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 Blobsright
- 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

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

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 Partitionright
- 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_ARCHIVEright
- 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

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)right
- 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[1]	Yes
A region-wide outage occurs in the primary region	No	No	Yes[1]	Yes[1]
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

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

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

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

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 - mean(x)/stdev(x)right
- B. z = x-min(x)/[max(x) - min(x)]
- C. z = 1/1 + exp(-x)
- D. z = Lognormal.CDF(x;;μ,σ

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:

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

- 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

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 Sparkright
- 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

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-locationsright
- 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

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 Hubsright
- 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

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.right
- 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

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 securityright
- 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

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 aright
- 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. createOrReplaceTempViewright
- 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

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 possibleright
- 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 joinsright
- E. Move joins that increase the number of rows after aggregations whenever possible.right
- F. The order of various types of joins matters when it comes to the resource consumptionright

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 prefering 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

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.right
- 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

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. Noright

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

Question 22

Domain: Design and implement data storage

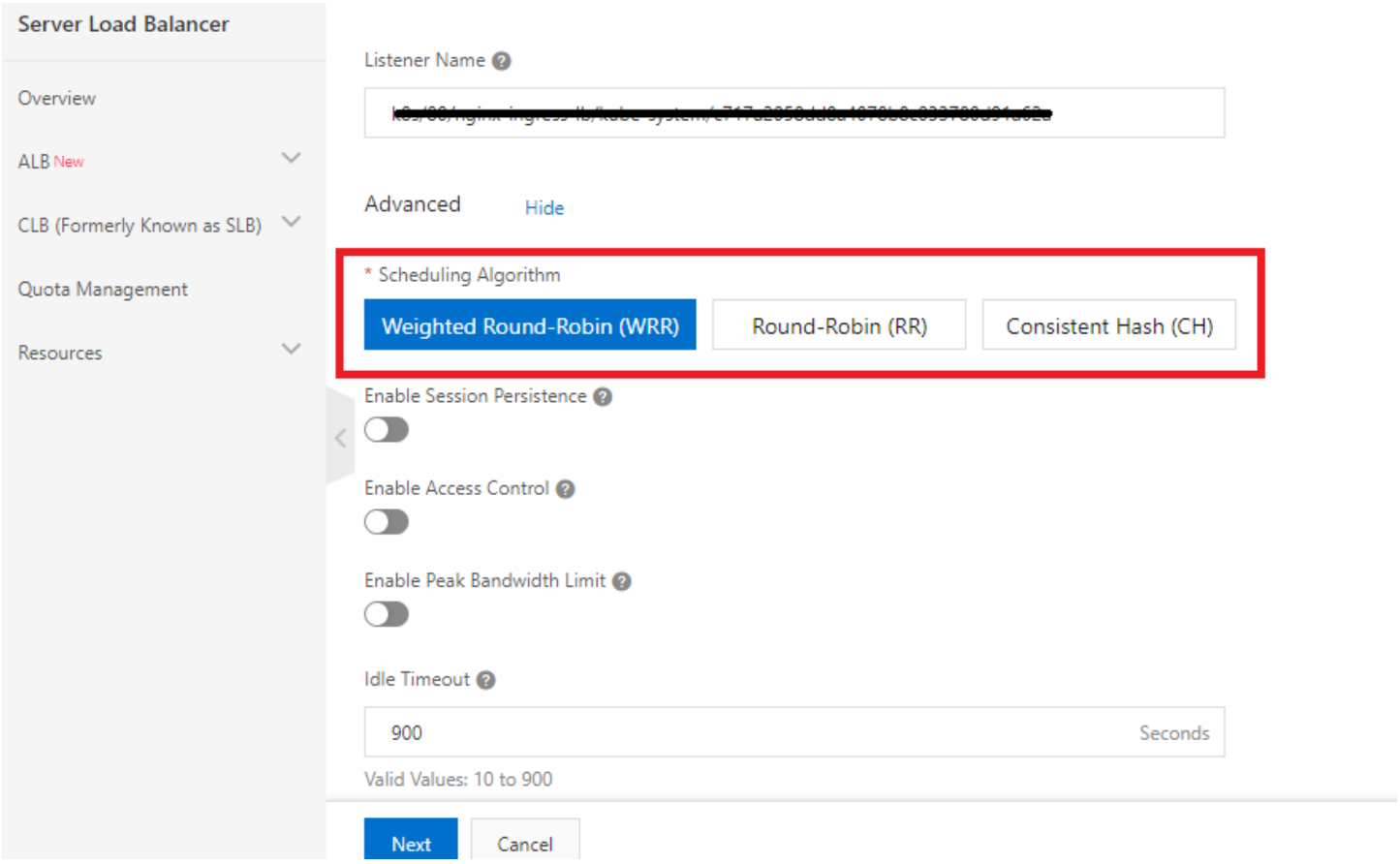
In Azure Synapse, the data is distributed in three different ways: Round-Robin, Hashing, and Replication. Which distribution type to be used depends upon the scenario and the requirements. Which of the following statement(s) is/are true about these distribution types? (Select all that are applicable)

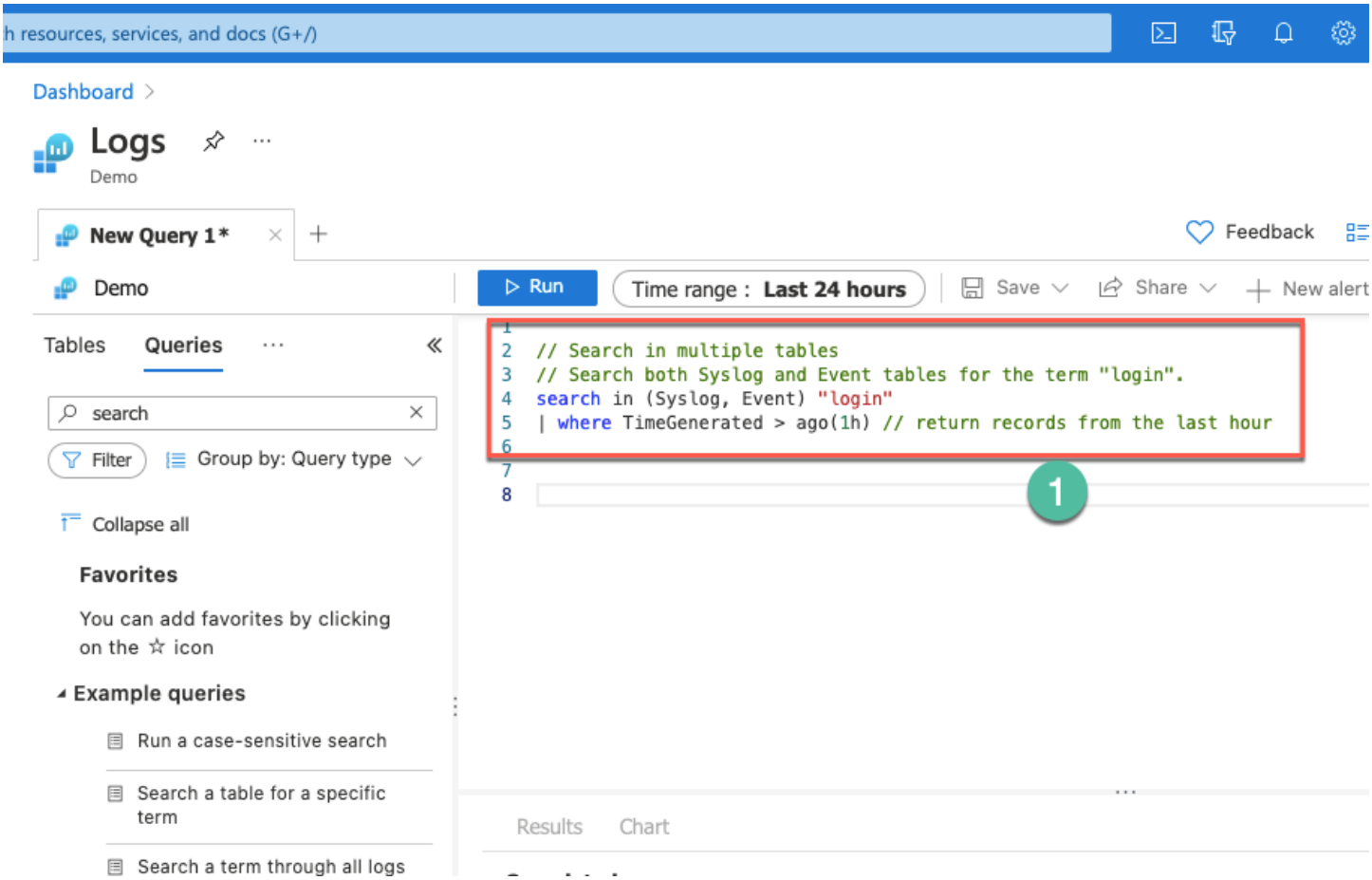
- A. Choose Hash distribution when you can't identify a single key for distributing your data.
- B. Choose Round Robin distribution when you can't identify a single key for distributing your data.right
- C. Choose Round Robin distribution if the table is having temporary data.right
- D. Choose Replicated 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.





- 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

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. Noright

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. XMLright
- D. JSON
- E. JDBC
- F. PARQUET
- G. LIBSVM
- H. MPV4right
- I. BINARYright

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

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 nameright
- 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

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 aboveright

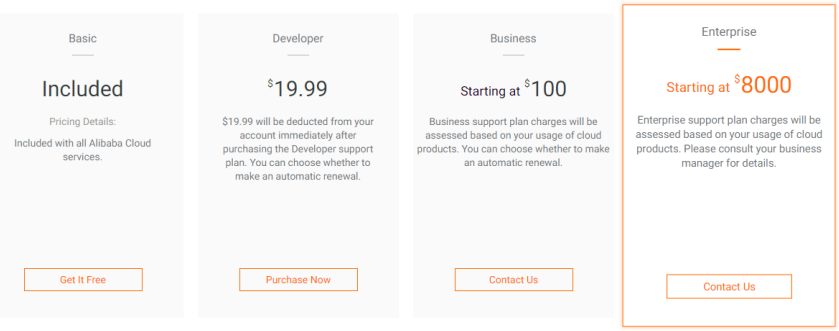
Explanation:

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:



- 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

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
```

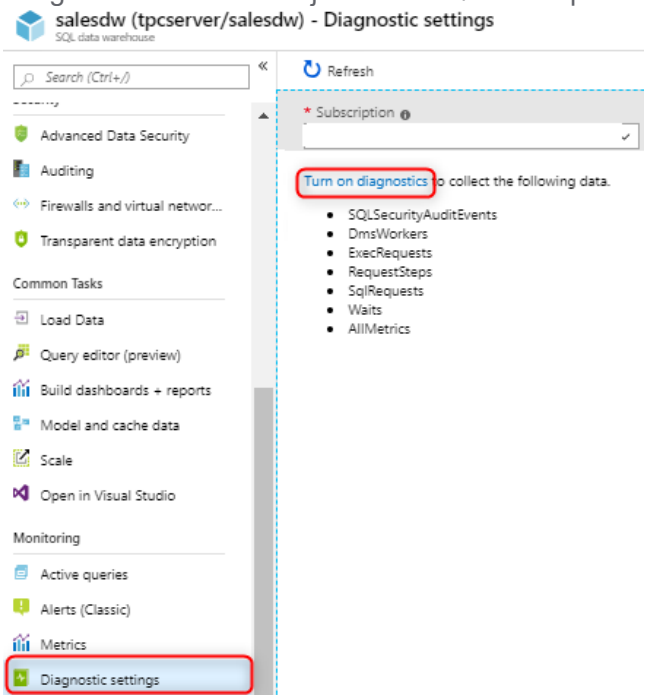
How will you resolve this error?

- A. Remove the environment variable _JAVA_OPTIONS and rerun the pipeline
- B. Add an environment variable _JAVA_OPTIONS and rerun the pipelineright
- 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.



- 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

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.

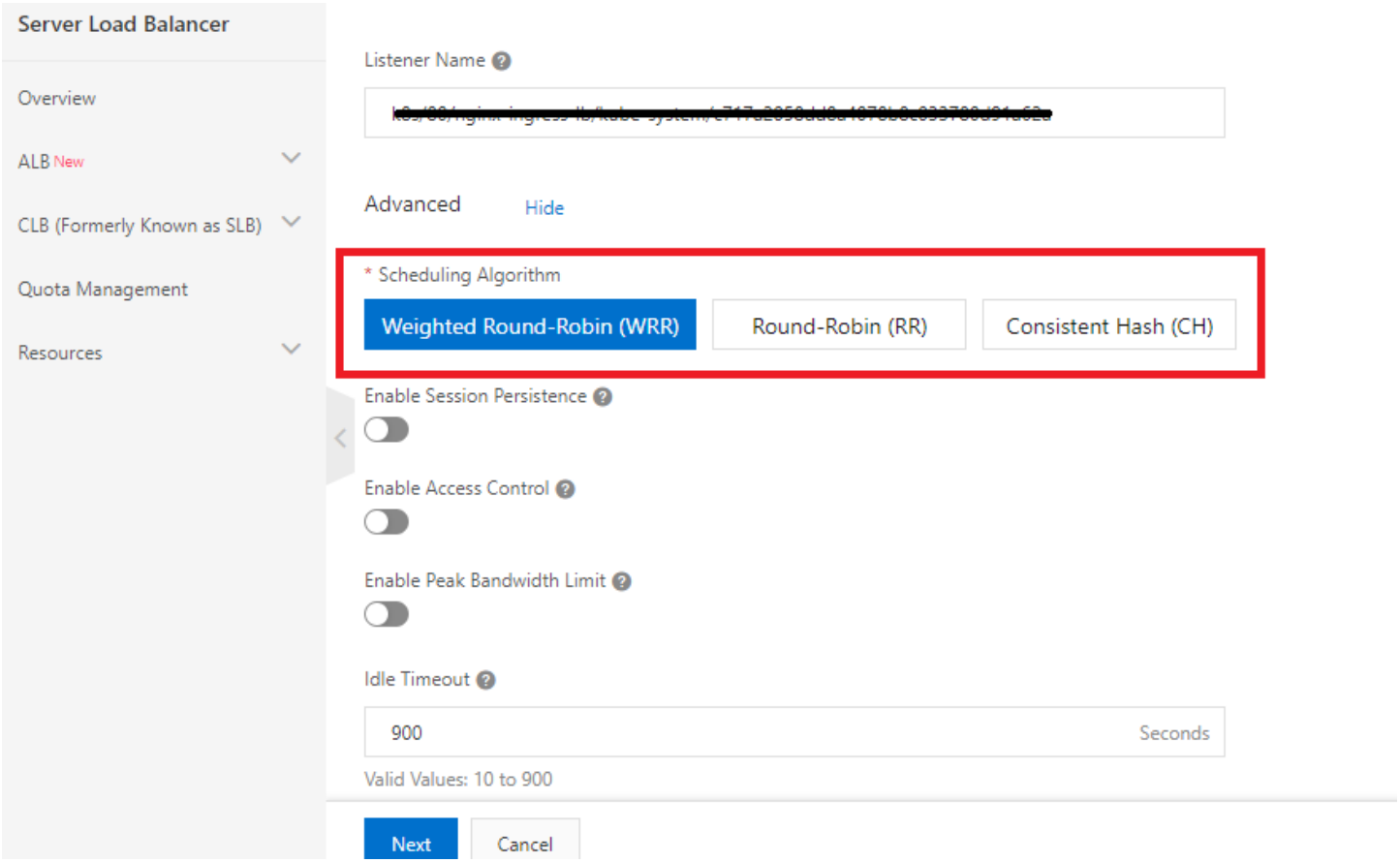
The correct sequence of steps is:

- 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-eright
- 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:



- 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

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. Validright
- 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

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 Splitright

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

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-SQLright
- 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. Integerright
- B. Doubleright
- 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 BIright
- 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 right
- 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-AzDataLakeGen2Itemright
- 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 subscriptionright
- 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.right
- 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

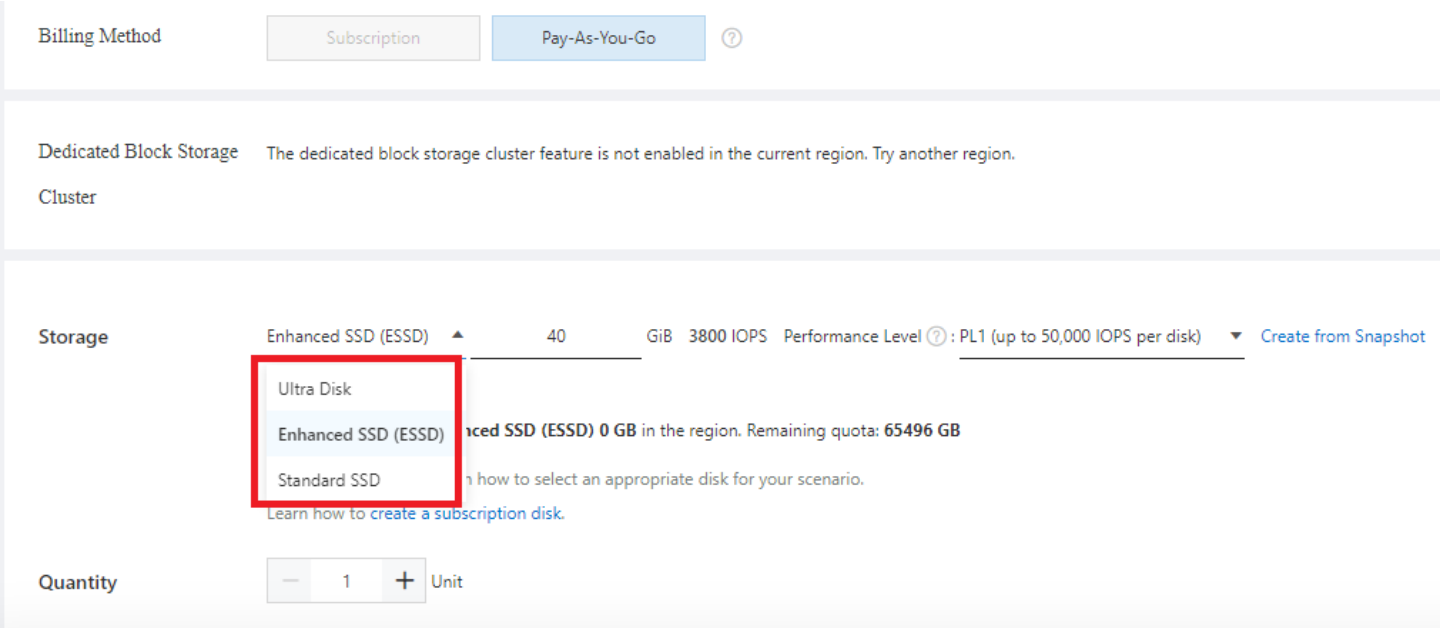
Azure Data Factory allows debugging and troubleshooting the pipelines using Azure PowerShell and Azure portal. You decide to use PowerShell to debug an error. As an important step of the process, you need to retrieve details about the activity run for the slice. Which of the following cmdlet would you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

- A. Get-AzDataFactorySlice
- B. Get-AzDataFactoryRunright
- 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:



- 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 aboveright

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 Filesright
- 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. Noright

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 partition9 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.

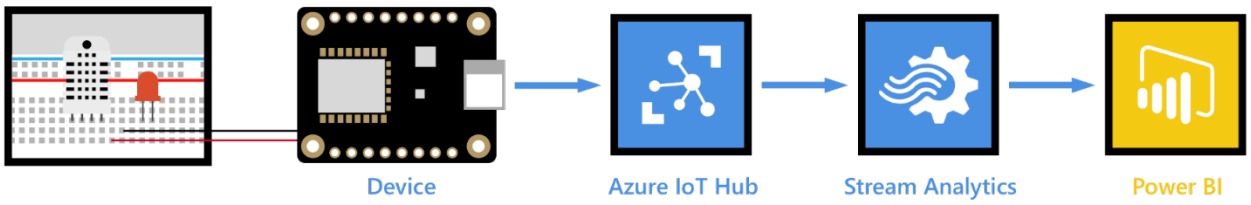
…… TABLE dbo.CustTemporal SET(SYSTEM_VERSIONING = ……)

- A. MODIFY, ON
- B. MODIFY, OFF
- C. MODIFY, NULL
- D. ALTER, OFFright
- 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

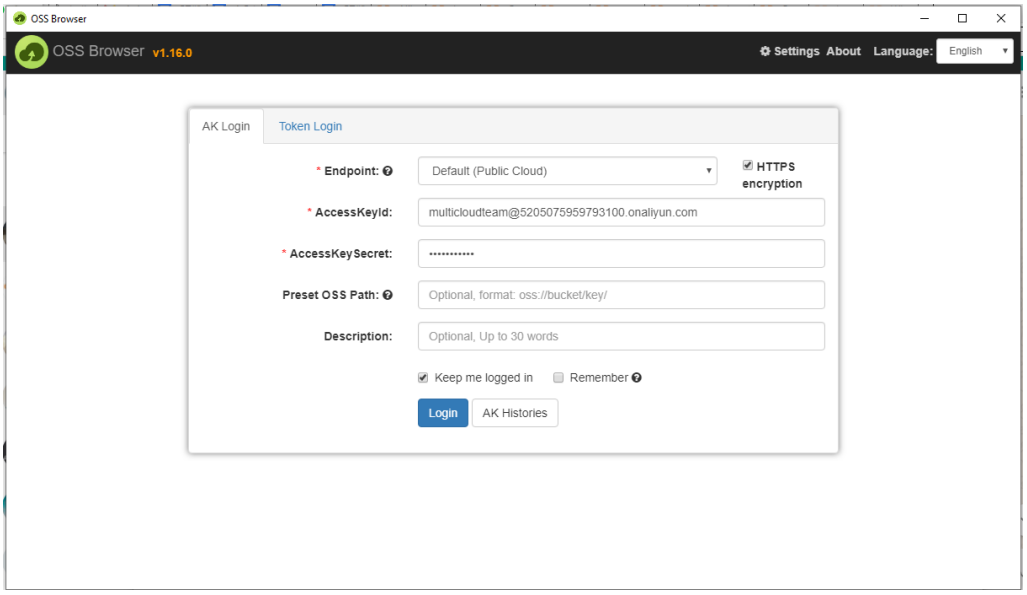
While working on Azure SQL Database, you want to use data compression to decrease the size of the database. From the given options, select the database objects on which you can configure row and page compression?

- 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 aboveright

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

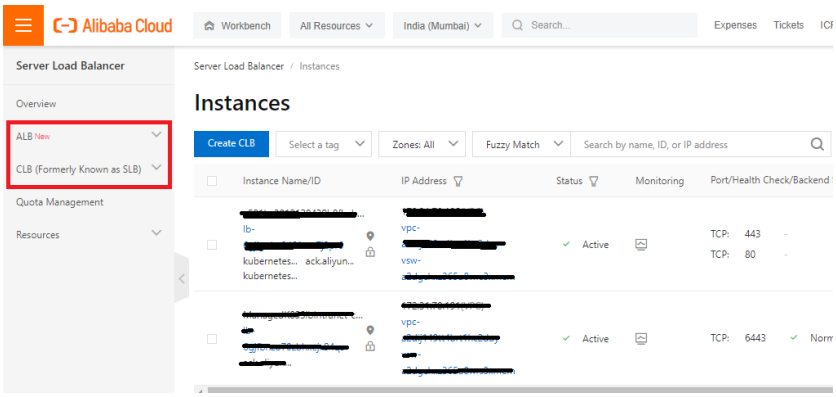
You have folders and files in Azure Blob Storage for the Azure Synapse. You develop a native external table "EtTable1" with LOCATION='/webdata/'. If you query EtTable1 with Azure Synapse Analytics Serverless SQL POOL, choose from the following files that will be returned?

- A. Only mydata.txtright
- 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