You are a data engineer implementing a lambda architecture on Microsoft Azure. You use an open-source big data solution to collect, process, and maintain data.

The analytical data store performs poorly.

You must implement a solution that meets the following requirements:

☞ Provide data warehousing

☞ Reduce ongoing management activities

☞ Deliver SQL query responses in less than one second

You need to create an HDInsight cluster to meet the requirements.

Which type of cluster should you create?

    A. Interactive Query

    B. Apache Hadoop

    C. Apache HBase

    D. Apache Spark

Suggested Answer: *D*

Lambda Architecture with Azure:

Azure offers you a combination of following technologies to accelerate real-time big data analytics:

1. Azure Cosmos DB, a globally distributed and multi-model database service.

2. Apache Spark for Azure HDInsight, a processing framework that runs large-scale data analytics applications.

3. Azure Cosmos DB change feed, which streams new data to the batch layer for HDInsight to process.

4. The Spark to Azure Cosmos DB Connector
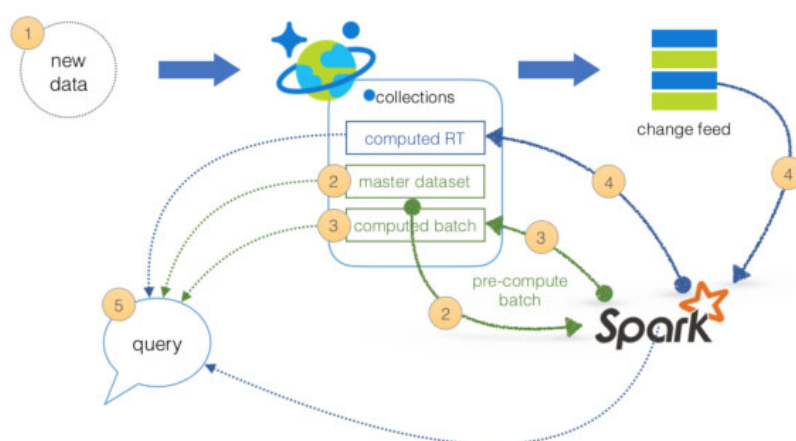
Note: Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch processing and stream processing methods, and minimizing the latency involved in querying big data.

References:

https://sqlwithmanoj.com/2018/02/16/what-is-lambda-architecture-and-what-azure-offers-with-its-new-cosmos-db/

☐ 👤 **dangal95** `Highly Voted 👍` 4 years, 2 months ago

Could the write answer be D becase Spark has;

1) Interactive queries through spark-sql

2) Datawarehousing capabilities through Delta Lake (and also spark-sql creates in memory tables)

3) Less management because these are out-of-the-box features?

upvoted 9 times

   ☐ 👤 **Amy007** 4 years, 1 month ago

   It also mentions SQL queries which is not Hive (Interactive Query)

   upvoted 2 times

☐ 👤 **jedi01** `Highly Voted 👍` 4 years, 10 months ago

I think the answer should be A. Interactive Query.

Here I am implementing Lambda architecture using a open source technology which can be Apache Spark and already in use. The prevailing issue

here Analytical Processing is very slow , in another words queries are slow. So I created an HDInsight Cluster of type "Interactive Query" to support Analytical processing/ fast query access, data warehousing etc. We can use HiveQL on Interactive Query. Refer to https://docs.microsoft.com/en-us/azure/hdinsight/interactive-query/apache-interactive-query-get-started

upvoted 8 times

agmadeira Most Recent ⊘ 4 years ago

A - Interactive Query - "Deliver SQL query responses in less than one second"

https://docs.microsoft.com/en-us/azure/hdinsight/interactive-query/apache-interactive-query-get-started

upvoted 2 times

Hinzzz 4 years ago

D is correct based on Data warehousing requirement.

upvoted 1 times

sandeep1111 4 years, 2 months ago

correct

upvoted 1 times

Satya217 4 years, 4 months ago

https://docs.microsoft.com/en-us/azure/cosmos-db/lambda-architectur

upvoted 1 times

Trivender 4 years, 5 months ago

Correct Answer in Spark because it is in memory

upvoted 1 times

dumpsm42 4 years, 6 months ago

hi to all,

answer:

https://azure.microsoft.com/pt-pt/blog/general-availability-of-hdinsight-interactive-query-blazing-fast-data-warehouse-style-queries-on-hyper-scale-data-2/

sub-second !

Summary

This week at Ignite, we are pleased to announce general availability of Azure HDInsight Interactive Query. Backed by our enterprise-grade SLA, HDInsight Interactive Query brings sub-second speed to data warehouse style SQL queries to the hyper-scale data stored in commodity cloud storage.

regards

upvoted 5 times

J_i_L_L 4 years, 7 months ago

Exam was updated on Nov 24, 2020. Didn't see too many questions from the test on ExamTopics...maybe 20-30% of the test questions. Suggest waiting a bit to take the test so that all the exam prep questions are updated. Exam definitely requires hands-on knowledge of the products. A lot of questions on CosmosDB consistency settings, encryption/security, monitoring/metrics.

upvoted 6 times

wahwah 4 years, 7 months ago

Do you passed the exam after 24 Nov, is there any difference regarding these questions and the updated ones ? I mean if I prepared the exam with these version of a questions what is my chances to pass it ? thank you

upvoted 4 times

uomer 4 years, 7 months ago

I also vote for Interactive Query as "An Interactive Query cluster is different from an Apache Hadoop cluster. It contains only the Hive service. Requirements:
☞ Provide data warehousing (Yes)
☞ Reduce ongoing management activities (Not sure)
☞ Deliver SQL query responses in less than one second (Yes)

upvoted 3 times

sunil08 4 years, 7 months ago

D: Apache spark

upvoted 1 times

👤 **nehab0101** 4 years, 8 months ago

https://azure.microsoft.com/en-in/blog/lambda-architecture-using-azure-cosmosdb-faster-performance-low-tco-low-devops/

upvoted 1 times

👤 **Mittun** 4 years, 8 months ago

Apache Spark is correct Answer !!

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview

upvoted 4 times

👤 **r8d1** 4 years, 10 months ago

i think the logic to answer this question is:

Lambda architecture:

https://databricks.com/glossary/lambda-architecture

Azure implementation:

https://azure.microsoft.com/en-us/services/databricks/

Azure Databricks = Fast, easy, and collaborative Apache SparkTM based analytics service

upvoted 1 times

👤 **dfrp92** 5 years ago

How does Spark meet the requirements? Spark does not provide data warehousing by itself, it is not a data store.

upvoted 4 times

  👤 **induna** 5 years ago

  Neither does any of the options, the last part of the question is key: Which type of cluster will you create, hence, Spark

  upvoted 9 times

👤 **AAJ** 5 years, 4 months ago

https://docs.microsoft.com/en-us/azure/cosmos-db/lambda-architecture

upvoted 2 times

  👤 **john_smith** 4 years, 11 months ago

  The link you provided is redirected to What is Azure Synapse Link for Azure Cosmos DB (Preview)?

  upvoted 1 times

👤 **Leonido** 5 years, 4 months ago

Would suggest to use the original link from MS: https://docs.microsoft.com/en-us/azure/cosmos-db/lambda-architecture as better background documentation

upvoted 2 times

  👤 **john_smith** 4 years, 11 months ago

  The link you provided is redirected to What is Azure Synapse Link for Azure Cosmos DB (Preview)? wy?

  upvoted 1 times

DRAG DROP -

You develop data engineering solutions for a company. You must migrate data from Microsoft Azure Blob storage to an Azure SQL Data Warehouse for further transformation. You need to implement the solution.

Which four 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.

Select and Place:

Actions

- Provision an Azure SQL Data Warehouse instance.
- Connect to the Blob storage container by using SQL Server Management Studio.
- Provision an Azure Blob storage container.
- Run Transact-SQL statements to load data.
- Connect to the Azure SQL Data Warehouse by using SQL Server Management Studio.
- Build external tables by using Azure portal.
- Build external tables by using SQL Server Management Studio.

Answer Area

(empty)

Suggested Answer:

Actions

- Provision an Azure SQL Data Warehouse instance.
- Connect to the Blob storage container by using SQL Server Management Studio.
- Provision an Azure Blob storage container.
- Run Transact-SQL statements to load data.
- Connect to the Azure SQL Data Warehouse by using SQL Server Management Studio.
- Build external tables by using Azure portal.
- Build external tables by using SQL Server Management Studio.

Answer Area

- Provision an Azure SQL Data Warehouse instance.
- Connect to the Blob storage container by using SQL Server Management Studio.
- Build external tables by using SQL Server Management Studio.
- Run Transact-SQL statements to load data.

Step 1: Provision an Azure SQL Data Warehouse instance.

Create a data warehouse in the Azure portal.

Step 2: Connect to the Azure SQL Data warehouse by using SQL Server Management Studio

Connect to the data warehouse with SSMS (SQL Server Management Studio)

Step 3: Build external tables by using the SQL Server Management Studio

Create external tables for data in Azure blob storage.

You are ready to begin the process of loading data into your new data warehouse. You use external tables to load data from the Azure storage blob.

Step 4: Run Transact-SQL statements to load data.

You can use the CREATE TABLE AS SELECT (CTAS) T-SQL statement to load the data from Azure Storage Blob into new tables in your data warehouse.

References:

https://github.com/MicrosoftDocs/azure-docs/blob/master/articles/sql-data-warehouse/load-data-from-azure-blob-storage-using-polybase.md

ssanka `Highly Voted 👍` 4 years, 2 months ago

Step 2 should be "Connect to SQL data warehouse using SSMS

upvoted 15 times

massnonn `Most Recent ⊙` 3 years, 7 months ago

also for me "Connect to SQL data warehouse using SSMS" on step 2

upvoted 1 times

Williammm 4 years ago

nope step 2 is point to Blob Storage (check polybase load steps!)

upvoted 1 times

fumaniM 4 years, 1 month ago

Step 2:Connect to the Azure SQL Data warehouse by using SQL Server Management Studio

upvoted 4 times

Aakansha01 4 years, 1 month ago

Why are we using sql data warehouse instance?

upvoted 1 times

tucho 4 years, 2 months ago

I've a doubt about step 2. Which is the correct one?

"Connect to the Blob Storage container by using SQL Server Management Studio."

or

"Connect to the Azure SQL Data Warehouse by using SQL Server Management Studio"

Thank you in advance.

upvoted 3 times

DongDuong 4 years, 2 months ago

look at the explanation, it should be Azure SQL Data Warehouse. Blob storage doesn't make sense here.

upvoted 9 times

You develop data engineering solutions for a company. The company has on-premises Microsoft SQL Server databases at multiple locations.

The company must integrate data with Microsoft Power BI and Microsoft Azure Logic Apps. The solution must avoid single points of failure during connection and transfer to the cloud. The solution must also minimize latency.

You need to secure the transfer of data between on-premises databases and Microsoft Azure.

What should you do?

    A. Install a standalone on-premises Azure data gateway at each location

    B. Install an on-premises data gateway in personal mode at each location

    C. Install an Azure on-premises data gateway at the primary location

    D. Install an Azure on-premises data gateway as a cluster at each location

Suggested Answer: *D*

You can create high availability clusters of On-premises data gateway installations, to ensure your organization can access on-premises data resources used in

Power BI reports and dashboards. Such clusters allow gateway administrators to group gateways to avoid single points of failure in accessing on-premises data resources. The Power BI service always uses the primary gateway in the cluster, unless it's not available. In that case, the service switches to the next gateway in the cluster, and so on.

References:

https://docs.microsoft.com/en-us/power-bi/service-gateway-high-availability-clusters

---

☐ 👤 **Treadmill** `Highly Voted 👍` 5 years, 1 month ago

https://docs.microsoft.com/en-us/data-integration/gateway/service-gateway-high-availability-clusters

upvoted 9 times

☐ 👤 **chaoxes** `Highly Voted 👍` 4 years, 6 months ago

D. Clusters is correct. The answer is ok.

upvoted 5 times

☐ 👤 **syu31svc** `Most Recent ⊘` 4 years, 7 months ago

D is correct

upvoted 2 times

☐ 👤 **uomer** 4 years, 7 months ago

IGNORE MY EARLIER COMMENT,

The answer is spark because SPARK has the functionality of INTERACTIVE QUERIES AS WELL.

Spark In-memory processing, interactive queries, micro-batch stream processing

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-provision-linux-clusters

upvoted 1 times

You are a data architect. The data engineering team needs to configure a synchronization of data between an on-premises Microsoft SQL Server database to
Azure SQL Database.
Ad-hoc and reporting queries are being overutilized the on-premises production instance. The synchronization process must:
✏ Perform an initial data synchronization to Azure SQL Database with minimal downtime
✏ Perform bi-directional data synchronization after initial synchronization
You need to implement this synchronization solution.
Which synchronization method should you use?

A. transactional replication

B. Data Migration Assistant (DMA)

C. backup and restore

D. SQL Server Agent job

E. Azure SQL Data Sync

Suggested Answer: *E*
SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you select bi-directionally across multiple SQL databases and SQL
Server instances.
With Data Sync, you can keep data synchronized between your on-premises databases and Azure SQL databases to enable hybrid applications.
Compare Data Sync with Transactional Replication

	Data Sync	Transactional Replication
Advantages	- Active-active support - Bi-directional between on-premises and Azure SQL Database	- Lower latency - Transactional consistency - Reuse existing topology after migration
Disadvantages	- 5 min or more latency - No transactional consistency - Higher performance impact	- Can't publish from Azure SQL Database single database or pooled database - High maintenance cost

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-sync-data

---

⊟ 👤 **promiseve** 4 years, 4 months ago
the answer is E
   upvoted 2 times

⊟ 👤 **chaoxes** 4 years, 6 months ago
Correct answer is E.
SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you select bi-directionally across multiple databases, both on-premises and in the cloud.

Source: https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-data-sync-data-sql-server-sql-database
   upvoted 4 times

   ⊟ 👤 **memo43** 4 years, 1 month ago
   keyword is bi-directional
      upvoted 1 times

⊟ 👤 **syu31svc** 4 years, 7 months ago
Can only be E
   upvoted 1 times

⊟ 👤 **jgil** 4 years, 10 months ago
https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-data-sync-data-sql-server-sql-database
   upvoted 2 times

An application will use Microsoft Azure Cosmos DB as its data solution. The application will use the Cassandra API to support a column-based database type that uses containers to store items.

You need to provision Azure Cosmos DB. Which container name and item name should you use? Each correct answer presents part of the solutions.

NOTE: Each correct answer selection is worth one point.

- A. collection
- B. rows
- C. graph
- D. entities
- E. table

Suggested Answer: *BE*

B: Depending on the choice of the API, an Azure Cosmos item can represent either a document in a collection, a row in a table or a node/edge in a graph. The following table shows the mapping between API-specific entities to an Azure Cosmos item:

Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos item	Document	Row	Document	Node or Edge	Item

E: An Azure Cosmos container is specialized into API-specific entities as follows:

Azure Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos container	Collection	Table	Collection	Graph	Table

References:

https://docs.microsoft.com/en-us/azure/cosmos-db/databases-containers-items

---

👤 **chaoxes** `Highly Voted 👍` 4 years, 6 months ago

Correct answer: B (row) & E (table)

Cassandra API: Table/Row

Table API: Table/Item

Mongo API: Collection/Document

SQL Core: Container/Item

Gremlin: Graph/Node&Edge

upvoted 23 times

👤 **bulbanos** `Highly Voted 👍` 4 years, 9 months ago

Not directly related but they've updated the screenshot in the question: https://docs.microsoft.com/en-us/azure/cosmos-db/databases-containers-items#azure-cosmos-containers

SQL API => Container and not Collection

upvoted 8 times

👤 **Rizmi** `Most Recent ⊘` 4 years, 7 months ago

B&E are the correct answers.

upvoted 2 times

👤 **syu31svc** 4 years, 7 months ago

100% B & E

upvoted 2 times

👤 **Egocentric** 4 years, 7 months ago

its obvious answer is B & E

upvoted 1 times

macvj87 4 years, 8 months ago

Answer is B & E

upvoted 2 times

macvj87 4 years, 8 months ago

Answer is B & E

upvoted 2 times

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization.

Customer organizations have peak usage at different periods during the year.

You need to implement the Azure SQL Database elastic pool to minimize cost.

Which option or options should you configure?

    A. Number of transactions only

    B. eDTUs per database only

    C. Number of databases only

    D. CPU usage only

    E. eDTUs and max data size

Suggested Answer: *E*

The best size for a pool depends on the aggregate resources needed for all databases in the pool. This involves determining the following:

☞ Maximum resources utilized by all databases in the pool (either maximum DTUs or maximum vCores depending on your choice of resourcing model).

☞ Maximum storage bytes utilized by all databases in the pool.

Note: Elastic pools enable the developer to purchase resources for a pool shared by multiple databases to accommodate unpredictable periods of usage by individual databases. You can configure resources for the pool based either on the DTU-based purchasing model or the vCore-based purchasing model.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-pool

---

👤 **induna** `Highly Voted 👍` 5 years ago

An additional note from the reference link:

"There is no per-database charge for elastic pools. You are billed for each hour a pool exists at the highest eDTU or vCores, regardless of usage or whether the pool was active for less than an hour."

So no need to figure out the needs of an individual database to size the pool

upvoted 10 times

👤 **JESUSBB** `Most Recent ⊘` 4 years, 4 months ago

Answer E

upvoted 4 times

👤 **chaoxes** 4 years, 6 months ago

Answer is correct - E.

How do I choose the correct pool size?

The best size for a pool depends on the aggregate resources needed for all databases in the pool. This involves determining the following:

-> Maximum resources utilized by all databases in the pool (either maximum DTUs or maximum vCores depending on your choice of purchasing model).

-> Maximum storage bytes utilized by all databases in the pool.

Source: https://docs.microsoft.com/en-us/azure/azure-sql/database/elastic-pool-overview

upvoted 3 times

👤 **syu31svc** 4 years, 7 months ago

Link provided supports E as the right answer

upvoted 1 times

👤 **ThieryLeLuronNadieline** 4 years, 7 months ago

indeed the correct answer is E

upvoted 1 times

👤 **Egocentric** 4 years, 7 months ago

answer is E

upvoted 1 times

☐ 👤 **Rickii** 4 years, 10 months ago

Answer is E

upvoted 3 times

☐ 👤 **Vijaya** 4 years, 10 months ago

so answer here is B i.e eDTU's per database only

upvoted 1 times

☐ 👤 **Varma_Saraswathula** 4 years, 10 months ago

The best size for a pool depends on the aggregate resources needed for all databases in the pool. This involves determining the following:

Maximum resources utilized by all databases in the pool (either maximum DTUs or maximum vCores depending on your choice of purchasing model).

Maximum storage bytes utilized by all databases in the pool.

upvoted 2 times

☐ 👤 **groy** 4 years, 9 months ago

Vijaya...please don't confuse people, answer is E!

upvoted 13 times

HOTSPOT -

You are a data engineer. You are designing a Hadoop Distributed File System (HDFS) architecture. You plan to use Microsoft Azure Data Lake as a data storage repository.

You must provision the repository with a resilient data schema. You need to ensure the resiliency of the Azure Data Lake Storage. What should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Requirement	Node
Provide data access to clients.	DataNode / NameNode
Run operations on files and directories of the file system.	DataNode / NameNode
Perform block creation, deletion, and replication.	DataNode / NameNode

Suggested Answer:

Answer Area

Requirement	Node
Provide data access to clients.	DataNode / **NameNode**
Run operations on files and directories of the file system.	**DataNode** / NameNode
Perform block creation, deletion, and replication.	**DataNode** / NameNode

Box 1: NameNode -

An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients.

Box 2: DataNode -

The DataNodes are responsible for serving read and write requests from the file system's clients.

Box 3: DataNode -

The DataNodes perform block creation, deletion, and replication upon instruction from the NameNode.

Note: HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

References:

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

Hassan_Mazhar_Khan `Highly Voted 👍` 4 years, 2 months ago

1.- NameNode (Its asking about data access meaning permission not data provider so name node has this responsibility)

2.- NameNode (Name Node perform namespace and etc on files system)

3.- DataNode

upvoted 22 times

   Steve92873197 4 years ago

   Agree, doc link supports this.

   upvoted 1 times

massnonn `Most Recent ⊙` 3 years, 7 months ago

i think that is the correct answer is 1) name node 2) name node 3) data node

upvoted 1 times

Kostali 3 years, 11 months ago

1.- DataNode

2.- NameNode

3.- DataNode

upvoted 1 times

eurekamike 4 years ago

NameNode, a master server that manages the file system namespace and regulates access to files by clients.

NameNode executes file system namespace operations like opening, closing, and renaming files and directories.

DataNodes are responsible for serving read and write requests from the file system's clients.
The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

upvoted 1 times

Williammm 4 years ago

NDD (last one >> Data Node, The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.)

upvoted 2 times

LordSnoek 4 years ago

1)DataNode

2)NameNode

3)DataNode

This is what Whizlab's exam answer is

upvoted 2 times

ThiruthuvaRajan 4 years ago

Run operations on Files and Directories.. DataNode has no work on these operations.. It has to serve to create, read,write blocks. Not name space work as per documentation.

It is Name Node, Name Node, Data Node

upvoted 3 times

davem0193 4 years, 1 month ago

The answer provided as a solution is correct.

1. Namenode

2. Datanode

3. Datanode

The apache ref link provided in the answer has the correct pointers

upvoted 1 times

   HrishiD 4 years, 1 month ago

   N, N, D is correct.

   Excerpt from the link provided:

   The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

   upvoted 4 times

Varunkris 4 years, 1 month ago

I believe the answers to be 1.DataNode 2.NameNode 3.DataNode

upvoted 4 times

IAMKPR 4 years, 2 months ago

As per the line "A master server that manages the file system namespace and regulates access to files by clients" so the first option should be NameNode.

upvoted 1 times

Psycho 4 years, 2 months ago

So what's the correct answer?

upvoted 1 times

cadio30 4 years, 2 months ago

NameNode, DataNode, DataNode

upvoted 3 times

cadio30 4 years, 1 month ago

It should be NameNode,NameNode,DataNode.

The second question is related to file namespace and according to the documentation "The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes."

Reference: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

upvoted 4 times

AngelRio 4 years, 2 months ago

1.- DataNode

2.- NameNode

3.- DataNode

upvoted 3 times

DRAG DROP -

You are developing the data platform for a global retail company. The company operates during normal working hours in each region. The analytical database is used once a week for building sales projections.

Each region maintains its own private virtual network.

Building the sales projections is very resource intensive and generates upwards of 20 terabytes (TB) of data.

Microsoft Azure SQL Databases must be provisioned.

☞ Database provisioning must maximize performance and minimize cost

☞ The daily sales for each region must be stored in an Azure SQL Database instance

☞ Once a day, the data for all regions must be loaded in an analytical Azure SQL Database instance

You need to provision Azure SQL database instances.

How should you provision the database instances? To answer, drag the appropriate Azure SQL products to the correct databases. Each Azure SQL product 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.

Select and Place:

Azure SQL products

- Azure SQL Database elastic pools
- Azure SQL Database Premium
- Azure SQL Database Managed Instance
- Azure SQL Database Hyperscale

Answer area

Database	Azure SQL product
Daily Sales	Azure SQL product
Weekly Sales Projections	Azure SQL product

Suggested Answer:

Azure SQL products

- Azure SQL Database elastic pools
- Azure SQL Database Premium
- Azure SQL Database Managed Instance
- Azure SQL Database Hyperscale

Answer area

Database	Azure SQL product
Daily Sales	Azure SQL Database elastic pools
Weekly Sales Projections	Azure SQL Database Hyperscale

Box 1: Azure SQL Database elastic pools

SQL Database elastic pools are a simple, cost-effective solution for managing and scaling multiple databases that have varying and unpredictable usage demands. The databases in an elastic pool are on a single Azure SQL Database server and share a set number of resources at a set price. Elastic pools in Azure

SQL Database enable SaaS developers to optimize the price performance for a group of databases within a prescribed budget while delivering performance elasticity for each database.

Box 2: Azure SQL Database Hyperscale

A Hyperscale database is an Azure SQL database in the Hyperscale service tier that is backed by the Hyperscale scale-out storage technology. A Hyperscale database supports up to 100 TB of data and provides high throughput and performance, as well as rapid scaling to adapt to the workload requirements. Scaling is transparent to the application ג€" connectivity, query processing, and so on, work like any other SQL database.

Incorrect Answers:

Azure SQL Database Managed Instance: The managed instance deployment model is designed for customers looking to migrate a large number of apps from on- premises or IaaS, self-built, or ISV provided environment to fully managed PaaS cloud environment, with as low migration effort as possible.

References:

memo43 `Highly Voted` 4 years, 1 month ago

ANSWER IS CORRECT

• Azure SQL Database Elastic Pools >>Simple, Cost-effective, scaling multiple db, deploy multiple db to a single logical instance

• Azure SQL Database Premium >> better capabilities for the database

• Azure SQL Database Managed Instance >> migrate large numbers of apps on-premise to Azure, use CLR features ,IASS

>> Back up and restore

• Azure SQL Database Hyperscale >> Up to 100TB, high performance

So;

Azure SQL Database Premium and Azure SQL Database Managed Instance are irrelavent

for Daily Sales >> Azure SQL Database Elastic Pools (from multiple region to single location)

for weekly projections >> Azure SQL Database Hyperscale (up to 100TB)

upvoted 6 times

ThiruthuvaRajan `Most Recent` 4 years ago

Daily sales itself will go around 20TB.. I don't see any model is providing that much limit as per the document.

Anyone have document to prove that "Elastic Pools" is the answer ?

https://docs.microsoft.com/en-us/azure/azure-sql/database/resource-limits-dtu-elastic-pools

upvoted 2 times

memo43 `Highly Voted` 4 years, 1 month ago

ANSWER IS CORRECT

• Azure SQL Database Elastic Pools >>Simple, Cost-effective, scaling multiple db, deploy multiple db to a single logical instance

• Azure SQL Database Premium >> better capabilities for the database

• Azure SQL Database Managed Instance >> migrate large numbers of apps on-premise to Azure, use CLR features ,IASS

>> Back up and restore

• Azure SQL Database Hyperscale >> Up to 100TB, high performance

A company manages several on-premises Microsoft SQL Server databases.
You need to migrate the databases to Microsoft Azure by using a backup process of Microsoft SQL Server.
Which data technology should you use?

    A. Azure SQL Database single database

    B. Azure SQL Data Warehouse

    C. Azure Cosmos DB

    D. Azure SQL Database Managed Instance

Suggested Answer: *D*
Managed instance is a new deployment option of Azure SQL Database, providing near 100% compatibility with the latest SQL Server on-premises (Enterprise
Edition) Database Engine, providing a native virtual network (VNet) implementation that addresses common security concerns, and a business model favorable for on-premises SQL Server customers. The managed instance deployment model allows existing SQL Server customers to lift and shift their on-premises applications to the cloud with minimal application and database changes.
References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-managed-instance

👤 **skiwi** `Highly Voted 👍` 5 years, 2 months ago
You can only do backup and restore to a managed instance. See page 60 https://azure.microsoft.com/mediahandler/files/resourcefiles/choosing-your-database-migration-path-to-azure/Choosing_your_database_migration_path_to_Azure.pdf
upvoted 26 times

👤 **Noor_S** `Highly Voted 👍` 5 years ago
when ever you are choosing to migrate indirectly we would need to do it in a cost effective way which is possible through azure sql database managed instance.
upvoted 6 times

👤 **dmnantilla9** `Most Recent ⊙` 4 years, 5 months ago
D: es correct
upvoted 2 times

👤 **PaulMD** 4 years, 6 months ago
The managed instance is the resource, which is closest to an on-prem SQL-server, whereas a SQL database is limited (e.g. support for R). The lift and shift is the least effort solution for a SQL server
upvoted 1 times

👤 **chaoxes** 4 years, 6 months ago
Correct answer is D: Azure SQL Managed Instance.

Azure SQL Managed Instance is designed for customers looking to migrate a large number of apps from an on-premises or IaaS, self-built, or ISV provided environment to a fully managed PaaS cloud environment, with as low a migration effort as possible.
upvoted 2 times

👤 **KakashiHatake** 4 years, 7 months ago
This Question came in DP-201 in Nov 2020
upvoted 2 times

    👤 **Abhishek_123** 4 years, 6 months ago
    On which date u gave the exam?
    upvoted 1 times

👤 **Egocentric** 4 years, 7 months ago
the keywords are "migrating databases" hence single db is wrong.the answer is SQL Managed Instance
upvoted 2 times

👤 **javaxxl** 4 years, 9 months ago

SQL Managed Instance allows existing SQL Server customers to lift and shift their on-premises applications to the cloud with minimal application and database changes. At the same time, SQL Managed Instance preserves all PaaS capabilities (automatic patching and version updates, automated backups, high availability) that drastically reduce management overhead and TCO.

upvoted 4 times

Vijaya 4 years, 10 months ago

the given answer is correct because it is several on-prem DB's

upvoted 2 times

avestabrzn 5 years, 3 months ago

Why isn't it single db? If the question asked a backup and restore then I would prefer to choose managed db.

upvoted 2 times

Yuri1101 5 years, 2 months ago

Because there are several databases. Single db is only good for very few databases.

upvoted 5 times

The data engineering team manages Azure HDInsight clusters. The team spends a large amount of time creating and destroying clusters daily because most of the data pipeline process runs in minutes.

You need to implement a solution that deploys multiple HDInsight clusters with minimal effort.

What should you implement?

A. Azure Databricks

B. Azure Traffic Manager

C. Azure Resource Manager templates

D. Ambari web user interface

Suggested Answer: *C*

A Resource Manager template makes it easy to create the following resources for your application in a single, coordinated operation:

☞ HDInsight clusters and their dependent resources (such as the default storage account).

☞ Other resources (such as Azure SQL Database to use Apache Sqoop).

In the template, you define the resources that are needed for the application. You also specify deployment parameters to input values for different environments.

The template consists of JSON and expressions that you use to construct values for your deployment.

References:

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-create-linux-clusters-arm-templates

☐ 👤 **dmnantilla9** 4 years, 5 months ago

C. Azure Resource Manager templates,

Is correct

upvoted 3 times

☐ 👤 **syu31svc** 4 years, 7 months ago

C is correct; afterall Templates are meant to make life easier ain't it?

upvoted 4 times

☐ 👤 **M0e** 4 years, 9 months ago

Are the topics related to HDInsight still part of DP-200 exam?

upvoted 3 times

☐ 👤 **JParzival** 4 years, 8 months ago

yes, they are

upvoted 2 times

You are the data engineer for your company. An application uses a NoSQL database to store data. The database uses the key-value and wide-column NoSQL database type.

Developers need to access data in the database using an API.

You need to determine which API to use for the database model and type.

Which two APIs should you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

    A. Table API

    B. MongoDB API

    C. Gremlin API

    D. SQL API

    E. Cassandra API

Suggested Answer: *BE*

B: Azure Cosmos DB is the globally distributed, multimodel database service from Microsoft for mission-critical applications. It is a multimodel database and supports document, key-value, graph, and columnar data models.

E: Wide-column stores store data together as columns instead of rows and are optimized for queries over large datasets. The most popular are Cassandra and

HBase.

References:

https://docs.microsoft.com/en-us/azure/cosmos-db/graph-introduction https://www.mongodb.com/scale/types-of-nosql-databases

---

👤 **mustaphaa** `Highly Voted 👍` 5 years, 7 months ago

it should ba A and E

upvoted 97 times

  👤 **Vijaya** 4 years, 10 months ago

  as per documentation mongo API is multi-model and supports document, key-value, graph, and columnar data models so answer is correct ref https://docs.microsoft.com/en-us/azure/cosmos-db/mongodb-introduction

  upvoted 2 times

    👤 **ThiruthuvaRajan** 4 years ago

    Answer is clearly A and E. No documentation refers to Mongo API .

    upvoted 1 times

    👤 **Sasidhar39** 4 years, 1 month ago

    No, the documentation didn't refer to that

    upvoted 1 times

👤 **avestabrzn** `Highly Voted 👍` 5 years, 3 months ago

key value --> Table API

graph --> Gremlin API

document --> SQL API and MongoDB API

columnar --> Cassandra API

so, the answer is A,E

upvoted 87 times

  👤 **Nayon** 4 years ago

  right answer is A,E but slight change in your description.

  key value --> Cassandra API

  graph --> Gremlin API

  document --> SQL API and MongoDB API

  columnar --> Table API

  https://cloud.netapp.com/blog/azure-cvo-blg-azure-nosql-types-services-and-a-quick-tutorial#H_H1

  upvoted 3 times

👤 **massnonn** `Most Recent ⊘` 3 years, 7 months ago

the correct answer is Table Api e Cassandra Api. so A - E

upvoted 1 times

 👤 **LG5** 4 years, 3 months ago

It's A and E.

upvoted 1 times

 👤 **deepz8** 4 years, 3 months ago

Key-Value -> Table API

Wide column -> Cassandra API

upvoted 1 times

 👤 **SilNilanjan** 4 years, 4 months ago

A & E certainly

upvoted 1 times

 👤 **ssantos** 4 years, 5 months ago

Table API and Cassandra API for sure

upvoted 1 times

 👤 **afonsomeireles** 4 years, 5 months ago

It's A (Table API for key-value pairs) and E (Cassandra API for wide-columns)

upvoted 1 times

 👤 **rmn900** 4 years, 6 months ago

AE. Table uses key/value. Not MongoDB. B is wrong.

upvoted 1 times

 👤 **chaoxes** 4 years, 6 months ago

Should be A (Table API) and E (Cassandra API).

Table API is the best solution for key-value model.

upvoted 1 times

 👤 **ViniJsr** 4 years, 7 months ago

Azure CosmosDB table API is a key-value storage hosted in the cloud. It's a part of Azure Cosmos DB, that is Microsoft's multi-model database. It's a globally distributed, low latency, high throughput solution with client SDKs available for . NET, Java, Python, and Node. So Answer will be A and E

upvoted 1 times

 👤 **syu31svc** 4 years, 7 months ago

It's A & E

Cassandra is wide-table and Table is key

upvoted 1 times

 👤 **Aditya167** 4 years, 7 months ago

it should be A and E

upvoted 1 times

 👤 **UKiran** 4 years, 7 months ago

Answer is Table APi and cassandra API

upvoted 2 times

 👤 **lingjun** 4 years, 7 months ago

I think it should be MongoDB, since it uses key-value (which is the json format in document). but not storing the key-value pairs(then should be Table API)

upvoted 1 times

> 👤 **lingjun** 4 years, 7 months ago
>
> https://www.mongodb.com/key-value-database
>
> upvoted 1 times

> 👤 **lingjun** 4 years, 7 months ago
>
> ignore it. I would go for Table
>
> upvoted 1 times

 👤 **narun** 4 years, 8 months ago

It should be D and E.

Before Azure Cosmos DB existed, Redis or the Table API might have been a good fit for this kind of data; however, Core (SQL) API is now the better choice, as it offers a richer query experience, with improved indexing over the Table API.

upvoted 4 times

☐ 👤 **nacondras** 4 years, 9 months ago

The answer should be Table API for key value and Cassandra API for wide column

upvoted 2 times

☐ 👤 **nacondras** 4 years, 9 months ago

The answer should be Table API for key value and Cassandra API for wide column

upvoted 2 times

Question #12   *Topic 1*

A company is designing a hybrid solution to synchronize data and on-premises Microsoft SQL Server database to Azure SQL Database.
You must perform an assessment of databases to determine whether data will move without compatibility issues. You need to perform the assessment.
Which tool should you use?

    A. SQL Server Migration Assistant (SSMA)

    B. Microsoft Assessment and Planning Toolkit

    C. SQL Vulnerability Assessment (VA)

    D. Azure SQL Data Sync

    E. Data Migration Assistant (DMA)

> **Suggested Answer:** *E*
> The Data Migration Assistant (DMA) helps you upgrade to a modern data platform by detecting compatibility issues that can impact database functionality in your new version of SQL Server or Azure SQL Database. DMA recommends performance and reliability improvements for your target environment and allows you to move your schema, data, and uncontained objects from your source server to your target server.
> References:
> https://docs.microsoft.com/en-us/sql/dma/dma-overview

👤 **Debjit** `Highly Voted 👍` 4 years, 9 months ago

Correct answer is DMA. Data Migration Assistant is a client-side tool that you can install on a Windows-compatible workstation or server. It has two major functions in the migration of the social database to the Azure SQL Database platform in this module. First, it assesses your existing database and identifies any incompatibilities between that database and Azure SQL Database. It then generates a report of the things you need to fix before you can migrate. As you make changes, you can rerun Data Migration Assistant to generate an updated report of changes that you need to make. This capability helps you to not only track your progress, but also catch any new issues that might have been introduced during your coding phase.
Refrence: https://docs.microsoft.com/en-us/learn/modules/migrate-sql-server-relational-data/3-migration-overview

upvoted 24 times

   👤 **memo43** 4 years, 1 month ago

   keyword is >> Assessment of db

   upvoted 1 times

👤 **deepz8** `Most Recent ⊘` 4 years, 3 months ago

DMA is the correct answer. It is used for compatibility check prior to migration.

upvoted 3 times

👤 **JESUSBB** 4 years, 4 months ago

Answer E Data Migration Assistant.

upvoted 1 times

👤 **Needium** 4 years, 6 months ago

The correct answer is definitely the chosen answer E. The question is about assessement not the actual migration and/or synchronization. Azure Data Sync (E) addresses the actual migration and synchronization; however, the question is about assessment before migration

upvoted 2 times

👤 **syu31svc** 4 years, 7 months ago

From link https://docs.microsoft.com/en-us/sql/dma/dma-overview?view=sql-server-ver15:
Assess on-premises SQL Server instance(s) migrating to Azure SQL database(s). The assessment workflow helps you to detect the following issues that can affect Azure SQL database migration and provides detailed guidance on how to resolve them.

upvoted 2 times

👤 **Egocentric** 4 years, 7 months ago

"perform an assessment" is the keyword here not sync . Data Migration Assistant is the correct one

upvoted 2 times

👤 **abhinavkj** 4 years, 11 months ago

I also feel it should be Azure Data Sync

upvoted 2 times

r8d1 4 years, 10 months ago

i think the question is about first migrating data, so we need to address this issue first. The sync comes later.

upvoted 6 times

SebK 4 years, 11 months ago

Should not be D: Azure Data Sync as it is said to synchronize data ? Data Migration Assistant should be used to migrate data from on-premise to Azure.

upvoted 2 times

cutebird 4 years, 11 months ago

It is mentioned that "perform an assessment of databases to determine whether data will move without compatibility issues".So E is correct

Ref:https://docs.microsoft.com/en-us/sql/dma/dma-overview?view=sql-server-ver15

upvoted 12 times

r8d1 4 years, 10 months ago

i think the question is about first migrating data, so we need to address this issue first. The sync comes later.

upvoted 6 times

SebK 4 years, 11 months ago

Should not be D: Azure Data Sync as it is said to synchronize data ? Data Migration Assistant should be used to migrate data from on-premise to Azure.

upvoted 2 times

cutebird 4 years, 11 months ago

It is mentioned that "perform an assessment of databases to determine whether data will move without compatibility issues".So E is correct

Ref:https://docs.microsoft.com/en-us/sql/dma/dma-overview?view=sql-server-ver15

DRAG DROP -

You manage a financial computation data analysis process. Microsoft Azure virtual machines (VMs) run the process in daily jobs, and store the results in virtual hard drives (VHDs.)

The VMs product results using data from the previous day and store the results in a snapshot of the VHD. When a new month begins, a process creates a new

VHD.

You must implement the following data retention requirements:

☞ Daily results must be kept for 90 days

☞ Data for the current year must be available for weekly reports

☞ Data from the previous 10 years must be stored for auditing purposes

☞ Data required for an audit must be produced within 10 days of a request.

You need to enforce the data retention requirements while minimizing cost.

How should you configure the lifecycle policy? To answer, drag the appropriate JSON segments to the correct locations. Each JSON segment may be used once, more than once, or not at all. You may need to drag the split bat between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Code segments

| delete |

| blockBob |

| baseBlob |

| snapshot |

| tierToCool |

| tierToArchive |

Answer Area

```
{
"version": "0.5",
"rules": [
{
"name":"dataRetention",
"type":"Lifecycle",
"definition":{
"actions":{
"            ":{

"            ":{daysAfterModificationGreaterThan":365},

"            ":{daysAfterModificationGreaterThan":3650}

},
    "            ":{

    "            ":{"daysAfterCreationGreaterThan": 90}

    }
   }
  }
 ]
}
```

Suggested Answer:

Code segments	Answer Area

Code segments:
- delete
- blockBob
- baseBlob
- snapshot
- tierToCool
- tierToArchive

Answer Area:

```
{
"version": "0.5",
"rules": [
{
"name":"dataRetention",
"type":"Lifecycle",
"definition":{
"actions":{
" baseBlob ":{

" tierToArchive ":{daysAfterModificationGreaterThan":365},

" delete ":{daysAfterModificationGreaterThan":3650}

},

" snapshot ":{

" tierToCool ":{"daysAfterCreationGreaterThan": 90}

}
}
}
}
}
```

The Set-AzStorageAccountManagementPolicy cmdlet creates or modifies the management policy of an Azure Storage account.

Example: Create or update the management policy of a Storage account with ManagementPolicy rule objects.

```
"Actions":  {
        "BaseBlob":  {
                "TierToCool":  {
                        "DaysAfterModificationGreaterThan":
                },
                "TierToArchive":  {
                        "DaysAfterModificationGreaterThan
                },
                "Delete":  {
                        "DaysAfterModificationGreaterThan":  100
                }
        },
        "Snapshot":  {
                "Delete":  {
                        "DaysAfterCreationGreaterThan":  100
                }
        }
}
```

Action -BaseBlobAction Delete -daysAfterModificationGreaterThan 100

PS C:\>$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject $action1 -BaseBlobAction TierToArchive - daysAfterModificationGreaterThan 50

PS C:\>$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject $action1 -BaseBlobAction TierToCool - daysAfterModificationGreaterThan 30

PS C:\>$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject $action1 -SnapshotAction Delete - daysAfterCreationGreaterThan 100

PS C:\>$filter1 = New-AzStorageAccountManagementPolicyFilter -PrefixMatch ab,cd

PS C:\>$rule1 = New-AzStorageAccountManagementPolicyRule -Name Test -Action $action1 -Filter $filter1

PS C:\>$action2 = Add-AzStorageAccountManagementPolicyAction -BaseBlobAction Delete -daysAfterModificationGreaterThan 100

PS C:\>$filter2 = New-AzStorageAccountManagementPolicyFilter

References:

https://docs.microsoft.com/en-us/powershell/module/az.storage/set-azstorageaccountmanagementpolicy

I would rather choose "delete" in the last section

upvoted 6 times

ⓔ 👤 **memo43** 4 years, 1 month ago

Remember: Data for the current year must be available for weekly reports.

if you delete after 90 days you cant prepare weekly reports! so you must not delete them before 10 years.

ANSWER IS CORRECT

upvoted 12 times

ⓔ 👤 **memo43** Most Recent ⊘ 4 years, 1 month ago

Remember: Data for the current year must be available for weekly reports.

if you delete after 90 days you cant prepare weekly reports! so you must not delete them before 10 years.

ANSWER IS CORRECT

upvoted 4 times

ⓔ 👤 **hello_there_** 4 years ago

I think "data for the current year" refers to de data on the baseBlob, which is why the json part for the baseBlob is archived after 356 days. It explicitly says that snapshot data should be kept for 90 days, whitch is analogous to saying it can be deleted after 90.

upvoted 1 times

ⓔ 👤 **Amy007** 4 years, 1 month ago

Data should be kept for current year for reporting. So Greater than 90 cannot be deleted.

upvoted 2 times

ⓔ 👤 **Internet_User** 4 years, 2 months ago

I agree. The daily snapshots are not needed after the 90-day period, so should be deleted instead of being moved to cool tier.

upvoted 2 times

A company plans to use Azure SQL Database to support a mission-critical application.

The application must be highly available without performance degradation during maintenance windows.

You need to implement the solution.

Which three technologies should you implement? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Premium service tier

- B. Virtual machine Scale Sets

- C. Basic service tier

- D. SQL Data Sync

- E. Always On availability groups

- F. Zone-redundant configuration

Suggested Answer: *AEF*

A: Premium/business critical service tier model that is based on a cluster of database engine processes. This architectural model relies on a fact that there is always a quorum of available database engine nodes and has minimal performance impact on your workload even during maintenance activities.

E: In the premium model, Azure SQL database integrates compute and storage on the single node. High availability in this architectural model is achieved by replication of compute (SQL Server Database Engine process) and storage (locally attached SSD) deployed in 4-node cluster, using technology similar to SQL

Server Always On Availability Groups.
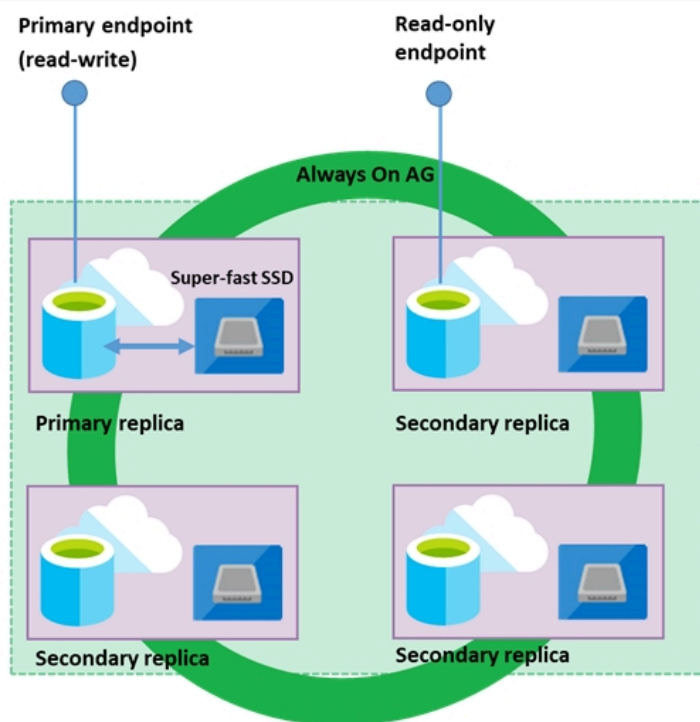
Business Critical service tier: collocated compute and storage

F: Zone redundant configuration -

By default, the quorum-set replicas for the local storage configurations are created in the same datacenter. With the introduction of Azure Availability Zones, you have the ability to place the different replicas in the quorum-sets to different availability zones in the same region. To eliminate a single point of failure, the control ring is also duplicated across multiple zones as three gateway rings (GW).

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-high-availability

👤 **bansal_vikrant** `Highly Voted 👍` 5 years ago

Look at the Conclusion section in the below link. It clearly mentions Always on and thus the provided answer is correct [A,E,F]

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-high-availability

upvoted 19 times

   ☐ 👤 **induna** 4 years, 11 months ago

   I think it is confusing because the Premium tier says "similar to Always On" but I agree with the answer because you can apply Always On AGs to the VMs that host the SQL Servers

   upvoted 4 times

   ☐ 👤 **hello_there_** 4 years ago

   The question asks "how should YOU configure". The documentation states that "High availability is implemented using a technology similar to SQL Server Always On availability groups", but this is an implementation detail of the premium/business critical service tiers, not something you have to configure yourself. That said, the other answers make even less sense.

   upvoted 1 times

☐ 👤 **sidharthamanu** `Highly Voted 👍` 5 years, 1 month ago

The given answer[A,E,F] is correct. Same can be understood from given link - https://docs.microsoft.com/en-us/azure/sql-database/sql-database-high-availability

upvoted 9 times

☐ 👤 **GeorgiP** `Most Recent ⊘` 4 years, 3 months ago

Correct answer is A & F; "Always On availability groups" are integral part of (already included in) Premium tier zone redundant availability:

"High availability is implemented using a technology similar to SQL Server Always On availability groups."

https://docs.microsoft.com/en-us/azure/azure-sql/database/high-availability-sla

For better understanding of the above once and for all, pls look at the illustrations too.

upvoted 2 times

☐ 👤 **syu31svc** 4 years, 7 months ago

Between Premium and Basic, Premium would be the better choice

Data Sync is definitely irrelevant for this question

VM Scale Sets is also wrong for sure

So answer is AEF

upvoted 4 times

☐ 👤 **Egocentric** 4 years, 7 months ago

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-high-availability. here there describe it nicely. A<E<F is the answer

upvoted 1 times

☐ 👤 **hart232** 4 years, 9 months ago

Re premium- High availability requirement - As an extra benefit, the premium availability model includes the ability to redirect read-only Azure SQL connections to one of the secondary replicas.

upvoted 1 times

☐ 👤 **LeandroAmore** 4 years, 11 months ago

The correct answer is A and F, in here https://docs.microsoft.com/en-us/azure/azure-sql/database/high-availability-sla microsoft says:

The underlying database files (.mdf/.ldf) are placed on the attached SSD storage to provide very low latency IO to your workload. High availability is implemented using a technology similar to SQL Server Always On availability groups.

but it says SIMILAR and you don't implement always on, Microsoft does it for you.

upvoted 3 times

   ☐ 👤 **zac874997967** 4 years, 10 months ago

   But the question asked for three technologies🤷🤷

   upvoted 4 times

☐ 👤 **Huepig** 5 years, 3 months ago

You can not implement Alway On Availability Group on Azure Database Premium or any SKU.

https://docs.microsoft.com/en-au/sql/database-engine/availability-groups/windows/overview-of-always-on-availability-groups-sql-server?view=sql-server-ver15

upvoted 8 times

   ☐ 👤 **zenomas** 5 years, 3 months ago

   Right, "Always On Available" only applicable to on-prem "SQL Server".

upvoted 1 times

abeworld 5 years, 2 months ago

Its on azure too : https://docs.microsoft.com/en-us/azure/virtual-machines/windows/sql/virtual-machines-windows-portal-sql-availability-group-overview

answer is correct
upvoted 7 times

abeworld 5 years, 2 months ago

actually this is for Azure VM - so my answer is not correct.
upvoted 2 times

M0e 4 years, 9 months ago

For Azure SQL DB it is called "Failover groups" so the correct answer is: Premium service tier, Failover groups and Zone-redundant configuration.
upvoted 1 times

soak 5 years, 1 month ago

This is true, always on availability groups is for on premise. Answer should be Premium service tier, SQL Data Sync and Zone-redundant configuration.
upvoted 1 times

A company plans to use Azure Storage for file storage purposes. Compliance rules require:

☞ A single storage account to store all operations including reads, writes and deletes

☞ Retention of an on-premises copy of historical operations

You need to configure the storage account.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

    A. Configure the storage account to log read, write and delete operations for service type Blob

    B. Use the AzCopy tool to download log data from $logs/blob

    C. Configure the storage account to log read, write and delete operations for service-type table

    D. Use the storage client to download log data from $logs/table

    E. Configure the storage account to log read, write and delete operations for service type queue

Suggested Answer: *AB*

Storage Logging logs request data in a set of blobs in a blob container named $logs in your storage account. This container does not show up if you list all the blob containers in your account but you can see its contents if you access it directly.

To view and analyze your log data, you should download the blobs that contain the log data you are interested in to a local machine. Many storage-browsing tools enable you to download blobs from your storage account; you can also use the Azure Storage team provided command-line Azure Copy Tool (AzCopy) to download your log data.

References:

https://docs.microsoft.com/en-us/rest/api/storageservices/enabling-storage-logging-and-accessing-log-data

---

👤 **Deepu1987** `Highly Voted 👍` 4 years, 5 months ago

I agree with the answer we have a new feature "Azure Storage Explorer" which is explained in detail in the video "https://www.youtube.com/watch?v=GJYAgi5eYYE" which confirms the answer with proof

upvoted 7 times

👤 **forrahul123** `Highly Voted 👍` 4 years, 6 months ago

i agree with the answer

upvoted 5 times

👤 **Piiri565** `Most Recent ⊘` 4 years, 7 months ago

I DIDN'T AGREE, When the question is asking for file storage, how the answer is related to blob storage.

upvoted 2 times

   👤 **mrsmjparker** 4 years, 7 months ago

Anything can go into blob storage, including files. However, files cannot go into tables or queues. Since there is no choice for files, blobs can be the only choice. AZ Copy is an excellent way to get data into a storage blob as well

upvoted 15 times

👤 **MHZ** 4 years, 7 months ago

I agree

upvoted 3 times

👤 **Sc2** 4 years, 8 months ago

I agree with the answer

upvoted 3 times

DRAG DROP -

You are developing a solution to visualize multiple terabytes of geospatial data.

The solution has the following requirements:

☞ Data must be encrypted.

☞ Data must be accessible by multiple resources on Microsoft Azure.

You need to provision storage for the solution.

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

Select and Place:

Actions

- Enable encryption on the Azure Data Lake using the Azure portal.
- Add an access policy for the new Azure Data Lake account to the key storage container.
- Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys.
- Select and configure an encryption key storage container.
- Create a new Azure Data Lake Storage account with Azure Key Vault managed encryption keys.
- Create a new Azure Data Lake Storage account with encryption disabled.

Answer Area

Suggested Answer:

Actions

- Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys.
- Create a new Azure Data Lake Storage account with encryption disabled.

Answer Area

1. Create a new Azure Data Lake Storage account with Azure Key Vault managed encryption keys.
2. Select and configure an encryption key storage container.
3. Add an access policy for the new Azure Data Lake account to the key storage container.
4. Enable encryption on the Azure Data Lake using the Azure portal.

Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys

For Azure services, Azure Key Vault is the recommended key storage solution and provides a common management experience across services. Keys are stored and managed in key vaults, and access to a key vault can be given to users or services. Azure Key Vault supports customer creation of keys or import of customer keys for use in customer-managed encryption key scenarios.

Note: Data Lake Storage Gen1 account Encryption Settings. There are three options:

☞ Do not enable encryption.

☞ Use keys managed by Data Lake Storage Gen1, if you want Data Lake Storage Gen1 to manage your encryption keys.

☞ Use keys from your own Key Vault. You can select an existing Azure Key Vault or create a new Key Vault. To use the keys from a Key Vault, you must assign permissions for the Data Lake Storage Gen1 account to access the Azure Key Vault.

References:
https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest

SebK `Highly Voted 👍` 5 years ago

In Data Lake Gen 2, encryption is enabled by default and can't be disabled.

https://docs.microsoft.com/en-us/learn/modules/secure-azure-storage-account/2-storage-security-features

upvoted 16 times

Miles19 `Highly Voted 👍` 5 years, 2 months ago

I think this question is obsolete as in the answer we can see that they are talking about Gen1. Microsoft recommends using Data Lake Gen2.

upvoted 13 times

hart232 `Most Recent ⊘` 4 years, 9 months ago

Confusing question.

upvoted 3 times

goodzilla 5 years, 3 months ago

One of the things that are not clear for me in this kind of questions is the right order of some answers.

In this case, 3rd and 4th answers couldn't those be switched as keep the right order of answers of this exercise?

upvoted 3 times

Leonido 5 years, 2 months ago

No, Actually you can't. If you don't have a policy pre-configured, you won't be able to designate the secret.

upvoted 5 times

You are developing a data engineering solution for a company. The solution will store a large set of key-value pair data by using Microsoft Azure Cosmos DB.

The solution has the following requirements:

☞ Data must be partitioned into multiple containers.

☞ Data containers must be configured separately.

☞ Data must be accessible from applications hosted around the world.

☞ The solution must minimize latency.

You need to provision Azure Cosmos DB.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

    A. Configure account-level throughput.

    B. Provision an Azure Cosmos DB account with the Azure Table API. Enable geo-redundancy.

    C. Configure table-level throughput.

    D. Replicate the data globally by manually adding regions to the Azure Cosmos DB account.

    E. Provision an Azure Cosmos DB account with the Azure Table API. Enable multi-region writes.

Suggested Answer: *E*

Scale read and write throughput globally. You can enable every region to be writable and elastically scale reads and writes all around the world. The throughput that your application configures on an Azure Cosmos database or a container is guaranteed to be delivered across all regions associated with your Azure Cosmos account. The provisioned throughput is guaranteed up by financially backed SLAs.

Reference:

https://docs.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally

---

☐ 👤 **JohnCrawford** `Highly Voted 👍` 4 years, 2 months ago

The answers are C, D and E.

• C. Configure table-level throughput. Requirements state that containers must be configured separately.

• D. Replicate the data globally by manually adding regions to the Azure Cosmos DB account. By adding extra regions our data is automatically copied to those regions reducing latency.

• E. Provision an Azure Cosmos DB account with the Azure Table API. Enable multi-region writes. By enabling multi-region writes this also reduces latency since we don't have a single master database, but rather would be implementing a multi-master model.

upvoted 37 times

☐ 👤 **Maky2365** `Highly Voted 👍` 4 years, 1 month ago

question doesn't mention the requirement for multi-region write, So as per my understanding answer should be B, C, D

Please suggest if my understanding is correct

upvoted 8 times

☐ 👤 **massnonn** `Most Recent ⊘` 3 years, 7 months ago

Cosmos db support multi region write so C-D-E

upvoted 1 times

☐ 👤 **Wendy_DK** 4 years, 2 months ago

Answer is CDE

upvoted 1 times

☐ 👤 **AnilKJ** 4 years, 2 months ago

B,D,E is the answer

upvoted 2 times

    ☐ 👤 **VeeraSekhar** 4 years, 1 month ago

    https://docs.microsoft.com/en-us/azure/cosmos-db/set-throughput

    From the above link CosmosDB allows throughput at two levels

    Azure Cosmos containers

    Azure Cosmos databases

    Hence B,D,E is correct answer. Sometimes we have to choose answer from list of provided answers.

upvoted 2 times

cadio30 4 years, 1 month ago

Propose solution is C, D and E.

When we say Azure Cosmos containers we are pertaining to the "CONTAINERS" of what we chose in the creation of DB. The option C pertains to configuration in "Cassandra API" in which the name of the container is "TABLE" and there are also other containers such as Container for SQL API, Collection for Mongo DB, graph for Gremlin API and Table for Table API

Reference: https://azure.microsoft.com/en-us/blog/sharing-provisioned-throughput-across-multiple-containers-in-azure-cosmosdb/

upvoted 4 times

princy18 4 years, 2 months ago

There must be 3 answers, any one knows what are those.

upvoted 1 times

Devendra00023 4 years, 2 months ago

Answer is CDE

upvoted 1 times

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution will have a dedicated database for each customer organization.

Customer organizations have peak usage at different periods during the year.

Which two factors affect your costs when sizing the Azure SQL Database elastic pools? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. maximum data size

B. number of databases

C. eDTUs consumption

D. number of read operations

E. number of transactions

Suggested Answer: *AC*

A: With the vCore purchase model, in the General Purpose tier, you are charged for Premium blob storage that you provision for your database or elastic pool.

Storage can be configured between 5 GB and 4 TB with 1 GB increments. Storage is priced at GB/month.

C: In the DTU purchase model, elastic pools are available in basic, standard and premium service tiers. Each tier is distinguished primarily by its overall performance, which is measured in elastic Database Transaction Units (eDTUs).

References:

https://azure.microsoft.com/en-in/pricing/details/sql-database/elastic/

---

chaoxes `Highly Voted` 4 years, 6 months ago

Number of databases is not affecting pricing.

What affects is:

Storage: maximum data size

eDTU: elasitc database throughput unit - unit of measurement used to determine scale in elastic pool.

Thus correct answers: A & C

upvoted 17 times

syu31svc `Highly Voted` 4 years, 7 months ago

A & C 100%

upvoted 5 times

EYIT `Most Recent` 4 years, 9 months ago

out of scope post July 2020

upvoted 1 times

azp 4 years, 7 months ago

nope its not

upvoted 4 times

HOTSPOT -

You are developing a solution using a Lambda architecture on Microsoft Azure.

The data at rest layer must meet the following requirements:

Data storage:

☞ Serve as a repository for high volumes of large files in various formats.

☞ Implement optimized storage for big data analytics workloads.

☞ Ensure that data can be organized using a hierarchical structure.

Batch processing:

☞ Use a managed solution for in-memory computation processing.

☞ Natively support Scala, Python, and R programming languages.

☞ Provide the ability to resize and terminate the cluster automatically.

Analytical data store:

☞ Support parallel processing.

☞ Use columnar storage.

☞ Support SQL-based languages.

You need to identify the correct technologies to build the Lambda architecture.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Architecture requirement	Technology
Data storage	Azure SQL Database / Azure Blob Storage / Azure Cosmos DB / Azure Data Lake Store
Batch processing	HDInsight Spark / HDInsight Hadoop / Azure Databricks / HDInsight Interactive Query
Analytical data store	HDInsight HBase / Azure Synapse Analytics / Azure Analysis Services / Azure Cosmos DB

Answer Area

Architecture requirement	Technology

Data storage

- Azure SQL Database
- Azure Blob Storage
- Azure Cosmos DB
- **Azure Data Lake Store**

Batch processing

- **HDInsight Spark**
- HDInsight Hadoop
- Azure Databricks
- HDInsight Interactive Query

Analytical data store

- HDInsight HBase
- **Azure Synapse Analytics**
- Azure Analysis Services
- Azure Cosmos DB

Data storage: Azure Data Lake Store

A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.

Batch processing: HD Insight Spark

Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.

HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP,

MapReduce.

Languages: R, Python, Java, Scala, SQL

Analytic data store: Azure Synapse Analytics

Azure Synapse Analytics Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).

Azure Synapse Analytics stores data into relational tables with columnar storage.

Note: As of November 2019, Azure SQL Data Warehouse is now Azure Synapse Analytics.

References:

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is

---

gallego82 `Highly Voted` 4 years, 2 months ago

I think that in batch processing the answer sould be Azure DataBricks, due to the link provided an it´s capabilities:

Azure Databricks is an Apache Spark-based analytics platform. You can think of it as "Spark as a service." It's the easiest way to use Spark on the Azure platform.

Languages: R, Python, Java, Scala, Spark SQL

Fast cluster start times, autotermination, autoscaling.

Manages the Spark cluster for you.

Built-in integration with Azure Blob Storage, Azure Data Lake Storage (ADLS), Azure Synapse, and other services. See Data Sources.

User authentication with Azure Active Directory.

Web-based notebooks for collaboration and data exploration.

Supports GPU-enabled cluster

upvoted 38 times

Pairon `Highly Voted 👍` 4 years, 2 months ago

Agree with the comments above. Databricks enables you to autoscale and autoterminate your cluster and enables also to in-memory processing because of the undelying Spark engine.

upvoted 6 times

Palp `Most Recent ⊘` 4 years ago

Batch processing is Spark as it provides in memory operations

upvoted 1 times

AZ20 4 years ago

"terminate the cluster automatically" - I think this line makes Databricks a more suitable choice

Rest requirements suits both HDInsight and Databricks equally.

upvoted 3 times

MYR55 4 years, 1 month ago

ADLS

HDInsight Spark (https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-autoscale-clusters) keyword here is in-memory processing

Azure Synapse Analytics

upvoted 1 times

eurekamike 4 years ago

Databricks has in-memory processing

upvoted 1 times

maciejt 4 years, 1 month ago

Why not cosmos for analytical datastore?

upvoted 1 times

NamishBansal 4 years, 1 month ago

For third one Synapse will work but why will Cosmos not work?

upvoted 2 times

ssanka 4 years, 2 months ago

I think answer should be cosmos db for 3rd one. Azure synapse doesn't support columnar storage right ?

upvoted 2 times

meswapnilspal 4 years, 2 months ago

it does. https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-what-is

upvoted 2 times

Wendy_DK 4 years, 2 months ago

Batch processing should be Azure Databricks

upvoted 2 times

Manoel_Benicio 4 years, 2 months ago

that´s correct so the answers would be: Azure DataBricks, Azure DataLake and ASA (Azure Synapse)

upvoted 4 times

LG5 4 years, 2 months ago

Batch processing should be Azure Databricks right?

upvoted 5 times

eliabsbueno 4 years, 2 months ago

Yes! HDInsight does not support autotermination natively

upvoted 2 times

DRAG DROP -

Your company has an on-premises Microsoft SQL Server instance.

The data engineering team plans to implement a process that copies data from the SQL Server instance to Azure Blob storage once a day. The process must orchestrate and manage the data lifecycle.

You need to create Azure Data Factory to connect to the SQL Server instance.

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.

Select and Place:

Actions	Answer Area
From the on-premises network, install and configure a self-hosted integration runtime.	
Configure a linked service to connect to the SQL Server instance.	
Create an Azure Data Factory.	
From the SQL Server, create a database master key.	
From the SQL Server, backup the database and then copy the database to Azure Blob storage.	

Suggested Answer:

Actions	Answer Area
	Create an Azure Data Factory.
	From the on-premises network, install and configure a self-hosted integration runtime.
	Configure a linked service to connect to the SQL Server instance.
From the SQL Server, create a database master key.	
From the SQL Server, backup the database and then copy the database to Azure Blob storage.	

Step 1: Create an Azure Data Factory

You need to create a data factory and start the Data Factory UI to create a pipeline in the data factory.

Step 2: From the on-premises network, install and configure a self-hosted runtime.

To use copy data from a SQL Server database that isn't publicly accessible, you need to set up a self-hosted integration runtime.

Step 3: Configure a linked service to connect to the SQL Server instance.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/connector-sql-server https://www.mssqltips.com/sqlservertip/5812/connect-to-onpremises-data-in-azure-data-factory-with-the-selfhosted-integration-runtime--part-1/

☐ 👤 **memo43** 4 years, 1 month ago

answer with sequences is CORRECT

upvoted 4 times

☐ 👤 **maciejt** 4 years, 1 month ago

linked service ust be created after crating data factory, but SH runtime is independent, so there are many possible sequences here.

upvoted 4 times

　☐ 👤 **AZ20** 4 years ago

　　step1 and step2 mentioned in are independent of each other , only step 3 is depedent on both of them

　　upvoted 1 times

☐ 👤 **Psycho** 4 years, 2 months ago

I've worked before with integration runtime and I've followed the steps in this order: Create Data Factory, Configure a link service, install and configure an Integration runtime

upvoted 3 times

cadio30 4 years, 1 month ago

Cannot setup link service to on-premise SQL Server if the self-hosted integration runtime is not configured.

upvoted 4 times

I've worked before with integration runtime and I've followed the steps in this order: Create Data Factory, Configure a link service, install and configure an Integration runtime

upvoted 3 times

cadio30 4 years, 1 month ago

Cannot setup link service to on-premise SQL Server if the self-hosted integration runtime is not configured.

upvoted 4 times

A company runs Microsoft SQL Server in an on-premises virtual machine (VM).

You must migrate the database to Azure SQL Database. You synchronize users from Active Directory to Azure Active Directory (Azure AD).

You need to configure Azure SQL Database to use an Azure AD user as administrator.

What should you configure?

A. For each Azure SQL Database, set the Access Control to administrator.

B. For each Azure SQL Database server, set the Active Directory to administrator.

C. For each Azure SQL Database, set the Active Directory administrator role.

D. For each Azure SQL Database server, set the Access Control to administrator.

Suggested Answer: *C*

There are two administrative accounts (Server admin and Active Directory admin) that act as administrators.

One Azure Active Directory account, either an individual or security group account, can also be configured as an administrator. It is optional to configure an Azure

AD administrator, but an Azure AD administrator must be configured if you want to use Azure AD accounts to connect to SQL Database.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-manage-logins

---

👤 **STH** `Highly Voted 👍` 5 years, 7 months ago

Admin user is defined at server level.

Answer is B

upvoted 72 times

> 👤 **deepz8** 4 years, 3 months ago
>
> +1 for this answer
>
> upvoted 2 times
>
> > 👤 **contactodonuno** 4 years, 3 months ago
> >
> > +1 for this answer
> >
> > upvoted 2 times
>
> 👤 **cadio30** 4 years, 1 month ago
>
> Appropriate answer
>
> upvoted 3 times

👤 **avestabrzn** `Highly Voted 👍` 5 years, 3 months ago

All resources --> Active Directory Admin --> Set Admin

Answer is B

upvoted 9 times

👤 **massnonn** `Most Recent ⊘` 3 years, 7 months ago

the correct answer is B 100%

upvoted 1 times

👤 **marc88** 3 years, 10 months ago

The question mentions migrating just 'one' database. Do we need to grant admin access to a DB server (group of DBs) ? Doesn't go with the minimal access practice. Don't agree with B.

upvoted 1 times

👤 **tejasjoshi** 4 years ago

correct answer is B. I did a practical implementation for the same

upvoted 1 times

👤 **Simon2021** 4 years ago

Just created a new Azure SQL server and a new Azure sql DB. Answer it B.

upvoted 1 times

👤 **Qrm_1972** 4 years, 1 month ago

Azure AD admin with a server in SQL Database
Each server in Azure (which hosts SQL Database or Azure Synapse) starts with a single server administrator account that is the administrator of the entire server. Create a second administrator account as an Azure AD account. This principal is created as a contained database user in the master database of the server. Administrator accounts are members of the db_owner role in every user database, and enter each user database as the dbo user.

https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-configure?tabs=azure-powershell

The correct answer is : B
  upvoted 2 times

⊟ 👤 **sandeep1111** 4 years, 2 months ago
Correct answer is B
  upvoted 1 times

⊟ 👤 **get2niz** 4 years, 6 months ago
C is Right , Under SQL Server -> select SQL database -> Select ACTIVE DIRETORY ADMIN -> Set Admin -> Select User or Azure AD group -> Save them as Admin role (Az AD Admin of the SQL database)
https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-configure?tabs=azure-powershell#provision-azure-ad-admin-sql-database
  upvoted 1 times

  ⊟ 👤 **AyeshJr** 4 years, 5 months ago
  my friend, you have accessed the sql server and not sql db. Answer is B
    upvoted 3 times

⊟ 👤 **sdas1** 4 years, 6 months ago
As per below link, the answer should be B.
https://docs.microsoft.com/en-us/azure/azure-sql/database/logins-create-manage
  upvoted 1 times

⊟ 👤 **BRW** 4 years, 6 months ago
EDIT. C is correct.
As mentioned in the answer there are two accounts, with admin permissions to a database. Server Admin Login and Active Directory Admin. Server Admin is I guess nothing to do with AD. Go to "Properties" tab for Azure Database for reference. It is mentioned there
  upvoted 1 times

  ⊟ 👤 **AyeshJr** 4 years, 5 months ago
  Well, if you go to the link and go to the properties as you've mentioned --> the option is dimmed and you cannot edit it. you have to do it from the Azure sql server and not Azure sql db
    upvoted 1 times

⊟ 👤 **BRW** 4 years, 6 months ago
C is correct.
As mentioned in the answer there are two accounts, with admin permissions to a database. Server Admin Login and Active Directory Admin. Server Admin is I guess nothing to do with AD. Go to "Performance" tab for Azure Database for reference. It is mentioned there
  upvoted 1 times

⊟ 👤 **syu31svc** 4 years, 7 months ago
From link https://docs.microsoft.com/en-us/azure/azure-sql/database/logical-servers:
When you create a server, you provide a server login account and password that has administrative rights to the master database on that server and all databases created on that server. This initial account is a SQL login account. Azure SQL Database and Synapse Analytics support SQL authentication and Azure Active Directory Authentication for authentication.
Answer is B
  upvoted 2 times

  ⊟ 👤 **dumpsm42** 4 years, 6 months ago
  B - the pic bellow says it all: server level (This article shows you how to create and populate an Azure Active Directory (Azure AD) instance, and then use Azure AD with Azure SQL Database, Azure SQL Managed Instance, and Azure Synapse Analytics)
  https://docs.microsoft.com/pt-pt/azure/azure-sql/database/media/authentication-aad-configure/set-admin.png
    upvoted 1 times

⊟ 👤 **hart232** 4 years, 9 months ago
C - Practically it makes sense to configure previliges at database level than SQL server level.

upvoted 1 times

    ☐ 👤 **hart232** 4 years, 9 months ago

      eh...Sorry, I cant delete the above. There is no option to configure the AD admin at database level. You can do it only at server level. So answer is B

      upvoted 5 times

☐ 👤 **kimhok** 4 years, 9 months ago

Using AD as administrator for SQL Databases should be on a Database level. (Check diagaram in this page) ---> https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-overview

upvoted 1 times

☐ 👤 **Varma_Saraswathula** 4 years, 10 months ago

Answer is C

An Azure AD administrator must be configured if you want to use Azure AD accounts to connect to SQL Database, SQL Managed Instance, or Azure Synapse.

Refer https://docs.microsoft.com/en-us/azure/azure-sql/database/logins-create-manage

upvoted 2 times

    ☐ 👤 **M0e** 4 years, 9 months ago

      To configure Azure AD admin account to a Database, it references to this link: https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-configure?tabs=azure-powershell ... in the screenshots you see that configuration has to be set on the server level. If you check the portal you can see that there is NO option to configure the AD admin at the database level. So the answer is B, but agree that the wording is incorrect.

      upvoted 2 times

☐ 👤 **r8d1** 4 years, 10 months ago

B: Admin is at server level.

not C: as each server in Azure (which hosts SQL Database) starts with a single server administrator account that is the administrator of the entire server.

i.e. Admin at server level.

not D: as role-based access control (RBAC) applies only to the Azure portal and isn't propagated to SQL Database.

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 an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

☞ The first two prefix characters must be exposed.

☞ The last four suffix characters must be exposed.

☞ All other characters must be masked.

Solution: You implement data masking and use a credit card function mask.

Does this meet the goal?

   A. Yes

   B. No

Suggested Answer: *B*

Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

Reference:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

---

👤 **memo43** 4 years, 1 month ago

answer is NO

  upvoted 1 times

👤 **rajeshdp200** 5 years ago

Email Masking:

Masking method, which exposes the first letter and replaces the domain with XXX.com using a constant string prefix in the form of an email address.

aXX@XXXX.com

Reference: https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

  upvoted 3 times

    👤 **naveen_nagavelly** 4 years, 2 months ago

    In the Question we have column "Customer_ID", I guess it is not related to Email

    also, the first two characters to show.

      upvoted 1 times

👤 **llt** 5 years ago

Either Q22 or Q23 must have an option of 'last for SUFFIX characters must be exposed'. Now both options are the same. In real exam they have 'SUFFIX'. Anyhow, this option also makes the answer 'NO'.

  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 an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

☞ The first two prefix characters must be exposed.

The last four suffix characters must be exposed.

▪

☞ All other characters must be masked.

Solution: You implement data masking and use an email function mask.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

---

☐ 👤 **ZodiaC** 4 years, 1 month ago

aXX@XXXX.com = NO, is correct!

  upvoted 1 times

☐ 👤 **vicnovice** 4 years, 2 months ago

this is the right answer

  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 an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

☞ The first two prefix characters must be exposed.

☞ The last four suffix characters must be exposed.

☞ All other characters must be masked.

Solution: You implement data masking and use a random number function mask.

Does this meet the goal?

    A. Yes

    B. No

> **Suggested Answer:** *B*
>
> Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.
>
> References:
>
> https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

---

☐ 👤 **JESUSBB** 4 years, 4 months ago

Answer: B. No.

Random number Masking method, which generates a random number according to the selected boundaries and actual data types. If the designated boundaries are equal, then the masking function is a constant number.

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

  upvoted 3 times

☐ 👤 **syu31svc** 4 years, 7 months ago

Answer to Qn 22-24 is Custom text (Masking method, which exposes the first and last characters and adds a custom padding string in the middle.)

From https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

  upvoted 1 times

☐ 👤 **Johnnien** 4 years, 7 months ago

Correct answer: You implement data masking and use a custom text mask.

  upvoted 1 times

☐ 👤 **rajeshdp200** 5 years ago

Reference: https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

  upvoted 1 times

DRAG DROP -

You are responsible for providing access to an Azure Data Lake Storage Gen2 account.

Your user account has contributor access to the storage account, and you have the application ID and access key.

You plan to use PolyBase to load data into an enterprise data warehouse in Azure Synapse Analytics.

You need to configure PolyBase to connect the data warehouse to the storage account.

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

Select and Place:

Components

| a database encryption key |
| an asymmetric key |
| an external data source |
| an external file format |
| a database scoped credential |

Answer Area

(empty boxes)

Suggested Answer:

Components

| a database encryption key |
| an asymmetric key |

Answer Area

| a database scoped credential |
| an external data source |
| an external file format |

Step 1: a database scoped credential

To access your Data Lake Storage account, you will need to create a Database Master Key to encrypt your credential secret used in the next step. You then create a database scoped credential.

Step 2: an external data source -

Create the external data source. Use the CREATE EXTERNAL DATA SOURCE command to store the location of the data. Provide the credential created in the previous step.

Step 3: an external file format -

Configure data format: To import the data from Data Lake Storage, you need to specify the External File Format. This object defines how the files are written in

Data Lake Storage.

☐ 👤 **Internet_User** `Highly Voted 👍` 4 years, 1 month ago

This answer is correct. A database scoped credential contains the authentication info to connect to external resources. Before creating a database scoped credential, the database must have a master key. https://docs.microsoft.com/en-us/sql/t-sql/statements/create-database-scoped-credential-transact-sql

upvoted 10 times

☐ 👤 **Internet_User** `Highly Voted 👍` 4 years, 1 month ago

This answer is correct. A database scoped credential contains the authentication info to connect to external resources. Before creating a database scoped credential, the database must have a master key. https://docs.microsoft.com/en-us/sql/t-sql/statements/create-database-scoped-credential-transact-sql

upvoted 10 times

You plan to create a dimension table in Azure Synapse Analytics that will be less than 1 GB.

You need to create the table to meet the following requirements:

☞ Provide the fastest query time.

☞ Minimize data movement during queries.

Which type of table should you use?

    A. hash distributed

    B. heap

    C. replicated

    D. round-robin

Suggested Answer: *D*

Usually common dimension tables or tables that doesn't distribute evenly are good candidates for round-robin distributed table.

Note: Dimension tables or other lookup tables in a schema can usually be stored as round-robin tables. Usually these tables connect to more than one fact tables and optimizing for one join may not be the best idea. Also usually dimension tables are smaller which can leave some distributions empty when hash distributed.

Round-robin by definition guarantees a uniform data distribution.

Reference:

https://blogs.msdn.microsoft.com/sqlcat/2015/08/11/choosing-hash-distributed-table-vs-round-robin-distributed-table-in-azure-sql-dw-service/

Community vote distribution

C (100%)

👤 **sandeep1111** `Highly Voted 👍` 4 years, 2 months ago

Answer is Replicated table here.

upvoted 56 times

👤 **oldpony** `Most Recent ⊘` 3 years ago

`Selected Answer: C`

dimension tables are normally replicated

Fact tables are normally hash due to large size of data

upvoted 1 times

👤 **massnonn** 3 years, 7 months ago

Absolutely wrong the round robin use all distrbution, the correct answer is replicated

upvoted 2 times

👤 **eurekamike** 4 years ago

dimension tables are normally replicated. also given that it is less than 1GB is another reason to use replicated.

staging is round-robin unless it's too big, then use hash.

fact tables are hash distributed.

upvoted 3 times

👤 **jayeshstudies** 4 years ago

minimize data movement --> no way it can be round robin

has to be replicated

upvoted 2 times

👤 **VinayakT** 4 years ago

Replicated should be correct answer

upvoted 3 times

👤 **Simon2021** 4 years ago

Answer is Replicated

upvoted 4 times

👤 **tej24** 4 years ago

Data movement is reduced with Replicated tables because a full copy of data is stored on each Compute node. As a result, queries that required data movement steps to complete now run faster with Replicated tables.
Answer is Replicated table
https://azure.microsoft.com/en-in/blog/replicated-tables-now-in-preview-for-azure-sql-data-warehouse/
upvoted 3 times

☐ 👤 **LordSnoek** 4 years ago
Replicated is the right choice
upvoted 2 times

☐ 👤 **AZ20** 4 years ago
I think answer should be replicated as size is already given 1GB (not that big) , plus no data movement required as present on all nodes
upvoted 1 times

☐ 👤 **AZ20** 4 years ago
to support it more .. from MS guide -
What is a replicated table?

A replicated table has a full copy of the table accessible on each Compute node. Replicating a table removes the need to transfer data among Compute nodes before a join or aggregation. Since the table has multiple copies, replicated tables work best when the table size is less than 2 GB compressed. 2 GB is not a hard limit. If the data is static and does not change, you can replicate larger tables.

My addition -
In dimension data doesn't change frequently
upvoted 1 times

☐ 👤 **Qrm_1972** 4 years, 2 months ago
The correct is Replicated ✅ Small-dimension tables in a star schema with less than 2GB of storage after compression (~5x compression).
upvoted 2 times

☐ 👤 **gkozla** 4 years, 2 months ago
Replicated table: A replicated table has a full copy of the table available on every Compute node. Queries run fast on replicated tables because joins on replicated tables don't require data movement. Replication requires extra storage, though, and isn't practical for large tables.
upvoted 2 times

☐ 👤 **JohnCrawford** 4 years, 2 months ago
See this link supporting my statement. https://docs.microsoft.com/en-gb/learn/modules/design-azure-sql-data-warehouse/6-table-geometries
upvoted 3 times

☐ 👤 **JohnCrawford** 4 years, 2 months ago
The correct answer is REPLICATED.

Fact tables should almost always use hash distribution, dimension tables almost always replicated and staging tables almost always round-robin. If the dimension table is large (2 GB+) and/or used in joins with the fact table on the partition key then it may make sense to also use hash distribution for that particular dimension table.
upvoted 3 times

☐ 👤 **SuperAlex** 4 years, 2 months ago
why not replicated table ?
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/design-guidance-for-replicated-tables
upvoted 2 times

☐ 👤 **GoldenStateWarriors** 4 years, 2 months ago
In my opinion, the correct answer is Hash-distributed based on the link below:
"A hash distributed table distributes rows based on the value in the distribution column. A hash distributed table is designed to achieve high performance for queries on large tables. There are several factors to consider when choosing a distribution column."
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-overview
upvoted 3 times

☐ 👤 **maciejt** 4 years, 1 month ago
This is not a large table, it;s very small. Hash distributed is good for large fact tables with evenly distributed column that is used in joins.
upvoted 1 times

You have an enterprise data warehouse in Azure Synapse Analytics.

Using PolyBase, you create an external table named [Ext].[Items] to query Parquet files stored in Azure Data Lake Storage Gen2 without importing the data to the data warehouse.

The external table has three columns.

You discover that the Parquet files have a fourth column named ItemID.

Which command should you run to add the ItemID column to the external table?

```
A.
    DROP EXTERNAL TABLE [Ext].[Items]
    CREATE EXTERNAL TABLE [Ext].[Items]
    ( [ItemID] [int] NULL,
      [ItemName] nvarchar(50) NULL,
      [ItemType] nvarchar(20) NULL,
      [ItemDescription] nvarchar(250))
    WITH
    (
        LOCATION='/Items/',
        DATA_SOURCE = AzureDataLakeStore,
        FILE_FORMAT = PARQUET,
        REJECT_TYPE = VALUE,
        REJECT_VALUE = 0
    );
B.
    ALTER TABLE [Ext].[Items]
    ADD [ItemID] int;
C.
    DROP EXTERNAL FILE FORMAT parquetfile1;
    CREATE EXTERNAL FILE FORMAT parquetfile1
    WITH (
        FORMAT_TYPE = PARQUET,
        DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
    );
D.
    ALTER EXTERNAL TABLE [Ext].[Items]
    ADD [ItemID] int;
```

A. Option A

B. Option B

C. Option C

D. Option D

Suggested Answer: *A*
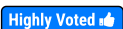
Incorrect Answers:

B, D: Only these Data Definition Language (DDL) statements are allowed on external tables:

☞ CREATE TABLE and DROP TABLE

☞ CREATE STATISTICS and DROP STATISTICS

☞ CREATE VIEW and DROP VIEW

Reference:

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-transact-sql

---

⊟ 👤 **sandeep1111** `Highly Voted 👍` 4 years, 2 months ago

Answer is correct

upvoted 8 times

⊟ 👤 **gaganmay26** `Most Recent ⊘` 3 years ago

Only DML queries are allowed on external tables

upvoted 1 times

⊟ 👤 **Ysandee** 4 years ago

Right Answer

upvoted 1 times

DRAG DROP -

You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics. SalesFact contains sales data from the past 36 months and has the following characteristics:

☞ Is partitioned by month

☞ Contains one billion rows

☞ Has clustered columnstore indexes

At the beginning of each month, you need to remove data from SalesFact that is older than 36 months as quickly as possible.

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

Select and Place:

Actions

Create an empty table named SalesFact_Work that has the same schema as SalesFact.
Drop the SalesFact_Work table.
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).
Truncate the partition containing the stale data
Switch the partition containing the stale data from SalesFact to SalesFact_Work.
Execute a DELETE statement where the value in the Date column is more than 36 months ago.

Answer Area

Suggested Answer:

Actions

Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).
Truncate the partition containing the stale data
Execute a DELETE statement where the value in the Date column is more than 36 months ago.

Answer Area

Create an empty table named SalesFact_Work that has the same schema as SalesFact.
Switch the partition containing the stale data from SalesFact to SalesFact_Work.
Drop the SalesFact_Work table.

Step 1: Create an empty table named SalesFact_work that has the same schema as SalesFact.

Step 2: Switch the partition containing the stale data from SalesFact to SalesFact_Work.

SQL Data Warehouse supports partition splitting, merging, and switching. To switch partitions between two tables, you must ensure that the partitions align on their respective boundaries and that the table definitions match.

Loading data into partitions with partition switching is a convenient way stage new data in a table that is not visible to users the switch in the new data.

Step 3: Drop the SalesFact_Work table.

Reference:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-partition

---

⊟ 👤 **LongBao** `Highly Voted 👍` 4 years, 2 months ago

The answer is correct

upvoted 19 times

⊟ 👤 **Kostali** `Most Recent ⊘` 3 years, 11 months ago

1.- Copy the data to a new table by using CTAS.

2.- Switch partition...

3.- Drop new table

upvoted 2 times

AngelRio 4 years, 2 months ago

1.- Copy the data to a new table by using CTAS.

2.- Switch partition...

3.- Drop new table

upvoted 2 times

Mily94 4 years, 2 months ago

are you sure? Why copy data?

upvoted 4 times

rajneesharora 4 years, 1 month ago

No need to copy data, when finally you have to drop the table, additionally, copying will take more time is not needed and is less efficient

upvoted 5 times

memo43 4 years, 1 month ago

No! you dont have to spend time for copying. to switch is more easy and take less time...

upvoted 1 times

You plan to implement an Azure Cosmos DB database that will write 100,000,000 JSON records every 24 hours. The database will be replicated to three regions.

Only one region will be writable.

You need to select a consistency level for the database to meet the following requirements:

☞ Guarantee monotonic reads and writes within a session.

☞ Provide the fastest throughput.

☞ Provide the lowest latency.

Which consistency level should you select?
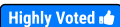
    A. Strong

    B. Bounded Staleness

    C. Eventual

    D. Session

    E. Consistent Prefix

Suggested Answer: *D*

Session: Within a single client session reads are guaranteed to honor the consistent-prefix (assuming a single ⅄€writer⅄€ session), monotonic reads, monotonic writes, read-your-writes, and write-follows-reads guarantees. Clients outside of the session performing writes will see eventual consistency.

Reference:

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

---

⊟ 👤 **ivanbtod** `Highly Voted 👍` 5 years, 2 months ago

I guess Andrea25 could have got this question which i found in another dump where there are three answers but the question is slightly different

You plan to deploy an Azure Cosmos DB database that supports multi-master replication.

You need to select a consistency level for the database to meet the following requirements:

– Provide a recovery point objective (RPO) of less than 15 minutes.

– Provide a recovery time objective (RTO) of zero minutes.

What are three possible consistency levels that you can select? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A . Strong

B . Bounded Staleness

C . Eventual

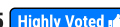D . Session

E . Consistent Prefix

  upvoted 34 times

  ⊟ 👤 **mickeyisacat615** 4 years, 5 months ago

    I saw this question in other dump too. CDE would be the three options.

    To this question, Session is the only option. D is the answer.

      upvoted 2 times

⊟ 👤 **Andrea25** `Highly Voted 👍` 5 years, 2 months ago

I had this question in my exam. The question ask to choice three of those:

- Eventual;

- Session;

- Consistent Prefix

  upvoted 18 times

  ⊟ 👤 **ivanbtod** 5 years, 2 months ago

Andrea25 Were all the questions on the exam from this dump?

upvoted 1 times

⊟ 👤 llt 5 years ago

No, the actual exam DOES NOT require to provide 3 answers, I know this question. The keyword is 'monotonic', so the correct answer is SESSION.

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

upvoted 12 times

⊟ 👤 Qrm_1972 **Most Recent** ⊙ 4 years, 2 months ago

The correct answer is D: Session ------>>> Session consistency is the most widely used consistency level for both single region as well as globally distributed applications. It provides write latencies, availability, and read throughput comparable to that of eventual consistency but also provides the consistency guarantees that suit the needs of applications written to operate in the context of a user

upvoted 3 times

⊟ 👤 syu31svc 4 years, 7 months ago

Link supports D as the answer

upvoted 2 times

⊟ 👤 zenomas 5 years, 3 months ago

Keyword here is "Guarantee monotonic reads and writes within a session."

upvoted 14 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 an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

☞ The first two prefix characters must be exposed.

☞ The last four suffix characters must be exposed.

☞ All other characters must be masked.

Solution: You implement data masking and use a custom text mask.

Does this meet the goal?

   A. Yes

   B. No

Suggested Answer: *A*

We must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

Reference:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

---

👤 **memo43** 4 years, 1 month ago

answer is CORRECT.

you should mask in custom... others are not appropriate

upvoted 3 times

Question #31 — Topic 1

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 an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is less than 1 MB.

Does this meet the goal?

   A. Yes

   B. No

> **Suggested Answer:** *B*
> Instead convert the files to compressed delimited text files.
> Reference:
> https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

---

JohnCrawford `Highly Voted` 4 years, 2 months ago

The only thing I've found mentioning row size when doing the loads relates to loads using Polybase. Then the row size needs to less than 1 MB.
https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse

upvoted 9 times

---

hsetin `Most Recent` 3 years, 9 months ago

REcommended approach as per this: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/data-loading-best-practices

upvoted 1 times

---

hsetin 3 years, 9 months ago

it is not just about the size, it is the approach. the recommended approach is about zipping the file.

upvoted 1 times

Simon2021 4 years ago

Answer is Yes. File size should be less than 1 MB.

upvoted 4 times

> **ricardoveedgee** 4 years ago
>
> I concur. The correct answer is 'yes' since the no size limitation update (PolyBase) applies to only SQL Server 2019+. To date, Azure Synapse Analytics still has a cap at 1 MB.
>
> upvoted 1 times

---

dumpi 4 years ago

correct answer is B

upvoted 1 times

JaNieWiem 4 years, 2 months ago

"Yes."

upvoted 4 times

Nevia 4 years, 2 months ago

Row size should be still maximum 1 MB. So answer is YES. Also according to latest updates https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-versioned-feature-summary?view=sql-server-ver15

upvoted 2 times

MayankSh 4 years, 2 months ago

Earlier, the answer to this question was "Yes", but now it shows as "No". Can someone confirm what is the correct answer ?

upvoted 4 times

> **dlamine9** 4 years, 1 month ago
>
> so what is the correct answer

⊟ 👤 **Anand_96** 4 years ago

Answer is "yes"

⊟ 👤 **Anand_96** 4 years ago

Answer is "yes"

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 an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an Azure SQL data warehouse.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is more than 1 MB.

Does this meet the goal?

   A. Yes

   B. No

Suggested Answer: *B*

Instead modify the files to ensure that each row is less than 1 MB.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

---

👤 **clownfishman** `Highly Voted 👍` 4 years, 9 months ago

one word, huge difference (less/more)

upvoted 6 times

👤 **dduque10** `Most Recent ⊘` 3 years, 9 months ago

This answer contradicts the last one

upvoted 1 times

👤 **memo43** 4 years, 1 month ago

the answer is correct... NO

upvoted 4 times

👤 **BungyTex** 4 years, 10 months ago

The answer is A. The description is literally what is said in A.

upvoted 1 times

   👤 **BungyTex** 4 years, 10 months ago

   Ignore my comment, the answer is B. I misread.

   upvoted 6 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 an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You copy the files to a table that has a columnstore index.

Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

Instead convert the files to compressed delimited text files.

Reference:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

---

👤 **memo43** 4 years, 1 month ago

answer is CORRECT!

upvoted 1 times

You plan to deploy an Azure Cosmos DB database that supports multi-master replication.

You need to select a consistency level for the database to meet the following requirements:

☞ Provide a recovery point objective (RPO) of less than 15 minutes.

☞ Provide a recovery time objective (RTO) of zero minutes.

What are three possible consistency levels that you can select? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

    A. Strong

    B. Bounded Staleness

    C. Eventual

    D. Session

    E. Consistent Prefix

Suggested Answer: *CDE*

Region(s)	Replication mode	Consistency level	RPO	RTO
1	Single or Multi-Master	Any Consistency Level	< 240 Minutes	<1 Week
>1	Single Master	Session, Consistent Prefix, Eventual	< 15 minutes	< 15 minutes
>1	Single Master	Bounded Staleness	K & T	< 15 minutes
>1	Single Master	Strong	0	< 15 minutes
>1	Multi-Master	Session, Consistent Prefix, Eventual	< 15 minutes	0
>1	Multi-Master	Bounded Staleness	K & T	0

Reference:

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels-choosing

memo43 4 years, 1 month ago

%100 CORRECT

upvoted 4 times

SIMULATION -
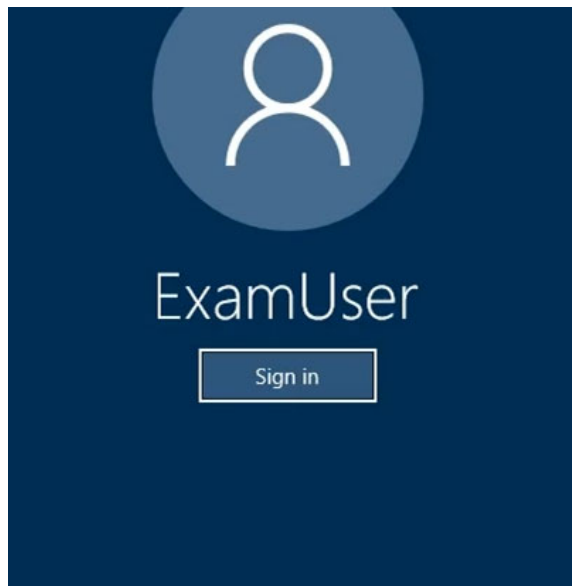
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
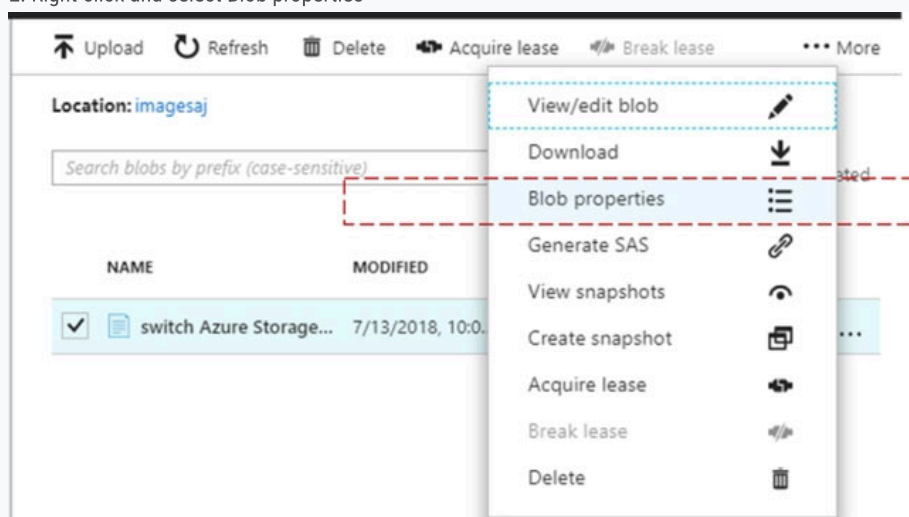The following information is for technical support purposes only:

Lab Instance: 10277521 -
You need to ensure that you can recover any blob data from an Azure Storage account named storage 10277521 up to 30 days after the data is deleted.
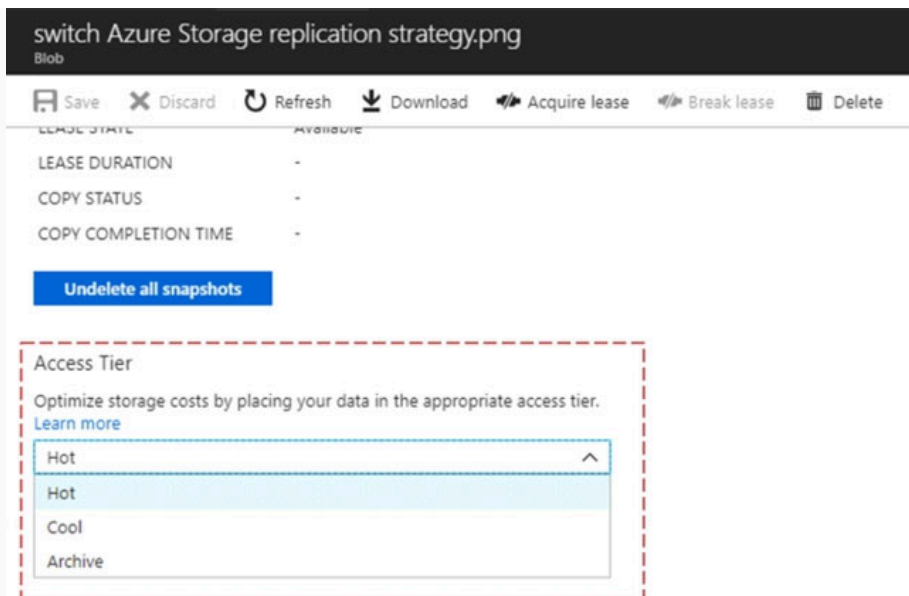To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*

1. Open Azure Portal and open the Azure Blob storage account named storage10277521.
2. Right-click and select Blob properties

3. From the properties window, change the access tier for the blob to Cool.

switch Azure Storage replication strategy.png
Blob

Save	Discard	Refresh	Download	Acquire lease	Break lease	Delete

LEASE STATE   Available
LEASE DURATION   -
COPY STATUS   -
COPY COMPLETION TIME   -

Undelete all snapshots

Access Tier

Optimize storage costs by placing your data in the appropriate access tier.
Learn more

Hot   ∧
Hot
Cool
Archive

Note: The cool access tier has lower storage costs and higher access costs compared to hot storage. This tier is intended for data that will remain in the cool tier for at least 30 days.
Reference:
https://dailydotnettips.com/how-to-update-access-tier-in-azure-storage-blob-level/

LongBao `Highly Voted` 4 years, 2 months ago
In the Azure portal, navigate to your storage account.
Locate the Data Protection option under Blob service.
Set the Blob soft delete property to Enabled.
Under Retention policies, specify how long soft-deleted blobs are retained by Azure Storage.
Save your changes.
Ref link https://docs.microsoft.com/en-us/azure/storage/blobs/soft-delete-blob-enable?tabs=azure-portal
upvoted 30 times

> **memo43** 4 years, 1 month ago
> you should set soft delete
> upvoted 3 times

menatallah `Most Recent ⊙` 4 years, 1 month ago
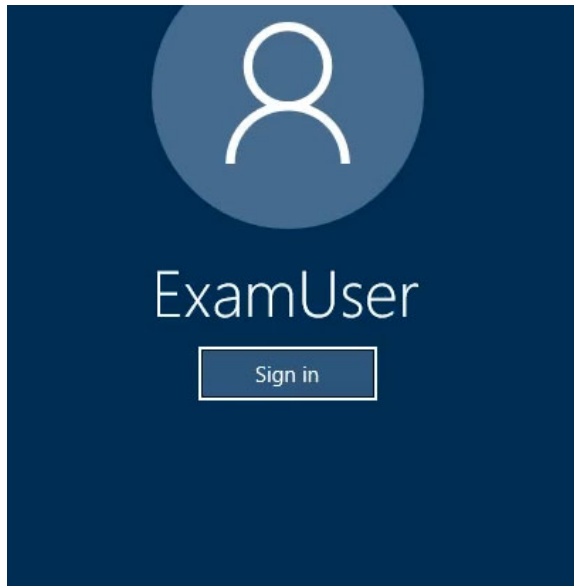Soft Delete is the option needed here , not the answer provided
upvoted 2 times

Maky2365 4 years, 1 month ago
wrong answer. the question is asking for recovery of deleted data and that can be achieve by setting soft delete property.
upvoted 4 times

SIMULATION -

Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -

The following information is for technical support purposes only:

Lab Instance: 10277521 -

You need to replicate db1 to a new Azure SQL server named REPL10277521 in the Central Canada region.

To complete this task, sign in to the Azure portal.

NOTE: This task might take several minutes to complete. You can perform other tasks while the task completes or ends this section of the exam.

To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*

1. In the Azure portal, browse to the database that you want to set up for geo-replication.

2. On the SQL database page, select geo-replication, and then select the region to create the secondary database.
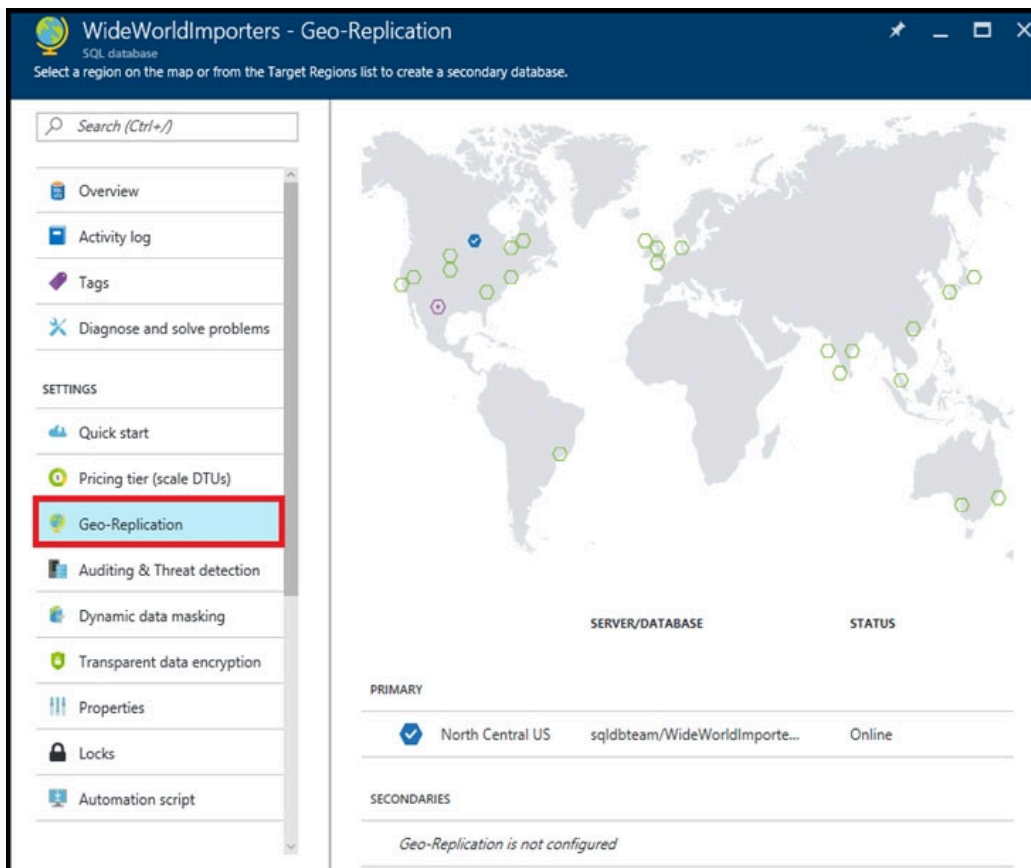
3. Select or configure the server and for the secondary database.

Region: Central Canada -

Target server: REPL10277521 -
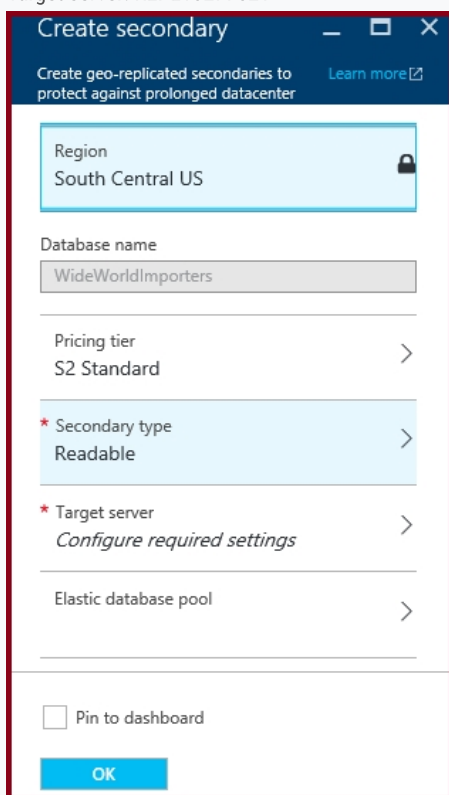
4. Click Create to add the secondary.

5. The secondary database is created and the seeding process begins.
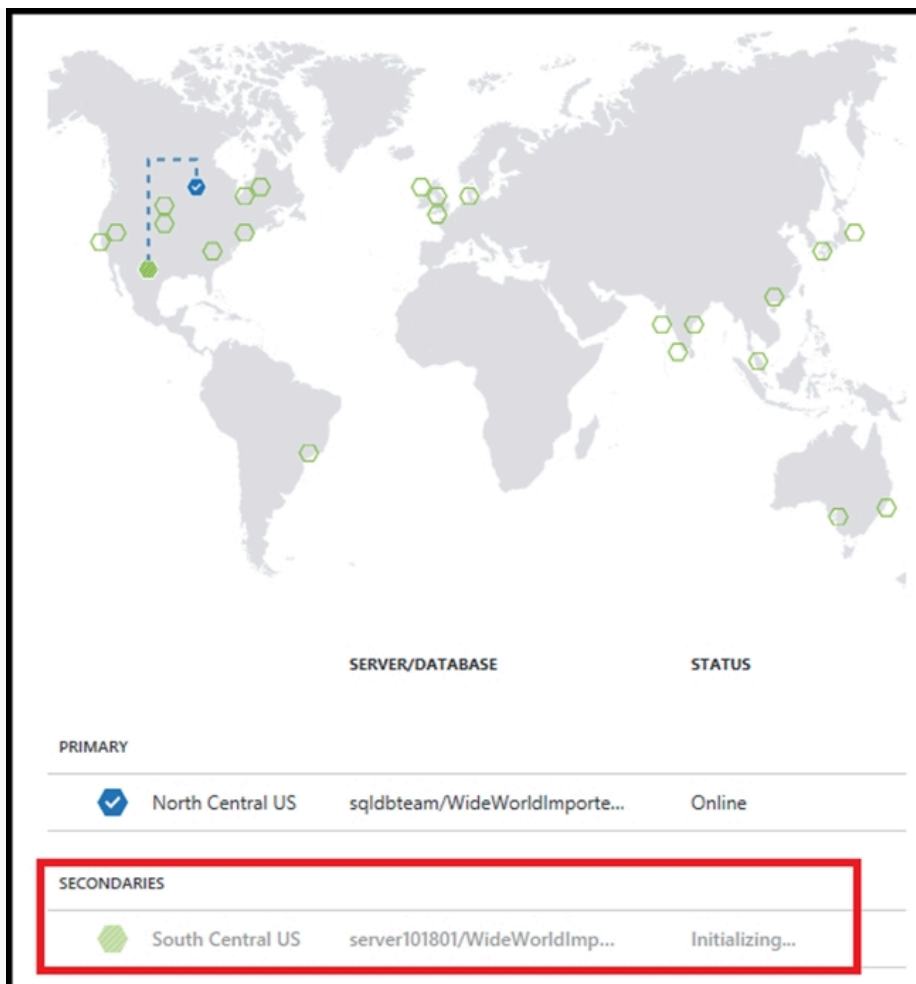
6. When the seeding process is complete, the secondary database displays its status.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-active-geo-replication-portal

---

☐ 👤 **hsetin** 3 years, 9 months ago

can i expect these in dP 203?

upvoted 3 times

☐ 👤 **dumpi** 4 years ago

is this type of question for now as well?

upvoted 2 times

☐ 👤 **meswapnilspal** 4 years, 2 months ago

In the solution, it is showing, South Central US, shouldn't it be Canada Central ?

upvoted 3 times

☐ 👤 **memo43** 4 years, 1 month ago

Yes, the showing is wrong. but the explanation is TRUE

upvoted 1 times

SIMULATION -
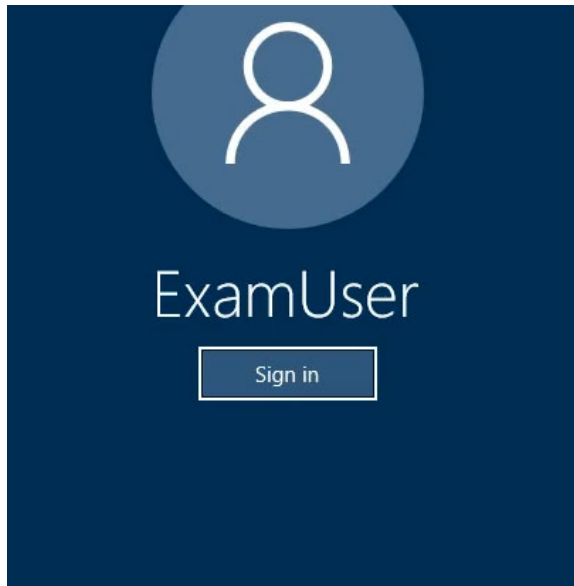
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10277521 -
You plan to query db3 to retrieve a list of sales customers. The query will retrieve several columns that include the email address of each sales customer.
You need to modify db3 to ensure that a portion of the email addresses is hidden in the query results.
To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*

1. Launch the Azure portal.

2. Navigate to the settings page of the database db3 that includes the sensitive data you want to mask.

3. Click the Dynamic Data Masking tile that launches the Dynamic Data Masking configuration page.

Note: Alternatively, you can scroll down to the Operations section and click Dynamic Data Masking.

4. In the Dynamic Data Masking configuration page, you may see some database columns that the recommendations engine has flagged for masking.

5. Click ADD MASK for the EmailAddress column

6. Click Save in the data masking rule page to update the set of masking rules in the dynamic data masking policy.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started-portal

HrishiD 4 years, 1 month ago

Shouldn't we select masking field format as 'Email'?

upvoted 1 times

memo43 4 years, 1 month ago

it is not mentioned in question. so in step 5 you can select format as; email, custom or default... it is not important in this question.

upvoted 1 times

HrishiD 4 years, 1 month ago

Shouldn't we select masking field format as 'Email'?

upvoted 1 times

memo43 4 years, 1 month ago

it is not mentioned in question. so in step 5 you can select format as; email, custom or default... it is not important in this question.

upvoted 1 times

HOTSPOT -

You have an enterprise data warehouse in Azure Synapse Analytics that contains a table named FactOnlineSales. The table contains data from the start of 2009 to the end of 2012.

You need to improve the performance of queries against FactOnlineSales by using table partitions. The solution must meet the following requirements:

☞ Create four partitions based on the order date.

☞ Ensure that each partition contains all the orders placed during a given calendar year.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
CREATE TABLE [dbo].FactOnlineSales
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime]    NOT NULL,
[StoreKey] [int]         NOT NULL,
[ProductKey] [int]        NOT NULL,
[CustomerKey] [int]       NOT NULL,
[SalesOrderNumber] [nvarchar](20) NOT NULL,
[SalesQuantity] [int]    NOT NULL,
[SalesAmount] [money]    NOT NULL,
[UnitPrice] [money]      NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE
```

| ▼ | FOR VALUES |
| RIGHT |
| LEFT |

(

| ▼ |) |
| 20090101,20121231 |
| 20100101,20110101,20120101 |
| 20090101,20100101,20110101,20120101 |

Answer Area

Suggested Answer:

```
CREATE TABLE [dbo].FactOnlineSales
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime]    NOT NULL,
[StoreKey] [int]          NOT NULL,
[ProductKey] [int]        NOT NULL,
[CustomerKey] [int]       NOT NULL,
[SalesOrderNumber] [nvarchar](20) NOT NULL,
[SalesQuantity] [int]     NOT NULL,
[SalesAmount] [money]     NOT NULL,
[UnitPrice] [money]       NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE [____▼] FOR VALUES
```

▼
RIGHT
LEFT

```
(  [____▼]  )
```

▼
20090101,20121231
20100101,20110101,20120101
20090101,20100101,20110101,20120101

Box 1: LEFT -

RANGE LEFT: Specifies the boundary value belongs to the partition on the left (lower values). The default is LEFT.

Box 2: 20090101, 20100101, 20110101, 20120101

FOR VALUES (boundary_value [,...n]) specifies the boundary values for the partition. boundary_value is a constant expression.

Reference:

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse

---

👤 **Akesh** `Highly Voted 👍` 3 years, 10 months ago

Considerations

Create four partitions based on the order date.

Ensure that each partition contains all the orders placed during a given calendar year.

If we chose the left option here, data buckets will be like below

Partition 1: OrderDateKey <= 20100101

Partition 2: 20100101 < OrderDateKey <= 20110101

Partition 3: 20110101 < OrderDateKey <= 20120101

Partition 4: 20120101 < OrderDateKey

The problem here is "Ensure that each partition contains all the orders placed during a given calendar year." will Fail in this case. The OrderDateKey first day of the year ex- 20110101 will be in the partition which contains 2010 dates.

If we chose the Right option here, data buckets will be like below

Partition 1: OrderDateKey < 20100101

Partition 2: 20100101 <= OrderDateKey < 20110101

Partition 3: 20110101 <= OrderDateKey < 20120101

Partition 4: 20120101 <= OrderDateKey

This will "Ensure that each partition contains all the orders placed during a given calendar year."

So Option is Right with (20100101,20110101,20120101)
upvoted 21 times

☐ 👤 **satyamkishoresingh** 3 years, 9 months ago

correct , clearly explained at

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7#PartitionedTable
upvoted 2 times

☐ 👤 **medsimus** 3 years, 8 months ago

correct
upvoted 3 times

☐ 👤 **MayankSh** `Highly Voted 👍` 4 years, 2 months ago

The second option should be RIGHT, since we need to store calendar data in each partition.

Partition 1: col < 20090101
Partition 2: 20090101 <= col < 20100101
Partition 3: 20100101 <= col < 20110101
Partition 4: 20110101 <= col < 20120101
Partition 5: 20120101 <= col

upvoted 19 times

☐ 👤 **memo43** 4 years, 1 month ago

same question in page 38 with RIGHT answer!!!
upvoted 2 times

☐ 👤 **rajneesharora** 4 years, 2 months ago

This creates 5 partitions and not 4... Question says to create 4 partitions... If we have to create 4 Partitions, it would be "LEFT" and not "RIGHT"

Partition 1: 20090101 <= col < 20100101

Partition 2: 20100101 <= col < 20110101

Partition 3: 20110101 <= col < 20120101

Partition 4: 20120101 <= col

upvoted 12 times

☐ 👤 **Marcus1612** `Most Recent ⊙` 3 years, 8 months ago

Box 1 = Right, Box 2 = 20090101, 20100101, 20110101, 20120101

In a range left partition function, all boundary values are upper boundaries, they are the last values in the partitions. If you partition by year, you use December 31st. If you partition by month, you use January 31st, February 28th / 29th, March 31st, April 30th and so on. In a range right partition function, all boundary values are lower boundaries, they are the first values in the partitions. If you partition by year, you use January 1st. If you partition by month, you use January 1st, February 1st, March 1st, April 1st and so on:
upvoted 1 times

☐ 👤 **hoangton** 4 years ago

LEFT is the Correct answer. We will use RIGHT when the value range is (20081231,20091231,20101231,20111231)
upvoted 1 times

☐ 👤 **Steviyke** 4 years ago

The answer is: LEFT, and 20100101, 20110101, 20120101. This will (1. give 4 partitions (2. Ensure each partitions contain orders made within a calendar year (E.g. 1st Jan 2009 to 31st Dec 2009)

PARTION 1: <= 20100101 --------> all orders before AND on 31st Dec 2009

PARTION 2: Col > 1st Jan 2010 to 31st Dec 2010

PARTION 3: 1st Jan 2011 to 31st Dec 2011

PARTION 4: 1st Jan 2012 and above

This satisfies the 2 conditions of 4 partitions and each partition comprising orders in (within) a calendar year.
upvoted 3 times

☐ 👤 **AZ20** 4 years ago

Answer should be

Right

2010 , 2011, 2012

This ensures all the dates of one calendar year falls in same partiton and 3 boundaries gives 4 partitons

upvoted 8 times

□ 👤 **vrmei** 4 years ago

Answer should be RIGHT and 20100101,20110101,20120101

Range Left or Right, both are creating similar partition but there is difference in comparison

For example: in this scenario, When you use LEFT and 20100101,20110101,20120101
Partition will be, datecol<=20100101 , datecol>20100101 and datecol<=20110101 , datecol>20110101 and datecol<=20120101, datecol>20120101

But if you use range RIGHT and 20100101,20110101,20120101
Partition will be, datecol<20100101 , datecol>=20100101 and datecol<20110101 , datecol>=20110101 and datecol<20120101, datecol>=20120101

In this example, Range RIGHT will be suitable for calendar comparison Jan 1st to Dec 31st

Reference: https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql?view=sql-server-ver15

upvoted 7 times

□ 👤 **Chiranjib** 4 years, 1 month ago

The total number of partitions is always the total number of boundary values + 1. As the question talks about four partions, second option should be the right answer.. And as it using is lower value of the right parition, it should be right.

upvoted 3 times

□ 👤 **111222333** 4 years, 1 month ago

Check the link provided in the solution: https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7

RANGE LEFT: left boundary is exclusive, while the right boundary is inclusive.
RANGE RIGHT: left boundary is inclusive, while the right boundary is exclusive.

=> RANGE RIGHT is certainly the correct option because 01-01-xxxx should be included in the right partition together with all the dates from year xxxx. We just need to play with the date values.

Precisely, IMO the correct answer is RANGE RIGHT FOR VALUES (20100101, 20110101, 20120101).
- Partition 1: col < 20100101
- Partition 2: 20100101 <= col < 20110101
- Partition 3: 20110101 <= col < 20120101
- Partition 4: 20120101 <= col

If the empty partitions are discarded, it could also be RANGE RIGHT FOR VALUES (20090101, 20100101, 20110101, 20120101) because the partition <20090101 is empty. But I think this partition still exists even though it's empty, and the question specifies for 4 partitions. Please provide me evidence if I'm wrong about the existing empty partitions.

upvoted 11 times

□ 👤 **jitu803** 4 years, 1 month ago

This is the actual correct Answer.

upvoted 1 times

□ 👤 **Taekook** 4 years, 1 month ago

this is correct answer. since in question they have mentioned 4 partitions imples 3 boundaries and right

upvoted 2 times

□ 👤 **maciejt** 4 years, 1 month ago

Answer cannot be RIGHT, because boundary value is INCLUSIVE, so RIGHT would be a range of 2nd Jan one year up to 1st Jan next year.
Both 2nd and 3rd option would be valid, because there are no rows with dates < 2009, so crating left boundary on 01.01.2009 will not create fifth partition, because it would had to be empty.

upvoted 1 times

□ 👤 **badetoro234** 4 years, 1 month ago

The range should be RIGHT if you follow the link below.
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7#PartitionedTable
upvoted 2 times

⊟ 👤 **tem1234** 4 years, 1 month ago
What is the correct answer??
upvoted 1 times

⊟ 👤 **Amy007** 4 years, 1 month ago
Correct answer is Left and option 2 .
PARTITION (id RANGE LEFT FOR VALUES (10, 20, 30, 40)),
CLUSTERED COLUMNSTORE INDEX

Partition 1: col <= 10
Partition 2: 10 < col <= 20
Partition 3: 20 < col <= 30
Partition 4: 30 < col <= 40
Partition 5: 40 < col

4 values creates 5 partitions , so we need to go for 3 values that will create 4 partitions.
upvoted 1 times

⊟ 👤 **Garnew** 4 years, 1 month ago
This can be a bit confusing if you don't read and interpret what's said in the link below with an open mind;
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7

scroll to "Table partition options"
partition_column_name - Specifies the column that Azure Synapse Analytics will use to partition the rows. This column can be any data type. Azure Synapse Analytics sorts the partition column values in ascending order. The low-to-high ordering goes from LEFT to RIGHT in the RANGE specification.

RANGE LEFT - Specifies the boundary value belongs to the partition on the left (lower values). The default is LEFT.

RANGE RIGHT - Specifies the boundary value belongs to the partition on the right (higher values).

I will go with the answers provided, "LEFT and the 4 date partitions" - because the questions state that it's to the end of 2012, so the lower boundary will be the start of that year (2012-01-01), therefore LEFT range.
upvoted 1 times

⊟ 👤 **KpKo** 4 years, 2 months ago
The answer should be LEFT and 2nd Line.
upvoted 2 times

⊟ 👤 **meswapnilspal** 4 years, 2 months ago
Correct answer is 'RIGHT'
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7
(Scroll towards the end)

LEFT/RIGHT is confusing, trick to remember is
1. Draw a hyphen among the values
20090101-20100101-20110101-20120101
2. pick one value (say 20100101)
3. Now understand, in which range do we want 20100101 to fall? here we want it in 20100101-20110101.
4. 20110101 is towards its right, Hence 'RIGHT'
upvoted 3 times

⊟ 👤 **cadio30** 4 years, 2 months ago
In a range left partition function, all boundary values are upper boundaries, they are the last values in the partitions. If you partition by year, you use December 31st. If you partition by month, you use January 31st, February 28th / 29th, March 31st, April 30th and so on. In a range right partition function, all boundary values are lower boundaries, they are the first values in the partitions. If you partition by year, you use January 1st. If you partition by month, you use January 1st, February 1st, March 1st, April 1st and so on.

Reference: https://www.cathrinewilhelmsen.net/table-partitioning-in-sql-server/#:~:text=Range%20Left%20and%20Range%20Right&text=Range%20left%20means%20that%20the,value%20in%20the%20right%20partition.

upvoted 1 times

⊟ 👤 **NamishBansal** 4 years, 2 months ago

The range is correct. We need to create 4 partitions, 2009-2010, 2010-2011, 2011-2012, and 2012-2013, because we want data till the end of 2012.

upvoted 1 times

SIMULATION -
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10543936 -

You need to create an elastic pool that contains an Azure SQL database named db2 and a new SQL database named db3.
To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*

Step 1: Create a new SQL database named db3

1. Select SQL in the left-hand menu of the Azure portal. If SQL is not in the list, select All services, then type SQL in the search box.

2. Select + Add to open the Select SQL deployment option page. Select Single Database. You can view additional information about the different databases by selecting Show details on the Databases tile.
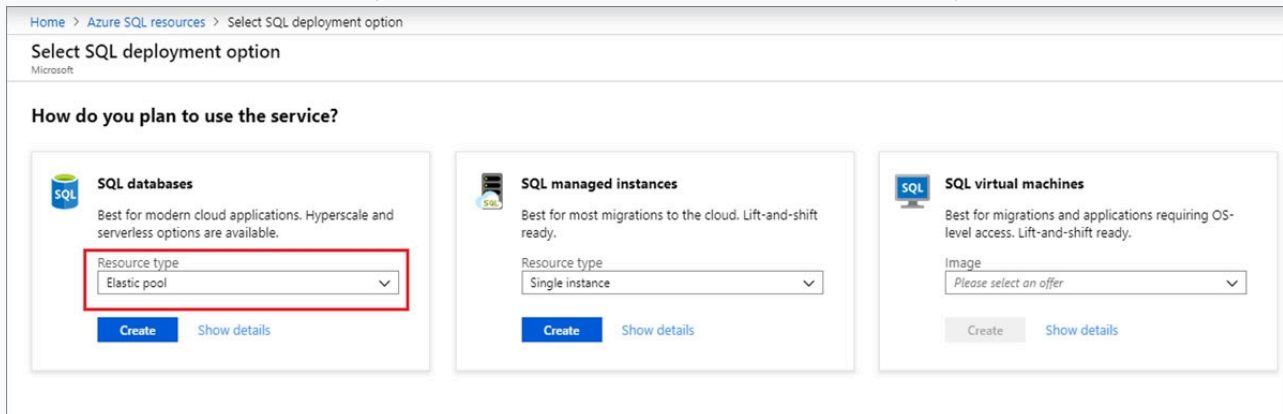
3. Select Create:
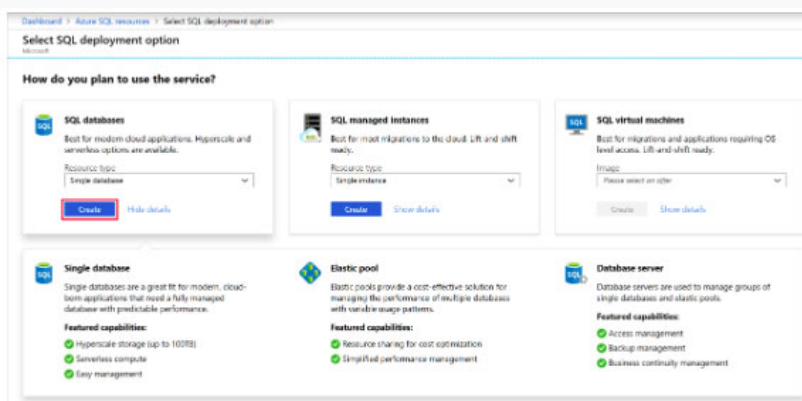
4. Enter the required fields if necessary.

5. Leave the rest of the values as default and select Review + Create at the bottom of the form.

6. Review the final settings and select Create. Use Db3 as database name.

On the SQL Database form, select Create to deploy and provision the resource group, server, and database.

Step 2: Create your elastic pool using the Azure portal.

1. Select Azure SQL in the left-hand menu of the Azure portal. If Azure SQL is not in the list, select All services, then type Azure SQL in the search box.

2. Select + Add to open the Select SQL deployment option page.

3. Select Elastic pool from the Resource type drop-down in the SQL Databases tile. Select Create to create your elastic pool.

4. Configure your elastic pool with the following values:

Name: Provide a unique name for your elastic pool, such as myElasticPool.

Subscription: Select your subscription from the drop-down.

ResourceGroup: Select the resource group.

Server: Select the server -

5. Select Configure elastic pool

6. On the Configure page, select the Databases tab, and then choose to Add database.

Configure

❤ Feedback

General Purpose
Scalable compute and storage options

Up to 7,000 IOPS
5-10 ms latency

< Looking for basic, standard, premium?

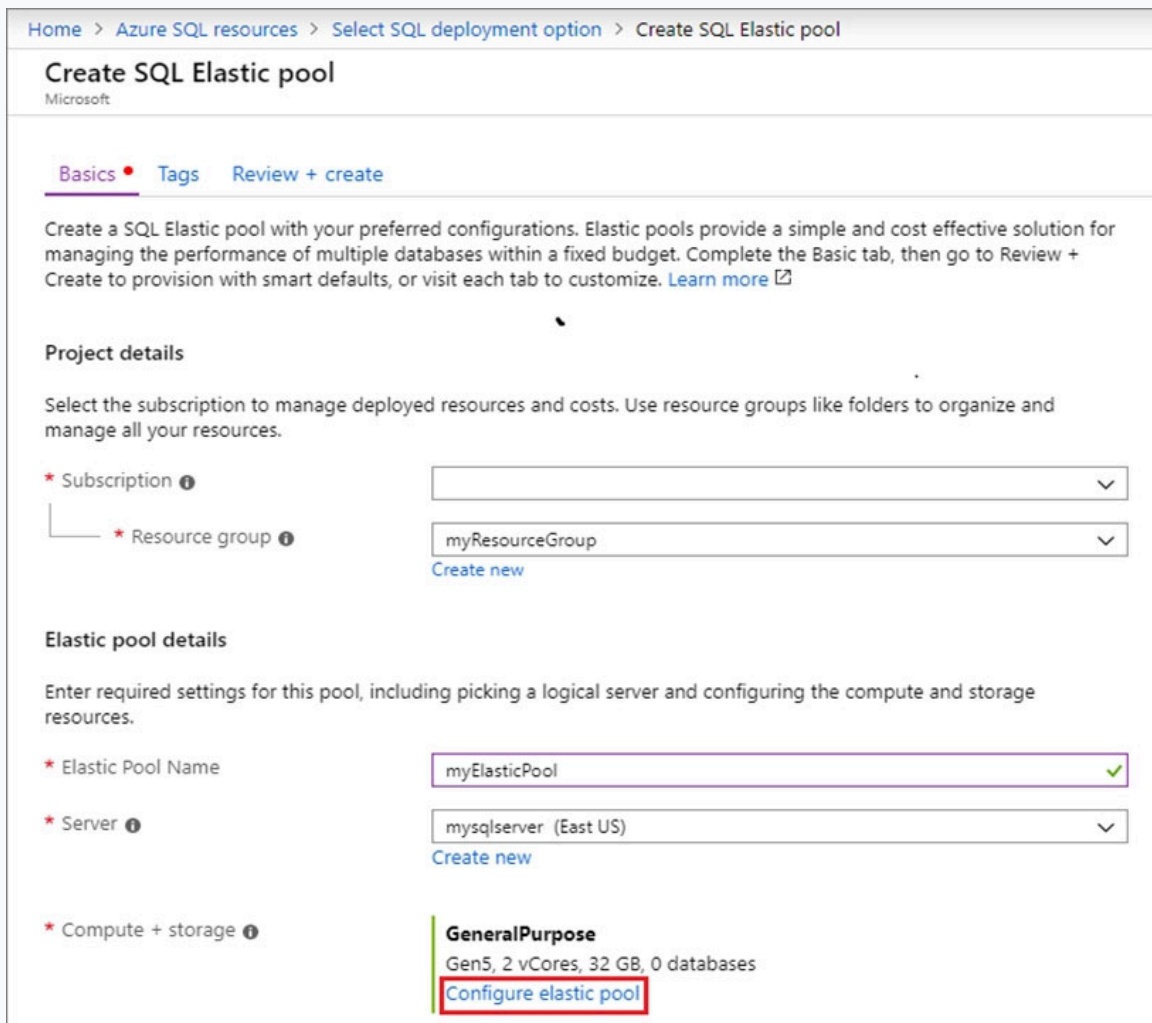Pool settings   Databases   Per database settings

+ Add databases   ↺ Revert selected

🔍 Search to filter databases...

DATABASE NAME	↑↓	PRICING TIER

Currently, there are no databases selected to be added to the pool. To add databases, click 'Add databases' above.

Add databases

⊟ Select all

🔍 Search to filter databases...

DATABASE NAME

☑   🗄 mySampleDatabase

7. Add the Azure SQL database named db2, and the new SQL database named db3 that you created in Step 1.

8. Select Review + create to review your elastic pool settings and then select Create to create your elastic pool.

Reference:

https://docs.microsoft.com/bs-latn-ba/azure/sql-database/sql-database-elastic-pool-failover-group-tutorial

---

⊟ 👤 **dduque10** 3 years, 9 months ago

The answer is obsolete

upvoted 1 times

SIMULATION -
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10543936 -

You need to create an Azure Storage account named account10543936. The solution must meet the following requirements:

☞ Minimize storage costs.

☞ Ensure that account10543936 can store many image files.

Ensure that account10543936 can quickly retrieve stored image files.

▪

To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*

Create a general-purpose v2 storage account, which provides access to all of the Azure Storage services: blobs, files, queues, tables, and disks.

1. On the Azure portal menu, select All services. In the list of resources, type Storage Accounts. As you begin typing, the list filters based on your input. Select
Storage Accounts.

2. On the Storage Accounts window that appears, choose Add.

3. Select the subscription in which to create the storage account.

4. Under the Resource group field, select Create new. Enter the name for your new resource group, as shown in the following image.
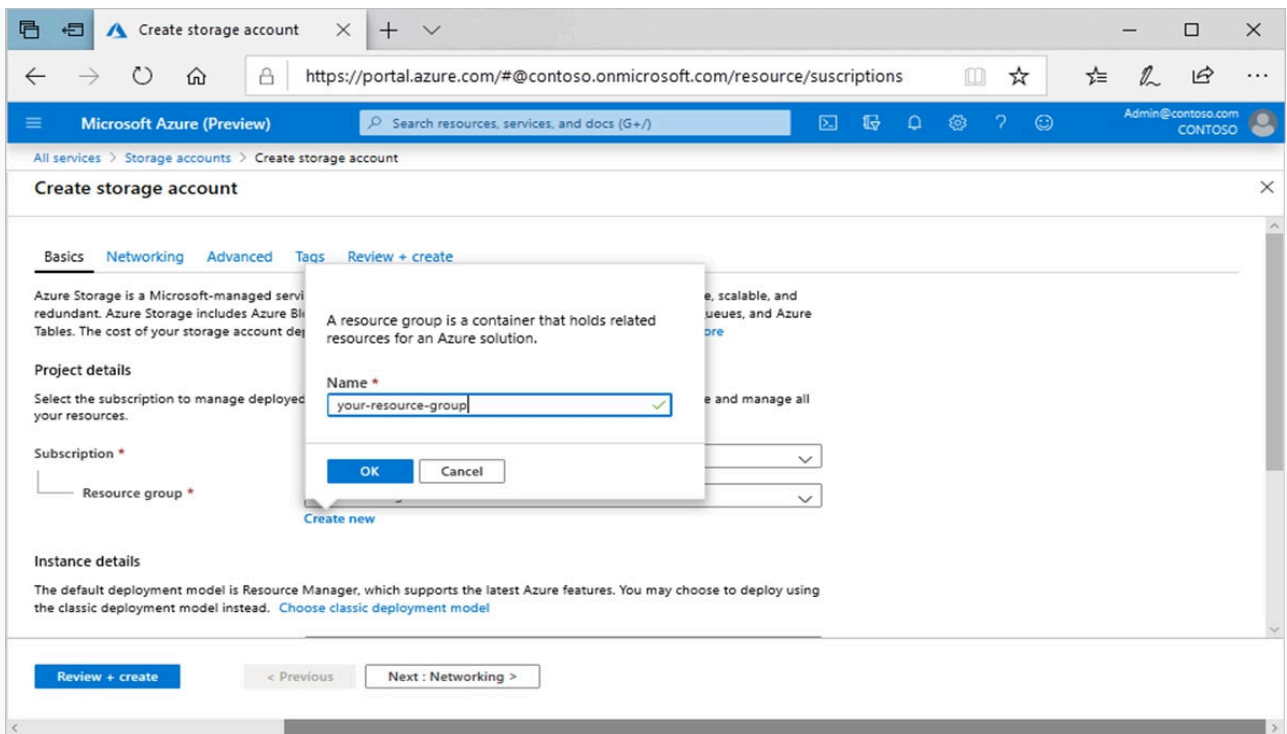
5. Next, enter the name account10543936 for your storage account.

6. Select a location for your storage account, or use the default location.

7. Leave these fields set to their default values:

Deployment model: Resource Manager

Performance: Standard -
Account kind: StorageV2 (general-purpose v2)
Replication: Read-access geo-redundant storage (RA-GRS)

Access tier: Hot -
8. Select Review + Create to review your storage account settings and create the account.
9. Select Create.
Reference:
https://docs.microsoft.com/en-us/azure/storage/common/storage-account-create

---

Wendy_DK `Highly Voted` 4 years, 2 months ago

LRS is thee cheapest one for cost.

upvoted 6 times

> **Internet_User** 4 years, 1 month ago
>
> I agree. Locally redundant storage is a cheaper option than globally redundant, and does not negatively impact any of the other requirements of the scenario.
>
> upvoted 6 times

Devendra00023 `Most Recent` 4 years, 2 months ago

I believe Storage tier should be Cool.. in order to minimize the cost

upvoted 2 times

> **JohnCrawford** 4 years, 2 months ago
>
> The question states "must be able to quickly retrieve...". Thus, storage has to be hot, BUT it also means we should use RA-GRS not LRS nor GRS.
>
> upvoted 2 times
>
> > **hichemck** 4 years, 1 month ago
> >
> > in the new azure there is no direct RA-GRS. you have to select Geo-zone-redundant storage (GZRS) and then check make read access to data available in the event of regional unavailability
> >
> > upvoted 1 times

SIMULATION -
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10543936 -

You need to ensure that users in the West US region can read data from a local copy of an Azure Cosmos DB database named cosmos10543936.
To complete this task, sign in to the Azure portal.
NOTE: This task might take several minutes to complete. You can perform other tasks while the task completes or end this section of the exam.

Suggested Answer: *See the explanation below.*

You can enable Availability Zones by using Azure portal when creating an Azure Cosmos account.

You can enable Availability Zones by using Azure portal.

Step 1: enable the Geo-redundancy, Multi-region Writes

1. In Azure Portal search for and select Azure Cosmos DB.

2. Locate the Cosmos DB database named cosmos10543936

3. Access the properties for cosmos10543936

4. enable the Geo-redundancy, Multi-region Writes.

Location: West US region -

Instance Details

* Account Name	myaccount1 ✓
* API ❶	Core (SQL) ⌄
└─ Apache Spark ❶	(Enable Disable)
	You're on the waitlist for Azure Cosmos with support for Apache Spark preview
* Location	(Asia Pacific) Southeast Asia ⌄
Geo-Redundancy ❶	(**Enable** Disable)
Multi-region Writes ❶	(**Enable** Disable)
Availability Zones ❶	(**Enable** Disable)

Step 2: Add region from your database account

1. In to Azure portal, go to your Azure Cosmos account, and open the Replicate data globally menu.

2. To add regions, select the hexagons on the map with the + label that corresponds to your desired region(s). Alternatively, to add a region, select the + Add region option and choose a region from the drop-down menu.

Add: West US region -

3. To save your changes, select OK.

Reference:

https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-manage-database-account

---

⊟ 👤 **marc88** 3 years, 10 months ago

The question specifies "users in the West US region can READ data". Geo redundancy should cut it and Multi region writes are unnecessary.

upvoted 1 times

⊟ 👤 **JohnCrawford** 4 years, 2 months ago

There is no need for multi-region writes only geo-redundancy. See https://docs.microsoft.com/en-us/azure/azure-sql/database/active-geo-replication-overview You can create the secondary across the country where it would have low latency for someone nearby and it would be readable.

upvoted 2 times

⠀⠀⊟ 👤 **SuperAlex** 4 years, 2 months ago

⠀⠀The question is asking the Azure Cosmos DB, not Azure SQL Database in your link.

⠀⠀upvoted 3 times

⠀⠀⠀⠀⊟ 👤 **JohnCrawford** 4 years, 2 months ago

⠀⠀⠀⠀I am also now wondering why they suggested using the availability zones feature. The way the question is worded, "ensure that users in the West region can read..." makes me wonder if "ensure" is the key word here and that's why they are also enabling the availability zones which I would not have done if I was only interested in saving money.

⠀⠀⠀⠀upvoted 3 times

cadio30 4 years, 1 month ago

It make sense to enable the "Availability Zones" as part of the requirement since there is a chance the replicated server could go down.

upvoted 1 times

JohnCrawford 4 years, 2 months ago

You're right. Sorry about that. In this link (https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability) look at the section on multi-region accounts with a single write region. There is no need given the stated objectives to enable multi-region writes. Thanks for the catch.

upvoted 4 times

cadio30 4 years, 1 month ago

It make sense to enable the "Availability Zones" as part of the requirement since there is a chance the replicated server could go down.

upvoted 1 times

JohnCrawford 4 years, 2 months ago

You're right. Sorry about that. In this link (https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability) look at the section on multi-region accounts with a single write region. There is no need given the stated objectives to enable multi-region writes. Thanks for the catch.

upvoted 4 times

SIMULATION -
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10543936 -

You plan to enable Azure Multi-Factor Authentication (MFA).
You need to ensure that User1-10543936@ExamUsers.com can manage any databases hosted on an Azure SQL server named SQL10543936 by signing in using his Azure Active Directory (Azure AD) user account.
To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*
Provision an Azure Active Directory administrator for your managed instance
Each Azure SQL server (which hosts a SQL Database or SQL Data Warehouse) starts with a single server administrator account that is the administrator of the entire Azure SQL server. A second SQL Server administrator must be created, that is an Azure AD account. This principal is created as a contained database user in the master database.
1. In the Azure portal, in the upper-right corner, select your connection to drop down a list of possible Active Directories. Choose the correct Active Directory as the default Azure AD. This step links the subscription-associated Active Directory with Azure SQL server making sure that the same subscription is used for both
Azure AD and SQL Server. (The Azure SQL server can be hosting either Azure SQL Database or Azure SQL Data Warehouse.)
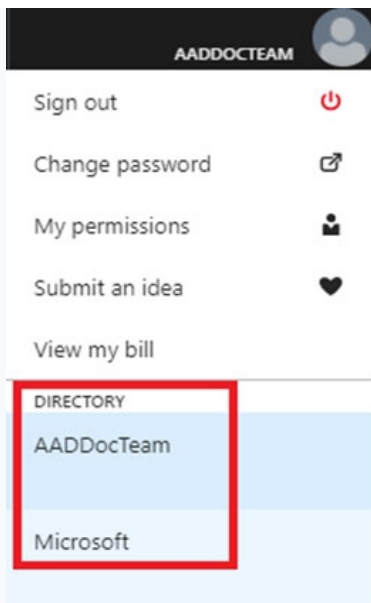
2. Search for and select the SQL server SQL10543936
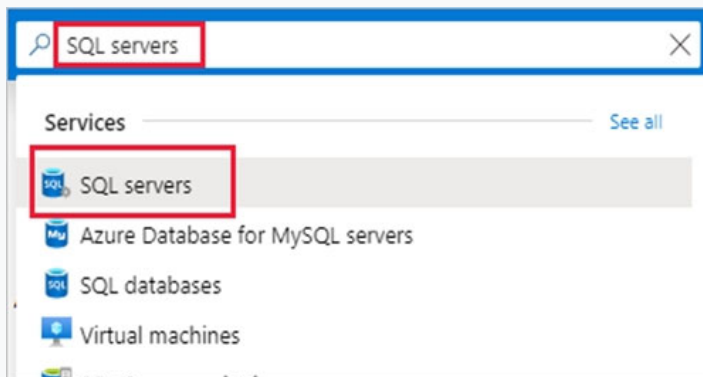
3. In SQL Server page, select Active Directory admin.

4. In the Active Directory admin page, select Set admin.
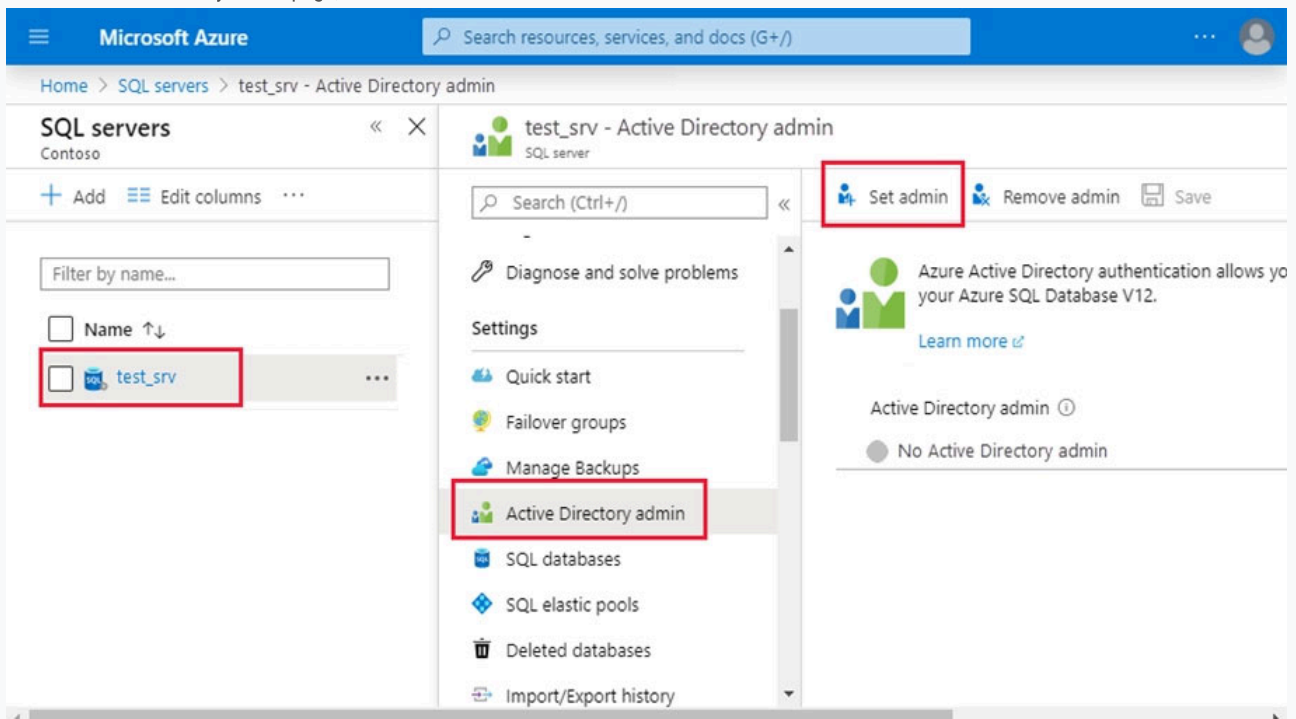
5. In the Add admin page, search for user User1-10543936@ExamUsers.com, select it, and then select Select. (The Active Directory admin page shows all members and groups of your Active Directory. Users or groups that are grayed out cannot be selected because they are not supported as Azure AD administrators.

6. At the top of the Active Directory admin page, select SAVE.

Reference:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-aad-authentication-configure?

👤 **mftp** 4 years, 1 month ago

Answer here: https://docs.microsoft.com/en-us/azure/active-directory/authentication/tutorial-enable-azure-mfa

upvoted 3 times

HOTSPOT -

You have the following Azure Stream Analytics query.

```
WITH

step1 AS (SELECT *
    FROM input1
    PARTITION BY StateID
    INTO 10),
step2 AS (SELECT *
    FROM input2
    PARTITION BY StateID
    INTO 10)

SELECT *
INTO output
FROM step1
PARTITION BY StateID
UNION step2
  BY StateID
```

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Statements	Yes	No
The query joins two streams of partitioned data.	○	○
The stream scheme key and count must match the output scheme.	○	○
Providing 60 streaming units will optimize the performance of the query.	○	○

Suggested Answer:

Answer Area

Statements	Yes	No
The query joins two streams of partitioned data.	○	◉
The stream scheme key and count must match the output scheme.	◉	○
Providing 60 streaming units will optimize the performance of the query.	◉	○

Box 1: No -

Note: You can now use a new extension of Azure Stream Analytics SQL to specify the number of partitions of a stream when reshuffling the data.

The outcome is a stream that has the same partition scheme. Please see below for an example:

WITH step1 AS (SELECT * FROM [input1] PARTITION BY DeviceID INTO 10), step2 AS (SELECT * FROM [input2] PARTITION BY DeviceID INTO 10)

SELECT * INTO [output] FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID

Note: The new extension of Azure Stream Analytics SQL includes a keyword INTO that allows you to specify the number of partitions for a stream when performing reshuffling using a PARTITION BY statement.

Box 2: Yes -

When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count.

Box 3: Yes -

Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

In general, the best practice is to start with 6 SUs for queries that don't use PARTITION BY.
Here there are 10 partitions, so 6x10 = 60 SUs is good.
Note: Remember, Streaming Unit (SU) count, which is the unit of scale for Azure Stream Analytics, must be adjusted so the number of physical resources available to the job can fit the partitioned flow. In general, six SUs is a good number to assign to each partition. In case there are insufficient resources assigned to the job, the system will only apply the repartition if it benefits the job.
Reference:
https://azure.microsoft.com/en-in/blog/maximize-throughput-with-repartitioning-in-azure-stream-analytics/ https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

---

👤 **sjsheikdawood** `Highly Voted 👍` 4 years, 2 months ago

Answer is No,No,No (source whizlab)
Here we are using the UNION clause which is different from the JOIN clause
We need to match the partition key in the input and output scheme, but not necessarily need to match the count.
You can scale up to 6 streaming units for each step in a job. If you have partitions, you need to
multiply the number of partitions by 6.
Now in the query, we have 2 select queries in the input streams. And each has a partition count of 10.
That means we can scale the job to the following number of streaming units.
Number of SELECT queries * Number of partitions * 6
= 2*10*6 = 120
In the query, we have one SELECT statement in the output with no partition count. Hence the
calculation for the maximum number of streaming units is
Number of SELECT queries * 6 = 6.
Hence the total number of streaming units that can be assigned to the job is 126.

upvoted 13 times

👤 **vaseva1** `Highly Voted 👍` 4 years, 2 months ago

Yes at 1st Question

upvoted 10 times

👤 **massnonn** `Most Recent ⊘` 3 years, 7 months ago

the answer is yes-yes-yes

upvoted 2 times

👤 **hello_there_** 4 years ago

The answer is yes, yes, no. The example comes almost directly from the microsoft documentation. Quoting from docs.microsoft.com/en-us/azure/stream-analytics/repartition:

The following example query joins two streams of repartitioned data. When joining two streams of repartitioned data, the streams must have the same partition key and count. The outcome is a stream that has the same partition scheme.

SQL
WITH step1 AS (SELECT * FROM input1 PARTITION BY DeviceID),
step2 AS (SELECT * FROM input2 PARTITION BY DeviceID)

SELECT * INTO output FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID

The documentation clearly states that this example consitutes a join of partitioned streams, even though traditionally one might not consider a union a join.
Others have explained why 60 stu is not optimal.

upvoted 3 times

👤 **hello_there_** 4 years ago

I forgot to include the following from the same documetation:
The output scheme should match the stream scheme key and count so that each substream can be flushed independently. I guess the question states "must" while the documentation states "should", so I guess "no" is the technically correct answer here.

upvoted 2 times

👤 **vrmei** 4 years ago

Second Option should be No.
The output scheme should match the stream scheme key and count so that each substream can be flushed independently.
https://docs.microsoft.com/en-us/azure/stream-analytics/repartition

upvoted 3 times

niwe 4 years ago

I don't think "Union step2 By StateID" it is valid.

upvoted 2 times

jasonoubre 4 years ago

so what is the answer??

upvoted 3 times

NamishBansal 4 years, 1 month ago

3rd question is correct. While 126 RUs might be the most ideal, 60 RUs will still optimize the preformance.

upvoted 2 times

Aragorn_2021 4 years, 2 months ago

Answer for RU unit is wrong. Ideal RU unit required is 126

upvoted 2 times

cadio30 4 years, 2 months ago

The following example query joins two streams of repartitioned data. When joining two streams of repartitioned data, the streams must have the same partition key and count. The outcome is a stream that has the same partition scheme.

WITH step1 AS (SELECT * FROM input1 PARTITION BY DeviceID),
step2 AS (SELECT * FROM input2 PARTITION BY DeviceID)

SELECT * INTO output FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID

Reference: https://docs.microsoft.com/en-us/azure/stream-analytics/repartition

upvoted 3 times

cadio30 4 years, 1 month ago

Propose solution is No, Yes, Yes

Referencing the query, the "PARTITION" keyword is missing after the "union step2"

upvoted 2 times

dangal95 4 years, 2 months ago

why are some people saying that the first one is also yes? A Union is not a join. A union will combine the rows from both steps together. A join would combine the columns based on some join condition

upvoted 2 times

Rob77 4 years, 1 month ago

and yet Microsoft calls union "join" see https://docs.microsoft.com/en-us/azure/stream-analytics/repartition

upvoted 1 times

kriskan1 4 years, 2 months ago

In the union query for step2 there is no partition by. so may be the query is wront. thats the reason for No for 1st question

upvoted 2 times

Sai2609 4 years, 2 months ago

It should be yes,yes,yes

upvoted 1 times

AngelRio 4 years, 2 months ago

No, No and No (source Whizlabs)

upvoted 2 times

Sherinm 4 years, 2 months ago

Box 1 is Yes.

When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count. The outcome is a stream that has the same partition scheme.

https://azure.microsoft.com/en-in/blog/maximize-throughput-with-repartitioning-in-azure-stream-analytics/

upvoted 4 times

DRAG DROP -

You have an Azure SQL database named DB1 in the East US 2 region.

You need to build a secondary geo-replicated copy of DB1 in the West US region on a new server.

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.

Select and Place:

Suggested Answer:

Step 1: From the Geo-replication settings of DB1, select West US

The following steps create a new secondary database in a geo-replication partnership.

1. In the Azure portal, browse to the database that you want to set up for geo-replication.

2. (Step 1) On the SQL database page, select geo-replication, and then select the region to create the secondary database.

3. (Step 2-3) Select or configure the server and pricing tier for the secondary database.

Step 2: Create a target server and select a pricing tier

Step 3: On the secondary server, create logins that match the SIDs on the primary server.

Incorrect Answers:

Not log shipping: Replication is used.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-active-geo-replication-portal

Jerrie86 2 years, 5 months ago

Answer looks good. Very similar question above in the form of Simulation.

upvoted 2 times

HOTSPOT -

You have an Azure SQL database that contains a table named Employee. Employee contains sensitive data in a decimal (10,2) column named Salary.

You need to ensure that nonprivileged users can view the table data, but Salary must display a number from 0 to 100.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Permission:
- CONNECT
- SELECT
- UNMASK

Mask:
- Custom text
- Default
- Random number

Suggested Answer:

Answer Area

Permission:
- CONNECT
- SELECT
- UNMASK

Mask:
- Custom text
- Default
- Random number

Box 1: SELECT -

Users with SELECT permission on a table can view the table data. Columns that are defined as masked, will display the masked data.

Incorrect:

Grant the UNMASK permission to a user to enable them to retrieve unmasked data from the columns for which masking is defined.

The CONTROL permission on the database includes both the ALTER ANY MASK and UNMASK permission.

Box 2: Random number -

Random number: Masking method, which generates a random number according to the selected boundaries and actual data types. If the designated boundaries are equal, then the masking function is a constant number.

Masking Field Format

Random number

From
0

To
0

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 an Azure subscription that contains an Azure Storage account.

You plan to implement changes to a data storage solution to meet regulatory and compliance standards.

Every day, Azure needs to identify and delete blobs that were NOT modified during the last 100 days.

Solution: You apply an Azure policy that tags the storage account.

Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

Instead apply an Azure Blob storage lifecycle policy.

Reference:

https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal

👤 **memo43** 4 years, 1 month ago

answer is 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 an Azure subscription that contains an Azure Storage account.

You plan to implement changes to a data storage solution to meet regulatory and compliance standards.

Every day, Azure needs to identify and delete blobs that were NOT modified during the last 100 days.

Solution: You apply an expired tag to the blobs in the storage account.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

Instead apply an Azure Blob storage lifecycle policy.

Reference:

https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal

---

☐ 👤 **kilowd** `Highly Voted 👍` 4 years, 10 months ago

Data sets have unique lifecycles. Early in the lifecycle, people access some data often. But the need for access drops drastically as the data ages. Some data stays idle in the cloud and is rarely accessed once stored. Some data expires days or months after creation, while other data sets are actively read and modified throughout their lifetimes. Azure Blob storage lifecycle management offers a rich, rule-based policy for GPv2 and Blob storage accounts. Use the policy to transition your data to the appropriate access tiers or expire at the end of the data's lifecycle.

The lifecycle management policy lets you:

Transition blobs to a cooler storage tier (hot to cool, hot to archive, or cool to archive) to optimize for performance and cost
Delete blobs at the end of their lifecycles
Define rules to be run once per day at the storage account level
Apply rules to containers or a subset of blobs (using name prefixes or blob index tags as filters)

upvoted 10 times

    ☐ 👤 **syu31svc** 4 years, 7 months ago

    Agree!

    upvoted 1 times

☐ 👤 **memo43** `Most Recent ⊙` 4 years, 1 month ago

answer is 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 an Azure subscription that contains an Azure Storage account.

You plan to implement changes to a data storage solution to meet regulatory and compliance standards.

Every day, Azure needs to identify and delete blobs that were NOT modified during the last 100 days.

Solution: You apply an Azure Blob storage lifecycle policy.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *A*

Azure Blob storage lifecycle management offers a rich, rule-based policy for GPv2 and Blob storage accounts. Use the policy to transition your data to the appropriate access tiers or expire at the end of the data's lifecycle.

The lifecycle management policy lets you:

☞ Transition blobs to a cooler storage tier (hot to cool, hot to archive, or cool to archive) to optimize for performance and cost

☞ Delete blobs at the end of their lifecycles

☞ Define rules to be run once per day at the storage account level

☞ Apply rules to containers or a subset of blobs (using prefixes as filters)

Reference:

https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal

---

  ☐ 👤 **memo43** 4 years, 1 month ago

answer is CORRECT

    upvoted 2 times

  ☐ 👤 **rajat009** 4 years, 6 months ago

duplicate

    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 an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

☞ The first two prefix characters must be exposed.

☞ The last four suffix characters must be exposed.

☞ All other characters must be masked.

Solution: You implement data masking and use a custom string function mask.

Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

Reference:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

---

🗆 👤 **zac874997967** `Highly Voted 👍` 4 years, 10 months ago

For this question the last sentence is "You implement data masking and use a custom text mask." and the answer is yes.

upvoted 17 times

🗆 👤 **diulin** `Highly Voted 👍` 5 years, 1 month ago

This question appeared as Topic 1 Question #22 earlier

upvoted 7 times

🗆 👤 **dumpi** `Most Recent ⊙` 4 years ago

Answer is Correct only .It should be "You implement data masking and use a custom text mask."

upvoted 1 times

🗆 👤 **vrmei** 4 years ago

It should be "Yes"; Just test by adding sample database while creating sql database in azure. Goto Dynamic Masking and click any field to add mask and edit the mask.

If you check, you can see only below types.

1. Default Value
2. Credit Card
3. Email
4. Number
5. Custom String

Check the latest document : 03/24/2021

https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking?view=sql-server-ver15

upvoted 3 times

🗆 👤 **Wendy_DK** 4 years, 2 months ago

Newly updated document (1/25/2021) online Masking function called custom text. The old document online it was called custom string.

upvoted 7 times

🗆 👤 **Wendy_DK** 4 years, 2 months ago

Answer should be Yes

upvoted 3 times

🗆 👤 **JohnCrawford** 4 years, 2 months ago

I think for this answer they were being too literal in their interpretation of first/last characters for the custom mask. That does not mean exactly 1 at front and back, but rather that those are parameters that let you determine how many to expose at the front and back. 0, 1 or more are all permissible.

upvoted 1 times

☐ 👤 **dev2dev** 4 years, 5 months ago

I was expecting a question with 'Yes' solution which is to use custom text masking. this is repeative.

upvoted 2 times

☐ 👤 **dmnantilla9** 4 years, 5 months ago

Repeated question.

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 an Azure subscription that contains an Azure Storage account.

You plan to implement changes to a data storage solution to meet regulatory and compliance standards.

Every day, Azure needs to identify and delete blobs that were NOT modified during the last 100 days.

Solution: You schedule an Azure Data Factory pipeline.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

Instead you can use the Delete Activity in Azure Data Factory to delete files or folders from on-premises storage stores or cloud storage stores or apply an Azure

Blob storage lifecycle policy.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/delete-activity https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal

---

👤 **kaigalmane** `Highly Voted 👍` 4 years, 1 month ago

This should be Yes right?

upvoted 5 times

    👤 **111222333** 4 years, 1 month ago

    I also think "yes": https://docs.microsoft.com/en-us/azure/data-factory/delete-activity#clean-up-the-expired-files-that-were-last-modified-before-201811

    "You can create a pipeline to clean up the old or expired files by leveraging file attribute filter: "LastModified" in dataset."

    And it uses blob files:

    "dataset": {

    "referenceName":"BlobFilesLastModifiedBefore201811",

    ...

    }

    upvoted 4 times

👤 **satyamkishoresingh** `Most Recent ⓘ` 3 years, 9 months ago

when it can be done with Azure Blob storage lifecycle policy, why data factory ?

upvoted 1 times

👤 **niwe** 4 years, 1 month ago

There is no activity mentioned, it should be No.

Am I right?

upvoted 2 times

    👤 **niwe** 4 years, 1 month ago

    Correct question is question #58

    upvoted 1 times

👤 **Saravjeet** 4 years, 1 month ago

It should be Yes.

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 an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You convert the files to compressed delimited text files.

Does this meet the goal?

A. Yes

B. No

Suggested Answer: *A*

All file formats have different performance characteristics. For the fastest load, use compressed delimited text files.

Reference:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

---

☐ 👤 **Marcus1612** 3 years, 8 months ago

It should be "Yes".

Compress the source file is good practice in this use case,. Polybase is not mandatory in this question and the data to transfert is not so huge. As mentionned by Microsoft below, we can turn off "using polybase" which avoid the overhead of splitting the files.

"Row size and data type limits

PolyBase loads are limited to rows smaller than 1 MB. It cannot be used to load to VARCHR(MAX), NVARCHAR(MAX), or VARBINARY(MAX). For more information, see Azure Synapse Analytics service capacity limits.

When your source data has rows greater than 1 MB, you might want to vertically split the source tables into several small ones. Make sure that the largest size of each row doesn't exceed the limit. The smaller tables can then be loaded by using PolyBase and merged together in Azure Synapse Analytics.

Alternatively, for data with such wide columns, you can use non-PolyBase to load the data by turning off "allow PolyBase" setting."

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse?tabs=data-factory

upvoted 1 times

☐ 👤 **vrmei** 4 years ago

It should be "No" as the row length should be less than 1 MB

upvoted 1 times

☐ 👤 **hello_there_** 4 years ago

The max 1 MB limit only applies if you plan to use polybase

upvoted 2 times

☐ 👤 **lgtiza** 4 years ago

Check answer 31. It seems that the 1 MB limit is no longer necessary.

upvoted 1 times

You are designing an enterprise data warehouse in Azure Synapse Analytics. You plan to load millions of rows of data into the data warehouse each day.

You must ensure that staging tables are optimized for data loading.

You need to design the staging tables.

What type of tables should you recommend?

    A. Round-robin distributed table

    B. Hash-distributed table

    C. Replicated table

    D. External table

Suggested Answer: *A*

👤 **syu31svc** `Highly Voted 👍` 4 years, 7 months ago

From https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-overview:

Use round-robin for the staging table. The load with CTAS is fast. Once the data is in the staging table, use INSERT...SELECT to move the data to production tables.

upvoted 17 times

  👤 **memo43** 4 years, 1 month ago

  keyword is staging table

  upvoted 1 times

👤 **jayeshstudies** `Most Recent ⊘` 4 years ago

keyword --> staging table --> round robin is best

upvoted 1 times

👤 **cadio30** 4 years, 2 months ago

round robin is entirely correct for staging tables

upvoted 3 times

👤 **cjh1912** 4 years, 5 months ago

answer is correct its specifically asked for fast loading , not read, which leads to round robin being the correct answer

upvoted 4 times

HOTSPOT -

You have a SQL pool in Azure Synapse.

You plan to load data from Azure Blob storage to a staging table. Approximately 1 million rows of data will be loaded daily. The table will be truncated before each daily load.

You need to create the staging table. The solution must minimize how long it takes to load the data to the staging table.

How should you configure the table? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Distribution:
- Hash
- Replicated
- Round-robin

Indexing:
- Clustered
- Clustered columnstore
- Heap

Partitioning:
- Date
- None

Answer Area

Suggested Answer:

Distribution:
- **Hash**
- Replicated
- Round-robin

Indexing:
- Clustered
- **Clustered columnstore**
- Heap

Partitioning:
- **Date**
- None

Box 1: Hash -

Hash-distributed tables improve query performance on large fact tables. hey can have very large numbers of rows and still achieve high performance.

Incorrect:

Round-robin tables are useful for improving loading speed.

Box 2: Clustered columnstore -

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal

mamhh `Highly Voted 👍` 4 years, 2 months ago

Round-Robin

Heap

None

upvoted 68 times

LongBao `Highly Voted 👍` 4 years, 2 months ago

From https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-overview:

"Use round-robin for the staging table. The load with CTAS is fast. Once the data is in the staging table, use INSERT...SELECT to move the data to production tables."

This is a staging table, not a fact, so I think the answer is Round robin, Heap, Date.

upvoted 10 times

> **maciejt** 4 years, 1 month ago
>
> If Round Robin, then you don't choose column to partition on, it can only be None
>
> upvoted 2 times

> > **uzairahm** 3 years ago
> >
> > As @TessieB has indicated quoting MS Docs "Partitioning is also supported on all distribution types, including both hash or round robin distributed."
> >
> > https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition
> >
> > upvoted 1 times

dark_one `Most Recent ⊘` 1 year, 2 months ago

round robin, heap, none

https://learn.microsoft.com/en-us/azure/synapse-analytics/sql/data-loading-best-practices#load-to-a-staging-table

upvoted 1 times

saranya23 3 years, 10 months ago

Round-Robin , Heap , None

upvoted 2 times

elimey 3 years, 11 months ago

round-robin, haep, non

Why should someone hash distribute a stage table?

upvoted 1 times

TessieB 3 years, 11 months ago

I think the answer should be:

Round robin

Heap

Date

See here: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

Partitioning and distribution are two different things! Partitioning is supported by all types of distribution and by all types of indexes! Partitioning can speed up the loading process and it's often done by using a date column. In this question it's not clear however, if the Date column is the right fit for the loading process, but since it is often used like that, I'm gonna go and say that Date might be the correct answer! :)

upvoted 2 times

alok1988 4 years ago

Round-Robin , Heap , None

upvoted 4 times

Saravjeet 4 years, 1 month ago

I think it should be Round-RObin, Heap and None. Refer the link

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-data-loading-guidance?view=azure-sqldw-latest

upvoted 4 times

Qrm_1972 4 years, 1 month ago

The correct answer is :

Round-Robin

Clustered columnstore

Data

Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column. Partitioning is supported on all dedicated SQL pool table types; including clustered columnstore, clustered index, and heap. Partitioning is also supported on all distribution types, including both hash or round robin distributed.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

upvoted 2 times

    maciejt 4 years, 1 month ago

    Round RObin is random equal distribution, it doesn't include choosing a column to partition on

    upvoted 1 times

robin_examtopics 4 years, 1 month ago

The answer should be Round-Robin/ Heap/ None.

Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/guidance-for-loading-data#loading-to-a-staging-table

"To achieve the fastest loading speed for moving data into a dedicated SQL pool table, load data into a staging table. Define the staging table as a heap and use round-robin for the distribution option."

upvoted 7 times

itmemememe 4 years, 1 month ago

If you are partitioning by Date and using Clustered Column Store, Would that be faster than using a Heap, None?

upvoted 1 times

cadio30 4 years, 2 months ago

Round-Robin

Heap

None

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-index

upvoted 3 times

cadio30 4 years, 2 months ago

round-robin

heap

Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-index

upvoted 2 times

jamorey 4 years, 2 months ago

Anyone know the correct answer?, I'm between (Round-Robin, Heap and None) the other option (Round-Robin, Heap, Date)

upvoted 1 times

    Garnew 4 years, 1 month ago

    I'll go with "Round-Robin, Heap and None", Round-Robin is the best and obvious choice, then Heap as there's no need for indexing since it's just for loading purposes, and then None as Round-Robin does not support partitioning.

    upvoted 1 times

Pairon 4 years, 2 months ago

I think that the first is round robin (fastest way to load data) and the third box should be None, since round robin doesn't need partitioning

upvoted 4 times

alf99 4 years, 2 months ago

Answer should be "round-robin"

upvoted 7 times

☐ 👤 **princy18** 4 years, 2 months ago

Consider using the round-robin distribution for your table in the following scenarios:

When getting started as a simple starting point since it is the default

If there is no obvious joining key

If there is no good candidate column for hash distributing the table

If the table does not share a common join key with other tables

If the join is less significant than other joins in the query

When the table is a temporary staging table

upvoted 3 times

☐ 👤 **princy18** 4 years, 2 months ago

Consider using the round-robin distribution for your table in the following scenarios:

When getting started as a simple starting point since it is the default

If there is no obvious joining key

If there is no good candidate column for hash distributing the table

If the table does not share a common join key with other tables

If the join is less significant than other joins in the query

When the table is a temporary staging table

You have an enterprise-wide Azure Data Lake Storage Gen2 account. The data lake is accessible only through an Azure virtual network named VNET1.

You are building a SQL pool in Azure Synapse that will use data from the data lake.

Your company has a sales team. All the members of the sales team are in an Azure Active Directory group named Sales. POSIX controls are used to assign the

Sales group access to the files in the data lake.

You plan to load data to the SQL pool every hour.

You need to ensure that the SQL pool can load the sales data from the data lake.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

    A. Create a managed identity.

    B. Use the shared access signature (SAS) as the credentials for the data load process.

    C. Add the managed identity to the Sales group.

    D. Add your Azure Active Directory (Azure AD) account to the Sales group.

    E. Create a shared access signature (SAS).

    F. Use the managed identity as the credentials for the data load process.

Suggested Answer: *ACD*

A: The managed identity grants permissions to the dedicated SQL pools in the workspace.

Note: Managed identity for Azure resources is a feature of Azure Active Directory. The feature provides Azure services with an automatically managed identity in

Azure AD -

Reference:

https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-managed-identity

Community vote distribution

ACF (100%)

☐ 👤 **ACSC** `Highly Voted 👍` 4 years, 5 months ago

Correct answers are A, C, F.

upvoted 69 times

☐ 👤 **ck8.kakade** `Most Recent ⊘` 11 months, 1 week ago

F. Use the managed identity as the credentials for the data load process.

A. Create a managed identity.

B. Use the shared access signature (SAS) as the credentials for the data load process.

upvoted 1 times

   ☐ 👤 **ck8.kakade** 11 months, 1 week ago

Sorry for the typo - I meant FAC

upvoted 1 times

☐ 👤 **FredNo** 3 years, 7 months ago

`Selected Answer: ACF`

ACF, or FAD in the correct order are the good answers

upvoted 1 times

   ☐ 👤 **FredNo** 3 years, 7 months ago

I meant FAD

upvoted 1 times

      ☐ 👤 **FredNo** 3 years, 7 months ago

FAC***

upvoted 1 times

hsetin 3 years, 9 months ago

ACD is fine. data load process has nothing to do with Sales Group. you just need to add sales group to Active directory.

upvoted 1 times

uzairahm 3 years ago

Please Azure AD Group Sales already created: "All the members of the sales team are in an Azure Active Directory group named Sales."

upvoted 1 times

hoangton 4 years, 1 month ago

I also think ACF

upvoted 3 times

cadio30 4 years, 1 month ago

Agree with A,C,F

Reference: https://docs.microsoft.com/en-us/azure/purview/register-scan-adls-gen2

upvoted 4 times

davita8 4 years, 2 months ago

Correct answers are C, D, F.

upvoted 2 times

Wendy_DK 4 years, 2 months ago

Correct answers are A, C, F.

upvoted 3 times

HOTSPOT -

You have two Azure Storage accounts named Storage1 and Storage2. Each account contains an Azure Data Lake Storage file system. The system has files that contain data stored in the Apache Parquet format.

You need to copy folders and files from Storage1 to Storage2 by using a Data Factory copy activity. The solution must meet the following requirements:

☞ No transformations must be performed.

☞ The original folder structure must be retained.

How should you configure the copy activity? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Source dataset type:

▼
Binary
Parquet
Delimited text

Copy activity copy behavior:

▼
FlattenHierarchy
MergeFiles
PreserveHierarchy

Suggested Answer:

Answer Area

Source dataset type:

▼
Binary
Parquet
Delimited text

Copy activity copy behavior:

▼
FlattenHierarchy
MergeFiles
PreserveHierarchy

Box 1: Parquet -

For Parquet datasets, the type property of the copy activity source must be set to ParquetSource..

Box 2: PreserveHierarchy -

PreserveHierarchy (default): Preserves the file hierarchy in the target folder. The relative path of the source file to the source folder is identical to the relative path of the target file to the target folder.

Incorrect Answers:

FlattenHierarchy: All files from the source folder are in the first level of the target folder. The target files have autogenerated names.

MergeFiles: Merges all files from the source folder to one file. If the file name is specified, the merged file name is the specified name. Otherwise, it's an autogenerated file name.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/format-parquet https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

---

🗅 👤 **SorinXp** `Highly Voted 👍` 4 years, 1 month ago

The first box should be "Binary". It says - no transformation.

upvoted 6 times

🗅 👤 **[Removed]** 4 years ago

Binary is only for Binary format: https://docs.microsoft.com/en-us/azure/data-factory/format-binary
upvoted 1 times

⊟ 👤 **lgtiza** 4 years ago
Every parquet file is also a binary file. I think the key is "no transformations", so why the extra work of interpreting a parquet file?! Binary and preserve hierarchy should do it imo.
upvoted 3 times

⊟ 👤 **medsimus** Most Recent ⊙ 3 years, 8 months ago
First box should be "Binary" . I tested it with the 2 options . using paquet i got an error with the following message :
"Dataset Parquet1 location is a folder, the wildcard file name is required for Copy data1"
upvoted 3 times

⊟ 👤 **Hinzzz** 4 years ago
The given answer is correct Parquet and preserve hierarchy
upvoted 4 times

⊟ 👤 **CarNama_IG** 4 years ago
You can use Binary dataset in Copy activity, GetMetadata activity, or Delete activity. When using Binary dataset, ADF does not parse file content but treat it as-is. When using Binary dataset in copy activity, you can only copy from Binary dataset to Binary dataset..so the ans should be parquet
upvoted 1 times

⊟ 👤 **hello_there_** 4 years ago
Why does it need to be parquet? Just configure the sink dataset as binary as well. This way ADF doesn't need to parse the files. You just need parquet if you want to do some transformation or when the sink dataset is an existing parquet dataset
upvoted 1 times

⊟ 👤 **maciejt** 4 years, 1 month ago
It should be Binary - it copies the files as they are, no need to parse the parquet format if you don't need to transform them.
upvoted 2 times

⊟ 👤 **cadio30** 4 years, 2 months ago
Agree with the answer as both source and sink can accommodate "parquet" extension files using the behavior as seen below. Try working it on ADFv2

File Format: Parquet (source and sink)
Copy behavior: Preserve Hierarchy
upvoted 4 times

⊟ 👤 **dangal95** 4 years, 2 months ago
Answer is correct.
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet
upvoted 3 times

⊟ 👤 **Dark12arrow** 4 years, 2 months ago
do u have any reference ? and if u cant use parquet to load parquet files whats the point of ever choosing parquet?
upvoted 2 times

⊟ 👤 **eliabsbueno** 4 years, 2 months ago
The first box should be "Binary". You can't use a parquet data source to load different parquet files.
upvoted 3 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 an Azure subscription that contains an Azure Storage account.

You plan to implement changes to a data storage solution to meet regulatory and compliance standards.

Every day, Azure needs to identify and delete blobs that were NOT modified during the last 100 days.

Solution: You schedule an Azure Data Factory pipeline with a delete activity.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *A*

You can use the Delete Activity in Azure Data Factory to delete files or folders from on-premises storage stores or cloud storage stores.

Azure Blob storage is supported.

Note: You can also apply an Azure Blob storage lifecycle policy.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/delete-activity https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal

---

👤 **Nevia** 4 years, 2 months ago

Answer A seems to be correct https://azure.microsoft.com/it-it/blog/clean-up-files-by-built-in-delete-activity-in-azure-data-factory/

upvoted 4 times

HOTSPOT -

You have the following Azure Stream Analytics query.

```
WITH

step1 AS (SELECT *
    FROM input1
    PARTITION BY StateID
    INTO 10),
step2 AS (SELECT *
    FROM input2
    PARTITION BY StateID
    INTO 10)

SELECT *
INTO output
FROM step1
PARTITION BY StateID
UNION
SELECT * INTO output
    FROM step2
    PARTITION BY StateID
```

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Statements	Yes	No
The query combines two streams of partitioned data.	○	○
The stream scheme key and count must match the output scheme.	○	○
Providing 60 streaming units will optimize the performance of the query.	○	○

Suggested Answer:

Answer Area

Statements	Yes	No
The query combines two streams of partitioned data.	●	○
The stream scheme key and count must match the output scheme.	●	○
Providing 60 streaming units will optimize the performance of the query.	●	○

Box 1: Yes -

You can now use a new extension of Azure Stream Analytics SQL to specify the number of partitions of a stream when reshuffling the data.

The outcome is a stream that has the same partition scheme. Please see below for an example:

WITH step1 AS (SELECT * FROM [input1] PARTITION BY DeviceID INTO 10), step2 AS (SELECT * FROM [input2] PARTITION BY DeviceID INTO 10)

SELECT * INTO [output] FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID

Note: The new extension of Azure Stream Analytics SQL includes a keyword INTO that allows you to specify the number of partitions for a stream when performing reshuffling using a PARTITION BY statement.

Box 2: Yes -

When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count.

Box 3: Yes -

Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs,

the more CPU and memory resources are allocated for your job.

In general, the best practice is to start with 6 SUs for queries that don't use PARTITION BY.

Here there are 10 partitions, so 6x10 = 60 SUs is good.

Note: Remember, Streaming Unit (SU) count, which is the unit of scale for Azure Stream Analytics, must be adjusted so the number of physical resources available to the job can fit the partitioned flow. In general, six SUs is a good number to assign to each partition. In case there are insufficient resources assigned to the job, the system will only apply the repartition if it benefits the job.

Reference:

https://azure.microsoft.com/en-in/blog/maximize-throughput-with-repartitioning-in-azure-stream-analytics/ https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

---

👤 **Amy007** `Highly Voted 👍` 4 years, 1 month ago

Yes , Yes and No.

upvoted 8 times

👤 **Chemmangat** `Most Recent ⊘` 1 year, 9 months ago

From DP 203 discussions, I found out the answer is Yes for all.

upvoted 1 times

👤 **hosan** 4 years ago

What is the correct answer?? Pl advise.

upvoted 1 times

> 👤 **captainbee** 3 years, 12 months ago
>
> Doesn't matter, no-one can seem to agree on the answer on here either
>
> upvoted 1 times

👤 **Williammm** 4 years ago

What is the answer?? WHy do they not give the correct answer in the first place..what is their problem?

]

upvoted 1 times

👤 **qiyan1982** 4 years ago

NO, YES, YES.

upvoted 3 times

👤 **jasonoubre** 4 years ago

After the survey, I have no ideas. Can I know what is the answer?

upvoted 1 times

👤 **akod** 4 years, 2 months ago

This question is repetitive.

upvoted 2 times

> 👤 **AngelRio** 4 years, 2 months ago
>
> It's different. My options are Yes, No and No (Source Whizlabs)
>
> upvoted 2 times

> 👤 **Nobody99** 4 years, 1 month ago
>
> Its not.
>
> Question 45: The query joins ...
>
> Question 59: The query combines ...
>
> upvoted 7 times

> > 👤 **cadio30** 4 years, 1 month ago
> >
> > The query from both items are different. Question # 45 doesn't have "partition" syntax that makes it incorrect. On the other hand, there are two select statements that is combined.
> >
> > upvoted 2 times

DRAG DROP -

You need to create an Azure Cosmos DB account that will use encryption keys managed by your organization.

Which four 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.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Select and Place:

Actions

- Generate a new key in the Azure key vault
- Create an Azure key vault and enable purge protection
- Create a new Azure Cosmos DB account and set Data Encryption to **Service-managed key**
- Add an Azure Key Vault access policy to grant permissions to the Azure Cosmos DB principal
- Create a new Azure Cosmos DB account, set Data Encryption to **Customer-managed key (Enter key URI)**, and enter the key URI

Answer Area

Suggested Answer:

Actions

- Create a new Azure Cosmos DB account and set Data Encryption to **Service-managed key**

Answer Area

1. Create an Azure key vault and enable purge protection
2. Create a new Azure Cosmos DB account, set Data Encryption to **Customer-managed key (Enter key URI)**, and enter the key URI
3. Add an Azure Key Vault access policy to grant permissions to the Azure Cosmos DB principal
4. Generate a new key in the Azure key vault
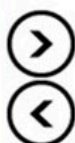
Step 1: Create an Azure key vault and enable purge protection

Using customer-managed keys with Azure Cosmos DB requires you to set two properties on the Azure Key Vault instance that you plan to use to host your encryption keys: Soft Delete and Purge Protection.

Step 2: Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed Key (Enter key URI), and enter the key URI

Data stored in your Azure Cosmos account is automatically and seamlessly encrypted with keys managed by Microsoft (service-managed keys). Optionally, you can choose to add a second layer of encryption with keys you manage (customer-managed keys).

Step 3: Add an Azure Key Vault access policy to grant permissions to the Azure Cosmos DB principal

Add an access policy to your Azure Key Vault instance

Step 4: Generate a new key in the Azure key vault

Generate a key in Azure Key Vault

Reference:

https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-setup-cmk

Wendy_DK `Highly Voted 👍` 4 years, 2 months ago

Step 1 Create an Azure Key vault and enable purge protection

Step 2 Add an Azure Key Vault access policy to grant permission to the Azure Cosmos DB principal

Step 3 Generate a new key in Azure Key Vault

Step 4 Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed key (Enter Key URI), and enter the key URI

upvoted 37 times

vrmei 4 years ago

Perfect. Microsoft.DocumentDB Resouce Provider to be registerd and then all the steps mentioned here.

upvoted 1 times

vaio 4 years, 2 months ago

this solution is correct. Check documentation here:

https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-setup-cmk

upvoted 2 times

vaseva1 `Highly Voted 👍` 4 years, 2 months ago

Step 1: Create an Azure key vault and enable purge protection

Step 2: Generate a new key in the Azure key vault

Step 3: Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed Key (Enter key URI), and enter the key URI

Step 4: Add an Azure Key Vault access policy to grant permissions to the Azure Cosmos DB principal

upvoted 22 times

cadio30 4 years, 2 months ago

this make sense. checked the documentation in the url below

https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-setup-cmk

upvoted 2 times

cadio30 4 years, 1 month ago

retracting my feedback here instead go for the solution below

Step 1 Create an Azure Key vault and enable purge protection

Step 2 Add an Azure Key Vault access policy to grant permission to the Azure Cosmos DB principal

Step 3 Generate a new key in Azure Key Vault

Step 4 Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed key (Enter Key URI), and enter the key URI

upvoted 4 times

nit687 `Most Recent ⊙` 4 years ago

right sequence looks like this :

Step 1 Create an Azure Key vault and enable purge protection

Step 2 Generate a new key in Azure Key Vault

Step 3 Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed key (Enter Key URI), and enter the key URI

Step 4 Add an Azure Key Vault access policy to grant permission to the Azure Cosmos DB principal

In discussions there is confusion going on whether step 4 should come above step 3 etc..but unless we dont create a cosmos DB resource , how can we create key vault access policy and grant permission to cosmos DB principal.so step 4 should be last

upvoted 2 times

hoangton 4 years, 1 month ago

Step 1:Create an Azure key vault and enable purge protection

Step 2:Add an Azure Key Vault access policy to grant permissions to the Azure Cosmos DB principal

Step 3:Generate a new key in the Azure key vault

Step 4:Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed Key (Enter key URI), and enter the key URI

upvoted 1 times

MMM777 4 years, 1 month ago

Step 1: Create an Azure key vault and enable purge protection

Step 2: Add an Azure Key Vault access policy to grant permissions to the Azure Cosmos DB principal (doesn't have to actually exist yet)

Step 3: Generate a new key in the Azure key vault

Step 4: Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed Key (Enter key URI), and enter the key URI

https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-setup-cmk

⊟ 👤 **dangal95** 4 years, 2 months ago

These are the correct steps:

Step 1 Create an Azure Key vault and enable purge protection
Step 2 Generate a new key in Azure Key Vault
Step 3 Create a new Azure Cosmos DB account, set Data Encryption to Customer-managed key (Enter Key URI), and enter the key URI
Step 4 Add an Azure Key Vault access policy to grant permission to the Azure Cosmos DB principal

You cannot add the key URI before you've even created the key so creating the Cosmos DB account AND inserting the key uri before the key even exists does not make sense. Also, you cannot add an access policy for a resource that does not exist yet so adding the access policy to the key vault before you even created the cosmos DB account does not make sense.

⊟ 👤 **alf99** 4 years, 2 months ago

The Cosmos DB account must be created as last step using previous created key. MS docs states that:

"When you create a new Azure Cosmos DB account from the Azure portal, choose Customer-managed key in the Encryption step. In the Key URI field, paste the URI/key identifier of the Azure Key Vault key that you copied from the previous step"

https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-setup-cmk

You have an Azure Storage account named storage1 that is configured as shown in the following exhibit.

Account kind
StorageV2 (general purpose v2)

Performance ⓘ
⦿ Standard ○ Premium
  ⓘ This setting cannot be changed after the storage account is created.

Secure transfer required ⓘ
⦿ Disabled ○ Enabled

Allow Blob public access ⓘ
○ Disabled ⦿ Enabled

Minimum TLS version ⓘ
| Version 1.0   ⌄ |

Blob access tier (default) ⓘ
○ Cool ⦿ Hot

Replication ⓘ
| Read-access geo-redundant storage (RA-GRS)   ⌄ |

Large file shares ⓘ
⦿ Disabled ○ Enabled

  ⓘ The current combination of subscription, storage account kind, performance,
  replication and location does not support large file shares.

Identity-based access for file shares
Azure Active Directory Domain Services (Azure AD DS) ⓘ
⦿ Disabled ○ Enabled

You need to ensure that all calls to an Azure Storage REST API operation on storage1 are made over HTTPS.
What should you do?

    A. Set Secure transfer required to Enabled.

    B. Set Allow Blob public access to Disabled.

    C. For the Blob service, create a shared access signature (SAS) that allows HTTPS only.

    D. Set Minimum TLS version to Version 1.2.

Suggested Answer: *A*

You can configure your storage account to accept requests from secure connections only by setting the Secure transfer required property for the storage account.
When you require secure transfer, any requests originating from an insecure connection are rejected. Microsoft recommends that you always require secure transfer for all of your storage accounts.
When secure transfer is required, a call to an Azure Storage REST API operation must be made over HTTPS. Any request made over HTTP is rejected.
Reference:
https://docs.microsoft.com/en-us/azure/storage/common/storage-require-secure-transfer

---

👤 **Podavenna** 3 years, 9 months ago

Correct Answer!

  upvoted 2 times

HOTSPOT -

You have an Azure Synapse Analytics dedicated SQL pool that contains the users shown in the following table.

Name	Role
User1	Server admin
User2	db_datareader

User1 executes a query on the database, and the query returns the results shown in the following exhibit.

```
1   SELECT c.name,
2       tbl.name as table_name,
3       typ.name as datatype,
4       c.is_masked,
5       c.masking_function
6   FROM sys.masked_columns AS c
7   INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
8   INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
9   WHERE is_masked = 1;
10
```

Results Messages

	name	table_name	datatype	is_masked	masking_function
1	BirthDate	DimCustomer	date	1	default()
2	Gender	DimCustomer	nvarchar	1	default()
3	EmailAddress	DimCustomer	nvarchar	1	email()
4	YearlyIncome	DimCustomer	money	1	default()

User1 is the only user who has access to the unmasked data.

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.

Hot Area:

Answer Area

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

a random number
the values stored in the database
XXXX
0

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

a random date
the values stored in the database
XXXX
1900-01-01

Answer Area

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

a random number
the values stored in the database
XXXX
0

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

a random date
the values stored in the database
XXXX
1900-01-01

Box 1: 0 -

The YearlyIncome column is of the money data type.

The Default masking function: Full masking according to the data types of the designated fields

☞ Use a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).

Box 2: the values stored in the database

Users with administrator privileges are always excluded from masking, and see the original data without any mask.

Reference:

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

---

🗏 👤 **marc88** 3 years, 10 months ago

The second question involves the admin user, so the answer is correct

upvoted 4 times

🗏 👤 **BlackBeesN** 3 years, 11 months ago

Zero and 01-01-1900

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

upvoted 3 times

🗏 👤 **tendaitakas** 3 years, 10 months ago

No. The provided answer is correct. It should show the value stored in the database and based on the graphic, we dont know what that value is

upvoted 2 times

Note: This question is a part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to an enterprise data warehouse in Azure Synapse Analytics. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create a workload group using the Azure storage account name as the pool name

3. Load the data using the INSERT`¦SELECT statement

Does the solution meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

You need to create an external file format and external table using the external data source.

You then load the data using the CREATE TABLE AS SELECT statement.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

---

👤 **certstowinirl** 3 years, 8 months ago

This is correct, please refer to: https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

  upvoted 1 times

You develop data engineering solutions for a company.

You must integrate the company's on-premises Microsoft SQL Server data with Microsoft Azure SQL Database. Data must be transformed incrementally.

You need to implement the data integration solution.

Which tool should you use to configure a pipeline to copy data?

    A. Use the Copy Data tool with Blob storage linked service as the source

    B. Use Azure PowerShell with SQL Server linked service as a source

    C. Use Azure Data Factory UI with Blob storage linked service as a source

    D. Use the .NET Data Factory API with Blob storage linked service as the source

Suggested Answer: *C*

The Integration Runtime is a customer managed data integration infrastructure used by Azure Data Factory to provide data integration capabilities across different network environments.

A linked service defines the information needed for Azure Data Factory to connect to a data resource. We have three resources in this scenario for which linked services are needed:

☞ On-premises SQL Server

☞ Azure Blob Storage

☞ Azure SQL database

Note: Azure Data Factory is a fully managed cloud-based data integration service that orchestrates and automates the movement and transformation of data. The key concept in the ADF model is pipeline. A pipeline is a logical grouping of Activities, each of which defines the actions to perform on the data contained in

Datasets. Linked services are used to define the information needed for Data Factory to connect to the data resources.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-sql-azure-adf

---

⊟ 👤 **brcdbrcd** `Highly Voted 👍` 4 years, 7 months ago

The answer is B. Use Azure PowerShell with SQL Server linked service as a source

https://docs.microsoft.com/en-us/azure/data-factory/tutorial-incremental-copy-multiple-tables-powershell

upvoted 35 times

    ⊟ 👤 **tankzzz** 4 years, 2 months ago

    This is the answer. 100%

    upvoted 5 times

⊟ 👤 **BungyTex** `Highly Voted 👍` 4 years, 10 months ago

C says the linked service is Blob storage. The question says it coming from SQL Server. Why would you use Blob storage as the source?

upvoted 7 times

⊟ 👤 **111222333** `Most Recent ⊙` 4 years, 1 month ago

Correct answer is B.

1. Both *Azure Data Factory UI (Azure Portal)* and *Azure PowerShell* can be used to incrementally load data from multiple tables in SQL Server to a database in Azure SQL Database.
2. Both *Azure Data Factory UI (Azure Portal)* and *Azure PowerShell* can use an SQL Server linked service as a source for this incremental copy.

Evidence for these two statements:
- *ADF GUI*: https://docs.microsoft.com/en-us/azure/data-factory/tutorial-incremental-copy-multiple-tables-portal
- *PowerShell*: https://docs.microsoft.com/en-us/azure/data-factory/tutorial-incremental-copy-multiple-tables-powershell

Hence, use *Azure Data Factory UI (Azure Portal) with SQL Server linked service* or use *Azure PowerShell with SQL Server linked service*.

Since *ADF with SQL Server linked service* is not an option among the answers ("Blob storage linked service as a source" in answer C is really not necessary), the correct answer is B: Use Azure PowerShell with SQL Server linked service as a source.

upvoted 5 times

👤 **Qrm_1972** 4 years, 1 month ago

The correct answer is : 100% is C

upvoted 2 times

👤 **cadio30** 4 years, 2 months ago

This should be ADF

upvoted 1 times

> 👤 **cadio30** 4 years, 1 month ago
>
> Disregard this and the answer is 'B' as the other connection string relies on azure blob storage and the source is from on-prem SQL Server
>
> upvoted 1 times

👤 **Sai2609** 4 years, 2 months ago

The answer is B since the catch is the movement of data incrementally which can be done easily through powershell

upvoted 3 times

👤 **Nevia** 4 years, 2 months ago

In my opinion the key word is "pipeline". ADF UI is the only one that creates a pipeline

upvoted 1 times

> 👤 **maciejt** 4 years, 1 month ago
>
> ADF is not good with incremental or delta loads. Besides all anwers except B refers to Blob as source and the source is on prem SQL, so it can only be B
>
> upvoted 1 times

👤 **felmasri** 4 years, 3 months ago

looks like the question is based on a scenario presented in one of Microsoft help docs :

https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-sql-azure-adf

the scenario assuming the following:

The Scenario

We set up an ADF pipeline that composes two data migration activities. Together they move data on a daily basis between a SQL Server database and Azure SQL Database. The two activities are:

Copy data from a SQL Server database to an Azure Blob Storage account

Copy data from the Azure Blob Storage account to Azure SQL Database.

upvoted 2 times

👤 **dumpsm42** 4 years, 6 months ago

hi to all,

the text says "...tool..." so for me that leaves out B and D.

answer C seems right but the source is SQL Server onprem, not the blob storage !

so for me it's A because its source is SQL onPrem and because of this link:

https://docs.microsoft.com/pt-pt/azure/data-factory/tutorial-hybrid-copy-data-tool

so it's A for me.

regards

upvoted 1 times

> 👤 **dumpsm42** 4 years, 6 months ago
>
> Under New Linked Service, search for SQL Server, and then select Continue.
>
> In the New Linked Service (SQL Server) dialog box, under Name, enter SqlServerLinkedService. Select +New under Connect via integration runtime. You must create a self-hosted integration runtime, download it to your machine, and register it with Data Factory. The self-hosted integration runtime copies data between your on-premises environment and the cloud.
>
> upvoted 1 times

👤 **anarvekar** 4 years, 9 months ago

Can someone please explain how blob storage comes into picture in this scenario? The data transformation is being carried out using ADF, and the data moves between on premise SQL Server instance and Azure SQL DB.

upvoted 3 times

> 👤 **poomazuretest** 4 years, 9 months ago
>
> ADF need staging area on cloud using Blob

upvoted 3 times

 big_data_au 4 years, 7 months ago

Not if you are using a self hosted integration runtime - ADF can draw directly from on-prem SQL

upvoted 4 times

 maciejt 4 years, 1 month ago

I confirm, ADF copy activity can draw from on prem, data flows need staging on cloud as they run on IR

upvoted 1 times

 big_data_au 4 years, 7 months ago

Not if you are using a self hosted integration runtime - ADF can draw directly from on-prem SQL

upvoted 4 times

 maciejt 4 years, 1 month ago

I confirm, ADF copy activity can draw from on prem, data flows need staging on cloud as they run on IR

upvoted 1 times

HOTSPOT -

A company runs Microsoft Dynamics CRM with Microsoft SQL Server on-premises. SQL Server Integration Services (SSIS) packages extract data from Dynamics

CRM APIs, and load the data into a SQL Server data warehouse.

The datacenter is running out of capacity. Because of the network configuration, you must extract on premises data to the cloud over https. You cannot open any additional ports. The solution must implement the least amount of effort.

You need to create the pipeline system.

Which component should you use? To answer, select the appropriate technology in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Action	Technology
Extract SQL data on-premises	Self-hosted integration runtime / Azure-SSIS integration runtime / Azure integration runtime / Source
Load SQL data warehouse	Self-hosted integration runtime / Azure-SSIS integration runtime / Azure integration runtime / Sink

Suggested Answer:

Action	Technology
Extract SQL data on-premises	Self-hosted integration runtime / Azure-SSIS integration runtime / Azure integration runtime / **Source**
Load SQL data warehouse	**Self-hosted integration runtime** / Azure-SSIS integration runtime / Azure integration runtime / Sink

Box 1: Source -

For Copy activity, it requires source and sink linked services to define the direction of data flow.

Copying between a cloud data source and a data source in private network: if either source or sink linked service points to a self-hosted IR, the copy activity is executed on that self-hosted Integration Runtime.

Box 2: Self-hosted integration runtime

A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network, and it can dispatch transform activities against compute resources in an on-premises network or an Azure virtual network. The installation of a self-hosted integration runtime needs on an on-premises machine or a virtual machine (VM) inside a private network.

References:

https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

---

⊟ 👤 **Pairon** `Highly Voted 👍` 4 years, 2 months ago

For me are:

- Self Hosted IR

- Sink

upvoted 17 times

⊟ 👤 **Williammm** `Highly Voted 👍` 4 years ago

So...what is the answer??? so anoying!

upvoted 7 times

Nnuujj `Most Recent ⊙` 1 year, 10 months ago

For me:

Extract SQL data on-premises: Use Azure-SSIS Integration runtime to run SSIS packages in ADF to extract data from Dynamics.

Load SQL data warehouse: Use Self-hosted integration runtime to perform operations between on premises and azure. In this case loading data using ADF to SQL Server data warehouse

upvoted 1 times

Marcus1612 3 years, 8 months ago

Bloc 1= Azure SSIS Integration runtine".

Bloc 2 = Azure SSIS Integration runtine".

The main problem is that The datacenter is running out of capacity. So we just need the cloud infrastructure with the underlying ressources (cpu, memory) to transfert data between the CRM and the Data warehouse. Since an existing SSIS package exists already, the minimum effort would be to lift and shift this exiting SSIS package to the cloud. To do that, the technology behind the scene is "Azure SSIS Integration runtine".

upvoted 3 times

eng1 4 years ago

The answer is correct!

upvoted 1 times

vrmei 4 years ago

Self Hosted IR for both answer

https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

upvoted 2 times

hoangton 4 years, 1 month ago

-Self Hosted

-Self Hosted

upvoted 2 times

cadio30 4 years, 2 months ago

The scenario retrieves data from on-prem database and loads it to on-prem data warehouse.

ADF function would be to orchestrate data and their link services would use "SELF-HOSTED" and this requires to implement gateway configuration on the server where the on-prem resides. Therefore, both SELF-HOSTED is the appropriate answer

upvoted 4 times

> **cadio30** 4 years, 1 month ago
>
> On second though, the question states that there is a SSIS involve and it requires minimal effort, hence, Azure SSIS and Self Hosted are the answers
>
> upvoted 2 times

Aragorn_2021 4 years, 2 months ago

- Self Hosted IR (to transfer the data from on premise to Azure)

- Azure-SSIS (to take care of the load from source to target warehouse)

upvoted 7 times

Prabhakaran94 4 years, 2 months ago

Correct Answer is:

- Azure-SSIS integration runtime

- Azure integration runtime

upvoted 4 times

LuBarba 4 years, 2 months ago

Correct answer is:

- Self Hosted IR

- Sink

Source: https://docs.microsoft.com/en-us/azure/data-factory/tutorial-hybrid-copy-data-tool

upvoted 4 times

> **memo43** 4 years, 1 month ago
>
> source is irrelevant. for me its;
>
> self host ir
>
> azure ssis
>
> upvoted 3 times

tucho 4 years, 2 months ago

I'm not sure that "Self Hosted IR" can be the right choice for the first option.

Keep in mind that on-premise data-center is already running "SQL Server Integration Services (SSIS)".

Some related doc: https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

upvoted 3 times

DRAG DROP -

You develop data engineering solutions for a company.

A project requires analysis of real-time Twitter feeds. Posts that contain specific keywords must be stored and processed on Microsoft Azure and then displayed by using Microsoft Power BI. You need to implement the solution.

Which five 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.

Select and Place:

Actions

Create an HDInsight cluster with the Hadoop cluster type.

Create a Jupyter Notebook.

Run a job that uses the Spark Streaming API to ingest data from Twitter.

Create a Runbook.

Create an HDInsight cluster with the Spark cluster type.

Create an table.

Load the hvac table into Power BI Desktop.

Answer Area

Suggested Answer:

Actions

Create an HDInsight cluster with the Hadoop cluster type.

Create a Jupyter Notebook.

Run a job that uses the Spark Streaming API to ingest data from Twitter.

Create a Runbook.

Create an HDInsight cluster with the Spark cluster type.

Create an table.

Load the hvac table into Power BI Desktop.

Answer Area

Create an HDInsight cluster with the Spark cluster type.

Create a Jupyter Notebook.

Create an table.

Run a job that uses the Spark Streaming API to ingest data from Twitter.

Load the hvac table into Power BI Desktop.

Step 1: Create an HDInisght cluster with the Spark cluster type

Step 2: Create a Jyputer Notebook

Step 3: Create a table -

The Jupyter Notebook that you created in the previous step includes code to create an hvac table.

Step 4: Run a job that uses the Spark Streaming API to ingest data from Twitter

Step 5: Load the hvac table into Power BI Desktop

You use Power BI to create visualizations, reports, and dashboards from the Spark cluster data.

References:

https://acadgild.com/blog/streaming-twitter-data-using-spark

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-use-with-data-lake-store

☐ 👤 **cadio30** `Highly Voted 👍` 4 years, 2 months ago

The propose solution is correct.

A table cannot be created if the notebook is not yet available, the scenario is in assumption the table is within the hdinsight spark cluster.

upvoted 6 times

    □ **cadio30** 4 years, 1 month ago

    Reference: https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-load-data-run-query

    upvoted 2 times

□ **Aragorn_2021** `Most Recent ⊘` 4 years, 2 months ago

Agree with Pairon. Ideal to have Target table table created first before Jupyter notebook

upvoted 1 times

□ **Pairon** 4 years, 2 months ago

I agree with the answer, but maybe we can swapp second and third step?

upvoted 1 times

□ **tucho** 4 years, 2 months ago

With the proposed solution, "who" runs the Jupyter notebook? :-(

upvoted 1 times

□ **JohnCrawford** 4 years, 2 months ago

answer appears correct. https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-use-bi-tools

upvoted 3 times

DRAG DROP -

Your company manages on-premises Microsoft SQL Server pipelines by using a custom solution.

The data engineering team must implement a process to pull data from SQL Server and migrate it to Azure Blob storage. The process must orchestrate and manage the data lifecycle.

You need to configure Azure Data Factory to connect to the on-premises SQL Server database.

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.

Select and Place:

Answer Area

Actions

Create an Azure Data Factory resource.

Configure a self-hosted integration runtime.

Create a virtual private network (VPN)connection from on-premises to Microsoft Azure.

Create a database master key on SQL Server.

Backup the database and send it Azure Blob storage.

Configure the on-premises SQL Server instance with an integration runtime.

Answer Area

Suggested Answer:

Answer Area

Actions

Create an Azure Data Factory resource.

Configure a self-hosted integration runtime.

Create a virtual private network (VPN)connection from on-premises to Microsoft Azure.

Create a database master key on SQL Server.

Backup the database and send it Azure Blob storage.

Configure the on-premises SQL Server instance with an integration runtime.

Answer Area

Create a virtual private network (VPN)connection from on-premises to Microsoft Azure.

Create an Azure Data Factory resource.

Configure a self-hosted integration runtime.

Step 1: Create a virtual private network (VPN) connection from on-premises to Microsoft Azure.

You can also use IPSec VPN or Azure ExpressRoute to further secure the communication channel between your on-premises network and Azure.

Azure Virtual Network is a logical representation of your network in the cloud. You can connect an on-premises network to your virtual network by setting up IPSec

VPN (site-to-site) or ExpressRoute (private peering).

Step 2: Create an Azure Data Factory resource.

Step 3: Configure a self-hosted integration runtime.
You create a self-hosted integration runtime and associate it with an on-premises machine with the SQL Server database. The self-hosted integration runtime is the component that copies data from the SQL Server database on your machine to Azure Blob storage.
Note: A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network, and it can dispatch transform activities against compute resources in an on-premises network or an Azure virtual network. The installation of a self-hosted integration runtime needs on an on- premises machine or a virtual machine (VM) inside a private network.
References:
https://docs.microsoft.com/en-us/azure/data-factory/tutorial-hybrid-copy-powershell

---

cnuusd `Highly Voted 👍` 5 years, 5 months ago
Create an Azure Data Factory
Configure a self-hosted integration runtime
Configure on-premises SQL Server Instance ....
upvoted 71 times

> **avestabrzn** 5 years, 4 months ago
> totally agree
> upvoted 5 times

> **zenomas** 5 years, 3 months ago
> curious, what sort of configuration do we need in on-prem SQL Server instance? creating user account and so on?
> upvoted 2 times

>> **Yuri1101** 5 years, 2 months ago
>> Just the installation of the IR in the on-premise server. I guess it can be called configuration as well.
>> upvoted 1 times

>>> **BRW** 4 years, 5 months ago
>>> You may need to create a db user to use in a link service in ADF.
>>> upvoted 1 times

epgd `Highly Voted 👍` 5 years, 7 months ago
It's not necessary to Create a virtual private network (VPN) connection from on-premises to Microsoft Azure.
upvoted 23 times

> **John123123** 5 years, 3 months ago
> "The installation of a self-hosted integration runtime needs an on-premises machine or a virtual machine inside a private network." - https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime
>
> Implies that a virtual private network is indeed required.
> upvoted 8 times

>> **zenomas** 5 years, 3 months ago
>> I think it is not required when the machine is internet-facing with connectivity to on-prem SQL Server.
>> upvoted 1 times

>>> **visakh** 5 years, 1 month ago
>>> Its not madatory. But its a security best practice to use VPN
>>>
>>> See
>>>
>>> https://docs.microsoft.com/en-us/azure/data-factory/data-movement-security-considerations
>>> upvoted 1 times

>> **cowtown** 4 years, 12 months ago
>> The company may have ExpressRoute setup.
>> upvoted 3 times

cadio30 `Most Recent ⊘` 4 years, 2 months ago
Create an Azure Data Factory
Configure a self-hosted integration runtime
Configure on-premises SQL Server Instance

These steps leads to installation of gateway to the on-prem server that links to ADF (Self-Hosted IR)

upvoted 1 times

sonu_agrawal 4 years, 3 months ago

What is the correct answer for this?

upvoted 1 times

joguerra 4 years, 5 months ago

You don't need to pick "Configure on-premises SQL Server Instance ... " because self-host runtime already does that. Look at the tuturial, you don't need to touch anything related to SQL Server Instance, you simply create a linked service through the self-host IR. I would say it is redundant.

Creating a VPN on the otherside, (although not mandatory) makes the solution more complete.

upvoted 4 times

syu31svc 4 years, 7 months ago

I would say the answer given is correct

VPN to establish connection as the first step followed by creating Azure Data Factory and configure self-hosted integration runtime

upvoted 3 times

groy 4 years, 9 months ago

1: Deploy an Azure Data Factory

2: From the on-premises network, install and configure a self-hosted runtime.

3: Configure a linked service to connect to the SQL Server instance.

upvoted 2 times

BRW 4 years, 5 months ago

Without link service there is no way to connect SQL Server from ADF. I guess this option is missing.

upvoted 1 times

ExamPwnr 4 years, 9 months ago

1. Create an Azure Data Factory resource

2. Configure a self-hosted integration runtime (configure means that you have ADF already running)

3. Configure the on-premises SQL server instance with an integration runtime (means that you go into ADF and connect to the database using IR)

You don't need VPN - all communication from IR to ADF is over HTTPS

upvoted 7 times

dumpsm42 4 years, 6 months ago

hi to all, this is the right answer 4 sure. please see this link, all the steps are there: https://docs.microsoft.com/pt-pt/azure/machine-learning/team-data-science-process/move-sql-azure-adf

vpn no needed, the IR onpremises uses https as default

regards

upvoted 1 times

dumpsm42 4 years, 6 months ago

Create an Azure Data Factory

The instructions for creating a new Azure Data Factory and a resource group in the Azure portal are provided Create an Azure Data Factory. Name the new ADF instance adfdsp and name the resource group created adfdsprg.

Install and configure Azure Data Factory Integration Runtime

The Integration Runtime is a customer-managed data integration infrastructure used by Azure Data Factory to provide data integration capabilities across different network environments. This runtime was formerly called "Data Management Gateway".

To set up, follow the instructions for creating a pipeline

Create linked services to connect to the data resources

A linked service defines the information needed for Azure Data Factory to connect to a data resource. We have three resources in this scenario for which linked services are needed:

On-premises SQL Server

Azure Blob Storage

Azure SQL Database

upvoted 1 times

□ 👤 **Abhitm** 5 years ago

1: Create an Azure Data Factory

2: Configure a self-hosted integration runtime

3: Backup the DB and send it to Azure Blob storage

upvoted 4 times

□ 👤 **maciejt** 4 years, 1 month ago

That's why you have ADF to use copy activity directly from table to blob, and you don't need to manually copy a backup to storage

upvoted 1 times

□ 👤 **Luke97** 5 years, 1 month ago

VPN is not a mandatory requirement for self-hosted IR. It is only needed when to perform data integration securely in a private network environment, which doesn't have a direct line-of-sight from the public cloud environment. (https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime).

So, the correct answers should be

1. Create an ADF;

2. Create a self-hosted IR;

3. Install and configure self-hosted IR on on-premise server;

upvoted 5 times

□ 👤 **TheCyanideLancer** 5 years, 1 month ago

Is VPN really a necessity?

upvoted 1 times

□ 👤 **kilowd** 4 years, 10 months ago

The installation of a self-hosted integration runtime needs an on-premises machine or a virtual machine inside a private network.

upvoted 1 times

□ 👤 **zenomas** 5 years, 3 months ago

Step 1 and 2 can be in any order.

upvoted 2 times

□ 👤 **AAJ** 5 years, 3 months ago

Given answer is correct

upvoted 5 times

□ 👤 **MilindD** 5 years, 4 months ago

correct VPN is not needed

upvoted 2 times

□ 👤 **kilowd** 4 years, 10 months ago

The installation of a self-hosted integration runtime needs an on-premises machine or a virtual machine inside a private network.

upvoted 1 times

□ 👤 **epgd** 5 years, 7 months ago

The answer could be

Step 1: Create an Azure Data Factory

You need to create a data factory and start the Data Factory UI to create a pipeline in the data factory.

Step 2: Configure a self-hosted integration runtime

Step 3: Configure a self-hosted integration runtime.

?¿?

upvoted 6 times

□ 👤 **Kampai787** 4 years, 8 months ago

Tu ere loko o k

upvoted 9 times

HOTSPOT -

You are designing a new Lambda architecture on Microsoft Azure.

The real-time processing layer must meet the following requirements:

Ingestion:

☞ Receive millions of events per second

☞ Act as a fully managed Platform-as-a-Service (PaaS) solution

☞ Integrate with Azure Functions

Stream processing:

☞ Process on a per-job basis

☞ Provide seamless connectivity with Azure services

☞ Use a SQL-based query language

Analytical data store:

☞ Act as a managed service

☞ Use a document store

☞ Provide data encryption at rest

You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Architecture requirement	Technology
Ingestion	▼
	HDInsight Kafka
	Azure Event Hubs
	HDInsight Storm
	HDInsight Spark
Stream Processing	▼
	Azure Stream Analytics
	HDInsight with Spark Streaming
	Azure Cosmos DB Change Feed
	Azure Analysis Services
Analytical Data Store	▼
	Hive LLAP on HDInsight
	Azure Analysis Services
	Azure Cosmos DB
	Azure Synapse Analytics

Answer Area

Architecture requirement	Technology
Ingestion	HDInsight Kafka
	Azure Event Hubs
	HDInsight Storm
	HDInsight Spark
Stream Processing	**Azure Stream Analytics**
	HDInsight with Spark Streaming
	Azure Cosmos DB Change Feed
	Azure Analysis Services
Analytical Data Store	Hive LLAP on HDInsight
	Azure Analysis Services
	Azure Cosmos DB
	Azure Synapse Analytics

Box 1: Azure Event Hubs -

This portion of a streaming architecture is often referred to as stream buffering. Options include Azure Event Hubs, Azure IoT Hub, and Kafka.

Incorrect Answers: Not HDInsight Kafka

Azure Functions need a trigger defined in order to run. There is a limited set of supported trigger types, and Kafka is not one of them.

Box 2: Azure Stream Analytics -

Azure Stream Analytics provides a managed stream processing service based on perpetually running SQL queries that operate on unbounded streams.

You can also use open source Apache streaming technologies like Storm and Spark Streaming in an HDInsight cluster.

Box 3: Azure Synapse Analytics -

Azure Synapse Analytics provides a managed service for large-scale, cloud-based data warehousing. HDInsight supports Interactive Hive, HBase, and Spark

SQL, which can also be used to serve data for analysis.

Reference:

https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/

☐ 👤 **samkslee** `Highly Voted 👍` 4 years, 2 months ago

The analytical data store should be Azure Cosmos DB, as the question said "use a document store".

upvoted 33 times

☐ 👤 **ak08** `Highly Voted 👍` 4 years, 2 months ago

Analytical data store should be Cosmos DB, since, Synapse analytics is not a document store.

upvoted 13 times

☐ 👤 **josegv** `Most Recent ⊘` 4 years ago

Agree. Analytical store should be Cosmos DB

upvoted 1 times

☐ 👤 **Saravjeet** 4 years, 1 month ago

it should be cosmos DB

upvoted 1 times

☐ 👤 **cadio30** 4 years, 2 months ago

Same sentiment with the Analytics data store should be Azure Cosmos DB

upvoted 2 times

☐ 👤 **Internet_User** 4 years, 2 months ago

Agree with samkslee & ak08 - Synapse is not a document store, Cosmos DB is.

You develop data engineering solutions for a company.

You need to ingest and visualize real-time Twitter data by using Microsoft Azure.

Which three technologies should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

    A. Event Grid topic

    B. Azure Stream Analytics Job that queries Twitter data from an Event Hub

    C. Azure Stream Analytics Job that queries Twitter data from an Event Grid

    D. Logic App that sends Twitter posts which have target keywords to Azure

    E. Event Grid subscription

    F. Event Hub instance

Suggested Answer: *BDF*

You can use Azure Logic apps to send tweets to an event hub and then use a Stream Analytics job to read from event hub and send them to PowerBI.

References:

https://community.powerbi.com/t5/Integrations-with-Files-and/Twitter-streaming-analytics-step-by-step/td-p/9594

👤 **zenomas** `Highly Voted 👍` 5 years, 3 months ago

Same as my thoughts. BDF

upvoted 12 times

👤 **cowtown** `Highly Voted 👍` 4 years, 12 months ago

Shouldn't be EventGrid because that is for subscribing to Azure events, ie. New blob in a container.

upvoted 5 times

   👤 **cadio30** 4 years, 1 month ago

   This should clear up the differences.

   Reference: https://www.cognizantsoftvision.com/blog/azure-event-grid-vs-event-hubs/

   upvoted 1 times

👤 **Ambujinee** `Most Recent ⊙` 4 years, 1 month ago

My ans is BDF

upvoted 2 times

👤 **kaigalmane** 4 years, 1 month ago

Azure Event Hubs is a more suitable solution when we need a service that can receive and process millions of events per second and provide low-latency event processing. It can handle data from concurrent sources and route it to a variety of stream-processing infrastructure and analytics services, as I have already mentioned. Azure Event Hubs are used more for telemetry scenarios.

On the other hand, Azure Event Grid is ideal for reactive scenarios, like when an item has been shipped or an item has been added or updated on storage. We have to take into account also its native integrations with Functions, Logic Apps and Webhooks. Moreover, Event Grid is cheaper than Event Hubs and more suitable when we don't have to deal with big data.

So I'll go with CDF

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 plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

☞ A workload for data engineers who will use Python and SQL

☞ A workload for jobs that will run notebooks that use Python, Scala, and SQL

☞ A workload that data scientists will use to perform ad hoc analysis in Scala and R

The enterprise architecture team at your company identifies the following standards for Databricks environments:

☞ The data engineers must share a cluster.

☞ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

☞ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

We would need a High Concurrency cluster for the jobs.

Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

References:

https://docs.azuredatabricks.net/clusters/configure.html

Community vote distribution

A (100%)

☺ **vaseva1** `Highly Voted 👍` 4 years, 2 months ago

A workload for jobs that will run notebooks that use Python, Scala, and SQL --> so Standard clusters because of Scala (High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.)

upvoted 13 times

    ☺ **Chiranjib** 4 years, 1 month ago

    Correct. Create New Cluster UI for Jobs allows either Standard or Single Node. It does list high concurrency as an option

    https://docs.microsoft.com/en-us/azure/databricks/clusters/create

    upvoted 1 times

☺ **sharma21** `Highly Voted 👍` 4 years, 2 months ago

A is correct

upvoted 7 times

☺ **FredNo** `Most Recent ⊙` 3 years, 7 months ago

`Selected Answer: A`

Standard for jobs and high concurrency for darta scientists and data engineers

upvoted 1 times

☺ **elimey** 3 years, 11 months ago

High Concurrency clusters can run workloads developed in SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

means job can not be high concurrency

so, ANSWER IS A
   upvoted 1 times

⊟ 👤 **cadio30** 4 years, 1 month ago
appropriate answer is A
   upvoted 3 times

⊟ 👤 **Wendy_DK** 4 years, 2 months ago
Correct answer is A
   upvoted 5 times

⊟ 👤 **Prabhakaran94** 4 years, 2 months ago
Correct answer is:
Yes
   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 plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

☞ A workload for data engineers who will use Python and SQL

☞ A workload for jobs that will run notebooks that use Python, Scala, and SQL

☞ A workload that data scientists will use to perform ad hoc analysis in Scala and R

The enterprise architecture team at your company identifies the following standards for Databricks environments:

☞ The data engineers must share a cluster.

☞ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

☞ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs.

Does this meet the goal?

A. Yes

B. No

Suggested Answer: *A*

We need a High Concurrency cluster for the data engineers and the jobs.
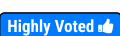
Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

References:

https://docs.azuredatabricks.net/clusters/configure.html

☐ 👤 **P_11** `Highly Voted 👍` 5 years ago

High Concurrency does not work for Scala...should B is correct answer

Note

High concurrency clusters work only for SQL, Python, and R. The performance, security, and fault isolation of high concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

The Table Access Control checkbox is available only for high concurrency cluster

https://docs.microsoft.com/en-gb/azure/databricks/clusters/configure

upvoted 31 times

☐ 👤 **MMM777** 4 years ago

actually Data Engineers do not require Scala - so High Concurrency is ok

upvoted 1 times

☐ 👤 **RyuHayabusa** 3 years, 11 months ago

What about reading the question again?

upvoted 1 times

☐ 👤 **Williammm** 4 years ago

it says that the will use scala...so A

upvoted 1 times

☐ 👤 **MLCL** 4 years, 11 months ago

Data Engineers do not use Scala.

upvoted 4 times

☐ 👤 **Ambujinee** 4 years, 1 month ago

Agreed with you

upvoted 1 times

☐ 👤 **ichacas** 5 years ago

"You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs"

Data engineers share same cluster and only use Python and SQL -> High Concurrency

Each Data Scientists needs a cluster and they are going to work with R and Scala -> Standard

Jobs are going to be executed with Scala -> Standard.

So correct answer is Yes

upvoted 39 times

☐ 👤 **HeywwooodJab** 4 years, 2 months ago

That's not what the question says ...

upvoted 6 times

☐ 👤 **Ambujinee** 4 years, 1 month ago

What you have explained is correct hence the ans is B. As the ans contain high concurrence cluster for job which is wrong.

upvoted 2 times

☐ 👤 **azmun** `Most Recent ⊘` 4 years, 1 month ago

Answer is No

High Concurrency clusters work only for SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

https://docs.microsoft.com/en-us/azure/databricks/clusters/configure

upvoted 1 times

☐ 👤 **cadio30** 4 years, 2 months ago

This statement "A workload for jobs that will run notebooks that use Python, Scala, and SQL" pertains to the propose solution "High Concurrency cluster for the jobs" and as we all know, high concurrency doesn't work for SCALA. Therefore, the answer is NO.

upvoted 2 times

☐ 👤 **sharma21** 4 years, 2 months ago

Answer is NO

upvoted 1 times

☐ 👤 **UmashankarJanakiraman** 4 years, 2 months ago

https://github.com/Azure/AzureDatabricksBestPractices/blob/master/Table2.PNG

Answer is NO

upvoted 1 times

☐ 👤 **Hassan_Mazhar_Khan** 4 years, 2 months ago

Correct Answer is B as High Concurrency does not work for Scala.

upvoted 1 times

☐ 👤 **brcdbrcd** 4 years, 7 months ago

No - because a single node cluster is appropriate for jobs. Not because of any other reasons...

A Single Node cluster has no workers and runs Spark jobs on the driver node. In contrast, Standard mode clusters require at least one Spark worker node in addition to the driver node to execute Spark jobs.

https://docs.microsoft.com/en-us/azure/databricks/clusters/configure#--single-node-clusters

upvoted 3 times

☐ 👤 **seaun** 4 years, 8 months ago

Jobs should be standard cluster, correct answer is Yes

upvoted 2 times

☐ 👤 **sandGrain** 4 years, 9 months ago

That is correct answer should be "Yes". High Concurrency does not support Scala

upvoted 3 times

☐ 👤 **avix** 4 years, 11 months ago

Answer is wrong as High Concurrency cluster can't support scala

upvoted 2 times

hart232 4 years, 9 months ago

There is no need to use Scala for the users of high concurrency. This is mentioned in the question.

upvoted 1 times

sirshanam 4 years, 8 months ago

hart232 its mentioned as a requirement:

A workload for jobs that will run notebooks that use Python, Spark, Scala, and SQL

upvoted 2 times

hart232 4 years, 9 months ago

There is no need to use Scala for the users of high concurrency. This is mentioned in the question.

upvoted 1 times

sirshanam 4 years, 8 months ago

hart232 its mentioned as a requirement:

A workload for jobs that will run notebooks that use Python, Spark, Scala, and SQL

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 plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

☞ A workload for data engineers who will use Python and SQL

☞ A workload for jobs that will run notebooks that use Python, Scala, and SQL

☞ A workload that data scientists will use to perform ad hoc analysis in Scala and R

The enterprise architecture team at your company identifies the following standards for Databricks environments:

☞ The data engineers must share a cluster.

☞ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

☞ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

No need for a High Concurrency cluster for each data scientist.

Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

References:

https://docs.azuredatabricks.net/clusters/configure.html

👤 **ACSC** `Highly Voted 👍` 4 years, 6 months ago

A workload that data scientists will use to perform ad hoc analysis in Scala and R. High Concurrency clusters don't support Scala. Answer is "No".

upvoted 9 times

👤 **cadio30** `Most Recent ⊘` 4 years, 2 months ago

Definitely the answer is NO

upvoted 1 times

👤 **Hassan_Mazhar_Khan** 4 years, 2 months ago

A workload that data scientists will use to perform ad hoc analysis in Scala and R. High Concurrency clusters don't support Scala. Answer is "No".

upvoted 2 times

👤 **watata** 4 years, 4 months ago

Answers should be "yes"...

upvoted 1 times

    👤 **watata** 4 years, 4 months ago

    sorry, its "no", because for data Scientist should be Standard cluster

    upvoted 1 times

👤 **karishura** 4 years, 4 months ago

No - Right answer

upvoted 2 times

👤 **M0e** 4 years, 9 months ago

"There is no need for a High concurrency cluster" does not mean, it does not fulfil the goal. High concurrency clusters can be used for data scientists workload. It is just more expensive. There is no requirement regarding keeping the costs low, mentioned in the question. So the correct answer to this question is also "A - Yes."

You have an Azure Stream Analytics query. The query returns a result set that contains 10,000 distinct values for a column named clusterID.

You monitor the Stream Analytics job and discover high latency.

You need to reduce the latency.

Which two actions should you perform? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

    A. Add a pass-through query.

    B. Add a temporal analytic function.

    C. Scale out the query by using PARTITION BY.

    D. Convert the query to a reference query.

    E. Increase the number of streaming units.

Suggested Answer: *CE*

C: Scaling a Stream Analytics job takes advantage of partitions in the input or output. Partitioning lets you divide data into subsets based on a partition key. A process that consumes the data (such as a Streaming Analytics job) can consume and write different partitions in parallel, which increases throughput.

E: Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job. This capacity lets you focus on the query logic and abstracts the need to manage the hardware to run your

Stream Analytics job in a timely manner.

References:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

---

👤 **syu31svc** `Highly Voted 👍` 4 years, 7 months ago
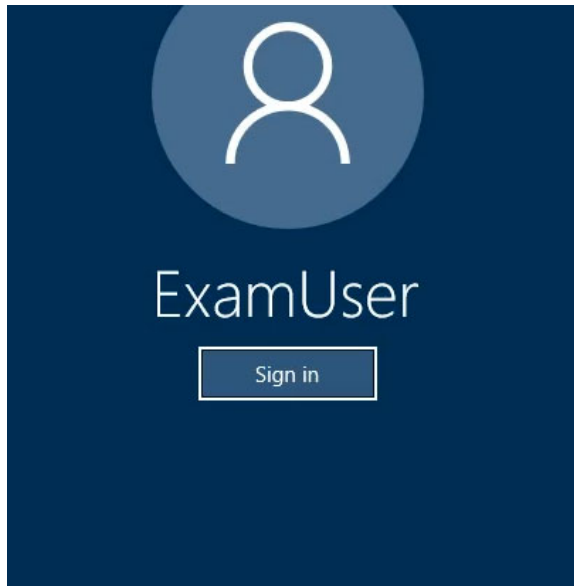
C and E are correct

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

https://docs.microsoft.com/en-us/azure/stream-analytics/repartition

Both touch on optimization

  upvoted 14 times

SIMULATION -

Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10277521 -
You plan to generate large amounts of real-time data that will be copied to Azure Blob storage.
You plan to create reports that will read the data from an Azure Cosmos DB database.
You need to create an Azure Stream Analytics job that will input the data from a blob storage named storage10277521 to the Cosmos DB database.
To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*
Step 1: Create a Stream Analytics job
1. Sign in to the Azure portal.
2. Select Create a resource in the upper left-hand corner of the Azure portal.
3. Select Analytics > Stream Analytics job from the results list.
4. Fill out the Stream Analytics job page.
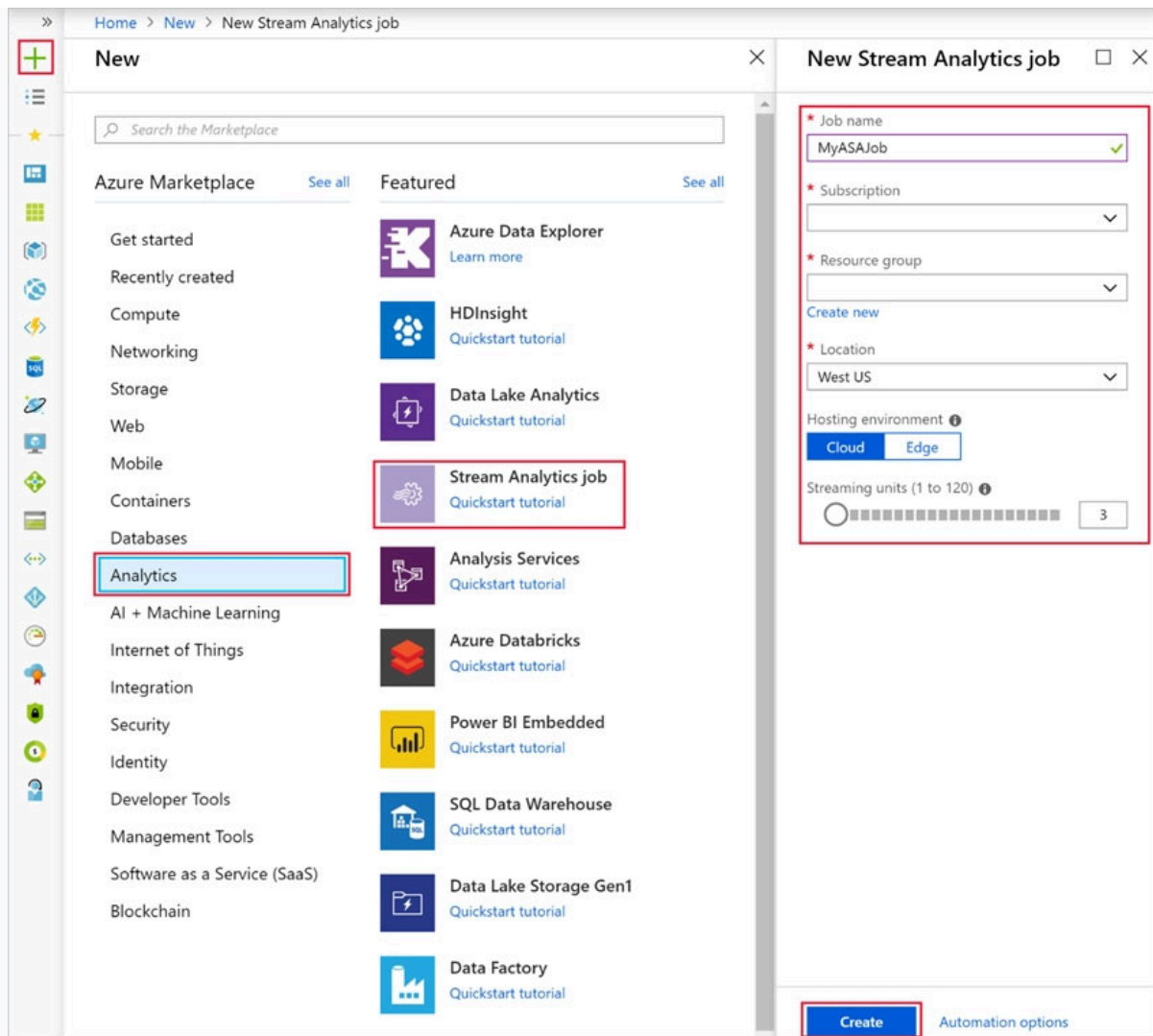
5. Check the Pin to dashboard box to place your job on your dashboard and then select Create.

6. You should see a Deployment in progress... notification displayed in the top right of your browser window.

Step 2: Configure job input -

1. Navigate to your Stream Analytics job.

2. Select Inputs > Add Stream input > Azure Blob storage
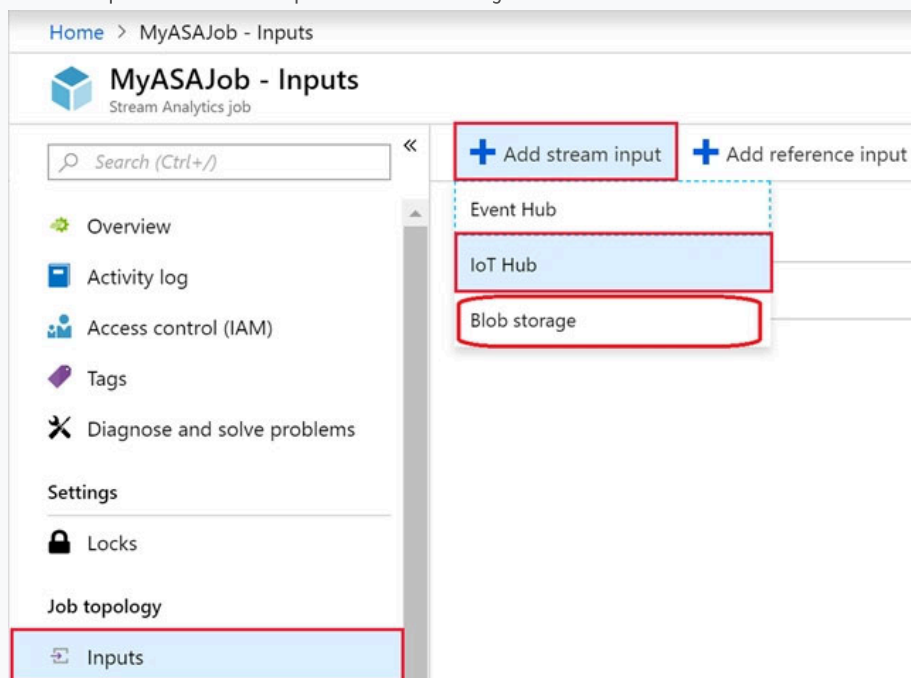
3. In the Azure Blob storage setting choose: storage10277521. Leave other options to default values and select Save to save the settings.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal

Sasidhar39 `Highly Voted 👍` 4 years, 1 month ago

Need to configure output section as well -- In the output select Azure cosmos db and select database and container or else need to create new ones

upvoted 8 times

Sasidhar39 `Highly Voted 👍` 4 years, 1 month ago

Need to configure output section as well -- In the output select Azure cosmos db and select database and container or else need to create new ones

upvoted 8 times

SIMULATION -
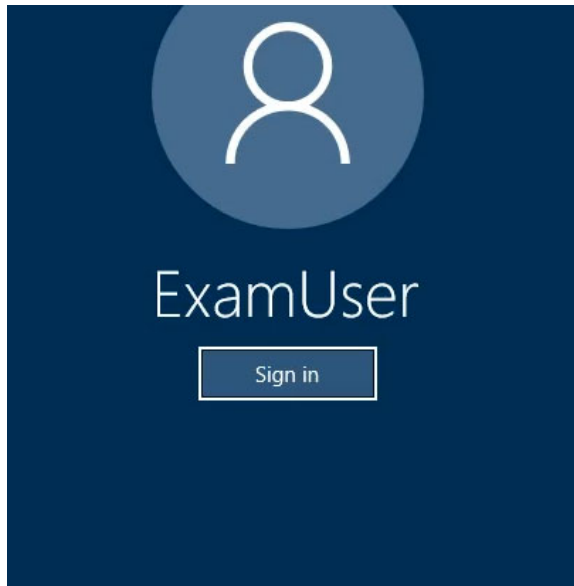
Use the following login credentials as needed:

Azure Username: xxxxx -

Azure Password: xxxxx -
The following information is for technical support purposes only:

Lab Instance: 10277521 -
You plan to create multiple pipelines in a new Azure Data Factory V2.
You need to create the data factory, and then create a scheduled trigger for the planned pipelines. The trigger must execute every two hours starting at 24:00:00.
To complete this task, sign in to the Azure portal.

Suggested Answer: *See the explanation below.*
Step 1: Create a new Azure Data Factory V2
1. Go to the Azure portal.
2. Select Create a resource on the left menu, select Analytics, and then select Data Factory.
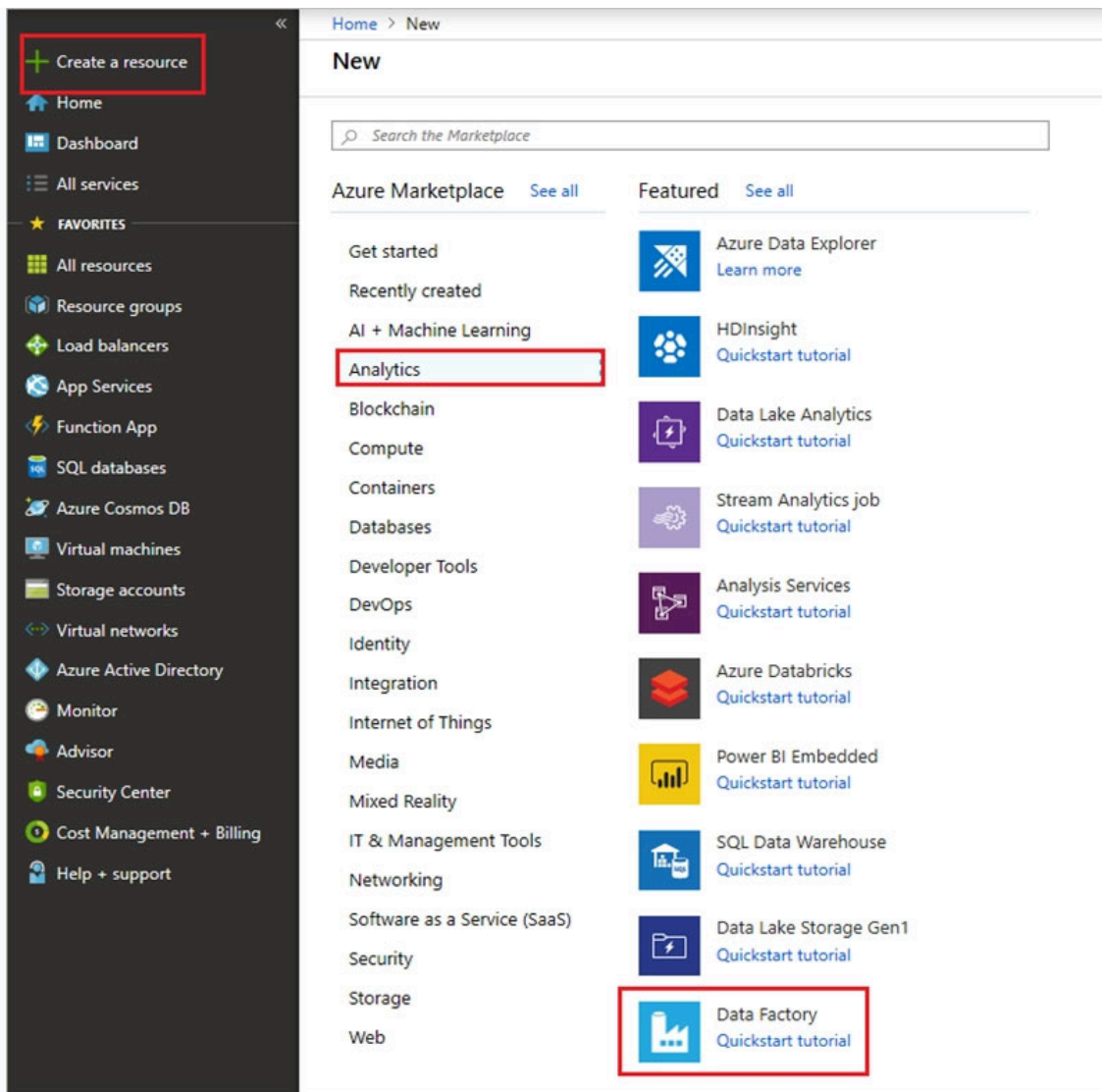
4. On the New data factory page, enter a name.

5. For Subscription, select your Azure subscription in which you want to create the data factory.

6. For Resource Group, use one of the following steps:

☞ Select Use existing, and select an existing resource group from the list.

☞ Select Create new, and enter the name of a resource group.

7. For Version, select V2.

8. For Location, select the location for the data factory.

9. Select Create.

10. After the creation is complete, you see the Data Factory page.

Step 2: Create a schedule trigger for the Data Factory

1. Select the Data Factory you created, and switch to the Edit tab.
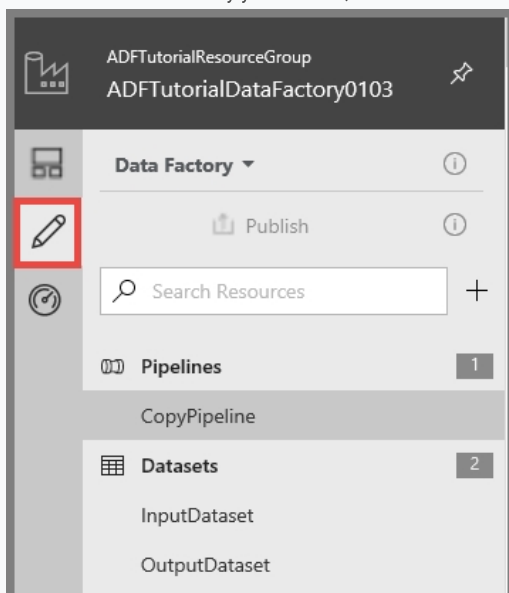
2. Click Trigger on the menu, and click New/Edit.
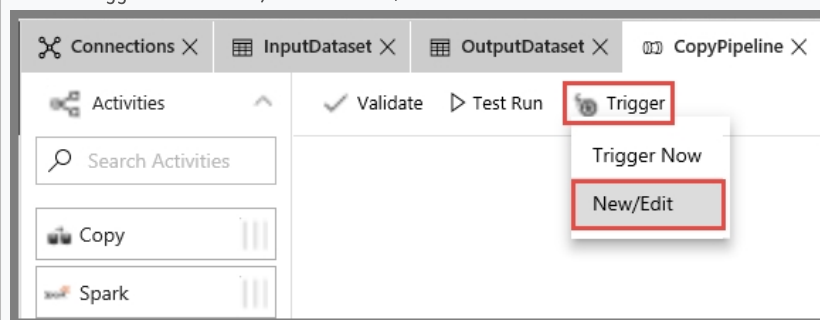
3. In the Add Triggers page, click Choose trigger..., and click New.
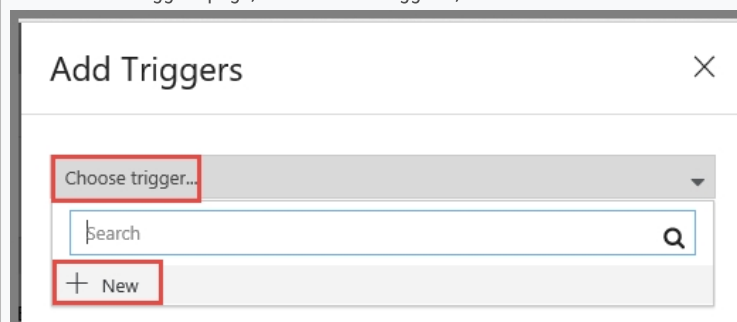
4. In the New Trigger page, do the following steps:

a. Confirm that Schedule is selected for Type.

b. Specify the start datetime of the trigger for Start Date (UTC) to: 24:00:00 c. Specify Recurrence for the trigger. Select Every Hour, and enter 2 in the text box.

5. In the New Trigger window, check the Activated option, and click Next.

6. In the New Trigger page, review the warning message, and click Finish.

7. Click Publish to publish changes to Data Factory. Until you publish changes to Data Factory, the trigger does not start triggering the pipeline runs.

References:

https://docs.microsoft.com/en-us/azure/data-factory/quickstart-create-data-factory-portal https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-schedule-trigger
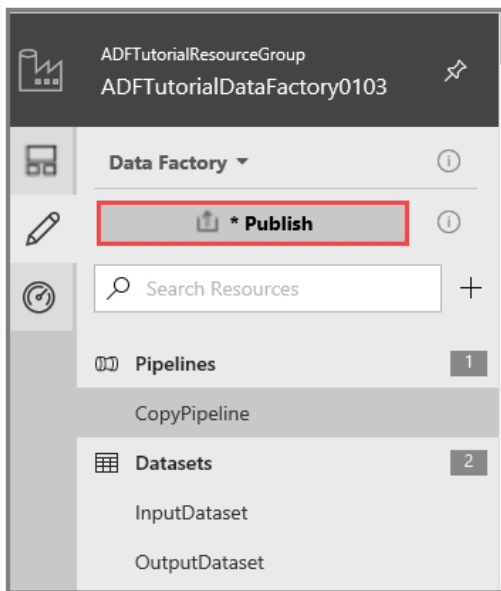
⊟ 👤 **jamorey** 3 years, 11 months ago

How appears this question type into the exam?, We need to do in the azure portal, or drag and drop box, or is in other way?. Help please

upvoted 2 times

⊟ 👤 **dragos_dragos62000** 4 years ago

Correct.

upvoted 1 times

Each day, company plans to store hundreds of files in Azure Blob Storage and Azure Data Lake Storage. The company uses the parquet format.
You must develop a pipeline that meets the following requirements:
☞ Process data every six hours
☞ Offer interactive data analysis capabilities
☞ Offer the ability to process data using solid-state drive (SSD) caching
☞ Use Directed Acyclic Graph(DAG) processing mechanisms
☞ Provide support for REST API calls to monitor processes
☞ Provide native support for Python
☞ Integrate with Microsoft Power BI
You need to select the appropriate data technology to implement the pipeline.
Which data technology should you implement?

    A. Azure SQL Data Warehouse

    B. HDInsight Apache Storm cluster

    C. Azure Stream Analytics

    D. HDInsight Apache Hadoop cluster using MapReduce

    E. HDInsight Spark cluster

Suggested Answer: *B*
Storm runs topologies instead of the Apache Hadoop MapReduce jobs that you might be familiar with. Storm topologies are composed of multiple components that are arranged in a directed acyclic graph (DAG). Data flows between the components in the graph. Each component consumes one or more data streams, and can optionally emit one or more streams.
Python can be used to develop Storm components.
References:
https://docs.microsoft.com/en-us/azure/hdinsight/storm/apache-storm-overview

---

☐ 👤 **c1265** `Highly Voted 👍` 5 years, 3 months ago
this really looks like spark, https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview
upvoted 30 times

☐ 👤 **wyxh** `Highly Voted 👍` 5 years, 2 months ago
Spark clusters in HDInsight provide connectors for BI tools such as Power BI for data analytics.
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview
Storm processes streams of data in real time, and in the question is stated the data must be Processed every six hours
upvoted 21 times

☐ 👤 **GData23** `Most Recent ⊘` 4 years, 1 month ago
hadoop is dead. who cares
upvoted 4 times

☐ 👤 **davita8** 4 years, 2 months ago
The answer is : "E "
upvoted 1 times

☐ 👤 **Hassan_Mazhar_Khan** 4 years, 2 months ago
The answer is : "E "
upvoted 1 times

☐ 👤 **rjile** 4 years, 3 months ago
The answer is : "E "
Spark clusters in HDInsight provide connectors for BI tools such as Power BI for data analytics.
upvoted 3 times

☐ 👤 **vidray** 4 years, 4 months ago
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview

Answer is : HDInsight Spark cluster

upvoted 2 times

⊟ 👤 **Shiva2** 4 years, 5 months ago

I think it must be spark too. I don't think there is a direct configuration from power bi to storm. You will need to process data in storm pass the processed data to SQL server, then visualize. Here there is no mention of SQL server , hence answer must be E. Spark.

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-use-bi-tools

upvoted 2 times

⊟ 👤 **Prada** 4 years, 5 months ago

It must be E. HDInsight Spark cluster due to Offer the ability to process data using solid-state drive (SSD) caching

https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-improve-performance-iocache

upvoted 1 times

⊟ 👤 **syu31svc** 4 years, 7 months ago

I would say E as per link https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview:

Read the parts on Caching on SSDs, The SparkContext connects to the Spark master and is responsible for converting an application to a directed graph (DAG) of individual tasks, REST APIs, Integration with BI Tools

Python is mentioned as well

upvoted 4 times

⊟ 👤 **dumpsm42** 4 years, 6 months ago

hi to all,

it's E mostly because of the power BI, the text says exactly "integrate with power BI" and in the microsoft documentatio for hd insight spark cluster => "...Integration with BI Tools Spark clusters in HDInsight provide connectors for BI tools such as Power BI for data analytics....", so for me it's E

regards

upvoted 1 times

⊟ 👤 **Sagja** 4 years, 8 months ago

Answer is correct. Checkout this - https://docs.microsoft.com/en-us/azure/hdinsight/storm/apache-storm-overview.

Python can also be used to develop Storm components.

Create solutions in multiple languages: You can write Storm components in the language of your choice, such as Java, C#, and Python.

upvoted 1 times

⊟ 👤 **M0e** 4 years, 9 months ago

Storm does not provide Python (https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing) and is not used for 6hr batch processing (https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing).

Spark supports all of the options. Hence, E is the correct answer.

upvoted 2 times

⊟ 👤 **EYIT** 4 years, 9 months ago

E. HDInsight Spark cluster

https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing

upvoted 2 times

⊟ 👤 **Arsa** 4 years, 10 months ago

it should be Spark because of (DAG + PowerBI integration)

upvoted 3 times

⊟ 👤 **rmk4ever** 4 years, 11 months ago

Ans is HDInsight Spark cluster

Caching on SSDs

Integration with BI Tools

Spark master is responsible for converting an application to a directed graph (DAG)

REST API-based Spark job server to remotely submit and monitor job

full reference: https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview

upvoted 7 times

⊟ 👤 **azrnovice** 5 years ago

Question says need python native support. Azure Storm don't support Python. Check out this comparison chart:

https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing

upvoted 12 times

Luke97 5 years, 1 month ago

I think HDInsight Apache Spark should be the correct answer.

1. Offer interactive data analysis

2. Offer caching

3. DirectQuery (live connection) to PowerBI

And Spark also use DAG too (https://docs.microsoft.com/en-au/azure/hdinsight/spark/apache-spark-streaming-exactly-once).

upvoted 8 times

Luke97 5 years, 1 month ago

I think HDInsight Apache Spark should be the correct answer.

1. Offer interactive data analysis

2. Offer caching

3. DirectQuery (live connection) to PowerBI

And Spark also use DAG too (https://docs.microsoft.com/en-au/azure/hdinsight/spark/apache-spark-streaming-exactly-once).

upvoted 8 times

HOTSPOT -

A company is deploying a service-based data environment. You are developing a solution to process this data.

The solution must meet the following requirements:

☞ Use an Azure HDInsight cluster for data ingestion from a relational database in a different cloud service

☞ Use an Azure Data Lake Storage account to store processed data

☞ Allow users to download processed data

You need to recommend technologies for the solution.

Which technologies should you use? To answer, select the appropriate options in the answer area.

Hot Area:

Answer Area

Data process	Technology
Ingest	RevoScaleR / Apache Sqoop / Apache DistCp / Azure CLI ∨
Process	Apache DistCp / Apache Kafka / C# / Apache Hive ∨
Download	Apache Sqoop / MapReduce / RevoScaleR / Ambari Hive View ∨

Suggested Answer:

Answer Area

Data process	Technology
Ingest	RevoScaleR / **Apache Sqoop** / Apache DistCp / Azure CLI ∨
Process	Apache DistCp / **Apache Kafka** / C# / Apache Hive ∨
Download	Apache Sqoop / MapReduce / RevoScaleR / **Ambari Hive View** ∨

Box 1: Apache Sqoop -

Apache Sqoop is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases.

Azure HDInsight is a cloud distribution of the Hadoop components from the Hortonworks Data Platform (HDP).

Incorrect Answers:

DistCp (distributed copy) is a tool used for large inter/intra-cluster copying. It uses MapReduce to effect its distribution, error handling and recovery, and reporting.

It expands a list of files and directories into input to map tasks, each of which will copy a partition of the files specified in the source list. Its MapReduce pedigree has endowed it with some quirks in both its semantics and execution.

RevoScaleR is a collection of proprietary functions in Machine Learning Server used for practicing data science at scale. For data scientists, RevoScaleR gives you data-related functions for import, transformation and manipulation, summarization, visualization, and analysis.

Box 2: Apache Kafka -
Apache Kafka is a distributed streaming platform.
A streaming platform has three key capabilities:
Publish and subscribe to streams of records, similar to a message queue or enterprise messaging system.
Store streams of records in a fault-tolerant durable way.
Process streams of records as they occur.
Kafka is generally used for two broad classes of applications:
Building real-time streaming data pipelines that reliably get data between systems or applications
Building real-time streaming applications that transform or react to the streams of data

Box 3: Ambari Hive View -
You can run Hive queries by using Apache Ambari Hive View. The Hive View allows you to author, optimize, and run Hive queries from your web browser.
References:
https://sqoop.apache.org/
https://kafka.apache.org/intro
https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-ambari-view

unidigm 4 years, 1 month ago
Apache Sqoop, Apache Hive, Ambari Hive View
upvoted 4 times

Kratik 4 years, 1 month ago
For Process, I think it should be hive
For Download, the answer seems correct. But instead of 'Ambari Hive View', I think it should be 'Apache Hive View'
upvoted 1 times

Hassan_Mazhar_Khan 4 years, 2 months ago
For Process it should be 'Hive' as it provide full storage mechanism
upvoted 1 times

tucho 4 years, 2 months ago
can't be "hive" for process task?
upvoted 2 times

A company uses Azure SQL Database to store sales transaction data. Field sales employees need an offline copy of the database that includes last year's sales on their laptops when there is no internet connection available.

You need to create the offline export copy.

Which three options can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

    A. Export to a BACPAC file by using Azure Cloud Shell, and save the file to an Azure storage account

    B. Export to a BACPAC file by using SQL Server Management Studio. Save the file to an Azure storage account

    C. Export to a BACPAC file by using the Azure portal

    D. Export to a BACPAC file by using Azure PowerShell and save the file locally

    E. Export to a BACPAC file by using the SqlPackage utility

Suggested Answer: *BCE*

You can export to a BACPAC file using the Azure portal.

You can export to a BACPAC file using SQL Server Management Studio (SSMS). The newest versions of SQL Server Management Studio provide a wizard to export an Azure SQL database to a BACPAC file.

You can export to a BACPAC file using the SQLPackage utility.

Incorrect Answers:

D: You can export to a BACPAC file using PowerShell. Use the New-AzSqlDatabaseExport cmdlet to submit an export database request to the Azure SQL

Database service. Depending on the size of your database, the export operation may take some time to complete. However, the file is not stored locally.

References:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-export

---

STH `Highly Voted 👍` 5 years, 7 months ago

Shouldn't the solution be C, D and E ?

All tools are valid way to create the BACPAC file, but A and B store it online to Azure Storage, so it cannot be accessed offline

upvoted 31 times

> **Frederi** 5 years, 7 months ago
>
> The question is not how to distribute the offline copy, but how to create the original export, which would then be distributed.
>
> upvoted 4 times
>
> > **goodzilla** 5 years, 3 months ago
> >
> > agree with STH
> >
> > "...includes last year's sales on their laptops when there is no internet connection available.". If there is no internet connection, no way to connect to Azure BLOB storage
> >
> > upvoted 3 times
> >
> > > **cowtown** 4 years, 12 months ago
> > >
> > > When it's not available != no network access ever.
> > >
> > > upvoted 3 times

> **AshCool** 4 years, 2 months ago
>
> Without internet how you access azure portal?
>
> upvoted 4 times

> **111222333** 4 years, 1 month ago
>
> You cannot really save a BACPAC file locally using PowerShell either - only to Azure Storage account.
>
> To export with PowerShell, you must use command "New-AzSqlDatabaseExport", and you have to specify a parameter "-StorageUri" which specifies Blob location:
>
> https://docs.microsoft.com/en-us/powershell/module/az.sql/new-azsqldatabaseexport?view=azps-5.9.0

So, answer D also stores it online and hence is not correct.
upvoted 2 times

cnuusd `Highly Voted 👍` 5 years, 5 months ago
Correct Answer : B , C and E
upvoted 12 times

LordSnoek `Most Recent ⊘` 4 years ago
its B,D,E - according to Whizlabs
upvoted 1 times

Hevz 4 years, 1 month ago
Azure SQL Managed Instance does not currently support exporting a database to a BACPAC file using Azure PowerShell. To export a managed instance into a BACPAC file, use SQL Server Management Studio or SQLPackage.
upvoted 1 times

hoangton 4 years, 1 month ago
C,D,E because A,B are save data to Azure storage account
upvoted 2 times

111222333 4 years, 1 month ago
Correct answer is B, C, E.

I came to this conclusion by eliminating incorrect ones.

There are 4 ways to export an SQL Database to a BACPAC file:
1. Export using Azure Portal -> only to Storage Account
2. Export using SQLPackage command-line utility -> to local destination
3. Export using SQL Server Management Studio -> to local or to Storage Account
4. Export using PowerShell -> only to Storage Account

So, answer A is incorrect because you cannot export a BACPAC file by using Cloud Shell.
Answer D is incorrect because you cannot save a BACPAC file locally using PowerShell, but only to Storage account.

Another thing to note is that I think it is assumed that people will download BACPAC file from Azure Storage Account right after it is saved there and have them locally when the Internet breaks.
upvoted 4 times

111222333 4 years, 1 month ago
Just to elaborate a bit more on D, I think that Microsoft purposely wrote the incorrect statement "and save the file locally" so we can interpret it as an invalid statement. (Even though you can save it to Storage Account and download it locally like with Azure portal.)
Notice how they do not emphasize where to save files in answers C and E, and how they emphasize Storage Account in answers A and B. That's why I think that the emphasis is on the part "and save the file locally".
upvoted 1 times

Maky2365 4 years, 1 month ago
answer should be C, D & E.
moreover Option A & B are itself saying to save the file to Azure Storage Account and getting the file from Azure Storage Account is practically impossible with internet connection.
upvoted 2 times

SorinXp 4 years, 1 month ago
"Field sales employees need an offline copy of the database that includes last yearג€™s sales on their laptops when there is no internet connection available." -> The BACPAC file contains only metadata, no data. How is it possible to include the last year sales?
upvoted 1 times

cadio30 4 years, 1 month ago
CDE are the appropriate answers
upvoted 2 times

cadio30 4 years, 1 month ago
B,C,D,E are ways to perform database export process though if the requirement is to have the file locally then the appropriate solution are C,D,E.
Both C and D requires azure blob storage to store the copy and in option D, it was stated that it will save the file locally. The option E is straight

forward the user indicates the target drive and folder before the export begins. As for option 'B' using it exports the data into the local drive but it was directed to azure storage account.

upvoted 1 times

🗕 👤 **dbdev** 4 years, 2 months ago

CDE:

https://www.sqlshack.com/how-to-perform-azure-sql-database-import-export-operations-using-powershell/

upvoted 1 times

🗕 👤 **Hassan_Mazhar_Khan** 4 years, 2 months ago

It should be C,D,E

upvoted 1 times

🗕 👤 **Jzerpa_ccs** 4 years, 3 months ago

the key is here

"on their laptops when there is no internet connection available"

If you don't have an internet connection, how do you access a blob storage?

I think C, D and E

upvoted 1 times

🗕 👤 **ck1729** 4 years, 3 months ago

B, D, E. Azure Portal cannot be an option. please see link below.

https://docs.microsoft.com/en-us/azure/azure-sql/database/database-export

upvoted 2 times

🗕 👤 **JohnCrawford** 4 years, 5 months ago

I expect the issue with the PowerShell option is that while it can be used to export the BACPAC it is not exported to local storage. From the wording of that choice it seems to imply it is a direct export to local storage which is not possible.

upvoted 1 times

🗕 👤 **KRV** 4 years, 5 months ago

Option B , D & E , since even though you can create a backup using the Azure Portal, the backup won't be available locally.

Export to a BACPAC file by using SQL Server Management Studio. Save the file to an Azure storage account

Export to a BACPAC file by using Azure PowerShell and save the file locally

Export to a BACPAC file by using the SqlPackage utility

upvoted 1 times

🗕 👤 **syu31svc** 4 years, 7 months ago

I would go for BCE as the link provided supports it

upvoted 2 times

🗕 👤 **hart232** 4 years, 9 months ago

As far as I have searched

- Export to a BACPAC file by using Azure Cloud Shell is not possible

- Cannot save the BACPAC file locally via a straight forward Azure powershell command

So, the answers given are correct in the sense that they can produce a BACPAC file for consumption.

upvoted 1 times

Note: This question is a part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to an enterprise data warehouse in Azure Synapse Analytics. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure Data Lake Gen 2 storage account

2. Create an external file format and external table using the external data source

3. Load the data using the CREATE TABLE AS SELECT statement

Does the solution meet the goal?

    A. Yes

    B. No

Suggested Answer: *A*

You need to create an external file format and external table using the external data source.

You load the data using the CREATE TABLE AS SELECT statement.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

---

Muishkin 3 years, 1 month ago

answer should be NO ..As it is from data source & fileformat

upvoted 2 times

edba 3 years, 6 months ago

should it be "No" as last step shall be Create external table ...with (location..,data_source=..,fileformat=..)?

upvoted 3 times

Note: This question is a part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to an enterprise data warehouse in Azure Synapse Analytics. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Data Warehouse.

Solution:

1. Create a remote service binding pointing to the Azure Data Lake Gen 2 storage account

2. Create an external file format and external table using the external data source

3. Load the data using the CREATE TABLE AS SELECT statement

Does the solution meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

You need to create an external file format and external table from an external data source, instead from a remote service binding pointing.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

---

👤 **wxlf23** 3 years, 10 months ago

1. Create an external data source pointing to the Azure Data Lake Gen 2 storage account

vs

1. Create a remote service binding pointing to the Azure Data Lake Gen 2 storage account

Which is correct?

upvoted 1 times

👤 **Ankush1994** 3 years, 10 months ago

A

Answer is correct

upvoted 2 times

👤 **niwe** 4 years, 1 month ago

This should be the correct answer

upvoted 1 times

You need to develop a pipeline for processing data. The pipeline must meet the following requirements:

☞ Scale up and down resources for cost reduction

☞ Use an in-memory data processing engine to speed up ETL and machine learning operations.

☞ Use streaming capabilities

☞ Provide the ability to code in SQL, Python, Scala, and R

Integrate workspace collaboration with Git

▪

What should you use?

    A. HDInsight Spark Cluster

    B. Azure Stream Analytics

    C. HDInsight Hadoop Cluster

    D. Azure SQL Data Warehouse

    E. HDInsight Kafka Cluster

    F. HDInsight Storm Cluster

Suggested Answer: *A*

Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.

HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP,

MapReduce.

Languages: R, Python, Java, Scala, SQL

You can create an HDInsight Spark cluster using an Azure Resource Manager template. The template can be found in GitHub.

References:

https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing

🗕 👤 **hoangton** 4 years ago

Answer is CORECT. key word is in-memory data processing

  upvoted 4 times

DRAG DROP -

You implement an event processing solution using Microsoft Azure Stream Analytics.

The solution must meet the following requirements:

☞ Ingest data from Blob storage

☞ Analyze data in real time

☞ Store processed data in Azure Cosmos DB

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.

Select and Place:

Actions

Create a query statement with the ORDER BY clause.

Create a query statement with the SELECT INTO statement.

Configure Blob storage for a reference data JOIN clause.

Configure Azure Event Hub as input; select items with the TIMESTAMP BY clause

Set up Cosmos DB as the output.

Configure Blob storage as input; select items with the TIMESTAMP BY clause.

Answer Area

Suggested Answer:

Actions

Create a query statement with the ORDER BY clause.

Configure Blob storage for a reference data JOIN clause.

Configure Azure Event Hub as input; select items with the TIMESTAMP BY clause

Answer Area

Configure Blob storage as input; select items with the TIMESTAMP BY clause.

Set up Cosmos DB as the output.

Create a query statement with the SELECT INTO statement.

Step 1: Configure Blob storage as input; select items with the TIMESTAMP BY clause

The default timestamp of Blob storage events in Stream Analytics is the timestamp that the blob was last modified, which is BlobLastModifiedUtcTime. To process the data as a stream using a timestamp in the event payload, you must use the TIMESTAMP BY keyword.

Example:

The following is a TIMESTAMP BY example which uses the EntryTime column as the application time for events:

SELECT TollId, EntryTime AS VehicleEntryTime, LicensePlate, State, Make, Model, VehicleType, VehicleWeight, Toll, Tag

FROM TollTagEntry TIMESTAMP BY EntryTime

Step 2: Set up cosmos DB as the output

Creating Cosmos DB as an output in Stream Analytics generates a prompt for information as seen below.

Step 3: Create a query statement with the SELECT INTO statement.

References:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-inputs

---

👤 **AidenPearce** 3 years, 9 months ago

correct

upvoted 1 times

👤 **promiseve** 4 years, 4 months ago

given answer is correct

upvoted 2 times

👤 **lingjun** 4 years, 7 months ago

query: Select .. INTO.. FROM..Having... https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal

upvoted 3 times

HOTSPOT -

A company plans to use Platform-as-a-Service (PaaS) to create the new data pipeline process. The process must meet the following requirements:

Ingest:

☞ Access multiple data sources.

☞ Provide the ability to orchestrate workflow.

☞ Provide the capability to run SQL Server Integration Services packages.

Store:

☞ Optimize storage for big data workloads

☞ Provide encryption of data at rest.

☞ Operate with no size limits.

Prepare and Train:

☞ Provide a fully-managed and interactive workspace for exploration and visualization.

☞ Provide the ability to program in R, SQL, Python, Scala, and Java.

☞ Provide seamless user authentication with Azure Active Directory.

Model & Serve:

☞ Implement native columnar storage.

☞ Support for the SQL language.

☞ Provide support for structured streaming.

You need to build the data integration pipeline.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Architecture requirement	Technology
Ingest	logic apps
	Azure Data Factory
	Azure Automation
Store	Azure Data Lake Storage
	Azure Blob storage
	Azure Files
Prepare and Train	HDInsight Apache Spark cluster
	Azure Databricks
	HDInsight Apache Storm cluster
Model and Serve	HDInsight Apache Kafka cluster
	Azure Synapse Analytics
	Azure Data Lake Storage

Answer Area

Architecture requirement	Technology
Ingest	▼ logic apps / **Azure Data Factory** / Azure Automation
Store	▼ **Azure Data Lake Storage** / Azure Blob storage / Azure Files
Prepare and Train	▼ HDInsight Apache Spark cluster / **Azure Databricks** / HDInsight Apache Storm cluster
Model and Serve	▼ HDInsight Apache Kafka cluster / **Azure Synapse Analytics** / Azure Data Lake Storage

Suggested Answer:

Ingest: Azure Data Factory -

Azure Data Factory pipelines can execute SSIS packages.

In Azure, the following services and tools will meet the core requirements for pipeline orchestration, control flow, and data movement: Azure Data Factory, Oozie on HDInsight, and SQL Server Integration Services (SSIS).

Store: Data Lake Storage -

Data Lake Storage Gen1 provides unlimited storage.

Note: Data at rest includes information that resides in persistent storage on physical media, in any digital format. Microsoft Azure offers a variety of data storage solutions to meet different needs, including file, disk, blob, and table storage. Microsoft also provides encryption to protect Azure SQL Database, Azure Cosmos

DB, and Azure Data Lake.

Prepare and Train: Azure Databricks

Azure Databricks provides enterprise-grade Azure security, including Azure Active Directory integration.

With Azure Databricks, you can set up your Apache Spark environment in minutes, autoscale and collaborate on shared projects in an interactive workspace.

Azure Databricks supports Python, Scala, R, Java and SQL, as well as data science frameworks and libraries including TensorFlow, PyTorch and scikit-learn.

Model and Serve: Azure Synapse Analytics

Azure Synapse Analytics/ SQL Data Warehouse stores data into relational tables with columnar storage.

Azure SQL Data Warehouse connector now offers efficient and scalable structured streaming write support for SQL Data Warehouse. Access SQL Data

Warehouse from Azure Databricks using the SQL Data Warehouse connector.

Note: Note: As of November 2019, Azure SQL Data Warehouse is now Azure Synapse Analytics.

References:

https://docs.microsoft.com/bs-latn-ba/azure/architecture/data-guide/technology-choices/pipeline-orchestration-data-movement

https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks

---

□ 👤 **syu31svc** `Highly Voted 👍` 4 years, 7 months ago

Answer given is correct

upvoted 15 times

□ 👤 **Arsa** `Most Recent ⊙` 4 years, 10 months ago

Featured snippet from the web

Azure SQL Data Warehouse connector now offers efficient and scalable structured streaming write support for SQL Data Warehouse.

upvoted 1 times

Treadmill 4 years, 11 months ago

Prepare and train: Databricks, also as Spark does not support Azure AD and Databricks does

upvoted 3 times

serger 5 years ago

Not sure Databricks is the good answer. Databricks does not support Java. Should be HDInsight Spark.

upvoted 3 times

Ikrom 5 years ago

Databricks supports Java and all mentioned languages since it's optimized Apache Spark: https://azure.microsoft.com/en-us/services/databricks/#overview

upvoted 7 times

vkmicrosoft 4 years, 12 months ago

I think the confusion of databricks not supporting java is because, there was a question in here answered saying so. But yes i agree. Databrick is the right answer as it supports the required languages including Java.

upvoted 4 times

induna 4 years, 8 months ago

Currently when I go to my Databricks cluster and I create a new notebook the dropdown does not list Java, I think that is the reason people think it is not supported. BUT you can create a java JAR, and upload that to Databricks THEN access the classes in it.

upvoted 3 times

zenomas 5 years, 3 months ago

Anyone, any idea about Azure SQL Data Warehouse feature for "structured streaming"?

upvoted 1 times

Yuri1101 5 years, 2 months ago

https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/synapse-analytics#streaming-support

upvoted 8 times

HOTSPOT -

A company plans to analyze a continuous flow of data from a social media platform by using Microsoft Azure Stream Analytics. The incoming data is formatted as one record per row.
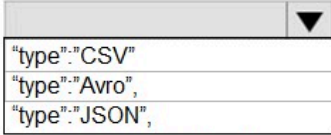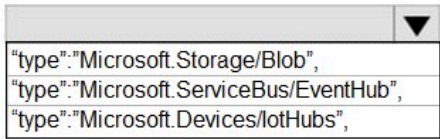
You need to create the input stream.

How should you complete the REST API segment? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
{
  "properties" : {
    "type" : "stream",
    "serialization" : {
```

▼
"type":"CSV"
"type":"Avro",
"type":"JSON",

```
      "properties" : {
        "fieldDelimiter" : ". ",
        "encoding" : "UTF8"
      }
  },
  "datasource":{
```

▼
"type":"Microsoft.Storage/Blob",
"type":"Microsoft.ServiceBus/EventHub",
"type":"Microsoft.Devices/IotHubs",
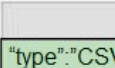
```
  "properties": {
        "serviceBusNamespace" : "sampleServiceBus",
        "sharedAccessPolicyName" : "SampleReceiver",
        "sharedAccessPolicyKey" : "<PolicyKey>"
        "eventHubName" : "sampleEventHub"
      }
  },
      "compression":{
      "type" : "GZip"
      }
    }
}
```

Answer Area

```
{
    "properties" : {
        "type" : "stream",
        "serialization" : {
```

▼
"type":"CSV"
"type":"Avro",
"type":"JSON",

```
            "properties" : {
                "fieldDelimiter" : ". ",
                "encoding" : "UTF8"
            }
    },
    "datasource":{
```

Suggested Answer:

▼
"type":"Microsoft.Storage/Blob",
"type":"Microsoft.ServiceBus/EventHub",
"type":"Microsoft.Devices/IotHubs",

```
        "properties": {
            "serviceBusNamespace" : "sampleServiceBus",
            "sharedAccessPolicyName" : "SampleReceiver",
            "sharedAccessPolicyKey" : "<PolicyKey>"
            "eventHubName" : "sampleEventHub"
        }
    },
        "compression":{
        "type" : "GZip"
        }
    }
}
```

Box 1: CSV -

A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. A CSV file stores tabular data (numbers and text) in plain text.

Each line of the file is a data record.

JSON and AVRO are not formatted as one record per row.

Box 2: "type":"Microsoft.ServiceBus/EventHub",

Properties include "EventHubName"

References:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-inputs https://en.wikipedia.org/wiki/Comma-separated_values

---

☐ 👤 **epgd** `Highly Voted 👍` 5 years, 7 months ago

Check this:

```
{
"properties":{
"type":"stream",
"serialization":{
"type":"CSV",
"properties":{
"fieldDelimiter":",",
"encoding":"UTF8"
}
},
"datasource":{
"type":"Microsoft.ServiceBus/EventHub",
"properties":{
"serviceBusNamespace":"sampleServiceBus",
"sharedAccessPolicyName":"SampleReceiver",
"sharedAccessPolicyKey":"***/*********/****************************",
```

"eventHubName":"sampleEventHub"
}
},
"compression":{
"type":"GZip"
}
}
}
https://docs.microsoft.com/es-es/rest/api/streamanalytics/stream-analytics-input

upvoted 17 times

⊟ 👤 **AAJ** `Highly Voted 👍` 5 years, 3 months ago

CSV does not have to be"only comma separated". Given answer is correct.

upvoted 15 times

⊟ 👤 **Prada** `Most Recent ⊙` 4 years, 5 months ago

AVRO is a serialized JSON. Therefore, it must be CSV

upvoted 2 times

⊟ 👤 **Luke97** 5 years, 1 month ago

fieldDelimiter is ONLY for CSV type data in ASA Input Serialization JSON file. So, CSV is the right answer.

upvoted 2 times

⊟ 👤 **STH** 5 years, 7 months ago

CSV is comma separated, not dot.

Avro is made for serialization, so it's probably the right format.

upvoted 3 times

⊟ 👤 **Gluckos** 4 years, 8 months ago

wrong response for me.. csv may have comma, dot, star so may have any characters

upvoted 1 times

⊟ 👤 **Pairon** 4 years, 3 months ago

"One record per row" -> CSV. Avro is in columnar format for reads optimization.

upvoted 3 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 are developing a solution that will use Azure Stream Analytics. The solution will accept an Azure Blob storage file named Customers. The file will contain both in-store and online customer details. The online customers will provide a mailing address.

You have a file in Blob storage named LocationIncomes that contains median incomes based on location. The file rarely changes.

You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure

Data Lake Storage Gen2 for long-term retention.

Solution: You implement a Stream Analytics job that has one streaming input, one query, and two outputs.

Does this meet the goal?

   A. Yes

   B. No

Suggested Answer: *B*

We need one reference data input for LocationIncomes, which rarely changes.

Note: Stream Analytics also supports input known as reference data. Reference data is either completely static or changes slowly.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs#stream-and-reference-inputs

---

⊟ 👤 **tucho** 4 years, 2 months ago

Correct answer is B) , "add reference input" is missing.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

  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 are developing a solution that will use Azure Stream Analytics. The solution will accept an Azure Blob storage file named Customers. The file will contain both in-store and online customer details. The online customers will provide a mailing address.

You have a file in Blob storage named LocationIncomes that contains median incomes based on location. The file rarely changes.

You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure

Data Lake Storage Gen2 for long-term retention.

Solution: You implement a Stream Analytics job that has one streaming input, one reference input, one query, and two outputs.

Does this meet the goal?

A. Yes

B. No

> **Suggested Answer:** *B*
> We need one reference data input for LocationIncomes, which rarely changes.
> We need two queries, on for in-store customers, and one for online customers.
> For each query two outputs is needed.
> Note: Stream Analytics also supports input known as reference data. Reference data is either completely static or changes slowly.
> References:
> https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs#stream-and-reference-inputs
> https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs

👤 **Uday0809** `Highly Voted 👍` 4 years, 7 months ago

This is the correct answer --> "You implement a Stream Analytics job that has one streaming input, one reference input, one query, and two outputs."

upvoted 21 times

👤 **dumpsm42** `Highly Voted 👍` 4 years, 6 months ago

hi to all,

it's correct: A

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs

in one query we can join reference data

two outputs because of two different database systems

regards

upvoted 6 times

　👤 **dumpsm42** 4 years, 6 months ago

　https://docs.microsoft.com/en-us/stream-analytics-query/reference-data-join-azure-stream-analytics

　upvoted 2 times

👤 **Akva** `Most Recent ⊘` 4 years, 8 months ago

We don't need streaming input, hence answer - no

upvoted 2 times

　👤 **victor90** 3 years, 7 months ago

　No, Stream analytics jobs must have at least one stream input.

　https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs#data-stream-input

　upvoted 1 times

👤 **BungyTex** 4 years, 10 months ago

According to the documentation you can have two output types in a query. "You can use a single output per job, or multiple outputs per streaming job (if you r the query." I think the answer is A.

upvoted 2 times

　👤 **sandGrain** 4 years, 9 months ago

I agree with BungyTex ... you only need one query with 2 output types .. i have done it in practice ... so can validate it works ...

upvoted 1 times

- **jumby** 4 years, 7 months ago

  To back BungyTex and sandGrain, here' the doc: https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs#:~:text=An%20Azure%20Stream%20Analytics%20job,%2C%20query%2C%20and%20an%20output.&text=When%20you%20design%20your%20S

  upvoted 1 times

  - **mohowzeh** 4 years, 6 months ago

    Hi Jumby, in the link you provide, the text says that there can be multiple INTO statements in the "query". However, in the following link, it can be se SELECT statements: https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns#query-example-s It seems that different authors at Microsoft disagree on exactly what a query is. If one query is equal to one SELECT statement, as most SQL literat would mean two queries. The author of the article on Streaming Analytics uses a different definition of "query". IMHO he should have said "job" inst particularly since an exam question rests entirely of which version of the definition one uses.

    upvoted 1 times

- **mohowzeh** 4 years, 6 months ago

  In your quote "multiple outputs per streaming job", the key word "job" seems different to me from a "query". The test question refers to number of queries, to one SELECT statement, then one job can have multiple queries.

  upvoted 1 times

- **singhadi003** 4 years, 11 months ago

  We need to two query to load in different targets as stated in question:

  "You must output the data to Azure SQL Database for immediate use and to Azure

  Data Lake Storage Gen2 for long-term retention."

  upvoted 3 times

  - **Pairon** 4 years, 2 months ago

    Yeah but that means you save the SAME result (only one query) in two services because of different purposes. So I think that one query is enough.

    upvoted 1 times

- **HeleneB** 4 years, 11 months ago

  I don't see why you couldn't use one query since both sources are in the one customer file. It doesn't say you need two sets of queries for each source in the scenario. I disagree.

  upvoted 3 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 are developing a solution that will use Azure Stream Analytics. The solution will accept an Azure Blob storage file named Customers. The file will contain both in-store and online customer details. The online customers will provide a mailing address.

You have a file in Blob storage named LocationIncomes that contains median incomes based on location. The file rarely changes.

You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure

Data Lake Storage Gen2 for long-term retention.

Solution: You implement a Stream Analytics job that has one streaming input, one reference input, two queries, and four outputs.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *A*

We need one reference data input for LocationIncomes, which rarely changes.

We need two queries, on for in-store customers, and one for online customers.

For each query two outputs is needed.

Note: Stream Analytics also supports input known as reference data. Reference data is either completely static or changes slowly.

References:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs#stream-and-reference-inputs

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs

---

👤 **BungyTex** `Highly Voted 👍` 4 years, 10 months ago

It would meet the goal, but why would you even create 2 queries? You can get there with just 1 query and two outputs.

upvoted 11 times

    👤 **H_S** 4 years, 4 months ago

    no you can"t, you have 2 "INTO"

    upvoted 1 times

👤 **cadio30** `Most Recent ⊙` 4 years, 1 month ago

NO is the answer

Reference: https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns#query-example-send-data-to-multiple-outputs

upvoted 1 times

    👤 **cadio30** 4 years, 1 month ago

    Here is the scenario, there is a file that contains two dataset in-store and online customer and it requires to load it into Azure SQL and Data Lake. In coding perspective each query is equivalent to ONE select into which results to two queries to load the requirement. However, if we are going to combine the in-store and online customers then it is only ONE query. Since it was not stated here to unify the datasets then it is best to assume that they will be loaded into two tables. In azure SQL it will be loaded to two tables then in Azure Datalake there will be also two placeholder. Therefore, there are 2 queries and 4 outputs in total.

    upvoted 2 times

👤 **azure_emumba** 4 years, 2 months ago

"NO". we can handle two inputs i.e two SELECT statements in one query. so no need of two queries.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data#joining-multiple-reference-datasets-in-a-job

upvoted 1 times

👤 **maynard13x8** 4 years, 6 months ago

I think the problem is that its a bit confusing what is "one query". Is it all you can do into query editor or is each time you use separated "select - into - from" ? If its the second one, so you need to create 2 queries, each one with different "INTO output"

upvoted 1 times

mohowzeh 4 years, 6 months ago

Good point about "what is a query"...

Two SELECT clauses are needed if there are two INTO statements, as per this link: https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns:

(quote from above link) "Multiple SELECT statements can be used to output data to different output sinks. For example, one SELECT can output a threshold-based alert while another one can output events to blob storage."

upvoted 1 times

---

dumpsm42 4 years, 6 months ago

hi to all,

for it's false

just because we have 1 query and not 2 queries; the query already joins the main data with the reference data so it's one query

the rest is OK

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-troubleshoot-query

regards

upvoted 3 times

> **H_S** 4 years, 4 months ago
>
> it's one query, with 2 outputs and not 4 ?
>
> upvoted 1 times

---

syu31svc 4 years, 7 months ago

Hmm...so long as it meets the goal the answer then the answer is yes then.

Yes the suggested solution does kind of "overdo" it but no harm in it I suppose.

Just my own thoughts about it so I will go with Yes in this case.

upvoted 3 times

---

hart232 4 years, 9 months ago

There is only one requirement in teh question "You need to use an address to look up a median income based on location". 1 Query and 2 outputs makes sense. Not sure why we need the second query for? anyone?

upvoted 2 times

Varma_Saraswathula 4 years, 9 months ago

The online customers will provide a mailing address.

You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure

Does this mean that we need to have 2 diferent queries to get this, if there are 2 queries then 4 outputs

upvoted 1 times

---

dcpavelescu 4 years, 10 months ago

Not clear why 4 outputs are required, as long as no specific requirement to provide separate outputs per customer type (in-store and online customers)

2 query, 2 outputs shall be the right answer accordingly also to:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns

Anyone, is an option with 2 queries and 2 outputs in the exam?

upvoted 3 times

Yan2x 4 years, 11 months ago

There are only two queries, why need for four outputs? I think the answer should be No.

upvoted 1 times

> **nivas143srinivas** 4 years, 11 months ago
>
> The query output need to be stored in two places: SQL DW and Data lake both.
>
> So it's 2x2
>
> upvoted 6 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 plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

☞ A workload for data engineers who will use Python and SQL

☞ A workload for jobs that will run notebooks that use Python, Scala, and SQL

☞ A workload that data scientists will use to perform ad hoc analysis in Scala and R

The enterprise architecture team at your company identifies the following standards for Databricks environments:

☞ The data engineers must share a cluster.

☞ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

☞ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a Standard cluster for the data engineers, and a High Concurrency cluster for the jobs.

Does this meet the goal?

    A. Yes

    B. No

Suggested Answer: *B*

We need a High Concurrency cluster for the data engineers and the jobs.

Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

References:

https://docs.azuredatabricks.net/clusters/configure.html

☐ 👤 **Nieswurz** `Highly Voted 👍` 4 years, 11 months ago

As job notebooks include scala and high concurrency clusters do not support scala, the answer should be no.

upvoted 26 times

   ☐ 👤 **Equalizer** 4 years, 10 months ago

   Correct, check: https://docs.microsoft.com/en-us/azure/databricks/clusters/configure

   upvoted 3 times

   ☐ 👤 **TashaP** 2 years, 9 months ago

   this is 100% correct, it's as simple as this, job requires scala, high concurrency does not support scala. The answer is no.

   upvoted 1 times

☐ 👤 **avix** `Highly Voted 👍` 4 years, 11 months ago

If job needs to use scala then high concurrency cluster can't be used as that wont support Scala

upvoted 9 times

☐ 👤 **ffgghhjj** `Most Recent ⊘` 10 months, 3 weeks ago

B. as high concurrency clusters do not support Scala.

upvoted 1 times

☐ 👤 **hoangton** 4 years, 1 month ago

B is corect

There are two options for cluster mode:

Standard: Single user / small group clusters - can use any language.

High Concurrency: A cluster built for minimizing latency in high concurrency workloads.

There are a few main reasons you would use a Standard cluster over a high concurrency cluster. The first is if you are a single user of Databricks

exploring the technology. For most PoCs and exploration, a Standard cluster should suffice. The second is if you are a Scala user, as high concurrency clusters do not support Scala. The third is if your use case simply does not require high concurrency processes.

High concurrency clusters, in addition to performance gains, also allow you utilize table access control, which is not supported in Standard clusters.

Please note that High Concurrency clusters do not automatically set the auto shutdown field, whereas standard clusters default it to 120 minutes.
https://www.mssqltips.com/sqlservertip/6604/azure-databricks-cluster-configuration/
upvoted 2 times

Note: This question is a part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to an enterprise data warehouse in Azure Synapse Analytics. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Data Warehouse.

Solution:

1. Use Azure Data Factory to convert the parquet files to CSV files

2. Create an external data source pointing to the Azure Data Lake Gen 2 storage account

3. Create an external file format and external table using the external data source

4. Load the data using the CREATE TABLE AS SELECT statement

Does the solution meet the goal?

A. Yes

B. No

Suggested Answer: *A*

It is not necessary to convert the parquet files to CSV files.

You need to create an external file format and external table using the external data source.

You load the data using the CREATE TABLE AS SELECT statement.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

Community vote distribution

B (100%)

👤 **vaseva1** `Highly Voted 👍` 4 years, 2 months ago

It is not necessary to convert the parquet files to CSV files.--> Answer No

upvoted 9 times

👤 **Mily94** 4 years, 2 months ago

indeed, it is not necessary to convert parquet to CSV, but the question is if that apporach meet the goal - A: yes

upvoted 14 times

👤 **FredNo** `Most Recent ⊘` 3 years, 7 months ago

`Selected Answer: B`

No need to convert to CSV File

upvoted 1 times

👤 **nishant_678993** 4 years ago

why would i convert to csv ..When my further steps are going to read data from ADLS.

Explanation says not required and ANSWER says YES.

upvoted 1 times

👤 **maciejt** 4 years, 1 month ago

You can just use data factory to read parquest and ingest to Synapse with auto creating table, but per se the given solution meets the goal even if in not optimal way

upvoted 1 times

👤 **Qrm_1972** 4 years, 1 month ago

Correct answer is (B-------- No)

upvoted 3 times

👤 **KpKo** 4 years, 2 months ago

The answer should be A: Yes. Since it meets the goal though it is not necessary to change the parquet to csv but you can change it.

upvoted 2 times

👤 **Wendy_DK** 4 years, 2 months ago

Correct Answer is B

upvoted 1 times

Ash001 4 years, 2 months ago

It should be No as per the details provided.

upvoted 2 times

Ash001 4 years, 2 months ago

It should be No as per the details provided.

upvoted 2 times

You need to implement complex stateful business logic within an Azure Stream Analytics service.

Which type of function should you create in the Stream Analytics topology?

A. JavaScript user-define functions (UDFs)

B. Azure Machine Learning

C. JavaScript user-defined aggregates (UDA)

Suggested Answer: *C*

Azure Stream Analytics supports user-defined aggregates (UDA) written in JavaScript, it enables you to implement complex stateful business logic. Within UDA you have full control of the state data structure, state accumulation, state decumulation, and aggregate result computation.

References:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-javascript-user-defined-aggregates

👤 **H_S** 4 years, 4 months ago

keyword statufull, otherwise it would be 1

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-javascript-user-defined-functions

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-javascript-user-defined-aggregates

upvoted 2 times

👤 **syu31svc** 4 years, 7 months ago

Link provided supports C as the answer

upvoted 4 times

You have an Azure virtual machine that has Microsoft SQL Server installed. The database on the virtual machine contains a table named Table1.
You need to copy the data from Table1 to an Azure Data Lake Storage Gen2 account by using an Azure Data Factory V2 copy activity.
Which type of integration runtime should you use?

    A. Azure integration runtime

    B. self-hosted integration runtime

    C. Azure-SSIS integration runtime

Suggested Answer: *B*

Copying between a cloud data source and a data source in private network: if either source or sink linked service points to a self-hosted IR, the copy activity is executed on that self-hosted Integration Runtime.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime#determining-which-ir-to-use

samkslee `Highly Voted 👍` 4 years, 2 months ago

Use the self-hosted integration runtime even if the data store is in the cloud on an Azure Infrastructure as a Service (IaaS) virtual machine.
https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime#considerations-for-using-a-self-hosted-ir
upvoted 15 times

    Ambujinee 4 years, 1 month ago

    Correct
    upvoted 2 times

cadio30 `Most Recent ⊘` 4 years, 1 month ago

The answer is correct either the SQL server resides in the traditional virtual server or azure virtual machine it requires "self hosted integration runtime" for the ADF to access the server.
upvoted 1 times

dangal95 4 years, 1 month ago

Why is B the correct answer? There is no mention that the VM running inside of a private network so shouldn't A be the correct answer?
upvoted 2 times

Wendy_DK 4 years, 2 months ago

correct answer is A
upvoted 1 times

    Chiranjib 4 years, 1 month ago

    This is incorrect. Correct Answer is B.. Please refer to samkslee's reply.
    upvoted 1 times

alf99 4 years, 2 months ago

should be "A". All components are on cloud
upvoted 1 times

    SuperAlex 4 years, 2 months ago

    It's SQL Server in the Azure VM server, not Azure SQL DB. Self-hosted integration runtime can be used for SQL Servers hosted on an on-premises machine or on a Azure VM.
    upvoted 6 times

DRAG DROP -

Your company plans to create an event processing engine to handle streaming data from Twitter.

The data engineering team uses Azure Event Hubs to ingest the streaming data.

You need to implement a solution that uses Azure Databricks to receive the streaming data from the Azure Event Hubs.

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

Select and Place:

Actions

| Create and configure a Notebook that consumes the streaming data. |
| Import data from Blob storage. |
| Use Environment variables to define the Apache Spark connection. |
| Configure an ODBC or JDBC Connector. |
| Deploy the Azure Databricks service. |
| Deploy a Spark cluster and then attach the required libraries to the cluster. |

Answer Area

Suggested Answer:

Actions

| |
| Import data from Blob storage. |
| Use Environment variables to define the Apache Spark connection. |
| Configure an ODBC or JDBC Connector. |
| |
| |

Answer Area

| Deploy the Azure Databricks service. |
| Deploy a Spark cluster and then attach the required libraries to the cluster. |
| Create and configure a Notebook that consumes the streaming data. |

Step 1: Deploy the Azure Databricks service

Create an Azure Databricks workspace by setting up an Azure Databricks Service.

Step 2: Deploy a Spark cluster and then attach the required libraries to the cluster.

To create a Spark cluster in Databricks, in the Azure portal, go to the Databricks workspace that you created, and then select Launch Workspace.

Attach libraries to Spark cluster: you use the Twitter APIs to send tweets to Event Hubs. You also use the Apache Spark Event Hubs connector to read and write data into Azure Event Hubs. To use these APIs as part of your cluster, add them as libraries to Azure Databricks and associate

them with your Spark cluster.

Step 3: Create and configure a Notebook that consumes the streaming data.

You create a notebook named ReadTweetsFromEventhub in Databricks workspace. ReadTweetsFromEventHub is a consumer notebook you use to read the tweets from Event Hubs.

References:

https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-stream-from-eventhubs

---

ck1729 `Highly Voted 👍` 4 years, 3 months ago

Given answer is correct

upvoted 6 times

---

maciejt `Most Recent ⊘` 4 years, 1 month ago

But you don't need a lbrary in Spark to connect to Twotter as the data is already in EvenHub that you only need to connect to

upvoted 1 times

---

syu31svc 4 years, 7 months ago

https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-stream-from-eventhubs supports answer provided

upvoted 1 times

UKiran 4 years, 7 months ago

Its mandatory to create Spark Cluster, and then create a Notebook. As the notebook needs a cluster to run.

upvoted 1 times

---

serger 5 years ago

Always mandatory to create a spark cluster from databricks and then you can run a notebook (the notebook should be attached to an existing spark cluster).

upvoted 4 times

Luke97 5 years, 1 month ago

Refer this article, https://docs.databricks.com/spark/latest/structured-streaming/streaming-event-hubs.html

In order to connect Azure Event Hub via Azure Event Hubs Spark Connector in Databricks, we need to deploy Spark Cluster and install the new created library in your Databricks workspace using the Maven coordinate.

upvoted 2 times

> **Luke97** 5 years, 1 month ago
>
> This tutorial explains more details. https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-stream-from-eventhubs
>
> upvoted 6 times

---

mclawson1966 5 years, 3 months ago

why is the Spark Cluster included in the answer?

upvoted 1 times

> **mclawson1966** 5 years, 3 months ago
>
> I guess Spark is the only way to connect Databricks to an Event Hub?
>
> upvoted 1 times

> **Yuri1101** 5 years, 2 months ago
>
> Spark has built-in API to retrieve Twitter data
>
> upvoted 4 times

> **gunjan075** 5 years, 2 months ago
>
> i guess to run notebook in databricks you need to first create spark cluster
>
> upvoted 4 times

> **SebK** 4 years, 11 months ago
>
> Databricks is a spark-based technology. It is always mandatory to have a spark cluster for Databricks to work.
>
> upvoted 8 times

HOTSPOT -

You develop data engineering solutions for a company.

A project requires an in-memory batch data processing solution.

You need to provision an HDInsight cluster for batch processing of data on Microsoft Azure.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
┌────────────────────────────────┬───▼───┐
│ New-AzureStorageContainer              │
│ New-AzureRmHDInsightClusterConfig      │
│ New-AzureRmHDInsightCluster            │
└────────────────────────────────────────┘
  -Name $clusterName -Context $defaultStorageContext
$objectConfig = New Object "System.Collections.Generic.Dictionary''2[System.String,System.String]"
&objectConfig.Add ┌──────────────────────────┬──▼──┐ ,"2.3")
                  │ "spark"                        │
                  │ "haddop"                       │
                  └──────────────────────────────────┘

(
┌────────────────────────────────┬───▼───┐ '
│ New-AzureRmHDInsightCluster            │
│ New-AzureRmHDInsightClusterConfig      │
│ New-AzureStorageContainer              │
└────────────────────────────────────────┘
-ResourceGroupName $resourceGroupName'
-ClusterName $clustername'
-Location $location `
-ClusterSizeInNodes $clusterSizeInNodes `
-ClusterType ┌──────────────────────────┬──▼──┐ ,
             │ "spark"                        │
             │ "hadoop"                       │
             │ "HBase"                        │
             │ "Storm"                        │
             └──────────────────────────────────┘
-OSType $clusterOS'
-Version $clusterVersion'
-ComponentVersion $objectConfig'
-HttpCredential $httpCredential'
-DefaultStorageAccountName
"$defaultStorageAccountName.blob.core.windows.net" `
-DefaultStorageAccountKey $defaultStorageAccountKey'
-DefaultStorageContainer $clusterName'
-SshCredential $sshCredentials
```

Suggested Answer:

Answer Area

Box 1: New-AzStorageContainer -
Example: Create a blob container. This holds the default data store for the cluster.
New-AzStorageContainer `
-Name $clusterName `
-Context $defaultStorageContext
$sparkConfig = New-Object "System.Collections.Generic.Dictionary``2[System.String,System.String]"
$sparkConfig.Add("spark", "2.3")

Box 2: Spark -
Spark provides primitives for in-memory cluster computing. A Spark job can load and cache data into memory and query it repeatedly. In-memory computing is much faster than disk-based applications than disk-based applications, such as Hadoop, which shares data through Hadoop distributed file system (HDFS).

Box 3: New-AzureRMHDInsightCluster
Create the HDInsight cluster. Example:
New-AzHDInsightCluster `
-ResourceGroupName $resourceGroupName `
-ClusterName $clusterName `
-Location $location `
-ClusterSizeInNodes $clusterSizeInNodes `
-ClusterType $"Spark" `
-OSType "Linux" `

Box 4: Spark -
HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP,
MapReduce.
References:

  SebK `Highly Voted 👍` 4 years, 11 months ago

This question is no more part of the DP-200 exam.

upvoted 13 times

    ZekroMancer 4 years, 7 months ago

why not?

upvoted 2 times

  cadio30 `Most Recent ⊘` 4 years, 1 month ago

Given answer is correct

upvoted 2 times

  mohowzeh 4 years, 6 months ago

imho the answer is correct, see https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-create-linux-clusters-azure-powershell

upvoted 1 times

  runningman 5 years, 1 month ago

second answer is clearly spark because haddop is misspelled :>}

upvoted 3 times

HOTSPOT -

A company plans to develop solutions to perform batch processing of multiple sets of geospatial data.

You need to implement the solutions.

Which Azure services should you use? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Scenario	Tool
Use a native client application to run interactive queries and batch processes.	▼
	HDInsight Tools for Visual Studio
	Hive View
	HDInsight REST API
	Azure Data Factory
Use a web browser to run interactive queries and batch processes.	▼
	HDInsight Tools for Visual Studio
	Hive View
	HDInsight REST API
	Azure PowerShell
Develop batch processing applications that use Azure HDInsight	▼
	HDInsight Tools for Visual Studio
	Hive View
	HDInsight REST API
	NoSQL database

Suggested Answer:

Answer Area

Scenario	Tool
Use a native client application to run interactive queries and batch processes.	▼
	HDInsight Tools for Visual Studio
	Hive View
	HDInsight REST API
	Azure Data Factory
Use a web browser to run interactive queries and batch processes.	▼
	HDInsight Tools for Visual Studio
	Hive View
	HDInsight REST API
	Azure PowerShell
Develop batch processing applications that use Azure HDInsight	▼
	HDInsight Tools for Visual Studio
	Hive View
	HDInsight REST API
	NoSQL database

Box 1: HDInsight Tools for Visual Studio

Azure HDInsight Tools for Visual Studio Code is an extension in the Visual Studio Code Marketplace for developing Hive Interactive Query, Hive Batch Job and

PySpark Job against Microsoft HDInsight.

Box 2: Hive View -

You can use Apache Ambari Hive View with Apache Hadoop in HDInsight. The Hive View allows you to author, optimize, and run Hive queries from your web browser.

Box 3: HDInsight REST API -
Azure HDInsight REST APIs are used to create and manage HDInsight resources through Azure Resource Manager.
References:
https://visualstudiomagazine.com/articles/2019/01/25/vscode-hdinsight.aspx https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-ambari-view https://docs.microsoft.com/en-us/rest/api/hdinsight/

Luke97 `Highly Voted 👍` 5 years, 1 month ago

I think box 3 should also be HDInsight Tools for VS. "Azure HDInsight Tools for Visual Studio Code is an extension in the Visual Studio Code Marketplace for developing Hive Interactive Query, Hive Batch Job and PySpark Job against Microsoft HDInsight."

https://marketplace.visualstudio.com/items?itemName=mshdinsight.azure-hdinsight

upvoted 16 times

Manue `Highly Voted 👍` 5 years, 2 months ago

Honestly, I think quesions and answes make no sense at all. "Hive View" is not a tool to run interactive queries and batch processes. On the other hand, "REST API" is not a tool to designed to develop "batch applications". It's very hard to try to make sense of these questions/answers...

upvoted 15 times

dangal95 `Most Recent ⊙` 4 years, 1 month ago

I think that the third answer should also be HDInsight Tools for Visual Studio simply because the questions starts with "Develop..." which means we need to develop the batch job (and possibly push it) to the cluster. The HDInsight API only lets us publish batch jobs through pre-existing JARs which container the batch processing logic / code within itself.

upvoted 1 times

syu31svc 4 years, 7 months ago

I would say HDInsight Tools for Visual Studio is the answer for both the 1st and 3rd dropdown

Hive view is answer for the 2nd dropdown

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-for-vscode

https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-ambari-view

upvoted 3 times

    syu31svc 4 years, 7 months ago

    On second thought, Azure HDInsight REST APIs for batch processing of applications that use HDInsight

    https://docs.microsoft.com/en-us/rest/api/hdinsight/hdinsight-application

    upvoted 2 times

wyxh 5 years, 2 months ago

Hive View is designed to help you author, optimize, and execute queries. With Hive Views you can:

Browse databases.

Write queries or browse query results in full-screen mode, which can be particularly helpful with complex queries or large query results.

Manage query execution jobs and history.

View existing databases, tables, and their statistics.

Create/upload tables and export table DDL to source control.

View visual explain plans to learn more about query plan.

upvoted 5 times

zenomas 5 years, 3 months ago

I think box 1 and box 3 need to switch.

- Visual Studio is a development environment where development is done.

- REST API is used to call from the application being developed to HDInsight Cluster to perform actions.

Any thoughts!

upvoted 6 times

    Leonido 5 years, 2 months ago

    Don't think so. The only native application in the list is VS code, and REST api could be used to create a job processing.

    upvoted 1 times

DRAG DROP -

You are creating a managed data warehouse solution on Microsoft Azure.

You must use PolyBase to retrieve data from Azure Blob storage that resides in parquet format and load the data into a large table called FactSalesOrderDetails.

You need to configure Azure Synapse Analytics to receive the data.

Which four 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.

Select and Place:

Actions

Create an external file format to map the parquet files.
Load the data to a staging table.
Create the external table FactSalesOrderDetails.
Enable Transparent Data Ecnryption.
Create an external data source for Azure Blob storage.
Create a master key on database.

Answer Area

Suggested Answer:

Actions

Load the data to a staging table.
Enable Transparent Data Ecnryption.

Answer Area

Create a master key on database.
Create an external data source for Azure Blob storage.
Create an external file format to map the parquet files.
Create the external table FactSalesOrderDetails.

Step 1: Create a master key on the database

Create a master key on the database. This is required to encrypt the credential secret.

Step 2: Create an external data source for Azure Blob storage

Create an external data source with CREATE EXTERNAL DATA SOURCE..

Step 3: Create an external file format to map parquet files.

Create an external file format with CREATE EXTERNAL FILE FORMAT.

FORMAT TYPE: Type of format in Hadoop (DELIMITEDTEXT, RCFILE, ORC, PARQUET).

Step 4: Create the external table FactSalesOrderDetails

To query the data in your Hadoop data source, you must define an external table to use in Transact-SQL queries.

Create an external table pointing to data stored in Azure storage with CREATE EXTERNAL TABLE.

Note: PolyBase is a technology that accesses and combines both non-relational and relational data, all from within SQL Server. It allows you to run queries on external data in Hadoop or Azure blob storage.

Reference:

https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-configure-azure-blob-storage

---

hoangton `Highly Voted 👍` 4 years ago

The answer is correct

Step to Configure an external table

1. Create a master key on the database. The master key is required to encrypt the credential secret.

2. Create a database scoped credential for Azure blob storage.

3. Create an external data source with CREATE EXTERNAL DATA SOURCE..

4. Create an external file format with CREATE EXTERNAL FILE FORMAT.

5. Create an external table pointing to data stored in Azure storage with CREATE EXTERNAL TABLE

6. Create statistics on an external table.

upvoted 5 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 are developing a solution that will use Azure Stream Analytics. The solution will accept an Azure Blob storage file named Customers. The file will contain both in-store and online customer details. The online customers will provide a mailing address.

You have a file in Blob storage named LocationIncomes that contains median incomes based on location. The file rarely changes.

You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure

Data Lake Storage Gen2 for long-term retention.

Solution: You implement a Stream Analytics job that has two streaming inputs, one query, and two outputs.

Does this meet the goal?

A. Yes

B. No

Suggested Answer: *B*

We need one reference data input for LocationIncomes, which rarely changes

Note: Stream Analytics also supports input known as reference data. Reference data is either completely static or changes slowly.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs#stream-and-reference-inputs

---

☐ 👤 **syu31svc** 4 years, 7 months ago

You need a reference so answer is no

upvoted 4 times

DRAG DROP -

You develop data engineering solutions for a company.

You need to deploy a Microsoft Azure Stream Analytics job for an IoT solution. The solution must:

☞ Minimize latency.

☞ Minimize bandwidth usage between the job and IoT device.

Which four 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.

Select and Place:

Actions

Configure routes.
Create an Azure Data Lake Storage container.
Create an IoT Hub and add the Azure Stream Analytics module to the IoT Hub namespace.
Create an Azure Stream Analytics edge job and configure job definition save location.
Create an Azure Stream Analytics cloud job and configure job definition save location.
Create an Azure Blob storage container.
Configure Streaming Units.

Answer Area

Suggested Answer:

Actions

Create an Azure Data Lake Storage container.
Create an Azure Stream Analytics cloud job and configure job definition save location.
Configure Streaming Units.

Answer Area

Create an Azure Blob storage container.
Create an Azure Stream Analytics edge job and configure job definition save location.
Create an IoT Hub and add the Azure Stream Analytics module to the IoT Hub namespace.
Configure routes.

Step 1: Create an Azure Blob Storage container

To prepare your Stream Analytics job to be deployed on an IoT Edge device, you need to associate the job with a container in a storage account.

When you go to deploy your job, the job definition is exported to the storage container.

Step 2: Create an Azure Stream Analytics edge job and configure job definition save location

When you create an Azure Stream Analytics job to run on an IoT Edge device, it needs to be stored in a way that can be called from the device.

Step 3: Create and IoT hub and add the Azure Stream Analytics module to the IoT Hub namespace
An IoT Hub in Azure is required.
Stream Analytics accepts data incoming from several kinds of event sources including Event Hubs, IoT Hub, and Blob storage.

Step 4: Configure routes -
You are now ready to deploy the Azure Stream Analytics job on your IoT Edge device.
The routes that you declare define the flow of data through the IoT Edge device.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs https://docs.microsoft.com/en-us/azure/iot-edge/tutorial-deploy-stream-analytics https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge

SAMBIT `Highly Voted 👍` 5 years, 4 months ago

Wrong order:

Correct order can be found in the article

1) storage
2) Edge
3) IOT
4) route

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge
upvoted 74 times

> **dev2dev** 4 years, 5 months ago
> but the link doesnt give any mention about the order
> upvoted 4 times

> > **Billybob0604** 2 years, 6 months ago
> > what would in sequence mean to you ??
> > upvoted 1 times

> **Sasidhar39** 4 years, 1 month ago
> This is wrong order, provided answer is correct. Refer the below link
> https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal
> upvoted 4 times

> **Alekx42** 4 years, 1 month ago
> SAMBIT's answer is correct. When you add the Stream Analytics module to the IoT hub you must select an existing Stream Analytics job. So you must create the Stream Analytics Edge job first
> upvoted 2 times

> **Ambujinee** 4 years, 1 month ago
> Correct ans
> upvoted 2 times

kova123 `Highly Voted 👍` 4 years, 9 months ago

1. Create Storage
2. Create ASA Job
3. Create IOT Hug and Add ASA Job
4. Configure Routes
This is correct order
upvoted 12 times

Wendy_DK `Most Recent ⊘` 4 years, 2 months ago

: Create an IoT hub and add the Azure Stream Analytics module to the IoT Hub namespace
An IoT Hub in Azure is required. It's a Prerequisites. So the answer is correct
upvoted 2 times

Lolo_alzin19 4 years, 3 months ago

answer is in correct.
upvoted 2 times

👤 **syu31svc** 4 years, 7 months ago

From https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge:

Azure Stream Analytics (ASA) on IoT Edge empowers developers to deploy near-real-time analytical intelligence closer to IoT devices so that they can unlock the full value of device-generated data. Azure Stream Analytics is designed for low latency, resiliency, efficient use of bandwidth, and compliance

1) Create a storage container (create blob container)

2) Create an ASA edge job (create edge job)

3) Setup your IoT Edge environment on your device(s) (create IoT hub)

4) Deploy ASA on your IoT Edge device(s) (configure routes)

upvoted 5 times

👤 **Arsa** 4 years, 10 months ago

Given answer is correct as per the article.. you need to create IOT hub first

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal..

upvoted 4 times

> 👤 **dumpsm42** 4 years, 6 months ago
>
> yes. this is the answer, the link is perfect, it has the sequence, the only difference is that we use a Edge instead of Cloud Job, like the text says.
>
> Regards
>
> upvoted 1 times

> 👤 **cadio30** 4 years, 1 month ago
>
> Stated clearly in the documentation the procedure on the configuration. Cheers!
>
> upvoted 2 times
>
> > 👤 **cadio30** 4 years, 1 month ago
> >
> > Reference: https://rangv.github.io/azureiotedgelab/streamanalytics/
> >
> > upvoted 2 times

👤 **krisspark** 4 years, 11 months ago

why Azure blob storage only why not Azure Data Lake storage?

upvoted 1 times

> 👤 **sandGrain** 4 years, 9 months ago
>
> Data lake is not a accepted input in ASA only Blob, Iot Hib and Event Hub
>
> upvoted 2 times

👤 **Leonido** 5 years, 2 months ago

Based on what I see, storage can be defined before or after the Edge, as long as it's defined before the input definition. However, it makes more sense to create Storage account before Hub, so it will be streamlines.

upvoted 1 times

> 👤 **wyxh** 5 years, 1 month ago
>
> i agree,
>
> Create a storage account
>
> When you create an Azure Stream Analytics job to run on an IoT Edge device, it needs to be stored in a way that can be called from the device. You can use an existing Azure Storage account, or create a new one now.
>
> https://docs.microsoft.com/en-us/azure/iot-edge/tutorial-deploy-stream-analytics
>
> upvoted 2 times

👤 **Amitkhanna** 5 years, 3 months ago

Given answer is correct refer below article

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal

upvoted 4 times

> 👤 **jdpl** 5 years, 3 months ago
>
> this article refers to stream analytics job hosting in the cloud (!). This question refers to the job on edge.
>
> upvoted 6 times

DRAG DROP -

You have data stored in thousands of CSV files in Azure Data Lake Storage Gen2. Each file has a header row followed by a property formatted carriage return (/r) and line feed (/n).

You are implementing a pattern that batch loads the files daily into an enterprise data warehouse in Azure Synapse Analytics by using PolyBase.

You need to skip the header row when you import the files into the data warehouse. Before building the loading pattern, you need to prepare the required database objects in Azure Synapse Analytics.

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.

Select and Place:

Actions

- Create an external data source that uses the abfs location.
- Create an external file format and set the First_Row option.
- Create a database scoped credential that uses Azure Active Directory Application and a Service Principal Key
- Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and create a view that removes the empty row.
- Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and configure the reject options to specify reject values or percentages.

Answer Area

(empty)

Suggested Answer:

Actions

- Create a database scoped credential that uses Azure Active Directory Application and a Service Principal Key
- Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and create a view that removes the empty row.

Answer Area

1. Create an external data source that uses the abfs location.
2. Create an external file format and set the First_Row option.
3. Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and configure the reject options to specify reject values or percentages.

Step 1: Create an external data source that uses the abfs location

Create External Data Source to reference Azure Data Lake Store Gen 1 or 2

Step 2: Create an external file format and set the First_Row option.

Create External File Format.

Step 3: Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and configure the reject options to specify reject values or percentages

To use PolyBase, you must create external tables to reference your external data.

Use reject options.

Note: REJECT options don't apply at the time this CREATE EXTERNAL TABLE AS SELECT statement is run. Instead, they're specified here so that the database can use them at a later time when it imports data from the external table. Later, when the CREATE TABLE AS SELECT statement selects data from the external table, the database will use the reject options to determine the number or percentage of rows that can fail to import before it stops the import.

Reference:

https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-t-sql-objects https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-as-select-transact-sql

👤 **javedjss** 3 years, 5 months ago

Correct Answer.

upvoted 1 times

https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-t-sql-objects https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-as-select-transact-sql

👤 **javedjss** 3 years, 5 months ago

Correct Answer.

upvoted 1 times

You are creating a new notebook in Azure Databricks that will support R as the primary language but will also support Scala and SQL.
Which switch should you use to switch between languages?

    A. %<language>

    B. \\[<language>]

    C. \\(<language>)

    D. @<Language>

Suggested Answer: A
You can override the primary language by specifying the language magic command %<language> at the beginning of a cell. The supported magic commands are:
%python, %r, %scala, and %sql.
References:
https://docs.databricks.com/user-guide/notebooks/notebook-use.html#mix-languages

⊟ 👤 **akash0680** Highly Voted 👍 4 years, 6 months ago
Correct Answer :)
upvoted 18 times

⊟ 👤 **Allapanda** Most Recent ⊙ 3 years, 11 months ago
not exactly. its %%
upvoted 1 times

   ⊟ 👤 **tanssive** 3 years, 9 months ago
in Databricks you use just '%' e.g. %scala ... CODE
upvoted 1 times

HOTSPOT -

You are implementing mapping data flows in Azure Data Factory to convert daily logs of taxi records into aggregated datasets.

You configure a data flow and receive the error shown in the following exhibit.

You need to resolve the error.

Which setting should you configure? To answer, select the appropriate setting in the answer area.

Hot Area:

Answer Area

Suggested Answer:

Answer Area

The Inspect tab provides a view into the metadata of the data stream that you're transforming. You can see column counts, the columns changed, the columns added, data types, the column order, and column references. Inspect is a read-only view of your metadata. You don't need

to have debug mode enabled to see metadata in the Inspect pane.

	Derived column's settings	Optimize	Inspect	Data Preview				Description

Output schema   Input schema

Number of columns		New* 1		Updated* 2		Unchanged 4		Total 7

Order ⬍	Column ⬍	Type ⬍	Updated ⬍	Based on ⬍
1	movie	abc string		
2	title	abc string	•	title
3	genres	abc string		
4	year	121 long	•	year
5	Rating	abc string		
6	Rotton Tomato	abc string		
7	Rotten Tomato	121 long	•	Rotton Tomato

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/concepts-data-flow-overview

---

👤 **tucho** `Highly Voted 👍` 4 years, 2 months ago

passengerCount column type

upvoted 6 times

> 👤 **Pairon** 4 years, 2 months ago
>
> I agree...with inspect you can also look at data types, you can't edit it. So I would edit data type for passengerCount.
>
> upvoted 4 times

👤 **Shrutii** `Most Recent ⊘` 1 year, 6 months ago

passengerCount

upvoted 1 times

👤 **hoangton** 4 years ago

Config PassengerCount column type to number

upvoted 2 times

👤 **maciejt** 4 years, 1 month ago

The anwer is incorrect as you cannot solve the error using read-only tab. The question was which section to configure and you cannot configure the read-only inspect. You need to configure schema in projection

upvoted 2 times

> 👤 **emski** 4 years ago
>
> https://docs.microsoft.com/en-us/azure/data-factory/data-flow-source
>
> upvoted 1 times

HOTSPOT -

You have an Azure SQL database named Database1 and two Azure event hubs named HubA and HubB. The data consumed from each source is shown in the following table.

Source	Data
Database1	Driver's name Driver's license number
HubA	Ride route Ride distance Ride duration
HubB	Ride fare Ride payment

You need to implement Azure Stream Analytics to calculate the average fare per mile by driver.

How should you configure the Stream Analytics input for each source? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

HubA:
- Stream
- Reference

HubB:
- Stream
- Reference

Database1:
- Stream
- Reference

Suggested Answer:

Answer Area

HubA:
- **Stream**
- Reference

HubB:
- **Stream**
- Reference

Database1:
- Stream
- **Reference**

HubA: Stream -

HubB: Stream -

Database1: Reference -

Reference data (also known as a lookup table) is a finite data set that is static or slowly changing in nature, used to perform a lookup or to augment your data streams. For example, in an IoT scenario, you could store metadata about sensors (which don't change often) in reference data and join it with real time IoT data streams. Azure Stream Analytics loads reference data in memory to achieve low latency stream

processing
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

hoangton `Highly Voted 👍` 4 years, 1 month ago

It is correct Answer

upvoted 8 times

hoangton `Highly Voted 👍` 4 years, 1 month ago

It is correct Answer

upvoted 8 times

HOTSPOT -

You have a self-hosted integration runtime in Azure Data Factory.

The current status of the integration runtime has the following configurations:

☞ Status: Running

☞ Type: Self-Hosted

☞ Version: 4.4.7292.1

☞ Running / Registered Node(s): 1/1

☞ High Availability Enabled: False

☞ Linked Count: 0

☞ Queue Length: 0

☞ Average Queue Duration: 0.00s

The integration runtime has the following node details:

☞ Name: X-M

☞ Status: Running

☞ Version: 4.4.7292.1

☞ Available Memory: 7697MB

☞ CPU Utilization: 6%

☞ Network (In/Out): 1.21KBps/0.83KBps

☞ Concurrent Jobs (Running/Limit): 2/14

☞ Role: Dispatcher/Worker

☞ Credential Status: In Sync

Use the drop-down menus to select the answer choice that completes each statement based on the information presented.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

If the X-M node becomes unavailable, all executed pipelines will:

- fail until the node comes back online
- switch to another integration runtime
- exceed the CPU limit

The number of concurrent jobs and the CPU usage indicate that the Concurrent Jobs (Running/Limit) value should be:

- raised
- lowered
- left as is

Suggested Answer:

Answer Area

If the X-M node becomes unavailable, all executed pipelines will:

- **fail until the node comes back online**
- switch to another integration runtime
- exceed the CPU limit

The number of concurrent jobs and the CPU usage indicate that the Concurrent Jobs (Running/Limit) value should be:

- raised
- **lowered**
- left as is

Box 1: fail until the node comes back online

We see: High Availability Enabled: False

Note: Higher availability of the self-hosted integration runtime so that it's no longer the single point of failure in your big data solution or cloud data integration with

Data Factory.

Box 2: lowered -
We see:
Concurrent Jobs (Running/Limit): 2/14
CPU Utilization: 6%
Note: When the processor and available RAM aren't well utilized, but the execution of concurrent jobs reaches a node's limits, scale up by increasing the number of concurrent jobs that a node can run
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

MMM777 `Highly Voted 👍` 4 years, 1 month ago
If each job is similar and each uses 3% CPU - then even at current max 14,only 42% CPU utilization. Should INCREASE the limit - which is what the answer is also saying.
upvoted 12 times

111222333 4 years, 1 month ago
Agree with "raise".
If we lower the number of concurrent jobs, the CPU will become even more *under*utilized. Because then when it reaches the limit number of concurrent jobs, it will occupy even less CPU -> CPU will never be optimally utilized.
RAISE the number of concurrent jobs so that their CPU consumption comes closer to CPU's max.

Reference: https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime#scale-up
upvoted 1 times

michalS 3 years, 10 months ago
You are not increasing currently running jobs, but the limit of jobs that can run concurrently, which is 14 at the moment. So increasing it would only reduce cpu utilization.
upvoted 1 times

Simon2021 `Most Recent ⊘` 4 years ago
It should lower the concurrent Jobs limit.
upvoted 1 times

Pierwiastek 4 years ago
I agree with the answer. We should lower Concurrent Jobs (Running/Limit) value. It means we should lower the limit value. We don't need so big value if we use only 2 of 14.
upvoted 2 times

dangal95 4 years, 1 month ago
Shouldn't the second option be "left as is" since the number of concurrent jobs is 2/14? The answer states that if the number of concurrent jobs reaches the limit (14 in this case) and the CPU is under utilized then we should increase the number of concurrent jobs that can run. In this case it seems that 6% is a good amount of utilization given that 2/14 jobs (14%) are running, no?
upvoted 3 times

cadio30 4 years, 1 month ago
"When the processor and available RAM aren't well utilized, but the execution of concurrent jobs reaches a node's limits, scale up by increasing the number of concurrent jobs that a node can run. You might also want to scale up when activities time out because the self-hosted IR is overloaded. As shown in the following image, you can increase the maximum capacity for a node."

The above statement was quoted from the link provided below and on this scenario, it requires to increase the concurrent jobs (running/limit)

Reference: https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime
upvoted 2 times

Igtiza 4 years ago
I agree with Dangal, "left as is", because here you are not reaching a node's limits (it's just using 6%). It it was using the 14 jobs and it would be using 20% then I would say do raise the number of possible jobs, otherwise not. And also I would not lower the number of possible jobs because I don't know how many concurrent jobs I would need to run, this is just a "snapshot" of what was happening at that moment.
upvoted 1 times

You have an Azure Stream Analytics job that receives clickstream data from an Azure event hub.

You need to define a query in the Stream Analytics job. The query must meet the following requirements:

☞ Count the number of clicks within each 10-second window based on the country of a visitor.

☞ Ensure that each click is NOT counted more than once.

How should you define the query?

    A. SELECT Country, Count(*) AS Count FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, TumblingWindow(second, 10)

    B. SELECT Country, Count(*) AS Count FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SessionWindow(second, 5, 10)

    C. SELECT Country, Avg(*) AS Average FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SlidingWindow(second, 10)

    D. SELECT Country, Avg(*) AS Average FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, HoppingWindow(second, 10, 2)

Suggested Answer: *A*

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

Example:

Incorrect Answers:

B: Session windows group events that arrive at similar times, filtering out periods of time where there is no data.

C: Sliding windows, unlike Tumbling or Hopping windows, output events only for points in time when the content of the window actually changes. In other words, when an event enters or exits the window. Every window has at least one event, like in the case of Hopping windows, events can belong to more than one sliding window.

D: Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

---

👤 **meswapnilspal** `Highly Voted 👍` 4 years, 1 month ago

answer is correct

https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

upvoted 6 times

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets from the last five minutes every minute.

Which windowing function should you use?

    A. Sliding

    B. Session

    C. Tumbling

    D. Hopping

Suggested Answer: *D*

Hopping window functions hop forward in time by a fixed period.

Incorrect Answers:

A: Sliding windows, unlike Tumbling or Hopping windows, output events only for points in time when the content of the window actually changes. In other words, when an event enters or exits the window.

B: Session window functions group events that arrive at similar times, filtering out periods of time where there is no data. A session window begins when the first event occurs. If another event occurs within the specified timeout from the last ingested event, then the window extends to include the new event. Otherwise if no events occur within the timeout, then the window is closed at the timeout.

C: Tumbling window functions are used to segment a data stream into distinct time segments. A Tumbling windows do not overlap, and an event cannot belong to more than one tumbling window.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

---

👤 **SankarMG** `Highly Voted 👍` 4 years ago

Hopping Window is the correct answer

upvoted 12 times

    👤 **victor90** 3 years, 7 months ago

    https://docs.microsoft.com/en-us/azure/stream-analytics/media/stream-analytics-window-functions/stream-analytics-window-functions-hopping-intro.png

    upvoted 1 times

👤 **dumpi** `Most Recent ⊘` 4 years ago

correct answer c only

upvoted 1 times

    👤 **captainbee** 3 years, 11 months ago

    Don't be ridiculous, doesn't help other people either. It's a hopping window as the windows will overlap.

    upvoted 7 times

👤 **Sudhansu21** 4 years, 1 month ago

Hopping

upvoted 4 times

👤 **kaigalmane** 4 years, 1 month ago

Sorry. Looks like D now.

upvoted 1 times

👤 **kaigalmane** 4 years, 1 month ago

I think C. Tumbling

upvoted 1 times

👤 **kaigalmane** 4 years, 1 month ago

So what is the answer?

upvoted 2 times

👤 **Mily94** 4 years, 2 months ago

correct answer: D

upvoted 3 times

imkoomin 4 years, 3 months ago

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

upvoted 3 times

imkoomin 4 years, 3 months ago

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

upvoted 3 times

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets during the last five minutes every five minutes. Each tweet must only be counted once.

Which windowing function should you use?

A. a five-minute Sliding window

B. a five-minute Session window

C. a five-minute Tumbling window

D. a five-minute Hopping window that has a one-minute hop

Suggested Answer: *C*

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.
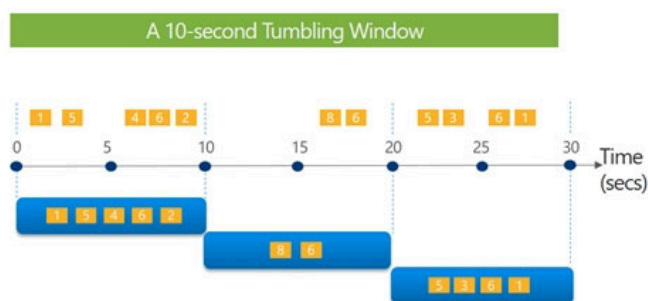
Incorrect Answers:

D: Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

👤 **09012020** `Highly Voted 👍` 4 years, 9 months ago

I think answer is correct. TUMBLING WINDOW as event muse be counted once.

Tumbling window

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window

Tumbling window

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window

upvoted 20 times

👤 **dinesh_tng** 4 years ago

Actually, with the given options, answer is correct. Ans - C

upvoted 1 times

👤 **maciejt** `Most Recent ⊘` 4 years, 1 month ago

Every tween must be counted only once - tumbling window, unlike hopping, guarantees the windows will not overlap

upvoted 3 times

☐ 👤 **tucho** 4 years, 2 months ago

I think C is the correct answer. Definition of tumbling window: "fixed size window with advance interval equals to the window size".

Answer D is clearly wrong: "a five-minute Hopping window that has a one-minute hop" doesnt fit "five minutes every five minutes" requirement.

upvoted 2 times

☐ 👤 **Rahul111** 4 years, 12 months ago

A hoping window can act as tumbling window, if time unit and window size is same.

upvoted 2 times

☐ 👤 **MLCL** 4 years, 11 months ago

D. a five-minute Hopping window that has one-minute hop

One minute hop Rahul, 1 min hop.

upvoted 4 times

☐ 👤 **oku** 5 years ago

It must be Hopping Window -To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size. So Answer is D

upvoted 1 times

☐ 👤 **Ikrom** 5 years ago

It would be if the hop is 5 minutes as well as the window, BUT the answer "D" says this:

"D. a five-minute Hopping window that has one-minute hop"

which means overlap will occur.

So, the answer is "C" and that's correct.

upvoted 9 times

☐ 👤 **avestabrzn** 5 years, 3 months ago

I would say just Hopping Window.

upvoted 1 times

☐ 👤 **z8zhong** 5 years, 3 months ago

Hopping Window has overlapped and an item can be counted more than once

upvoted 9 times

You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.
Which input type should you use for the reference data?

    A. Azure Cosmos DB

    B. Azure Event Hubs

    C. Azure Blob storage

    D. Azure IoT Hub

Suggested Answer: *C*
Stream Analytics supports Azure Blob storage and Azure SQL Database as the storage layer for Reference Data.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

👤 **Yaswant** `Highly Voted 👍` 4 years, 11 months ago

For reference data Azure SQL database and Blob storage are supported.

upvoted 12 times

👤 **syu31svc** `Most Recent ⊘` 4 years, 7 months ago

Link provided supports C as the answer

upvoted 2 times

👤 **sunil_kalra** 4 years, 7 months ago

That is true, but it SQL DB is not in options

upvoted 2 times

HOTSPOT -

You are implementing Azure Stream Analytics windowing functions.

Which windowing function should you use for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap:
- Hopping
- Sliding
- Tumbling

Segment the data stream into distinct time segments that repeat and can overlap:
- Hopping
- Sliding
- Tumbling

Segment the data stream to produce an output only when an event occurs:
- Hopping
- Sliding
- Tumbling

Suggested Answer:

Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap:
- Hopping
- Sliding
- **Tumbling**

Segment the data stream into distinct time segments that repeat and can overlap:
- **Hopping**
- Sliding
- Tumbling

Segment the data stream to produce an output only when an event occurs:
- Hopping
- **Sliding**
- Tumbling

Box 1: Tumbling -

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

Tell me the count of tweets per time zone every 10 seconds
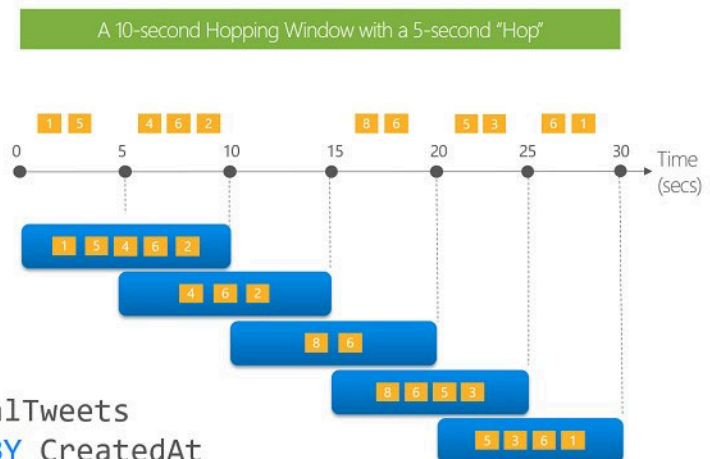
A 10-second Tumbling Window
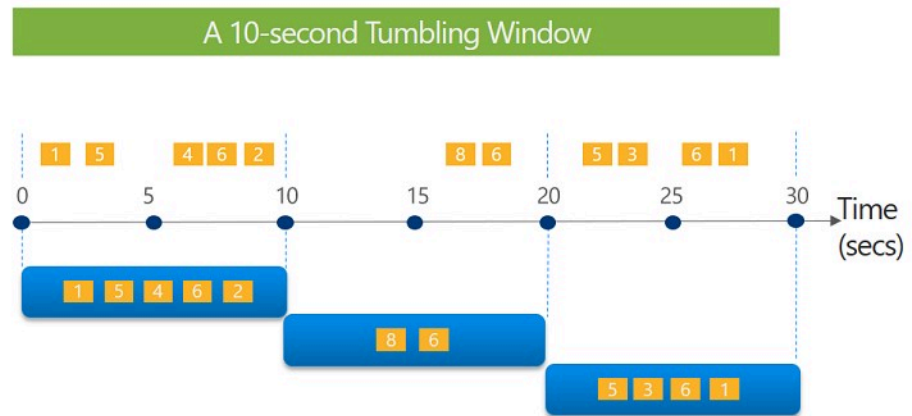
```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Box 2: Hopping -

Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

Every 5 seconds give me the count of tweets over the last 10 seconds

A 10-second Hopping Window with a 5-second "Hop"

```
SELECT Topic, COUNT(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10 , 5)
```

Box 3: Sliding -

Sliding window functions, unlike Tumbling or Hopping windows, produce an output only when an event occurs. Every window will have at least one event and the window continuously moves forward by an ג,¬ (epsilon). Like hopping windows, events can belong to more than one sliding window.

Give me the count of tweets for a single topic in the last 10 seconds.

A 10-second Sliding Window
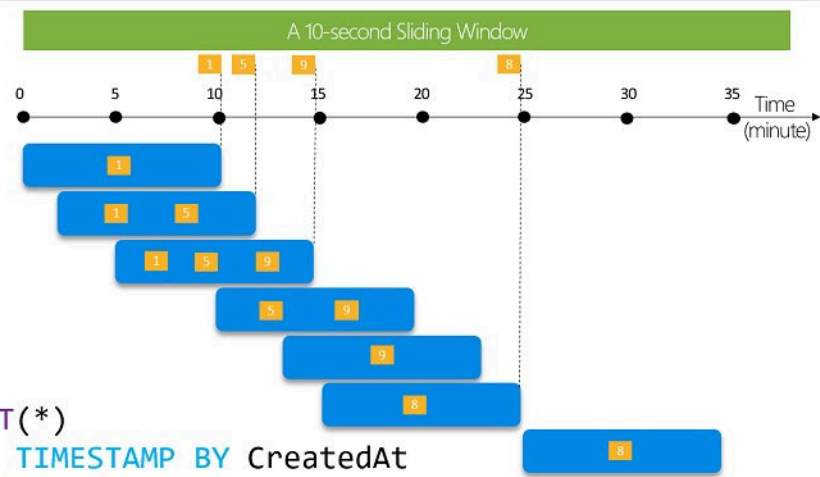
```
SELECT Topic, COUNT(*)
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, SlidingWindow(second, 10)
```

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

---

memo43 Highly Voted 👍 4 years, 1 month ago

answer is CORRECT

upvoted 6 times

DRAG DROP -

You have an Azure Data Lake Storage Gen2 account that contains JSON files for customers. The files contain two attributes named FirstName and LastName.

You need to copy the data from the JSON files to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

☞ A destination table in Azure Synapse

☞ An Azure Blob storage container

☞ A service principal

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

Select and Place:

Answer Area

Actions
Specify a temporary folder to stage the data.
Write the results to a table in Azure Synapse
Write the results to Data Lake Storage.
Drop the data frame.
Perform transformations on the data frame.
Mount the Data Lake Storage onto DBFS.
Perform transformations on the file.
Read the file into a data frame.

Answer Area (empty boxes)

Suggested Answer:

Answer Area

Left (unused) list:
- Write the results to Data Lake Storage.
- Mount the Data Lake Storage onto DBFS.
- Perform transformations on the file.

Right (answer) list:
1. Read the file into a data frame.
2. Perform transformations on the data frame.
3. Specify a temporary folder to stage the data.
4. Write the results to a table in Azure Synapse
5. Drop the data frame.

Step 1: Read the file into a data frame.

You can load the json files as a data frame in Azure Databricks.

Step 2: Perform transformations on the data frame.

Step 3:Specify a temporary folder to stage the data
Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse.
Step 4: Write the results to a table in Azure Synapse.
You upload the transformed data frame into Azure Synapse. You use the Azure Synapse connector for Azure Databricks to directly upload a dataframe as a table in a Azure Synapse.

Step 5: Drop the data frame -
Clean up resources. You can terminate the cluster. From the Azure Databricks workspace, select Clusters on the left. For the cluster to terminate, under Actions, point to the ellipsis (...) and select the Terminate icon.
Reference:
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse

cadio30 Highly Voted 👍 4 years, 1 month ago
It requires to mount the ALDS gen 2 thus the sequence is right "FHEAB".
upvoted 15 times

niwe 4 years, 1 month ago
Can you explain what is "FHEAB"?
upvoted 1 times

maciejt 4 years, 1 month ago
letter-numbers of steps to choose
upvoted 2 times

niwe 4 years ago
Thanks!
upvoted 1 times

hoangton Highly Voted 👍 4 years ago
Correct answer should be:
Step 1: Mount the Data Lake Storage onto DBFS
Step 2: Read the file into a data frame.
Step 3: Perform transformations on the data frame.
Step 4: Specify a temporary folder to stage the data
Step 5: Write the results to a table in Azure Synapse.
upvoted 10 times

Bhagya123456 Most Recent ⊘ 3 years, 10 months ago
The Answer is Perfect.
Mounting is not Required.
Drop Data Frame should be there.

The question never mentioned that you have to use Service Principle. Had it be 6 steps I would have added Mounting Steps. But Considering only 5 steps, the below 5 steps have more priority then Mounting (not an essential).
upvoted 1 times

satyamkishoresingh 3 years, 9 months ago
why drop dataframe ?cleanup the resource is about cluster not the DF .
upvoted 1 times

vrmei 4 years ago
Mount Data Lake Storage onto DBFS (Service Principal)
Read the file into data frame
Perform Transformation on the data frame
Specify the temp folder to stage data
write results to synapse table
https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse
upvoted 4 times

vrmei 4 years ago
small correction: I don't see the mount option in ADLS account configuration in the given URL.
I feel the given answer might correct. The last one should be Drop the data frame which will do cleanup ...

upvoted 1 times

unidigm 4 years, 1 month ago

Do we really need to stage the data? We could directly write the dataframe to Synapse.

https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/synapse-analytics

upvoted 1 times

Rob77 4 years, 1 month ago

Yes, we do. tempDir (that stages data) MUST be specified for Synapse write method.

upvoted 2 times

Aragorn_2021 4 years, 2 months ago

I would go for FHEAB. Mount the storage -> Read the file to a dataframe ->transform it further -> write the data to temporary folder in storage -> and load to DWH

upvoted 5 times

111222333 4 years, 1 month ago

Agree.

Service Principal (which is given in the task) is used for mounting.

Mount an Azure Data Lake Gen 2 to DBFS (Databricks File System) using a Service Principal:
https://kyleake.medium.com/mount-an-adls-gen-2-to-databricks-file-system-using-a-service-principal-and-oauth-2-0-ep-5-73172dd0ddeb

upvoted 2 times

tucho 4 years, 2 months ago

I agree with HEAB. But I don' know which is the missing one. I think there is no need to "drop the DF" or to "mount the DL storage"... :-(Does anybody know the right full answer?

upvoted 1 times

alf99 4 years, 2 months ago

wrong, should be F,H,E,A,B. The DataLake storage has to be mounted onto DBFS before red the file

upvoted 2 times

DongDuong 4 years, 2 months ago

Based on the provided link, I think the keyword here is "mounted". Datalake storage is not mounted onto DBFS, instead, it is called by Databricks via API. So the given answer is correct

upvoted 2 times

DongDuong 4 years, 2 months ago

After revising, I think FHEAB makes more sense

upvoted 2 times