

## Practice Test 1

### Question 1

Domain: Design and implement data storage

You are working in a cloud company and you have been assigned the responsibility of building an enterprise data lake on Azure and accomplish big data analytics. Which of the following Azure Service would you use in this scenario?

- A. Azure Files
- B. Azure Blobs right
- C. Azure Disks
- D. Azure Queues
- E. Azure Tables

### Explanation:

#### Correct Answer: B

Azure blobs allow storing and accessing the unstructured data at a massive scale in block blobs. Azure blobs are recommended to use:

- When you want your applications to support streaming and random-access scenarios.
- When you want to access application data from anywhere.
- When you want to develop an enterprise data lake on Azure and carry out big data analytics.
- Option A is incorrect. Azure Files offer fully managed cloud file shares that can be accessed from anywhere using the industry-standard Server Message Block (SMB) protocol. Azure files are not the right choice for the given scenario.
- Option B is correct. Azure Blobs is the best choice to be used in the given scenario.
- Option C is incorrect. Azure Disks help in persistently storing and accessing the data from an attached virtual hard disk. In the given scenario, using Azure Disks is not the right choice.
- Option D is incorrect. Azure Queues is the best choice for decoupling the application components and using asynchronous messaging to communicate among them.
- Option E is incorrect. Azure tables should be used to store flexible data. In the given scenario, using Azure Tables is not the right choice.

#### Reference:

To know more about core Azure Storage services, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction#example-scenarios>

Ask our Experts

[View Queries](#)

### Question 2

Domain: Design and implement data storage

You have created an external table named ExtTable in Azure Data Explorer. Now, a database user needs to query this external table. Which of the following function should he use to refer to this table?

- A. external\_table() right

- B. access\_table()
- C. refer\_table()
- D. Only the table administrator can query the table.

### Explanation:

#### Correct Answer: A

After an external table is defined, the function external\_table() should be used to refer to this table. Any database reader or user can query an external table.

- Option A is correct. external\_table() should be used to refer to the external table.
- Option B is incorrect. Access\_table() is not the valid function.
- Option C is incorrect. There is no function like refer\_table().
- Option D is incorrect. Any database reader or user can query an external table.

#### Reference:

To know more about how to query data in Azure Data Lake using Azure Data Explorer, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-explorer/data-lake-query-data>

Ask our Experts

[View Queries](#)

### Question 3

Domain: Design and implement data storage

You have been assigned the task of partitioning the FactOnlineSales table on the OrderDateKey column in the dedicated SQL pool. For this purpose, you decide to use the CREATE TABLE statement.

```

CREATE TABLE [dbo].[FactOnlineSales]
(
    [ProductKey]          int          NOT NULL
,   [OrderDateKey]       int          NOT NULL
,   [CustomerKey]        int          NOT NULL
,   [PromotionKey]       int          NOT NULL
,   [SalesOrderNumber]   nvarchar(20) NOT NULL
,   [OrderQuantity]      smallint     NOT NULL
,   [UnitPrice]          money        NOT NULL
,   [SalesAmount]        money        NOT NULL
)
WITH
( CLUSTERED COLUMNSTORE INDEX
,   ..... = HASH([ProductKey])
,   ..... ( | [OrderDateKey] RANGE RIGHT FOR VALUES
            (20000101,20010101,20020101
            ,20030101,20040101,20050101
            )
        )
)
;

```

Complete the statement by filling the blanks with the right words.

- A. Distribution and Partition<sup>right</sup>
- B. DistributionTable and PartitionTable
- C. Distribution and Collate
- D. Partition and Distribution

## Explanation:

### Correct Answer: A

DISTRIBUTION = HASH ( *distribution\_column\_name* ) is the distribution method that assigns every row to one distribution by hashing the value present in *distribution\_column\_name*. The right syntax to use partition method is PARTITION ( *partition\_column\_name* RANGE [ LEFT | RIGHT ] FOR VALUES ([*boundary\_value* [,...*n*]])).

- Option A is correct. Distribution and Partition are the right options to be used to complete the given Create Table statement.
- Option B is incorrect. The right syntax is to use only Distribution and Partition, not DistributionTable and PartitionTable.
- Option C is incorrect. The partition should be used instead of Collate.
- Option D is incorrect. Distribution and Partition are the right options to use.

### References:

To know more about partitioning the tables, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition>
- <https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7>

Ask our Experts

## Question 4

Domain: Design and implement data storage

When you create a temporal table in Azure SQL Database, it automatically creates a history table in the same database for capturing the historical records. Which of the following statements are true about the temporal table and history table? [Select all options that are applicable]

- A. A temporal table must have 1 primary key.**right**
- B. To create a temporal table, System Versioning needs to be set to On.**right**
- C. To create a temporal table, System Versioning needs to be set to Off.
- D. It is mandatory to mention the name of the history table when you create the temporal table.
- E. If you don't specify the name for the history table, the default naming convention is used for the history table.**right**
- F. You can specify the table constraints for the history table.

### Explanation:

#### Correct Answers: A, B and E

Here are some key points to note before creating the temporal table:

- A temporal table must have 1 primary key.
- The period for system time must be defined with appropriate valid from and to fields with *datetime2* datatype.
- Set System Versioning to ON.
- If you don't specify the name for the history table, the default naming convention is used for the history table.
- Other optional parameters such as data consistency check and retention period etc can be declared in the syntax; if needed.
- The history table is page compressed.
- history table can't have any table constraints.
- Option A is correct. A temporal table must have 1 primary key.
- Option B is correct. To create a temporal table, System Versioning needs to be set to On.
- Option C is incorrect. To create a temporal table, System Versioning needs to be set to On, not Off.
- Option D is incorrect. It is up to you to specify the name for the history table or not.
- Option E is correct. If you don't specify the name for the history table, the default naming convention is used for the history table.
- Option F is incorrect. The history table can't have any table constraints. Although you can create statistics or indexes to optimize the performance.

#### Reference:

To know more about temporal tables, please visit the below-given link:

- <https://visualbi.com/blogs/microsoft/azure/designing-slowly-changing-dimension-scd-azure-data-factory-using-sql-server-temporal-tables/>

Ask our Experts

## Question 5

Domain: Design and implement data storage

You are working on a Columnstore table. Although the columnstore indexes and tables are saved with the columnstore compression always, you are interested in further decreasing the columnstore data size. For this purpose, you decide to configure an add-on compression known as archival compression. Which of the following method would you use to compress the data by using the archival compression?

- A. COLUMNSTORE
- B. COLUMNSTORE\_ARCHIVE<sup>right</sup>
- C. COLUMNSTORE\_COMPRESS
- D. COLUMNSTORE\_ARCHIVECOMPRESS
- E. Archival compression can't be configured for the columnstore data.

### Explanation:

#### Correct Answer: B

COLUMNSTORE\_ARCHIVE data compression is used to compress columnstore data with the help of archival compression.

To configure archival compression, write **ALTER INDEX (Transact-SQL)** or **ALTER TABLE (Transact-SQL)** including the REBUILD option and DATA COMPRESSION = COLUMNSTORE\_ARCHIVE.

For example:

```
ALTER TABLE ColumnstoreTable1  
REBUILD PARTITION = 1 WITH (DATA_COMPRESSION = COLUMNSTORE_ARCHIVE) ;
```

- Option A is incorrect. COLUMNSTORE data compression is used to decompress the archival compression.
- Option B is correct. COLUMNSTORE\_ARCHIVE data compression is used to compress columnstore data using archival compression.
- Option C is incorrect. There is no valid method like COLUMNSTORE\_COMPRESS.
- Option D is incorrect. The right method is COLUMNSTORE\_ARCHIVE, not COLUMNSTORE\_ARCHIVECOMPRESS.
- Option E is incorrect. Archival compression can be configured for the columnstore data. It will further decrease the columnstore data size.

#### Reference:

To know more about Data Compression, please visit the below-given link:

<https://docs.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression?view=sql-server-ver15>

Ask our Experts

[View Queries](#)

### Question 6

Domain: Design and implement data storage

You are implementing an Azure Data Lake Gen2 storage account. You need to ensure that data will be accessible for write and read operations both even if an entire data center (zonal or non-zonal) becomes unavailable. Which kind of replication would you use for the storage account? (Choose the solution with minimum cost)

- A. Locally-redundant storage (LRS)
- B. Zone-redundant storage (ZRS)<sup>right</sup>
- C. Geo-redundant storage (GRS)

- D. Geo-zone-redundant storage (GZRS)

## Explanation:

### Correct Answer: B

Zone-redundant storage replicates the Azure Storage data in a synchronous manner across 3 Azure availability zones in the primary region. With Zone-redundant storage, the data remains accessible for write and read operations both even if a zone is not available.

The following table describes the durability and availability by outage scenario:

Outage scenario	LRS	ZRS	GRS/RA-GRS	GZRS/RA-GZRS
A node within a data center becomes unavailable	Yes	Yes	Yes	Yes
An entire data center (zonal or non-zonal) becomes unavailable	No	Yes	Yes <sup>1</sup>	Yes
A region-wide outage occurs in the primary region	No	No	Yes <sup>1</sup>	Yes <sup>1</sup>
Read access to the secondary region is available if the primary region becomes unavailable	No	No	Yes (with RA-GRS)	Yes (with RA-GZRS)

Account failover is needed for restoring the write availability if the primary region becomes unavailable.

- Option A is incorrect. LRS ensures availability only if a node within a data center becomes unavailable.
- Option B is correct. Zone-redundant storage replicates the Azure Storage data in a synchronous manner around 3 Azure availability zones within primary region.
- Option C is incorrect. GRS is not a cost-effective method. ZRS will be a more suitable option in the given scenario.
- Option D is incorrect. GZRS will also ensure availability but it is not the redundant method with minimum cost. ZRS will achieve the goal in the given scenario.

### Reference:

To know more about Azure Storage Redundancy, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy>

Ask our Experts

[View Queries](#)

## Question 7

Domain: Design and implement data storage

You are working on Azure Data Lake Store Gen1. Suddenly, you realize the need to know the schema of the external data. Which of the following plug-in would you use to know the external data schema?

- A. ipv4\_lookup
- B. mysql\_request
- C. Pivot
- D. Narrow
- E. infer\_storage\_schema right

## Explanation:

**Correct Answer: E**

`infer_storage_schema` is the plug-in that helps infer the schema based on the external file contents; when the external data schema is unknown.

- Option A is incorrect. The `ipv4_lookup` plugin checks for an IPv4 value in a lookup table and returns the matched rows.
- Option B is incorrect. The `mysql_request` plugin transfers a SQL query to a MySQL Server network endpoint and returns the 1st row set in the result.
- Option C is incorrect. Pivot plug-in is used to rotate a table by changing the unique values from 1 column in the input table into a number of different columns in the output table and perform aggregations wherever needed on any remaining column values that are desired in the final output.
- Option D is incorrect. This plug-in is used to unpivot a wide table into a table with only three columns.
- Option E is correct. `infer_storage_schema` plug-in can be used to infer the schema of external data and return it as a CSL schema string.

**References:**

To know more about the external tables and plug-in, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-explorer/kusto/management/external-tables-azurestorage-azuredatalake>
- <https://docs.microsoft.com/en-us/azure/data-explorer/kusto/query/inferstorageschemaplugin>

Ask our Experts

[View Queries](#)

**Question 8**

Domain: Design and implement data storage

You need to create data semantic models in SQL Server Analysis Services. There are some recommended best practices for data modeling that one should follow. Which of the following practices are considered as the best practices that you would mind while creating data semantic models? (Select three options)

- A. Never create a dimension model snowflake as well as a star while ingesting data from various sources.
- B. Create a dimension model snowflake or/and star, even if you need to ingest data from various sources.**right**
- C. Only include the integer surrogate keys or value encoding in the model and exclude all the natural keys from the dimension tables.**right**
- D. Only include the natural keys in the model and exclude the integer surrogate keys or value encoding.
- E. Decrease cardinality to reduce the uniqueness of the values and allow much better compression.**right**
- F. Increase the cardinality to reduce the uniqueness of the values and allow much better compression.

**Explanation:****Correct Answers: B, C and E**

The below image describes the best practices that can be considered while creating data semantic models in, Azure Analysis Services, SQL Server Analysis Services, or Power BI:

- Create a dimension model star and/or snowflake, even if you are ingesting data from different sources.
- Ensure that you create integer surrogate keys on dimension tables. Natural keys are not best practice and can cause issues if you need to change them at a later date. Natural keys are generally strings, so larger in size and can perform poorly when joining to other tables. The key point in regards to performance with tabular models is that natural keys are not optimal for compression. The process with natural keys is that they are:
  - Encoded, hash/dictionary encoding.
  - Foreign keys encoded on the fact table relating to the dimension table, again hash/dictionary encoding.
  - Build the relationships.
- This has an impact on performance and reduces the available memory for data as a proportion, which will be needed for the dictionary encoding.
- Only bring into the model the integer surrogate keys or value encoding and exclude any natural keys from the dimension tables.
- Only bring into the model the foreign keys or integer surrogate keys on the fact table from the dimension tables.
- Only bring columns into your model that are required for analysis, this may be excluding columns that are not needed or filter on data to only bring the data in that is being analyzed.
- Reduce cardinality so that the values uniqueness can be reduced, allowing for much greater compression.
- Add a date dimension into your model.
- Ideally, we should run calculations at the compute layer if possible.
- Option A is incorrect. It is recommended to create a dimension model even if you need to ingest data from various sources.
- Option B is correct. You need to create a dimension model snowflake or/and star, even if you need to ingest data from various sources.
- Option C is correct. The best practices ask to only include the integer surrogate keys or value encoding in the model and exclude all the natural keys from the dimension tables.
- Option D is incorrect. The best practices ask to only include the integer surrogate keys or value encoding in the model and exclude all the natural keys from the dimension tables.
- Option E is correct. You should decrease the cardinality to reduce the uniqueness of the values and allow much better compression.
- Option F is incorrect. Decreasing the cardinality will help in reducing the uniqueness of the values and allowing much better compression.

#### Reference:

To know more about Data models, please visit the below-given link:

- <https://azure.microsoft.com/en-in/blog/data-models-within-azure-analysis-services-and-power-bi/>

Ask our Experts

[View Queries](#)

## Question 9

Domain: Design and develop data processing

SQLite differs from commercial relational database systems in terms of features as it does not support a number of features supported by commercial relational database systems. Which of the following statement(s) is/are true about SQLite?

- A. SQLite also assigns a type to the columns like most relational database systems.



- B. SQLite supports all – LEFT OUTERJOIN, RIGHT OUTERJOIN and FULL OUTERJOIN.
- C. SQLite supports no type of OUTERJOIN.
- D. SQLite supports only LEFT OUTERJOIN, not RIGHT OUTERJOIN or FULL OUTERJOIN.right
- E. You can't create views in SQLite

### Explanation:

#### Correct Answer: D

SQLite puts some restrictions while creating the SQL statements. SQLite utilizes dynamic typing for values. It supports only LEFT OUTERJOIN, not RIGHT or FULL OUTERJOIN. SQLite allows creating views but here views are read only.

- Option A is incorrect. SQLite utilizes dynamic typing for values rather than allotting a type to the columns.
- Option B is incorrect. SQLite supports only LEFT OUTERJOIN.
- Option C is incorrect. SQLite supports LEFT OUTERJOIN.
- Option D is correct. SQLite supports only LEFT OUTERJOIN, not RIGHT OUTERJOIN or FULL OUTERJOIN.
- Option E is incorrect. SQLite allows creating views but here views are read-only. It does not allow executing INSERT, DELETE, or UPDATE statements on a view.

#### Reference:

To know more about SQL Transformation, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/apply-sql-transformation>

Ask our Experts

[View Queries](#)

### Question 10

Domain: Design and develop data processing

While configuring Normalize data, you decide to apply the Zscore mathematical function from the Transformation method dropdown list to apply on the chosen columns. From the given options, choose the right formula that is used to convert all values to a z-score.

- A.  $z = x - \text{mean}(x) / \text{stdev}(x)$ right
- B.  $z = x - \min(x) / [\max(x) - \min(x)]$
- C.  $z = 1 / 1 + \exp(-x)$
- D.  $z = \text{Lognormal.CDF}(x; \mu, \sigma)$

### Explanation:

#### Correct Answer: A

Zscore function converts all values for the chosen columns to Z-score. The formula used to transform the values in a column is as given below:

$$z = \frac{x - \text{mean}(x)}{\text{stdev}(x)}$$

- Option A is correct. the given formula is the right formula for the Zscore function.
- Option B is incorrect. The given formula is for the MinMax function.
- Option C is incorrect. The given formula is for Logistic mathematical function.

- Option D is incorrect. The given formula is for the LogNormal function.

**Reference:**

To know more about Normalizing Data Module, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/normalize-data>

Ask our Experts

[View Queries](#)

## Question 11

Domain: Design and develop data processing

You need to decide on the technology choice that your team should use for batch processing in Azure. The requirements demand the technology to meet the following capabilities:

- Autoscaling
- In-memory caching of data
- Query from external relational stores
- Support for firewall

Which of the following techniques would you choose?

- A. Azure Data Lake Analytics
- B. Azure Synapse Analytics
- C. Azure HDInsight with Spark<sup>right</sup>
- D. Azure Databricks

### Explanation:

**Correct Answer: C**

HDInsight is a managed Hadoop service that can be used for deploying and managing the Hadoop clusters in Azure. HDInsight can be used with **Spark**, **MapReduce**, **Hive**, or **Hive LLAP** for batch processing. HDInsight meets all the capabilities mentioned in the scenario.

- Option A is incorrect. Azure Data Lake Analytics does not support Autoscaling and In-memory caching of data.
- Option B is incorrect. Azure Synapse Analytics does not support Autoscaling and Query from external relational stores.
- Option C is correct. HDInsight with Spark supports all the given capabilities: Autoscaling, In-memory caching of data, Query from the external relational store, and support for the firewall.
- Option D is incorrect. Azure Databricks does not support a firewall.

**Reference:**

To know more about batch processing technologies in Azure, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing>

Ask our Experts

[View Queries](#)

## Question 12

Domain: Design and develop data processing

You need to set your default Azure region. You know that the region can be set using the following command:

z config set defaults.location=<REGION>

Here, you need to replace the <REGION> with the name of the region that is available for your subscription and you want to set. Which of the following command would you run in Cloud Shell to check the regions that are available from your Azure subscription?

- A. az account locations-list
- B. az account list-locations<sup>right</sup>
- C. az account regions-list
- D. az account list-regions
- E. az storage account create

### Explanation:

#### Correct Answer: B

*az account list-locations* command in Cloud Shell lists the available regions from your Azure subscription.

Azure CLI

```
az account list-locations \
  --query "[].{Name: name, DisplayName: displayName}" \
  --output table
```

- Option A is incorrect. The right command is *az account list-locations*, not *az account locations-list*.
- Option B is correct. the right command to be used to know the available regions is *az account list-locations*.
- Option C is incorrect. There is no command like *az account regions-list* in Azure.
- Option D is incorrect. *az account list-regions* is not the right command to find the regions.
- Option E is incorrect. *az storage account create* is used to create a new storage account.

#### Reference:

To know more about building a data pipeline, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/devops/pipelines/apps/cd/azure/build-data-pipeline?view=azure-devops>

Ask our Experts

[View Queries](#)

### Question 13

Domain: Design and develop data processing

While working on one of your company's projects, your teammate wants to check the options for input to an Azure Stream Analytics task that needs high throughput and low latencies. He is confused about the input that he should use in this case. He approaches you and asks you to help him. Which of the following Azure product would you suggest to him?

- A. Azure Table Storage
- B. Azure Blob Storage
- C. Azure Queue Storage

- D. Azure Event Hubs<sup>right</sup>
- E. Azure IoT Hub

## Explanation:

### Correct Answer: D

Azure Event Hub is considered as a highly scalable event ingestion service that can take and process over million events within a second. You can transform and store the data that is sent to the event hubs with the help of storage/batching adapters or real-time analytics provider. Event Hubs are known for consuming the data streams from applications with high throughput and low latencies.

- Option A is incorrect. Azure table storage is a NoSQL store that is used for schemaless storage of structured data.
- Option B is incorrect. Azure Blob Storage should be used when you desire your application to support streaming and random-access scenarios.
- Option C is incorrect. Azure Queue Storage is used to allow asynchronous message queueing among application components.
- Option D is correct. In the given scenario, You should suggest Azure Event Hubs to your teammate as Event Hubs are the primary choice for consuming the data streams from applications with high throughput and low latencies.
- Option E is incorrect. Azure IoT Hub offers a cloud-hosted solution back end for connecting any device virtually.

### Reference:

To know more about Azure Event Hubs, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-about>

Ask our Experts

[View Queries](#)

## Question 14

Domain: Design and develop data processing

You work in TCT company and you are having the responsibility to manage the jobs in Azure. You decide to add a new job. While specifying the job constraints, you set maxWallClockTime property to value 30 minutes. What does it mean?

- A. The job can be in a ready state for a maximum of 30 minutes.
- B. The job can be in an inactive state for a maximum of 30 minutes.
- C. The job can be in the active or running state for a maximum of 30 minutes.<sup>right</sup>
- D. The job will automatically start in 30 minutes.
- E. None of these

## Explanation:

### Correct Answer: C

The maxWallClockTime property is the optional property that sets the maximum amount of time a job can run for. As the time exceeds or the job doesn't complete in the specified time, the job is terminated. If you don't set this property, then the job has no time limit.

- Option A is incorrect. The maxWallClockTime property specifies the maximum time for which a job can be in a running state.
- Option B is incorrect. maxWallClockTime property does not specify the time limit for the inactive state of any job.
- Option C is correct. Setting maxWallClockTime property to 30 minutes specifies that the job can be in the active or running state for a maximum of 30 minutes

- Option D is incorrect. maxWallClockTime property does not specify the time after which the job will automatically start.
- Option E is incorrect. Setting maxWallClockTime property to 30 minutes specifies that the job can be in the active or running state for a maximum of 30 minutes

#### References:

To know more about jobs and scheduling the jobs, please visit the below-given links:

- <https://docs.microsoft.com/en-us/azure/batch/batch-job-task-error-checking>
- <https://docs.microsoft.com/en-us/azure/batch/batch-job-schedule>

Ask our Experts

[View Queries](#)

## Question 15

Domain: Design and implement data security

You work in Azure Synapse Analytics dedicated SQL pool that has a table titled Pilots. Now you want to restrict the user access in such a way that users in 'IndianAnalyst' role can see only the pilots from India. Which of the following would you add to the solution?

- A. Table partitions
- B. Encryption
- C. Column-Level security
- D. Row-level security right
- E. Data Masking

### Explanation:

#### Correct Answer: D

Row-level security is applicable on databases to allow fine-grained access to the rows in a database table for restricted control upon who could access which type of data.

- Option A is incorrect. Table partitions are generally used to group similar data.
- Option B is incorrect. Encryption is used for security purposes.
- Option C is incorrect. Column level security is used to restrict data access at the column level. In the given scenario, we need to restrict access at the row level.
- Option D is correct. In this scenario, we need to restrict access on a row basis, i.e only for the pilots from India, there Row-level security is the right solution.
- Option E is incorrect. Sensitive data exposure can be limited by masking it to unauthorized users using SQL Database dynamic data masking.

#### References:

To know more about Row-level security, please visit the below-given links:

- <https://azure.microsoft.com/en-in/resources/videos/row-level-security-in-azure-sql-database/>
- <https://techcommunity.microsoft.com/t5/azure-synapse-analytics/how-to-implement-row-level-security-in-serverless-sql-pools/ba-p/2354759>

Ask our Experts

[View Queries](#)

## Question 16

Domain: Design and implement data security

You have been assigned the *Storage Blob Data Contributor* role at a container level. Here are two statements regarding this:

1. You have been granted write, read, and delete access to all blobs in that container.
2. You can view a blob within Azure portal.

Which of the above-given statements are true?

- A. Only a **right**
- B. Only b
- C. Both a and b
- D. None

### Explanation:

#### Correct Answer: A

If the Storage Blob Data Contributor role is assigned to a user at the container level, titled sample-container, and the user is granted write, read, and delete permission to the blobs present in that specific container. But Storage Blob Data Contributor role itself does not provide enough permission to navigate to reach the blob through the Azure portal for view purposes. Extra permission is needed in order to perform a navigation through the Azure portal and see the additional resources which are available or visible there.

- Option A is correct. Storage Blob Data Contributor role at container level grants the write, read, and delete permission for all the blobs in that container.
- Option B is incorrect. With only the Storage Blob Data Contributor role, you can't perform navigation to the blobs via Azure portal. Therefore, statement a is correct while statement b is incorrect.
- Option C is incorrect. Statement b is incorrect.
- Option D is incorrect. Statement a is correct while statement b is incorrect.

#### Reference:

To know more about assigning Azure roles for data access, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/blobs/assign-azure-role-data-access?tabs=portal>

Ask our Experts

[View Queries](#)

## Question 17

Domain: Design and implement data security

You have created a *data* DataFrame. But before issuing SQL Queries, you decide to save your DataFrame as a temporary view. Which of the following DataFrame method will help you in creating the temporary view?

- A. createTempView
- B. createOrReplaceTempView **right**
- C. ReplaceTempView
- D. createTempViewForDataFrame
- E. CreateorReplaceDFTempView

### Explanation:

#### Correct Answer: B

Before issuing SQL queries, DataFrame needs to be saved as a table or temporary view. createOrReplaceTempView function is used to create the temporary view.

Python

```
# Register table so it is accessible via SQL Context
%python
data.createOrReplaceTempView("data_geo")
```

- Option A is incorrect. There is no DataFrame method like createTempView.
- Option B is correct. createOrReplaceTempView function is used to create the temporary view.
- Option C is incorrect. There is no such method as ReplaceTempView.
- Option D is incorrect. There is no such method as createTempViewForDataFrame.
- Option E is incorrect. There is no DataFrame method like CreateorReplaceDFTempView.

**Reference:**

To know more about DataFrames, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/databricks/getting-started/spark/dataframes>

Ask our Experts

[View Queries](#)

## Question 18

Domain: Monitor and optimize data storage and data processing

Log Analytics Workspaces store the data collected by Azure Monitor Logs. Log query is the query that is used to retrieve the data from a Log Analytics workspace. From the given list of languages, select the language in which these Log queries are written?

- A. Structured Query Language (SQL)
- B. PL/SQL (Procedural Language for SQL)
- C. Kusto Query Language (KQL) **right**
- D. Python
- E. PostgreSQL

## Explanation:

**Correct Answer: C**

Log queries are written in KQL (**Kusto Query Language**). Azure Data Explorer also uses the same query language. Log queries can be written in Log Analytics to analyze the results interactively, use them in alert rules for proactive notification of issues, or add their results in dashboards or workbooks.

- Option A is incorrect. KQL, not SQL is the language used in log queries.
- Option B is incorrect. Log queries are written in KQL, not PL/SQL.
- Option C is correct. Log queries are written in KQL(**Kusto Query Language**).
- Option D is incorrect. Log queries are written in KQL, not Python.
- Option E is incorrect. KQL (**Kusto Query Language**) is the language used for Log Queries.

**Reference:**

To know more about Azure Monitor Logs, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/azure-monitor/logs/data-platform-logs>

Ask our Experts

[View Queries](#)

## Question 19

Domain: Monitor and optimize data storage and data processing0

After checking the monitor tab in the Azure Synapse Studio environment, you realize that you can improve the performance of the run. Now, you decide to use bucketed tables to improve the performance. Which of the following are the recommended practices to consider while using bucketed tables? (Select all options that are applicable)

- A. Avoid the use of SortMerge join whenever possible<sup>right</sup>
- B. Prefer the use of SortMerge join as much as you can
- C. Never consider the most selective joins
- D. Start with the most selective joins<sup>right</sup>
- E. Move joins that increase the number of rows after aggregations whenever possible.<sup>right</sup>
- F. The order of various types of joins matters when it comes to the resource consumption<sup>right</sup>

### Explanation:

**Correct Answers: A, D, E and F**

While using bucketed tables, you need to deal with Merge join. A correctly pre-sorted and pre-partitioned dataset will skip the costly sort phase from a SortMerge join.

The order of joins does matter, especially in more complex queries. Start with the most selective joins. You should also consider moving the joins that increase the number of rows after aggregations, whenever possible.

- Option A is correct. While using bucketed tables, you should avoid the use of SortMerge join whenever possible.
- Option B is incorrect. You should avoid using expensive SortMerge join while using bucketed tables.
- Option C is incorrect. Instead of preferring the use of SortMerge join as much as you can, you should start with the most selective joins.
- Option D is correct. You should start with the most selective joins to improve the performance.
- Option E is correct. To increase the performance using bucketed tables, you should move joins that increase the number of rows after aggregations whenever possible
- Option F is correct. The order of various types of joins matters when it comes to the resource consumption

### Reference:

To know more about Apache Spark Performance, please visit the below-given performance:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-performance>

Ask our Experts

[View Queries](#)

## Question 20

Domain: Monitor and optimize data storage and data processing0

While copying a Data Factory pipeline, you get a "DelimitedTextMoreColumnsThanDefined" error. What might be the possible cause behind this error? (Choose the best possible option)

- A. You have reached the capacity limit of integration runtime.
- B. The folder you are copying has files with different schemas.<sup>right</sup>
- C. Browser cache issue.
- D. You have invoked REST API in a Web activity.

### Explanation:

**Correct Answer: B**



If a folder you are copying consists of the files with different schemas, like different delimiters, quote char settings, variable number of columns, or some data issue, the Data Factory pipeline may return the following error:

```
Operation on target Copy_sks failed: Failure happened on 'Sink' side. ErrorCode=DelimitedTextMoreColumnsThanDefined, 'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException, Message=Error found when processing 'Csv/Tsv Format Text' source '0_2020_11_09_11_43_32.avro' with row number 53: found more columns than expected column count 27., Source=Microsoft.DataTransfer.Common, '
```

- Option A is incorrect. When you reach the capacity limit of the integration runtime, you might be running a huge amount of data flow through the same integration runtime at the same time that makes the pipeline run get fail.
- Option B is correct. "DelimitedTextMoreColumnsThanDefined" error is generally returned when the folder you are copying has the files with different schemas.
- Option C is incorrect. Browser cache issue might result in a situation where you cancel a pipeline run, but pipeline monitoring often still demonstrates the progress status.
- Option D is incorrect. The option does not state the right cause for the given issue.

**Reference:**

- <https://docs.microsoft.com/en-us/azure/data-factory/pipeline-trigger-troubleshoot-guide>

Ask our Experts

[View Queries](#)

## Question 21

Domain: Design and implement data storage

You work in the PWO company as a data administrator. You need to store the user data for the web application and you decide to use Azure Table storage (which is the ideal storage in the given case). Do you need to define a schema for the table?

- A. Yes
- B. No right

### Explanation:

**Correct Answer: B**

Azure Table storage allows storing structured NoSQL data in the cloud environment, offering a key/attribute store with a schemaless design. Therefore, Azure Table Storage needs no schema. Data can be adapted as the requirements of the application evolve.

**Reference:**

To know more about core Azure Storage services, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction#example-scenarios>

Ask our Experts

[View Queries](#)

## Question 22

Domain: Design and implement data storage

In Azure Synapse, the data is distributed in three different ways: Round-Robin, Hashing, and Replication. Which distribution type to be used depends upon the scenario and the requirements. Which of the following statement(s) is/are true about these distribution types? (Select all that are applicable)

- A. Choose Hash distribution when you can't identify a single key for distributing your data.

- B. Choose Round Robin distribution when you can't identify a single key for distributing your data.**right**
- C. Choose Round Robin distribution if the table is having temporary data.**right**
- D. Choose Replicated distribution if the table is having temporary data.
- E. Choose Hash distribution if the table is having temporary data.
- F. Replicated distribution is ideal for dimension tables that are very frequently joined with other big tables.**right**
- G. Hash distribution is ideal for dimension tables that are very frequently joined with other big tables.

## Explanation:

### Correct Answers: B, C and F

The following table describes which distribution to use and not to use in which scenario.

Server Load Balancer

Overview

ALB New

CLB (Formerly Known as SLB)

Quota Management

Resources

Listener Name ?

Advanced

Hide

\* Scheduling Algorithm

Weighted Round-Robin (WRR)

Round-Robin (RR)

Consistent Hash (CH)

Enable Session Persistence ?

Enable Access Control ?

Enable Peak Bandwidth Limit ?

Idle Timeout ?

900

Seconds

Valid Values: 10 to 900

Next

Cancel

h resources, services, and docs (G+)

Dashboard >

**Logs**  
Demo

New Query 1\* x +

Feedback

Run Time range : Last 24 hours Save Share + New alert

Tables Queries

search x

Filter Group by: Query type

Collapse all

**Favorites**

You can add favorites by clicking on the ☆ icon

**Example queries**

- Run a case-sensitive search
- Search a table for a specific term
- Search a term through all logs

```

1
2 // Search in multiple tables
3 // Search both Syslog and Event tables for the term "login".
4 search in (Syslog, Event) "login"
5 | where TimeGenerated > ago(1h) // return records from the last hour
6
7
8

```

Results Chart

- Option A is incorrect. Round Robin Distribution, not Hash distribution, is ideal when you can't identify a single key for distributing your data.
- Option B is correct. Round Robin Distribution is recommended when you can't identify a single key for distributing your data.
- Option C is correct. You should choose the Round Robin distribution if the table is having temporary data.
- Option D is incorrect. Choosing a Replicated distribution is not the ideal choice if the table is having temporary data.
- Option E is incorrect. Choosing Hash distribution is not the ideal choice if the table is having temporary data.
- Option F is correct. Replicated distribution is ideal for dimension tables that are very frequently joined with other big tables.
- Option G is incorrect. Replicated distribution, not Hash distribution, is ideal for dimension tables that are very frequently joined with other big tables.

#### Reference:

To know more about the Right distribution strategy, please visit the below-given link:

- <https://rajanieshkaushikk.com/2020/09/09/how-to-choose-right-data-distribution-strategy-for-azure-synapse/>

Ask our Experts

[View Queries](#)

## Question 23

Domain: Design and implement data storage

While working on the project, you realize that the delta table is not correct. One of your friends suggests deleting the whole directory of the table and creating a new table on the same path. Would you follow the suggested solution?

- A. Yes

- B. No<sup>right</sup>

## Explanation:

### Correct Answer: B

Deleting the whole directory of a Delta table and creating a new table on the same path is not a recommended solution as:

- A directory may consist of very large files and deleting the directory can consume days or even hours. Therefore, it is not an efficient solution.
- All the content of the deleted files is lost and if by mistake you delete a wrong file, it is very hard to recover it.
- Deleting the directory is not atomic. While table deletion is in progress, a concurrent query reading the table can view a partial table or even fail.

### Reference:

To know more about best practices while using Delta Lake, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/databricks/delta/best-practices>

Ask our Experts

[View Queries](#)

## Question 24

Domain: Design and implement data storage

Spark offers several options for storing the data in the managed tables. From the below-given options, choose the formats that are/are not supported by Spark. (Select all that are applicable).

- A. TEXT
- B. CSV
- C. XML<sup>right</sup>
- D. JSON
- E. JDBC
- F. PARQUET
- G. LIBSVM
- H. MPV4<sup>right</sup>
- I. BINARY<sup>right</sup>

## Explanation:

### Correct Answers: C, H and I

Spark offers several options for storing the data in the managed tables, such as TEXT, JSON, CSV, JDBC, ORC, PARQUET, DELTA, LIBSVM and HIVE. These files are typically saved in the warehouse directory where data for the managed table is stored.

- Option A is incorrect. Spark supports TEXT format.
- Option B is incorrect. Spark supports CSV format.
- Option C is correct. XML format is not supported by Spark.
- Option D is incorrect. Spark supports JSON format.
- Option E is incorrect. JDBC format is supported by Spark.
- Option F is incorrect. PARQUET format is supported by Spark.
- Option G is incorrect. Spark supports LIBSVM format.

- Option H is correct. Spark does not support MPV4 format for data storage.
- Option I is correct. Spark does not support a Binary format for data storage.

**Reference:**

To know more about Spark created tables, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/metadata/table>

Ask our Experts

[View Queries](#)

## Question 25

Domain: Design and implement data storage

The partition specifies how Azure storage load balances entities, messages, and blobs across servers to achieve the traffic requirements of these objects. Which of the following represents the partition key for a blob?

- A. Account name + Table Name + blob name
- B. Account name + container name + blob name **right**
- C. Account name + Queue name + blob name
- D. Account Name + Table Name + Partition Key
- E. Account Name + Queue Name

### Explanation:

**Correct Answer: B**

For a blob, the partition key consists of account name + container name + blob name. Data is partitioned into ranges using these partition keys and these ranges are load balanced throughout the system.

- Option A is incorrect. For a blob, the partition key includes account name + container name + blob name.
- Option B is correct. For a blob, the partition key includes account name + container name + blob name.
- Option C is incorrect. Account name + Queue name + blob name is not the right partition key for a blob.
- Option D is incorrect. For an entity in a table, the partition key includes the table name and the partition key.
- Option E is incorrect. For a message in a queue, the queue name is the partition key itself.

**Reference:**

To know more about Partitioning Azure Blob Storage, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/best-practices/data-partitioning-strategies>

Ask our Experts

[View Queries](#)

## Question 26

Domain: Design and implement data storage

It might be possible to switch a storage account from one type of replication to another type using Azure portal, Azure CLI or PowerShell (just by updating the replication settings) or you might need to perform a manual migration, all depending upon scenarios. From the below-given options, choose the scenarios where you must perform manual migration.

- A. When the data needs to be migrated into a ZRS storage account that is located in a region other than the source account.
- B. If the storage account is a premium block blob or page blob account.

- C. When data needs to be migrated from ZRS to GRS, RA-GRS or LRS.
- D. When the storage account involves data in the archive tier.
- E. All the above<sup>right</sup>

## Explanation:

### Correct Answer: E

You need to perform the manual migration in the below cases:

- When the data needs to be migrated into a ZRS storage account that is located in a region other than the source account.
- If the storage account is a premium block blob or page blob account.
- When data needs to be migrated from ZRS to GRS, RA-GRS or LRS.
- When the storage account has data in the archive tier.

Also, the below table shows how one can switch from one replication type to another:

Basic	Developer	Business	Enterprise
<b>Included</b>	<b>\$19.99</b>	<b>Starting at \$100</b>	<b>Starting at \$8000</b>
Pricing Details: Included with all Alibaba Cloud services.	\$19.99 will be deducted from your account immediately after purchasing the Developer support plan. You can choose whether to make an automatic renewal.	Business support plan charges will be assessed based on your usage of cloud products. You can choose whether to make an automatic renewal.	Enterprise support plan charges will be assessed based on your usage of cloud products. Please consult your business manager for details.
<a href="#">Get It Free</a>	<a href="#">Purchase Now</a>	<a href="#">Contact Us</a>	<a href="#">Contact Us</a>

- Option A is incorrect. Not only this, all given options are correct.
- Option B is incorrect. Although it is true that you need to perform manual migration if the storage account is a premium block blob or page blob account, all other statements are also true.
- Option C is incorrect. From the above table, it is clear that you must perform manual migration When data needs to be migrated from ZRS to GRS, RA-GRS or LRS. But all other given options are also true.
- Option D is incorrect. It is true that you need to perform manual migration When the storage account involves data in the archive tier but all other statements are also true.
- Option E is correct. For all the given scenarios, you must perform manual migration.

### Reference:

To know more about replication, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/common/redundancy-migration?tabs=portal>

Ask our Experts

[View Queries](#)

## Question 27

Domain: Design and implement data storage

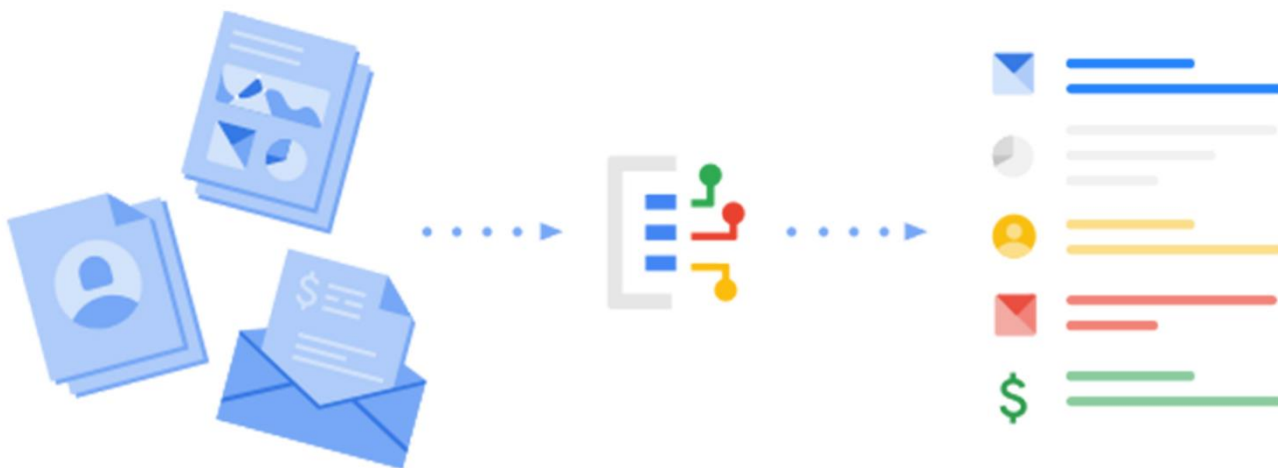
There are two types of external tables – Hadoop and Native external tables that can be used to read and export data depending upon the type and format of the external data source. Which of the following statements are true about these external tables? (Select all that are applicable)

- A. Hadoop external tables are only available in dedicated SQL pools, but not in serverless SQL pools.right
- B. Hadoop external tables are only available in serverless SQL pools, but not in dedicated SQL pools.
- C. Native external tables are available in serverless SQL pools but are not available in dedicated Synapse SQL pools.
- D. Native external tables are available in serverless SQL pools, but in Synapse SQL pools, they are only in gated preview.right
- E. The files having the name started with a period (.) or an underline (\_) are skipped by both Hadoop as well as native external tables while reading the data.right
- F. Only the files having the name started with underline (\_), not period(.) are skipped by both external tables.
- G. Hadoop considers all files and skips no file irrespective of its name.

## Explanation:

**Correct Answers: A, D and E**

The below table highlights the major differences between Hadoop external table and the Native external table:



- Option A is correct. Hadoop external tables are available only in dedicated SQL pools, but not in serverless SQL pools.
- Option B is incorrect. Hadoop external tables are only available in dedicated SQL pools, but not in serverless SQL pools.
- Option C is incorrect. Native external tables are available in serverless SQL pools, and also available in dedicated Synapse SQL pools but only in gated preview.
- Option D is correct. Native external tables are available in serverless SQL pools, but in Synapse SQL pools, they are only in gated preview.
- Option E is correct. The files having the name started with a period (.) or an underline (\_) are skipped by both Hadoop as well as native external tables while reading the data.
- Option F is incorrect. Both types of files having the name started with a period (.) or an underline (\_) are skipped by both types of external tables.
- Option G is incorrect. Hadoop external tables skip the files having the name started with a period (.) or an underline (\_).

**Reference:**

To know more about using external tables with Synapse SQL, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables?tabs=hadoop>

Ask our Experts

[View Queries](#)

## Question 28

Domain: Design and implement data storage

You are copying data from Parquet format in Azure Data Factory using Self-hosted Integration Runtime and you get the following error:

An error occurred when invoking java, message: `java.lang.OutOfMemoryError:Java heap space`

How will you resolve this error?

- A. Remove the environment variable `_JAVA_OPTIONS` and rerun the pipeline
- B. Add an environment variable `_JAVA_OPTIONS` and rerun the pipeline **right**
- C. Restart the machine
- D. You can't copy the data from Parquet format in Azure Data Factory

### Explanation:

#### Correct Answer: B

While copying the data from or to Parquet format using Self-hosted Integration Runtime(IR), you might get the error stating "An error occurred when invoking java, message: **java.lang.OutOfMemoryError:Java heap space**". This error can be resolved by adding an environment variable `_JAVA_OPTIONS` in the machine hosting the Self-hosted Integrated Runtime to adjust the max/min heap size for JVM to allow such copy and rerun the pipeline.



Search (Ctrl+/)

Refresh

\* Subscription

Turn on diagnostics to collect the following data.

- SQLSecurityAuditEvents
- DmsWorkers
- ExecRequests
- RequestSteps
- SqlRequests
- Waits
- AllMetrics

Diagnostic settings

- Option A is incorrect. You need to add the environment variable `_JAVA_OPTIONS` and rerun the pipeline to resolve the issue.
- Option B is correct. You need to add the environment variable `_JAVA_OPTIONS` and rerun the pipeline to resolve the issue.
- Option C is incorrect. Only restarting the machine won't solve the issue.
- Option D is incorrect. Azure Data Factory allows copying data from/to Parquet format.

#### Reference:

To know more about Parquet format in Azure Data Factory, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/format-parquet>

Ask our Experts

[View Queries](#)

## Question 29

Domain: Design and develop data processing

You need to upload a Python script to an Azure Blob storage account. Below given are the steps (not necessarily in the right sequence) you need to follow to upload the script. Arrange these steps in the right sequence.

1. Create WordCount\_Spark.py python file
2. Create a folder named spark.
3. Under the folder Spark, make a subfolder named as the script.

4. In Blob storage, create a container named adftutorial if it doesn't exist.
5. Upload the created WordCount\_Spark.py python file to the subfolder script.
6. Replace <storageAccountName> with the name of the Azure storage account where the file is to upload. Finally, save the file.

The correct sequence of steps is:

- A. a-b-c-d-e-f
- B. a-c-b-d-e-f
- C. a-b-c-e-d-f
- D. a-f-d-b-c-e<sup>right</sup>
- E. b-a-c-d-e-f

### Explanation:

#### Correct Answer: D

The below figure describes the steps involved in uploading the python script to the Blob Storage Account:

The screenshot shows the 'Server Load Balancer' configuration page in the Azure Portal. The left sidebar contains navigation links: Overview, ALB New, CLB (Formerly Known as SLB), Quota Management, and Resources. The main content area shows the 'Listener Name' field with a placeholder URL. Below this, the 'Advanced' tab is selected, and the 'Scheduling Algorithm' section is highlighted with a red box. This section contains three buttons: 'Weighted Round-Robin (WRR)' (selected), 'Round-Robin (RR)', and 'Consistent Hash (CH)'. Below the scheduling algorithm, there are three toggle switches: 'Enable Session Persistence', 'Enable Access Control', and 'Enable Peak Bandwidth Limit', all of which are currently disabled. An 'Idle Timeout' field is set to 900 seconds. The 'Next' button is visible at the bottom.

- Option A is incorrect. The option does not state the right sequence of the steps.
- Option B is incorrect. The given sequence of steps is not correct.
- Option C is incorrect. The given sequence is incorrect.
- Option D is correct. The given sequence of steps is correct.
- Option E is incorrect. The given sequence is not the correct one.

### Reference:

To know more about transforming data in the cloud, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/tutorial-transform-data-spark-portal>

Ask our Experts

[View Queries](#)

## Question 30

Domain: Design and develop data processing

Data Cleansing in Data Quality Services (DQS) includes a 2-step process for data cleansing: *computer-assisted* and *interactive cleansing*. Depending upon the computer-assisted cleansing process, during the interactive cleansing, DQS gives the data steward with information that is needed to make a decision about modifying or changing the data. For this purpose, DQS classifies the data into 5 tabs. From the below-given option, choose the tab that is not among these tabs.

- A. Invalid
- B. Valid<sup>right</sup>
- C. Suggested
- D. New
- E. Correct
- F. Corrected

### Explanation:

#### Correct Answer: B

During the interactive cleansing, Data Quality Services (DQS) classifies the data in these 5 tabs: Suggested, New, Invalid, Corrected, and Correct.

- Option A is incorrect. Invalid tab has the values that were specified as invalid in the domain in the knowledge base or values that failed reference data or a domain rule.
- Option B is correct. There is no such tab with the name valid.
- Option C is incorrect. The suggested tab consists of the values having a confidence level greater than the *auto-suggestion threshold* value but lesser than the *auto-correction threshold* value for which Data Quality Services (DQS) found suggestions.
- Option D is incorrect. The new tab consists of the valid values for which Data Quality Services (DQS) doesn't have sufficient information (suggestion), and hence can't be mapped to any other tab.
- Option E is incorrect. The correct tab is for the values which were found correct.
- Option F is incorrect. The corrected tab is for the values that are corrected by Data Quality Services (DQS) during the automated cleansing.

#### Reference:

To know more about data cleansing, please visit the below-given link:

- <https://docs.microsoft.com/en-us/sql/data-quality-services/data-cleansing?view=sql-server-ver15>

Ask our Experts

[View Queries](#)

## Question 31

Domain: Design and develop data processing

You have a dataset internet\_sales that represents the online sale of the products. Your manager asks you to divide this dataset by date i.e. in two datasets one having the sales before or equal to 03/31/2021 and one after that.

Which of the following split modes will you use? (Note: the dataset is having a column named date with format mmddyyyy)

- A. Split Rows
- B. Split Columns
- C. Recommender Split
- D. Regular Expression Split
- E. Relative Expression Split<sup>right</sup>

### Explanation:

#### Correct Answer: E

Relative Expression Split option is used wherever you need to apply a condition to a number column. This number can be a time/date field, a column representing dollar amount or age, or even a percentage. For example: The expression `"Date" > 03/31/2010` will select all the rows with the sale after 31st march.

- Option A is incorrect. This option is used for dividing the data into 2 parts. You can mention the percentage that how much data should be there in each split. If you don't mention it, by default data is split 50-50.
- Option B is incorrect. There is no such splitting mode as split columns.
- Option C is incorrect. The recommender Split option is used to prepare the data for a recommender system.
- Option D is incorrect. Regular Expression Split is used to divide the dataset by testing a single column for a value.
- Option E is correct. In the given scenario, as we need to apply a condition on a number column i.e Date, the Relative Expression Split option should be used.

#### References:

To know more about splitting a dataset, please visit the below-given links:

- <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data>
- <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/split-data-using-relative-expression>

Ask our Experts

[View Queries](#)

## Question 32

Domain: Design and develop data processing

Azure Data Lake Analytics is a recommended analytics job service for distributed processing of huge datasets saved in the Azure Data Lake Store. Which of the following languages is used by Data Lake Analytics for query processing?

- A. U-SQL<sup>right</sup>
- B. Scala
- C. Spark SQL
- D. Lake SQL
- E. Java

### Explanation:

Correct Answer: A

Azure Data Lake Analytics uses U-SQL language for query processing. U-SQL combines the declarative feature of SQL (Structured Query language) with the procedural extensibility of language C#, and takes the benefit of parallelism for enabling efficient data processing at a massive scale.

- Option A is correct. U-SQL is the query processing language that is used by Azure Data Lake Analytics.
- Option B is incorrect. Data Lake Analytics uses U-SQL, not Scala.
- Option C is incorrect. Spark SQL is the language used by the Azure Databricks.
- Option D is incorrect. There is no such language as Lake SQL.
- Option E is incorrect. Data Lake Analytics uses U-SQL, not Scala for query processing.

**References:**

To know more about batch processing, please visit the below-given links:

- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing>
- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/batch-processing>

Ask our Experts

[View Queries](#)

### Question 33

Domain: Design and develop data processing

*Replace with median option* calculates the median value for the column, and replace each missing value in the column with that median value. From the below-given list, choose the Data types to which this *Replace with median option* can be applied.

- A. Integer<sup>right</sup>
- B. Double<sup>right</sup>
- C. Boolean
- D. Categorical
- E. All the above

**Explanation:**

**Correct Answers: A and B**

Replace with median can be applied only to the columns having Integer or double data types. It can't be applied to Boolean or categorical data types.

- Option A is correct. Replace with median can be applied to a column with Integer data type.
- Option B is correct. Replace with median can be applied to a column with the double data type.
- Option C is incorrect. Replace with median can't be applied to Boolean data type.
- Option D is incorrect. Replace with median can't be applied to categorical data types.
- Option E is incorrect. Replace with median can't be applied to Boolean or categorical data types.

**Reference:**

To know more about cleaning missing data, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/clean-missing-data>

Ask our Experts

[View Queries](#)

### Question 34

Domain: Design and develop data processing

Which of the following services would you use as an event consumer to create the complex linked visualizations and analyze the aggregated data in near-real-time? (Choose the best possible option)

- A. Power BI<sup>right</sup>
- B. Event Hubs
- C. Azure cosmos DB
- D. Azure Blob Storage

### Explanation:

#### Correct Answer: A

Power BI provides a platform to visualize and analyze the aggregated data in near-real-time. Azure Stream Analytics can target Power BI as an output destination. Processed data is passed into Power BI to enable near-real-time dashboard updates.

- Option A is correct. It is Power BI that should be used as an event consumer to create the complex linked visualizations and analyze the aggregated data in near-real-time.
- Option B is incorrect. Azure Event Hubs is a large data streaming platform and event ingestion service, that can be used to feed events from event producers into Azure Stream Analytics.
- Option C is incorrect. Azure Cosmos DB is a fully managed NoSQL database for contemporary application development. It can be utilized for storing the output of data stream processing in Azure Stream Analytics but doesn't provide the facility to create the visualizations or dashboard.
- Option D is incorrect. Azure Blob storage offers an ingestion point for data streaming in an event processing solution that utilizes static data as a source.

#### Reference:

To know more about event processing, please visit the below-given link:

- <https://docs.microsoft.com/en-us/learn/modules/introduction-to-data-streaming/3-understand-event-processing>

Ask our Experts

[View Queries](#)

### Question 35

Domain: Design and implement data security

You need to design an enterprise data warehouse in Azure SQL Database with a table titled customers. You need to ensure that the customer supportive staff can identify the customers by matching the few characters of their email addresses but the full email addresses of the customers should not be visible to them. Which of the following would you include in the solution?

- A. Row-level security
- B. Encryption
- C. Column Level Security
- D. Dynamic Data Masking<sup>right</sup>
- E. Any of the above can be used

### Explanation:

#### Correct Answer: D

Dynamic data masking is helpful in preventing unauthorized access to sensitive data by empowering the clients to specify how much of the sensitive data to disclose with minimum impact on the application layer. In this policy-based security feature, the sensitive data is hidden in the output of a query over specified database fields, but there is no change in the data in the database.

For example: \*\*\*\*\*abc@gmail.com

- Option A is incorrect. Row-level security is used to enable the restricted access i.e who can access what type of data.
- Option B is incorrect. Encryption is not the right solution.
- Option C is incorrect. Column level security won't help in limiting the exposure of sensitive data.
- Option D is correct. In the given scenario, there is a need to use Dynamic data masking to limit the sensitive data exposure to non-privileged users.
- Option E is incorrect. Dynamic Data Masking is the right answer.

**Reference:**

To know more about dynamic data masking, please visit the below-given link:

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

Ask our Experts

[View Queries](#)

## Question 36

Domain: Design and implement data security

You work in a cloud-based company where you have been assigned a task to manage files and directories in Azure Data Lake Storage Gen2. You need to create a directory named first-directory. Which of the following cmdlet would you use in PowerShell?

- A. New-AzDataLakeGen2Item **right**
- B. NewDirectory-AzDataLakeGen2Item
- C. New-Dir-AzDataLakeGen2Item
- D. Create\_directory-AzDataLakeGen2Item

## Explanation:

**Correct Answer: A**

A directory reference can be created with cmdlet *New-AzDataLakeGen2Item* in PowerShell.

The following example shows how a directory *first-directory* can be added to the container.

```
$filesystemName = "myfirst-file-system"
```

```
$dirname = "first-directory/"
```

```
New-AzDataLakeGen2Item -Context $ctx -FileSystem $filesystemName -Path $dirname -Directory
```

- Option A is correct. New-AzDataLakeGen2Item cmdlet is used to create a directory.
- Option B is incorrect. There is no command like NewDirectory-AzDataLakeGen2Item.
- Option C is incorrect. There is no command like New-Dir-AzDataLakeGen2Item.
- Option D is incorrect. Cmdlet New-AzDataLakeGen2Item, not Create\_directory-AzDataLakeGen2Item is used to create a directory.

**Reference:**

To know more about managing files and directories using PowerShell, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-directory-file-acl-powershell#manage-access-control-lists-acls>

Ask our Experts

[View Queries](#)

## Question 37

Domain: Design and implement data security

Auditing for **Azure Synapse Analytics** and Azure SQL Database first tracks database events and then writes these events to an audit log under your Log Analytics workspace, Azure storage account, or Event Hubs. Which of the following statements are true about Auditing and Audit logs? (Select two options)

- A. Audit logs are written to Append Blobs in Blob storage on your Azure subscription<sup>right</sup>
- B. While using Azure AD Authentication, even the failed logins records appear in the SQL audit log.
- C. The format for Audit logs is .xml.
- D. You can use SQL Server Management Studio (SSMS) to open the audit logs.<sup>right</sup>
- E. If you want auditing on read-only replicas, you need to manually configure it.

### Explanation:

#### Correct Answers: A and D

Audit logs are written to Append Blobs in Blob storage on your Azure subscription. In Azure AD Authentication, the failed login records won't appear in SQL audit logs. The format of Audit logs is .xel and can be viewed using SQL Server Management Studio. Auditing on Read-Only Replicas is enabled automatically.

- Option A is correct. Audit logs are written to Append Blobs in Blob storage on your Azure subscription.
- Option B is incorrect. In Azure AD Authentication, the failed login records won't appear in SQL audit logs. If you want to see the failed login audit records, go to the Azure Active Directory portal, which logs the specifics of these events.
- Option C is incorrect. Audit logs are in .xel format.
- Option D is correct. Audit logs can be opened using SSMS.
- Option E is incorrect. Auditing on Read-Only Replicas is enabled automatically.

#### Reference:

To know more about Auditing, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/azure-sql/database/auditing-overview#setup-auditing>

Ask our Experts

[View Queries](#)

### Question 38

Domain: Monitor and optimize data storage and data processing

Azure Data Factory allows debugging and troubleshooting the pipelines using Azure PowerShell and Azure portal. You decide to use PowerShell to debug an error. As an important step of the process, you need to retrieve details about the activity run for the slice. Which of the following cmdlet would you use to retrieve the details?

- A. Get-AzDataFactorySlice
- B. Get-AzDataFactoryRun<sup>right</sup>
- C. Set-AzDataFactorySliceStatus
- D. Get-AzDataFactoryActivityRun
- E. Set-AzDataFactoryActivityRun

### Explanation:

#### Correct Answer: B

**Get-AzDataFactoryRun** cmdlet is used to retrieve the details about the activity run for the slice. The syntax for the command is:



Billing Method Subscription Pay-As-You-Go ?

---

Dedicated Block Storage The dedicated block storage cluster feature is not enabled in the current region. Try another region.

Cluster

---

Storage Enhanced SSD (ESSD) ▲ 40 GiB 3800 IOPS Performance Level ? : PL1 (up to 50,000 IOPS per disk) ▼ [Create from Snapshot](#)

Ultra Disk

Enhanced SSD (ESSD)

Standard SSD

Learn how to [create a subscription disk](#).

Quantity — 1 + Unit

- Option A is incorrect. Get-AzDataFactorySlice cmdlet is used to view the slices and their status.
- Option B is correct. Get-AzDataFactoryRun cmdlet is used to retrieve the details about the activity run for the slice.
- Option C is incorrect. Set-AzDataFactorySliceStatus cmdlet is used in Azure Data Factory to set the status of slices for a dataset.
- Option D is incorrect. There is no command like Get-AzDataFactoryActivityRun in Azure Data Factory.
- Option E is incorrect. There is no command like Set-AzDataFactoryActivityRun in Azure Data Factory.

#### Reference:

To know more about monitoring and managing Azure Data Factory pipelines, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-monitor-manage-pipelines>

Ask our Experts

[View Queries](#)

## Question 39

Domain: Monitor and optimize data storage and data processing

Automatic tuning continuously monitors the queries executed on a database using its in-built intelligence and enhances the performance. Which of the following Automatic tuning options are used by Azure SQL database to improve the database performance?

- A. Create Index
- B. Drop Index
- C. FORCE LAST GOOD PLAN
- D. All the above right

### Explanation:

#### Correct Answer: D

Create Index, Drop Index and FORCE LAST GOOD PLAN are three automatic tuning options available in Azure SQL Managed Instance and Azure SQL Database. The following table describes these automatic tuning options:



- Option A is incorrect. Create Index, Drop Index, and FORCE LAST GOOD PLAN are three automatic tuning options available in Azure SQL Database.
- Option B is incorrect. Not only Drop Index, but Create Index and FORCE LAST GOOD PLAN automatic tuning options are also available in Azure SQL Database.
- Option C is incorrect. Not only FORCE LAST GOOD PLAN, but Create Index and Drop Index automatic tuning options are also available in Azure SQL Database.
- Option D is correct. All the given are the automatic tuning options available in Azure SQL Database.

**Reference:**

To know more about Automatic tuning in the Azure SQL database, please visit the below-given link:

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

Ask our Experts

[View Queries](#)

## Question 40

Domain: Monitor and optimize data storage and data processing

You need to diagnose your slowly running cluster and you decide to regenerate the error state on another cluster. As a part of the process, you have performed below two steps:

- Gathering data about the issue.
- Validating the HDInsight cluster environment

Which of the following is the right next step that you would follow to achieve the goal?

- A. Reviewing the environment stack and versions

- B. Checking the configuration settings.
- C. Checking the health of your cluster.**right**
- D. Examining the cluster log files.
- E. Find K-mean for the clusters

### Explanation:

#### Correct Answer: C

Reproducing the error state on another cluster typically involves the following steps:

- Gathering the data regarding the issue.
- Validating the HDInsight cluster environment.
- Viewing the health of your cluster.
- Reviewing the environment stack & versions.
- Examining the cluster log files.
- Checking configuration settings.
- Reproducing the failure on another cluster.
- Option A is incorrect. Reviewing the environment stack and versions is one of the steps from the process but it is not the right next step to perform.
- Option B is incorrect. Checking the configuration settings is one of the steps from the process but it is not the right next step to perform.
- Option C is correct. viewing or checking the health of your cluster is the right next step to perform.
- Option D is incorrect. Examining the cluster log files is not the right next step in the process.
- Option E is incorrect. Reproducing the error state does not involve such a step.

#### Reference:

To know more about troubleshooting the clusters, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-troubleshoot-failed-cluster>

Ask our Experts

[View Queries](#)

### Question 41

Domain: Design and implement data storage

One of your friends has an application that is supposed to run on Windows and needs access to a mapped drive. He asks you about the Azure service that he should use. Which of the following Azure services would you recommend to him?

- A. Azure Blob Storage
- B. Azure Table Storage
- C. Azure Files**right**
- D. Azure Queues
- E. Azure Disk

### Explanation:

**Correct Answer: C**

Azure Files offer fully managed cloud file shares that can be accessed from anywhere using the industry-standard Server Message Block (SMB) protocol.

Azure file shares can be used with windows either by mounting it or accessing it through its UNC path.

- Option A is incorrect. Azure blobs are used to store the unstructured data and access it at a massive scale in block blobs.
- Option B is incorrect. Azure tables can be used to store structured NoSQL data in the cloud.
- Option C is correct. Azure file shares can be used with windows either by mounting it or accessing it through its UNC path.
- Option D is incorrect. Azure Queues allow asynchronous message queueing among various application components.
- Option E is incorrect. Azure disk allows persistently storing and accessing the data from an attached virtual hard disk.

**References:**

To know more about Azure Storage Devices, please visit the below-given links:

- <https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction#example-scenarios>
- <https://docs.microsoft.com/en-us/azure/storage/files/storage-how-to-use-files-windows>

Ask our Experts

[View Queries](#)

**Question 42**

Domain: Design and implement data storage

You need to design an Azure SQL Database based on elastic pools. You are planning to create a table Employee having the data about the employees. For each record, there is an EmployeeID. You need to implement a data partitioning strategy based on the values in EmployeeID. You discuss with your friend and he suggests you to separate the data into Employee Departments via vertical partitioning. Will it help you in achieving the goal?

- A. Yes
- B. No right

**Explanation:****Correct Answer: B**

Using vertical partitioning in the given scenario won't help in achieving the goal. Vertical partitioning is implemented for cross-database queries. As we need to partition the data based on EmployeeID, here horizontal partitioning should be used. Horizontal partitioning divides the data into partitions. Each partition is known as a shard and holds a subset of the specific data.

**Reference:**

To know more about partitioning, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/best-practices/data-partitioning#designing-partitions-for-query-performance>

Ask our Experts

[View Queries](#)

**Question 43**

Domain: Design and implement data storage

You need to change the schema for the temporal table you created earlier. The below is the command that is used to change the schema or drop the temporal table. Fill in the blanks with suitable words or values in the command.

..... TABLE dbo.CustTemporal SET(SYSTEM\_VERSIONING = .....)

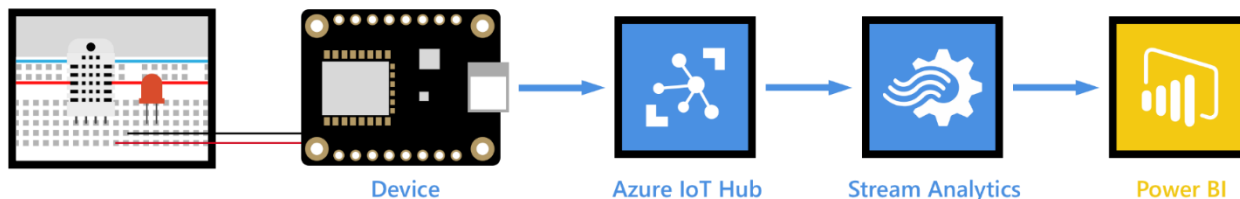
- A. MODIFY, ON
- B. MODIFY, OFF
- C. MODIFY, NULL
- D. ALTER, OFF<sup>right</sup>
- E. ALTER, ON
- F. ALTER, NULL

### Explanation:

#### Correct Answer: D

To drop the temporal table or change the Schema, first, you must set System Versioning to OFF.

The below is an example of the command:



- Option A is incorrect. ALTER, not MODIFY, is used to change the schema. Also, System Versioning needs to be set to OFF.
- Option B is incorrect. ALTER, not MODIFY, is used to change the schema.
- Option C is incorrect. ALTER, not MODIFY, is used to change the schema. Also, System Versioning needs to be set to OFF.
- Option D is correct. ALTER and OFF are the right solution.
- Option E is incorrect. System versioning needs to be set to OFF, not ON.
- Option F is incorrect. System versioning needs to be set to OFF, not NULL.

#### Reference:

To know more about Temporal Tables, please visit the below-given link:

- <https://visualbi.com/blogs/microsoft/azure/designing-slowly-changing-dimension-scd-azure-data-factory-using-sql-server-temporal-tables/>

Ask our Experts

[View Queries](#)

### Question 44

Domain: Design and implement data storage

While working on Azure SQL Database, you want to use data compression to decrease the size of the database. From the given options, select the database objects on which you can configure row and page compression?

- A. A whole nonclustered index

- B. A whole indexed view
- C. A whole table that has been stored as a heap.
- D. A whole table that has been stored as a clustered index.
- E. All the above<sup>right</sup>

## Explanation:

### Correct Answer: E

Row and page compression can be configured on the following database objects:

- Option A is incorrect. You can configure row and page compression not only on a whole nonclustered index but all given database objects.
- Option B is incorrect. Row and page compression can be configured not only on a whole indexed view but all given database objects.
- Option C is incorrect. Row and page compression can be configured not only on A whole table that has been stored as a heap but also all given database objects.
- Option D is incorrect. Row and page compression can be configured not only on A whole table that has been stored as a clustered index but also all given database objects.
- Option E is correct. Row and page compression can be configured for all the given database objects.

### Reference:

To know more about Data Compression, please visit the below-given link:

- <https://docs.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression?view=sql-server-ver15>

Ask our Experts

[View Queries](#)

## Question 45

Domain: Design and implement data storage

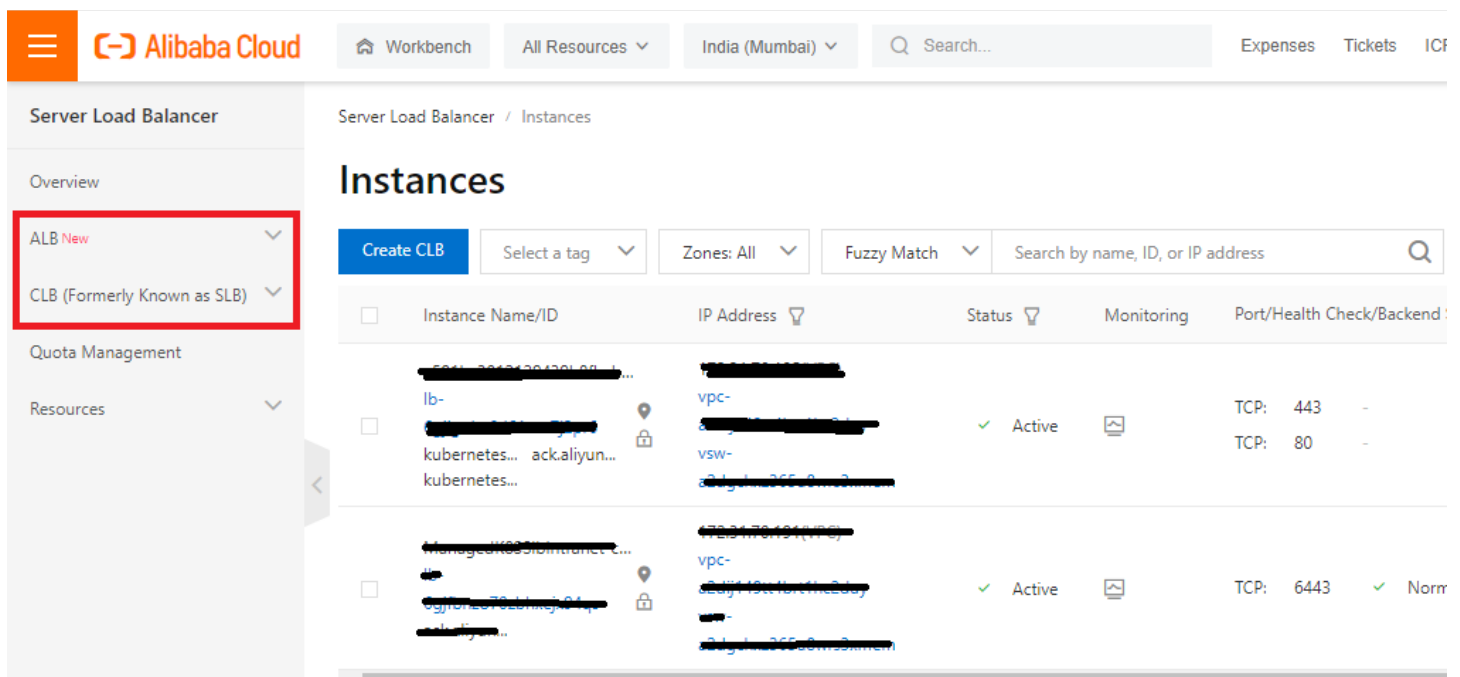
You have folders and files in Azure Blob Storage for the Azure Synapse. You develop a native external table "EtTable1" with LOCATION='/webdata/'. If you query EtTable1 with Azure Synapse Analytics Serverless SQL POOL, choose from the following files that will be returned?

- A. Only mydata.txt<sup>right</sup>
- B. Only mydata2.txt
- C. Only mydata3.txt
- D. mydata.txt, mydata2.txt, and mydata3.txt
- E. Only \_hidden.txt
- F. All of these i.e mydata.txt, mydata2.txt, mydata3.txt and \_hidden.txt

### Explanation:

#### Correct Answer: A

You have folders and files in Azure Blob Storage for the Azure Synapse workspace just like the below diagram:



If you want to return subfolders from Native external tables, you need to mention /\*\* at the end of the path. In the given case scenario, LOCATION='/webdata/' will make the query to return only mydata.txt, not mydata2.txt and mydata3.txt as these both files exist in the subfolders. On the other hand, A Hadoop table returns all files within any subfolder.

Both Native and external tables skip the files with the names beginning with a period (.) or an underline (\_). Therefore, it won't return \_hidden.txt as well.

- Option A is correct. In the given Scenario, the only mydata.txt will be returned.
- Option B is incorrect. mydata2.txt won't be returned as it is present in the subfolder and the location does not involve /\*\* at the end.
- Option C is incorrect. mydata3.txt won't be returned as it is present in the subfolder and the location does not involve /\*\* at the end.
- Option D is incorrect. The only mydata.txt will be returned.

- Option E is incorrect. Both Native and external tables skip the files with the names beginning with a period (.) or an underline (\_). Therefore, \_hidden.txt won't be returned.
- Option F is incorrect. In the given Scenario, the only mydata.txt will be returned.

**Reference:**

To know more about using external tables with Synapse SQL, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables?tabs=hadoop>

Ask our Experts

[View Queries](#)

## Question 46

Domain: Design and implement data storage

You work in ATR company as Azure Data Engineer. You have been assigned a task to create and manage the hierarchies in Azure Analysis Services. Here is the list of some tasks:

1. Change the name of a hierarchy
2. Change the name of a child level
3. Change the order of a child level in a hierarchy
4. Add a child level to the given hierarchy
5. Remove a child level from the given hierarchy

Choose the tasks that you can perform while editing the hierarchies.

- A. Only I, II, and III
- B. Only I, IV, and V
- C. Only II and III
- D. Only IV and V
- E. All the above right

## Explanation:

**Correct Answer: E**

Hierarchies can be created, edited, or even deleted in Azure Analysis Services. While editing the hierarchies, Azure Analysis Services allows renaming a hierarchy, changing the order for the child levels, renaming a child level, removing/deleting a child level from a hierarchy, adding extra columns as child levels, displaying the source name of a child level (the column name), and hiding a child level if it has the same name as the hierarchy parent level.

- Option A is incorrect. Not only I, II, and III, but all the given tasks can be performed while editing the hierarchies.
- Option B is incorrect. Not only I, IV, and V, but all the given tasks can be performed while editing the hierarchies.
- Option C is incorrect. Not only II and III, but all of the given tasks can also be performed while editing the hierarchies.
- Option D is incorrect. Not only IV and V, but all the given tasks can also be performed while editing the hierarchies.
- Option E is correct. All of the given tasks can be performed while editing the hierarchies.

**Reference:**

To know more about creating and managing hierarchies, please visit the below-given link:

- <https://docs.microsoft.com/en-us/analysis-services/tabular-models/create-and-manage-hierarchies-ssas-tabular?view=asallproducts-allversions>



Ask our Experts

[View Queries](#)

## Question 47

Domain: Design and develop data processing

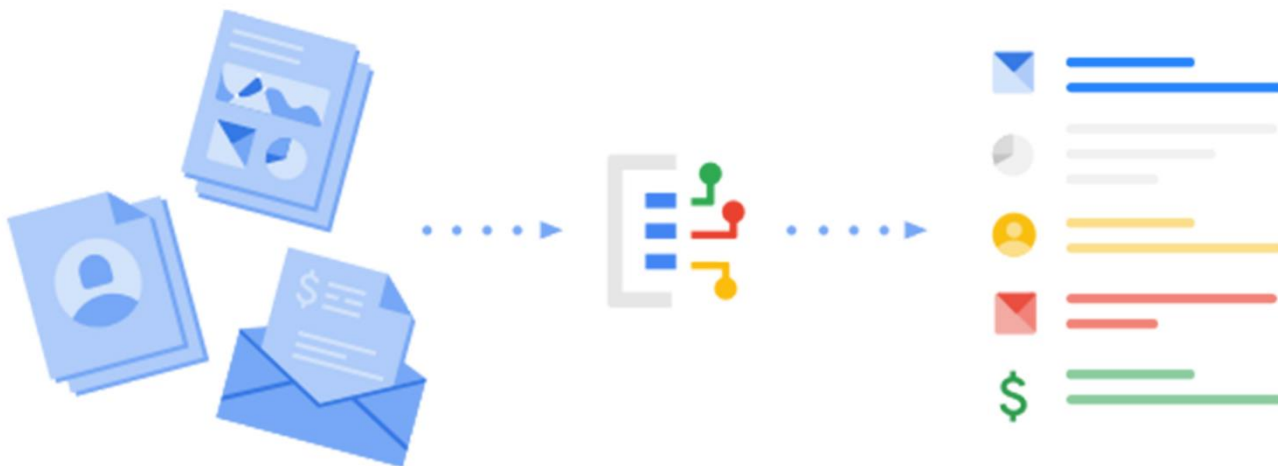
There are a number of JSON properties used in the JSON definition of a Spark Activity. From the given list of properties, choose the properties that are required/mandatory in the JSON definition. (Select all that are applicable)

- A. LinkedServiceName<sup>right</sup>
- B. SparkJobLinkedService
- C. rootPath<sup>right</sup>
- D. className
- E. proxyUser
- F. All of the above

### Explanation:

#### Correct Answers: A and C

Check the below table to know the JSON properties used in the JSON definition. The table also provides the description for each property and also whether that property is required or not.



- Option A is correct. linkedServiceName is a required property in JSON definition.
- Option B is incorrect. SparkJobLinkedService is not a required property in JSON definition.
- Option C is correct. rootPath is a required property in JSON definition.
- Option D is incorrect. className is not a required property in JSON definition.
- Option E is incorrect. proxyUser is not a required property in JSON definition.
- Option F is incorrect. Out of the given options, only linkedServiceName and rootPath are the required JSON properties.

#### Reference:

To know more about transforming data using spark, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/transform-data-using-spark>

Ask our Experts

[View Queries](#)

## Question 48

Domain: Design and develop data processing

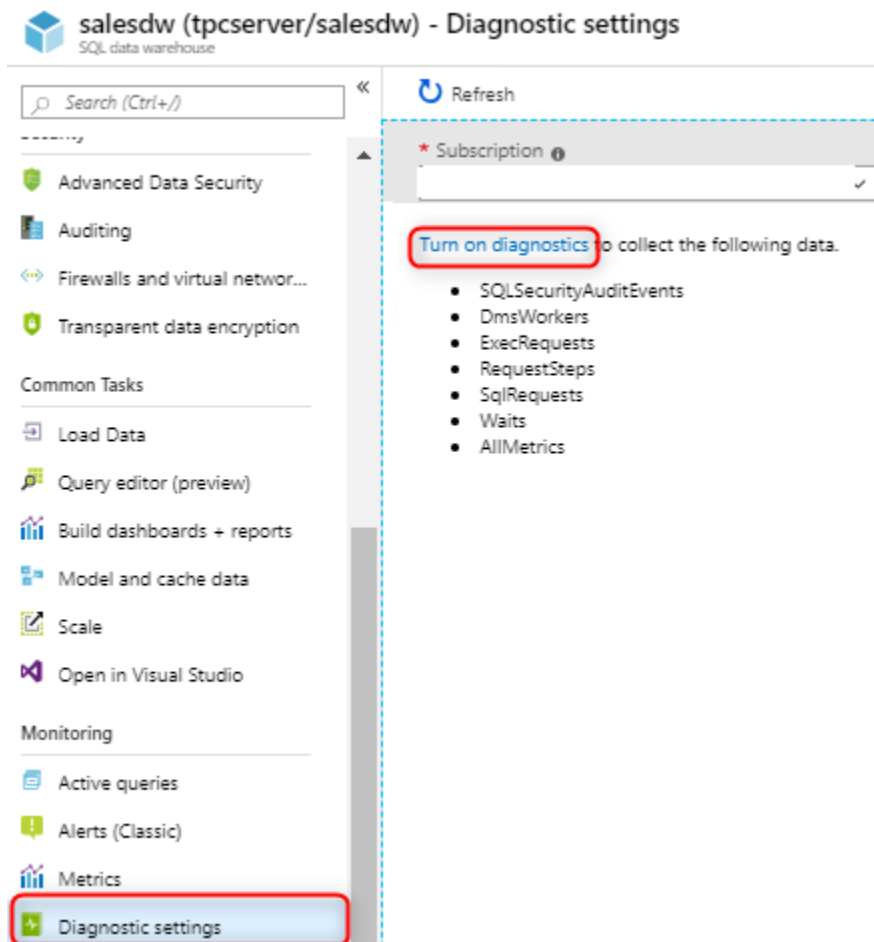
You are an Azure data engineer in a well-reputed multinational company. You need to create a workspace to accomplish the assigned task. Which of the following command would you use to create a new Azure Databricks workspace?

- A. az databricks workspace create **right**
- B. azure databricks workspace create
- C. create az databricks workspace
- D. create azure databricks workspace
- E. None of these

### Explanation:

#### Correct Answer: A

az databricks workspace create command is used to create a new azure databricks workspace. The syntax for creating the new workspace is as given below:



- Option A is correct. az databricks workspace create command is used to create an Azure databricks workspace.
- Option B is incorrect. azure databricks workspace create is not the right syntax to create the workspace.
- Option C is incorrect. There is no command like create az databricks workspace.
- Option D is incorrect. There is no command like create azure databricks workspace.
- Option E is incorrect. az databricks workspace create command is used to create an Azure databricks workspace.

**Reference:**

To know more about creating azure resources, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/devops/pipelines/apps/cd/azure/build-data-pipeline?view=azure-devops>

Ask our Experts

[View Queries](#)

## Question 49

Domain: Design and develop data processing

Which of the following Read-only service-defined variables would you use in the AutoScale formula when you want scaling depending upon the number of tasks that are running at a specific point in time?

- A. \$ActiveTasks
- B. \$RunningTasks<sup>right</sup>
- C. \$SucceededTasks
- D. \$PendingTasks

### Explanation:

**Correct Answer: B**

\$RunningTasks variable should be used while scaling depending upon the number of tasks running at a point in time.

- Option A is incorrect. \$ActiveTasks variable is used when scaling depending upon the number of tasks that are queued up to run.
- Option B is correct. \$RunningTasks variable should be used when scaling depending upon the number of tasks running at a point in time.
- Option C is incorrect. \$SucceededTasks is used for the number of tasks completed successfully.
- Option D is incorrect. \$PendingTasks is used for the number of tasks that are queued or running.

**Reference:**

To know more about creating automatic formulas, please visit the following link:

- <https://docs.microsoft.com/en-us/azure/batch/batch-automatic-scaling>

Ask our Experts

[View Queries](#)

## Question 50

Domain: Design and develop data processing

Azure Stream Analytics is a PaaS event processing engine. Below given is the list of key features or benefits of using Azure Stream Analytics to process streaming data. From the list, choose the incorrect statement.

- A. It eases writing complex time-based queries and aggregations over the data produced by connected devices, sensors, or applications.

- B. Stream Analytics processes the data in real-time, empowering powerful insights that further help in real-time decision-making.
- C. It enables writing and testing transformation queries in the Azure portal
- D. Integration with Azure Blob storage **right**
- E. None of these

## Explanation:

### Correct Answer: D

Integration with Azure Blob storage is not the main feature or benefit of using Azure Stream Analytics to process streaming data. The integration with Blob storage can be used for the processing of static data.

- Option A is incorrect. Azure Stream Analytics eases writing complex time-based queries and aggregations over the data produced by connected devices, sensors, or applications.
- Option B is incorrect. It is true that Stream Analytics processes the data in real-time, empowering powerful insights that further help in real-time decision-making.
- Option C is incorrect. Azure Stream Analytics enables writing and testing transformation queries in the Azure portal.
- Option D is correct. Integration with Azure Blob storage is not the main feature or benefit of using Azure Stream Analytics to process streaming data.
- Option E is incorrect. The statement "Integration with Azure Blob storage" is incorrect.

### Reference:

To know more about processing Azure Stream Analytics, please visit the below-given link:

- <https://docs.microsoft.com/en-us/learn/modules/introduction-to-data-streaming/4-process-events-azure-stream-analytics>

Ask our Experts

[View Queries](#)

## Question 51

Domain: Design and develop data processing

An Azure Data Factory pipeline needs to be scheduled in such a way that it executes with the deletion of a file in Azure Data Lake Storage Gen2 container. Which kind of trigger would you prefer?

- A. Schedule Trigger
- B. On-demand Trigger
- C. Event Trigger **right**
- D. Tumbling Window

## Explanation:

### Correct Answer: C

Event-driven architecture is a general data integration pattern that includes production, detection, consumption, and reaction to the events. Data integration scenarios generally need Data Factory consumers to trigger pipelines depending upon the events occurring in the storage account, for example, the deletion or arrival of any file in the Blob Storage account.

- Option A is incorrect. Schedule trigger is used to schedule a pipeline to run periodically i.e daily, hourly, etc.)
- Option B is incorrect. In the given scenario, an Event trigger, not an on-demand trigger, should be used.
- Option C is correct. Event trigger should be used to schedule the pipeline to execute with the deletion of the file.
- Option D is incorrect. A tumbling window is a particular type of trigger that fires at a periodic time interval from a particular start time while retaining state.

## Reference:

To know more about creating event triggers, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-event-trigger>

Ask our Experts

[View Queries](#)

## Question 52

Domain: Design and implement data security

Static Data Masking and Dynamic Data Masking are two data masking options offered by Microsoft. Which of the following statements(s) are true about Static and Dynamic Data Masking. (Select all that are applicable)

- A. Static data masking takes place on the original database whereas dynamic data masking takes place on a copy of the database.
- B. Dynamic data masking takes place on the original database whereas static data masking takes place on a copy of the database.**right**
- C. In Static Data Masking, the mask takes place at the storage level while in the case of dynamic data masking, the mask takes place on the fly at query time.**right**
- D. In Dynamic Data Masking, the mask takes place at the storage level while in the case of static data masking, the mask takes place on the fly at query time.
- E. In static as well as Dynamic Data Masking, all users have access to the same masked data.

## Explanation:

### Correct Answers: B and C

The below table highlights the differences between Static Data Masking and Dynamic Data Masking.

## GoChina ICP Filing Assistant Service Scope



### End-to-end Solution

- One-on-one Expert Service
- ICP Filing Delivery
- Qualification Compliance Assessment



### Company Registration

- Process Q&A
- Delivery Progress Follow-up
- Efficient Completion of Required Paperwork



### Domain Name Consultation

- Registration and Ownership Transfer
- Transfer-in and Transfer-out Assistance
- Real-name Verification



### Qualification Assessment

- Operations Compliance
- Qualification Document Compliance
- Qualification Compliance



### Assistance in ICP Filing Application

- Support from Experts
- Reduced Learning Costs
- Improved Pass Rate

- Option A is incorrect. Dynamic data masking takes place on the original database whereas static data masking takes place on a copy of the database.
- Option B is correct. Dynamic data masking takes place on the original database whereas static data masking takes place on a copy of the database.
- Option C is correct. In Static Data Masking, the mask takes place at the storage level while in the case of dynamic data masking, the mask takes place on the fly at query time.
- Option D is incorrect. The given statement is not true.
- Option E is incorrect. Only in static data masking, all users have access to the same masked data. In Dynamic Data Masking, the mask varies depending upon the user's permission.

## References:

To know more about static data masking and dynamic data masking, please visit the below-given links:

- <https://azure.microsoft.com/en-in/blog/static-data-masking-preview/>
- <https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

Ask our Experts

[View Queries](#)

## Question 53

Domain: Design and implement data security

Creating an Azure Storage account does not automatically assign you the permissions to access data using Azure AD. An Azure role needs to be explicitly assigned to yourself for Azure Storage. At which of the following levels you can assign it?

- A. Subscription only
- B. Resource Group only
- C. Storage Account only
- D. Container only
- E. Any the above **right**

### Explanation:

#### Correct Answer: E

You can assign the Azure role to yourself at the level of your subscription, storage account, resource group, or container.

- Option A is incorrect. An Azure role can be assigned at any level which includes your subscription, storage account, resource group, or container.
- Option B is incorrect. It is not true that you can assign an Azure role only on the Resource group level but any of your subscription, storage account, resource group, or container.
- Option C is incorrect. It is not true that you can assign Azure role only on storage account but any of your subscription, storage account, resource group, or container.
- Option D is incorrect. An azure role can be assigned at any level which includes your subscription, storage account, resource group, or container.
- Option E is correct. An azure role can be assigned at any level which includes your subscription, storage account, resource group, or container.

#### Reference:

To know more about assigning an Azure role, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/blobs/assign-azure-role-data-access?tabs=portal>

Ask our Experts

[View Queries](#)

## Question 54

Domain: Monitor and optimize data storage and data processing

You are working as Data Administrator in the TDC company. You need to retrieve data from a Log Analytics workspace using log queries. In which of the following languages these log queries are written?

- A. PL/SQL
- B. SQL **right**
- C. Kusto Query Language (KQL)

- D. Transact – SQL

## Explanation:

### Correct Answer: C

To retrieve the data from a Log Analytics workspace, use a log query which is a read-only request for data processing. Kusto Query Language (KQL) is the query language that is used to write these log queries.

- Option A is incorrect. Log queries are written in KQL, not PL/SQL.
- Option B is incorrect. Log queries are written in KQL, not SQL.
- Option C is correct. Kusto Query Language (KQL) is the query language that is used to write the log queries.
- Option D is incorrect. Log queries are written in KQL, not Transact-SQL.

### Reference:

To know more about Azure Monitor Logs, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/azure-monitor/logs/data-platform-logs>

Ask our Experts

[View Queries](#)

## Question 55

Domain: Monitor and optimize data storage and data processing

While working with pipelines in Azure Data Factory, you get an issue “Expression builder fails to load”. What might be the possible reason behind it? (Select the option that suits best)

- A. Chaining so many activities
- B. network or cache issues with the web browser **right**
- C. Not implementing time to live feature for Data Flow or optimized SHIR
- D. not scaling up SHIR according to your workload
- E. Files in a folder with different schema

## Explanation:

### Correct Answer: B

The expression builder can fail to load because of network or cache issues with the web browser. This problem can be resolved by upgrading the web browser to the latest updated version of the supported browser, clearing cookies for the site, and refreshing the page.

- Option A is incorrect. Chaining so many activities result in error  
“Code”:“BadRequest”,“message”:“ErrorCode=FlowRunSizeLimitExceeded.”
- Option B is correct. The expression builder can fail to load because of network or cache issues with the web browser.
- Option C is incorrect. Not implementing time to live feature for Data Flow or optimized SHIR results in Longer startup times for activities in ADF Copy and Data Flow.
- Option D is incorrect. The given is the possible cause of Hitting capacity issues in SHIR.
- Option E is incorrect. It is the "DelimitedTextMoreColumnsThanDefined" error which is generally returned when the folder you are copying has filed with different schemas.

### Reference:

To know more about troubleshooting pipeline orchestration and triggers in Azure Data Factory, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/pipeline-trigger-troubleshoot-guide>

## Practice Test 2

### Question 1

Domain: Design and implement data storage

You are an Azure Data Engineer. To accomplish your task, you need to design your table service solution to be read efficient. Which of the following are the possible guidelines/ recommendations that you would follow to make your solution to be read efficient?

- A. Normalize the data as possible
- B. Don't create hot partitions
- C. Use compound key values<sup>right</sup>
- D. Use Query projection<sup>right</sup>
- E. Specify either Rowkey or Partitionkey (not both) in the queries

### Explanation:

#### Correct Answers: C and D

Here are some guidelines for designing table service solution to be read-efficient:

- Design for querying in read-heavy applications
- Specify both RowKey and PartitionKey in the queries
- Embrace storing duplicate copies for entities
- Embrace denormalizing your data
- Use compound key values
- Use query projection
- Option A is incorrect. Read efficient solution advocates denormalizing the data, not normalizing as table storage is a cheap solution.
- Option B is incorrect. Not creating hot partitions is a guideline to design your Table service solution to be write-efficient.
- Option C is correct. Using compound key value is a recommended guideline for being the table service solution read -efficient.
- Option D is correct. Using query projection is a recommended guideline for being the table service solution read -efficient. It enables reducing the amount of data that you transfer over the network with the help of queries selecting only the fields that are required.
- Option E is incorrect. The recommended guideline to be the table service solution read efficient is to specify both RowKey and PartitionKey in the queries.

#### Reference:

To know more about guidelines for table design, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/tables/table-storage-design-guidelines>

Ask our Experts

[View Queries](#)

### Question 2

Domain: Design and implement data storage

ADLS (Azure Data Lake Storage) Gen2 destination writes data depending upon the specified data format and makes a separate file for each partition. From the below given list, choose the format(s) that can't be used by ADLS Gen2 destination? (Select all the applicable options)

- A. Avro



- B. Delimited
- C. ORC
- D. XML
- E. None of these **right**

### Explanation:

#### Correct Answer: E

The destination can write using all the given data formats like Avro, Delimited, ORC, JSON, Text, Parquet, and XML.

- Option A is incorrect. An Avro file is written by destination for every partition. Also, Avro schema is included in every file.
- Option B is incorrect. For every partition, a delimited file is written by the destination.
- Option C is incorrect. For every partition, an ORC file is written by the destination.
- Option D is incorrect. For each partition, the destination writes an XML file.
- Option E is correct. The destination can write using all the given data formats (Avro, Delimited, ORC, and XML).

#### Reference:

To know various data formats used by ADLS Gen2 and their naming convention, please visit the below-given link:

- <https://docs.streamsets.com/portal/#transformer/latest/help/transformer/Destinations/ADLS-G2-D.html>

Ask our Experts

[View Queries](#)

### Question 3

Domain: Design and implement data storage

There are a number of different options for data serving storage in Azure. These options vary based on the capability they offer. Which of the following data serving storage options in Azure does not provide SQL language support?

- A. Azure Synapse SQL pool
- B. Azure Synapse Spark pool
- C. Hive LLAP on HDInsight
- D. Cosmos DB
- E. Azure Analysis Services **right**

### Explanation:

#### Correct Answer: E

Except Azure Analysis services, all data serving storage options like SQL Database, Azure Synapse Spark pool, Azure Synapse SQL pool, Azure Data Explorer, Azure Data Explorer, Hive LLAP on HDInsight, HBase/Phoenix on HDInsight, and Cosmos DB provide SQL language support.

- Option A is incorrect. Azure Synapse SQL pool supports SQL language.
- Option B is incorrect. Azure Synapse Spark pool supports SQL language.
- Option C is incorrect. Hive LLAP on HDInsight is one of the data serving storage options that supports SQL language.
- Option D is incorrect. Cosmos DB is one of the data serving storage options that supports SQL language.
- Option E is correct. Azure Analysis Services does not support SQL.

#### Reference:

To know more about analytical data store in Azure, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/analytical-data-stores>

Ask our Experts

[View Queries](#)

#### Question 4

Domain: Design and implement data storage

There is an application that frequently needs to find all the orders delivered in a particular month. Which of the following sharding strategies would you implement to divide the data store to enable quick data retrieval?

- A. Lookup strategy
- B. Range strategy **right**
- C. Hash Strategy
- D. Normalized Strategy

#### Explanation:

##### Correct Answer: B

Range strategy results in grouping the related items together in the same shard, and ordering them by shard key. In the case of the application that frequently requires finding all orders delivered in a particular month, the data could be quickly retrieved/accessed if all orders associated with a month are stored in time and date order in the same shard.

- Option A is incorrect. Lookup strategy is not the best sharding strategy in the given scenario.
- Option B is correct. range strategy will put all the related orders (i.e. orders for a month) in the same shard which will result in quick data retrieval.
- Option C is incorrect. Hash strategy distributes the data across the shards to achieve a balance among the size of every shard and the average load to be encountered by each shard.
- Option D is incorrect. There is no such sharding strategy as a normalized strategy.

##### Reference:

To know more about sharding patterns, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/patterns/sharding#sharding-strategies>

Ask our Experts

[View Queries](#)

#### Question 5

Domain: Design and implement data storage

You have an Azure SQL database. You have created two geometry instances and now you are interested in returning the point where these two geometry instances intersect? Which of the following methods will you use to achieve the goal?

- A. Intersection()
- B. GeoIntersection()
- C. STIntersection() **right**
- D. IntersectionPoint()
- E. None of these

#### Explanation:

##### Correct Answer: C

STIntersection() method is used to return the point where two geometry instances intersect each other.

Example:

The screenshot shows the Alibaba Cloud console for Server Load Balancer. The left sidebar has a red box around 'ALB New'. The main content area is titled 'Instances' and shows a table of instances. The table has columns: Instance Name/ID, IP Address, Status, Monitoring, and Port/Health Check/Backend. Two instances are listed, both with a status of 'Active'.

- Option A is incorrect. Intersection is not the right function to return the point where two geometry instances intersect.
- Option B is incorrect. There is no function like GeoIntersection().
- Option C is correct. STIntersection() method is used to return the point where two geometry instances intersect each other.
- Option D is incorrect. IntersectionPoint() is not a valid function in Azure.
- Option E is incorrect. STIntersection() is the right choice.

#### Reference:

To know more about geometry types, please visit the below-given link:

- <https://docs.microsoft.com/en-us/sql/t-sql/spatial-geometry/spatial-types-geometry-transact-sql?view=sql-server-ver15>

Ask our Experts

[View Queries](#)

#### Question 6

Domain: Design and implement data storage

If you need to export the data, you can define an external table and export the data to it. Is it possible to override the properties of the external table using the export command?

- A. Yes
- B. No right

#### Explanation:

##### Correct Answer: B

The export command references the external table by name. You can't override the external table properties with the help of the export command. For example, data in the Parquet form can't be exported to the external table having CSV data format as CSV.

#### Reference:

To know more about exporting data to an external table, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-explorer/kusto/management/data-export/export-data-to-an-external-table>

Ask our Experts

[View Queries](#)

### Question 7

Domain: Design and implement data storage

You need to create a quick object in a test environment and therefore you decide to create a temporal table with an "anonymous" history table. From the given below statement/remarks about the history table in this context, choose the statement(s) that is/are true.

- A. You need to manually create the anonymous history table and provide its specific schema.
- B. The history table is created as a rowstore table.**right**
- C. The history table is created as a columnstore table.
- D. A default clustered index is developed for the history table**right**
- E. A history table is always uncompressed. No compression is ever applied on the history table.

### Explanation:

#### Correct Answers: B and D

An anonymous history table is automatically built in the same schema as the temporal or current table. The history table is built as a rowstore table. If possible, page compression is applied on the history table otherwise the table remains uncompressed. For example, few table configurations, like SPARSE columns, don't allow compression.

- Option A is incorrect. An anonymous history table is automatically built in the same schema as the temporal or current table.
- Option B is correct. It is true that the history table is built as a rowstore table.
- Option C is incorrect. The history table is created as a rowstore table, not columnstore table.
- Option D is correct. A default clustered index is developed for the history table with an auto-generated name with the format *IX\_<history\_table\_name>*. This index has the PERIOD columns (end, start).
- Option E is incorrect. It is not true that the history table always remains uncompressed. If possible, page compression is applied on the history table otherwise the table remains uncompressed. For example, few table configurations, like SPARSE columns, don't allow compression.

#### Reference:

To know more about creating a system-versioned temporal table, please visit the below-given link:

- <https://docs.microsoft.com/en-us/sql/relational-databases/tables/creating-a-system-versioned-temporal-table?view=sql-server-ver15>

Ask our Experts

[View Queries](#)

### Question 8

Domain: Design and implement data storage

You are writing some data on Hadoop in the parquet format using spark. You need to enable compression. Which of the following is/are the valid compression type(s) that you can use for parquet format?

- A. none
- B. gzip
- C. snappy
- D. lzo

- E. All of these **right**

## Explanation:

### Correct Answer: E

In Spark 2.1, the supported compression types of Parquet data types are: none, gzip, snappy and lzo. For Spark 2.4 / 3.0, the supported compression types are: uncompressed, none, snappy, lzo, gzip, brotli, lz4, and zstd.

- Option A is incorrect. none is not the only supported compression format. The complete Parquet dataset supports all given compression types.
- Option B is incorrect. gzip is not the only supported compression format. The complete Parquet dataset supports all given compression types.
- Option C is incorrect. snappy is not the only supported compression format. The complete Parquet dataset supports all given compression types.
- Option D is incorrect. lzo is not the only supported compression format. The complete Parquet dataset supports all given compression types.
- Option E is correct. The Parquet dataset supports all given compression types.

### Reference:

To know more about parquet format in Azure Data Factory, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/format-parquet>

Ask our Experts

[View Queries](#)

## Question 9

Domain: Design and develop data processing

You are working in Azure Data Explorer Web UI. You want to enable Error level highlighting for interpreting the verbosity or severity level of every row in the result panel and have them coloured accordingly. To enable the error level highlighting, the column needs to be of one of some supported types. From the below-given list, choose the column types that are supported?

- A. Int **right**
- B. Double
- C. Long **right**
- D. String **right**
- E. Float
- F. Boolean
- G. Enum

## Explanation:

### Correct Answers: A, C and D

There are some specific column requirements for error level highlighting. For highlighted error levels, the column needs to be of type long, int, or string.

- If the column type is int or long:
  - The name of the column must be *Level*
  - It may only contain numbers between 1 and 5.
- If the column type is the string:
  - Having the name of the column *Level* is optional. It will help in improving the performance.
  - The column may have only the following values:
    - crit, critical, assert, fatal, high
    - error, e

- warning, s
- w, warning, monitor
- information
- verbose, verb, d

- Option A is correct. Int is a valid datatype for the column for error level highlighting.
- Option B is incorrect. Double is not a valid data type for the column for error level highlighting.
- Option C is correct. Long is a valid datatype for the column for error level highlighting.
- Option D is correct. String is a valid datatype for the column for error level highlighting.
- Option E is incorrect. Float is not a valid datatype for the column for error level highlighting.
- Option F is incorrect. Boolean is not a valid datatype for the column for error level highlighting.
- Option G is incorrect. Enum is not a valid datatype for the column for error level highlighting.

**Reference:**

To know more about how to Query data in Azure Data Explorer Web UI, visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-explorer/web-query-data>

Ask our Experts

[View Queries](#)

## Question 10

Domain: Design and develop data processing

There are a number of various analytical data stores that use different languages, models, and provide different capabilities. Which of the following is a low-latency NoSQL data store that provides a high-performance and flexible option to query structured and semi-structured data?

- A. Azure Synapse Analytics
- B. HBase right
- C. Spark SQL
- D. Hive
- E. None of these

## Explanation:

### Correct Answer: B

HBase is a low-latency NoSQL data store that provides a high-performance and flexible option to query structured and semi-structured data. The primary data model used by HBase is the Wide column store.

- Option A is incorrect. Azure Synapse is a managed service depending upon the SQL Server database technologies and is optimized for supporting large-scale data warehousing workloads.
- Option B is correct. HBase is a low-latency NoSQL data store that provides a high-performance and flexible option to query structured and semi-structured data.
- Option C is incorrect. Spark SQL is an API developed on Spark that enables the creation of data frames and tables which are possible to be queried using SQL syntax.
- Option D is incorrect. It is HBase, not Hive that is a low-latency NoSQL data store that provides a high-performance and flexible option to query structured and semi-structured data.
- Option E is incorrect. HBase is the right answer.

**Reference:**

To know more about batch processing, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/batch-processing>

Ask our Experts

[View Queries](#)

## Question 11

Domain: Design and develop data processing

The below given are the steps (not necessarily in the right sequence) to run the continuous integration & continuous delivery (CI/CD) pipeline (in a random sequence).

1. Choose Azure Repos Git as the location for source code.
2. Go to the Pipelines page. Then to create a new pipeline, select the action.
3. While setting up your pipeline, choose the Existing Azure Pipelines YAML file.
4. Choose your repository from the list of your repositories.
5. Run the pipeline.

Choose the correct sequence of steps.

- A. a-b-c-d-e
- B. b-a-d-c-e **right**
- C. a-b-d-c-e
- D. a-b-c-e-d
- E. b-a-c-d-e

### Explanation:

#### Correct Answer: B

The below steps are required to be followed to run the CI/CD pipeline:

The screenshot shows the Alibaba Cloud console interface. At the top, there is a navigation bar with the Alibaba Cloud logo, a 'Workbench' tab, a search bar, and various service links (Expenses, Tickets, ICP, Enterprise, Support). Below the navigation bar, the 'Console' tab is selected. The main content area displays the 'Hotlink Protection' configuration page. The page title is 'Hotlink Protection' and it includes a description: 'Uses an HTTP Referer whitelist to prevent unauthorized users from accessing your data in OSS. [Learn more.](#)'. The 'Referer Whitelist' section is highlighted with a red box, showing a text input field containing 'https://10.10.10.10.com/'. Below this, there is a toggle switch for 'Allow Empty Referer' which is currently turned on. At the bottom of the configuration section, there are 'Save' and 'Cancel' buttons. Below the Hotlink Protection section, there is a 'Cross-Origin Resource Sharing (CORS)' section with a description: 'Resolves CORS issues in JavaScript. [Learn more.](#)' and a 'Configure' button. A hamburger menu icon is visible in the bottom right corner.

- Option A is incorrect. As can be checked from the explanation, the option doesn't give the right sequence.
- Option B is correct. The given sequence is right to run CI/CD pipeline.
- Option C is incorrect. The given sequence is not correct.
- Option D is incorrect. The given sequence is not the right sequence to run CI/CD pipeline.
- Option E is incorrect. The given sequence is not correct.

### Reference:

To know more about building a data pipeline, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/devops/pipelines/apps/cd/azure/build-data-pipeline?view=azure-devops>

Ask our Experts

[View Queries](#)

## Question 12

Domain: Design and develop data processing

Steve Warner's Car Dealership is an establishment in Wellington, New Zealand, which deals in purchasing and selling of cars and autos. Steve has hired you as a skilled consultant for Azure work and projects. You are chairing a team session and, in that session, you need to describe the Tumbling Window. Which of the following statements can you use to explain the Tumbling window?

- A. A window function that models scheduled overlapping windows, jumping forward in time by a fixed/defined period.
- B. A window function that segments a data stream into a contiguous series of fixed size, non-overlapping time segments and works against them. Events can't belong to more than 1 tumbling window.**right**
- C. A window function that generates events for specific times when the content of the window is really changed.
- D. A window function that clusters together events arriving at similar times, filtering out periods of time with no data.
- E. A windowing function that groups events by same timestamp values.

## Explanation:

### Correct Answer: B

A Tumbling window function segments a data stream into a contiguous series of fixed size, non-overlapping time segments and works against them. Events can't belong to more than 1 tumbling window.

- Option A is incorrect. It is the Hopping window function that models scheduled overlapping windows, jumping forward in time by a fixed/defined period.
- Option B is correct. A Tumbling window function segments a data stream into a contiguous series of fixed size, non-overlapping time segments and works against them. Events can't belong to more than 1 tumbling window.
- Option C is incorrect. It is the sliding window function that generates events for specific times when the content of the window is really changed.
- Option D is incorrect. A session window function clusters together events arriving at similar times, filtering out periods of time with no data.
- Option E is incorrect. A Snapshot window function groups events by the same timestamp values.

### Reference:

To know more about Windowing functions, please visit the below-given link:

- <https://docs.microsoft.com/en-us/learn/modules/ingest-data-streams-with-azure-stream-analytics/6-describe-windowing-functions>

Ask our Experts

[View Queries](#)

## Question 13

Domain: Design and develop data processing

Azure Stream Analytics allows using functions for scenarios like real-time scoring with machine learning models, complex mathematical calculations, string manipulations, encoding & decoding data. Which of the following function types are supported by Azure Stream Analytics?



- A. JavaScript user-defined functions
- B. JavaScript user-defined aggregates
- C. C# user defined functions (with Visual Studio)
- D. Azure Machine Learning
- E. All of these right

### Explanation:

#### Correct Answer: E

Azure Stream Analytics supports the below mentioned 4 function types:

- JavaScript user-defined functions
- JavaScript user-defined aggregates
- C# user defined functions (with Visual Studio)
- Azure Machine Learning
- Option A is incorrect. Azure Stream Analytics supports not only JavaScript user-defined functions but all the given function types.
- Option B is incorrect. Azure Stream Analytics supports not only JavaScript user-defined aggregates but all the given function types.
- Option C is incorrect. Azure Stream Analytics supports not only C# user defined functions (with Visual Studio) but all the given function types.
- Option D is incorrect. Azure Stream Analytics supports not only Azure Machine Learning but all the given function types.
- Option E is correct. All given function types are supported by Azure Stream Analytics.

#### Reference:

To know more about user-defined functions in Azure Stream Analytics, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/stream-analytics/functions-overview#exception-handling>

Ask our Experts

[View Queries](#)

### Question 14

Domain: Design and develop data processing

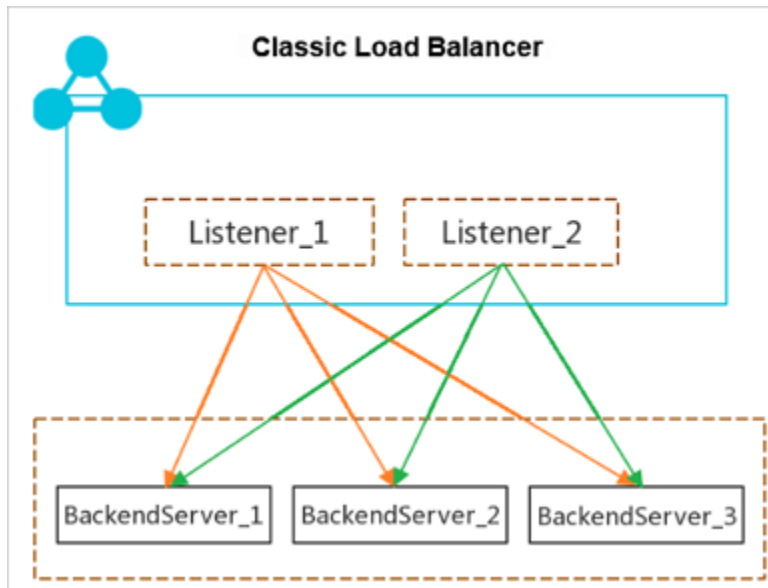
While running and managing your pipelines, you want to ensure that your pipelines don't run until an issue/error is fixed. Which of the following PowerShell cmdlet would you use to suspend/pause your pipelines?

- A. Pause- AzDataFactoryPipeline
- B. Stop- AzDataFactoryPipeline
- C. Resume- AzDataFactoryPipeline
- D. Skip- AzDataFactoryPipeline
- E. Suspend- AzDataFactoryPipeline right

### Explanation:

#### Correct Answer: E

Pipelines can be paused/suspended using the PowerShell cmdlet **Suspend-AzDataFactoryPipeline**. This command helps a lot when you do not want to run the pipelines until an error/issue is resolved. The syntax for the command is:



- Option A is incorrect. Pause- AzDataFactoryPipeline is not the right command to suspend/pause the pipeline.
- Option B is incorrect. Stop- AzDataFactoryPipeline is not the right command to suspend/pause the pipeline.
- Option C is incorrect. Resume- AzDataFactoryPipeline is used to resume the pause/suspended pipeline but is used to suspend/pause the pipeline.
- Option D is incorrect. There is no command like Skip- AzDataFactoryPipeline.
- Option E is correct. Suspend-AzDataFactoryPipeline is the right command to suspend/pause the pipelines.

#### Reference:

To know more about monitoring and managing pipelines, please visit the below-given link:

- <https://github.com/MicrosoftDocs/azure-docs/blob/master/articles/data-factory/v1/data-factory-monitor-manage-pipelines.md>

Ask our Experts

[View Queries](#)

### Question 15

Domain: Design and implement data security

You work as an expert consultant for Advanced Azure Learning where the IT team is working on an Azure SQL database titled AAL\_Targets which has a table titled Targets\_2021. In the table, there is a field Targets\_ID having type varchar(22).

Required: The team is to implement masking for the Targets\_ID field as given below:

- Setting the initial 3 prefix characters as "exposed"
- Setting the final 3 suffix characters as "exposed"
- Setting the rest of the characters as "masked"

The IT team is planning to use data masking with a credit card function mask.

As an expert you need to tell if the given solution meets the requirements and achieves the goal.

- A. Yes
- B. No<sup>right</sup>

## Explanation:

### Correct Answer: B

The below figure described the various masking functions with their masking logic:

Elastic IP

Billing Method	<input type="button" value="Subscription"/> <input checked="" type="button" value="Pay-As-You-Go"/>					
Region	<input type="button" value="China (Qingdao)"/> <input type="button" value="China (Beijing)"/> <input type="button" value="China (Zhangjiakou)"/> <input type="button" value="China (Hohhot)"/> <input type="button" value="China (Hangzhou)"/> <input type="button" value="China (Shanghai)"/>					
	<input type="button" value="China (Shenzhen)"/> <input type="button" value="China (Heyuan)"/> <input type="button" value="China (Guangzhou)"/> <input type="button" value="China (Chengdu)"/> <input type="button" value="China (Hong Kong)"/> <input type="button" value="Japan (Tokyo)"/>					
	<input type="button" value="Singapore"/> <input type="button" value="Australia (Sydney)"/> <input type="button" value="Malaysia (Kuala Lumpur)"/> <input type="button" value="Indonesia (Jakarta)"/> <input checked="" type="button" value="India (Mumbai)"/> <input type="button" value="US (Virginia)"/>					
	<input type="button" value="US (Silicon Valley)"/> <input type="button" value="UAE (Dubai)"/> <input type="button" value="Germany (Frankfurt)"/> <input type="button" value="UK (London)"/> <input type="button" value="China (Ulanqab)"/>					
ISP	<input checked="" type="button" value="BGP"/>					
Network Mode	<input checked="" type="button" value="Public"/>					
Network traffic	<input checked="" type="button" value="By traffic"/>					

Total Configuration Cost **\$0.006**/Hour <sup>?</sup> | Public Traffic Fee: **\$0.090**/GB

As can be seen from the above image, utilizing data masking with a credit card function mask won't be successful. To accomplish the requirements, AAL needs to use Custom Text data masking, which will expose the first and last characters as mentioned and add a custom padding string in the middle.

### Reference:

To know more about dynamic data masking, please visit the below given link:

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

Ask our Experts

[View Queries](#)

## Question 16

Domain: Design and implement data security

You work in a company that provides financial services to its customers. The company wants to allow only its account managers to access the Social Service Number, Emails and other personal information. Which of the following types of security will suit best in this scenario?

- A. Row-level security
- B. Column-level security<sup>right</sup>
- C. Table Level Security
- D. Dynamic Data Masking

## Explanation:

### Correct Answer: B

Column-level security allows restricting column access to protect the private or sensitive data. For instance, if you need to ensure that a particular user 'Smith' can only access specific columns of a table, column-level security helps here. Column-level security can be implemented with the GRANT T-SQL statement.

- Option A is incorrect. Row-level security can be applied on databases to allow fine-grained access over the rows in a table for restricted control upon who can access which type of data.
- Option B is correct. As we need to restrict column access and allow only the account managers to access the Social Service Number, Emails and other personal information, column-level security will work here.
- Option C is incorrect. Table Level Security is not the right option.
- Option D is incorrect. Dynamic data masking masks the data but here we need to restrict the column data access therefore column-level security is the right option.

**Reference:**

To know more about column level security, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/column-level-security>

Ask our Experts

[View Queries](#)

## Question 17

Domain: Design and implement data security

While working in an Azure Databricks workspace, you need to filter depending upon the end of a column value utilizing the Column Class. You are specifically looking at a column titled name and filtered by the words ending with "ka".

Which command filters based on the end of a column value as required?

- A. `df.filter("name like '_ka')`
- B. `df.filter("name like '%ka')`
- C. `df.filter(col("name").endswith("ka"))` **right**
- D. `df.filter( ).col("name").endswith("%ka")`

## Explanation:

### Correct Answer: C

The Column Class supports both the `endswith()` method and the `like()` method to filter based on the end of a column value.

- Option A is incorrect. The given command won't help in filtering based on the end of a column value.
- Option B is incorrect. Column Class support is the function to filter based on the end of a column value. Therefore, the given option is not correct.
- Option C is correct. `df.filter(col("name").endswith("ka"))` is the right command to filter based on the end of a column value.
- Option D is incorrect. The Column Class supports `endswith()` method but the given is not the right syntax.

**Reference:**

To know more about Dataframes, visit the below-given link:

- <https://docs.databricks.com/spark/latest/dataframes-datasets/introduction-to-dataframes-python.html>

Ask our Experts

[View Queries](#)

## Question 18

Domain: Monitor and optimize data storage and data processing

Activity log offers insight into the operations/activities on every Azure resource in the subscription from the outside (*the management plane*) in addition to the updates on Service Health events. At which of the following layers, Activity Log is available?

- A. Azure Resources
- B. Azure Subscription **right**
- C. Azure Tenant

- D. None of the above

## Explanation:

### Correct Answer: B

Activity Log is present on the Azure Subscription layer. Activity log offers insight into the operations/activities on every Azure resource in the subscription from the outside i.e *the management plane*, in addition to the updates on Service Health events. Activity Log can be used to know who, *what*, *when* and *who* for any write operations (DELETE, POST, PUT) performed on the resources in your Azure subscription.

- Option A is incorrect. These are the Resource Logs that are available at Azure Resources layer.
- Option B is correct. Activity Log is present on the Azure Subscription layer.
- Option C is incorrect. Azure Active Directory logs are available at the Azure Tenant layer.
- Option D is incorrect. Activity Log is present on the Azure Subscription layer.

### Reference:

To know more about Azure platform logs, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/azure-monitor/essentials/platform-logs-overview>

Ask our Experts

[View Queries](#)

## Question 19

Domain: Monitor and optimize data storage and data processing

You work in an Azure Transformational Logics (ATL) company and you have been given the responsibility to create and update query-optimization statistics utilizing the Synapse SQL resources in a dedicated SQL pool. The following are the guiding principles recommended for updating the statistics during the load process. Which of the following is/are not true?

- A. Ensure that every loaded table is having at least 1 statistics object updated.
- B. Focus on the columns participating in ORDER BY, GROUP BY, JOIN and DISTINCT clauses.
- C. Update "ascending key" columns like order dates more frequently as these values are not considered/included in the statistics histogram.
- D. Update static distribution columns more frequently.right
- E. None of these

## Explanation:

### Correct Answer: D

The below-given guiding principles are recommended to update the statistics during the load process:



- Option A is incorrect. It is true that you should ensure that every loaded table is having at least 1 statistics object updated.
- Option B is incorrect. The given one is also a guiding principle to update the statistics during the load process.
- Option C is incorrect. The given one is also a guiding principle to update the statistics during the load process.
- Option D is correct. Rather than more frequently, you need to update static distribution columns less frequently.
- Option E is incorrect. Option D describes the wrong principle.

**Reference:**

To know more about Statistics in Synapse SQL, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-statistics#update-statistics>

Ask our Experts

[View Queries](#)

## Question 20

Domain: Monitor and optimize data storage and data processing

While working on a project, you notice that your Apache Spark Job is underperforming. Which of the following can be a possible reason for a slower performance on such join or shuffle jobs?

- A. Bucketing
- B. Using the Cache Option
- C. Data Skew<sup>right</sup>
- D. Enabling Autoscaling
- E. None of these

## Explanation:

### Correct Answer: C

The data skew can be the most common reason for the slower performance of your join or shuffle jobs because of existing asymmetry in your job data.

Being a distributed system in Spark, Data is divided into several pieces, known as partitions, moved into the diverse cluster nodes, and processed in parallel. If a partition gets much larger than the other, the node processing it is likely to face resource issues and slow down the whole execution. This type of data imbalance is known as data skew.

- Option A is incorrect. Bucketing does not result in the slow performance of join or shuffle jobs.
- Option B is incorrect. Using the Cache option is likely to increase, not decrease the performance.
- Option C is correct. The data skew is the most common reason for the slower performance of your join or shuffle jobs.
- Option D is incorrect. Enabling Auto scaling can't be the possible cause of slow performance on Join or Shuffle jobs.
- Option E is incorrect. Option C Data Skew is the correct choice.

### Reference:

To know more about data-skew and how to resolve data skew problems, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-data-lake-tools-data-skew-solutions>

Ask our Experts

[View Queries](#)

## Question 21

Domain: Design and implement data storage

Fill in the blanks in the context of Microsoft Azure.

..... data is generally stored in a relational database like SQL Server or Azure SQL Database.

- A. Structured<sup>right</sup>
- B. Unstructured
- C. Semi-structured
- D. JSON format
- E. ORC format

## Explanation:

### Correct Answer: A

Based on the type of data such as unstructured, semi-structured, or structured, data is stored differently. Structured data is generally stored in a relational database like SQL Server or Azure SQL Database.

- Option A is correct. Relational databases like SQL Server or Azure SQL Database store structured data.
- Option B is incorrect. Unstructured data is generally stored in NoSQL databases.
- Option C is incorrect. Relational databases like SQL Server or Azure SQL Database store structured data.
- Option D is incorrect. JSON format is not the right answer.
- Option E is incorrect. Structured data is the correct answer.

### Reference:

To know more about understanding Data Storage Model, please visit the below given link:

- <https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview>

Ask our Experts

[View Queries](#)

## Question 22

Domain: Design and implement data storage

You need to determine the type of Azure service required to fit the following requirements and specifications:

Data classification: Unstructured

Operations:

- retrieve only by ID
- Customers need a high number of read operations with low latency
- updates and creates will be somewhat infrequent operations and can have higher latency as compared to read operations

Latency & throughput: Retrievals need to have high throughput and low latency. Updates and creates can have higher latency as compared to read operations.

Transactional support: Not needed

- A. Azure Cosmos DB
- B. Azure Route Table
- C. Azure SQL Database
- D. Azure Blob Storage **right**
- E. Azure Queue Storage

### Explanation:

#### Correct Answer: D

Azure Blob storage allows storing the files like videos and photos. It works in collaboration with Azure Content Delivery Network (CDN) by caching the content that is used more frequently and storing this content on edge servers. It decreases latency in providing those images to the users.

Azure Blob storage also allows moving images from the hot storage tier to the archive or cool storage tier, to decrease the costs and increase throughput on the most frequently retrieved or viewed photos and videos.

- Option A is incorrect. Using Azure cosmos DB won't meet the requirements.
- Option B is incorrect. Using Azure Route Table is not the right option.
- Option C is incorrect. Azure SQL Database stores the structured data.
- Option D is correct. Azure Blob Storage is the right choice.
- Option E is incorrect. Azure Queue Storage is not the right choice as per the requirements.

#### References:

To more about Azure blob storage, please visit the below given links:

- <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>
- <https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction#example-scenarios>

Ask our Experts

[View Queries](#)

## Question 23

Domain: Design and implement data storage



Being the expert consultant, you are chairing a team session in which you need to describe Azure Data Factory (ADF). Which of the following statements would you use to describe the Azure Data Factory?

- A. A storage that can store data in a highly compressed and in-memory cache for optimized query performance
- B. Low-latency NoSQL data store that provides a high-performance and flexible option to query structured and semi-structured data
- C. A managed service depending upon the SQL Server database technologies and is optimized for supporting the large-scale data warehousing workloads
- D. An ELT tool to orchestrate data from various sources to the target<sup>right</sup>

### Explanation:

#### Correct Answer: D

Azure Data Factory is an ELT (Extract, Transfer, and Load Data) tool to orchestrate data coming from different sources to the destination (target). ADF allows us to extract data from various sources and targets like Azure Data Warehouse, SQL Server, etc.

- Option A is incorrect. These are Azure Analysis Services tabular models that can store data in a highly compressed and in-memory cache for optimized query performance.
- Option B is incorrect. HBase is a low-latency NoSQL data store that provides a high-performance and flexible option to query structured and semi-structured data.
- Option C is incorrect. Azure Synapse is a managed service depending upon the SQL Server database technologies and is optimized for supporting large-scale data warehousing workloads.
- Option D is correct. Azure Data Factory is an ELT tool to orchestrate data from various sources to the target.

#### Reference:

To know more about Azure Data Factory, please visit the below-given link:

- <https://visualbi.com/blogs/microsoft/azure/file-partition-using-azure-data-factory/>

Ask our Experts

[View Queries](#)

### Question 24

Domain: Design and implement data storage

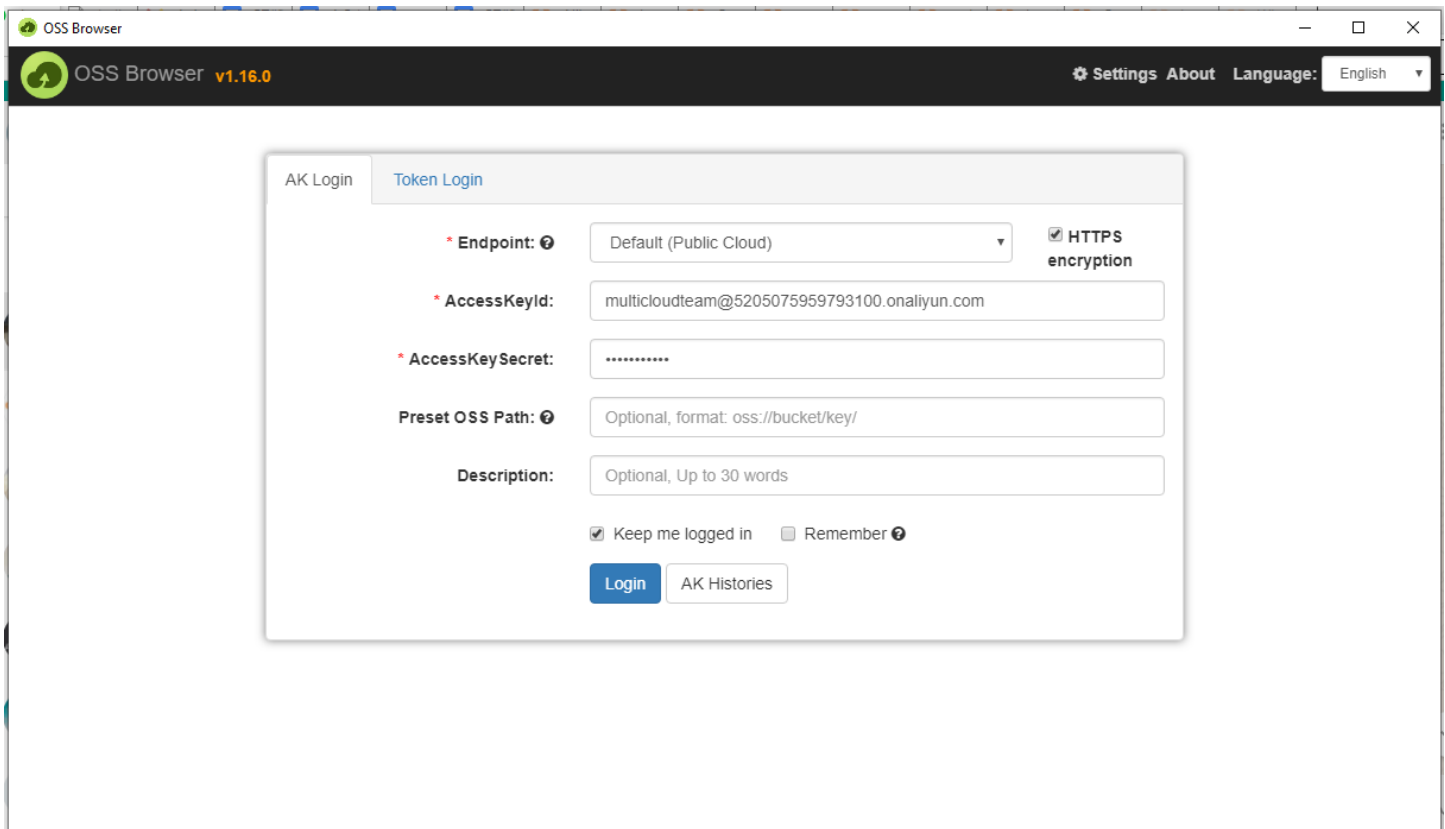
There are a number of different options for data serving storage in Azure. These options vary based on the capability they offer. Which of the below-given options don't offer Row-Level security? (Select two options)

- A. SQL Database
- B. Azure Data Explorer<sup>right</sup>
- C. HBase/Phoenix on HDInsight
- D. Hive LLAP on HDInsight
- E. Azure Analysis Services
- F. Cosmos DB<sup>right</sup>

### Explanation:

#### Correct Answers: B and F

The below table mentions the various security capabilities offered by different data serving storage options.



- Option A is incorrect. SQL Database offers Row-level security.
- Option B is correct. Azure Data Explorer doesn't provide Row-level security.
- Option C is incorrect. HBase/Phoenix on HDInsight offers Row-level security with domain-joined HDInsight clusters.
- Option D is incorrect. Hive LLAP on HDInsight offers Row level security with domain-joined HDInsight clusters.
- Option E is incorrect. Azure Analysis Services offers Row-level security.
- Option F is correct. Cosmos DB doesn't provide Row-level security.

#### Reference:

To know more about analytical data stores in Azure, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/analytical-data-stores>

Ask our Experts

[View Queries](#)

## Question 25

Domain: Design and implement data storage

You need a NoSQL database of a supported API model, with low latency performance and at planet scale.

Which of the following would you use?

- A. Azure SQL Database
- B. Azure Cosmos DB **right**
- C. Azure DB Server
- D. Azure Database for PostgreSQL
- E. Azure Database for MYSQL

#### Explanation:

**Correct Answer: B**

Use Azure Cosmos DB when you need a NoSQL database of the supported API model, at planet scale, and with low latency performance. Nowadays, Cosmos Databases support five-nines uptime (99.999 percent). It is likely to support response times below 10 ms when it is provisioned correctly.

- Option A is incorrect. The requirements demand a NoSQL database.
- Option B is correct. Using Azure Cosmos DB is the right choice.
- Option C is incorrect. Azure DB server is not the right choice.
- Option D is incorrect. The requirements ask for a NoSQL database of a supported API model, with low latency performance and at planet scale which is Azure cosmos Database.

Option E is incorrect. The Question asks for a NoSQL Database.

**Reference:**

To know more about Azure Cosmos DB, please visit the below-given link:

- <https://azure.microsoft.com/en-us/services/cosmos-db/>

Ask our Experts

[View Queries](#)

**Question 26**

Domain: Design and implement data storage

You are writing a PowerShell script to copy an archive blob to a new blob within the same storage account. In the script, you need to initialize some variables. From the below options, choose the variables that must be initialized with your blob names. (Choose all that are applicable)

- A. \$accountName
- B. \$rgName
- C. \$srcBlobName<sup>right</sup>
- D. \$destBlobName<sup>right</sup>
- E. \$newBlobName
- F. \$achiveBlobName
- G. \$srcdestBlob

**Explanation:****Correct Answers: C and D**

You can use a PowerShell Script to copy an archive blob to a new blob within the same storage account. As can be seen from the given diagram, you need to initialize a number of variables with your resource group, container, storage account and blob names.

Here you must initialize the following variables with the corresponding names:

Variable	Name used to initialize the variables
\$rgName	Your resource group name.
\$accountName	Your storage account name

\$srcContainerName and \$destContainerName	Your container name
\$srcBlobName and \$destBlobName	Your Blob Names

- Option A is incorrect. \$accountName variable is initialized with your storage account name.
- Option B is incorrect. \$rgName variable is initialized with resource group name.
- Option C is correct. \$srcContainerName and \$destContainerName variables are needed to be initialized with the blob names.
- Option D is correct. \$srcContainerName and \$destContainerName variables are needed to be initialized with the blob names.
- Option E is incorrect. There is no need to use or initialize \$newBlobName to copy an archive blob to a new blob.
- Option F is incorrect. There is no need to use or initialize \$archiveBlobName to copy an archive blob to a new blob.
- Option G is incorrect. There is no need to use or initialize \$srcdestBlob to copy an archive blob to a new blob.

#### Reference:

To know more about Rehydrating blob data from the archive tier, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-rehydration?tabs=azure-portal>

Ask our Experts

[View Queries](#)

## Question 27

Domain: Design and implement data storage

You need to export the data and for that purpose, you decide to define an external table. Choose the formats from the below-given options that can be used for the external tables in export scenarios? (Choose all that all applicable)

- A. CSV<sup>right</sup>
- B. XML
- C. ORC
- D. TSV<sup>right</sup>
- E. Delimited
- F. JSON<sup>right</sup>
- G. Parquet<sup>right</sup>
- H. AVRO

## Explanation:

### Correct Answers: A, D, F and G

Utilizing external table in export scenarios is limited only to the four formats: CSV, JSON ,TSV, and Parquet.

- Option A is correct. CSV is one of the 4 formats that is valid for the export tables in export scenarios.
- Option B is incorrect. XML is not a valid format.
- Option C is incorrect. ORC is not a valid format.
- Option D is correct. TSV is a valid format for the export tables in export scenarios.
- Option E is incorrect. Delimited is not a valid format.
- Option F is correct. JSON is a valid format for the export tables in export scenarios.
- Option G is correct. Parquet is a valid format for the export tables in export scenarios.
- Option H is incorrect. Avro is not a valid format.

#### Reference:

To know more about the external tables, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-explorer/kusto/management/external-tables-azurestorage-azuredatalake>

Ask our Experts

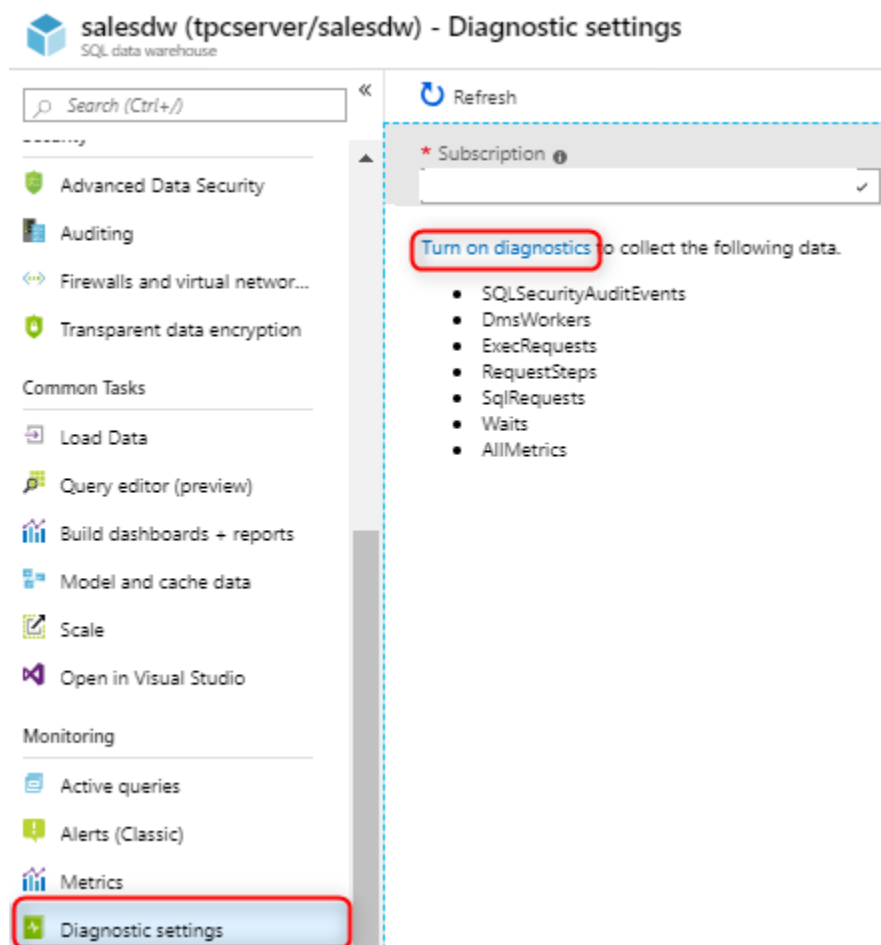
[View Queries](#)

## Question 28

Domain: Design and implement data storage

Advanced Insight Mechanics (AIM) has hired you as an expert consultant. You have called a meeting with the IT Team where you are discussing Azure Synapse. AIM has a Synapse workspace titled aimWorkspace that has an Apache Spark DB titled aimtestdb.

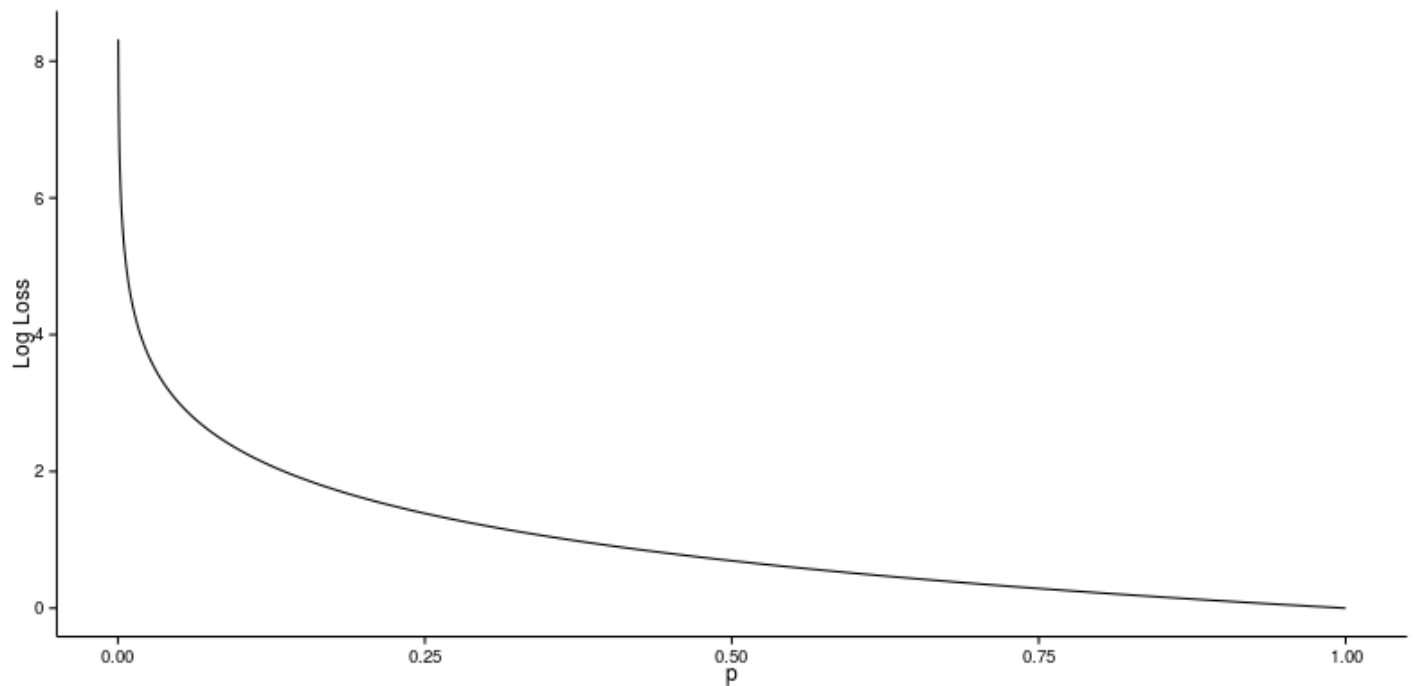
The senior developer of the team runs the below given query in Spark pool in aimWorkspace.



Using Parquet, the senior developer then employs Spark for inserting the below given row in aimtestdb.aimParquetTable. The row has the below-mentioned data:

EmployeeName	EmployeeID	EmployeeStartDate
Steve Warner	1234	2019-03-21

After 5 minutes, the senior developer executes the below given statement through a serverless SQL pool into aimWorkspace.



What will the query return?

- A. Null
- B. Steve Warner
- C. 1234
- D. 2019-03-21
- E. An error<sup>right</sup>

### Explanation:

#### Correct Answer: E

The given query will return an error as the query mentions a column 'name' in the WHERE clause. There is no such column in the actual table.

The actual query should be written as:

```
SELECT EmployeeID -  
FROM aimtestdb.dbo.aimParquetTable  
WHERE employeename = 'Steve Warner';
```

- Option A is incorrect. The query uses the column that actually does not exist in the table. So, it will return an error.
- Option B is incorrect. The query won't return Steve Warner.
- Option C is incorrect. The query won't return 1234.
- Option D is incorrect. The query won't return 2019-03-2021.

- Option E is correct. The query uses the column that actually does not exist in the table. So, it will return an error.

**Reference:**

To know more about Azure Synapse Analytics shared metadata tables, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/metadata/table>

Ask our Experts

[View Queries](#)

### Question 29

Domain: Design and develop data processing

You have created a VIEW within SQLite. From the given below options, choose the operation(s)/statement(s) that is/are possible to apply on view. (Choose all applicable)

- A. Read **right**
- B. Insert
- C. Delete
- D. Update
- E. All of the above

**Explanation:**

**Correct Answer: A**

The VIEWS you create within SQLite are read-only. It is not possible to execute INSERT, DELETE, or UPDATE statements on a view.

- Option A is correct. After a view is created in SQLite, only Read operation can be performed on it.
- Option B is incorrect. It is not possible to execute INSERT, DELETE, or UPDATE statements on a view in SQLite.
- Option C is incorrect. It is not possible to execute INSERT, DELETE, or UPDATE statements on a view in SQLite.
- Option D is incorrect. It is not possible to execute INSERT, DELETE, or UPDATE statements on a view in SQLite.
- Option E is incorrect. After a view is created in SQLite, it becomes Read-only. INSERT, DELETE, or UPDATE statements can't be performed on a view in SQLite.

**Reference:**

To know more about how to apply SQL Transformation, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/apply-sql-transformation>

Ask our Experts

[View Queries](#)

### Question 30

Domain: Design and develop data processing

There are a number of different technology choices for batch processing. These choices vary based on the capability they offer. Which of the following provide the facility of dynamic data masking? (Select all that are applicable)

- A. Azure Data Lake Analytics
- B. Azure Synapse **right**
- C. HDInsight **right**
- D. Azure Databricks

- E. Azure Data Factory

## Explanation:

### Correct Answers: B and C

Azure Synapse offers Dynamic Data Masking. HDInsight (with Hive and Hive LLAP) provides the dynamic data masking facility. Azure Databricks does not offer dynamic data masking facility.

- Option A is incorrect. Azure Data Lake Analytics does not offer dynamic data masking.
- Option B is correct. Azure Synapse offers dynamic data masking.
- Option C is correct. HDInsight( with Hive and Hive LLAP) provides the dynamic data masking facility.
- Option D is incorrect. Azure Databricks does not offer dynamic data masking.
- Option E is incorrect. Azure Synapse and HDInsight are the correct answers.

### Reference:

To know more about batch processing, please visit the below given link:

- <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing>

Ask our Experts

[View Queries](#)

## Question 31

Domain: Design and develop data processing

When you implement the Clean Missing Data module to a set of data, the Minimum missing value ratio and Maximum missing value ratio are two important factors in replacing the missing values. If the Maximum missing value is set to 1, what does it mean? [ select the one that suits best]

- A. missing values are cleaned only when 100% of the values in the column are missing.
- B. missing values are cleaned even if there is only one missing value.
- C. missing values are cleaned only when there is only one missing value.
- D. missing values won't be cleaned.
- E. missing values are cleaned even if 100% of the values in the column are missing.right

## Explanation:

### Correct Answer: E

Maximum missing value ratio is specified as the maximum number of missing values that can be present for the operation that is to be executed. By default, the Maximum missing value ratio is set to 1 which indicates that missing values will be cleaned even if 100% of the values in the column are missing.

- Option A is incorrect. The use of the word "Only When" does not rightly state the meaning.
- Option B is incorrect. Setting Minimum missing value ratio property to 0 actually means that missing values are cleaned even if there is only one missing value.
- Option C is incorrect. Minimum and Maximum missing value ratios talk only about minimum and maximum ratios, not a specific number.
- Option D is incorrect. The given statement is not right.
- Option E is correct. Setting the Maximum missing value ratio to 1 indicates that missing values will be cleaned even if 100% of the values in the column are missing.

### Reference:

To know more about the clean missing Data Module, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/clean-missing-data>

Ask our Experts

[View Queries](#)



### Question 32

Domain: Design and develop data processing

You are the team leader and while working on the project, one of your team members is confused about which command to invoke to view the list of active streams? What will you advise him?

- A. spark.view.active
- B. spark.view.activeStreams
- C. spark.streams.show
- D. spark.streams.active<sup>right</sup>

### Explanation:

#### Correct Answer: D

Invoke the command spark.streams.active to see the list of active streams.

- Option A is incorrect. spark.view.active is not the right command.
- Option B is incorrect. spark.view.activeStreams is not a valid command.
- Option C is incorrect. Invoking spark.streams.show won't help in getting the list of active streams.
- Option D is correct. spark.streams.active should be invoked to see the list of active streams.

#### Reference:

To know more about structured streaming programming, please visit the below-given link:

- <https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html>

Ask our Experts

[View Queries](#)

### Question 33

Domain: Design and develop data processing

You are leading an IT team in Azure company and a new member joined the team. He reviews the options for an input to an Azure Stream Analytics job that your IT team is working on (which requires high throughput and low latency). He seems confused about input he should use and therefore asks you “Which Azure product should I plan to use for the job's input?”

What would be your answer?

- A. Azure Table Storage
- B. Azure Blob Storage
- C. Azure Event Hubs<sup>right</sup>
- D. Azure Data Lake Storage
- E. Azure Queue Storage

### Explanation:

#### Correct Answer: C

Azure Event Hubs are the Azure product that consumes data streams from applications at high throughput and low latency.

The following diagram demonstrates how data is sent to Azure Stream Analytics, analyzed, and sent for further actions like presentation or storage.

Create Rule

File Lifecycle

Validity Period (Days)

Expiration Date

Disabled

Transit to IA Storage Class < Delete

Transit to IA Storage Class

☒

180

The storage class of the object is converted from Standard to IA 180 days after it is last modified.

Transit to Archive Storage Class

☒

180

The storage class of the object is converted from Standard or IA to Archive 180 days after it is last modified.

Delete

☐

240

Delete Parts

Part Lifecycle

Validity Period (Days)

Expiration Date

Disabled

Delete

☐

30

OK

Cancel

- Option A is incorrect. Using Azure Table storage for job's input is not the right option.
- Option B is incorrect. Azure Blob Storage is not a recommended choice for high throughput and low latencies. Azure event hub is a better choice.
- Option C is correct. Azure Event Hubs are the Azure product that consumes data streams from applications at high throughput and low latency.
- Option D is incorrect. Using Azure Data Lake storage for job's input is not the right option.
- Option E is incorrect. Using Azure Queue storage for job's input is not the right option.

#### Reference:

To know more about Azure Event hubs, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-about>
- <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction>

Ask our Experts

[View Queries](#)

### Question 34

Domain: Design and develop data processing

Which of the following windowing features would you use to output the events only when the content of the window really changed in Stream Analytics jobs?

- A. Tumbling Window
- B. Topping Window

- C. Sliding Window<sup>right</sup>
- D. Session Window
- E. Snapshot Window

## Explanation:

### Correct Answer: C

It is the sliding window that outputs the events only when the content of the window really changed in Stream Analytics jobs.

- Option A is incorrect. A Tumbling window function segments a data stream into a contiguous series of fixed size, non-overlapping time segments and works against them.
- Option B is incorrect. There is no windowing function like Topping Window.
- Option C is correct. The sliding window function is used to output the events only when the content of the window really changed in Stream Analytics jobs.
- Option D is incorrect. Session window function groups events arriving at similar times, filtering out points/periods of time where there exists no data.
- Option E is incorrect. The snapshot window function is used to group the events having the same timestamp.

### Reference:

To know more about Windowing functions, please visit the below-given link:

- <https://docs.microsoft.com/en-us/learn/modules/ingest-data-streams-with-azure-stream-analytics/6-describe-windowing-functions>

Ask our Experts

[View Queries](#)

## Question 35

Domain: Design and develop data processing

On each file upload, Batch writes 2 log files to the compute node. These log files can be examined to know more about a specific failure. These two files are:

- A. fileuploadin.txt and fileuploaderr.txt
- B. fileuploadout.txt and fileuploadin.txt
- C. fileuploadout.txt and fileuploaderr.txt<sup>right</sup>
- D. fileuploadout.JSON and fileuploaderr.JSON
- E. fileupload.txt and fileuploadout.txt

## Explanation:

### Correct Answer: C

When you upload a file, 2 log files are written by Batch to the compute node, named - fileuploadout.txt and fileuploaderr.txt. These log files help to get information about a specific failure. The scenarios where file upload is not done, these fileuploadout.txt and fileuploaderr.txt log files don't exist.

- Option A is incorrect. fileuploadin.txt and fileuploaderr.txt are not the right files.
- Option B is incorrect. fileuploadout.txt and fileuploadin.txt are not the right log files.
- Option C is correct. On each file upload, Batch writes 2 log files to the compute node. These files are fileuploadout.txt and fileuploaderr.txt.
- Option D is incorrect. fileuploadout.JSON and fileuploaderr.JSON are not the right log files.
- Option E is incorrect. fileupload.txt and fileuploadout.txt are not the right files.

### Reference:

To know more about job and task error checking, please visit the below given link:

- <https://docs.microsoft.com/en-us/azure/batch/batch-job-task-error-checking>

Ask our Experts

### Question 36

Domain: Design and implement data security

You are working on a project for a healthcare provider that provides health care services to the customers. The healthcare provider wants that only nurses and doctors must be able to access medical records. It should not be visible to the billing department. Which of the following types of security would you implement in this scenario?

- A. Column-level security<sup>right</sup>
- B. Row-level security
- C. Dynamic Data Masking
- D. Table Level Security

### Explanation:

#### Correct Answer: A

Column level security is used to restrict the column access to protect the sensitive data. If you want that only some specific person or department should be able to view some data, column-level security works the best.

- Option A is correct. As we want only nurses and doctors to be able to access medical records and records should not be visible to the billing department, column-level security will do the required.
- Option B is incorrect. Row-level security is the feature that can be applied on databases to enable fine-grained access over rows in a database table for restricted control upon who can access which type of data.
- Option D is incorrect. Dynamic data masking won't help in restricting the column access only to some specific persons.
- Option D is incorrect. Table Level Security is not a good choice.

#### References:

To know more about column level security, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/column-level-security>
- <https://azure.microsoft.com/en-us/blog/column-level-security-is-now-supported-in-azure-sql-data-warehouse/>

Ask our Experts

[View Queries](#)

### Question 37

Domain: Design and implement data security

Azure role-based access control (RBAC) is the authorization technique that can be used for managing access to Azure resources. From the below-given list, select four top-level classifications to which you can assign roles to grant access.

- A. Managed Identities<sup>right</sup>
- B. Assets
- C. Devices
- D. Attributes
- E. Users<sup>right</sup>
- F. Groups<sup>right</sup>
- G. Service Principals<sup>right</sup>
- H. Workflows
- I. Orchestrations

## Explanation:

### Correct Answers: A, E, F and G

RBAC is the authorization technique that is used for managing access to Azure resources. To grant access, roles can be assigned to groups, service principals, users, or managed identities at a particular scope.

- Option A is correct. A Role can be assigned to Managed Identities.
- Option B is incorrect. Role can't be assigned to Assets.
- Option C is incorrect. Role can't be assigned to devices.
- Option D is incorrect. Role can't be assigned to Attributes.
- Option E is correct. To grant access, roles can be assigned to groups, service principals, users, or managed identities at a particular scope.
- Option F is correct. To grant access, roles can be assigned to groups, service principals, users, or managed identities at a particular scope.
- Option G is correct. A role can be assigned to Service Principals.
- Option H is incorrect. Role can't be assigned to workflows.
- Option I is incorrect. Role can't be assigned to Orchestrations.

### Reference:

To know more about Identity and Access Management, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/ready/enterprise-scale/identity-and-access-management>

Ask our Experts

[View Queries](#)

## Question 38

Domain: Design and implement data security

While working in Azure Learning Institute, you are working with Azure SQL Managed Instance. Here is the list of some scenarios. Choose from the list the scenarios where you might need to provide a public endpoint connection with Azure SQL Managed Instance.

- A. The managed instance must integrate with multi-tenant-only PaaS offerings.
- B. You require higher throughput of data exchange than possible with VPN.
- C. Company policies/terms prohibit PaaS inside corporate networks.
- D. All the above right

## Explanation:

### Correct Answer: D

Here are the scenarios for public endpoint connection:

Billing Method Subscription Pay-As-You-Go ?

---

Dedicated Block Storage The dedicated block storage cluster feature is not enabled in the current region. Try another region.

Cluster

---

Storage Enhanced SSD (ESSD) ▲ 40 GiB 3800 IOPS Performance Level ? : PL1 (up to 50,000 IOPS per disk) ▼ [Create from Snapshot](#)

Ultra Disk

Enhanced SSD (ESSD) Enhanced SSD (ESSD) 0 GB in the region. Remaining quota: 65496 GB

Standard SSD Learn how to select an appropriate disk for your scenario.

Learn how to [create a subscription disk](#).

Quantity — 1 + Unit

- Option A is incorrect. Not only this, but all scenarios need to provide a public endpoint connection.
- Option B is incorrect. Not only this, but all scenarios need to provide a public endpoint connection.
- Option C is incorrect. Not only this, but all scenarios need to provide a public endpoint connection.
- Option D is correct. You need to provide a public endpoint connection for all the given scenarios.

#### References:

To know more about public endpoints, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/azure-sql/managed-instance/public-endpoint-overview>
- <https://github.com/Jayvardhan-Reddy/Azure-Certification-DP-201>

Ask our Experts

[View Queries](#)

### Question 39

Domain: Monitor and optimize data storage and data processing

As the queries are submitted, a dedicated SQL pool query optimizer attempts to check which Data access paths will result in the minimum amount of effort to get the data needed to resolve the query. This cost-based optimizer will compare the cost of different query plans, and will then select the plan that has the minimum cost.

Statistics in a serverless SQL pool have the same aim of utilizing a cost-based optimizer to select a plan with the fastest execution. The thing is that how it generates the statistics is different.

Statement: In a serverless SQL pool, if statistics are missing, the query optimizer creates statistics on the whole tables in the query predicate or join condition to enhance the cardinality estimates for the query plan.

Select whether the statement is true or false.

- A. True
- B. False right

#### Explanation:

**Correct Answer: B**

As per the Azure documentation:



**Reference:**

To know more about Statistics in Synapse SQL, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-statistics>

Ask our Experts

[View Queries](#)

**Question 40**

Domain: Monitor and optimize data storage and data processing

Combiner mode attempts to distribute huge skewed-key value sets to several vertices that support concurrent execution of the work. There are various attributes of combiner mode? What is the meaning of `SqlUserDefinedCombiner(Mode=CombinerMode.Inner)`.

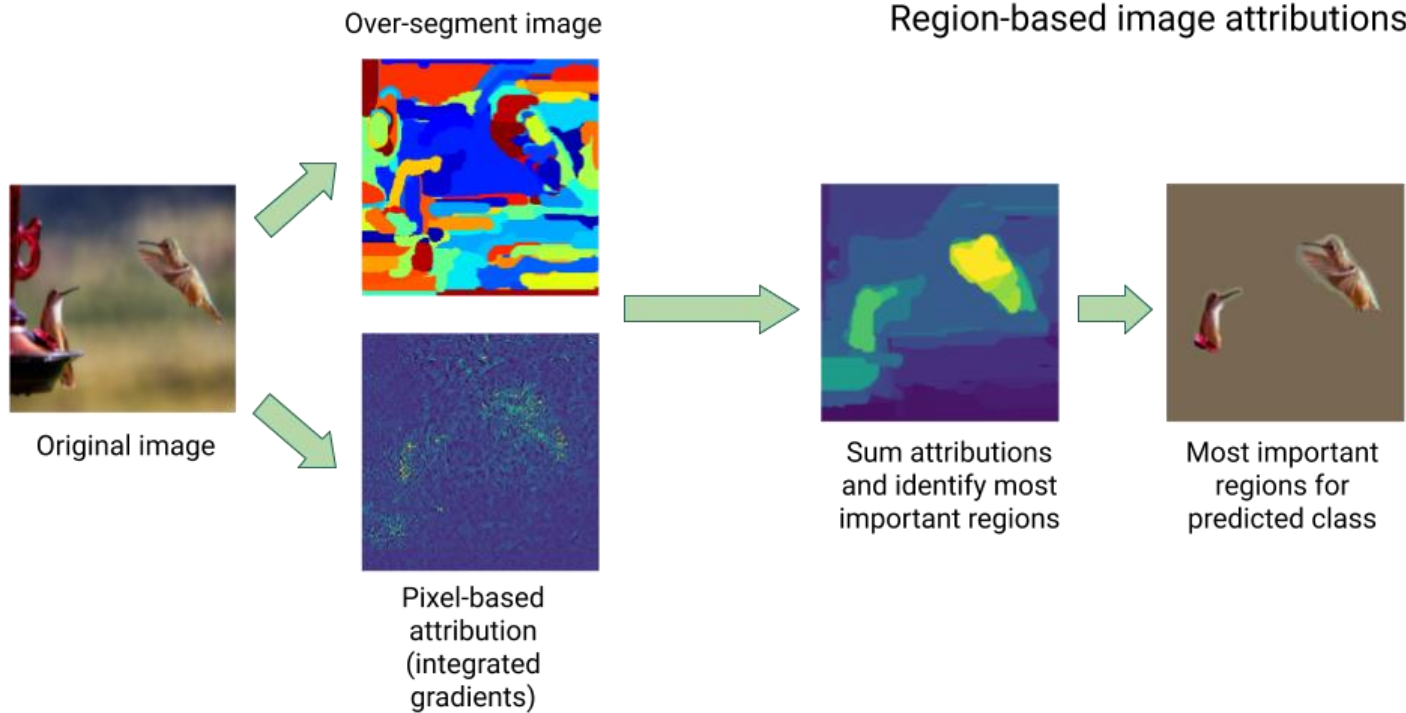
- A. Every output row potentially depends upon all the input rows from left and right with the same key value.
- B. Every output row depends upon a single input row from the left (and potentially all rows with the same key value from the right).
- C. Every output row depends upon a single input row from the right (and potentially all rows with the same key value from the left).
- D. Every output row depends upon a single input row from the left and the right with the same value.right
- E. None of these

**Explanation:**

**Correct Answer: D**

The following are the attributes of the combiner mode:

**Region-based image attributions**



- Option A is incorrect. This is the description for `SqlUserDefinedCombiner(Mode=CombinerMode.Full)`.
- Option B is incorrect. This is the description for `SqlUserDefinedCombiner(Mode=CombinerMode.Left)`.
- Option C is incorrect. This is the description for `SqlUserDefinedCombiner(Mode=CombinerMode.Right)`.
- Option D is correct. This is the description for `SqlUserDefinedCombiner(Mode=CombinerMode.Inner)`.
- Option E is incorrect. D is the right answer.

**Reference:**

To know more about Resolve data-skew problems, please visit the below link:

- <https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-data-lake-tools-data-skew-solutions>

Ask our Experts

[View Queries](#)

**Question 41**

Domain: Design and implement data storage

You have been assigned the task to manage the storage of consumers profiles and Sales data. A general request is to create a list of *“the top 100 consumers including name, account number and sales around for a specific time period”* or *“who are the consumers within a particular geographic region?”*

Is Azure Blob storage a recommended choice for this data?

- A. Yes
- B. No right

**Explanation:**

**Correct Answer: B**

Blob is not a recommended choice for structured data that needs to be queried regularly. Blobs have higher latency than memory and local disk and also do not have the indexing feature that increases the databases' efficiency at running queries.

**Reference:**

To know more about blobs, please visit the below-given link:



- <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>

Ask our Experts

[View Queries](#)

## Question 42

Domain: Design and implement data storage

One of your friends needs to replace the content of a table. He is thinking of deleting the entire directory of the Delta table and creating a new table on the same path. Is this a recommended solution?

- A. Yes
- B. No right

### Explanation:

#### Correct Answer: B

The given solution is not recommended as:

- Deleting a directory is not effective. A directory with very large files can consume hours or even days to delete.
- You lose all content of the deleted files and it is quite hard to recover if you delete the wrong table.
- The directory deletion is not atomic. When you delete the table, a concurrent query reading the table might fail or see a partial table.

#### Reference:

To know more about best practices in Delta Lake, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/databricks/delta/best-practices>

Ask our Experts

[View Queries](#)

## Question 43

Domain: Design and implement data storage

Fill in the given blanks:

Star schema is a modeling technique widely adopted by relational data warehouses. In Star Schema, ..... tables store events or observations and ..... tables define the business entities.

- A. Fact, Dimension right
- B. Dimension, Fact
- C. 2D, 3D
- D. Structured and unstructured
- E. Unstructured and structured

### Explanation:

#### Correct Answer: A

Dimension tables define business entities i.e the *things*, you model. Entities include people, products, concepts and places including time itself. Fact tables store events or observations and can be stock balances, sales orders, temperatures and exchange rates, etc. A fact table consists of dimension key columns related to dimension tables and numeric measure columns.

- Option A is correct. Fact tables and Dimension tables are the right answers.

- Option B is incorrect. Fact tables store events or observations and Dimension tables define business entities.
- Option C is incorrect. 2D and 3D is not the right answer.
- Option D is incorrect. Dimension and fact are two different types of tables used in Star Schema.
- Option E is incorrect. Dimension and fact are two different types of tables used in Star Schema.

**Reference:**

To know more about Star Schema, please visit the below-given link:

- <https://docs.microsoft.com/en-us/power-bi/guidance/star-schema#star-schema-overview>

Ask our Experts

[View Queries](#)

## Question 44

Domain: Design and implement data storage

Azure Synapse Link for Azure Cosmos DB develops a tight integration between Azure Synapse Analytics and Azure Cosmos DB. How would you disable Azure Synapse Link for Azure Cosmos DB?

- A. Deleting the Azure Cosmos DB Container
- B. Deleting the Azure Cosmos DB account **right**
- C. Setting the Azure Synapse Link option to disable on Azure Cosmos DB Container
- D. Setting the Azure Synapse Link option to disable on Azure Cosmos DB Instance

## Explanation:

### Correct Answer: B

The screenshot shows the OSS Browser v1.16.0 interface. The 'Token Login' tab is active. The form contains the following fields and options:

- Endpoint:** A dropdown menu set to 'Default (Public Cloud)'.
- AccessKeyId:** A text input field containing 'multicloudteam@5205075959793100.onaliyun.com'.
- AccessKeySecret:** A text input field with masked characters '\*\*\*\*\*'.
- Preset OSS Path:** A text input field with the placeholder 'Optional, format: oss://bucket/key/'.
- Description:** A text input field with the placeholder 'Optional, Up to 30 words'.
- HTTPS encryption:** A checked checkbox.
- Keep me logged in:** A checked checkbox.
- Remember:** An unchecked checkbox.
- Login:** A blue button.
- AK Histories:** A white button.

Deleting the Azure Cosmos DB account with disable and removing Azure Synapse Link is the right option.

- Option A is incorrect. Deleting the Azure Cosmos DB Container is not the right choice.
- Option B is correct. Deleting the Azure Cosmos DB account with disabling and removing Azure Synapse Link is the right option.

- Option C is incorrect. Setting the Azure Synapse Link option to disabling on Azure Cosmos DB Container is not the right choice.
- Option D is incorrect. Setting the Azure Synapse Link option to disable on Azure Cosmos DB Instance

**Reference:**

To know more about Azure Synapse Link for Azure Cosmos DB, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/cosmos-db/synapse-link-frequently-asked-questions>

Ask our Experts

[View Queries](#)

## Question 45

Domain: Design and implement data storage

Azure Data Factory (ADF) is made up of 4 core components. These components work in collaboration to provide a platform that allows composing data-driven workflows with steps to move & transform data.

Which component can be described by:

*“It is created to perform a specific task by composing the different activities in the task in a single workflow. This can be scheduled to execute, or a trigger can be defined that determines when an execution needs to be kicked off.”*

- A. Activity
- B. Pipeline **right**
- C. Dataset
- D. Linked Service
- E. None of these

## Explanation:

### Correct Answer: B

An Azure subscription can consist of 1 or more Azure Data Factory (ADF) instances. ADF is composed of 4 core components i.e Dataset, Activity, pipeline and Linked Services. These components work in collaboration to provide a platform where data-driven workflows can be composed with steps to move & transform data.

Alibaba Cloud Workbench All Resources India (Mumbai) Search... Expenses Tickets ICf

Server Load Balancer / Instances

## Instances

Create CLB Select a tag Zones: All Fuzzy Match Search by name, ID, or IP address

Instance Name/ID	IP Address	Status	Monitoring	Port/Health Check/Backend
lb-...	vpc-...	Active		TCP: 443 -
kubernetes... ack.aliyun...	vsw-...			TCP: 80 -
ManagedECSInstances...	vpc-...	Active		TCP: 6443 Norm

- Option A is incorrect. Activity is a specific action executed on the data in a pipeline like the ingestion or transformation of the data. Each pipeline may consist of 1 or more activities in it.
- Option B is correct. The given description is for Pipeline.
- Option C is incorrect. This is the data collected by users which are utilized as input for the ETL process. Datasets can be in various formats like CSV, JSON, text or ORC format.
- Option D is incorrect. Linked Service is not the right answer.
- Option E is incorrect. The given description is for Pipeline.

### Reference:

To know more about Azure Data Factory, please visit the below-given link:

- <https://www.educba.com/azure-data-factory/>

Ask our Experts

[View Queries](#)

## Question 46

Domain: Design and implement data storage

For the migration of data from one to another data lake by involving **Azure Data Lake Storage Gen2**, **Azure Blob**, and **Azure File Storage**, you have the option to preserve the file metadata along with data. Which of the following data store built-in system properties can be preserved by the copy activity during data copy?

- A. contentType
- B. contentLanguage
- C. contentEncoding
- D. contentDisposition
- E. cacheControl
- F. All the above

## Explanation:

### Correct Answer: F

The below given attributes can be preserved by the Copy activity during Data Copy:

Basic	Developer	Business	Enterprise
<b>Included</b>	<b>\$19.99</b>	<b>Starting at \$100</b>	<b>Starting at \$8000</b>
<p>Pricing Details:</p> <p>Included with all Alibaba Cloud services.</p>	<p>\$19.99 will be deducted from your account immediately after purchasing the Developer support plan. You can choose whether to make an automatic renewal.</p>	<p>Business support plan charges will be assessed based on your usage of cloud products. You can choose whether to make an automatic renewal.</p>	<p>Enterprise support plan charges will be assessed based on your usage of cloud products. Please consult your business manager for details.</p>
<a href="#">Get It Free</a>	<a href="#">Purchase Now</a>	<a href="#">Contact Us</a>	<a href="#">Contact Us</a>

- Option A is incorrect. Not only contentType, all the given datastore built-in system properties are preserved by Copy activity.
- Option B is incorrect. Not only contentLanguage, all the given datastore built-in system properties are preserved by Copy activity.
- Option C is incorrect. Not only contentEncoding, all the given datastore built-in system properties are preserved by Copy activity.
- Option D is incorrect. Not only contentDisposition, all the given datastore built-in system properties are preserved by Copy activity.
- Option E is incorrect. Not only cacheControl, all the given datastore built-in system properties are preserved by Copy activity.
- Option F is correct. All the given datastore built-in system properties i.e contentLanguage (except for Amazon S3), contentType, contentEncoding, cacheControl and contentDisposition are preserved by Copy activity.

#### Reference:

To know more about copy activity in ADF, please visit the below given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-preserve-metadata>

Ask our Experts

[View Queries](#)

#### Question 47

Domain: Design and develop data processing

The isolationLevel property in the Copy activity source section describes the transaction locking behavior for the SQL source. Which of the following are possible values for isolationLevel? (Select all that are applicable)

- A. ReadCommitted right
- B. ReadUncommitted right
- C. RepeatableRead right
- D. Serializable right
- E. Repeatable Write
- F. Non-Serializable
- G. Write committed
- H. Write uncommitted

#### Explanation:

**Correct Answers: A, B, C and D**

isolationLevel describes the transaction locking behavior for the SQL source. The allowed values for this property are: ReadUncommitted, ReadCommitted, , RepeatableRead, Snapshot and Serializable. If not mentioned, the default isolation level of the database is used.

- Option A is correct. Read committed is one of the allowed values for isolationLevel.
- Option B is correct. Read uncommitted is one of the allowed values for isolationLevel.
- Option C is correct. Repeatable Read is one of the allowed values for isolationLevel.
- Option D is correct. Serializable is one of the allowed values for isolationLevel.
- Option E is incorrect. Repeatable Write is not the allowed value for isolationLevel.
- Option F is incorrect. Non-Serializable is not the allowed value for isolationLevel.
- Option G is incorrect. Write Committed is not the allowed value for isolationLevel.
- Option H is incorrect. Write uncommitted is not the allowed value for isolationLevel.

**Reference:**

To know more about Copying and transforming data in Azure SQL Database, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-database#error-row-handling>

Ask our Experts

[View Queries](#)

**Question 48**

Domain: Design and develop data processing

When you create a notebook, you need to mention the pool either SQL or Spark Pool that needs to be connected to the notebook. In terms of languages, a notebook needs to be set with a primary language.

Statement: It is possible to utilize multiple languages in one notebook.

Choose the correct option regarding the statement above.

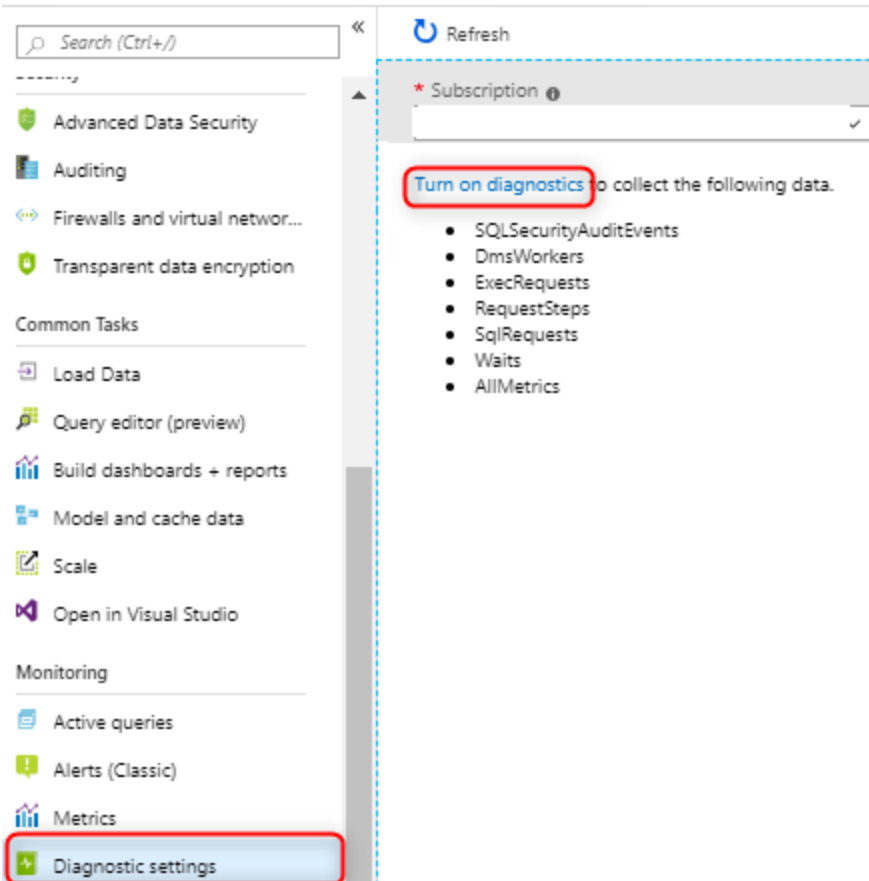
- A. True<sup>right</sup>
- B. False

**Explanation:****Correct Answer: A**

Here is the list of primary languages that are available within the notebook environment:

- PySpark (Python)
- NET Spark (C#)
- Spark (Scala)
- Spark SQL

It is possible to utilize multiple languages in one notebook by mentioning the language through a magic command at the beginning of a cell. The magic commands for switching the cell languages are given as below:



It is not possible to reference variables or data directly across multiple languages in a Synapse Studio notebook.

#### Reference:

To know more about creating, developing, and maintaining notebooks, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-development-using-notebooks?tabs=classical>

Ask our Experts

[View Queries](#)

#### Question 49

Domain: Design and develop data processing

If you don't specify the command option ("checkpointLocation", pointer-to-checkpoint directory) in Structured Streaming, what will happen?

- A. With the stoppage of the streaming job, all state data around the streaming job is lost and on the restart, the job must start from scratch.right
- B. With the stoppage of the streaming job, all state data around the streaming job is dumped to a default location and on the restart, the job must start from the aggregated data rather than tuned specific data.
- C. It won't be possible to create more than 1 streaming query that utilizes the same streaming source because of conflict issues.
- D. The streaming job will function as expected because of the non-existence of the checkpointLocation option.

#### Explanation:

**Correct Answer: A**

You need to set checkpointLocation for many sinks utilized in Structured Streaming. For the sinks where this setting is optional, if you don't set this value, you are at risk of losing your place in the stream.

- Option A is correct. If you don't specify the command option ("checkpointLocation", pointer-to-checkpoint directory) in Structured Streaming, With the stoppage of the streaming job, all state data around the streaming job is lost and on the restart, the job must start from scratch.
- Option B is incorrect. The given statement is not the potential result of not specifying the command option ("checkpointLocation", pointer-to-checkpoint directory) in Structured Streaming.
- Option C is incorrect. The statement is not a potential outcome of the given scenario.
- Option D is incorrect. The given statement is not the potential result of not specifying the command option ("checkpointLocation", pointer-to-checkpoint directory) in Structured Streaming.

**Reference:**

To know more about checkpoint storage in structured streaming, please visit the below-given link:

- <https://www.waitingforcode.com/apache-spark-structured-streaming/checkpoint-storage-structured-streaming/read>

Ask our Experts

[View Queries](#)

## Question 50

Domain: Design and develop data processing

Your friend is new to the task and is asking you to advise which Azure product is the best choice for an ingestion point for data streaming in an event processing solution that utilizes static data as a source.

Which of the following would you advise to your friend?

- A. Azure Event Hub
- B. Azure Blob Storage **right**
- C. Azure Cosmos DB
- D. Power BI
- E. Azure IoT Hub

## Explanation:

**Correct Answer: B**

Azure Blob storage offers an ingestion point for data streaming in an event processing solution that utilizes static data as a source.

- Option A is incorrect. Azure Event Hub is a large data streaming platform and event ingestion service, that can be used to feed events from event producers into Azure Stream Analytics.
- Option B is correct. Azure Blob storage offers an ingestion point for data streaming in an event processing solution that utilizes static data as a source.
- Option C is incorrect. Azure Cosmos Database is a fully managed NoSQL database for the development of modern applications. It can be utilized for storing the output of data stream processing in Azure Stream Analytics.
- Option D is incorrect. Power BI provides a platform to visualize and analyze the aggregated data in near-real-time. Azure Stream Analytics can target Power BI as an output destination.
- Option E is incorrect. Azure IoT Hub is not the right choice.

**Reference:**

To know more about event processing, please visit the below-given link:

- <https://docs.microsoft.com/en-us/learn/modules/introduction-to-data-streaming/3-understand-event-processing>

Ask our Experts

[View Queries](#)



## Question 51

Domain: Design and develop data processing

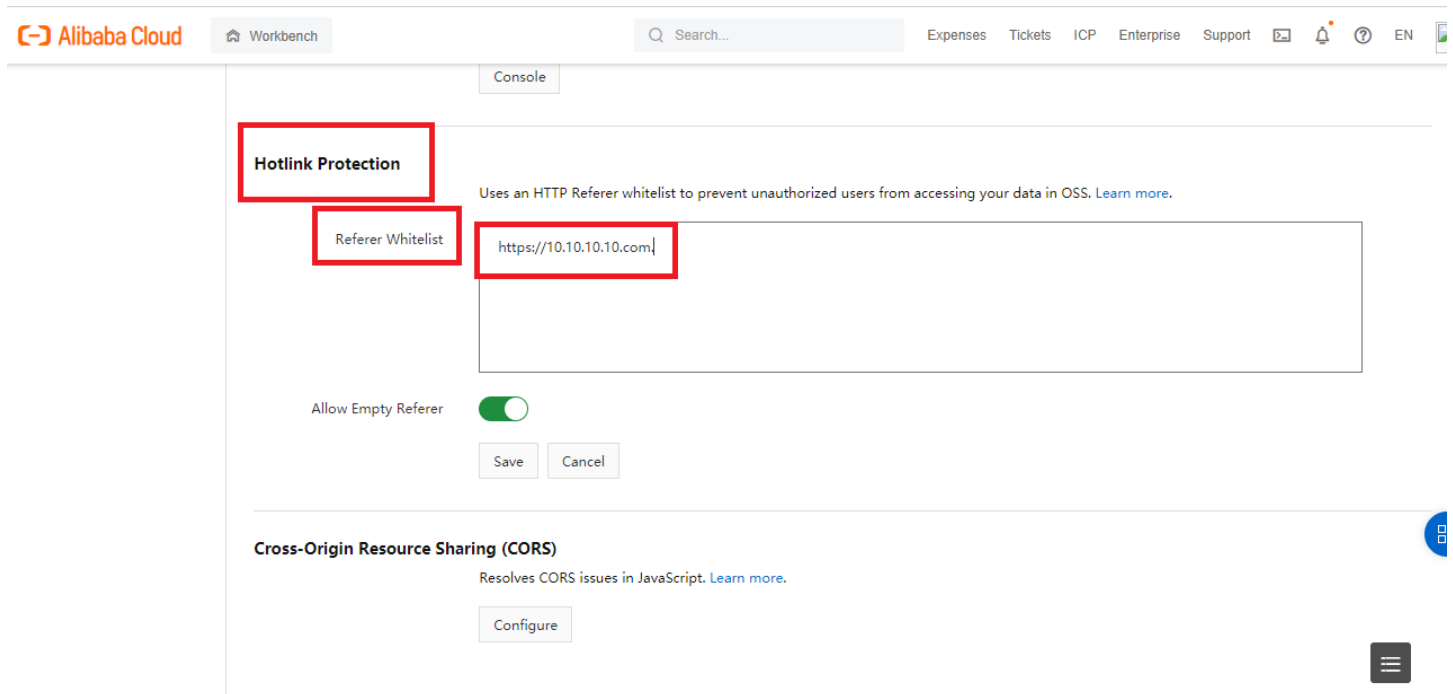
You have been tasked to create, start, and monitor a schedule trigger using Azure PowerShell. Which of the following cmdlet would you use in PowerShell to create a trigger?

- A. Set-AzDataFactoryV2Trigger **right**
- B. Set-ADFV2Trigger
- C. Get-AzDataFactoryV2Trigger
- D. Start-AzDataFactoryV2Trigger
- E. Start-ADFV2Trigger

### Explanation:

**Correct Answer: A**

You need to use Set-AzDataFactoryV2Trigger cmdlet for creating trigger.



- Option A is correct. A trigger is created by using the Set-AzDataFactoryV2Trigger cmdlet.
- Option B is incorrect. Set-ADFV2Trigger is not the right command to create a trigger.
- Option C is incorrect. Get-AzDataFactoryV2Trigger is used to confirm the status of the trigger.
- Option D is incorrect. Start-AzDataFactoryV2Trigger is the right command to start a trigger.
- Option E is incorrect. Start-ADFV2Trigger is not the right command.

### Reference:

To know more about schedule trigger, please visit the below given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-schedule-trigger?tabs=data-factory>

Ask our Experts

[View Queries](#)

## Question 52

Domain: Design and implement data security0

To access data in your establishment storage account, your client makes requests over HTTPS or HTTP. Every request to a secure resource must be authorized.

Which of the following services would you use to ensure that the client has the required permission to access the data? (Choose the most suitable option)

- A. Private Link
- B. Azure AD
- C. Key vault
- D. RBAC<sup>right</sup>
- E. Encryption

### Explanation:

#### Correct Answer: D

The most suitable option is role-based access.

Azure Storage supports role-based access control (RBAC) and Azure Active Directory for both resource management and data operations. To security principals, RBAC roles can be assigned that are scoped to the storage account. Active Directory(AD) should be used to authorize resource management operations like configuration. AD is supported for data operations on Queue and blob storage.

To a security principal or a managed identity for Azure resources, you can use role-based access control (RBAC) roles that are scoped to a resource group, a subscription, a storage account, or an individual queue or container.

- Option A is incorrect. Private Link is not the right choice.
- Option B is incorrect. As clear from the explanation, Azure AD is not the best choice.
- Option C is incorrect. RBAC should be used in the given scenario.
- Option D is correct. To a security principal or a managed identity for Azure resources, you can use role-based access control (RBAC) roles that are scoped to a resource group, a subscription, a storage account, or an individual queue or container.
- Option E is incorrect. Encryption is not the correct option as it is used for the protection of sensitive data/information.

#### Reference:

To know more about role-based access control, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/role-based-access-control/overview>

Ask our Experts

[View Queries](#)

### Question 53

Domain: Design and implement data security

While working with DataFrames, you need to create a DataFrame object. Which of the following functions can you use to create the objects. (Select all that are applicable)

- A. Introduce a variable name and assign it to something like `myDataFrameDF=`<sup>right</sup>
- B. Execute `createOrReplaceObject()`
- C. Use `DF.create()` syntax
- D. Use function `createDataFrame()`<sup>right</sup>
- E. All the above

### Explanation:

### Correct Answers: A and D

First, introduce a variable name and then equate it to `myDataFrameDF =` it is the right way to create the DataFrame objects. Also, DataFrame object can be created using `createDataFrame()` function.

- Option A is correct. Introduce a variable name and then equate it to `myDataFrameDF =` it is the right way to create the DataFrame objects.
- Option B is incorrect. `createOrReplaceObject()` won't help in creating the DataFrame objects.
- Option C is incorrect. The given function is not the right way to create a DataFrame object.
- Option D is correct. DataFrame object can be created using *createDataFrame()* function.
- Option E is incorrect. Not E, A and D are the correct options.

### Reference:

To know more about dataframes, please visit the below given links:

- <https://docs.microsoft.com/en-us/azure/databricks/getting-started/spark/dataframes>
- <https://docs.microsoft.com/en-us/dotnet/api/microsoft.spark.sql.sparksession.createdataframe?view=spark-dotnet>

Ask our Experts

[View Queries](#)

### Question 54

Domain: Monitor and optimize data storage and data processing

Data Factory keeps pipeline-run data stored only for a specific time. Using Azure Monitor is the best option If you desire to keep the data for a longer time. By default, how long are the Azure Data Factory diagnostic logs retained for?

- A. 10 days
- B. 15 days
- C. 40 days
- D. 60 days
- E. None of these right

### Explanation:

#### Correct Answer: E

Pipeline-run data is stored by Data Factory only for 45 days. Using Azure Monitor is the best option If you desire to keep the data for a longer time. Azure Monitor, diagnostic logs can be routed for analysis to many different targets.

- Option A is incorrect. Pipeline-run data is stored by Data Factory for 45 days, not 10 days.
- Option B is incorrect. Pipeline-run data is stored by Data Factory for 45 days, not 15 days.
- Option C is incorrect. Pipeline-run data is stored by Data Factory for 45 days.
- Option D is incorrect. Pipeline-run data is stored by Data Factory only for 45 days, not 60 days.
- Option E is correct. Pipeline-run data is stored by Data Factory for 45 days.

### Reference:

To know more about monitoring Data Factory, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor>

Ask our Experts

[View Queries](#)

### Question 55

Domain: Monitor and optimize data storage and data processing

After an index is created in HyperSpace, several actions can be performed on it like

- Refresh the index if the underlying data changes
- Delete if the index is not needed
- Vacuum if an index is no longer needed

What do you mean by vacuum index?

- A. Delete Only the content of the index
- B. Restructure the index
- C. Physical deletion of the index contents and related metadata completely from Hyperspace's metadata.right
- D. Only remove the data while preserving the metadata
- E. Update the index

### Explanation:

#### Correct Answer: C

Vacuuming an index means performing a hard delete i.e. removing files and the metadata entry completely for that index using the vacuumIndex command.

- Option A is incorrect. Vacuum the index does not mean to delete only the content.
- Option B is incorrect. Vacuum the index does not mean to restructure the index.
- Option C is correct. Vacuum the index means physical deletion of the index contents and related metadata completely from Hyperspace's metadata.
- Option D is incorrect. Vacuum the index deletes not only data but also metadata.
- Option E is incorrect. Updating the index is not the right answer.

#### Reference:

To know more about HyperSpace, please visit the below-given link:

- <https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-performance-hyperspace?pivots=programming-language-csharp>