Data Lake Storage Gen2 accounts.
ERP data: Daily CSV extracts of 50 to 100 GB contain equipment metadata, work orders, and purchase order information.

Problem Statements -
The company’s existing analytics environment has several issues:

Ingestion -
Telemetry pipelines fall behind during peak loads.
Telemetry ingestion fails when schema drift occurs.
Streaming pipelines reprocess events after a pipeline restarts.

Compute -
Production and development workloads run on the same all-purpose clusters.
Production and development workloads do NOT support autoscaling or workload isolation.

Governance -
The ERP data is duplicated across systems and development teams.

Naming conventions are inconsistent across development teams, regions, and products.
Ownership of the IoT sensors changes over time, and analysts must track the full history of the ownership.
Occasionally, equipment manufacturers must correct data-entry mistakes in equipment names. Historical values are NOT required.

Pipeline operations -
Pipelines lack resiliency, alerting, and centralized scheduling.

Requirements -

Planned Changes -
Contoso plans to implement the following changes:
Implement scalable data pipeline orchestration.
Create a managed analytics catalog in Unity Catalog.
Implement a consistent approach to creating curated datasets.
Establish a centralized governance model across ingestion, cleansed, and curated layers.
Grant data engineers access to the ERP tables by using minimal development effort.
Adopt a compute strategy that isolates production workloads and supports autoscaling.
Adopt a slowly changing dimension (SCD) approach to address current data modeling issues.

Technical Requirements -
Contoso identifies the following environment and compute requirements:
Ensure that production ingestion workloads run on compute clusters that can scale automatically during telemetry spikes.
Provide fast and consistent performance for business intelligence (BI) workloads.
Prevent development activity from affecting production pipelines.
Production ingestion workloads must run as scheduled, non-interactive pipelines rather than on shared interactive development clusters.
Contoso identifies the following data ingestion and processing requirements:
Auto-scale ingestion pipelines to handle bursty workloads.
Handle schema drift for the maintenance and telemetry data.
Ingest file-based telemetry data by using minimal operational effort.
Store all the ingested data in a format that supports incremental processing.
Support the continuous ingestion of telemetry data from the event hubs by using exactly-once semantics.
Support the ingestion of the structured maintenance data from the Azure Database for PostgreSQL server.
Build a new telemetry pipeline that ingests raw events from the event hubs, cleanses the data, and publishes curated tables to Unity Catalog.
Ensure that the Apache Spark Structured Streaming pipelines reading from the event hubs write the data into a managed Delta table named telemetry.raw_events. The pipelines must support schema drift and resume processing after failures without reprocessing the data.
Contoso identifies the following data modeling and optimization requirements:
Build curated tables that standardize business logic.
Overwrite equipment metadata attributes, such as name, manufacturer, model, and commissioning date, when the attributes change. Historical values are NOT required.
Contoso identifies the following pipeline deployment and operation requirements:
Orchestrate multi-step ingestion and transformation workflows.
Define a clear execution order and dependencies.
Automatically retry failed steps and notify operators.
Schedule ingestion and transformation workloads consistently.

Governance Requirements -
Contoso identifies the following governance requirements:
Centralize the metadata catalog.
Provide isolated development areas that follow standard naming conventions.
Establish a consistent structure for organizing raw, cleansed, and curated data.
Provide a read-only mechanism to reference the ERP data through a foreign catalog.

Business Requirements -
Contoso identifies the following business requirements:
Improve ingestion reliability and reduce operational effort.
Standardize data definitions across development teams.
You need to organize Unity Catalog. The solution must meet the governance requirements.
What should you do?

A. Use a single shared schema for all the development teams and rely on table-level permissions for isolation.

B. Enable the development teams to create objects directly in the default catalog and schema.

C. Create a separate catalog for each development team and enable each team to choose its own schema names.

D. Create a shared development catalog, enforce a standardized naming convention, and assign each development team its own schema.

Suggested Answer: *D*

Currently there are no comments in this discussion, be the first to comment!

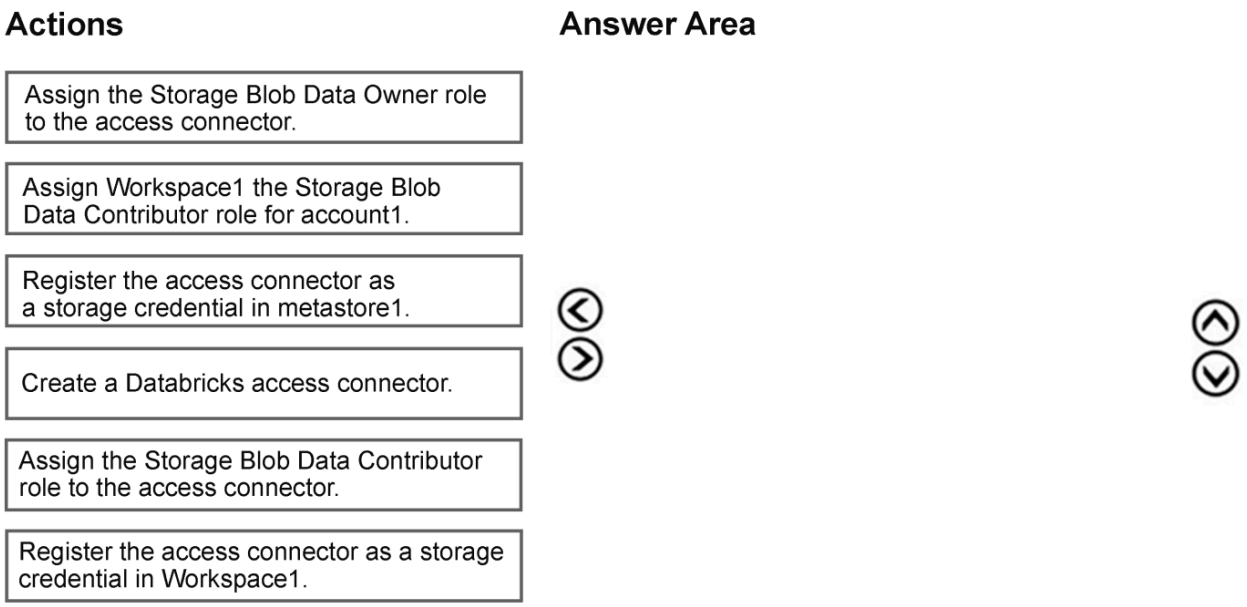
Question #2 *Topic 2*

DRAG DROP -

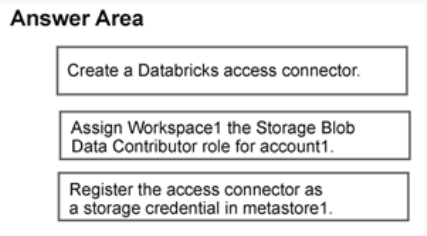
You have an Azure Databricks workspace named Workspace1 that is attached to a Unity Catalog metastore named metastore1.

You need to register an Azure Storage account named account1 that has a hierarchical namespace enabled as an external location. The external location must use a managed identity to authenticate to account1 and the solution must follow the principle of least privilege.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #3 *Topic 2*

HOTSPOT -

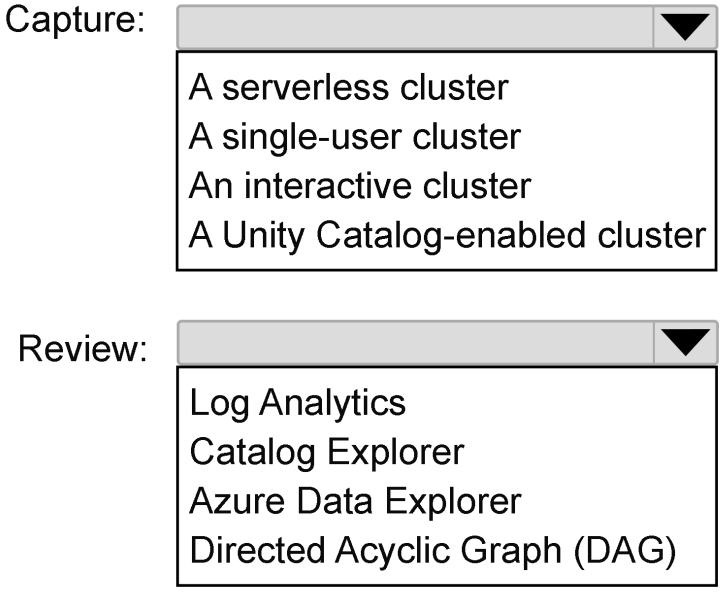
You have an Azure Databricks workspace that is enabled for Unity Catalog.

You need to ensure that data lineage is captured and can be reviewed for tables accessed by Databricks notebooks and jobs. The solution must minimize administrative effort.

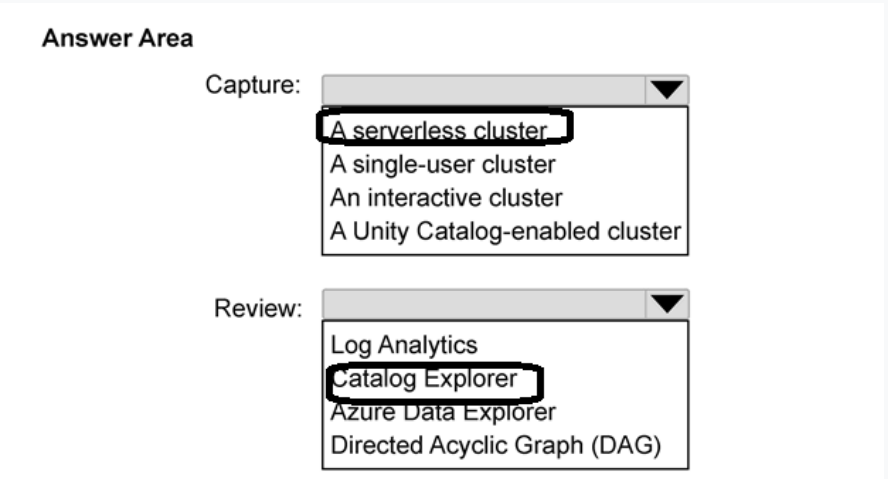
Which compute configuration should you use to capture the data lineage and what should you use to review the data lineage? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #4 *Topic 2*

HOTSPOT -

You have an Azure Databricks workspace.

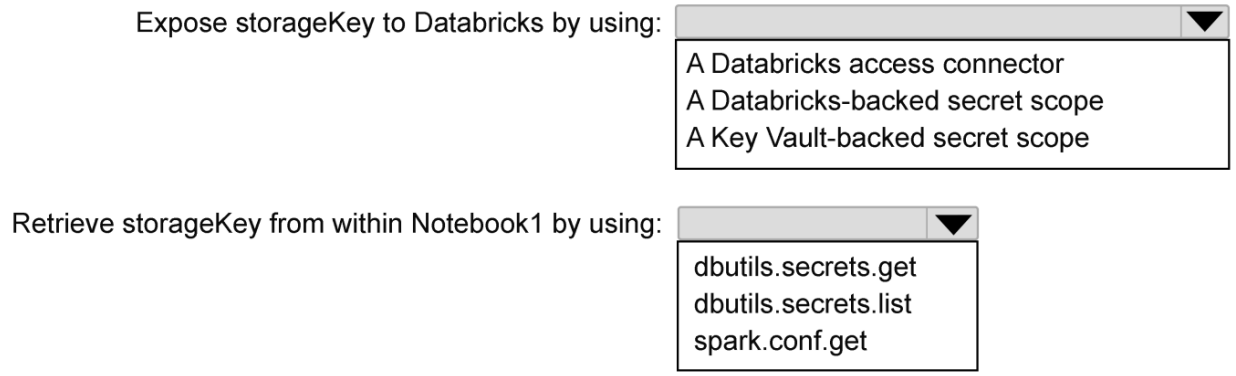
You have an Azure key vault named kv-secure that stores a secret named storageKey. The value of storageKey is managed and updated by the cloud security team at your company.

You need to enable a Databricks notebook named Notebook1 to retrieve the value of storageKey securely at runtime. The solution must follow the principle of least privilege and always retrieve the latest value.

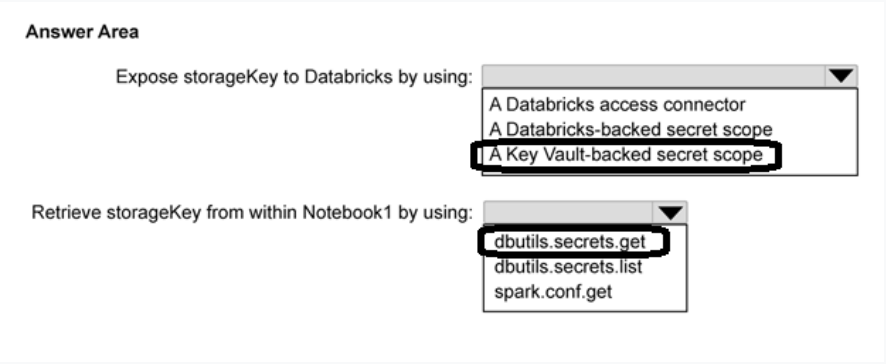
What should you do? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #5 *Topic 2*

HOTSPOT -

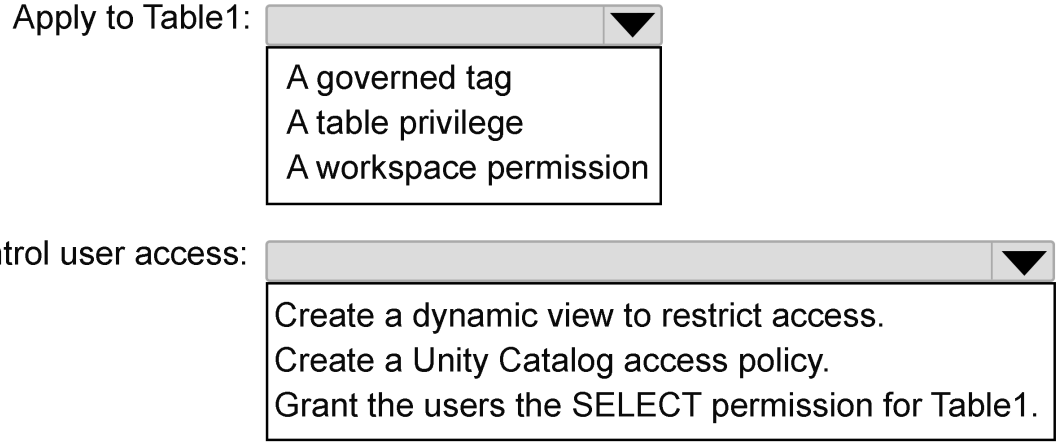
You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a catalog named Catalog1. Catalog1 contains a schema named Schema1 and a table named Table1.

You need to ensure that access to the data in Table1 is controlled by using attribute-based access control (ABAC).

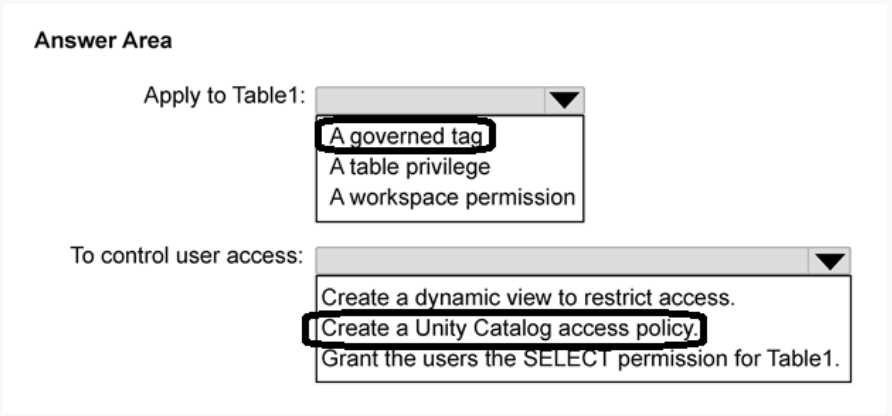
What should you apply to Table1, and how should you control access for users? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #6 *Topic 2*

HOTSPOT -

You have an Azure Databricks workspace that is enabled for Unity Catalog.

You need to implement a data lifecycle and expiration solution that meets the following requirements:

Transaction logs and deleted data files that are older than 90 days must be removed from Delta tables to reclaim storage.

All the tables must remain available for querying during the cleanup process.

Administrative effort must be minimized.

What should you do for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Data lifecycle control:

- Set timeUntilArchived to 90.
- Set logRetentionDuration to 90.
- Set deletedFileRetentionDuration to 90.
- Tag tables that contain retention metadata.

Expiration cleanup:

- Run the VACUUM command.
- Run the OPTIMIZE command.
- Drop and recreate the tables.
- Use scheduled DELETE operations in jobs.

Suggested Answer:

Answer Area

Data lifecycle control:

- Set timeUntilArchived to 90.
- Set logRetentionDuration to 90.
- **Set deletedFileRetentionDuration to 90.**
- Tag tables that contain retention metadata.

Expiration cleanup:

- **Run the VACUUM command.**
- Run the OPTIMIZE command.
- Drop and recreate the tables.
- Use scheduled DELETE operations in jobs.

Currently there are no comments in this discussion, be the first to comment!

Question #7 *Topic 2*

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a managed Delta table named Table1. Table1 stores customer data.
You need to implement a data retention solution that meets the following requirements:
Deleted data must be retained for 30 days to support audits.
Deleted data that is older than 30 days must be removed permanently.
The solution must minimize administrative effort
Which two properties should you configure? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. delta.timeUntilArchived

B. delta.enableDeletionVectors

C. delta.logRetentionDuration

D. delta.deletedFileRetentionDuration

E. delta.autoOptimize.autoCompact

Suggested Answer: *BC*

Currently there are no comments in this discussion, be the first to comment!

Question #8 *Topic 2*

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains two catalogs named Catalog1 and Catalog2.
An external application uses a service principal named SP1 to connect to a SQL warehouse.
You need to ensure that SP1 can query the data in Catalog1 and Catalog2. The solution must follow the principle of least privilege.
Which permissions should you grant to SP1 for the catalogs?

A. USE SCHEMA and SELECT

B. USE CATALOG and SELECT

C. USE CATALOG, USE SCHEMA, and SELECT

D. USE CATALOG and USE SCHEMA

Suggested Answer: *C*

Currently there are no comments in this discussion, be the first to comment!

Question #9 *Topic 2*

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a catalog named Catalog1. Catalog1 contains a table named Transactions. Transactions contains the following columns: transaction_id customer_name email_address credit_card_number transaction_amount

You need to ensure that business analysts can query all the rows in the Transactions table. The solution must meet the following requirements:

Prevent the analysts from seeing the full values in the email_address and credit_card_number columns.

Ensure that the analysts can see only the values after the @ character in each email address.

Ensure that the analysts can see only the last four digits of each credit card number.

Enable the analysts to query the table without errors.

Follow the principle of least privilege.

What should you do?

A. Grant the analysts the SELECT permission for the Transactions table and apply column-level encryption.

B. Grant the analysts the SELECT permission for columns that do NOT contain sensitive data.

C. Grant the analysts the SELECT permission for the Transactions table and implement row-level filters.

D. Grant the analysts the SELECT permission for the Transactions table and apply column masks to email_address and credit_card_number.

Suggested Answer: *D*

Currently there are no comments in this discussion, be the first to comment!

Question #10 *Topic 2*

HOTSPOT -

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a catalog named catalog1.

You have a group named group1.

You plan to create a schema named schema1 in catalog1.

You need to ensure that group1 meets the following requirements:

Can create tables in schema1 -

Can modify and query tables -

Cannot grant permissions for the schema and its objects

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

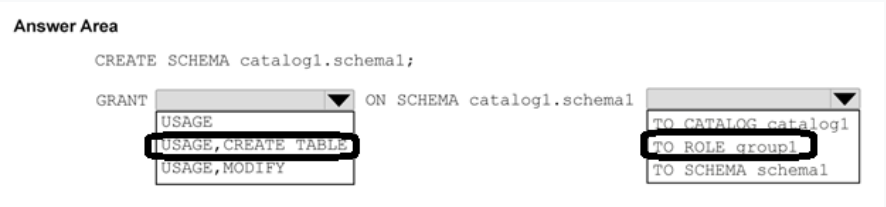
NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE SCHEMA catalog1.schema1;

GRANT [ USAGE | USAGE,CREATE TABLE | USAGE,MODIFY ] ON SCHEMA catalog1.schema1 [ TO CATALOG catalog1 | TO ROLE group1 | TO SCHEMA schema1 ]
```

Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #11 *Topic 2*

HOTSPOT -

You have an Azure Databricks workspace that is enabled for Unity Catalog.

You need to create an external volume named Volume1 in an existing schema. Volume1 must expose files from an Azure Storage container. The solution must meet the following requirements:

Ensure that authentication does NOT require storing credentials in Databricks.

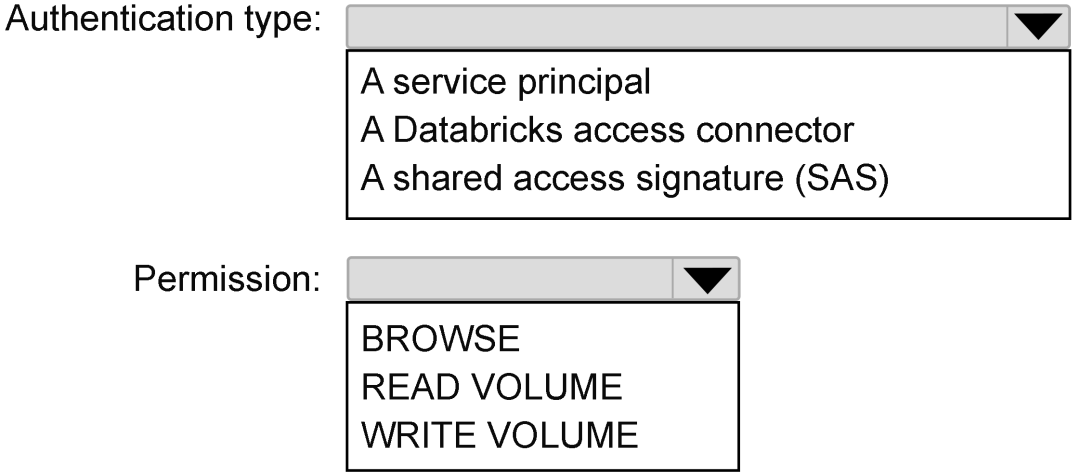
Ensure that users can access the files, but NOT modify the files.

Follow the principle of least privilege.

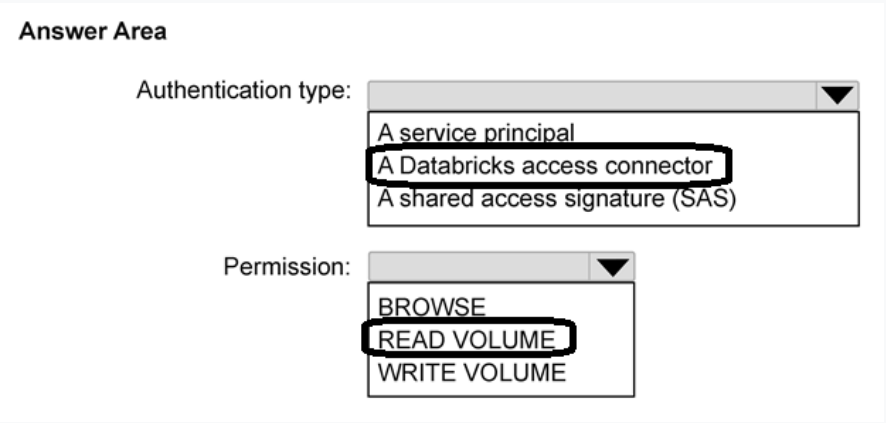
Which type of authentication should you configure, and which permission should you grant to the users? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #12 *Topic 2*

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.
You have an Azure Databricks workspace named Workspace1 that contains a lakehouse and is enabled for Unity Catalog.
You have a connection to a Microsoft SQL Server database named DB1.
You need to expose the schemas and tables of DB1 to meet the following requirements:
The schemas and tables can be queried in Databricks.
The schemas and tables appear alongside other Unity Catalog objects.
The data is NOT copied into Databricks-managed storage.
Solution: You create a Lakeflow Connect pipeline and connect it to DB1.
Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

Currently there are no comments in this discussion, be the first to comment!

Question #13 *Topic 2*

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.
You have an Azure Databricks workspace named Workspace1 that contains a lakehouse and is enabled for Unity Catalog.
You have a connection to a Microsoft SQL Server database named DB1.
You need to expose the schemas and tables of DB1 to meet the following requirements:
The schemas and tables can be queried in Databricks.
The schemas and tables appear alongside other Unity Catalog objects.
The data is NOT copied into Databricks-managed storage.
Solution: You create a Databricks access connector.
Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

Currently there are no comments in this discussion, be the first to comment!

Question #14 *Topic 2*

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.
You have an Azure Databricks workspace named Workspace1 that contains a lakehouse and is enabled for Unity Catalog.
You have a connection to a Microsoft SQL Server database named DB1.
You need to expose the schemas and tables of DB1 to meet the following requirements:
The schemas and tables can be queried in Databricks.
The schemas and tables appear alongside other Unity Catalog objects.
The data is NOT copied into Databricks-managed storage.
Solution: You create a foreign catalog in Catalog Explorer.
Does this meet the goal?

A. Yes

B. No

Suggested Answer: *A*

Currently there are no comments in this discussion, be the first to comment!

Question #15 *Topic 2*

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.
You have an Azure Databricks workspace named Workspace1 that contains a lakehouse and is enabled for Unity Catalog.
You have a connection to a Microsoft SQL Server database named DB1.
You need to expose the schemas and tables of DB1 to meet the following requirements:
The schemas and tables can be queried in Databricks.
The schemas and tables appear alongside other Unity Catalog objects.
The data is NOT copied into Databricks-managed storage.
Solution: You create a new native catalog in Unity Catalog.
Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

Currently there are no comments in this discussion, be the first to comment!

Question #16 *Topic 2*

Your company has sales, finance, and HR departments.
You have an Azure Databricks workspace that is enabled for Unity Catalog.
You need to implement Unity Catalog to meet the following requirements:
Access to Unity Catalog for each department must be managed independently from that of the other departments.
The data assets of each department must be isolated logically from those of the other departments.
The solution must support centralized governance.
What should you do for each department?

A. Create a separate Azure Data Lake Storage Gen2 account.

B. Create dynamic views.

C. Create a new workspace.

D. Create a new catalog.

Suggested Answer: *D*

Currently there are no comments in this discussion, be the first to comment!

Question #17 *Topic 2*

DRAG DROP -

You have an Azure Databricks workspace named Workspace1.

You have a user named User1 that is a non-admin user for Workspace1.

You need to ensure that User1 can perform the following tasks:

Provision clusters of any size.

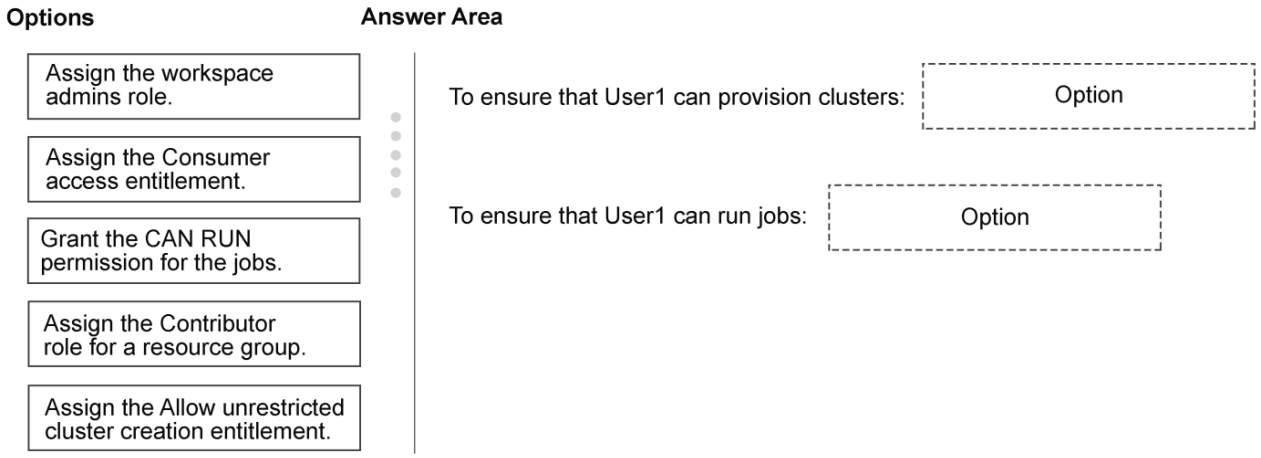
Run Databricks jobs.

The solution must follow the principle of least privilege.

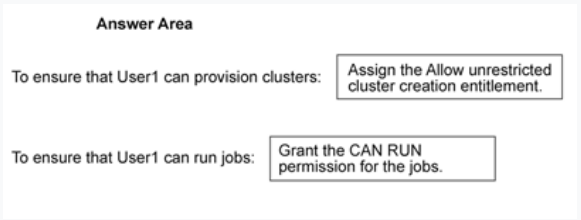
What should you do?

To answer, drag the appropriate options to the correct requirements. 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.



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #18 *Topic 2*

You have an Azure Databricks workspace that is enabled for Unity Catalog.
You have a CSV file stored in an Azure Data Lake Storage Gen2 container.
You plan to ingest the data into an existing table by running the following SQL statement.

COPY INTO Customer -
FROM 'abfss://container1@storageaccount.dfs.core.windows.net/data/customer'

FILEFORMAT = CSV -
You need to ensure that the statement can access the data in the container.
What should you configure?

A. an external table named Customer in the CSV file format

B. a Lakeflow ingestion gateway

C. an external location that uses a storage credential

D. a volume in a catalog

Suggested Answer: *C*

Currently there are no comments in this discussion, be the first to comment!

Question #19 *Topic 2*

HOTSPOT -

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a catalog named catalog1.

You have a group named group1. Group1 already has the USE CATALOG privilege on catalog1.

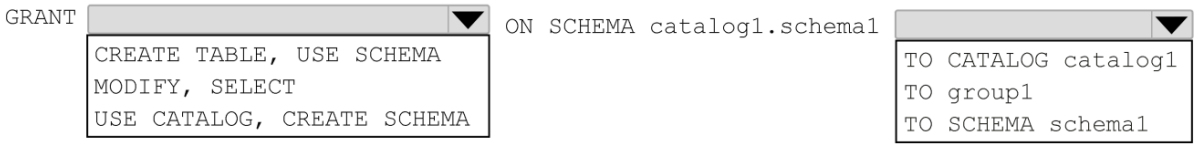
You create a schema named schema1 in catalog1.

You need to ensure that group1 can create tables in schema1. Group1 must not be able to grant permissions on the schema or its objects. The solution must follow the principle of least privilege.

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

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:

Currently there are no comments in this discussion, be the first to comment!

HOTSPOT -

You have an Azure Databricks workspace that is attached to a Unity Catalog metastore named metastore1.metastore1 contains:

A catalog named Sales -

A schema named Customers in the Sales catalog

A table named Customer_details in the Customers schema

You need to ensure that a user named User1 can update the data in Customer_details. The solution must meet the following requirements:

Ensure that User1 cannot create new tables.

Follow the principle of least privilege.

Which permission should you grant to User1 for each object? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

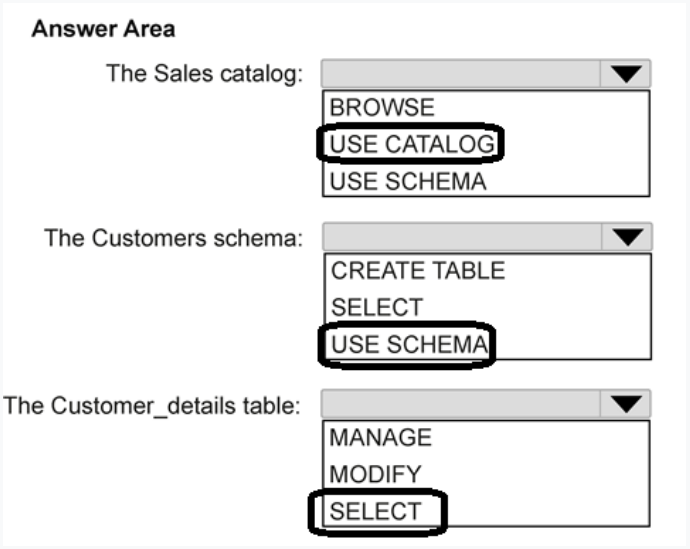
Answer Area

The Sales catalog: [▼]
- BROWSE
- USE CATALOG
- USE SCHEMA

The Customers schema: [▼]
- CREATE TABLE
- SELECT
- USE SCHEMA

The Customer_details table: [▼]
- MANAGE
- MODIFY
- SELECT

Suggested Answer:



fd21624 1 week, 6 days ago

the sales catalog - use schema

the customers schema - use schema

the customer_details table - modify

MANAGE Ability to grant/revoke privileges on the object Violates least privilege — User1 would gain control over others' access

MODIFY INSERT, UPDATE, DELETE on the table Correct — update data without any schema/table creation rights

SELECT Read-only query access Insufficient — cannot update data

upvoted 1 times

Question #1 *Topic 3*

HOTSPOT -

You have an Azure Databricks account that contains a single workspace named Workspace1. Workspace1 is enabled for Unity Catalog.

You discover that data access events for Unity Catalog tables fail to appear in the logs.

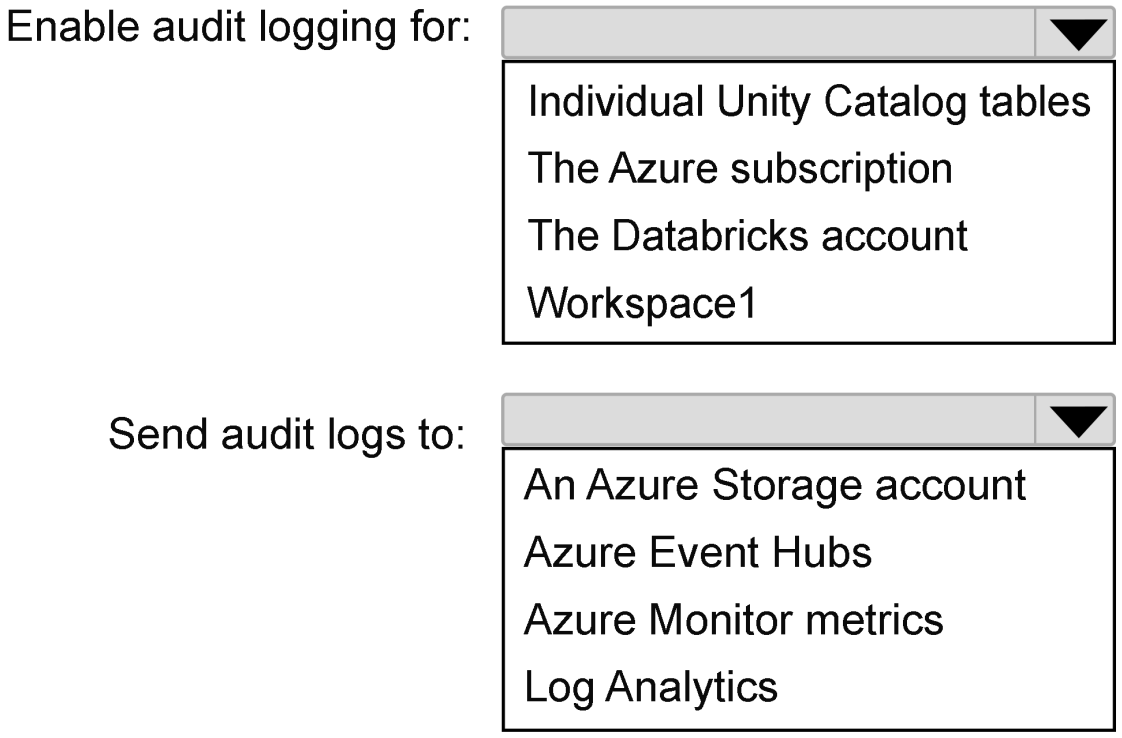
You need to ensure that all the data access events are captured centrally for auditing purposes. The log data must be available for analysis as quickly as possible.

What should you do?

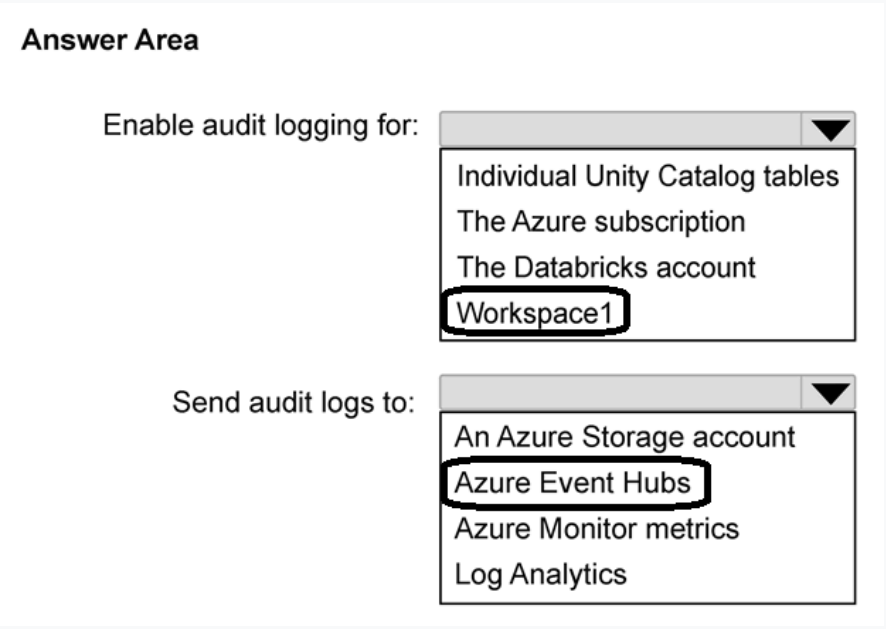
To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

DRAG DROP -
This is a case study. Case studies are not timed separately from other exam sections. You can use as much exam time as you would like to complete each case study. However, there might be additional case studies or other exam sections. Manage your time to ensure that you can complete all the exam sections in the time provided. Pay attention to the Exam Progress at the top of the screen so you have sufficient time to complete any exam sections that follow this case study.
To answer the case study questions, you will need to reference information that is provided in the case. Case studies and associated questions might contain exhibits or other resources that provide more information about the scenario described in the case. Information provided in an individual question does not apply to the other questions in the case study.
A Review Screen will appear at the end of this case study. From the Review Screen, you can review and change your answers before you move to the next exam section. After you leave this case study, you will NOT be able to return to it.

To start the case study -
To display the first question in this case study, select the “Next” button. To the left of the question, a menu provides links to information such as business requirements, the existing environment, and problem statements. Please read through all this information before answering any questions. When you are ready to answer a question, select the “Question” button to return to the question.

Overview -

Company Information -
Contoso, Inc. is a renewable energy provider that operates solar and wind farms across North America.

Existing Environment -

Azure Environment -
Contoso has a single Azure Databricks workspace named Workspace1 in the West US Azure region. Workspace1 is enabled for Unity Catalog. Workspace1 contains all-purpose clusters for both development and production workloads.
The company's Azure environment contains:
In the West US, Central US, and East US Azure regions, Azure event hubs that stream telemetry data and an Azure Data Lake Storage Gen2 account in each region for each hub
A single Azure SQL database in the West US region that hosts enterprise resource planning (ERP) data
An Azure Database for PostgreSQL server in the West US region that stores operational maintenance data

Data Environment -
Contoso ingests the following operational and business data:
Telemetry data: More than 40,000 IoT sensors across 28 sites emit JSON telemetry events every few seconds. Each site sends the events to the nearest event hub, which writes the data into the corresponding Data Lake Storage Gen2 account. These files frequently experience schema drift.
Maintenance logs: Maintenance systems generate historical repair logs, daily incremental updates, technician notes, and unstructured attachments that are stored in the Data Lake Storage Gen2 accounts.
Operational maintenance data: Structured operational maintenance data is stored on the Azure Database for PostgreSQL server.
External weather data: Hourly weather forecasts are retrieved from a REST API and written to the Data Lake Storage Gen2 accounts.
ERP data: Daily CSV extracts of 50 to 100 GB contain equipment metadata, work orders, and purchase order information.

Problem Statements -
The company’s existing analytics environment has several issues:

Ingestion -
Telemetry pipelines fall behind during peak loads.
Telemetry ingestion fails when schema drift occurs.
Streaming pipelines reprocess events after a pipeline restarts.

Compute -
Production and development workloads run on the same all-purpose clusters.
Production and development workloads do NOT support autoscaling or workload isolation.

Governance -

The ERP data is duplicated across systems and development teams.
Naming conventions are inconsistent across development teams, regions, and products.
Ownership of the IoT sensors changes over time, and analysts must track the full history of the ownership.
Occasionally, equipment manufacturers must correct data-entry mistakes in equipment names. Historical values are NOT required.

Pipeline operations -
Pipelines lack resiliency, alerting, and centralized scheduling.

Requirements -

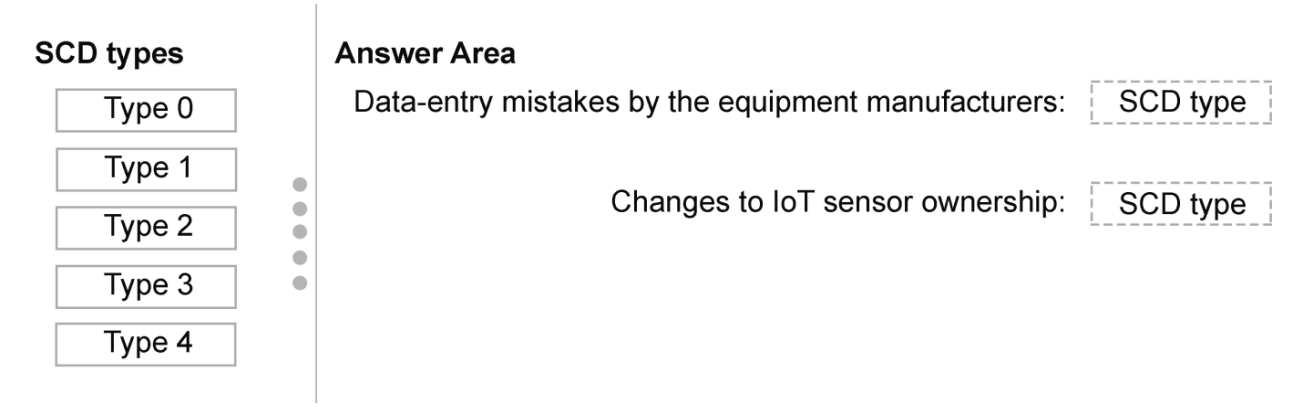
Planned Changes -
Contoso plans to implement the following changes:
Implement scalable data pipeline orchestration.
Create a managed analytics catalog in Unity Catalog.
Implement a consistent approach to creating curated datasets.
Establish a centralized governance model across ingestion, cleansed, and curated layers.
Grant data engineers access to the ERP tables by using minimal development effort.
Adopt a compute strategy that isolates production workloads and supports autoscaling.
Adopt a slowly changing dimension (SCD) approach to address current data modeling issues.

Technical Requirements -
Contoso identifies the following environment and compute requirements:
Ensure that production ingestion workloads run on compute clusters that can scale automatically during telemetry spikes.
Provide fast and consistent performance for business intelligence (BI) workloads.
Prevent development activity from affecting production pipelines.
Production ingestion workloads must run as scheduled, non-interactive pipelines rather than on shared interactive development clusters.
Contoso identifies the following data ingestion and processing requirements:
Auto-scale ingestion pipelines to handle bursty workloads.
Handle schema drift for the maintenance and telemetry data.
Ingest file-based telemetry data by using minimal operational effort.
Store all the ingested data in a format that supports incremental processing.
Support the continuous ingestion of telemetry data from the event hubs by using exactly-once semantics.
Support the ingestion of the structured maintenance data from the Azure Database for PostgreSQL server.
Build a new telemetry pipeline that ingests raw events from the event hubs, cleanses the data, and publishes curated tables to Unity Catalog.
Ensure that the Apache Spark Structured Streaming pipelines reading from the event hubs write the data into a managed Delta table named telemetry.raw_events. The pipelines must support schema drift and resume processing after failures without reprocessing the data.
Contoso identifies the following data modeling and optimization requirements:
Build curated tables that standardize business logic.
Overwrite equipment metadata attributes, such as name, manufacturer, model, and commissioning date, when the attributes change. Historical values are NOT required.
Contoso identifies the following pipeline deployment and operation requirements:
Orchestrate multi-step ingestion and transformation workflows.
Define a clear execution order and dependencies.
Automatically retry failed steps and notify operators.
Schedule ingestion and transformation workloads consistently.

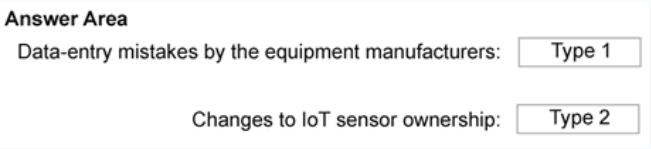
Governance Requirements -
Contoso identifies the following governance requirements:
Centralize the metadata catalog.
Provide isolated development areas that follow standard naming conventions.
Establish a consistent structure for organizing raw, cleansed, and curated data.
Provide a read-only mechanism to reference the ERP data through a foreign catalog.

Business Requirements -
Contoso identifies the following business requirements:
Improve ingestion reliability and reduce operational effort.
Standardize data definitions across development teams.
Which SCD type should you use to support the planned data modeling changes? To answer, drag the appropriate types to the correct issues. Each

type 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.



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

DRAG DROP -
This is a case study. Case studies are not timed separately from other exam sections. You can use as much exam time as you would like to complete each case study. However, there might be additional case studies or other exam sections. Manage your time to ensure that you can complete all the exam sections in the time provided. Pay attention to the Exam Progress at the top of the screen so you have sufficient time to complete any exam sections that follow this case study.
To answer the case study questions, you will need to reference information that is provided in the case. Case studies and associated questions might contain exhibits or other resources that provide more information about the scenario described in the case. Information provided in an individual question does not apply to the other questions in the case study.
A Review Screen will appear at the end of this case study. From the Review Screen, you can review and change your answers before you move to the next exam section. After you leave this case study, you will NOT be able to return to it.

To start the case study -
To display the first question in this case study, select the “Next” button. To the left of the question, a menu provides links to information such as business requirements, the existing environment, and problem statements. Please read through all this information before answering any questions. When you are ready to answer a question, select the “Question” button to return to the question.

Overview -

Company Information -
Contoso, Inc. is a renewable energy provider that operates solar and wind farms across North America.

Existing Environment -

Azure Environment -
Contoso has a single Azure Databricks workspace named Workspace1 in the West US Azure region. Workspace1 is enabled for Unity Catalog. Workspace1 contains all-purpose clusters for both development and production workloads.
The company's Azure environment contains:
In the West US, Central US, and East US Azure regions, Azure event hubs that stream telemetry data and an Azure Data Lake Storage Gen2 account in each region for each hub
A single Azure SQL database in the West US region that hosts enterprise resource planning (ERP) data
An Azure Database for PostgreSQL server in the West US region that stores operational maintenance data

Data Environment -
Contoso ingests the following operational and business data:
Telemetry data: More than 40,000 IoT sensors across 28 sites emit JSON telemetry events every few seconds. Each site sends the events to the nearest event hub, which writes the data into the corresponding Data Lake Storage Gen2 account. These files frequently experience schema drift.
Maintenance logs: Maintenance systems generate historical repair logs, daily incremental updates, technician notes, and unstructured attachments that are stored in the Data Lake Storage Gen2 accounts.
Operational maintenance data: Structured operational maintenance data is stored on the Azure Database for PostgreSQL server.
External weather data: Hourly weather forecasts are retrieved from a REST API and written to the Data Lake Storage Gen2 accounts.
ERP data: Daily CSV extracts of 50 to 100 GB contain equipment metadata, work orders, and purchase order information.

Problem Statements -
The company’s existing analytics environment has several issues:

Ingestion -
Telemetry pipelines fall behind during peak loads.
Telemetry ingestion fails when schema drift occurs.
Streaming pipelines reprocess events after a pipeline restarts.

Compute -
Production and development workloads run on the same all-purpose clusters.
Production and development workloads do NOT support autoscaling or workload isolation.

Governance -

The ERP data is duplicated across systems and development teams.
Naming conventions are inconsistent across development teams, regions, and products.
Ownership of the IoT sensors changes over time, and analysts must track the full history of the ownership.
Occasionally, equipment manufacturers must correct data-entry mistakes in equipment names. Historical values are NOT required.

Pipeline operations -
Pipelines lack resiliency, alerting, and centralized scheduling.

Requirements -

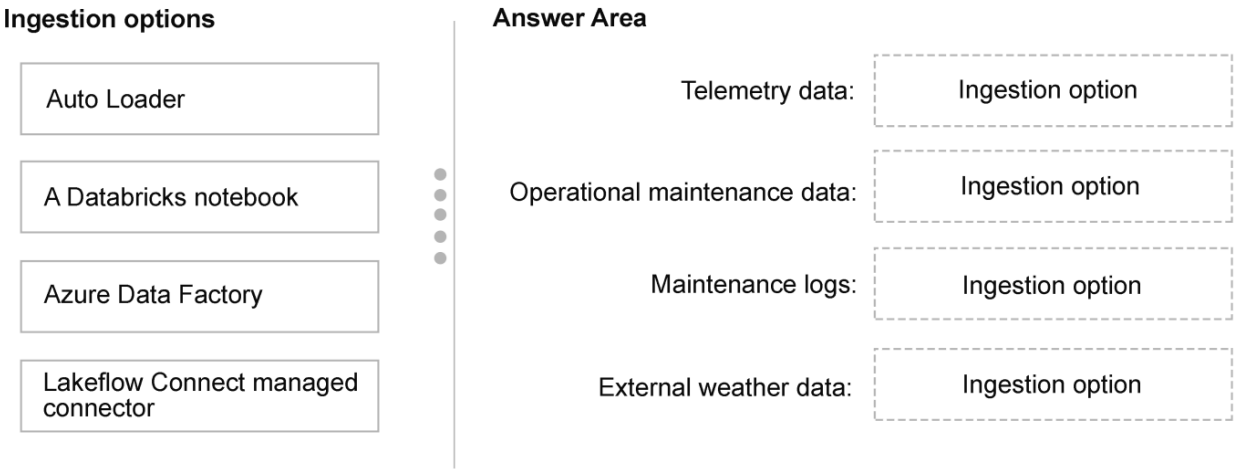
Planned Changes -
Contoso plans to implement the following changes:
Implement scalable data pipeline orchestration.
Create a managed analytics catalog in Unity Catalog.
Implement a consistent approach to creating curated datasets.
Establish a centralized governance model across ingestion, cleansed, and curated layers.
Grant data engineers access to the ERP tables by using minimal development effort.
Adopt a compute strategy that isolates production workloads and supports autoscaling.
Adopt a slowly changing dimension (SCD) approach to address current data modeling issues.

Technical Requirements -
Contoso identifies the following environment and compute requirements:
Ensure that production ingestion workloads run on compute clusters that can scale automatically during telemetry spikes.
Provide fast and consistent performance for business intelligence (BI) workloads.
Prevent development activity from affecting production pipelines.
Production ingestion workloads must run as scheduled, non-interactive pipelines rather than on shared interactive development clusters.
Contoso identifies the following data ingestion and processing requirements:
Auto-scale ingestion pipelines to handle bursty workloads.
Handle schema drift for the maintenance and telemetry data.
Ingest file-based telemetry data by using minimal operational effort.
Store all the ingested data in a format that supports incremental processing.
Support the continuous ingestion of telemetry data from the event hubs by using exactly-once semantics.
Support the ingestion of the structured maintenance data from the Azure Database for PostgreSQL server.
Build a new telemetry pipeline that ingests raw events from the event hubs, cleanses the data, and publishes curated tables to Unity Catalog.
Ensure that the Apache Spark Structured Streaming pipelines reading from the event hubs write the data into a managed Delta table named telemetry.raw_events. The pipelines must support schema drift and resume processing after failures without reprocessing the data.
Contoso identifies the following data modeling and optimization requirements:
Build curated tables that standardize business logic.
Overwrite equipment metadata attributes, such as name, manufacturer, model, and commissioning date, when the attributes change. Historical values are NOT required.
Contoso identifies the following pipeline deployment and operation requirements:
Orchestrate multi-step ingestion and transformation workflows.
Define a clear execution order and dependencies.
Automatically retry failed steps and notify operators.
Schedule ingestion and transformation workloads consistently.

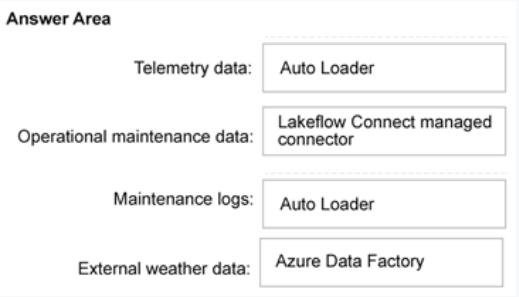
Governance Requirements -
Contoso identifies the following governance requirements:
Centralize the metadata catalog.
Provide isolated development areas that follow standard naming conventions.
Establish a consistent structure for organizing raw, cleansed, and curated data.
Provide a read-only mechanism to reference the ERP data through a foreign catalog.

Business Requirements -
Contoso identifies the following business requirements:
Improve ingestion reliability and reduce operational effort.
Standardize data definitions across development teams.
Which ingestion option should you recommend for each data source? To answer, drag the appropriate options to the correct data sources. 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.



Suggested Answer:



Currently there are no comments in this discussion, be the first to comment!

Question #4 *Topic 3*

You have an Azure Databricks workspace that uses serverless compute.
You need to ingest data by using Lakeflow Jobs. New records must be processed as soon as they become available.
Which type of job trigger should you use for the ingestion?

A. manual

B. file arrival

C. scheduled

D. continuous

Suggested Answer: *D*

Currently there are no comments in this discussion, be the first to comment!

Question #5 *Topic 3*

You have an Azure Databricks workspace that is enabled for Unity Catalog and contains a managed Delta table named Sales.
Sales stores transaction data and contains the following columns: transaction_id (string) transaction_date (date) amount (decimal)
You need to implement the following data quality requirements by using table-level data quality enforcement: amount must be greater than 0. transaction_id must never be null.
Invalid records must be rejected when data is written to the Sales table.
What should you do?

A. Use a SELECT statement with WHERE conditions to validate the data before querying.

B. Create a view that filters out rows where transaction_id is null or amount is less than or equal to 0.

C. Add a NOT NULL constraint to transaction_id and a CHECK constraint to amount.

D. Configure row-level security (RLS) where transaction_id is null or amount is less than or equal to 0.

Suggested Answer: *C*

Currently there are no comments in this discussion, be the first to comment!