# Practice Test 1

### **Question 1**

Domain :Monitor and optimize data solutions

A company has an Azure SQL database. The database contains tables that have 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.

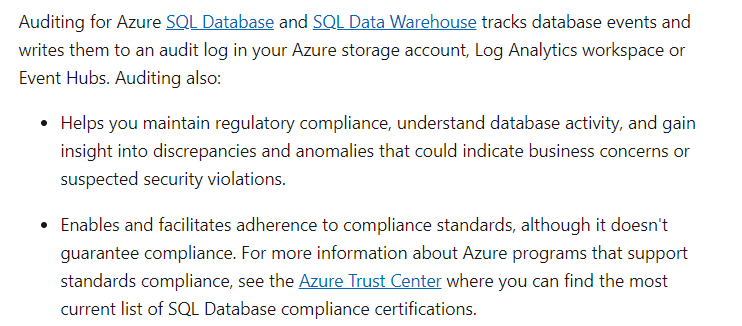
**Azure Monitor Audit Logs**

**Explanation:**

Answer – B

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

The Microsoft documentation mentions the following.



Option A is incorrect since this is used for enabling features such as 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>

### **Question 2**

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 that its 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.

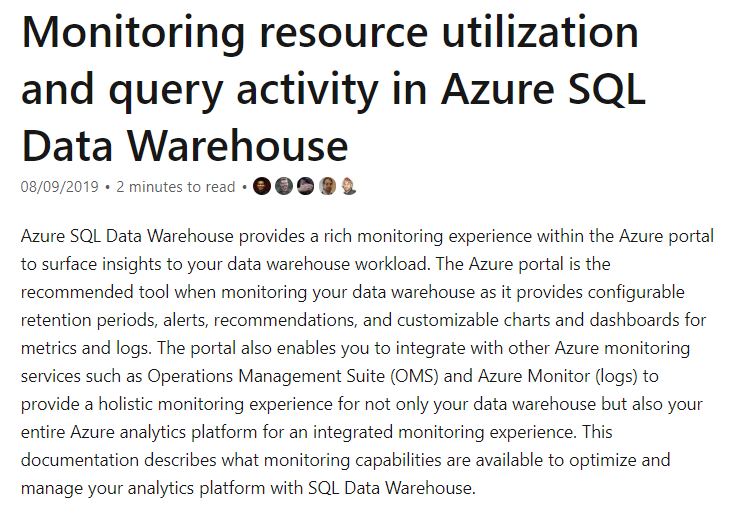
**Subscription**

**Explanation:**

Answer – C

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.



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**

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 that its support department gets an alert when the Data Warehouse consumes the maximum allotted resources.

Which of the following would you use as the signal type for the alert?

]A.

**CPU used**

]B.

**DWU limit**

]C.

**DWU used**

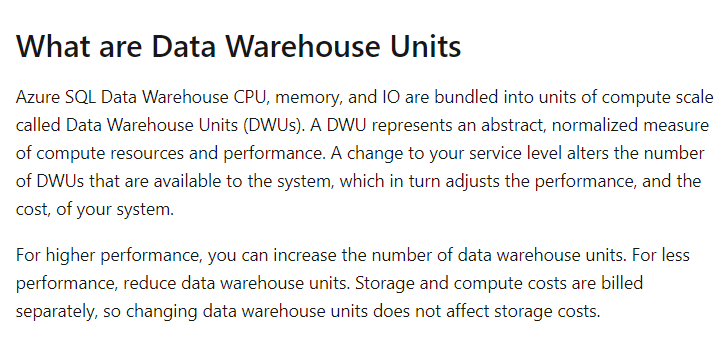
]D.

**Data IO Percentage**

**Explanation:**

Answer – C

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



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>

### **Question 4**

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database defined as part of its 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?

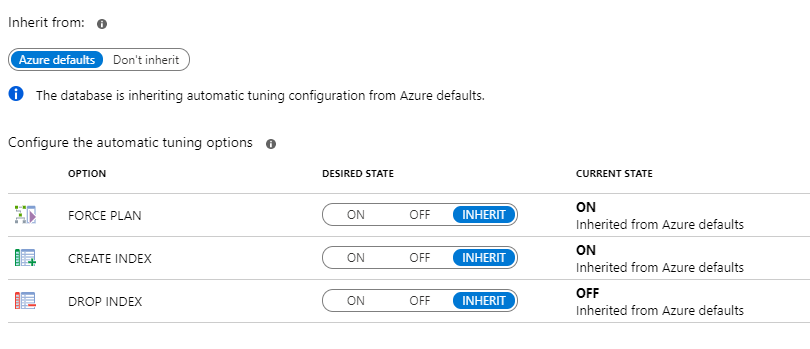
]A.**Yes**

]B.**No**

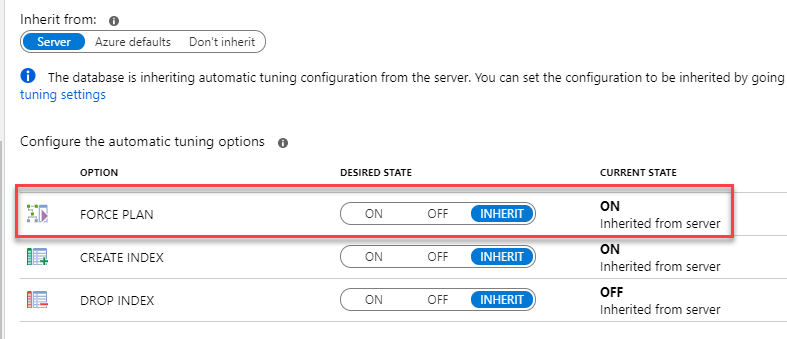
**Explanation:**

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>

### **Question 5**

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database defined as part of its 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?

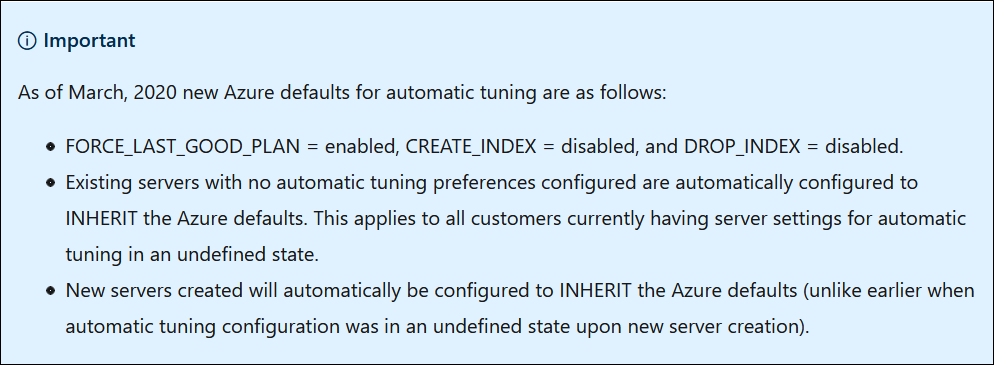
]A.**Yes**

]B.**No**

**Explanation:**

Answer – B

 By default, the Create Index is disabled.



For more information, please refer to the below links:

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

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

### **Question 6**

Domain :Monitor and optimize data solutions

A company has an Azure SQL Database defined as part of its 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?

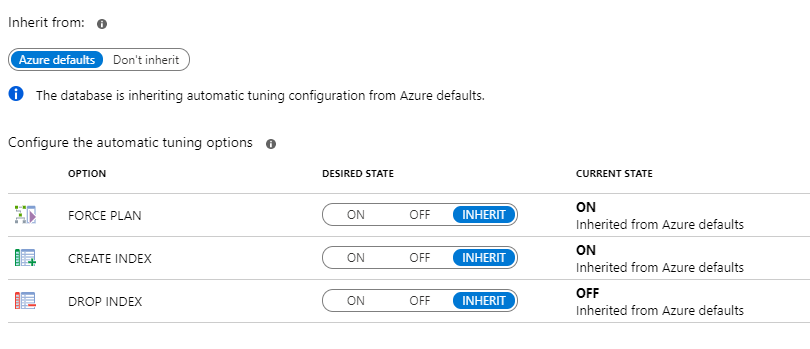
]A.**Yes**

]B.**No**

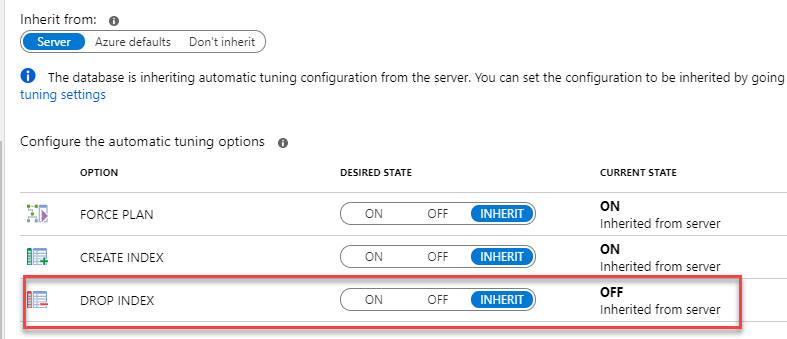
**Explanation:**

Answer – B

If you implement the below auditing settings for the server

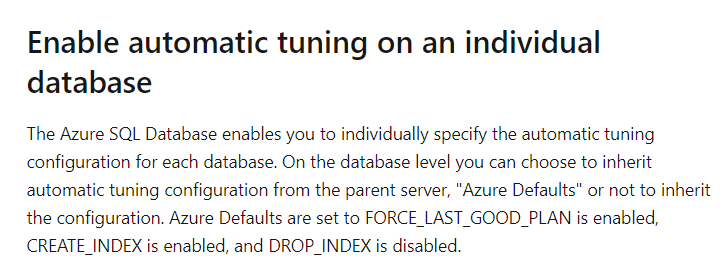


And also place the following settings for the database



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>

### **Question 7**

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 report that queries take a long time to execute when retrieving 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 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/>

### **Question 8**

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 report that queries take a long time to execute when retrieving 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 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/>

### **Question 9**

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 report that queries take a long time to execute when retrieving 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 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/>

### **Question 10**

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**

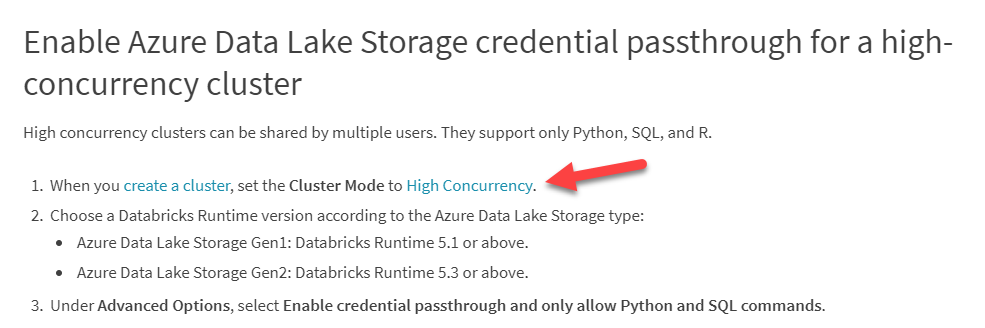
]D.

**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>

### **Question 11**

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 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**

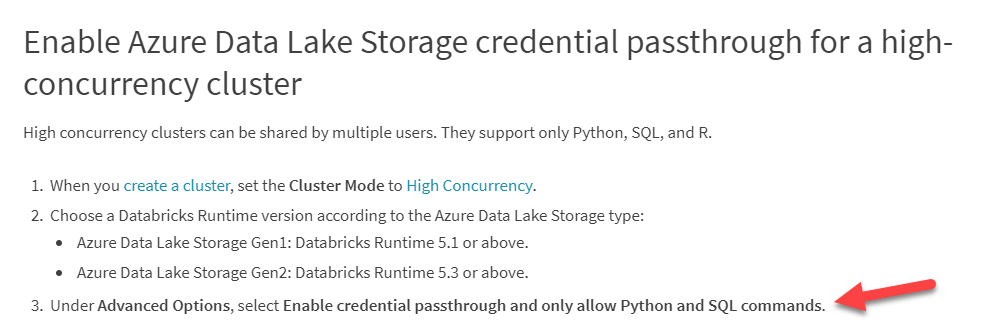
]D.

**Single Sign-On**

**Explanation:**

Answer – C

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



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>

### **Question 12**

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 that 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.

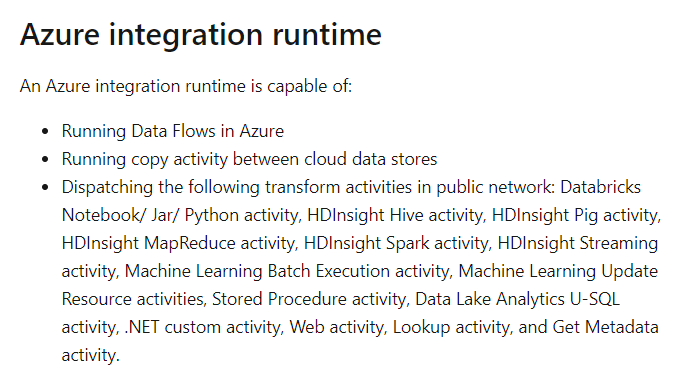
**Azure-SSIS**

**Explanation:**

Answer – A

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

The Microsoft documentation mentions the following.



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>

### **Question 13**

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.

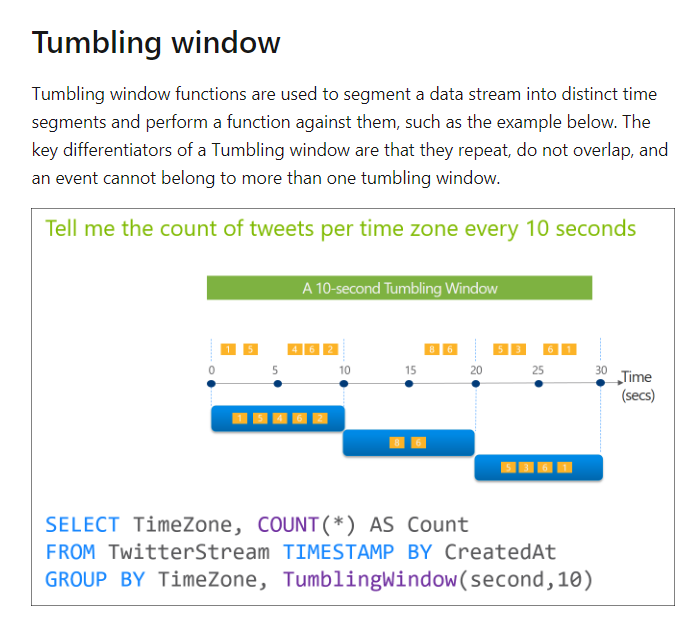
**Tumbling**

**Explanation:**

Answer – D

You need to use the Tumbling windowing function for this requirement.

The Microsoft documentation mentions the following.



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>

### **Question 14**

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.

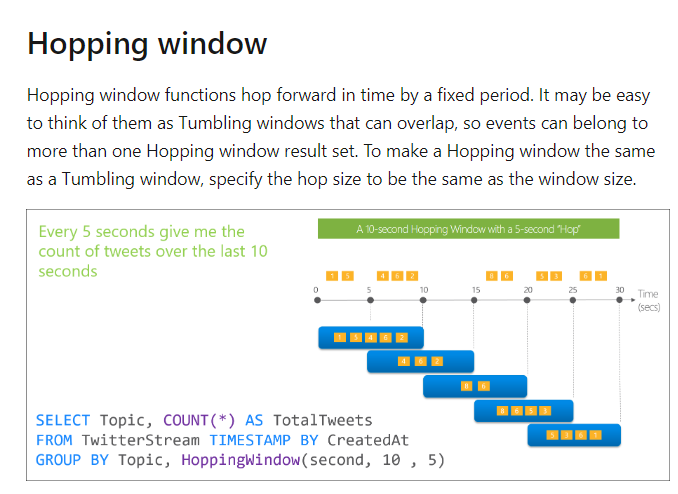
**Tumbling**

**Explanation:**

Answer – A

You need to use the Hopping windowing function for this requirement.

The Microsoft documentation mentions the following.



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>

### **Question 15**

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.

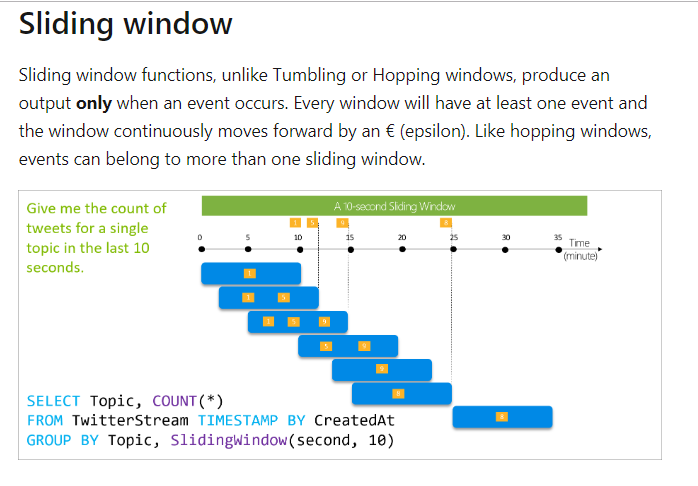
**Tumbling**

**Explanation:**

Answer – C

You need to use the Sliding windowing function for this requirement.

The Microsoft documentation mentions the following.



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>

### **Question 16**

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.**

C.

**Perform transformations on the data frame.**

D.

**Mount the Data Lake Storage onto DBFS.**

E.

**Perform transformations on the file.**

F.

**Read the file into a data frame.**

G.

**Specify a temporary folder to stage the data.**

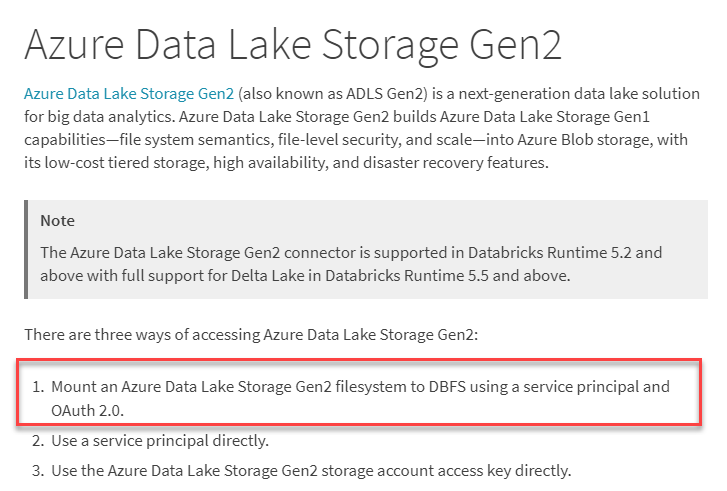
H.

**Write the results to a table in SQL Data Warehouse.**

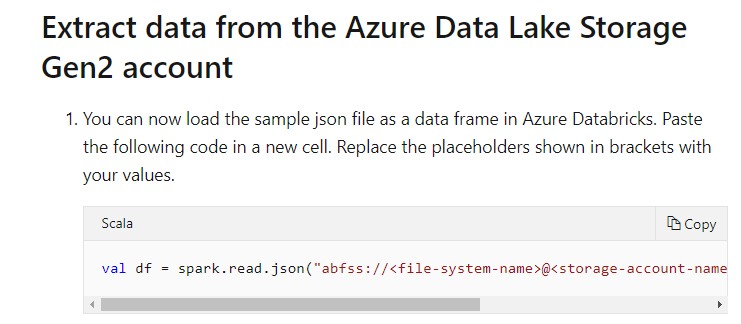
**Explanation:**

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.

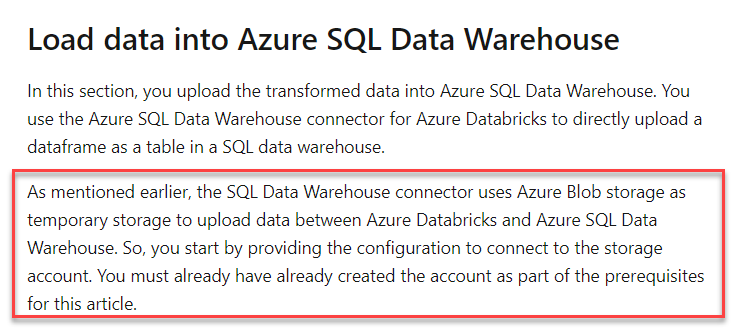


Next, we need to load the data from the Azure Data Lake Storage Gen2 account. An example is also given in the Microsoft documentation.



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 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**

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.

**%**

]B.

**#**

]C.

**@**

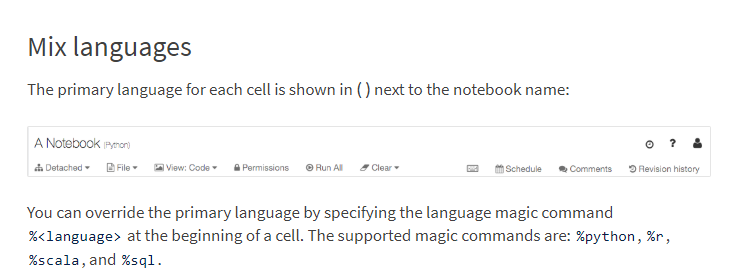
]D.

**&**

**Explanation:**

Answer – A

This can be done with the use of the % 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>

### **Question 18**

Domain :Manage and develop data processing

You have an Azure Data Lake Storage Gen 2 account. You have several CSV files loaded into the account. Each file has a header row. After the header row is a property 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.

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.**

C.

**Create an external file format and set the First\_row option.**

D.

**Create a database scoped credential that uses OAuth2 token and a key.**

E.

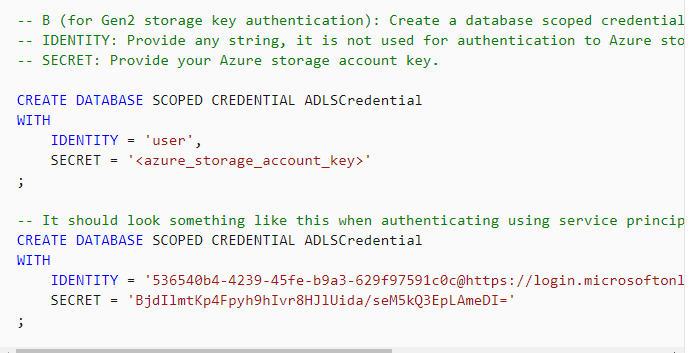
**Use the CREATE EXTERNAL TABLE AS SELECT and create a view that removes the empty row.**

**Explanation:**

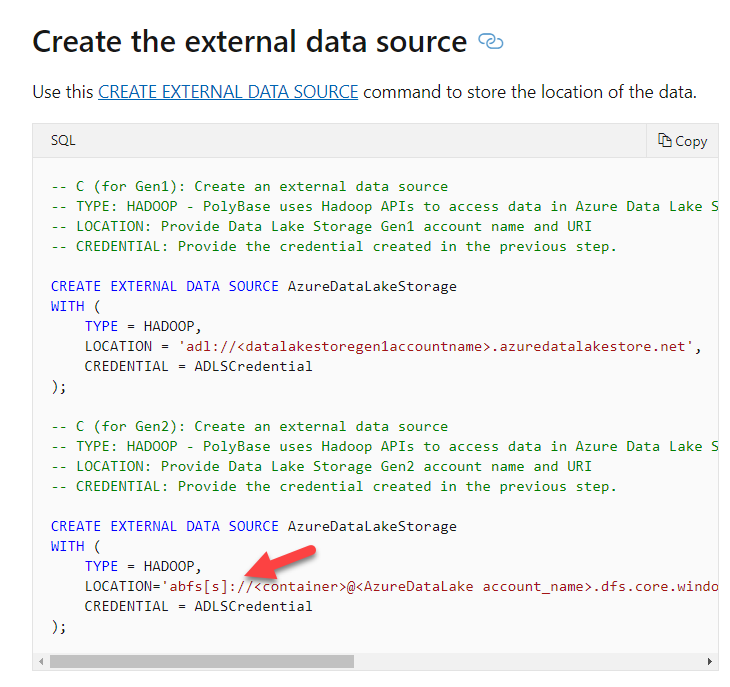
Answer – A, C and D

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

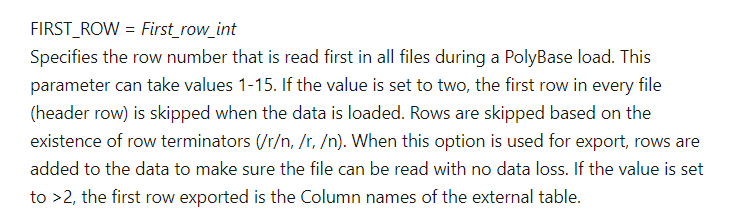
One of the steps is to create a database scoped credential.



Another step is to create the external data source using ‘abfs’ as the file location.



And you can use the FIRST\_ROW parameter to skip the first row of the file.



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>

### **Question 19**

Domain :Implement data storage solutions

A company is planning to create 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**

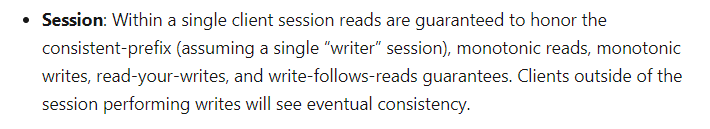
]E.

**Consistent Prefix**

**Explanation:**

Answer – D

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



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>

### **Question 20**

Domain :Implement data storage solutions

A company has an Azure SQL Datawarehouse. They have a table named comp\_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 comp\_salesfact\_new that has the same schema as comp\_salesfact.**

B.

**Drop the comp\_salesfact\_new table.**

C.

**Copy the data to the new table by using CREATE TABLE AS SELECT (CTAS).**

D.

**Truncate the partition containing the stale data.**

E.

**Switch the partition containing the stale data from comp\_salesfact to comp\_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 to 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/>

### **Question 21**

Domain :Implement data storage solutions

You have an Azure SQL data warehouse. You have used Polybase to create a table named [Ext].[compitems] 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 contain 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].[compitems] ADD [ItemID] int;**

]B.**ALTER TABLE [Ext].[compitems] ADD [ItemID] int;**

]C.**DROP TABLE [Ext].[compitems] CREATE EXTERNAL FILE FORAMT parquetfilenew WITH ( FORMAT\_TYPE=PARQUET, DATA\_COMPRESSION = ‘org.apache.hadoop.io.compress.SnappyCodec’ );**

]D.

**DROP TABLE [Ext].[compitems]**

**CREATE EXTERNAL TABLE [Ext].[compitems]**

**( [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 to create 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>

### **Question 22**

Domain :Implement data storage solutions

You are planning to create a dimension table in an Azure SQL Data Warehouse. The data in the table will be less than 1 GB. You need to ensure that 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**

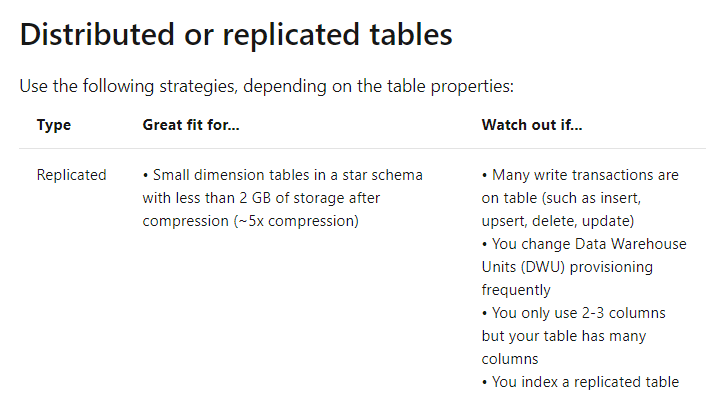
]D.

**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>

### **Question 23**

Domain :Implement data storage solutions

You have an Azure SQL Database named compdb. The database contains a table named compcustomer. 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 fulfill the requirement?

]A.**Yes**

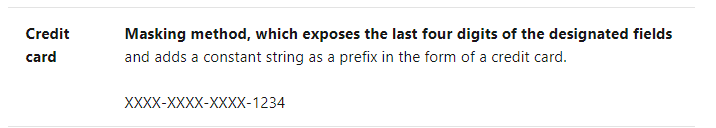
]B.**No**

**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.



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>

### **Question 24**

Domain :Implement data storage solutions

You have an Azure SQL Database named compdb. The database contains a table named compcustomer. 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 fulfill the requirement?

]A.**Yes**

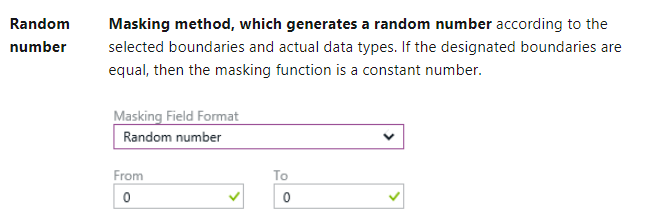
]B.**No**

**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>

### **Question 25**

Domain :Implement data storage solutions

You have an Azure SQL Database named compdb. The database contains a table named compcustomer. 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 fulfill the requirement?

]A.**Yes**

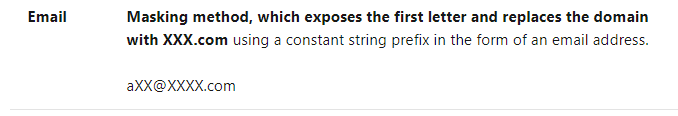
]B.**No**

**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.



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>

### **Question 26**

Domain :Implement data storage solutions

You have an Azure Data Lake Storage Gen 2 account. Your user account has contributor access to the storage account. 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**

C.

**An external data source**

D.

**An external file format**

E.

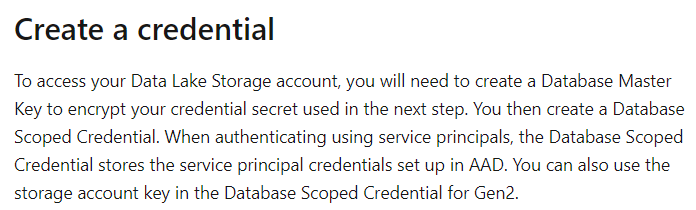
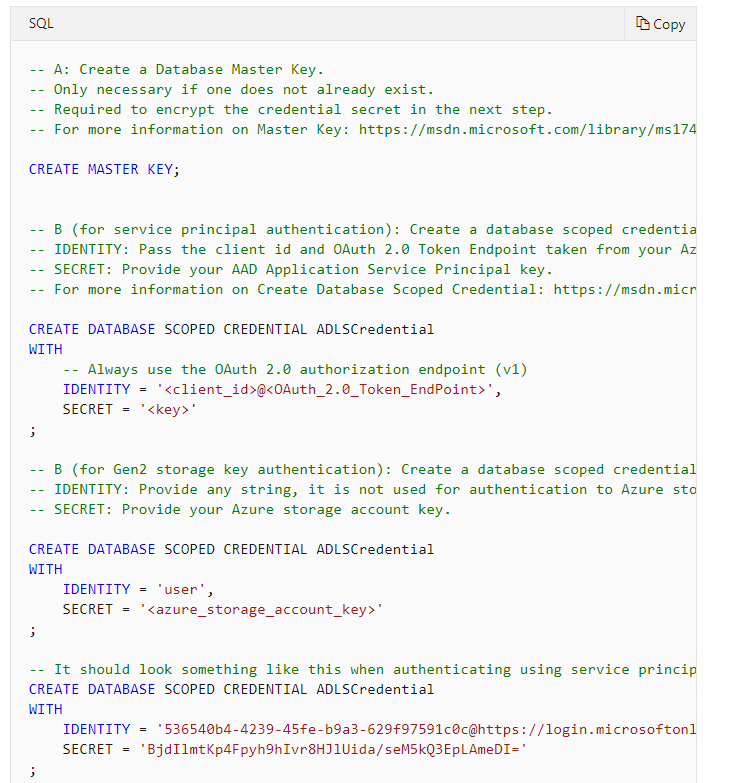
**A database scoped credential**

**Explanation:**

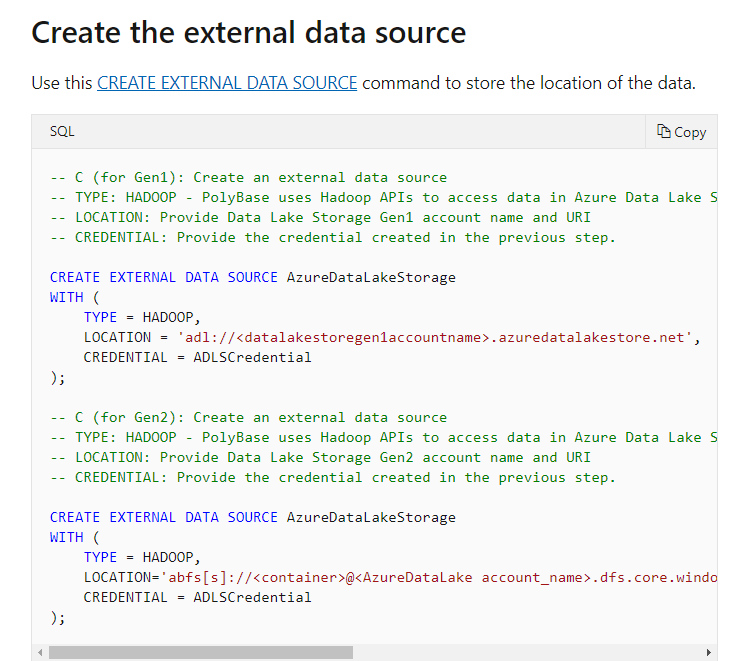
Answer – C, D and E

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

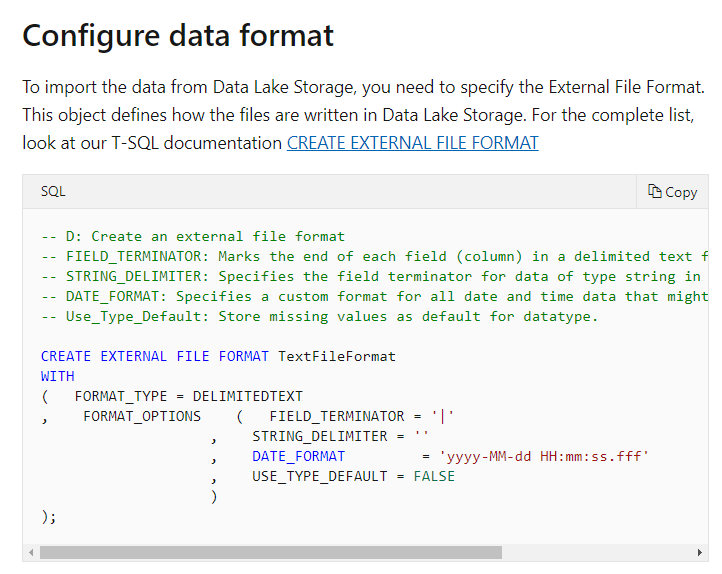
First, you have to create a database master key and a database scoped credential.

Then you have to create an external data source.



You then need to configure the external data format.



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**

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 to search for historical conversations and code snippets. Matches to previous code snippets also take place in the application. This comparison is made 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.

* Delays occur during live conversations.
* Delay 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.

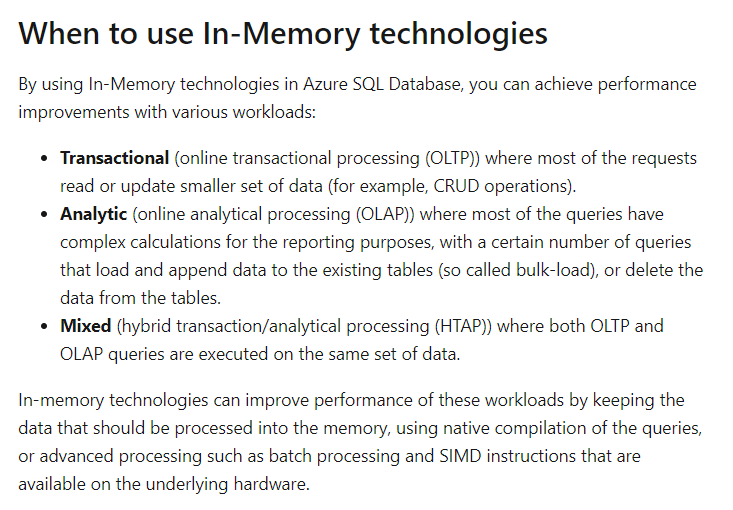
**memory-optimized table**

**Explanation:**

Answer – D

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

The Microsoft documentation mentions the following.



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**

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 to search for historical conversations and code snippets. Matches to previous code snippets also take place in the application. This comparison is made 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.

* Some delays occur during live conversations.
* A delay 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.

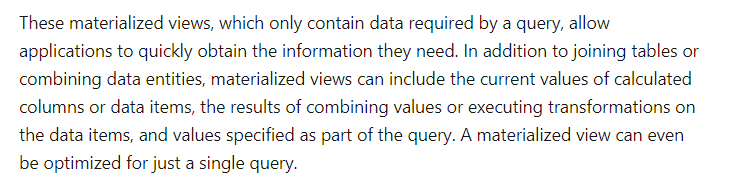
**memory-optimized table**

**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.



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**

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 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.’**

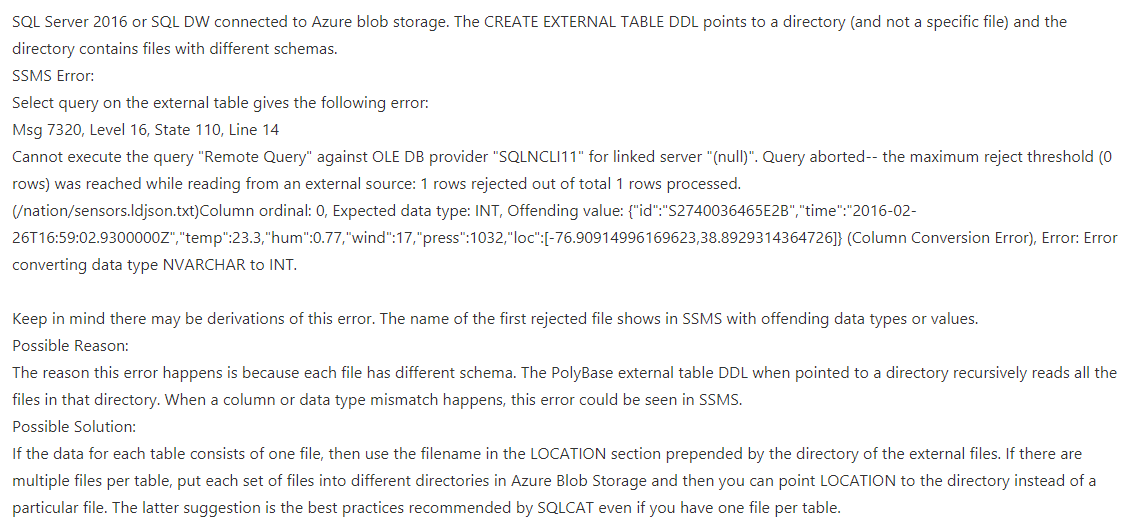
]D.

**‘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.



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>

### **Question 30**

Domain :Monitor and optimize data solutions

The security team in your company currently uses Azure Databricks to analyze 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.**

B.

**Configure the current Azure Databricks cluster to use the monitoring library.**

C.

**Deploy an instance of Grafana to a new Azure virtual machine.**

D.

**Build the spark-listeners-loganalytics-1.0-SNAPSHOT.jar file.**

E.

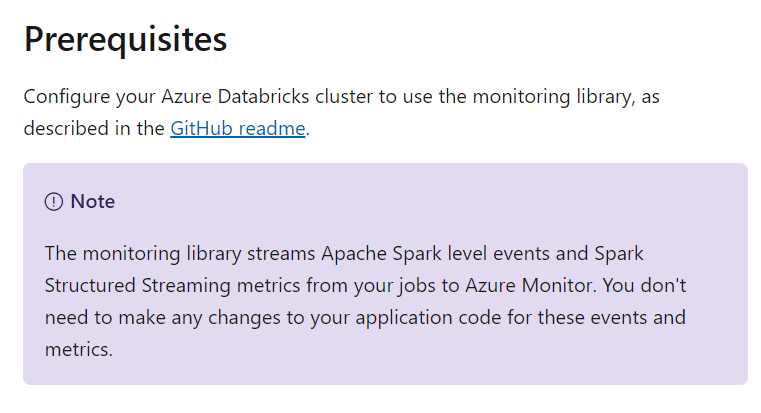
**Create the required Dropwizard counters in the application code.**

**Explanation:**

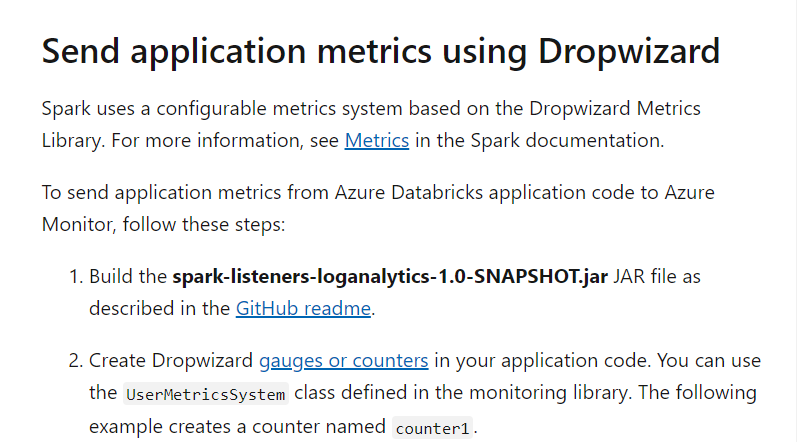
Answer – B, D and E

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

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



For application level metrics, you have to use Dropwizard.



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>

### **Question 31**

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.**

B.

**Create a certificate protected by the master key.**

C.

**Set the context to the master database.**

D.

**Create a master key using a password.**

E.

**Set the context to the company database.**

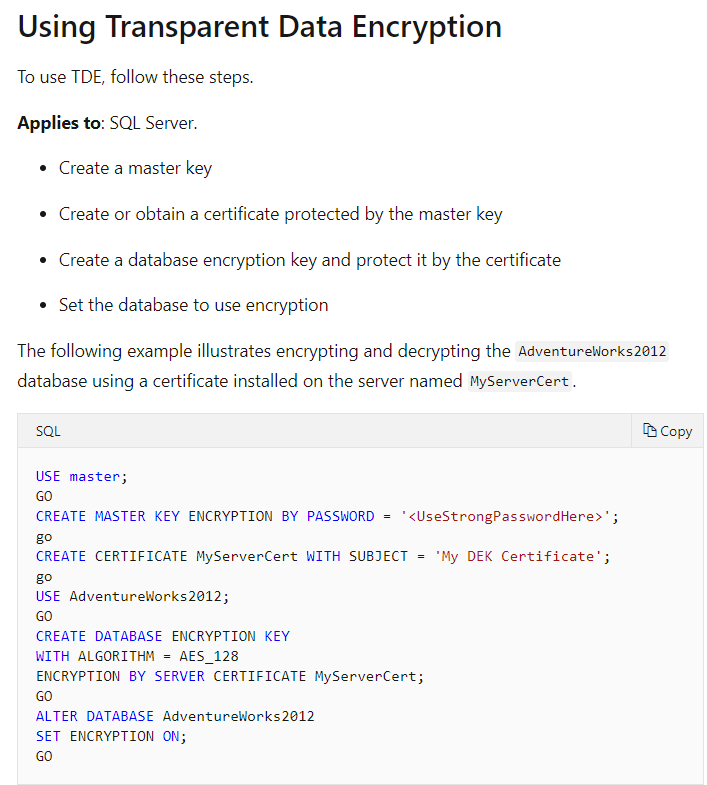
F.

**Enable Encryption.**

**Explanation:**

Answer – A, B, D and F

The list of steps is given in the Microsoft documentation.



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>

### **Question 32**

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 compendpointUrl = ConfigurationManager.AppSettings[“EndpointUrl”];

private static readonly SecureString compkey=ToSecureString(ConfigurationManager.AppSettings[“AuthorizationKey”]);

var comp\_client= new CosmosClient(new Url(compendpointUrl), compkey);

Database database= await comp\_client.CreateDatabaseAsync(new Database { Id=”compdb” );

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.



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>

### **Question 33**

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 compendpointUrl = ConfigurationManager.AppSettings[“EndpointUrl”];

private static readonly SecureString compkey=ToSecureString(ConfigurationManager.AppSettings[“AuthorizationKey”]);

var comp\_client= new DocumentClient(new Url(compendpointUrl), compkey);

Database database= await comp\_client.CreateDatabaseAsync(new Database { Id=”compdb” );

Which of the following is the key type used in the code?

]A.

**Resource token**

]B.

**Master Key**

]C.

**Certificate**

]D.

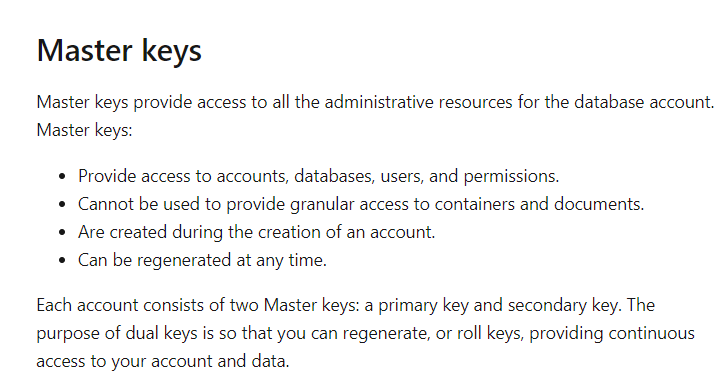
**Password**

**Explanation:**

Answer – B

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

The documentation mentions the following.



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>

### **Question 34**

Domain :Implement data storage solutions

A company is planning to set up an 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 the Azure SQL database to email a specific user when the data leakage occurs. Which of the following activities would you need to perform? Choose 3 answers from the options given below.

A.

**In Auditing, enable the auditing feature.**

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”.**

D.

**Enable Advanced threat protection.**

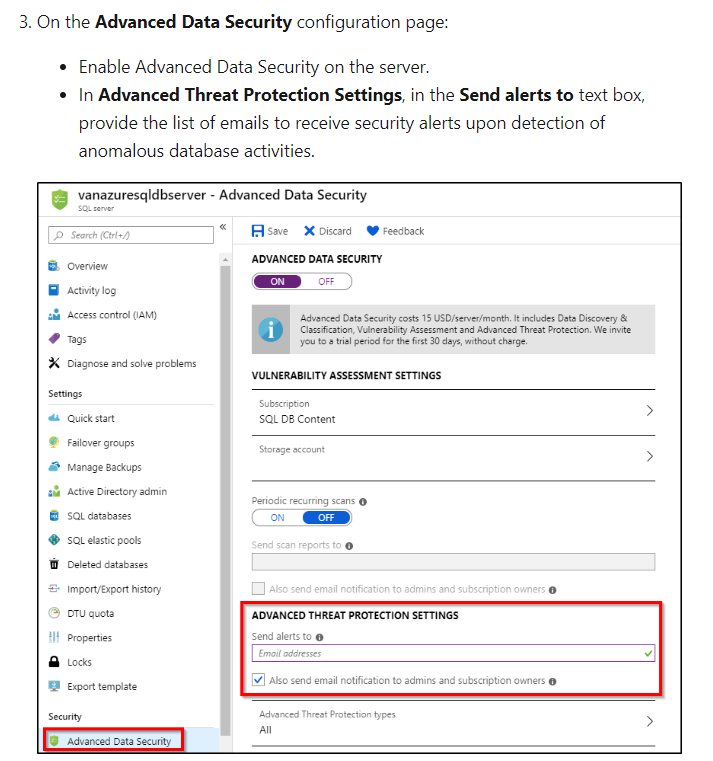
E.

**Configure the service to send email alerts to the IT security administrator.**

**Explanation:**

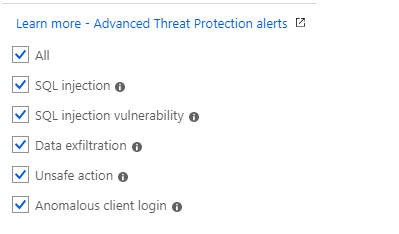
Answer – B, D and E

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



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>

### **Question 35**

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.

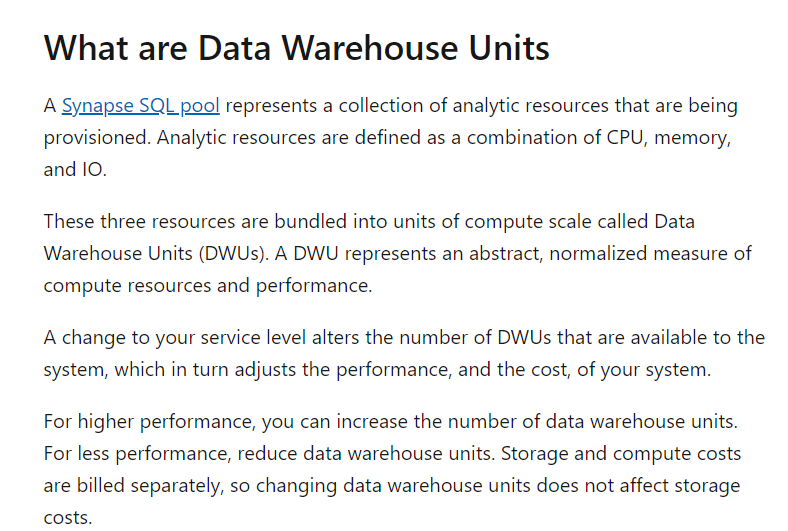
**Data IO percentage**

**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.



Option A and D are incorrect since DWU’s are used for allocated 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>

### **Question 36**

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.

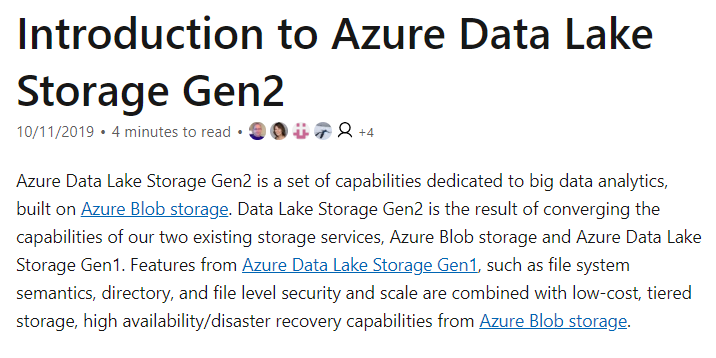
**Azure Data Lake Storage**

**Explanation:**

Answer – D

Azure Data Lake Storage fulfills 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.



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>

### **Question 37**

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.

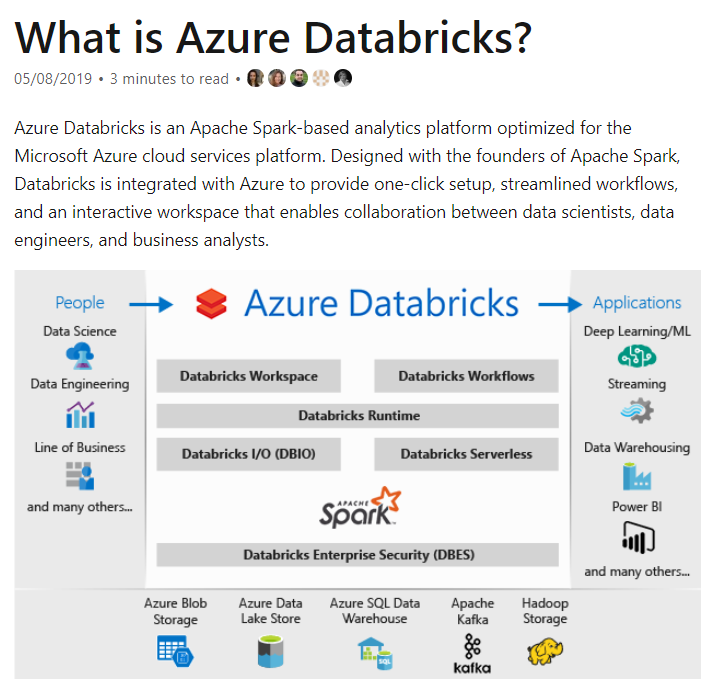
**HDInsight Interactive Query**

**Explanation:**

Answer – C

With Azure Databricks, you can set up 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.



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>

### **Question 38**

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 datastore**

* 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.

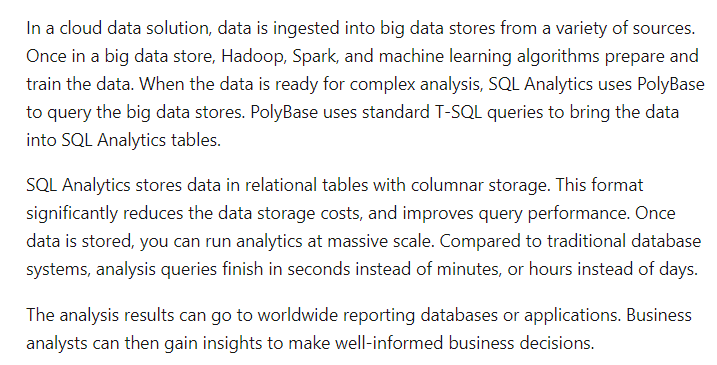
**Azure Cosmos DB**

**Explanation:**

Answer – B

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

The Microsoft documentation mentions the following.



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**

Domain :Implement data storage solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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>

### **Question 40**

Domain :Implement data storage solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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

]A.

**Eventual**

]B.

**Session**

]C.

**Strong**

]D.

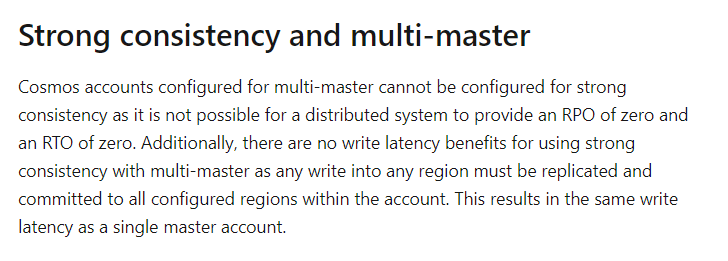
**Consistent Prefix**

**Explanation:**

Answer – A

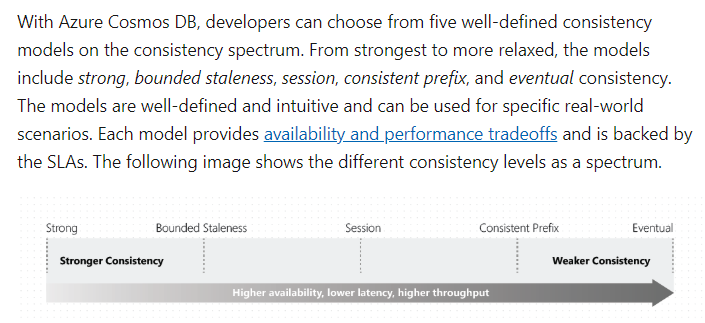
There is a requirement for data to be written to the closest data center for Cosmos DB. We need to ensure a multi-master setup for Cosmos DB wherein data can be written from multiple regions. For such accounts, we can’t set the consistency level to Strong.

The Microsoft documentation mentions the following.



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.



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>

### **Question 41**

Domain :Monitor and optimize data solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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.

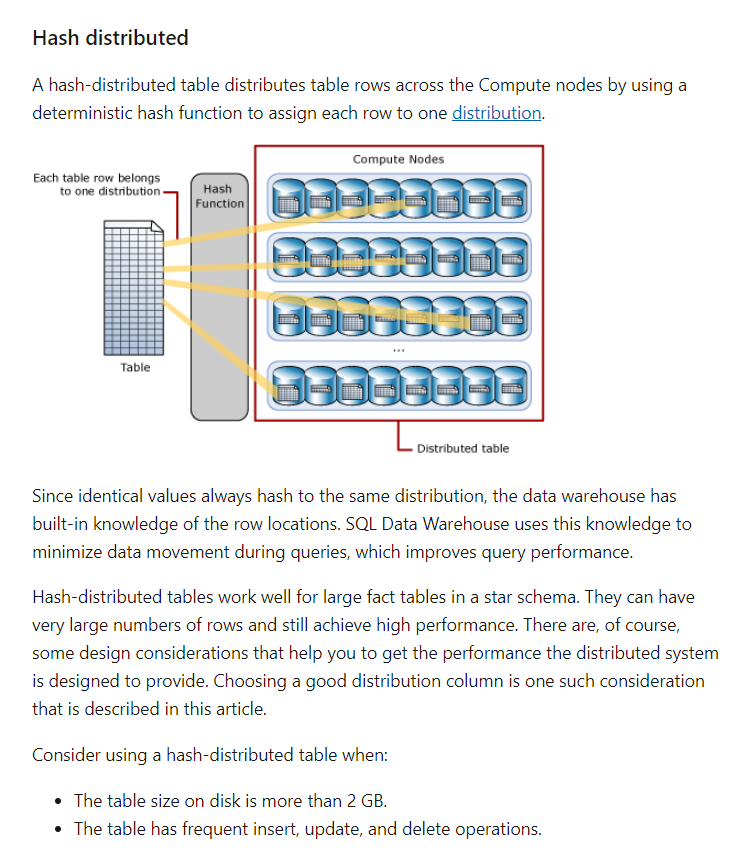
**Primary**

**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.



Since this is a 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>

### **Question 42**

Domain :Monitor and optimize data solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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.

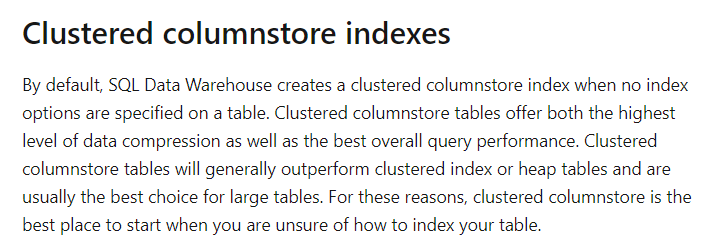
**Nonclustered**

**Explanation:**

Answer – B

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

The Microsoft documentation mentions the following.



Since this is a 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>

### **Question 43**

Domain :Implement data storage solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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

]A.

**Credit Card**

]B.

**Default**

]C.

**Email**

]D.

**Random number**

**Explanation:**

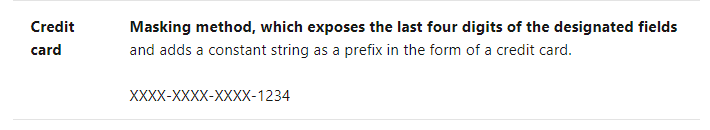
Answer – A

The case study mentions the following in 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.



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>

### **Question 44**

Domain :Implement data storage solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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

]A.

**Credit Card**

]B.

**Default**

]C.

**Email**

]D.

**Random number**

**Explanation:**

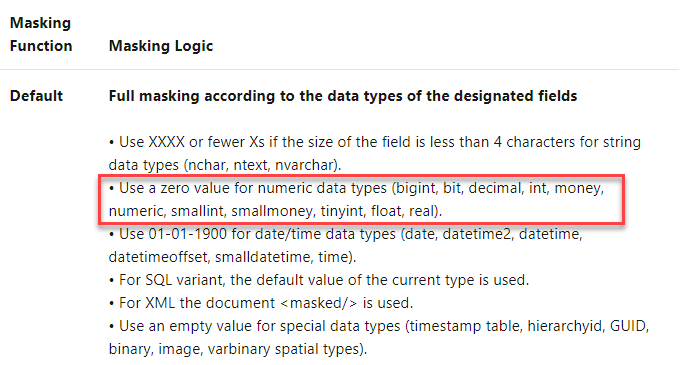
Answer – B

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.



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>

### **Question 45**

Domain :Manage and develop data processing

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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.

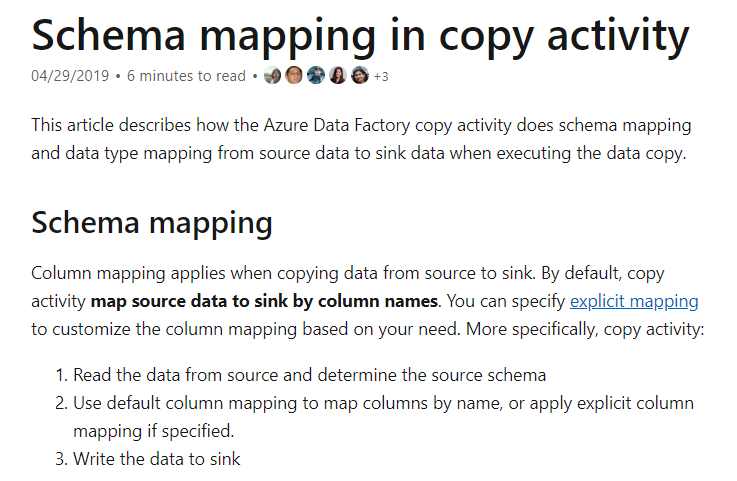
**A filter activity that needs to have a condition**

**Explanation:**

Answer – B

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.



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**

Domain :Monitor and optimize data solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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 that the copy activity uses a staged copy.**

]C.

**Ensure that the copy activity performs compression.**

]D.

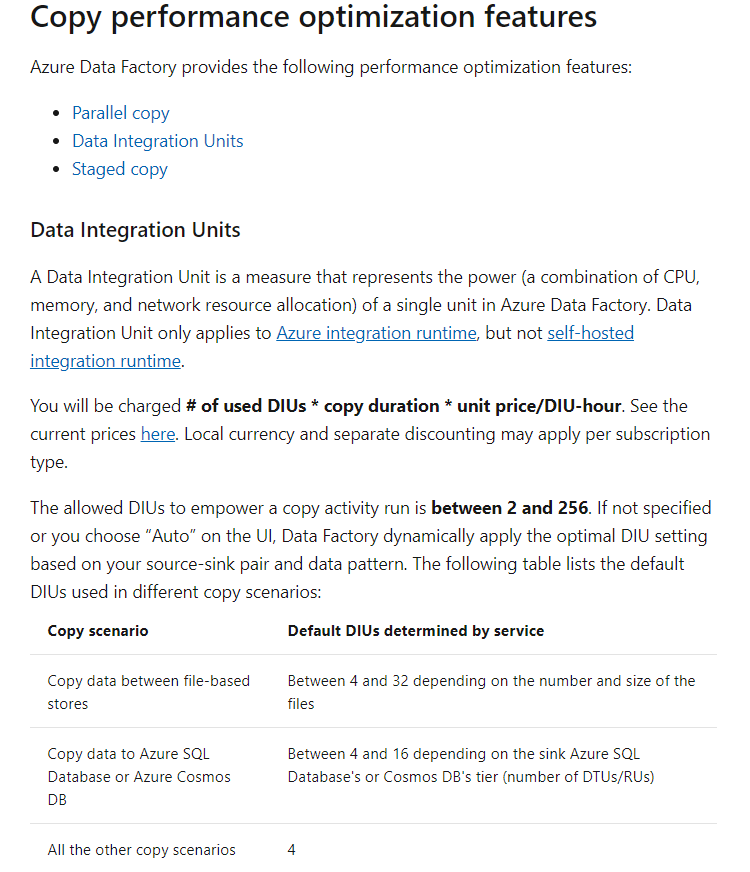
**Ensure to decrease the amount pf parallelism in the copy activities.**

**Explanation:**

Answer – A

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

The Microsoft documentation mentions the following.



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>

### **Question 47**

Domain :Monitor and optimize data solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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 for administrative overhead.”**

]A.

**sp\_update stored procedure**

]B.

**Using the Query store**

]C.

**Using the dbcc checkdb command**

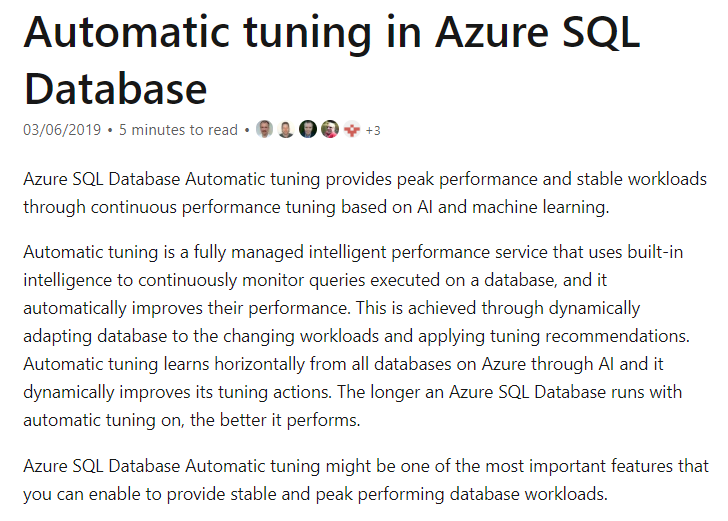
]D.

**Using automatic tuning**

**Explanation:**

Answer – D

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



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**

Domain :Monitor and optimize data solutions

[**View Case Study**](javascript:;)

**Overview**

Comps is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations.

**Current Environment**

The company currently has the following environment in place

* The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.
* A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.
* Another application named “Comp workflow” is then used to perform analytics on the telemetry data to look for improvements on the racing cars.
* The SQL Server 2017 database has a table named “cardata” which has around 1 TB of data. “Comp workflow” performs the required analytics on the data in this table. Large aggregations are performed on a column of the table.

**Proposed Environment**

The company now wants to move the environment to Azure. Below are the key requirements

* The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.
* The query performance for data in the Azure SQL database must be stable without the need of administrative overhead
* The data for analytics will be moved to an Azure SQL Data warehouse
* Transparent data encryption must be enabled for all data stores wherever possible
* An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configuration changes need to be made to the pipeline workflow.
* The telemetry data must be monitored for any sort of performance issues.
* The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.
* The data in the Azure SQL Server database must be protected via the following requirements
  + Only the last four digits of the values in the column carID must be shown
  + A zero value must be shown for all values in the column carWeight

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.

A.

**The number of requests**

B.

**The Session consistency**

C.

**The data and index storage consumed**

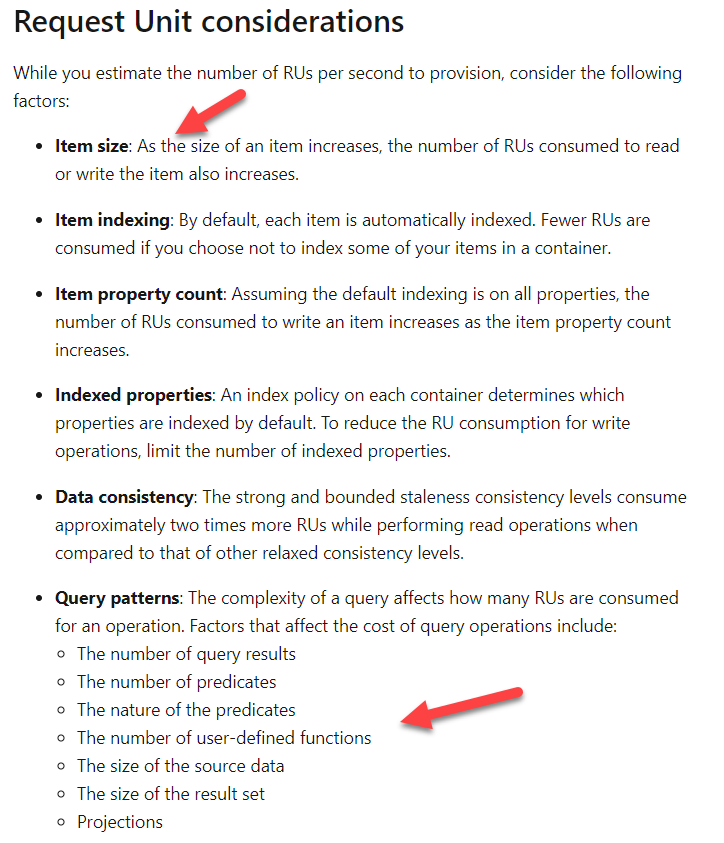
D.

**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.



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>

### **Question 49**

Domain :Manage and develop data processing

You have the following query defined in Azure Stream Analytics

WITH

step1 AS (SELECT \*

FROM compinput1

PARTITION BY OrderId

INTO 10),

step1 AS (SELECT \*

FROM compinput2

PARTITION BY OrderId

INTO 10)

SELECT \*

INTO compoutput

FROM step1

PARTITION BY OrderId

UNION step2

BY OrderId

Would the above query join two streams of partitioned data?

]A.**Yes**

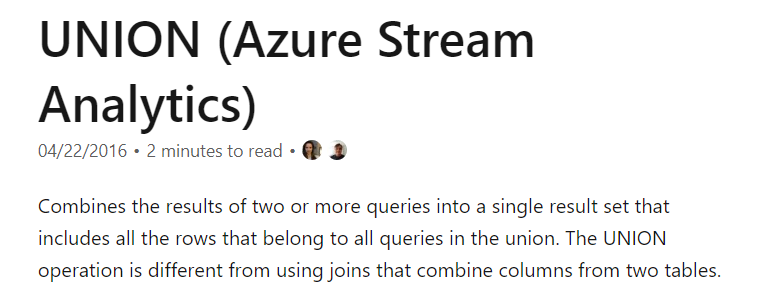
]B.**No**

**Explanation:**

Answer – B

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

The Microsoft documentation mentions the following.



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>

### **Question 50**

Domain :Manage and develop data processing

You have the following query defined in Azure Stream Analytics

WITH

step1 AS (SELECT \*

FROM compinput1

PARTITION BY OrderId

INTO 10),

step1 AS (SELECT \*

FROM compinput2

PARTITION BY OrderId

INTO 10)

SELECT \*

INTO compoutput

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**

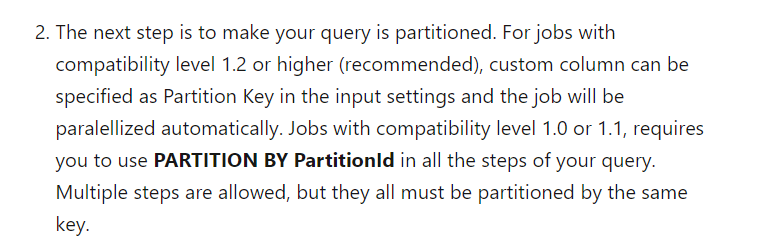
]B.**No**

**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.



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 51**

Domain :Manage and develop data processing

You have the following query defined in Azure Stream Analytics

WITH

step1 AS (SELECT \*

FROM compinput1

PARTITION BY OrderId

INTO 10),

step1 AS (SELECT \*

FROM compinput2

PARTITION BY OrderId

INTO 10)

SELECT \*

INTO compoutput

FROM step1

PARTITION BY OrderId

UNION step2

BY OrderId

Would the supply of 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**

Domain :Manage and develop data processing

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

* ClusterA – Here the cluster needs 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.

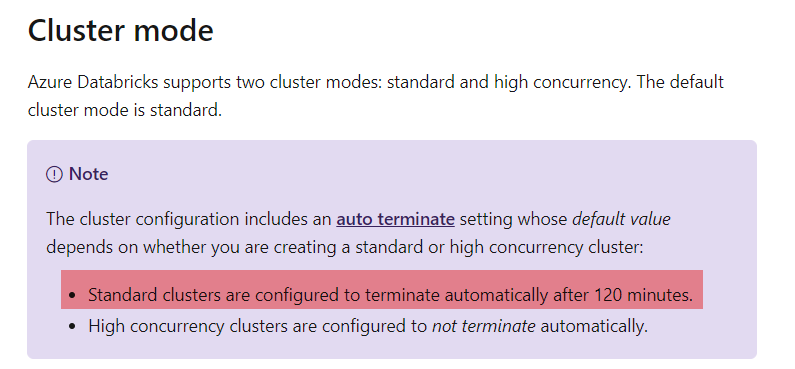
**High concurrency**

**Explanation:**

Answer – A

Here you have to set the cluster type as “Standard”.

The Microsoft documentation mentions the following.



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>

### **Question 53**

Domain :Manage and develop data processing

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

* ClusterA – Here the cluster needs 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.

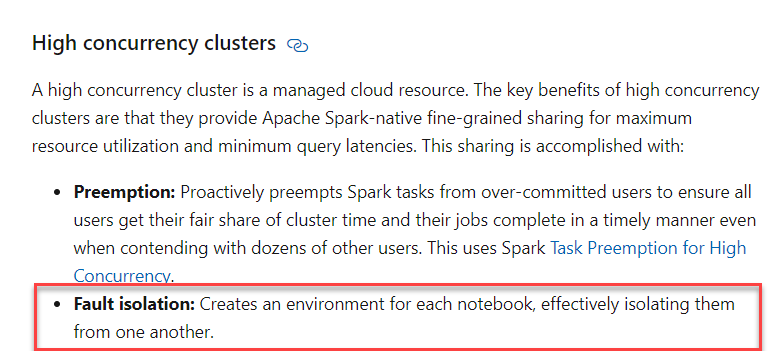
**High concurrency**

**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>

### **Question 54**

Domain :Manage and develop data processing

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

* ClusterA – Here the cluster needs 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.

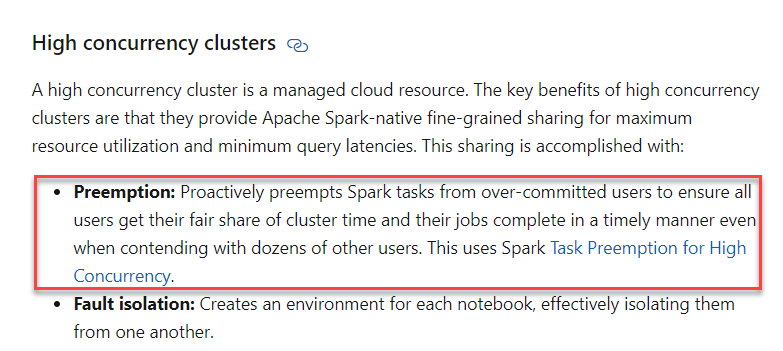
**High concurrency**

**Explanation:**

Answer – D

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

The Microsoft documentation mentions the following.



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>

### **Question 55**

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.

A.

**Azure Event Hubs**

B.

**Azure Storage**

C.

**Azure Cosmos DB**

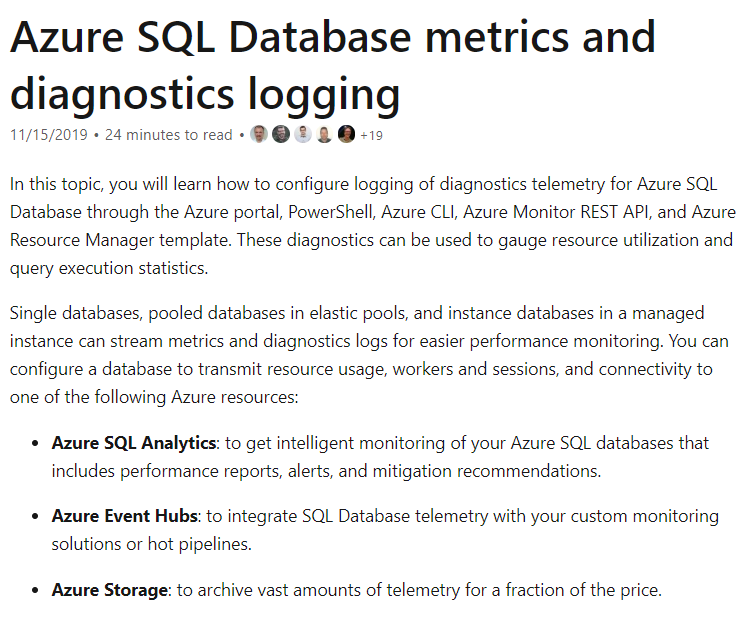
D.

**Azure SQL Data warehouse**

**Explanation:**

Answer – A and B

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



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>