## Practice Test 1

Completed on 17-June-2020

| Attempt | Marks Obtained | Your score | Time Taken | Result |
|---|---|---|---|---|
| 02 | 51 / 55 | 92.73% | 00 H 06 M 03 S | Congratulations! Passed |

Share your Result with your friends

### Domains wise Quiz Performance Report

| No | Domain | Total Question | Correct | Incorrect | Unattempted | Marked as Review |
|---|---|---|---|---|---|---|
| 1 | Monitor and optimize data solutions | 17 | 17 | 0 | 0 | 0 |
| 2 | Implement data storage solutions | 19 | 18 | 1 | 0 | 0 |
| 3 | Manage and develop data processing | 19 | 16 | 3 | 0 | 0 |
| Total | All Domain | 55 | 51 | 4 | 0 | 0 |

### Review the Answers

Sorting by All

**Question 1** **Correct**

Domain :Monitor and optimize data solutions

A company has an Azure SQL database. The database contains tables that has masked columns. The company wants to identify when a user tries to attempt to access the data from any one of the masked columns. Which of the following would you use for this requirement?

- A. **Azure Advanced Threat Protection**
- B. **Auditing** ✓
- C. **Transparent Data Encryption**
- D. **Azure Monitor Audit Logs**

---

**Explanation:**

Answer – B

You can use Azure SQL Database auditing to get insights on how the data in your database is being accessed.

The Microsoft documentation mentions the following

> Auditing for Azure SQL Database and SQL Data Warehouse tracks database events and writes them to an audit log in your Azure storage account, Log Analytics workspace or Event Hubs. Auditing also:
>
> - Helps you maintain regulatory compliance, understand database activity, and gain insight into discrepancies and anomalies that could indicate business concerns or suspected security violations.
> - Enables and facilitates adherence to compliance standards, although it doesn't guarantee compliance. For more information about Azure programs that support standards compliance, see the Azure Trust Center where you can find the most current list of SQL Database compliance certifications.

Option A is incorrect since this is used for enabling features such detecting vulnerabilities on your Azure SQL Database system.

Option C is incorrect since this is used to encrypt data at rest in SQL databases

Option D is incorrect since this is used to look at the control plane activities carried out on Azure resources

For more information on Azure SQL Database auditing, please visit the following URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 2

Correct

Domain :Monitor and optimize data solutions

A company has an Azure SQL Data Warehouse defined as part of their Azure subscription. The company wants to ensure their support department gets an alert when the Data Warehouse consumes the maximum allotted resources to it?
Which of the following would they use as the resource type when configuring the alert in Azure Monitor?

A. **Resource Group**

B. **SQL Server**

C. **SQL Data Warehouse**

D. **Subscription**

**Explanation:**

Answer – C

Here since we need to create an alert based on the consumption of the Data Warehouse itself, we should create an alert on the Warehouse itself.

The Microsoft documentation mentions the following

# Monitoring resource utilization and query activity in Azure SQL Data Warehouse

08/09/2019 • 2 minutes to read • 

Azure SQL Data Warehouse provides a rich monitoring experience within the Azure portal to surface insights to your data warehouse workload. The Azure portal is the recommended tool when monitoring your data warehouse as it provides configurable retention periods, alerts, recommendations, and customizable charts and dashboards for metrics and logs. The portal also enables you to integrate with other Azure monitoring services such as Operations Management Suite (OMS) and Azure Monitor (logs) to provide a holistic monitoring experience for not only your data warehouse but also your entire Azure analytics platform for an integrated monitoring experience. This documentation describes what monitoring capabilities are available to optimize and manage your analytics platform with SQL Data Warehouse.

The other options are incorrect since we need to ensure the monitoring is enabled on the Data Warehouse itself.

For more information on querying Azure SQL Data Warehouse, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-concept-resource-utilization-query-activity

Question 3 Correct

Domain :Monitor and optimize data solutions

A company has an Azure SQL Data Warehouse defined as part of its Azure subscription. The company wants to ensure its support department gets an alert when the Data Warehouse consumes the maximum allotted resources to it.
Which of the following would you use as the signal type for the alert?

A. **CPU used**

B. **DWU limit**

C. **DWU used** ✓

D. **Data IO Percentage**

**Explanation:**

Answer – C

You would monitor the number of DWU (Data Warehousing Units) being consumed.

## What are Data Warehouse Units

Azure SQL Data Warehouse CPU, memory, and IO are bundled into units of compute scale called Data Warehouse Units (DWUs). A DWU represents an abstract, normalized measure of compute resources and performance. A change to your service level alters the number of DWUs that are available to the system, which in turn adjusts the performance, and the cost, of your system.

For higher performance, you can increase the number of data warehouse units. For less performance, reduce data warehouse units. Storage and compute costs are billed separately, so changing data warehouse units does not affect storage costs.

Since this is a measurement for performance for Azure SQL Data Warehouses, all other options are incorrect.

For more information on Azure SQL Data Warehouse DWU's, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/what-is-a-data-warehouse-unit-dwu-cdwu

Ask our Experts

Rate this Question?

View Queries

open

Question 4

Correct

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database defined as part of their Azure subscription. The Automatic tuning settings are configured as shown below

| Option | Server level | Database level |
|---|---|---|
| **Force Plan** | Inherited | Inherited |
| **Create Index** | Inherited | Inherited |
| **Drop Index** | Inherited | Inherited |

Would the setting of "Force Plan" be ON for the database?

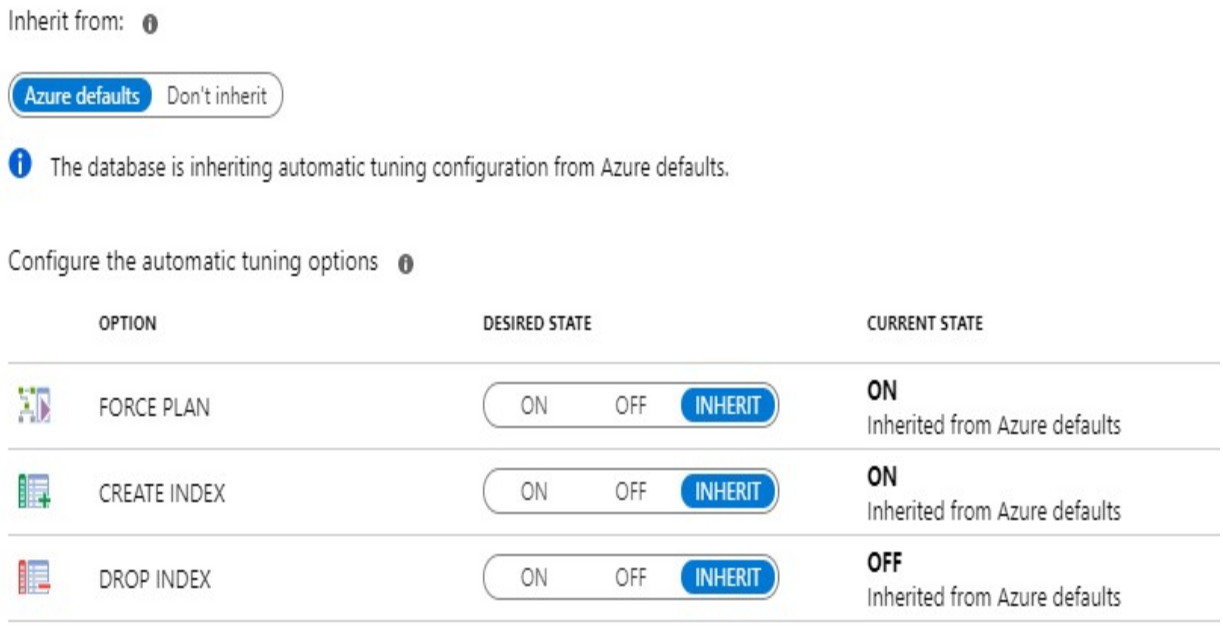
A. **Yes**

B. **No**

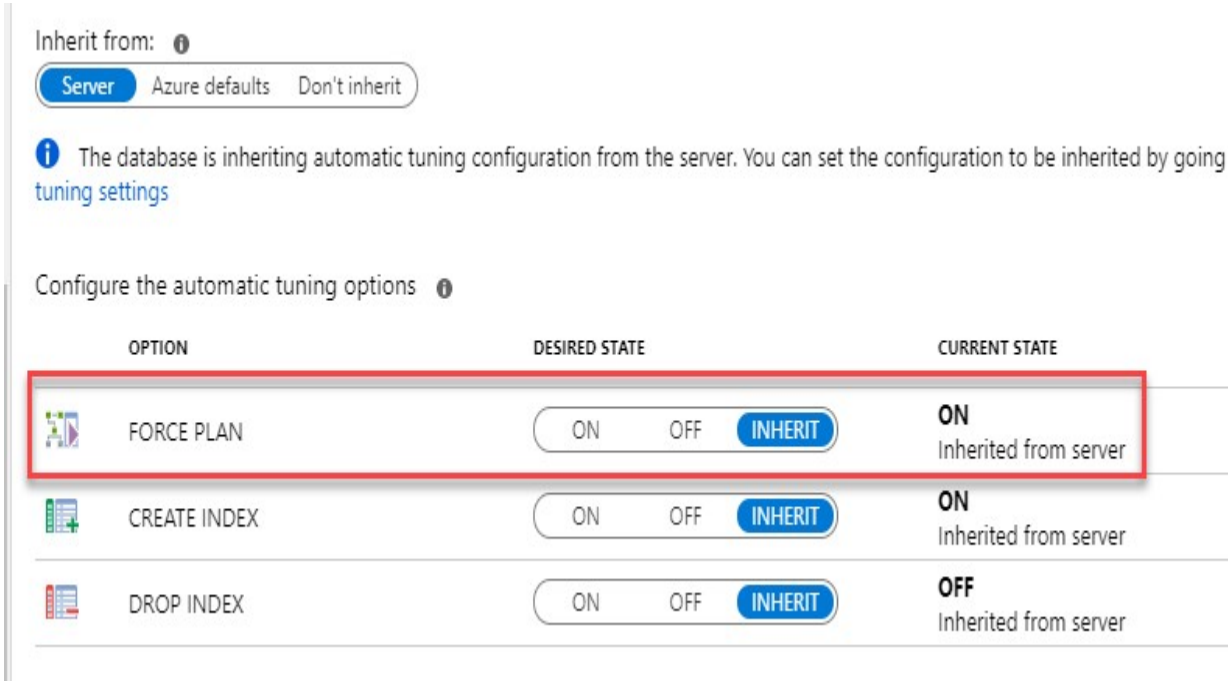
**Explanation:**

Answer – A

If you implement the below auditing settings for the server




And also place the following settings for the database




You can see that the “FORCE PLAN” setting is ON.

For more information on Automatic Tuning, please visit the following URL

> https://docs.microsoft.com/en-us/azure/sql-database/sql-database-automatic-tuning-enable

Ask our Experts

Rate this Question?

View Queries

open

Question 5

Correct

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database defined as part of their Azure subscription. The Automatic tuning settings are configured as shown below

| Option | Server level | Database level |
|---|---|---|
| **Force Plan** | Inherited | Inherited |
| **Create Index** | Inherited | Inherited |
| **Drop Index** | Inherited | Inherited |

Would the setting of "Create Index" be ON for the database?

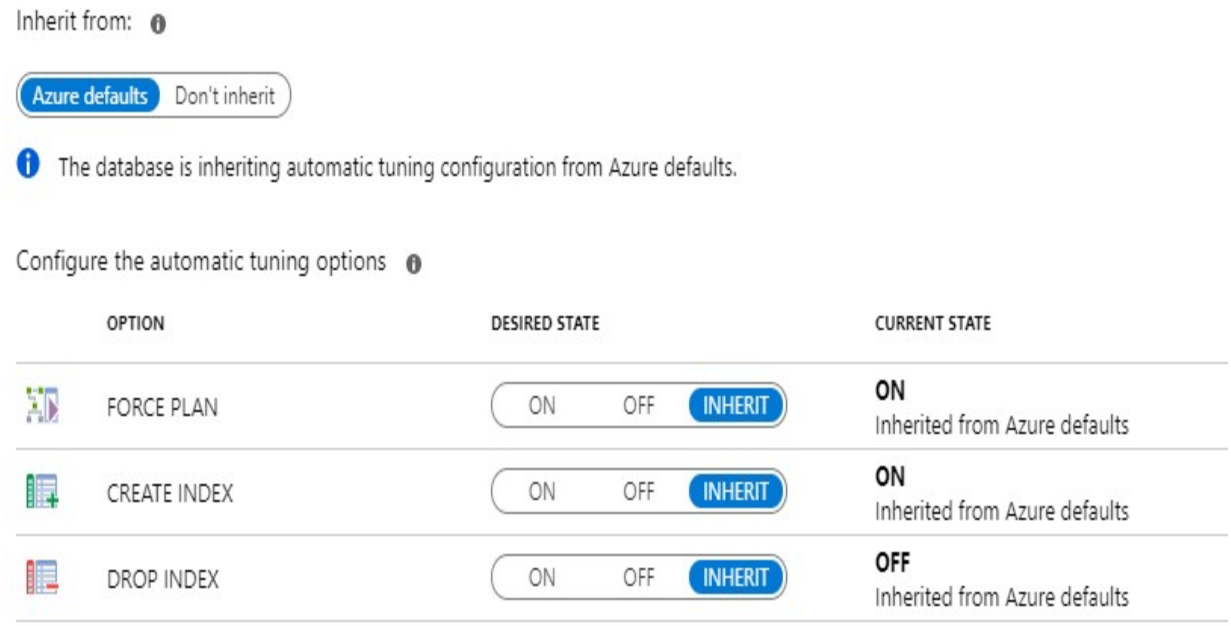
A. **Yes**

B. **No**

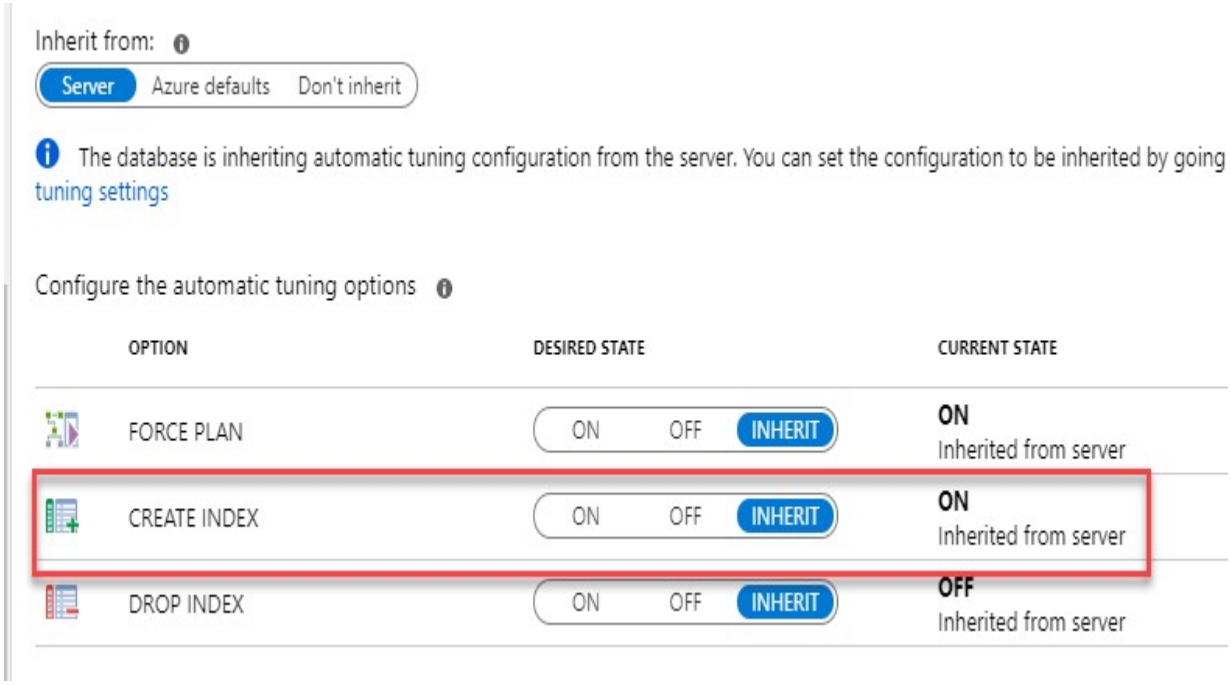
**Explanation:**

Answer – A

If you implement the below auditing settings for the server




And also place the following settings for the database




You can see that the “CREATE INDEX” setting is ON.

For more information on Automatic Tuning, please visit the following URL

> https://docs.microsoft.com/en-us/azure/sql-database/sql-database-automatic-tuning-enable

Ask our Experts

Rate this Question?

View Queries

open

Question 6

Correct

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database defined as part of their Azure subscription. The Automatic tuning settings are configured as shown below

| Option | Server level | Database level |
|---|---|---|
| **Force Plan** | Inherited | Inherited |
| **Create Index** | Inherited | Inherited |
| **Drop Index** | Inherited | Inherited |

Would the setting of "Drop Index" be ON for the database?

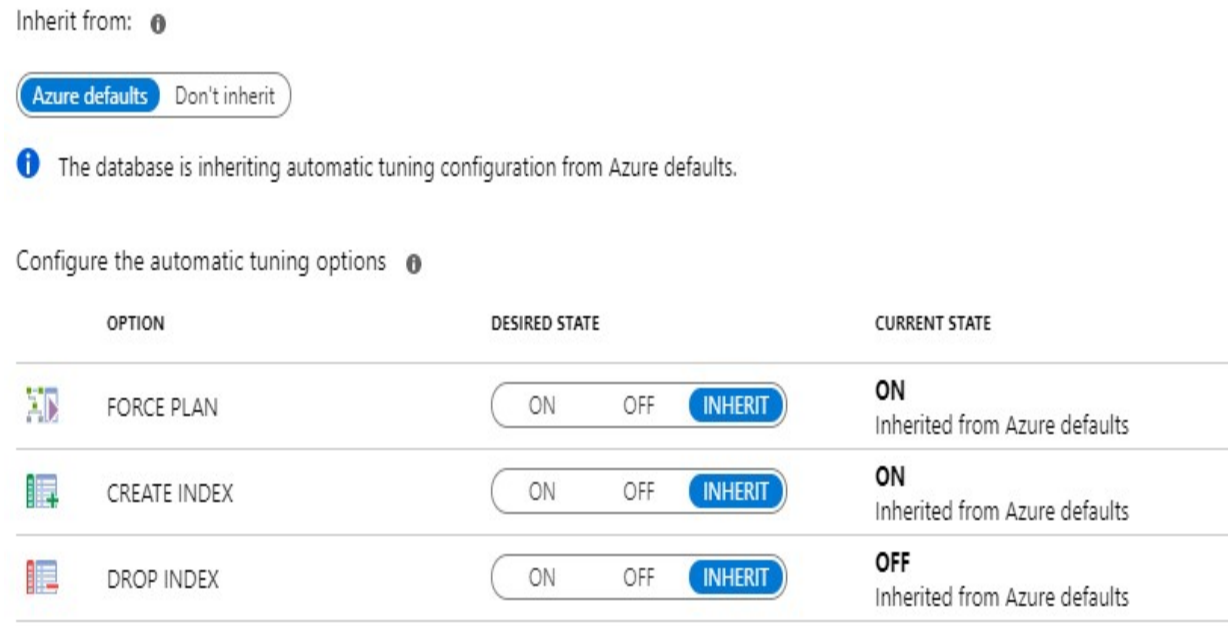
A. **Yes**

B. **No**

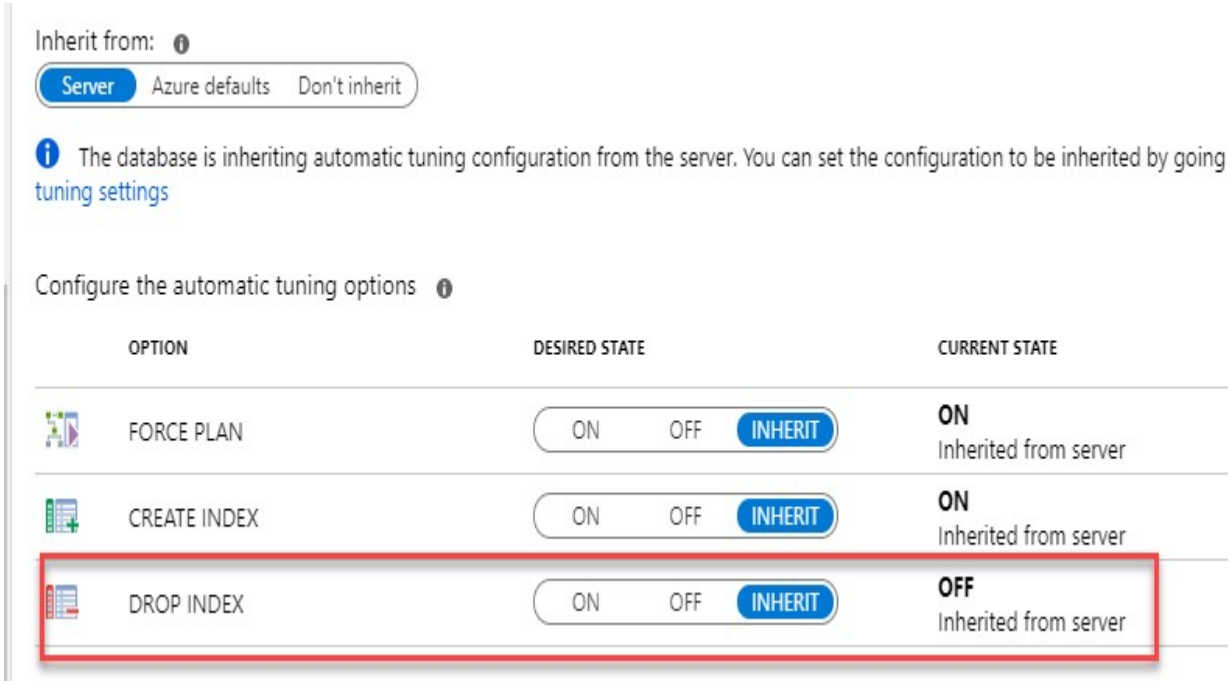
**Explanation:**

Answer – B

If you implement the below auditing settings for the server

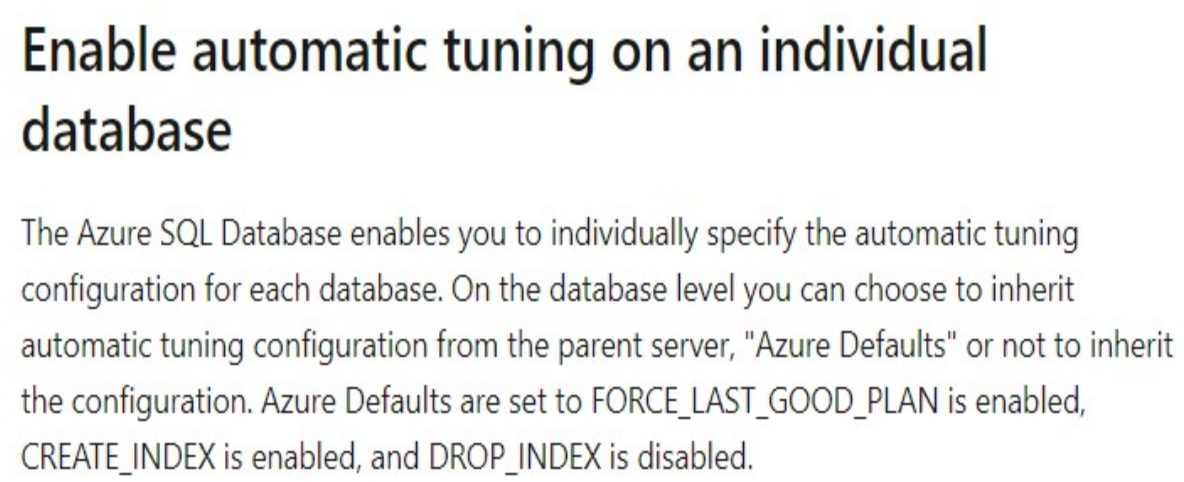



And also place the following settings for the database




You can see that the “DROP INDEX” setting is OFF.

The Microsoft documentation mentions the following




For more information on Automatic Tuning, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-automatic-tuning-enable

Ask our Experts

Rate this Question?

View Queries

open

**Question 7**

**Correct**

**Domain :Implement data storage solutions**

A company has an application that is storing its data in an Azure Cosmos DB Account. The database currently has around 100 GB worth of data. Each entry in a collection in the database is shown below

{

OrderId: number,
OrderDescriptionId: number,
ProductName: string
OrderValue: number

}

The partition key for the collection is set as OrderId.

Users reports that queries take a long time to execute when they try to retrieve data using the Product Name attribute.

You have to resolve the issue.
You decide to create a lookup collection that uses ProductName as the partition key
Would this resolve the issue?

A. Yes

B. No

**Explanation:**

Answer – B

The right option is to have a lookup collection and ideally use the ProductName as the partition key, but you also need to use OrderId as the value.

For more information on Cosmos DB partitioning, please visit the following URL

https://azure.microsoft.com/en-us/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

Ask our Experts

Rate this Question?

View Queries open

**Question 8** Correct

**Domain :Implement data storage solutions**

A company has an application that is storing its data in an Azure Cosmos DB Account. The database currently has around 100 GB worth of data. Each entry in a collection in the database is shown below

```
{
        OrderId: number,
        OrderDescriptionId: number,
        ProductName: string
        OrderValue: number
}
```

The partition key for the collection is set as OrderId.

Users reports that queries take a long time to execute when they try to retrieve data using the Product Name attribute.
You have to resolve the issue.
You decide to create a lookup collection that uses ProductName as the partition key and OrderId as a value
Would this resolve the issue?

A. Yes

B. No

**Explanation:**

Answer – A

The right option is to have a lookup collection and ideally use the ProductName as the partition key and use OrderId as the value.

For more information on Cosmos DB partitioning, please visit the following URL

https://azure.microsoft.com/en-us/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

Ask our Experts

Rate this Question?

View Queries

open

**Question 9**

**Correct**

**Domain :Implement data storage solutions**

A company has an application that is storing its data in an Azure Cosmos DB Account. The database currently has around 100 GB worth of data. Each entry in a collection in the database is shown below

```
{
        OrderId: number,
        OrderDescriptionId: number,
        ProductName: string
```

OrderValue: number
}
The partition key for the collection is set as OrderId.
Users reports that queries take a long time to execute when they try to retrieve data using the Product Name attribute.
You have to resolve the issue.
You decide to change the partition key to include the ProductName
Would this resolve the issue?

A. Yes

B. No

**Explanation:**

Answer – B

The Partition Key can only be based on a single attribute. The right option is to have a lookup collection and ideally use the ProductName as the partition key and use OrderId as the value.

For more information on Cosmos DB partitioning, please visit the following URL

https://azure.microsoft.com/en-us/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

Ask our Experts

Rate this Question?

View Queries

open

**Question 10**

**Correct**

**Domain :Manage and develop data processing**

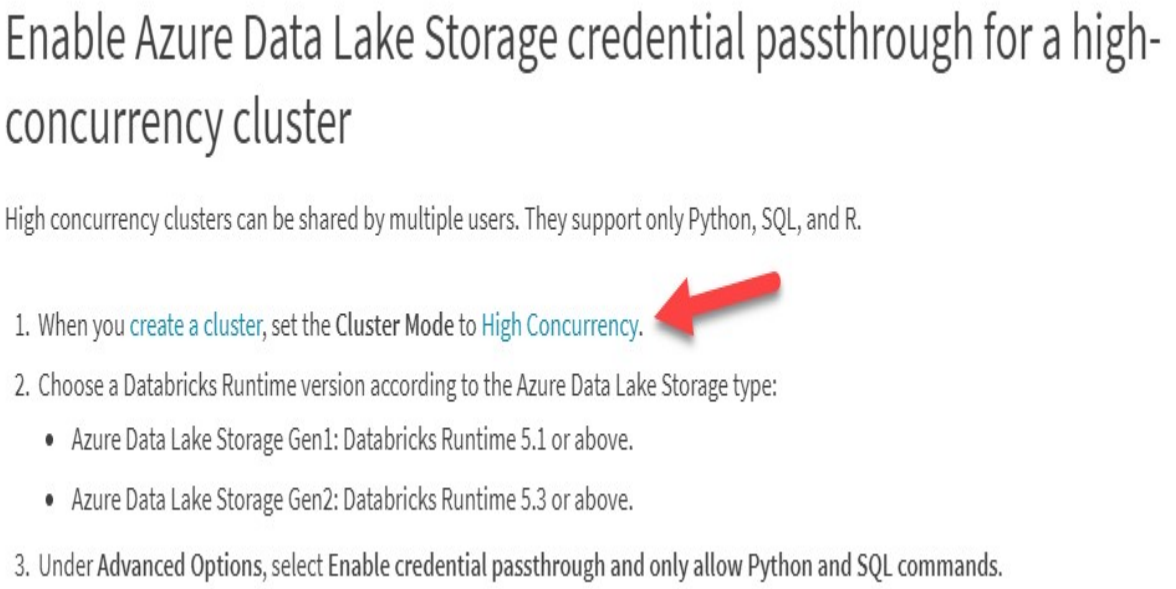
You need to create a new Azure Databricks cluster. This cluster would connect to Azure Data Lake Storage Gen2 by using Azure Active Directory (Azure AD) integration.
Which of the following would you use as the Cluster Mode?

- A. **High Concurrency** ✅
- B. **Low Concurrency**
- C. **Premium**
- D. **Standard**

**Explanation:**

Answer – A

The documentation for Azure Data bricks mentions that you should ideally set the Cluster Mode to High Concurrency




Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on using Azure AD for Azure Data bricks, please visit the following URL

https://docs.azuredatabricks.net/data/data-sources/azure/adls-passthrough.html

Ask our Experts

Rate this Question?

View Queries

open

Question 11

Correct

Domain :Manage and develop data processing

You need to create a new Azure Databricks cluster. This cluster would connect to Azure Data Lake Storage Gen2 by using Azure Active Directory (Azure AD) integration.
Which of the following advanced option would you enable?

A. **Blob access control**

B. **Table access control**

C. **Credential Passthrough**

D. **Single Sign-On**

**Explanation:**

Answer – C

The documentation for Azure Data bricks mentions that you should set Credential Passthrough in the Advanced option.

# Enable Azure Data Lake Storage credential passthrough for a high-concurrency cluster

High concurrency clusters can be shared by multiple users. They support only Python, SQL, and R.

1. When you create a cluster, set the **Cluster Mode** to High Concurrency.
2. Choose a Databricks Runtime version according to the Azure Data Lake Storage type:
   - Azure Data Lake Storage Gen1: Databricks Runtime 5.1 or above.
   - Azure Data Lake Storage Gen2: Databricks Runtime 5.3 or above.
3. Under **Advanced Options**, select **Enable credential passthrough and only allow Python and SQL commands.**

Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on using Azure AD for Azure Data bricks, please visit the following URL

https://docs.azuredatabricks.net/data/data-sources/azure/adls-passthrough.html

Ask our Experts

Rate this Question?

View Queries

open

**Question 12**

**Correct**

**Domain :Manage and develop data processing**

You currently have an Azure Storage Account and an Azure SQL Database defined as part of your Azure subscription. You need to move data from Azure Storage Account to the SQL database using Azure Data Factory. You have to ensure the following requirements are met

Ensure that the data remains in the same region as the Azure Storage Account and the Azure SQL Database at all times

Minimize administrative effort

Which of the following would you use as the Integration runtime type?

- A. **Azure** ✓
- B. **Self-Hosted**
- C. **Primary**
- D. **Azure-SSIS**

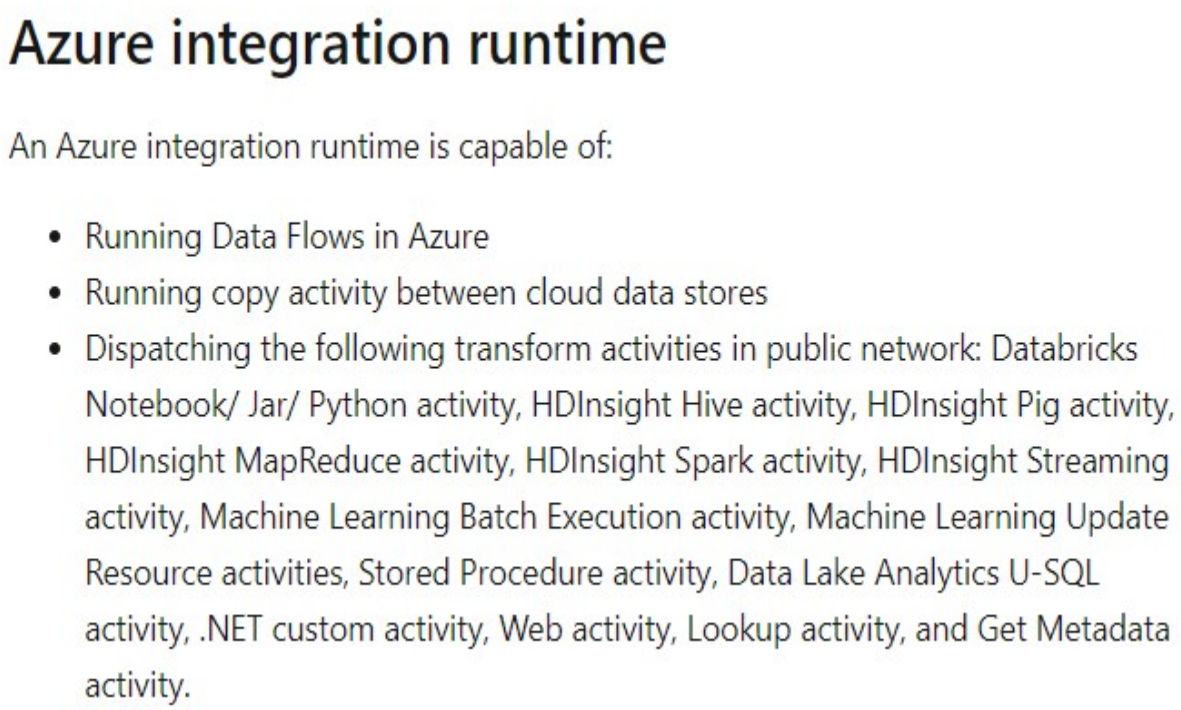
---

**Explanation:**

Answer – A

You can use the Azure Integration runtime type to copy data between Azure based resources.

The Microsoft documentation mentions the following



## Azure integration runtime

An Azure integration runtime is capable of:

- Running Data Flows in Azure
- Running copy activity between cloud data stores
- Dispatching the following transform activities in public network: Databricks Notebook/ Jar/ Python activity, HDInsight Hive activity, HDInsight Pig activity, HDInsight MapReduce activity, HDInsight Spark activity, HDInsight Streaming activity, Machine Learning Batch Execution activity, Machine Learning Update Resource activities, Stored Procedure activity, Data Lake Analytics U-SQL activity, .NET custom activity, Web activity, Lookup activity, and Get Metadata activity.

Option B is invalid since this is used to copy data between cloud data stores and data stores in private networks

Option C is invalid since this is not a valid integration runtime type

Option D is invalid this is used primarily for SSIS workloads.

For more information on the Integration runtime, please visit the following URL

https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

Ask our Experts

Rate this Question?

View Queries

open

**Question 13**

**Correct**

**Domain :Manage and develop data processing**

You have to implement Azure Stream Analytics Functions as part of your data streaming solution. The solution has the following requirements

Segment the data stream into distinct time segments that do not repeat or overlap

Segment the data stream into distinct time segments that repeat and can overlap

Segment the data stream to produce an output when an event occurs

Which of the following windowing function would you use for the following requirement?
***"Segment the data stream into distinct time segments that do not repeat or overlap"***

- A. **Hopping**
- B. **Session**
- C. **Sliding**
- D. **Tumbling** ✓

**Explanation:**

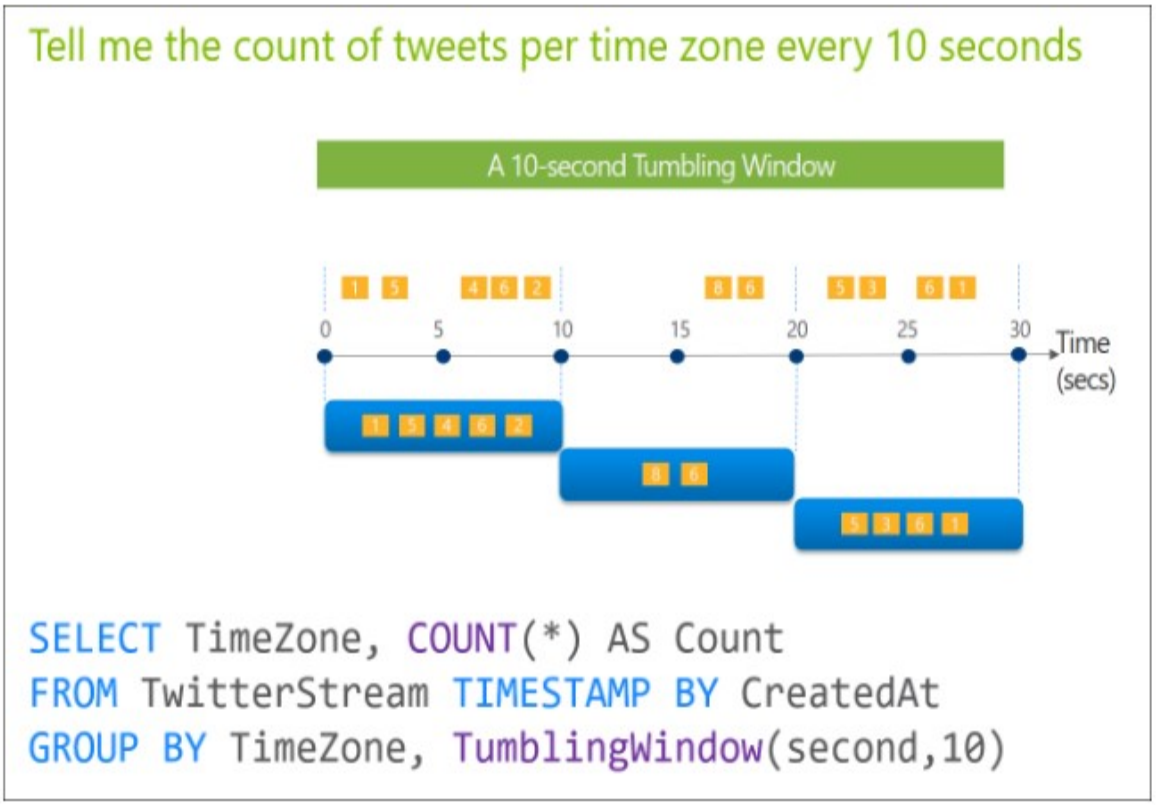
Answer – D

You need to use the Tumbling windowing function for this requirement

The Microsoft documentation mentions the following

# 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.




Since this is clearly given in the documentation, all other options are incorrect

For more information on the windowing functions, please visit the following URL

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

Ask our Experts

Rate this Question?

Question 14

Correct

Domain :Manage and develop data processing

You have to implement Azure Stream Analytics Functions as part of your data streaming solution. The solution has the following requirements

Segment the data stream into distinct time segments that do not repeat or overlap

Segment the data stream into distinct time segments that repeat and can overlap

Segment the data stream to produce an output when an event occurs

Which of the following windowing function would you use for the following requirement?
***"Segment the data stream into distinct time segments that repeat and can overlap"***

- A. **Hopping** ✓
- B. **Session**
- C. **Sliding**
- D. **Tumbling**

**Explanation:**

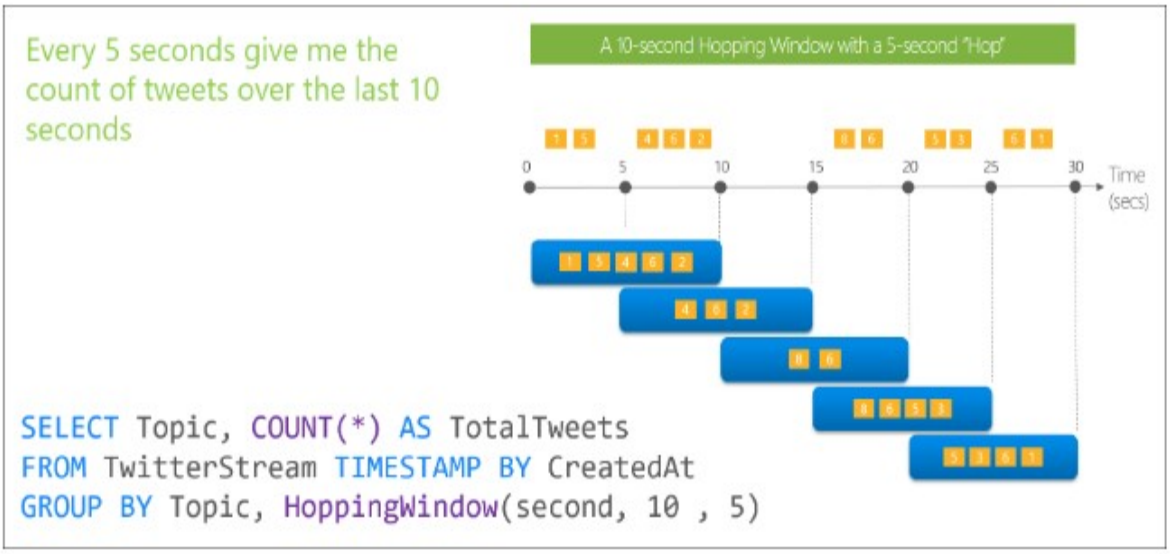
Answer – A

You need to use the Hopping windowing function for this requirement

The Microsoft documentation mentions the following

# Hopping window

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.




Since this is clearly given in the documentation, all other options are incorrect

For more information on the windowing functions, please visit the following URL

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

Ask our Experts

Rate this Question?

View Queries

open

Question 15

Correct

Domain :Manage and develop data processing

You have to implement Azure Stream Analytics Functions as part of your data streaming solution. The solution has the following requirements

Segment the data stream into distinct time segments that do not repeat or overlap

Segment the data stream into distinct time segments that repeat and can overlap

Segment the data stream to produce an output when an event occurs

Which of the following windowing function would you use for the following requirement?
***"Segment the data stream to produce an output when an event occurs"***

- A. **Hopping**
- B. **Session**
- C. **Sliding** ✓
- D. **Tumbling**

---

**Explanation:**

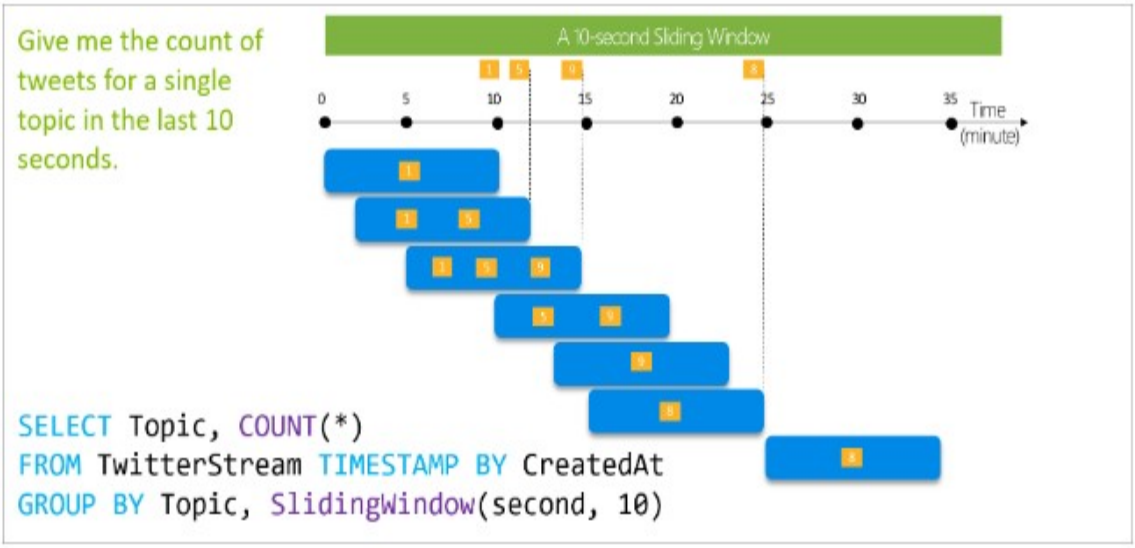
Answer – C

You need to use the Sliding windowing function for this requirement

The Microsoft documentation mentions the following

# Sliding window

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 € (epsilon). Like hopping windows, events can belong to more than one sliding window.




Since this is clearly given in the documentation, all other options are incorrect

For more information on the windowing functions, please visit the following URL

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

Ask our Experts

Rate this Question?

View Queries

open

Question 16

Correct

Domain :Manage and develop data processing

You have JSON files stored in an Azure Data Lake Storage Gen2 account. The JSON file contains the FirstName and LastName of customers. You need to use Azure Data bricks to copy the data in the JSON files to an Azure data warehouse. A new column must be created which concatenates the FirstName and LastName values. You have the following components in place in Azure

A destination table in the SQL Data Warehouse

An Azure Blob storage container

A service principal

Which of the following are actions you would perform to transfer the data onto the Azure SQL Data warehouse table? Choose 5 answers from the options given below

- [ ] A. **Write the results onto Azure Data Lake Storage**
- [ ] B. **Drop the data frame**
- [x] C. **Perform transformations on the data frame**
- [x] D. **Mount the Data Lake Storage onto DBFS**
- [ ] E. **Perform transformations on the file**
- [x] F. **Read the file into a data frame**
- [x] G. **Specify a temporary folder to stage the data**
- [x] H. **Write the results to a table in SQL Data Warehouse**

**Explanation:**

Answer - C, D, F,G and H

To Connect to Azure Data Lake Gen 2 storage, we first need to mount the file system. This is also given in the Azure data bricks documentation

# Azure Data Lake Storage Gen2

Azure Data Lake Storage Gen2 (also known as ADLS Gen2) is a next-generation data lake solution for big data analytics. Azure Data Lake Storage Gen2 builds Azure Data Lake Storage Gen1 capabilities—file system semantics, file-level security, and scale—into Azure Blob storage, with its low-cost tiered storage, high availability, and disaster recovery features.

> **Note**
>
> The Azure Data Lake Storage Gen2 connector is supported in Databricks Runtime 5.2 and above with full support for Delta Lake in Databricks Runtime 5.5 and above.

There are three ways of accessing Azure Data Lake Storage Gen2:

1. Mount an Azure Data Lake Storage Gen2 filesystem to DBFS using a service principal and OAuth 2.0.
2. Use a service principal directly.
3. Use the Azure Data Lake Storage Gen2 storage account access key directly.

Next, we need to load the data from the Azure Data Lake Storage Gen2 account. An example is also given in the Microsoft documentation

## Extract data from the Azure Data Lake Storage Gen2 account

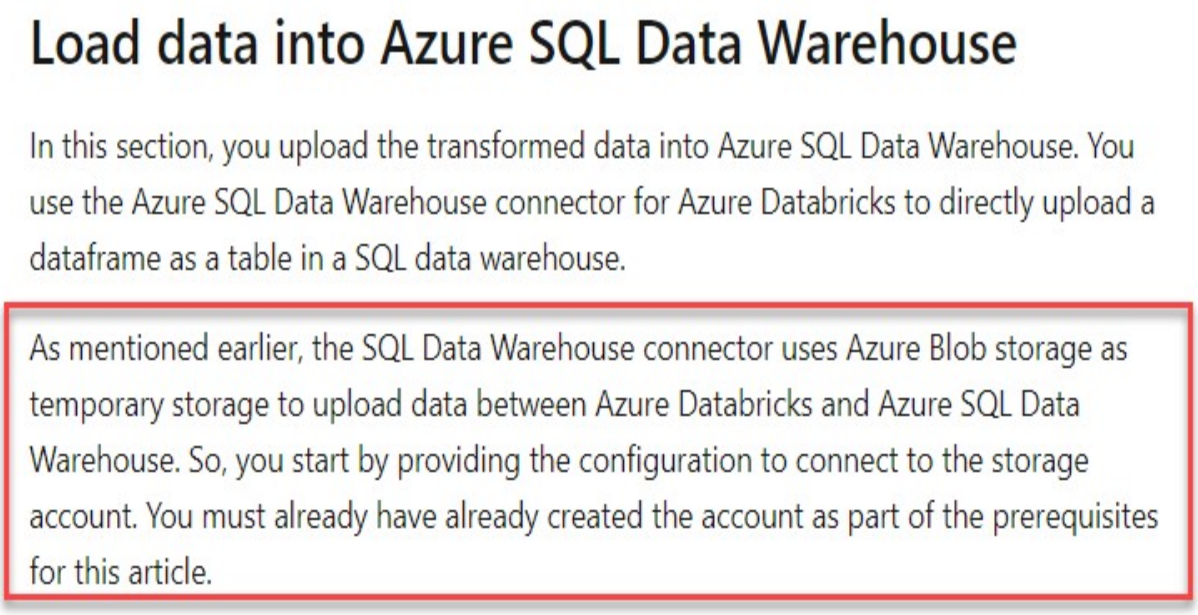
1. You can now load the sample json file as a data frame in Azure Databricks. Paste the following code in a new cell. Replace the placeholders shown in brackets with your values.

Scala | Copy

```
val df = spark.read.json("abfss://<file-system-name>@<storage-account-name
```

Next, we need to transform the data in the data frame to combine the FirstName and LastName values

You need to mention an Azure storage account as a temporary staging area.




And then finally you copy the data onto the Azure SQL data warehouse.

Option A is incorrect since you don't need to write the results onto Azure Data Lake storage. The results are written to the Azure SQL Data warehouse table.

Option B is incorrect since we don't need to drop the data frames

Option B is incorrect since transformations need to be carried out on the data frames

For more information on loading data from an Azure SQL Data warehouse, please visit the following URL

https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse#load-data-into-azure-sql-data-warehouse

**Question 17** **Correct**

**Domain :Manage and develop data processing**

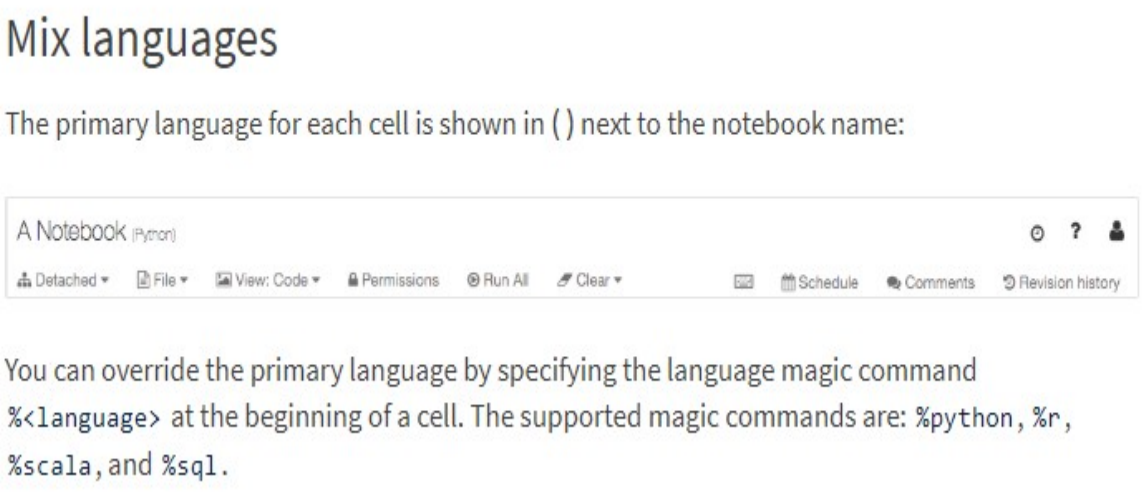
You have created an instance of Azure Data Bricks. You have gone ahead and created a cluster and a notebook. The notebook will use R as the primary language. But you also need to be able to switch the notebook to support Scala and SQL. Which of the following can be used to switch between languages in the notebook?

- A. **%<language>** ✓
- B. **#<language>**
- C. **@<language>**
- D. **&<language>**

**Explanation:**

Answer – A

This can be done with the use of the %<language> command. This is also mentioned in the databricks documentation.




Since this is clearly given in the documentation, all other options are incorrect

For more information on developing notebooks, please visit the following URL

https://docs.databricks.com/notebooks/notebooks-use.html#develop-notebooks

Ask our Experts

Rate this Question?

View Queries open

Question 18

Correct

Domain :Manage and develop data processing

You have an Azure Data Lake Storage Gen 2 account. You have a number of CSV files loaded in the account. Each file has a header row. After the header row is a property that is formatted by carriage return (/r) and line feed (/n).
You need to load the files daily as a batch into Azure SQL Data warehouse using Polybase. You have to skip the header row when the files are imported. Which of the following actions would you take to implement this requirement? Choose 3 answers from the options given below

- [x] A. **Create an external data source and ensure to use the abfs location**
- [ ] B. **Create an external data source and ensure to use the Hadoop location**
- [x] C. **Create an external file format and set the First_row option**
- [x] D. **Create a database scoped credential that uses OAuth2 token and a key**
- [ ] E. **Use the CREATE EXTERNAL TABLE AS SELECT and create a view that removes the empty row**

**Explanation:**

Answer – A,C and D

The Microsoft documentation highlights the steps required to load data from Azure Data Lake Gen 2 to an Azure SQL Data warehouse.

One of the steps is to create a database scoped credential

```
-- B (for Gen2 storage key authentication): Create a database scoped credential
-- IDENTITY: Provide any string, it is not used for authentication to Azure sto
-- SECRET: Provide your Azure storage account key.

CREATE DATABASE SCOPED CREDENTIAL ADLSCredential
WITH
    IDENTITY = 'user',
    SECRET = '<azure_storage_account_key>'
;

-- It should look something like this when authenticating using service princip
CREATE DATABASE SCOPED CREDENTIAL ADLSCredential
WITH
    IDENTITY = '536540b4-4239-45fe-b9a3-629f97591c0c@https://login.microsoftonl
    SECRET = 'BjdIlmtKp4Fpyh9hIvr8HJlUida/seM5kQ3EpLAmeDI='
;
```

Another step is to create the external data source using 'abfs' as the file location

## Create the external data source

Use this CREATE EXTERNAL DATA SOURCE command to store the location of the data.

SQL Copy

```
-- C (for Gen1): Create an external data source
-- TYPE: HADOOP - PolyBase uses Hadoop APIs to access data in Azure Data Lake S
-- LOCATION: Provide Data Lake Storage Gen1 account name and URI
-- CREDENTIAL: Provide the credential created in the previous step.

CREATE EXTERNAL DATA SOURCE AzureDataLakeStorage
WITH (
    TYPE = HADOOP,
    LOCATION = 'adl://<datalakestoregen1accountname>.azuredatalakestore.net',
    CREDENTIAL = ADLSCredential
);

-- C (for Gen2): Create an external data source
-- TYPE: HADOOP - PolyBase uses Hadoop APIs to access data in Azure Data Lake S
-- LOCATION: Provide Data Lake Storage Gen2 account name and URI
-- CREDENTIAL: Provide the credential created in the previous step.

CREATE EXTERNAL DATA SOURCE AzureDataLakeStorage
WITH (
    TYPE = HADOOP,
    LOCATION='abfs[s]://<container>@<AzureDataLake account_name>.dfs.core.windo
    CREDENTIAL = ADLSCredential
);
```

And you can use the FIRST_ROW parameter to skip the first row of the file.

FIRST_ROW = *First_row_int*

Specifies the row number that is read first in all files during a PolyBase load. This parameter can take values 1-15. If the value is set to two, the first row in every file (header row) is skipped when the data is loaded. Rows are skipped based on the existence of row terminators (/r/n, /r, /n). When this option is used for export, rows are added to the data to make sure the file can be read with no data loss. If the value is set to >2, the first row exported is the Column names of the external table.

For more information on an example of loading data from Azure Data Lake store and for the format of the external file format, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-file-format-transact-sql?view=sql-server-ver15

Ask our Experts

Rate this Question?

View Queries

open

**Question 19**

**Correct**

**Domain :Implement data storage solutions**

A company is planning on creating an Azure Cosmos DB account. This account will contain a database and a collection. Around 10,000 JSON records will be written to the collection every 24 hours. The company wants to set a consistency level for the database that would meet the following requirements

Enable monotonic reads and writes within a session

Provide fast throughput

Provide the lowest latency

Which of the following should be set as the consistency level for the database?

A. **Strong**

B. **Bounded Staleness**

C. **Eventual**

D. **Session**

E. **Consistent Prefix**

**Explanation:**

Answer – D

If you look at the Microsoft documentation, the Session consistency level matches the requirement

- **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.

Options A and B are incorrect since these would lead to more latency and less throughput

Option C and E are incorrect since these would not provide session-based consistency

For more information on Cosmos DB consistency levels, please visit the following URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 20

Correct

**Domain :Implement data storage solutions**

A company has an Azure SQL Datawarehouse. They have a table named whizlab_salesfact that contains data for the past 12 months. The data is partitioned by month. The table contains around a billion rows. The table has clustered columnstore indexes. At the beginning of each month you need to remove the data from the table that is older than 12 months. Which of the following actions would you implement for this requirement? Choose 3 answers from the options given below

- [ ] A. **Create a new empty table named whizlab_salesfact_new that has the same schema as whizlab_salesfact**
- [x] B. **Drop the whizlab_salesfact_new table**
- [x] C. **Copy the data to the new table by using CREATE TABLE AS SELECT (CTAS)**
- [ ] D. **Truncate the partition containing the stale data**
- [x] E. **Switch the partition containing the stale data from whizlab_salesfact to whizlab_salesfact_new**
- [ ] F. **Execute the DELETE statement where the value in the Date column is greater than 12 months**

---

**Explanation:**

Answer – B,C and E

An example of this is given in a blog post. To achieve this, we first need to copy the data onto a new table using the "CREATE TABLE AS SELECT" command. Then we switch the partition and then delete the staging table.

Option A is incorrect because we also need to copy the data onto the new table

Option D is incorrect because we need to switch the partition

Option F is incorrect because issuing the DELETE statement would take time.

For more information on the blog article, please visit the following URL

https://blogs.msdn.microsoft.com/apsblog/2018/06/18/azure-sql-dw-performance-ctaspartition-switching-vs-updatedelete/

Ask our Experts

Rate this Question?

View Queries

open

**Question 21**

**Correct**

**Domain :Implement data storage solutions**

You have an Azure SQL data warehouse. You have use Polybase to create a table named [Ext].[whizlabitems] to query Parquet files stored in Azure Data lake Storage Gen 2. The external table has been defined with 3 columns. You have now discovered that the Parquet files contains a fourth column named ItemID. Which of the following command can you use to add the fourth column to the external table?

A. ALTER EXTERNAL TABLE [Ext].[whizlabitems] ADD [ItemID] int;

B. ALTER TABLE [Ext].[whizlabitems] ADD [ItemID] int;

C.
```
DROP TABLE [Ext].[whizlabitems]
CREATE EXTERNAL FILE FORAMT parquetfilenew
WITH (
FORMAT_TYPE=PARQUET,
DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec' );
```

D.
```
DROP TABLE [Ext].[whizlabitems]
CREATE EXTERNAL TABLE [Ext].[whizlabitems]
( [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
);
```

**Explanation:**

Answer – D

You can't alter the external table to add a new column. You would need to drop the external table and recreate the table again.

For more information on creating the external table, please visit the following URL

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-file-format-transact-sql?view=sql-server-ver15

Ask our Experts

Rate this Question?

View Queries

open

**Question 22**

**Correct**

**Domain :Implement data storage solutions**

You are planning on creating a dimension table in an Azure SQL Data Warehouse. The data in the table will be less than 1 GB. You need to ensure the table meets the following requirements

Minimize data movement

Provide the fastest query time

Which of the following would you choose as the table type?

A. **Hash distributed**

B. **Heap**

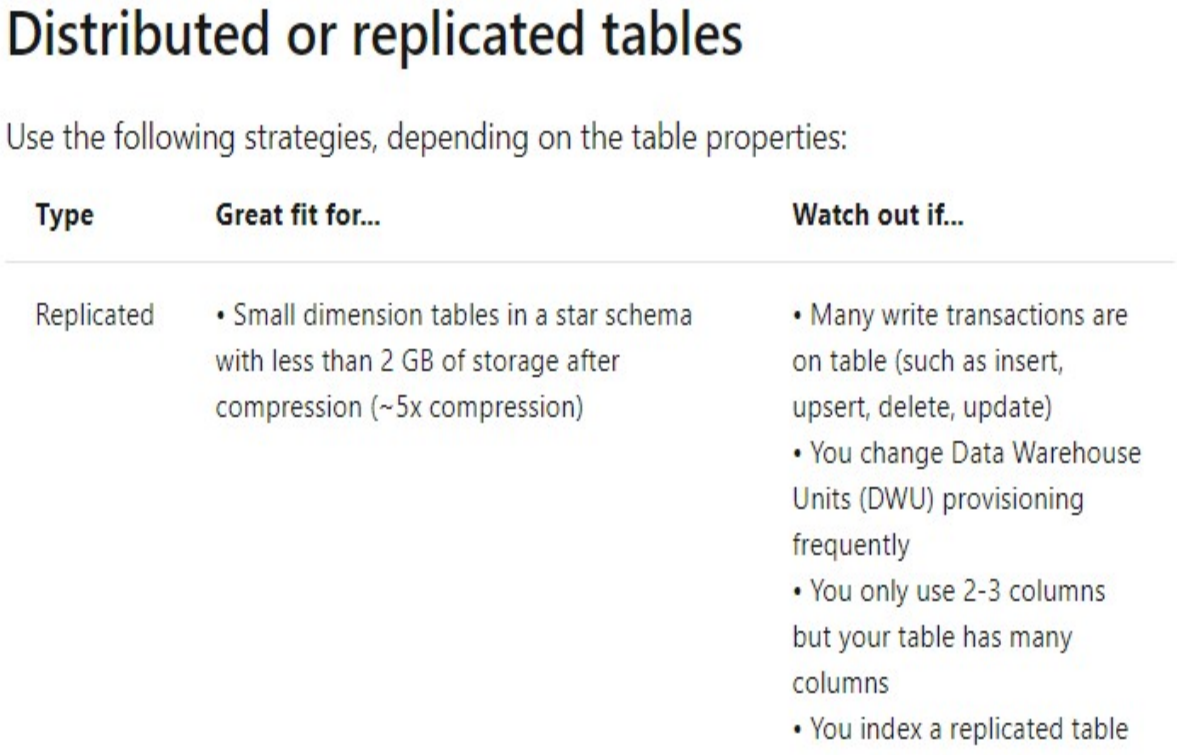
C. **Replicated**

D. **Round-Robin**

**Explanation:**

Answer – C

The Microsoft documentation mentions that the 'Replicated' table type would be ideal for tables less than 2 GB of size.




Since this is clearly given in the documentation, all other options are incorrect

For more information on a cheat sheet for designing tables, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/cheat-sheet

Ask our Experts

Rate this Question?

View Queries

open

Question 23

Correct

Domain :Implement data storage solutions

You have an Azure SQL Database named whizlabdb. The database contains a table named whizlabcustomer. The table has a column named customerID that is of the type varchar(22). You have to implement masking for the customerID which would meet the following requirements

The first two prefix characters must be exposed

The last four prefix characters must be exposed

All other characters must be masked

You decide to implement data masking and use a credit card function mask
Would this fulfil the requirement?

A. Yes

B. No

**Explanation:**

Answer – B

The Credit card masking function is specifically required for those columns that store credit card information and cannot be used for this requirement.

The Microsoft documentation mentions the following on the function mask

| | |
|---|---|
| **Credit card** | **Masking method, which exposes the last four digits of the designated fields** and adds a constant string as a prefix in the form of a credit card.<br><br>XXXX-XXXX-XXXX-1234 |

For more information on dynamic data masking, please visit the following URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 24

Correct

Domain :Implement data storage solutions

You have an Azure SQL Database named whizlabdb. The database contains a table named whizlabcustomer. The table has a column named customerID that is of the type varchar(22). You have to implement masking for the customerID which would meet the following requirements

The first two prefix characters must be exposed

The last four prefix characters must be exposed

All other characters must be masked

You decide to implement data masking and use a random number function mask. Would this fulfil the requirement?

A. Yes

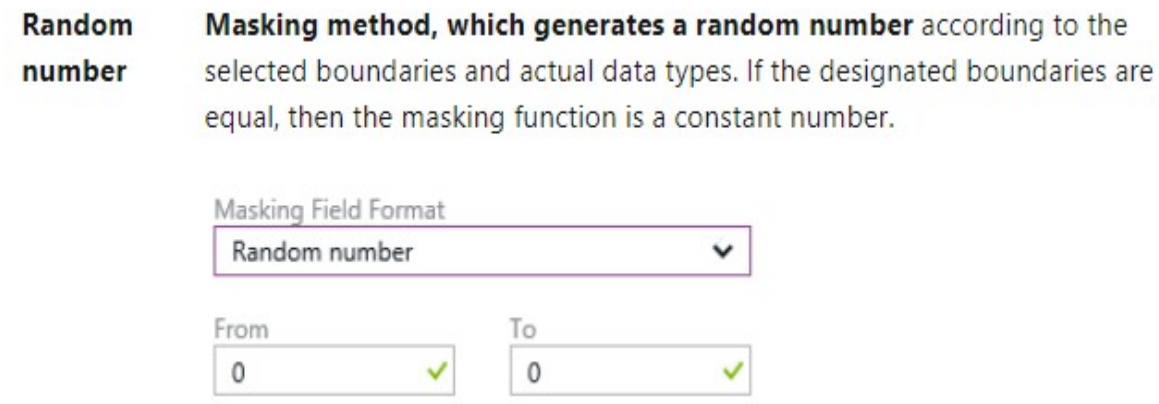
B. No

**Explanation:**

Answer - B

This masking function is used to mask specific boundaries

The Microsoft documentation mentions the following on the function mask




For more information on dynamic data masking, please visit the following URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 25

Correct

Domain :Implement data storage solutions

You have an Azure SQL Database named whizlabdb. The database contains a table named whizlabcustomer. The table has a column named customerID that is of the type varchar(22). You have to implement masking for the customerID which would meet the following requirements

The first two prefix characters must be exposed

The last four prefix characters must be exposed

All other characters must be masked

You decide to implement data masking and use an email function mask.
Would this fulfil the requirement?

A. Yes

B. No

**Explanation:**

Answer - B

The email masking function is specifically required for those columns that store email information and cannot be used for this requirement.

The Microsoft documentation mentions the following on the function mask

| | |
|---|---|
| **Email** | **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.<br><br>aXX@XXXX.com |

For more information on dynamic data masking, please visit the following URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 26

Correct

Domain :Implement data storage solutions

You have an Azure Data Lake Storage Gen 2 account. Your use account has contributor access to the storage account, and you have the application ID and access key. You need to use PolyBase to load data into the Azure SQL Data warehouse.
You need to configure PolyBase to connect the data warehouse to the storage account.
Which of the following would you need to create for this requirement? Choose 3 answers from the options given below

- [ ] A. A database encryption key
- [ ] B. An Asymmetric key
- [x] C. An external data source
- [x] D. An external file format
- [x] E. A database scoped credential

**Explanation:**

Answer – C,D and E

The Microsoft documentation mentions the steps required to load data into Azure SQL Data warehouse from Azure Data Lake storage accounts

First you have to create a database master key and a database scoped credential

# Create a 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. When authenticating using service principals, the Database Scoped Credential stores the service principal credentials set up in AAD. You can also use the storage account key in the Database Scoped Credential for Gen2.

SQL

Copy

```sql
-- A: Create a Database Master Key.
-- Only necessary if one does not already exist.
-- Required to encrypt the credential secret in the next step.
-- For more information on Master Key: https://msdn.microsoft.com/library/ms174

CREATE MASTER KEY;


-- B (for service principal authentication): Create a database scoped credentia
-- IDENTITY: Pass the client id and OAuth 2.0 Token Endpoint taken from your Az
-- SECRET: Provide your AAD Application Service Principal key.
-- For more information on Create Database Scoped Credential: https://msdn.micr

CREATE DATABASE SCOPED CREDENTIAL ADLSCredential
WITH
    -- Always use the OAuth 2.0 authorization endpoint (v1)
    IDENTITY = '<client_id>@<OAuth_2.0_Token_EndPoint>',
    SECRET = '<key>'
;

-- B (for Gen2 storage key authentication): Create a database scoped credential
-- IDENTITY: Provide any string, it is not used for authentication to Azure sto
-- SECRET: Provide your Azure storage account key.

CREATE DATABASE SCOPED CREDENTIAL ADLSCredential
WITH
    IDENTITY = 'user',
    SECRET = '<azure_storage_account_key>'
;

-- It should look something like this when authenticating using service princip
CREATE DATABASE SCOPED CREDENTIAL ADLSCredential
WITH
    IDENTITY = '536540b4-4239-45fe-b9a3-629f97591c0c@https://login.microsoftonl
    SECRET = 'BjdIlmtKp4Fpyh9hIvr8HJlUida/seM5kQ3EpLAmeDI='
;
```

Then you have to create an external data source.

## Create the external data source

Use this CREATE EXTERNAL DATA SOURCE command to store the location of the data.

SQL Copy

```
-- C (for Gen1): Create an external data source
-- TYPE: HADOOP - PolyBase uses Hadoop APIs to access data in Azure Data Lake S
-- LOCATION: Provide Data Lake Storage Gen1 account name and URI
-- CREDENTIAL: Provide the credential created in the previous step.

CREATE EXTERNAL DATA SOURCE AzureDataLakeStorage
WITH (
    TYPE = HADOOP,
    LOCATION = 'adl://<datalakestoregen1accountname>.azuredatalakestore.net',
    CREDENTIAL = ADLSCredential
);

-- C (for Gen2): Create an external data source
-- TYPE: HADOOP - PolyBase uses Hadoop APIs to access data in Azure Data Lake S
-- LOCATION: Provide Data Lake Storage Gen2 account name and URI
-- CREDENTIAL: Provide the credential created in the previous step.

CREATE EXTERNAL DATA SOURCE AzureDataLakeStorage
WITH (
    TYPE = HADOOP,
    LOCATION='abfs[s]://<container>@<AzureDataLake account_name>.dfs.core.windo
    CREDENTIAL = ADLSCredential
);
```

You then need to configure the external data 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. For the complete list, look at our T-SQL documentation CREATE EXTERNAL FILE FORMAT

SQL Copy

```
-- D: Create an external file format
-- FIELD_TERMINATOR: Marks the end of each field (column) in a delimited text f
-- STRING_DELIMITER: Specifies the field terminator for data of type string in
-- DATE_FORMAT: Specifies a custom format for all date and time data that might
-- Use_Type_Default: Store missing values as default for datatype.

CREATE EXTERNAL FILE FORMAT TextFileFormat
WITH
(   FORMAT_TYPE = DELIMITEDTEXT
,    FORMAT_OPTIONS    (   FIELD_TERMINATOR = '|'
                     ,    STRING_DELIMITER = ''
                     ,    DATE_FORMAT         = 'yyyy-MM-dd HH:mm:ss.fff'
                     ,    USE_TYPE_DEFAULT = FALSE
                     )
);
```

Since this is clearly given in the documentation, all other options are incorrect

For more information on loading data from Azure Data Lake storage onto SQL Data warehouse, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

**Question 27** **Correct**

**Domain :Monitor and optimize data solutions**

A company has an application that allows developers to share and compare code. The conversations for the code snippets, the code snippets themselves and the linked shared are all stored in an Azure SQL database instance. The application also allows uses to search for historical conversations and code snippets. Matches to previous code snippets also take place in the application. This comparison is done via Transact-SQL functions. If the application finds a match, a link to the match is added to the conversation.
Currently the following issues are occurring within the application

There are delays which occur during live conversations

There is a delay which occurs before the matching link appears after the code snippet is added to the conversation

Which of the following can be used to resolve the below issue?
**"There are delays which occur during live conversations"**

- A. **columnstore index**
- B. **non-durable table**
- C. **materialized view**
- D. **memory-optimized table** ✓

**Explanation:**

Answer – D

You can use Memory optimized technologies to speed up read operations on transactional data

The Microsoft documentation mentions the following

# When to use In-Memory technologies

By using In-Memory technologies in Azure SQL Database, you can achieve performance improvements with various workloads:

- **Transactional** (online transactional processing (OLTP)) where most of the requests read or update smaller set of data (for example, CRUD operations).
- **Analytic** (online analytical processing (OLAP)) where most of the queries have complex calculations for the reporting purposes, with a certain number of queries that load and append data to the existing tables (so called bulk-load), or delete the data from the tables.
- **Mixed** (hybrid transaction/analytical processing (HTAP)) where both OLTP and OLAP queries are executed on the same set of data.

In-memory technologies can improve performance of these workloads by keeping the data that should be processed into the memory, using native compilation of the queries, or advanced processing such as batch processing and SIMD instructions that are available on the underlying hardware.

Option A is incorrect since this is ideal when you want to perform analytical queries on the table

Option B is incorrect since this is used for storing temporary data which is used for caching

Option C is incorrect since this is used for creating a view from multiple tables

For more information on using In-Memory technologies, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-in-memory

**Question 28** **Correct**

**Domain :Monitor and optimize data solutions**

A company has an application that allows developers to share and compare code. The conversations for the code snippets, the code snippets themselves and the linked shared are all stored in an Azure SQL database instance. The application also allows uses to search for historical conversations and code snippets. Matches to previous code snippets also take place in the application. This comparison is done via Transact-SQL functions. If the application finds a match, a link to the match is added to the conversation.
Currently the following issues are occurring within the application

There are delays which occur during live conversations

There is a delay which occurs before the matching link appears after the code snippet is added to the conversation

Which of the following can be used to resolve the below issue?
**"There is a delay which occurs before the matching link appears after the code snippet is added to the conversation"**

- A. **columnstore index**
- B. **non-durable table**
- C. **materialized view** ✓
- D. **memory-optimized table**

**Explanation:**

Answer – C

Here it is better to create a materialized view which can be used for faster comparison of code snippets

The documentation mentions the following on materialized views

These materialized views, which only contain data required by a query, allow applications to quickly obtain the information they need. In addition to joining tables or combining data entities, materialized views can include the current values of calculated columns or data items, the results of combining values or executing transformations on the data items, and values specified as part of the query. A materialized view can even be optimized for just a single query.

Option A is incorrect since this is ideal when you want to perform analytical queries on the table

Option B is incorrect since this is used for storing temporary data which is used for caching

Option D is incorrect since this should be used as the storage design for the underlying tables

For more information on Materialized views, please visit the following URL

https://docs.microsoft.com/en-us/azure/architecture/patterns/materialized-view



**Question 29**

**Correct**

**Domain :Monitor and optimize data solutions**

You have an Azure SQL Data Warehouse. You plan to use PolyBase to load data from CSV files located in Azure Data Lake Gen 2 by using an external table. You need to monitor for files with invalid schema errors. Which of the following is an error you would monitor for?

A. **'Java exception raised on call to HdfsBridge_Connect:Error [com.microsoft.polybase.client.KerberosSecureLogin] occurred while accessing external files'**

B. Cannot execute the query “Remote Query” against OLE DB provider “SQLNCLI11”: for linked server “(null)”, Query aborted-the maximum reject threshold (0 rows) was reached while reading from external source

C. ‘Java exception raised on call to HdfsBridge_Connect:Error[Unable to Instantiate LoginClass] occurred while accessing external files.’

D. ‘Java exception raised on call to HdfsBridge_Connect:Error[No FileSystem for schema:wabs]

**Explanation:**

Answer – B

The Microsoft tech community mentions the error to look for when there is an invalid schema

SQL Server 2016 or SQL DW connected to Azure blob storage. The CREATE EXTERNAL TABLE DDL points to a directory (and not a specific file) and the directory contains files with different schemas.
SSMS Error:
Select query on the external table gives the following error:
Msg 7320, Level 16, State 110, Line 14
Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11" for linked server "(null)". Query aborted-- the maximum reject threshold (0 rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.
(/nation/sensors.ldjson.txt)Column ordinal: 0, Expected data type: INT, Offending value: {"id":"S2740036465E2B","time":"2016-02-26T16:59:02.9300000Z","temp":23.3,"hum":0.77,"wind":17,"press":1032,"loc":[-76.90914996169623,38.8929314364726]} (Column Conversion Error), Error: Error converting data type NVARCHAR to INT.

Keep in mind there may be derivations of this error. The name of the first rejected file shows in SSMS with offending data types or values.
Possible Reason:
The reason this error happens is because each file has different schema. The PolyBase external table DDL when pointed to a directory recursively reads all the files in that directory. When a column or data type mismatch happens, this error could be seen in SSMS.
Possible Solution:
If the data for each table consists of one file, then use the filename in the LOCATION section prepended by the directory of the external files. If there are multiple files per table, put each set of files into different directories in Azure Blob Storage and then you can point LOCATION to the directory instead of a particular file. The latter suggestion is the best practices recommended by SQLCAT even if you have one file per table.

Since this is clearly mentioned, all other options are invalid

For more information on the article, please visit the following URL

https://techcommunity.microsoft.com/t5/DataCAT/PolyBase-Setup-Errors-and-Possible-Solutions/ba-p/305297

Ask our Experts

Rate this Question?

View Queries

open

Question 30

Correct

Domain :Monitor and optimize data solutions

The security team in your company currently uses Azure Databricks to analyse data emitted from various sources. You have to send the Apache Spark level events, the Spark structured streaming metrics and application metrics to Azure Monitor.
Which of the following would you implement for this requirement? Choose 3 answers from the options given below

- [ ] A. **In Azure Monitor, go ahead and create a new data source**
- [x] B. **Configure the current Azure Databricks cluster to use the monitoring library.**
- [ ] C. **Deploy an instance of Grafana to a new Azure virtual machine**
- [x] D. **Build the spark-listeners-loganalytics-1.0-SNAPSHOT.jar file**
- [x] E. **Create the required Dropwizard counters in the application code**

**Explanation:**

Answer – B, D and E

The steps for sending application logs and metrics to Azure Monitor is given in the Microsoft documentation

For sending Apache Spark level events and Spark Structured Streaming metrics, you can use the monitoring library

## Prerequisites

Configure your Azure Databricks cluster to use the monitoring library, as described in the GitHub readme.

> **Note**
>
> The monitoring library streams Apache Spark level events and Spark Structured Streaming metrics from your jobs to Azure Monitor. You don't need to make any changes to your application code for these events and metrics.

For application level metrics, you have to use Dropwizard

## Send application metrics using Dropwizard

Spark uses a configurable metrics system based on the Dropwizard Metrics Library. For more information, see Metrics in the Spark documentation.

To send application metrics from Azure Databricks application code to Azure Monitor, follow these steps:

1. Build the **spark-listeners-loganalytics-1.0-SNAPSHOT.jar** JAR file as described in the GitHub readme.

2. Create Dropwizard gauges or counters in your application code. You can use the `UserMetricsSystem` class defined in the monitoring library. The following example creates a counter named `counter1`.

Since this is clearly given in the documentation, all other options are incorrect

For more information on Azure Databricks monitoring, please refer to the following link

https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs

Ask our Experts

Rate this Question?

View Queries

open

Question 31

Incorrect

Domain :Implement data storage solutions

You need to enable Transparent Data Encryption for an Azure SQL database. Which of the following steps would you perform for this requirement? Choose 4 answers from the options given below

- [ ] A. Create a database encryption key using a certificate
- [x] B. Create a certificate protected by the master key
- [ ] C. Set the context to the master database
- [x] D. Create a master key using a password
- [x] E. Set the context to the company database
- [x] F. Enable Encryption

**Explanation:**

Answer – A, B, D and F

The list of steps is given in the Microsoft documentation

# Using Transparent Data Encryption

To use TDE, follow these steps.

**Applies to**: SQL Server.

- Create a master key
- Create or obtain a certificate protected by the master key
- Create a database encryption key and protect it by the certificate
- Set the database to use encryption

The following example illustrates encrypting and decrypting the `AdventureWorks2012` database using a certificate installed on the server named `MyServerCert`.

SQL Copy

```
USE master;
GO
CREATE MASTER KEY ENCRYPTION BY PASSWORD = '<UseStrongPasswordHere>';
go
CREATE CERTIFICATE MyServerCert WITH SUBJECT = 'My DEK Certificate';
go
USE AdventureWorks2012;
GO
CREATE DATABASE ENCRYPTION KEY
WITH ALGORITHM = AES_128
ENCRYPTION BY SERVER CERTIFICATE MyServerCert;
GO
ALTER DATABASE AdventureWorks2012
SET ENCRYPTION ON;
GO
```

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on Transparent Data Encryption, please visit the following URL

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-ver15

Ask our Experts

Rate this Question?

View Queries

open

Question 32

Correct

Domain :Implement data storage solutions

An application is currently making use of a database on the Azure platform. Below is a snippet of the code base
private static readonly string whizlabendpointUrl =
ConfigurationManager.AppSettings["EndpointUrl"];
private static readonly SecureString
whizlabkey=ToSecureString(ConfigurationManager.AppSettings["AuthorizationKey"]);
var whizlab_client= new CosmosClient(new Url(whizlabendpointUrl), whizlabkey);
Database database= await whizlab_client.CreateDatabaseAsync(new Database {
Id="whizlabdb" );
Which of the following is the type of database the code is connecting to?

A. **Azure Cosmos DB**

B. **Azure SQL Database**

C. **Azure Storage Account – Blob**

D. **Azure SQL Datawarehouse**

**Explanation:**

Answer – A

This code is used to connect to an Azure Cosmos DB database. An example of the code is given in the Microsoft documentation

2. Copy and paste the code below where you instantiate the CosmosClient to call the **CreateDatabaseAsync** method you just added.

C# Copy

```
public async Task GetStartedDemoAsync()
{
    // Create a new instance of the Cosmos Client
    this.cosmosClient = new CosmosClient(EndpointUri, PrimaryKey);

    //ADD THIS PART TO YOUR CODE
    await this.CreateDatabaseAsync();
}
```

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on working with Cosmos DB from .Net, please visit the following URL

https://docs.microsoft.com/en-us/azure/cosmos-db/sql-api-get-started

Ask our Experts Rate this Question?

View Queries open

Question 33 Correct

Domain :Implement data storage solutions

An application is currently making use of a database on the Azure platform. Below is a snippet of the code base
private static readonly string whizlabendpointUrl =
ConfigurationManager.AppSettings[“EndpointUrl”];
private static readonly SecureString
whizlabkey=ToSecureString(ConfigurationManager.AppSettings[“AuthorizationKey”]);
var whizlab_client= new DocumentClient(new Url(whizlabendpointUrl), whizlabkey);

Database database= await whizlab_client.CreateDatabaseAsync(new Database {
Id="whizlabdb" );
Which of the following is the key type used in the code?

- A. **Resource token**
- B. **Master Key** ✓
- C. **Certificate**
- D. **Password**

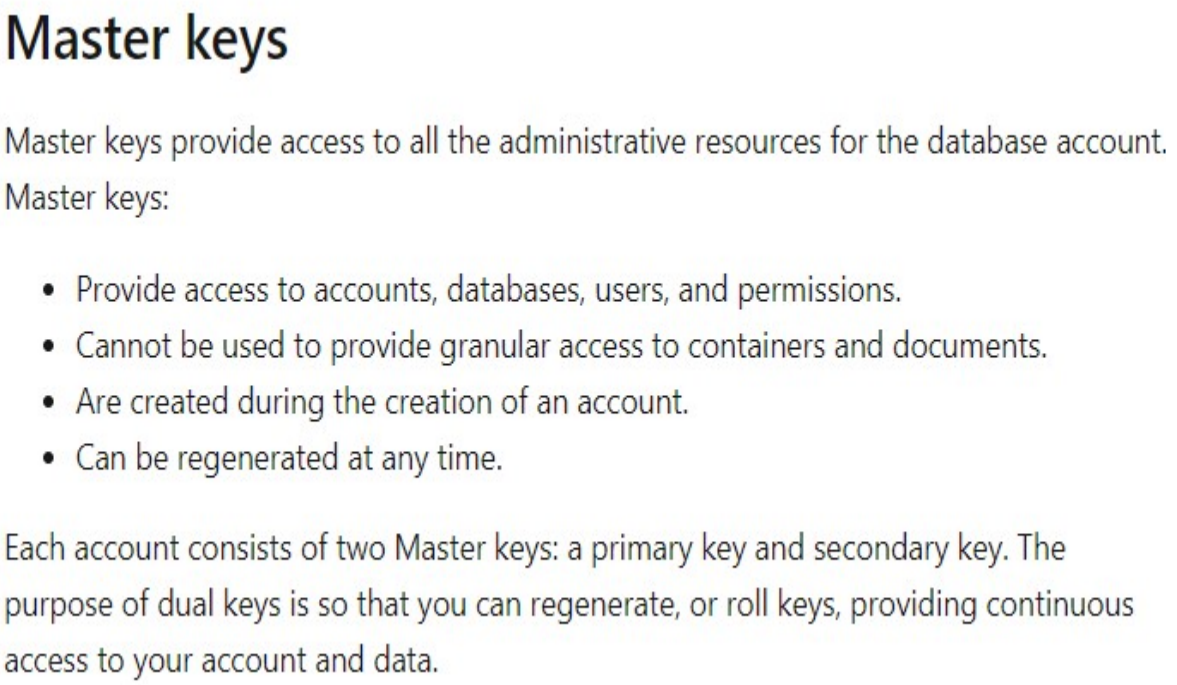
---

**Explanation:**

Answer – B

Here the master keys of the account are being used for authentication

The documentation mentions the following



## Master keys

Master keys provide access to all the administrative resources for the database account. Master keys:

- Provide access to accounts, databases, users, and permissions.
- Cannot be used to provide granular access to containers and documents.
- Are created during the creation of an account.
- Can be regenerated at any time.

Each account consists of two Master keys: a primary key and secondary key. The purpose of dual keys is so that you can regenerate, or roll keys, providing continuous access to your account and data.

Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on securing access to data, please visit the following URL

https://docs.microsoft.com/en-us/azure/cosmos-db/secure-access-to-data

Ask our Experts

Rate this Question?

View Queries

open

Question 34

Correct

Domain :Implement data storage solutions

A company is planning on setting up Azure SQL database to store sensitive data. The company wants to monitor data usage and data copied from the system to prevent data leakage. The company also wants to configure Azure SQL database to email a specific user when the data leakage occurs. Which of the following action would you need to perform? Choose 3 answers from the options given below

- [ ] A. **In Auditing, enable the auditing feature**
- [x] B. **Configure the service to create alerts for threat detections of the type "Data Exfiltration"**
- [ ] C. **In the Firewalls and virtual networks section, enable "Allow access to Azure services"**
- [x] D. **Enable Advanced threat protection**
- [x] E. **Configure the service to send email alerts to the IT security administrator**

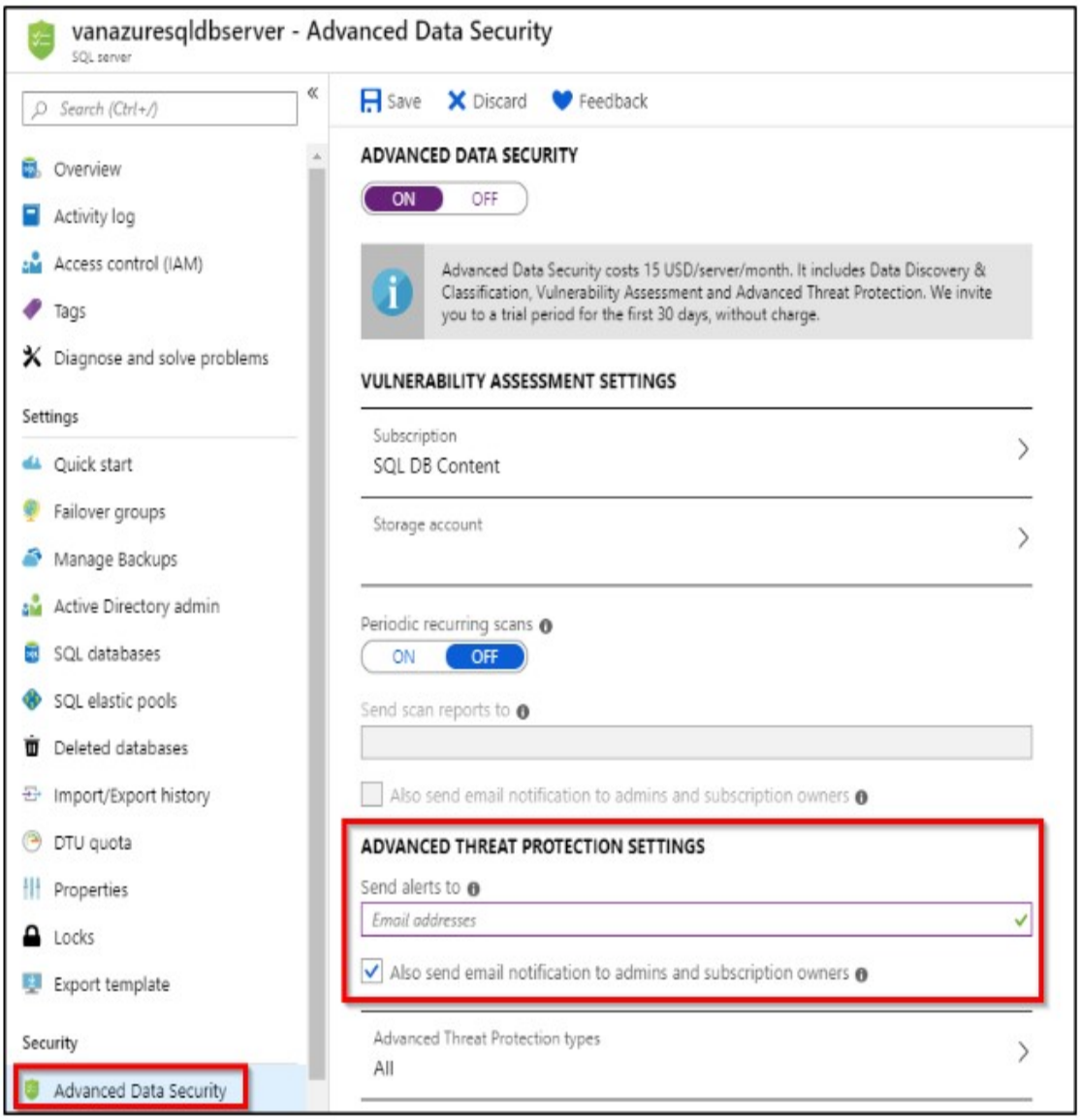
**Explanation:**

Answer – B,D and E

An example is given in the Microsoft documentation of enabling Advanced threat protection

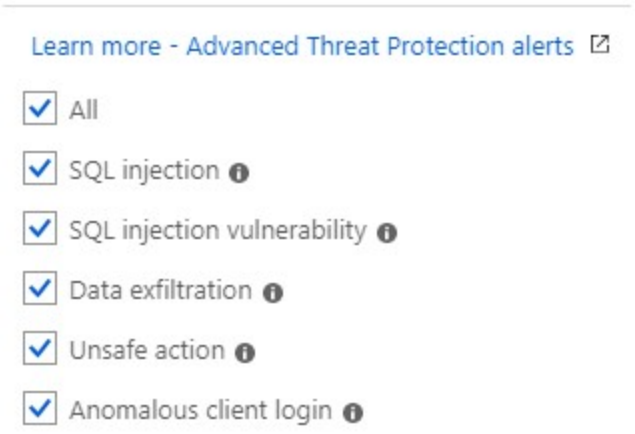
3. On the **Advanced Data Security** configuration page:

- Enable Advanced Data Security on the server.
- In **Advanced Threat Protection Settings**, in the **Send alerts to** text box, provide the list of emails to receive security alerts upon detection of anomalous database activities.




This setting can help detect when data leakage occurs.

One of the settings that can be set is “Data Exfiltration”




Since this is clearly mentioned in the Microsoft documentation, all other options are incorrect

For more information on Advanced threat protection, please visit the following URL

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-threat-detection

Ask our Experts

Rate this Question?

View Queries

open

Question 35

Correct

Domain :Monitor and optimize data solutions

Your company currently has an enterprise data warehouse in Azure Synapse Analytics.
You have to monitor the solution to see whether the data warehouse needs to be scaled up based on the current workloads.
Which of the following metric would you monitor for this requirement?

A. CPU Percentage

B. DWU used

C. DWU percentage

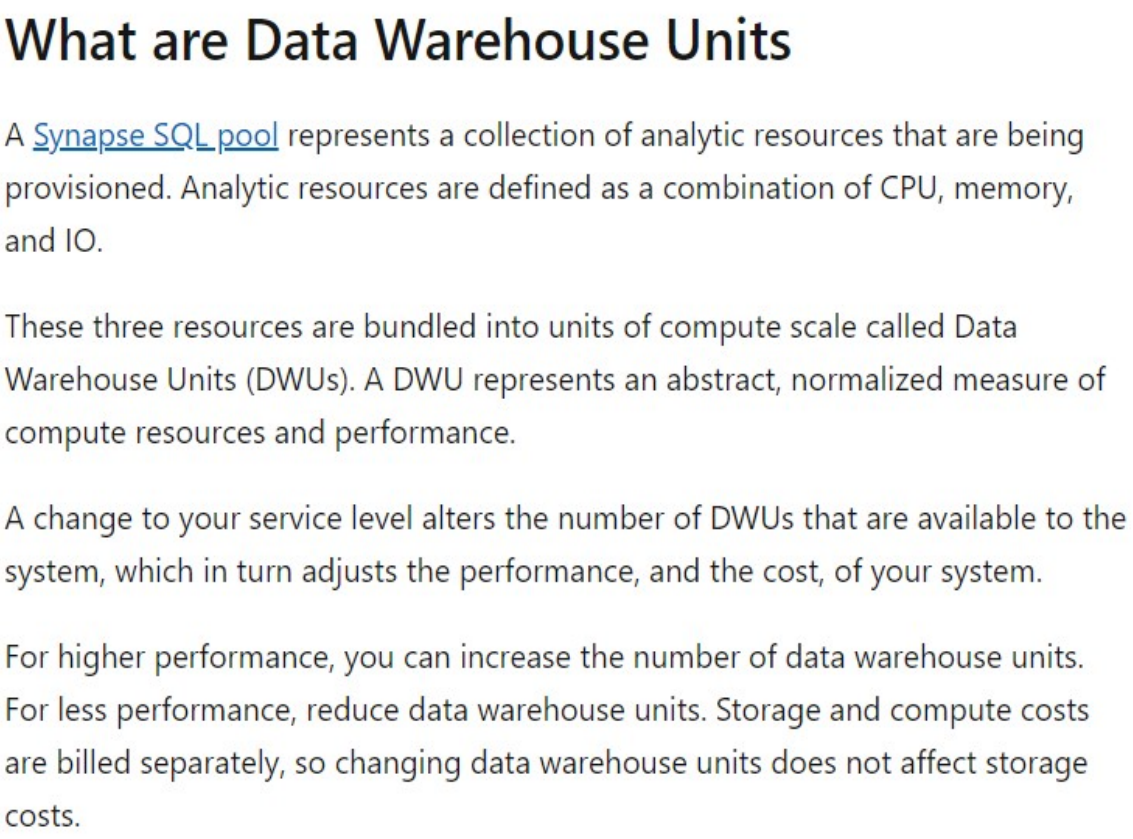
D. Data IO percentage

**Explanation:**

Answer – B

The resources allocated to the data warehouse are based on a metric called DWU or data Warehouse Units. This determines the resources allocated to the data warehouse

The Microsoft documentation mentions the following on Data Warehouse Units



## What are Data Warehouse Units

A Synapse SQL pool represents a collection of analytic resources that are being provisioned. Analytic resources are defined as a combination of CPU, memory, and IO.

These three resources are bundled into units of compute scale called Data Warehouse Units (DWUs). A DWU represents an abstract, normalized measure of compute resources and performance.

A change to your service level alters the number of DWUs that are available to the system, which in turn adjusts the performance, and the cost, of your system.

For higher performance, you can increase the number of data warehouse units. For less performance, reduce data warehouse units. Storage and compute costs are billed separately, so changing data warehouse units does not affect storage costs.

Option A and D are incorrect since DWU's are used for allocated of resources. But please note that you can still monitor these aspects as well for your data warehouse

Option C is incorrect because the right DWU to measure is the amount that has been used up

For more information on Data Warehouse Units, please refer to the following link

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/what-is-a-data-warehouse-unit-dwu-cdwu

Ask our Experts

Rate this Question?

View Queries

open

Question 36

Correct

Domain :Manage and develop data processing

A company wants to implement a lambda architecture on Microsoft Azure. The following are the key requirements for each architecture layer

**Data storage**

The data store should serve as a repository for high volumes of files.

The files can be large and of different formats

It should be optimized for big data analytics workloads

The data should be organized using a hierarchical structure

**Batch processing**

This layer should provide a managed solution for in-memory computation processing

It should provide support for a variety of programming languages

It should provide the ability to resize and terminate the cluster automatically

**Analytical data store**

This layer must provide support for SQL language

It must implement native columnar storage

It should support parallel processing

Which of the following should be used as a technology for the "Data Storage" layer?

A. **Azure SQL Database**

B. **Azure Blob storage**

C. **Azure Cosmos DB**

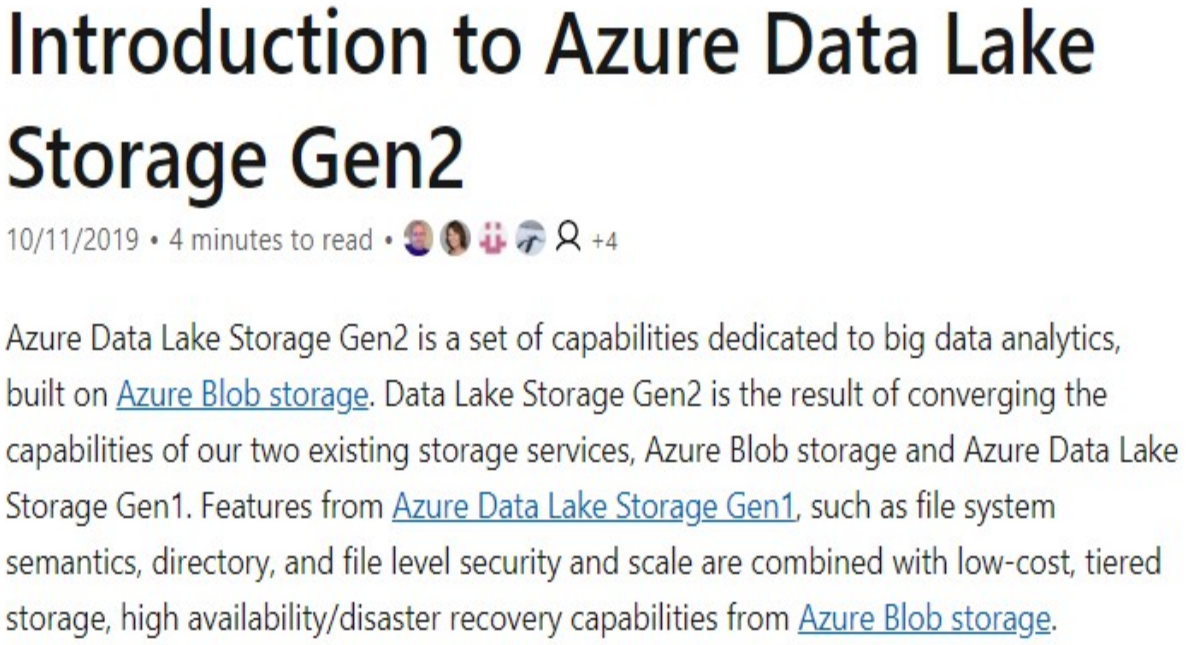
D. **Azure Data Lake Storage**

**Explanation:**

Answer – D

Azure Data Lake Storage fulfils all of the right aspects as being built for Big Data Analytics. It can also scale in terms of storage. Here you can store different types of files, different file sizes as well.

The Microsoft documentation mentions the following



# Introduction to Azure Data Lake Storage Gen2

10/11/2019 • 4 minutes to read • +4

Azure Data Lake Storage Gen2 is a set of capabilities dedicated to big data analytics, built on Azure Blob storage. Data Lake Storage Gen2 is the result of converging the capabilities of our two existing storage services, Azure Blob storage and Azure Data Lake Storage Gen1. Features from Azure Data Lake Storage Gen1, such as file system semantics, directory, and file level security and scale are combined with low-cost, tiered storage, high availability/disaster recovery capabilities from Azure Blob storage.

Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure Data Lake Storage, one can visit the below URL

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

Ask our Experts

Rate this Question?

View Queries

open

Question 37

Correct

**Domain :Manage and develop data processing**

A company wants to implement a lambda architecture on Microsoft Azure. The following are the key requirements for each architecture layer

**Data storage**

The data store should serve as a repository for high volumes of files.

The files can be large and of different formats

It should be optimized for big data analytics workloads

The data should be organized using a hierarchical structure

**Batch processing**

This layer should provide a managed solution for in-memory computation processing

It should provide support for a variety of programming languages

It should provide the ability to resize and terminate the cluster automatically

**Analytical data store**

This layer must provide support for SQL language

It must implement native columnar storage

It should support parallel processing

Which of the following should be used as a technology for the "Batch processing" layer?

- A. **HDInsight Spark**
- B. **HDInsight Hadoop**
- C. **Azure Databricks** ✓
- D. **HDInsight Interactive Query**

---

**Explanation:**

Answer – C

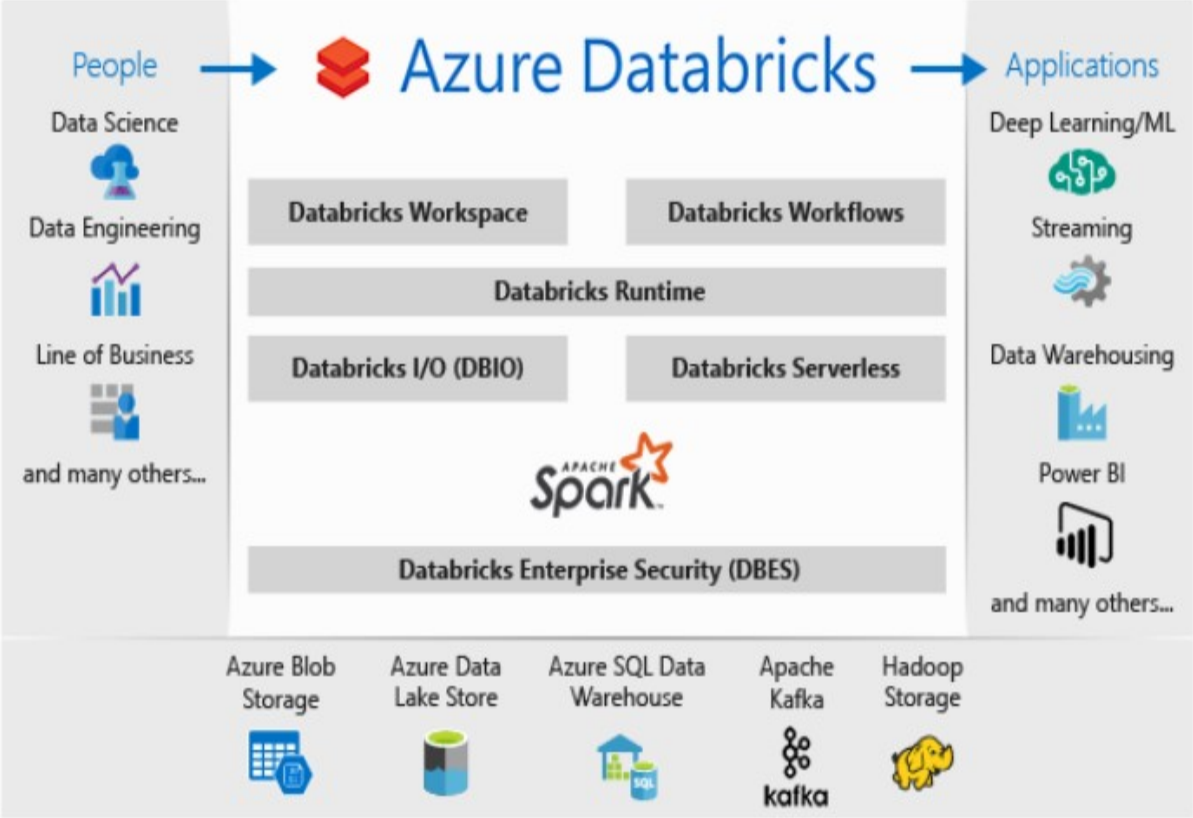
With Azure Databricks, you can setup clusters that can be terminated automatically. You can configure these clusters to perform computational processing. You can make use of Notebooks based on different programming languages

The Microsoft documentation mentions the following

# What is Azure Databricks?

05/08/2019 • 3 minutes to read •

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform. Designed with the founders of Apache Spark, Databricks is integrated with Azure to provide one-click setup, streamlined workflows, and an interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.




Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure Databricks, one can visit the below URL

https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks

Ask our Experts

Rate this Question?

View Queries

open

Question 38

Correct

Domain :Manage and develop data processing

A company wants to implement a lambda architecture on Microsoft Azure. The following are the key requirements for each architecture layer

**Data storage**

The data store should serve as a repository for high volumes of files.

The files can be large and of different formats

It should be optimized for big data analytics workloads

The data should be organized using a hierarchical structure

**Batch processing**

This layer should provide a managed solution for in-memory computation processing

It should provide support for a variety of programming languages

It should provide the ability to resize and terminate the cluster automatically

**Analytical data store**

This layer must provide support for SQL language

It must implement native columnar storage

It should support parallel processing

Which of the following should be used as a technology for the "Analytical data store" layer?

A. **HDInsight Base**

B. **Azure SQL Data warehouse** ✓

C. **Azure Analysis services**

D. **Azure Cosmos DB**

**Explanation:**

Answer – B

For columnar storage you can make use of Azure SQL data warehouse

The Microsoft documentation mentions the following

> In a cloud data solution, data is ingested into big data stores from a variety of sources. Once in a big data store, Hadoop, Spark, and machine learning algorithms prepare and train the data. When the data is ready for complex analysis, SQL Analytics uses PolyBase to query the big data stores. PolyBase uses standard T-SQL queries to bring the data into SQL Analytics tables.
>
> SQL Analytics stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves query performance. Once data is stored, you can run analytics at massive scale. Compared to traditional database systems, analysis queries finish in seconds instead of minutes, or hours instead of days.
>
> The analysis results can go to worldwide reporting databases or applications. Business analysts can then gain insights to make well-informed business decisions.

Since this is the perfect fit for the requirement, all other options are incorrect

For more information on Azure SQL data warehouse, one can visit the below URL

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



Question 39

Correct

Domain :Implement data storage solutions

View Case Study

Which of the following should be used as the API for the Cosmos DB account?

- A. Cassandra
- B. Gremlin
- C. MongoDB
- D. SQL
- E. Table

**Explanation:**

Answer – C

Since the on-premise data store is a MongoDB database, it makes logical sense to migrate it to the MongoDB API in the Cosmos DB account.

Since this is the most logical solution, all other options are incorrect

For more information on Azure Cosmos DB MongoDB API, one can visit the below URL

https://docs.microsoft.com/en-us/azure/cosmos-db/mongodb-introduction

Ask our Experts

Rate this Question?

View Queries open

Question 40

Correct

Domain :Implement data storage solutions

View Case Study

Which of the following would you use for the consistency level for the database?

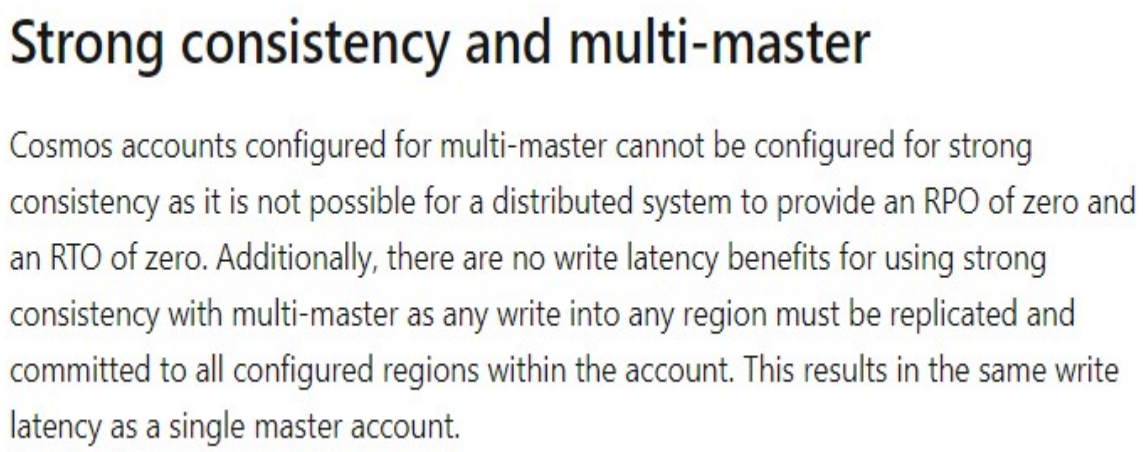
- A. **Eventual** ✅
- B. **Session**
- C. **Strong**
- D. **Consistent Prefix**

**Explanation:**

Answer – A

Since there is a requirement for data to be written to the closest data center for Cosmos DB, we need to ensure there is a multi-master setup for Cosmos DB wherein data can be written from multiple regions. For such accounts, we can't set the consistency level to Strong.

The Microsoft documentation mentions the following



## Strong consistency and multi-master

Cosmos accounts configured for multi-master cannot be configured for strong consistency as it is not possible for a distributed system to provide an RPO of zero and an RTO of zero. Additionally, there are no write latency benefits for using strong consistency with multi-master as any write into any region must be replicated and committed to all configured regions within the account. This results in the same write latency as a single master account.

Hence if we want data to converge in the least amount of time, we need to use Eventual consistency. This offers the least latency in terms of consistency.

The Microsoft documentation mentions the following on the consistency levels.

With Azure Cosmos DB, developers can choose from five well-defined consistency models on the consistency spectrum. From strongest to more relaxed, the models include *strong*, *bounded staleness*, *session*, *consistent prefix*, and *eventual* consistency. The models are well-defined and intuitive and can be used for specific real-world scenarios. Each model provides availability and performance tradeoffs and is backed by the SLAs. The following image shows the different consistency levels as a spectrum.


Because of the proposed logic to the consistency level, all other options are incorrect

For more information on Azure Cosmos DB consistency levels and trade-offs, one can visit the below URL

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

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

Ask our Experts

Rate this Question?

View Queries

open

**Question 41**

**Correct**

**Domain :Monitor and optimize data solutions**

View Case Study

You need to build the Azure SQL Data warehouse data store. Which of the following would you use as the underlying table type?

- A. **Hash distributed** ✓
- B. **Replicated**
- C. **Round-Robin**
- D. **Primary**

---

**Explanation:**

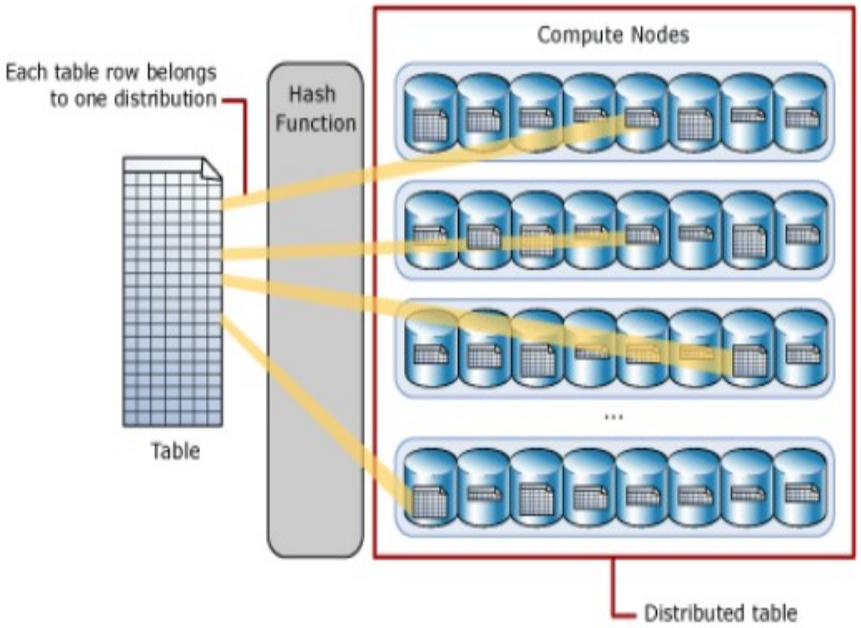
Answer – A

Since the data size is large and data is aggregated using a column, we can use the Hash distributed table type to get the best performance

The Microsoft documentation mentions the following

## Hash distributed

A hash-distributed table distributes table rows across the Compute nodes by using a deterministic hash function to assign each row to one distribution.




Since identical values always hash to the same distribution, the data warehouse has built-in knowledge of the row locations. SQL Data Warehouse uses this knowledge to minimize data movement during queries, which improves query performance.

Hash-distributed tables work well for large fact tables in a star schema. They can have very large numbers of rows and still achieve high performance. There are, of course, some design considerations that help you to get the performance the distributed system is designed to provide. Choosing a good distribution column is one such consideration that is described in this article.

Consider using a hash-distributed table when:

- The table size on disk is more than 2 GB.
- The table has frequent insert, update, and delete operations.

Since this is clear candidate for the table type, all other options are incorrect

For more information on Azure SQL Data warehouse table types, one can visit the below URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 42

Correct

Domain :Monitor and optimize data solutions

View Case Study

You need to build the Azure SQL Data warehouse data store. Which of the following would you use as the underlying index type?

A. Clustered

B. Clustered column store

C. Heap

D. Nonclustered

**Explanation:**

Answer – B

If you have large tables, you should consider a Clustered column store index.

The Microsoft documentation mentions the following

## Clustered columnstore indexes

By default, SQL Data Warehouse creates a clustered columnstore index when no index options are specified on a table. Clustered columnstore tables offer both the highest level of data compression as well as the best overall query performance. Clustered columnstore tables will generally outperform clustered index or heap tables and are usually the best choice for large tables. For these reasons, clustered columnstore is the best place to start when you are unsure of how to index your table.

Since this is clear candidate for the index type, all other options are incorrect

For more information on Azure SQL Data warehouse indexes, one can visit the below URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 43

Correct

Domain :Implement data storage solutions

View Case Study

Which of the following masking functions should be used for the "carID" column?

- A. **Credit Card** ✓
- B. **Default**
- C. **Email**
- D. **Random number**

**Explanation:**

Answer – A

The case study mentions the following on this column

> Only the last four digits of the values in the column carID must be shown

Hence for this we can use the Credit Card masking type

The Microsoft documentation mentions the following

| | |
|---|---|
| **Credit card** | **Masking method, which exposes the last four digits of the designated fields** and adds a constant string as a prefix in the form of a credit card.<br><br>XXXX-XXXX-XXXX-1234 |

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on dynamic masking, one can visit the below URL

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

Ask our Experts

Rate this Question? ☺ ☹

View Queries

open

Question 44

Correct

Domain :Implement data storage solutions

View Case Study

Which of the following masking functions should be used for the "carWeight" column?

- A. **Credit Card**
- B. **Default** ✅
- C. **Email**
- D. **Random number**

---

**Explanation:**

Answer – B

The case study mentions the following on this column

> A zero value must be shown for all values in the column carWeight

Hence for this we can use the Default masking type

The Microsoft documentation mentions the following

| Masking Function | Masking Logic |
|---|---|
| **Default** | **Full masking according to the data types of the designated fields**<br><br>• Use XXXX or fewer Xs if the size of the field is less than 4 characters for string data types (nchar, ntext, nvarchar).<br>• Use a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).<br>• Use 01-01-1900 for date/time data types (date, datetime2, datetime, datetimeoffset, smalldatetime, time).<br>• For SQL variant, the default value of the current type is used.<br>• For XML the document <masked/> is used.<br>• Use an empty value for special data types (timestamp table, hierarchyid, GUID, binary, image, varbinary spatial types). |

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on dynamic masking, one can visit the below URL

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

Ask our Experts

Rate this Question?

View Queries

open

Question 45

Correct

Domain :Manage and develop data processing

View Case Study

Which of the following should be included in the Data Factory Pipeline?

A. A copy activity that needs to use a stored procedure as the source

B. A copy activity that needs to use schema mappings

C. A delete activity that needs to have logging enabled

D. A filter activity that needs to have a condition

**Explanation:**

Answer – B

Since in the case study it mentions that for on-premise the attributes used to change when data was transferred from MongoDB to SQL Server, the same needs to be done for Azure CosmosDB to Azure SQL Server.

The Microsoft documentation mentions the following

# Schema mapping in copy activity

04/29/2019 • 6 minutes to read • +3

This article describes how the Azure Data Factory copy activity does schema mapping and data type mapping from source data to sink data when executing the data copy.

## Schema mapping

Column mapping applies when copying data from source to sink. By default, copy activity **map source data to sink by column names**. You can specify explicit mapping to customize the column mapping based on your need. More specifically, copy activity:

1. Read the data from source and determine the source schema
2. Use default column mapping to map columns by name, or apply explicit column mapping if specified.
3. Write the data to sink

Since this is the logical step as per the case study, all other options are incorrect

For more information on schema mapping in Azure Data Factory, one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-schema-and-type-mapping



Question 46

Correct

Domain :Monitor and optimize data solutions

View Case Study

The team is monitoring the Data Factory pipeline. They can see that the Cosmos DB to SQL database run time is taking 45 minutes. Which of the following can be carried out to improve the performance of the job?

- A. **Increase in the number of data integration units** ✓
- B. **Ensure the copy activity uses a staged copy**
- C. **Ensure the copy activity performs compression**
- D. **Ensure to decrease the amount pf parallelism in the copy activities**

**Explanation:**

Answer – A

One of the ways to increase performance is to use more Data Integration Units.

The Microsoft documentation mentions the following

# Copy performance optimization features

Azure Data Factory provides the following performance optimization features:

- Parallel copy
- Data Integration Units
- Staged copy

## Data Integration Units

A Data Integration Unit is a measure that represents the power (a combination of CPU, memory, and network resource allocation) of a single unit in Azure Data Factory. Data Integration Unit only applies to Azure integration runtime, but not self-hosted integration runtime.

You will be charged **# of used DIUs * copy duration * unit price/DIU-hour**. See the current prices here. Local currency and separate discounting may apply per subscription type.

The allowed DIUs to empower a copy activity run is **between 2 and 256**. If not specified or you choose "Auto" on the UI, Data Factory dynamically apply the optimal DIU setting based on your source-sink pair and data pattern. The following table lists the default DIUs used in different copy scenarios:

| Copy scenario | Default DIUs determined by service |
|---|---|
| Copy data between file-based stores | Between 4 and 32 depending on the number and size of the files |
| Copy data to Azure SQL Database or Azure Cosmos DB | Between 4 and 16 depending on the sink Azure SQL Database's or Cosmos DB's tier (number of DTUs/RUs) |
| All the other copy scenarios | 4 |

Option B is incorrect since this is used for scenarios like copying data to an Azure SQL data warehouse or when you copy from on-premise to Azure.

Option C is incorrect since applying compression might slow down the process

Option D is incorrect since having parallel copies can actually increase throughput

For more information on increasing performance for Azure Data Factory copy activities, one can visit the below URL

https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-performance

Ask our Experts

Rate this Question?

View Queries

open

Question 47

Correct

Domain :Monitor and optimize data solutions

View Case Study

Which of the following can be used to satisfy the case study requirement?
**"The query performance for data in the Azure SQL database must be stable without the need of administrative overhead"**

A. **sp_update stored procedure**

B. **Using the Query store**

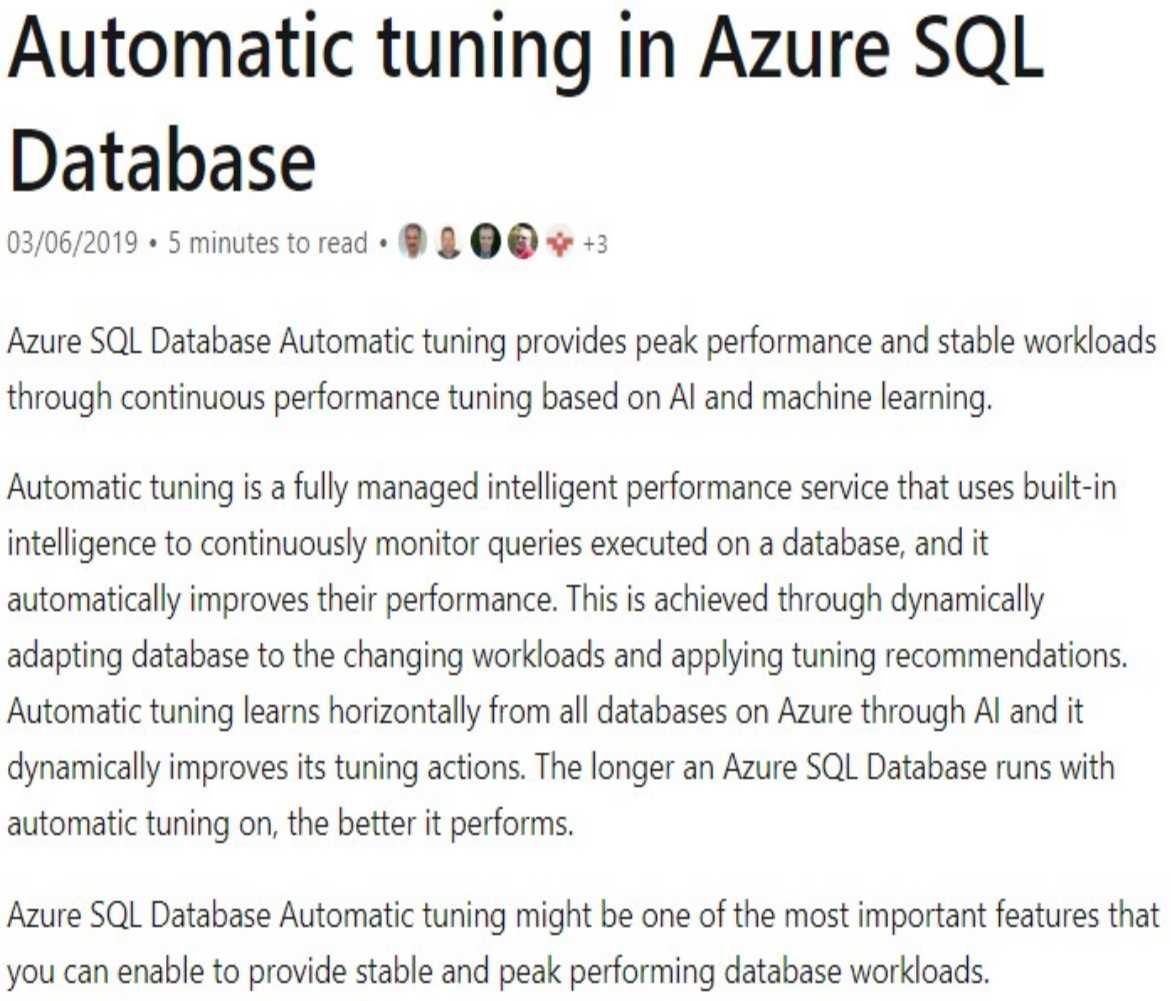
C. **Using the dbcc checkdb command**

D. **Using automatic tuning**

**Explanation:**

Answer – D

You can use automatic tunning which can ensure optimum query performance without the need of any sort of administrative overhead



# Automatic tuning in Azure SQL Database

03/06/2019 • 5 minutes to read • +3

Azure SQL Database Automatic tuning provides peak performance and stable workloads through continuous performance tuning based on AI and machine learning.

Automatic tuning is a fully managed intelligent performance service that uses built-in intelligence to continuously monitor queries executed on a database, and it automatically improves their performance. This is achieved through dynamically adapting database to the changing workloads and applying tuning recommendations. Automatic tuning learns horizontally from all databases on Azure through AI and it dynamically improves its tuning actions. The longer an Azure SQL Database runs with automatic tuning on, the better it performs.

Azure SQL Database Automatic tuning might be one of the most important features that you can enable to provide stable and peak performing database workloads.

All of the other methods all need administrative intervention

For more information on Azure SQL database automatic tuning, one can visit the below URL

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-automatic-tuning

Question 48 Correct

Domain :Monitor and optimize data solutions

View Case Study

You need to monitor the telemetry data being sent to Cosmos DB so that you can decide on the amount of Request Units to provision for Cosmos DB. Which of the following metrics must you monitor? Choose 2 answers from the options given below

- [x] A. **The number of requests**
- [ ] B. **The Session consistency**
- [x] C. **The data and index storage consumed**
- [ ] D. **The Average Throughput/second**

**Explanation:**

Answer – A and C

The Microsoft documentation gives the different elements to consider for the Request Units. Here we need to consider the number of requests and the amount of storage being consumed.

# Request Unit considerations

While you estimate the number of RUs per second to provision, consider the following factors:

- **Item size**: As the size of an item increases, the number of RUs consumed to read or write the item also increases.
- **Item indexing**: By default, each item is automatically indexed. Fewer RUs are consumed if you choose not to index some of your items in a container.
- **Item property count**: Assuming the default indexing is on all properties, the number of RUs consumed to write an item increases as the item property count increases.
- **Indexed properties**: An index policy on each container determines which properties are indexed by default. To reduce the RU consumption for write operations, limit the number of indexed properties.
- **Data consistency**: The strong and bounded staleness consistency levels consume approximately two times more RUs while performing read operations when compared to that of other relaxed consistency levels.
- **Query patterns**: The complexity of a query affects how many RUs are consumed for an operation. Factors that affect the cost of query operations include:
  - The number of query results
  - The number of predicates
  - The nature of the predicates
  - The number of user-defined functions
  - The size of the source data
  - The size of the result set
  - Projections

Since this is clear from the Microsoft documentation, all other options are incorrect

For more information on Cosmos DB request units, one can visit the below URL

https://docs.microsoft.com/en-us/azure/cosmos-db/request-units

Ask our Experts

Rate this Question?

View Queries open

Question 49

Incorrect

Domain :Manage and develop data processing

You have the following query defined in Azure Stream Analytics
WITH
step1 AS (SELECT *
FROM whizlabinput1
PARTITION BY OrderId
INTO 10),
step1 AS (SELECT *
FROM whizlabinput2
PARTITION BY OrderId
INTO 10)

SELECT *
INTO whizlaboutput
FROM step1
PARTITION BY OrderId
UNION step2
BY OrderId
Would the above query join two streams of partitioned data?

A. Yes

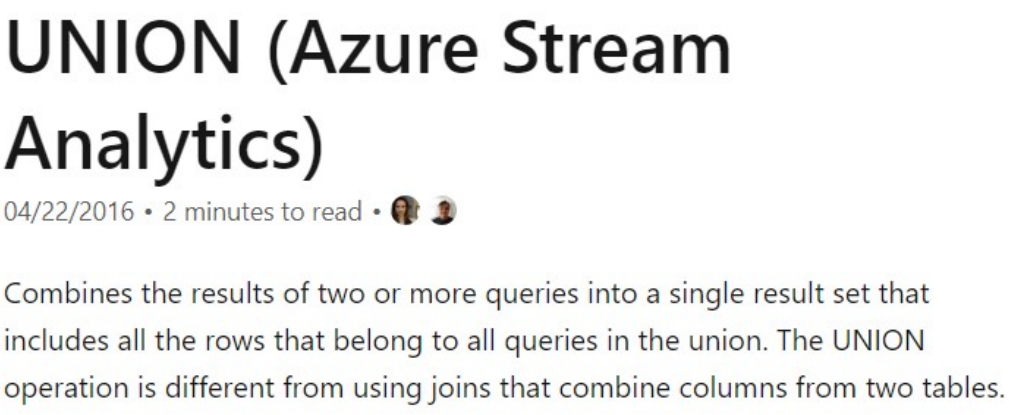
B. No

**Explanation:**

Answer – B

Here we are using the UNION clause which is different from the JOIN clause

The Microsoft documentation mentions the following



# UNION (Azure Stream Analytics)

04/22/2016 • 2 minutes to read •

Combines the results of two or more queries into a single result set that includes all the rows that belong to all queries in the union. The UNION operation is different from using joins that combine columns from two tables.

For more information on the UNION clause, please refer to the following link

https://docs.microsoft.com/en-us/stream-analytics-query/union-azure-stream-analytics

Ask our Experts

Rate this Question?

View Queries

open

**Question 50**

**Incorrect**

**Domain :Manage and develop data processing**

You have the following query defined in Azure Stream Analytics
WITH
step1 AS (SELECT *
FROM whizlabinput1
PARTITION BY OrderId
INTO 10),
step1 AS (SELECT *
FROM whizlabinput2
PARTITION BY OrderId
INTO 10)

SELECT *
INTO whizlaboutput
FROM step1
PARTITION BY OrderId
UNION step2
BY OrderId
Should the stream scheme key and the count match the one in the output scheme?

- A. Yes
- B. No

**Explanation:**

Answer – B

We need to match the partition key in the input and output scheme, but not necessarily need to match the count.

The Microsoft documentation mentions the following when it comes to the steps for performing parallel processing

> 2. The next step is to make your query is partitioned. For jobs with compatibility level 1.2 or higher (recommended), custom column can be specified as Partition Key in the input settings and the job will be paralellized automatically. Jobs with compatibility level 1.0 or 1.1, requires you to use **PARTITION BY PartitionId** in all the steps of your query. Multiple steps are allowed, but they all must be partitioned by the same key.

For more information on parallel processing in Azure Stream Analytics, please refer to the following link

> https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization

Ask our Experts

Rate this Question?

View Queries

open

Question 51

Incorrect

Domain :Manage and develop data processing

You have the following query defined in Azure Stream Analytics
WITH
step1 AS (SELECT *
FROM whizlabinput1
PARTITION BY OrderId
INTO 10),
step1 AS (SELECT *
FROM whizlabinput2
PARTITION BY OrderId
INTO 10)

SELECT *
INTO whizlaboutput
FROM step1
PARTITION BY OrderId
UNION step2
BY OrderId
Would providing 60 streaming units optimize the performance of the query?

A. Yes

B. No

**Explanation:**

Answer - B

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.

Hence this could be the ideal value for the optimization of the job.

For more information on parallel processing in Azure Stream Analytics, please refer to the following link

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization



**Question 52**

**Correct**

**Domain :Manage and develop data processing**

A company wants to use the Azure Databricks service. There is need to create clusters based on the following configuration

ClusterA – Here the cluster need to be configured to terminate automatically after 120 minutes

ClusterB – Here an environment needs to be created for each notebook

ClusterC – Here a group of data engineers will be sharing the same cluster

Which of the following cluster type would you set for ClusterA?

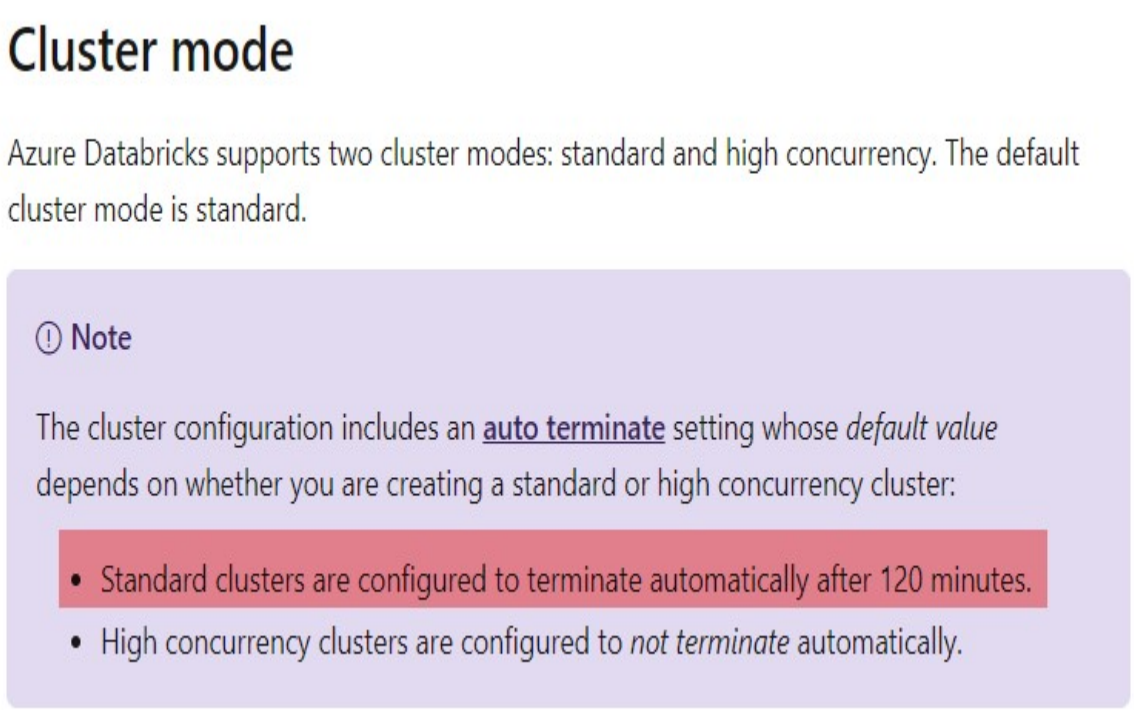
- A. **Standard** ✅
- B. **Basic**
- C. **Job**
- D. **High concurrency**

**Explanation:**

Answer – A

Here you have to set the cluster type as "Standard".

The Microsoft documentation mentions the following



## Cluster mode

Azure Databricks supports two cluster modes: standard and high concurrency. The default cluster mode is standard.

**Note**

The cluster configuration includes an auto terminate setting whose *default value* depends on whether you are creating a standard or high concurrency cluster:

- Standard clusters are configured to terminate automatically after 120 minutes.
- High concurrency clusters are configured to *not terminate* automatically.

Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on cluster configurations, one can visit the below URL

https://docs.microsoft.com/en-us/azure/databricks/clusters/configure

Ask our Experts

Rate this Question?

View Queries

open

Question 53

Correct

Domain :Manage and develop data processing

A company wants to use the Azure Databricks service. There is need to create clusters based on the following configuration

ClusterA – Here the cluster need to be configured to terminate automatically after 120 minutes

ClusterB – Here an environment needs to be created for each notebook

ClusterC – Here a group of data engineers will be sharing the same cluster

Which of the following cluster type would you set for ClusterB?

A. **Standard**

B. **Basic**

C. **Job**

D. **High concurrency**

**Explanation:**

Answer – D

Here you have to use the cluster type as “High concurrency”

The Microsoft documentation mentions the following


Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on cluster configurations, one can visit the below URL

https://docs.microsoft.com/en-us/azure/databricks/clusters/configure

Ask our Experts

Rate this Question?

View Queries

open

Question 54

Correct

Domain :Manage and develop data processing

A company wants to use the Azure Databricks service. There is need to create clusters based on the following configuration

ClusterA – Here the cluster need to be configured to terminate automatically after 120 minutes

ClusterB – Here an environment needs to be created for each notebook

ClusterC – Here a group of data engineers will be sharing the same cluster

Which of the following cluster type would you set for ClusterC?

- A. **Standard**
- B. **Basic**
- C. **Job**
- D. **High concurrency** ✅

**Explanation:**

Answer – D

Here you have to use the cluster type as "High concurrency"

The Microsoft documentation mentions the following

## High concurrency clusters

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. This sharing is accomplished with:

- **Preemption:** Proactively preempts Spark tasks from over-committed users to ensure all users get their fair share of cluster time and their jobs complete in a timely manner even when contending with dozens of other users. This uses Spark Task Preemption for High Concurrency.
- **Fault isolation:** Creates an environment for each notebook, effectively isolating them from one another.

Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on cluster configurations, one can visit the below URL

https://docs.microsoft.com/en-us/azure/databricks/clusters/configure

Ask our Experts

Rate this Question?

View Queries

open

Question 55

Correct

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database. They want to enable diagnostics logging for the database. Which of the following can be used to store the diagnostic logs for the database? Choose 2 answers from the options given below

- [x] A. Azure Event Hubs
- [x] B. Azure Storage
- [ ] C. Azure Cosmos DB
- [ ] D. Azure SQL Data warehouse

**Explanation:**

Answer – A and B

The Microsoft documentation provides the ways you can stream the diagnostics log data.

# Azure SQL Database metrics and diagnostics logging

11/15/2019 • 24 minutes to read •  +19

In this topic, you will learn how to configure logging of diagnostics telemetry for Azure SQL Database through the Azure portal, PowerShell, Azure CLI, Azure Monitor REST API, and Azure Resource Manager template. These diagnostics can be used to gauge resource utilization and query execution statistics.

Single databases, pooled databases in elastic pools, and instance databases in a managed instance can stream metrics and diagnostics logs for easier performance monitoring. You can configure a database to transmit resource usage, workers and sessions, and connectivity to one of the following Azure resources:

- **Azure SQL Analytics**: to get intelligent monitoring of your Azure SQL databases that includes performance reports, alerts, and mitigation recommendations.
- **Azure Event Hubs**: to integrate SQL Database telemetry with your custom monitoring solutions or hot pipelines.
- **Azure Storage**: to archive vast amounts of telemetry for a fraction of the price.

Since this is clearly mentioned in the documentation, all other options are incorrect

For more information on Azure SQL database diagnostic logging, one can visit the below URL

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-metrics-diag-logging

View Queries open

Finish Review

**Certification**

Cloud Certification

Java Certification

PM Certification

Big Data Certification

**Company**

Become Our Instructor

Support

Discussions

Blog

Business

**Follow us**