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

QUESTION 28
DRAG DROP
You develop data engineering solutions for a company. You must migrate data from Microsoft Azure Blob storage to an Azure SQL Data Warehouse for further transformation. You need to implement the solution.

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

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

Correct Answer:

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

QUESTION 29
DRAG DROP
You are creating a managed data warehouse solution on Microsoft Azure.

You must use PolyBase to retrieve data from Azure Blob storage that resides in parquet format and toad the data into a large table called FactSalesOrderDetails.

You need to configure Azure SQL Data Warehouse to receive the data.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
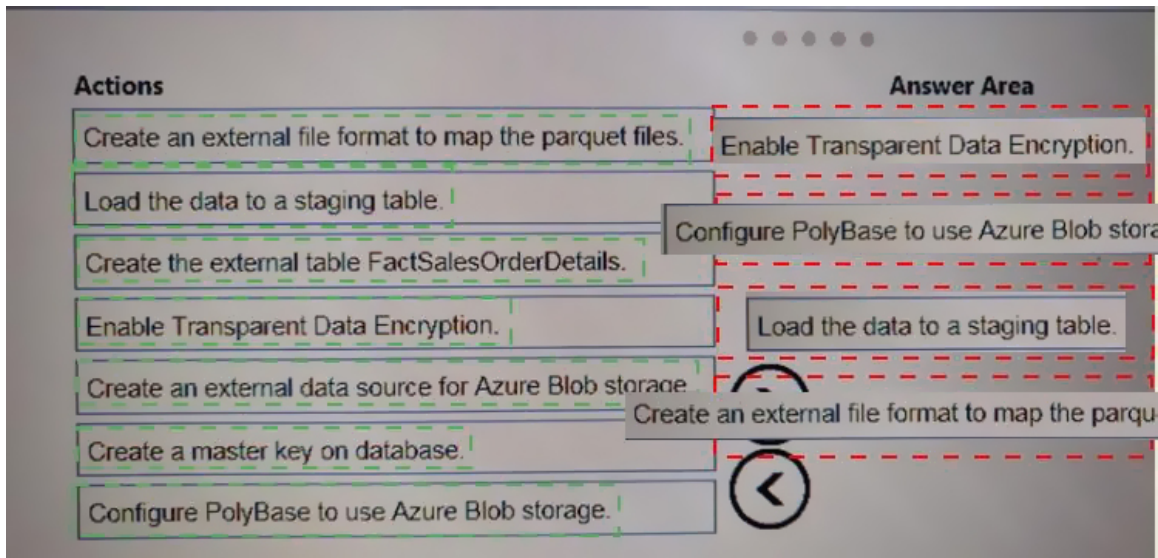
Actions	Answer Area
Create an external file format to map the parquet files.	
Load the data to a staging table.	
Create the external table FactSalesOrderDetails.	
Enable Transparent Data Encryption.	
Create an external data source for Azure Blob storage.	
Create a master key on database.	
Configure PolyBase to use Azure Blob storage.	

Correct Answer:

QUESTION 30
You are the data engineer tor your company. An application uses a NoSQL database to store data. The database uses the key-value and wide-column NoSQL database type.

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

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

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

NOTE: Each correct selection s worth one point.

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

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

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

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

QUESTION 31
A company has a real-lime data analysis solution that is hosted on Microsoft Azure the solution uses Azure Event Hub to ingest data and an Azure Stream Analytics cloud job to analyze the data. The cloud job is configured to use 120 Streaming Units (SU).

You need to optimize performance for the Azure Stream Analytics job.

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

NOTE: Each correct selection is worth one port.

A. Implement event ordering
B. Scale the SU count for the job up
C. Implement Azure Stream Analytics user-defined functions (UDF)
D. Scale the SU count for the job down
E. Implement query parallelization by partitioning the data output
F. Implement query parallelization by partitioning the data input

Correct Answer: BF
Explanation:
Scale out the query by allowing the system to process each input partition separately.

F: A Stream Analytics job definition includes inputs, a query, and output. Inputs are where the job reads the data stream from.

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

QUESTION 32
You configure monitoring for a Microsoft Azure SQL Data Warehouse implementation. The implementation uses PolyBase to load data from comma-separated value (CSV) files stored in Azure Data Lake Gen 2 using an external table.

Files with an invalid schema cause errors to occur.

You need to monitor for an invalid schema error.

For which error should you monitor?

A. EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [com.microsoft.polybase.client.KerberosSecureLogin] occurred while accessing external files.'
B. EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [No FileSystem for scheme: wasbs] occurred while accessing external file.'
C. Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11": for linked server "(null)", Query aborted- the maximum reject threshold (o rows) was reached while regarding from an external source: 1 rows rejected out of total 1 rows processed.
D. EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [Unable to instantiate LoginClass] occurred while accessing external files.'

Correct Answer: C
Explanation:
Customer Scenario:
SQL Server 2016 or SQL DW connected to Azure blob storage. The CREATE EXTERNAL TABLE DDL points to a directory (and not a specific file) and the directory contains files with different schemas.

SSMS Error:

Select query on the external table gives the following error:

Msg 7320, Level 16, State 110, Line 14

Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11" for linked server "(null)". Query aborted-- the maximum reject threshold (0 rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.

Possible Reason:
The reason this error happens is because each file has different schema. The PolyBase external table DDL when pointed to a directory recursively reads all the files in that directory. When a column or data type mismatch happens, this error could be seen in SSMS.

Possible Solution:
If the data for each table consists of one file, then use the filename in the LOCATION section prepended by the directory of the external files. If there are multiple files per table, put each set of files into different directories in Azure Blob Storage and then you can point LOCATION to the directory instead of a particular file. The latter suggestion is the best practices recommended by SQLCAT even if you have one file per table.

QUESTION 33
You are a data architect. The data engineering team needs to configure a synchronization of data between an on-premises Microsoft SQL Server database to Azure SQL Database.

Ad-hoc and reporting queries are being overutilized the on-premises production instance. The synchronization process must:

▪ Perform an initial data synchronization to Azure SQL Database with minimal downtime
▪ Perform bi-directional data synchronization after initial synchronization

You need to implement this synchronization solution.

Which synchronization method should you use?

A. transactional replication
B. Data Migration Assistant (DMA)
C. backup and restore
D. SQL Server Agent job
E. Azure SQL Data Sync

Correct Answer: E
Explanation:
SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you