Microsoft MCSA Certification 70-776 Exam

- **Vendor: Microsoft**
- **Exam Code: 70-776**
- **Exam Name: Perform Big Data Engineering on Microsoft Cloud Services**

QUESTION 1
You are building a Microsoft Azure Stream Analytics job definition that includes inputs, queries, and outputs. You need to create a job that automatically provides the highest level of parallelism to the compute instances. What should you do?

A. Configure event hubs and blobs to use the PartitionKey field as the partition ID.
B. Set the partition key for the inputs, queries, and outputs to use the same partition folders. Configure the queries to use uniform partition keys.
C. Set the partition key for the inputs, queries, and outputs to use the same partition folders. Configure the queries to use different partition keys.
D. Define the number of input partitions to equal the number of output partitions.

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

QUESTION 2
You manage an on-premises data warehouse that uses Microsoft SQL Server. The data warehouse contains 100 TB of data. The data is partitioned by month. One TB of data is added to the data warehouse each month. You create a Microsoft Azure SQL data warehouse and copy the on-premises data to the data warehouse. You need to implement a process to replicate the on-premises data warehouse to the Azure SQL data warehouse. The solution must support daily incremental updates and must provide error handling. What should you use?

A. Azure Import/Export service
B. SQL Server log shipping
C. Azure Data Factory
D. AzCopy utility

Answer: C

QUESTION 3
You plan to use Microsoft Azure Data factory to copy data daily from an Azure SQL data warehouse to an Azure Data Lake Store. You need to define a linked service for the Data Lake Store. The solution must prevent the access token from expiring. Which type of authentication should you use?

A. OAuth
B. service-to-service
C. Basic
D. service principal

Answer: D
Explanation:
https://docs.microsoft.com/en-gb/azure/data-factory/v1/data-factory-azure-datalake-connector#azure-data-lake-store-linked-service-properties

QUESTION 4
You have a Microsoft Azure Data Lake Store and an Azure Active Directory tenant. You are developing an application that will access the Data Lake Store by using end-user credentials. You need to ensure that the application uses end-user authentication to access the Data Lake Store. What should you create?

A. a Native Active Directory app registration

B. a policy assignment that uses the Allowed resource types policy definition
C. a Web app/API Active Directory app registration
D. a policy assignment that uses the Allowed locations policy definition

Answer: A
Explanation:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-end-user-authenticate-using-active-directory

QUESTION 5
You are developing an application that uses Microsoft Azure Stream Analytics. You have data structures that are defined dynamically. You want to enable consistency between the logical methods used by stream processing and batch processing. You need to ensure that the data can be integrated by using consistent data points. What should you use to process the data?

A. a vectorized Microsoft SQL Server Database Engine
B. directed acyclic graph (DAG)
C. Apache Spark queries that use updateStateByKey operators
D. Apache Spark queries that use mapWithState operators

Answer: D

QUESTION 6
You need to use the Cognition.Vision.FaceDetector() function in U-SQL to analyze images. Which attribute can you detect by using the function?

A. gender
B. race
C. weight
D. hair color

Answer: A

QUESTION 7
You have a Microsoft Azure SQL data warehouse that contains information about community events. An Azure Data Factory job writes an updated CSV file in Azure Blob storage to Community/{date}/events.csv daily. You plan to consume a Twitter feed by using Azure Stream Analytics and to correlate the feed to the community events. You plan to use Stream Analytics to retrieve the latest community events data and to correlate the data to the Twitter feed data. You need to ensure that when updates to the community events data is written to the CSV files, the Stream Analytics job can access the latest community events data. What should you configure?

A. an output that uses a blob storage sink and has a path pattern of Community/{date}
B. an output that uses an event hub sink and the CSV event serialization format
C. an input that uses a reference data source and has a path pattern of Community/{date}/events.csv
D. an input that uses a reference data source and has a path pattern of Community/{date}

Answer: C

QUESTION 8

You have a Microsoft Azure Data Lake Analytics service. You need to write a U-SQL query to extract from a CSV file all the users who live in Boston, and then to save the results in a new CSV file. Which U-SQL script should you use?

A.

```
@users =
EXTRACT Name    string,
        Age     int,
        City    string,
        State   string
FROM "/users.csv"
USING Extractors.Csv();

@BostonUsers =
SELECT Name, Age, City, State
FROM @users
WHERE City == "Boston";

OUTPUT @BostonUsers
      TO "/output/BostonUsers.csv"
      USING Outputters.Csv();
```

B.

```
@users =
EXTRACT UserName    string,
        Age         int,
        City        string,
        State       string
FROM '/users.csv'
USING Outputters.Csv();

@BostonUsers =
SELECT UserName, Age, City, State
FROM @users
WHERE City == "Boston";

OUTPUT @BostonUsers
      TO '/output/BostonUsers.csv'
      USING Extractors.Csv();
```

C.

```
@users =
EXTRACT Name    string,
        Age     int,
        City    string,
        State   string
FROM '/users.csv'
USING Extractors.Csv();

@BostonUsers =
SELECT UserName, Age, City, State
FROM @users
WHERE City == "Boston";

OUTPUT @BostonUsers
      TO '/output/BostonUsers.csv'
      USING Outputters.Csv();
```

D.

```
@users =
EXTRACT UserName   string,
        Age        int,
        City       string,
        State      string
From "/users.csv"
Using Extractors.Csv();

@BostonUsers =
SELECT UserName, Age, City, State
From @users
Where City == "Boston";

OUTPUT @BostonUsers
      TO "/output/BostonUsers.csv"
      Using Outputters.Csv();
```

Answer: A

QUESTION 9

You have an on-premises deployment of Active Directory named contoso.com. You plan to deploy a Microsoft Azure SQL data warehouse. You need to ensure that the data warehouse can be accessed by contoso.com users. Which two components should you deploy? (Each correct answer presents part of the solution. Choose two.)

A. Azure AD Privileged Identity Management

B. Azure Information Protection
C. Azure Active Directory
D. Azure AD Connect
E. Cloud App Discovery
F. Azure Active Directory B2C

Answer: CD

QUESTION 10
Note: This question is part of a series of questions that present the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario.
You are migrating an existing on-premises data warehouse named LocalDW to Microsoft Azure. You will use an Azure SQL data warehouse named AzureDW for data storage and an Azure Data Factory named AzureDF for extract, transformation, and load (ETL) functions. For each table in LocalDW, you create a table in AzureDW. On the on-premises network, you have a Data Management Gateway. Some source data is stored in Azure Blob storage. Some source data is stored on an on-premises Microsoft SQL Server instance. The instance has a table named Table1. After data is processed by using AzureDF, the data must be archived and accessible forever. The archived data must meet a Service Level Agreement (SLA) for availability of 99 percent. If an Azure region fails, the archived data must be available for reading always.
End of repeated scenario.
You need to configure Azure Data Factory to connect to the on-premises SQL Server instance. What should you do first?

A. Deploy an Azure virtual network gateway.
B. Create a dataset in Azure Data Factory.
C. From Azure Data Factory, define a data gateway.
D. Deploy an Azure local network gateway.

Answer: C
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-move-data-between-onprem-and-cloud

QUESTION 11
You have a Microsoft Azure SQL data warehouse. The following statements are used to define file formats in the data warehouse:

```
CREATE EXTERNAL FILE FORMAT FileFormat_ORC
WITH (
    FORMAT_TYPE = ORC
, DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);

CREATE EXTERNAL FILE FORMAT FileFormat_PARQUET
WITH (
    FORMAT_TYPE = PARQUET
, DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);
```

You have an external PolyBase table named file_factPowerMeasurement that uses the FileFormat_ORC file format. You need to change file_ factPowerMeasurement to use the FileFormat_PARQUET file format. Which two statements should you execute? (Each correct answer presents part of the solution. Choose two.)

A. CREATE EXTERNAL TABLE
B. ALTER TABLE
C. CREATE EXTERNAL TABLE AS SELECT
D. ALTER EXTERNAL DATA SOURCE
E. DROP EXTERNAL TABLE

Answer: AE

QUESTION 12
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are troubleshooting a slice in Microsoft Azure Data Factory for a dataset that has been in a waiting state for the last three days. The dataset should have been ready two days ago. The dataset is being produced outside the scope of Azure Data Factory. The dataset is defined by using the following JSON code:

```
{
"name": "CustomerTable",
"properties": {
"type": "AzureBlob",
"linkedServiceName": "MyLinkedService",
"typeProperties": {
"folderPath": "MyContainer/MySubFolder/",
"format": {
"type": "TextFormat",
"columnDelimiter": ",",
"rowDelimiter": ";"
}
},
"external": false,
"availability": {
"frequency": "Hour",
"interval": 1
},
"policy": {
}
}
}
}
```

You need to modify the JSON code to ensure that the dataset is marked as ready whenever there is data in the data store.
Solution: You change the interval to 24.
Does this meet the goal?

A. Yes
B. No

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-create-datasets

QUESTION 13
You plan to deploy a Microsoft Azure virtual machine that will a host data warehouse. The data warehouse will contain a 10 TB database. You need to provide the fastest read and writes times for the database. Which disk configuration should you use?

A. storage pools with mirrored disks
B. RAID 5 volumes
C. spanned volumes
D. stripped volumes
E. storage pools with striped disks

Answer: E

QUESTION 14

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are troubleshooting a slice in Microsoft Azure Data Factory for a dataset that has been in a waiting state for the last three days. The dataset should have been ready two days ago. The dataset is being produced outside the scope of Azure Data Factory. The dataset is defined by using the following JSON code:

```
{
"name": "CustomerTable",
"properties": {
"type": "AzureBlob",
"linkedServiceName": "MyLinkedService",
"typeProperties": {
"folderPath": "MyContainer/MySubFolder/",
"format": {
"type": "TextFormat",
"columnDelimiter": ",",
"rowDelimiter": ";"
}
},
"external": false,
"availability": {
"frequency": "Hour",
"interval": 1
},
"policy": {
}
}
}
}
```

You need to modify the JSON code to ensure that the dataset is marked as ready whenever there is data in the data store.
Solution: You add a structure property to the dataset.
Does this meet the goal?

A. Yes
B. No

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-create-datasets

QUESTION 15

You have a fact table named PowerUsage that has 10 billion rows. PowerUsage contains data about customer power usage during the last 12 months. The usage data is collected every minute. PowerUsage contains the columns configured as shown in the following table:

Column name	Data type	Nullable
MeasurementId	bigint	No
CustomerId	int	No
LocationNumber	int	No
MinuteOfMonth	int	No
MonthKey	int	No
Usage	int	Yes

LocationNumber has a default value of 1. The MinuteOfMonth column contains the relative minute within each month. The value resets at the beginning of each month. A sample of the fact table data is shown in the following table:

Measurement Id	CustomerId	Location Number	MinuteOf Month	MonthKey	Usage
1	1	1	1	1	100
2	1	1	2	1	66
3	2	2	1	1	88
4	1	1	1	2	93
5	1	1	2	2	0
6	2	2	1	2	47
7	1	1	1	2	52
8	1	1	2	2	22

There is a related table named Customer that joins to the PowerUsage table on the CustomerId column. Sixty percent of the rows in PowerUsage are associated to less than 10 percent of the rows in Customer. Most queries do not require the use of the Customer table. Many queries select on a specific month. You need to minimize how long it takes to find the records for a specific month. What should you do?

A. Implement partitioning by using the MonthKey column.
Implement hash distribution by using the CustomerId column.
B. Implement partitioning by using the CustomerId column.
Implement hash distribution by using the MonthKey column.
C. Implement partitioning by using the MonthKey column.
Implement hash distribution by using the MeasurementId column.
D. Implement partitioning by using the MinuteOfMonth column.
Implement hash distribution by using the MeasurementId column.

Answer: C

QUESTION 16

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a table named Table1 that contains 3 billion rows. Table1 contains data from the last 36

months. At the end of every month, the oldest month of data is removed based on a column named DateTime. You need to minimize how long it takes to remove the oldest month of data.
Solution: You specify DateTime as the hash distribution column.
Does this meet the goal?

A. Yes
B. No

Answer: B

QUESTION 17
You have an extract, transformation, and load (ETL) process for a Microsoft Azure SQL data warehouse. You run the following statements to create the logon and user for an account that will run the nightly data load for the data warehouse:

```
CREATE LOGIN LoaderLogin WITH PASSWORD = 'mypassword';
CREATE USER LoaderUser for LOGIN LoaderLogin;
```

You connect to the data warehouse. You need to ensure that the user can access the highest resource class. Which statement should you execute?

A. ALTER SERVER ROLE xLargeRC ADD MEMBER LoaderLogin;
B. EXEC sp_addrolemember 'xlargerc', 'LoaderUser'
C. ALTER SERVER ROLE LargeRC ADD MEMBER LoaderUser;
D. EXEC sp_addrolemember 'largerc', 'LoaderLogin'

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-develop-concurrency

QUESTION 18
You have a Microsoft Azure Data Lake Analytics service. You plan to configure diagnostic logging. You need to use Microsoft Operations Management Suite (OMS) to monitor the IP addresses that are used to access the Data Lake Store. What should you do?

A. Stream the request logs to an event hub.
B. Send the audit logs to Log Analytics.
C. Send the request logs to Log Analytics.
D. Stream the audit logs to an event hub.

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-diagnostic-logs
https://docs.microsoft.com/en-us/azure/security/azure-log-audit

QUESTION 19
You plan to add a file from Microsoft Azure Data Lake Store to Azure Data Catalog. You run the Data Catalog tool and select Data Lake Store as the data source. Which information should you enter in the Store Account field to connect to the Data Lake Store?

A. an email alias
B. a server name
C. a URL

D. a subscription ID

Answer: C

QUESTION 20
You have a Microsoft Azure SQL data warehouse that has a fact table named FactOrder. FactOrder contains three columns named CustomerId, OrderId, and OrderDateKey. FactOrder is hash distributed on CustomerId. OrderId is the unique identifier for FactOrder. FactOrder contains 3 million rows. Orders are distributed evenly among different customers from a table named dimCustomers that contains 2 million rows. You often run queries that join FactOrder and dimCustomers by selecting and grouping by the OrderDateKey column. You add 7 million rows to FactOrder. Most of the new records have a more recent OrderDateKey value than the previous records. You need to reduce the execution time of queries that group on OrderDateKey and that join dimCustomers and FactOrder. What should you do?

A. Change the distribution for the FactOrder table to round robin.
B. Update the statistics for the OrderDateKey column.
C. Change the distribution for the FactOrder table to be based on OrderId.
D. Change the distribution for the dimCustomers table to OrderDateKey.

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-statistics

QUESTION 21
You need to define an input dataset for a Microsoft Azure Data Factory pipeline. Which properties should you include when you define the dataset?

A. name, type, typeProperties, and availability
B. name, typeProperties, structure, and availability
C. name, policy, structure, and external
D. name, type, policy, and structure

Answer: A
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-create-datasets

QUESTION 22
You have a file in a Microsoft Azure Data Lake Store that contains sales data. The file contains sales amounts by salesperson, by city, and by state. You need to use U-SQL to calculate the percentage of sales that each city has for its respective state. Which code should you use?

A.

```
@result=
    SELECT
        City, State,
        SUM(SalesAmount)
            OVER( PARTITION BY City ) / SUM(SalesAmount)
            OVER( PARTITION BY State )
        AS CitySalesPercent
    FROM @Sales;
```

B.

```
@result=
    SELECT City, SUM(SalesAmount)
    AS CitySalesPercent
    FROM @Sales;
    GROUP BY City;
```

C.

```
@result=
    SELECT
        Salesperson, City, State,
        SUM(SalesAmount)
            OVER( PARTITION BY City ) / SUM(SalesAmount)
            OVER()
        AS CitySalesPercent
    FROM @Sales;
```

D.

```
@result=
    SELECT
        City, State,
        SUM(SalesAmount)
            OVER( ) / SUM(SalesAmount)
            OVER( )
        AS CitySalesPercent
    FROM @Sales;
```

Answer: A

QUESTION 23

You have an on-premises data warehouse that uses Microsoft SQL Server 2016. All the data in the data warehouse comes from text files stored in Azure Blob storage. The text files are imported into the data warehouse by using SQL Server Integration Services (SSIS). The text files are not transformed. You need to migrate the data to an Azure SQL data warehouse in the least amount of time possible. Which two actions should you perform? (Each correct answer presents part of the solution. Choose two.)

A. Use SSIS to upload the files in Azure Blob storage to tables in the Azure SQL data warehouse.

B. Execute the CREATE EXTERNAL TABLE AS SELECT statement to export the data.
C. Use AzCopy to transfer the data from the on-premises data warehouse to Azure SQL data warehouse.
D. Execute the CREATE TABLE AS SELECT statement to load the data.
E. Define external tables in the Azure SQL data warehouse that map to the existing files in Azure Blob storage.

Answer: DE
Explanation:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-blob-storage-with-polybase

QUESTION 24
You have a Microsoft Azure Data Factory that recently ran several activities in parallel. You receive alerts indicating that there are insufficient resources. From the Activity Windows list in the Monitoring and Management app, you discover the statuses described in the following table:

Activity name	Status	Substatus
Activity1	Failed	Canceled
Activity2	Waiting	DatasetDependencies
Activity3	Waiting	ComputeResources
Activity4	Waiting	ActivityResume
Activity5	Waiting	ConcurrencyLimit
Activity6	Skipped	Not applicable
Activity7	In progress	Validating
Activity8	Waiting	ValidationRetry

Which activity cannot complete because of insufficient resources?

A. Activity2
B. Activity4
C. Activity5
D. Activity7

Answer: C

QUESTION 25
Note: This question is part of a series of questions that present the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario.
You are migrating an existing on-premises data warehouse named LocalDW to Microsoft Azure. You will use an Azure SQL data warehouse named AzureDW for data storage and an Azure Data Factory named AzureDF for extract, transformation, and load (ETL) functions. For each table in LocalDW, you create a table in AzureDW. On the on-premises network, you have a Data Management Gateway. Some source data is stored in Azure Blob storage. Some source data is stored on an on-premises Microsoft SQL Server instance. The instance has a table named Table1. After data is processed by using AzureDF, the data must be archived and accessible forever. The archived data must meet a Service Level Agreement (SLA) for availability of 99 percent. If an Azure region fails, the archived data must be available for reading always. The storage solution for the archived data must minimize costs.

End of repeated scenario.
You need to define the schema of Table1 in AzureDF. What should you create?

A. a gateway
B. a linked service
C. a dataset
D. a pipeline

Answer: C

QUESTION 26
You are developing an application by using the Microsoft .NET SDK. The application will access data from a Microsoft Azure Data Lake folder. You plan to authenticate the application by using service-to-service authentication. You need to ensure that the application can access the Data Lake folder. Which three actions should you perform? (Each correct answer presents part of the solution. Choose three.)

A. Register an Azure Active Directory app that uses the Web app/API application type.
B. Configure the application to use the application ID, authentication key, and tenant ID.
C. Assign the Azure Active Directory app permission to the Data Lake Store folder.
D. Configure the application to use the OAuth 2.0 token endpoint.
E. Register an Azure Active Directory app that uses the Native application type.
F. Configure the application to use the application ID and redirect URI.

Answer: ABC
Explanation:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-service-to-service-authenticate-using-active-directory

QUESTION 27
You ingest data into a Microsoft Azure event hub. You need to export the data from the event hub to Azure Storage and to prepare the data for batch processing tasks in Azure Data Lake Analytics. Which two actions should you perform? (Each correct answer presents part of the solution. Choose two.)

A. Run the Avro extractor from a U-SQL script.
B. Create an Azure Storage account.
C. Add a shared access policy.
D. Enable Event Hubs Archive.
E. Run the CSV extractor from a U-SQL script.

Answer: BD

QUESTION 28
You are creating a Data Factory workflow. The workflow must run a stored procedure from a SQL database, perform a Hive query, and copy data from an on-premises SQL database into an Azure SQL database. These steps must be managed and performed as an atomic unit. You need to configure the workflow. Which Data Factory entity should you use?

A. Dataset
B. Linked service
C. Activity

D. Pipeline

Answer: D

QUESTION 29
You are designing an Azure Data Lake managed cluster for your organization's realtime data processing. The streaming data could range from 10 MB to 100 GB per message. You need to choose the correct storage model to store this data. Which storage model should you use?

A. Block Blobs
B. Page Blobs
C. HBase
D. Azure Storage Queues

Answer: A

QUESTION 30
You have a Microsoft Azure SQL data warehouse. You have an Azure Data Lake Store that contains data from ORC, RC, Parquet, and delimited text files. You need to load the data to the data warehouse in the least amount of time possible. Which two actions should you perform? (Each correct answer presents part of the solution. Choose two.)

A. Use Microsoft SQL Server Integration Services (SSIS) to enumerate from the Data Lake Store by using a for loop.
B. Use AzCopy to export the files from the Data Lake Store to Azure Blob storage.
C. For each file in the loop, export the data to Parallel Data Warehouse by using a Microsoft SQL Server Native Client destination.
D. Load the data by executing the CREATE TABLE AS SELECT statement.
E. Use bcp to import the files.
F. In the data warehouse, configure external tables and external file formats that correspond to the Data Lake Store.

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

QUESTION 31
Note: This question is part of a series of questions that present the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario.
You are migrating an existing on-premises data warehouse named LocalDW to Microsoft Azure. You will use an Azure SQL data warehouse named AzureDW for data storage and an Azure Data Factory named AzureDF for extract, transformation, and load (ETL) functions. For each table in LocalDW, you create a table in AzureDW. On the on-premises network, you have a Data Management Gateway. Some source data is stored in Azure Blob storage. Some source data is stored on an on-premises Microsoft SQL Server instance. The instance has a table named Table1. After data is processed by using AzureDF, the data must be archived and accessible forever. The archived data must meet a Service Level Agreement (SLA) for availability of 99 percent. If an Azure region fails, the archived data must be available for reading always.
End of repeated scenario.
You need to connect AzureDF to the storage account. What should you create?

A. a gateway
B. a dataset
C. a linked service
D. a pipeline

Answer: C
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-azure-blob-connector

QUESTION 32
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are monitoring user queries to a Microsoft Azure SQL data warehouse that has six compute nodes. You discover that compute node utilization is uneven. The rows_processed column from sys.dm_pdw_workers shows a significant variation in the number of rows being moved among the distributions for the same table for the same query. You need to ensure that the load is distributed evenly across the compute nodes.
Solution: You add a clustered columnstore index.
Does this meet the goal?

A. Yes
B. No

Answer: B

QUESTION 33
You have a Microsoft Azure subscription that contains an Azure Data Factory pipeline. You have an RSS feed that is published on a public website. You need to configure the RSS feed as a data source for the pipeline. Which type of linked service should you use?

A. Web
B. OData
C. Azure Search
D. Azure Data Lake Store

Answer: A

QUESTION 34
You have sensor devices that report data to Microsoft Azure Stream Analytics. Each sensor reports data several times per second. You need to create a live dashboard in Microsoft Power BI that shows the performance of the sensor devices. The solution must minimize lag when visualizing the data. Which function should you use for the time-series data element?

A. LAG
B. SlidingWindow
C. System.TimeStamp
D. TumblingWindow

Answer: D

QUESTION 35
You have a Microsoft Azure SQL data warehouse that has 10 compute nodes. You need to export 10 TB of data from a data warehouse table to several new flat files in Azure Blob storage. The solution must maximize the use of the available compute nodes. What should you do?

A. Use the bcp utility.
B. Execute the CREATE EXTERNAL TABLE AS SELECT statement.
C. Create a Microsoft SQL Server Integration Services (SSIS) package that has a data flow task.
D. Create a Microsoft SQL Server Integration Services (SSIS) package that has an SSIS Azure Blob Storage task.

Answer: D

QUESTION 36
You plan to use Microsoft Azure Event Hubs to ingest sensor data. You plan to use Azure Stream Analytics to analyze the data in real time and to send the output directly to Azure Data Lake Store. You need to write events to the Data Lake Store in batches. What should you use?

A. Apache Storm in Azure HDInsight
B. Stream Analytics
C. Microsoft SQL Server Integration Services (SSIS)
D. The Azure CLI

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-data-scenarios

QUESTION 37
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a table named Table1 that contains 3 billion rows. Table1 contains data from the last 36 months. At the end of every month, the oldest month of data is removed based on a column named DateTime. You need to minimize how long it takes to remove the oldest month of data.
Solution: You implement a columnstore index on the DateTime column.
Does this meet the goal?

A. Yes
B. No

Answer: A

QUESTION 38
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a table named Table1 that contains 3 billion rows. Table1 contains data from the last 36 months. At the end of every month, the oldest month of data is removed based on a column named DateTime. You need to minimize how long it takes to remove the oldest month of data.

Solution: You implement round robin for table distribution.
Does this meet the goal?

A. Yes
B. No

Answer: B

QUESTION 39
You are using a Microsoft Azure Stream Analytics query language. You are outputting data from an input click stream. You need to ensure that when you consecutively receive two rows from the same IP address within one minute, only the first row is outputted. Which functions should you use in the WHERE statement?

A. Last and HoppingWindow
B. Last and SlidingWindow
C. LAG and HoppingWindow
D. LAG and Duration

Answer: B

QUESTION 40
You are designing a solution that will use Apache HBase on Microsoft Azure HDInsight. You need to design the row keys for the database to ensure that client traffic is directed over all of the nodes in the cluster. What are two possible techniques that you can use? (Each correct answer presents a complete solution. Choose two.)

A. padding
B. trimming
C. hashing
D. salting

Answer: AC

QUESTION 41
You have a Microsoft Azure Data Factory pipeline. You discover that the pipeline fails to execute because data is missing. You need to rerun the failure in the pipeline. Which cmdlet should you use?

A. `Set-AzureAutomationJob`
B. `Resume-AzureDataFactoryPipeline`
C. `Resume-AzureAutomationJob`
D. `Set-AzureDataFactotySliceStatus`

Answer: B

QUESTION 42
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are monitoring user queries to a Microsoft Azure SQL data warehouse that has six compute nodes. You discover that compute node utilization is uneven. The rows_processed column from sys.dm_pdw_workers shows a significant variation in the number of rows being moved among the distributions for the same table for the same query. You need to ensure that the load is distributed evenly across the compute nodes.
Solution: You change the table to use a column that is not skewed for hash distribution.
Does this meet the goal?

A. Yes
B. No

Answer: A

QUESTION 43
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are monitoring user queries to a Microsoft Azure SQL data warehouse that has six compute nodes. You discover that compute node utilization is uneven. The rows_processed column from sys.dm_pdw_workers shows a significant variation in the number of rows being moved among the distributions for the same table for the same query. You need to ensure that the load is distributed evenly across the compute nodes.
Solution: You add a nonclustered columnstore index.
Does this meet the goal?

A. Yes
B. No

Answer: B

QUESTION 44
You have a Microsoft Azure SQL data warehouse named DW1 that is used only from Monday to Friday. You need to minimize Data Warehouse Unit (DWU) usage during the weekend. What should you do?

A. From the Azure CLI, run the account set command.
B. Run the ALTER DATABASE statement.
C. Call the Create or Update Database REST API.
D. Run the Suspend-AzureRmSqlDatabase Azure PowerShell cmdlet.

Answer: D

QUESTION 45
You have a Microsoft Azure Data Lake Analytics service. You have a CSV file that contains employee salaries. You need to write a U-SQL query to load the file and to extract all the employees who earn salaries that are greater than $100,000. You must encapsulate the data for reuse. What should you use?

A. a table-valued function
B. a view
C. the extract command

D. the output command

Answer: A
Explanation:
https://docs.microsoft.com/en-au/azure/data-lake-analytics/data-lake-analytics-u-sql-catalog

QUESTION 46
You have a Microsoft Azure Data Lake Analytics service and an Azure Data Lake Store. You need to use Python to submit a U-SQL job. Which Python module should you install?

A. `azure-mgmt-datalake-store`
B. `azure-mgmt-datalake-analytics`
C. `azure-datalake-store`
D. `azure-mgmt-resource`

Answer: B
Explanation:
https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-manage-use-python-sdk

QUESTION 47
You have a Microsoft Azure Data Lake Analytics service. You need to provide a user with the ability to monitor Data Lake Analytics jobs. The solution must minimize the number of permissions assigned to the user. Which role should you assign to the user?

A. Reader
B. Owner
C. Contributor
D. Data Lake Analytics Developer

Answer: A
Explanation:
https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-manage-use-portal

QUESTION 48
You plan to use Microsoft Azure Event Hubs to ingest data. You plan to use Azure Stream Analytics to analyze the data in real time and to send the output directly to Azure Data Lake Store. You discover duplicate records in the output data. What is a possible cause of the duplicate records?

A. There are connectivity issues with the output adapter.
B. There is a connectivity issue between the data source and the event hub.
C. There are multiple deliveries to the output adapter that writes the output events.
D. The Stream Analytics output adapter writes the output events transactionally.

Answer: A
Explanation:
https://msdn.microsoft.com/en-us/library/azure/mt721300.aspx

QUESTION 49
You have an on-premises Microsoft SQL Server instance. You plan to copy a table from the instance to a Microsoft Azure Storage account. You need to ensure that you can copy the table by using Azure Data Factory. Which service should you deploy?

A. An on-premises Data Gateway
B. Azure Application Gateway
C. Data Management Gateway
D. A Virtual Network Gateway

Answer: C

QUESTION 50
You are using Cognitive capabilities in U-SQL to analyze images that contain different types of objects. You need to identify which objects might be people. Which two reference assemblies should you use? (Each correct answer presents part of the solution. Choose two.)

A. ExtPython
B. ImageCommon
C. ImageTagging
D. ExtR
E. FaceSdk

Answer: BC

QUESTION 51
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You are troubleshooting a slice in Microsoft Azure Data Factory for a dataset that has been in a waiting state for the last three days. The dataset should have been ready two days ago. The dataset is being produced outside the scope of Azure Data Factory. The dataset is defined by using the following JSON code:

```
{
"name": "CustomerTable",
"properties": {
"type": "AzureBlob",
"linkedServiceName": "MyLinkedService",
"typeProperties": {
"folderPath": "MyContainer/MySubFolder/",
"format": {
"type": "TextFormat",
"columnDelimiter": ",",
"rowDelimiter": ";"
}
},
"external": false,
"availability": {
"frequency": "Hour",
"interval": 1
},
"policy": {
}
}
}
}
```

You need to modify the JSON code to ensure that the dataset is marked as ready whenever there is data in the data store.
Solution: You change the external attribute to true.
Does this meet the goal?

A. Yes
B. No

Answer: A
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-create-datasets

QUESTION 52
You have a Microsoft Azure SQL data warehouse. You need to configure Data Warehouse Units (DWUs) to ensure that you have six compute nodes. The solution must minimize costs. Which value should set for the DWUs?

A. DW200
B. DW400
C. DW600
D. DW1000

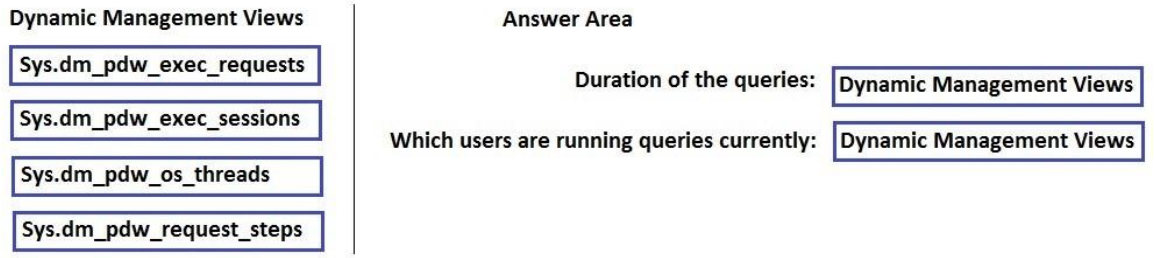
Answer: C
Explanation:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-manage-compute-overview

QUESTION 53
Note: This question is part of a series of questions that present the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is exactly the same in each question in this series.
Start of repeated scenario.
You are migrating an existing on-premises data warehouse named LocalDW to Microsoft Azure. You will use an Azure SQL data warehouse named AzureDW for data storage and an Azure Data Factory named AzureDF for extract, transformation, and load (ETL) functions. For each table in LocalDW, you create a table in AzureDW. On the on-premises network, you have a Data Management Gateway. Some source data is stored in Azure Blob storage. Some source data is stored on an on-premises Microsoft SQL Server instance. The instance has a table named Table1. After data is processed by using AzureDF, the data must be archived and accessible forever. The archived data must meet a Service Level Agreement (SLA) for availability of 99 percent. If an Azure region fails, the archived data must be available for reading always.
End of repeated scenario.
You need to configure an activity to move data from blob storage to AzureDW. What should you create?

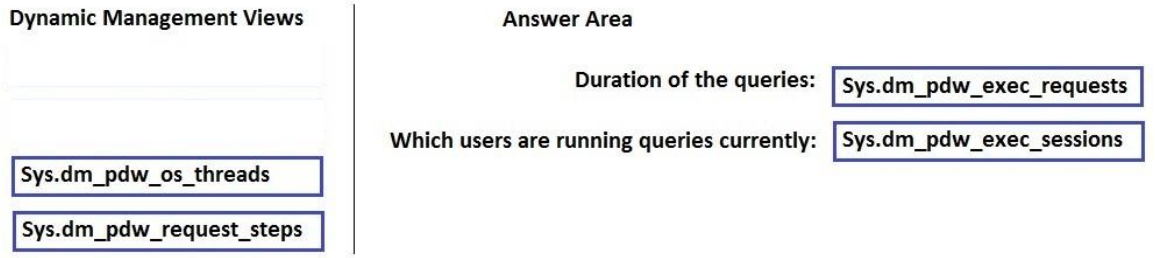
A. a pipeline
B. a linked service
C. an automation runbook
D. a dataset

Answer: A
Explanation:
https://docs.microsoft.com/en-us/azure/data-factory/v1/data-factory-azure-blob-connector

QUESTION 54
Drag and Drop
You have a Microsoft Azure SQL data warehouse. Users discover that reports running in the data warehouse take longer than expected to complete. You need to review the duration of the queries and which users are running the queries currently. Which dynamic management view should you review for each requirement? (To answer, drag the appropriate dynamic management views to the correct requirements. Each dynamic management view may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.)



Answer:



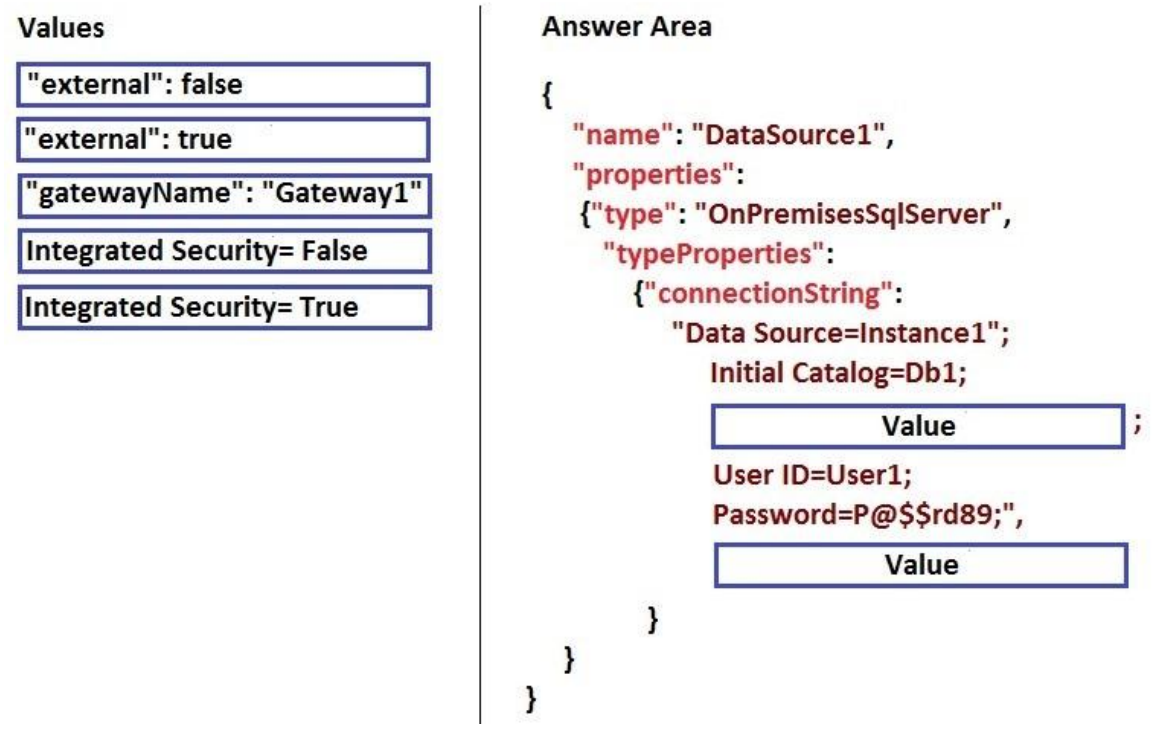
Explanation:

https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-pdw-exec-requests-transact-sql
https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-pdw-exec-sessions-transact-sql

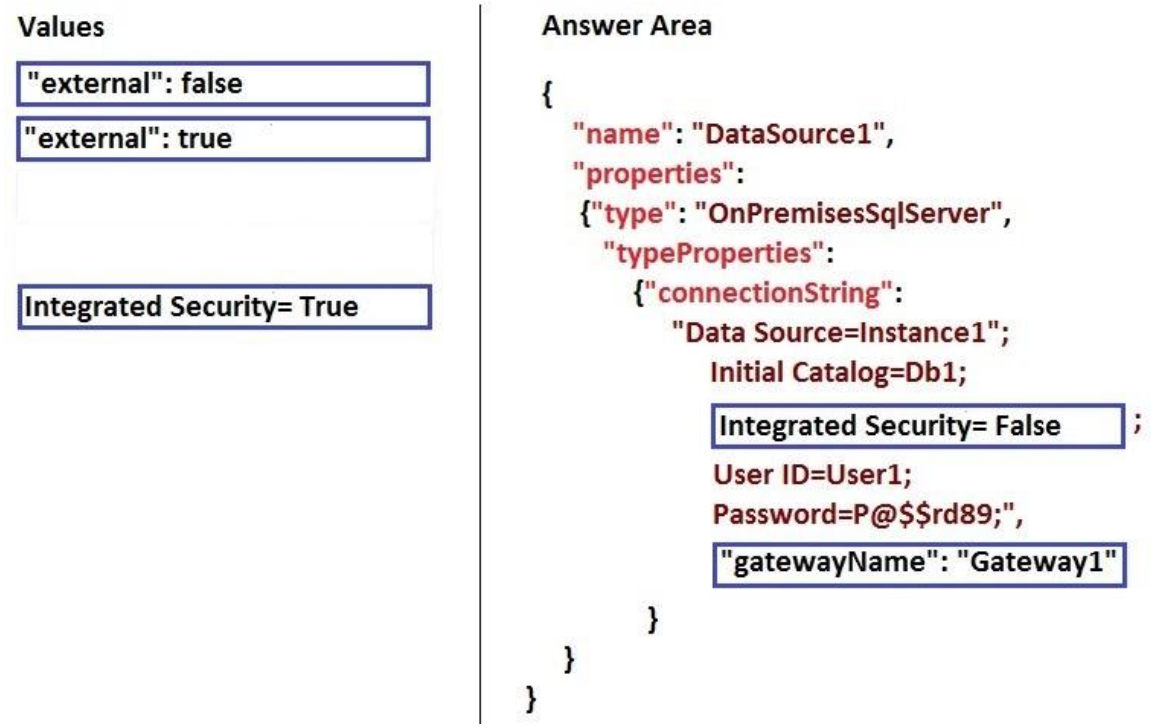
QUESTION 55

Drag and Drop

You have an on-premises Microsoft SQL Server instance named Instance1 that contains a database named DB1. You have a Data Management Gateway named Gateway1. You plan to create a linked service in Azure Data Factory for DB1. You need to connect to DB1 by using standard SQL Server Authentication. You must use a username of User1 and a password of P@$$w0rd89. How should you complete the JSON code? (To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.)



Answer:



Explanation:
https://github.com/uglide/azure-content/blob/master/articles/data-factory/data-factory-move-data-between-onprem-and-cloud.md

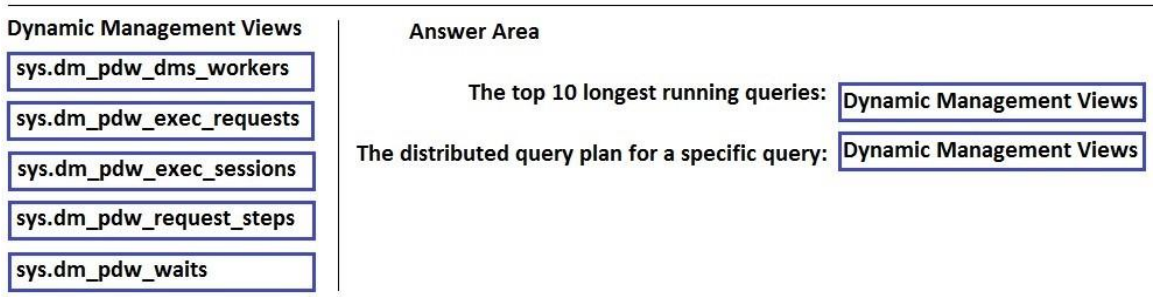
QUESTION 56

Drag and Drop

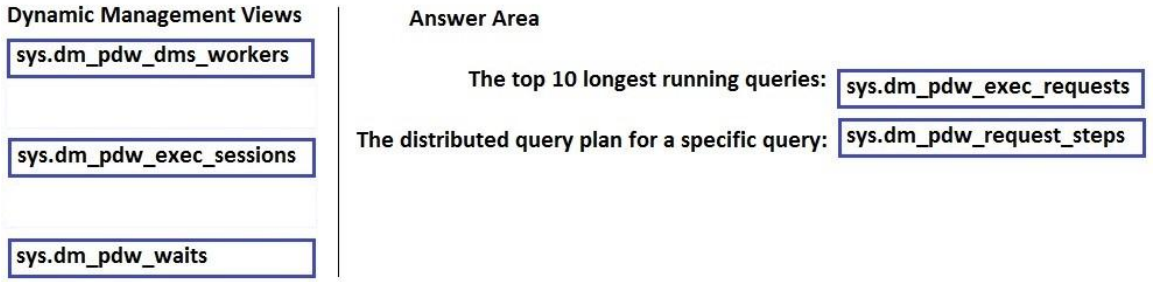
You are monitoring a Microsoft Azure SQL data warehouse. You need to get the following information:

- The top 10 longest running queries
- The distributed query plan for a specific query

Which dynamic management view should you use for each piece of information? (To answer, drag the appropriate dynamic management views to the correct pieces of information. Each dynamic management view may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.)



Answer:



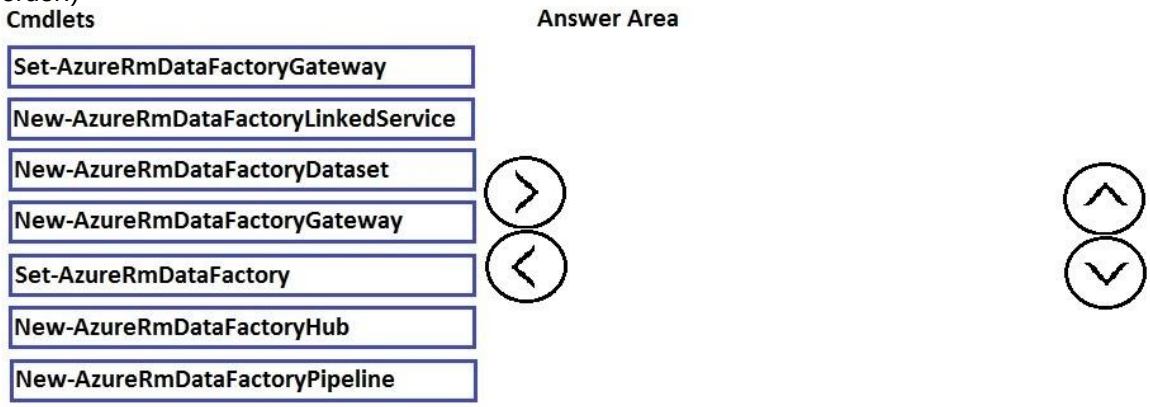
Explanation:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-manage-monitor

QUESTION 57

Drag and Drop

You have an Apache Hive database in a Microsoft Azure HDInsight cluster. You create an Azure Data Factory named DF1. You need to transform the data in the Hive database and to output the data to Azure Blob storage. Which three cmdlets should you run in sequence? (To answer, move the appropriate cmdlets from the list of cmdlets to the answer area and arrange them in the correct order.)



Answer:

Cmdlets

- Set-AzureRmDataFactoryGateway
- New-AzureRmDataFactoryGateway
- Set-AzureRmDataFactory
- New-AzureRmDataFactoryHub

Answer Area

- New-AzureRmDataFactoryLinkedService
- New-AzureRmDataFactoryDataset
- New-AzureRmDataFactoryPipeline

Explanation:

https://docs.microsoft.com/en-us/powershell/module/azurerm.datafactories/new-azurermdatafactorypipeline?view=azurermps-4.4.0
https://github.com/aelij/azure-content/blob/master/articles/data-factory/data-factory-build-your-first-pipeline-using-powershell.md

QUESTION 58

Drag and Drop

You have a Microsoft Azure Stream Analytics solution that captures website visits and user interactions on the website. You have the sample input data described in the following table:

username	feature	EventType	EventTime
User1@contoso.com	Shopping cart	Start	2017-01-01T00:00:01.0000000Z
User1@contoso.com	Shopping cart	End	2017-01-01T00:00:01.0000000Z

You have the sample output described in the following table:

username	feature	duration in sec
User1@contoso.com	Shopping cart	7

How should you complete the script? (To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.)

Values

- DATEADD
- DURATION
- LAG
- DATEDIFF
- FIRST(EventTime)
- LAST(EventTime)

Answer Area

```
SELECT
        username,
        feature,
        [Value] (second,
            [Value] OVER (
        PARTITION BY username,
        feature LIMIT [Value] (hour, 1) WHEN EventType = 'start'),
        EventTime)
        as 'duration in sec'
    FROM input TIMESTAMP BY Time
        WHERE EventType = 'end'
```

Answer:

Values	Answer Area
DATEADD	SELECT username, feature, DATEDIFF (second, LAST(EventTime) OVER (PARTITION BY username, feature LIMIT DURATION (hour, 1) WHEN EventType = 'start'), EventTime) as 'duration in sec' FROM input TIMESTAMP BY Time WHERE EventType = 'end'
LAG	
FIRST(EventTime)	

Explanation:

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

QUESTION 59

Drag and Drop

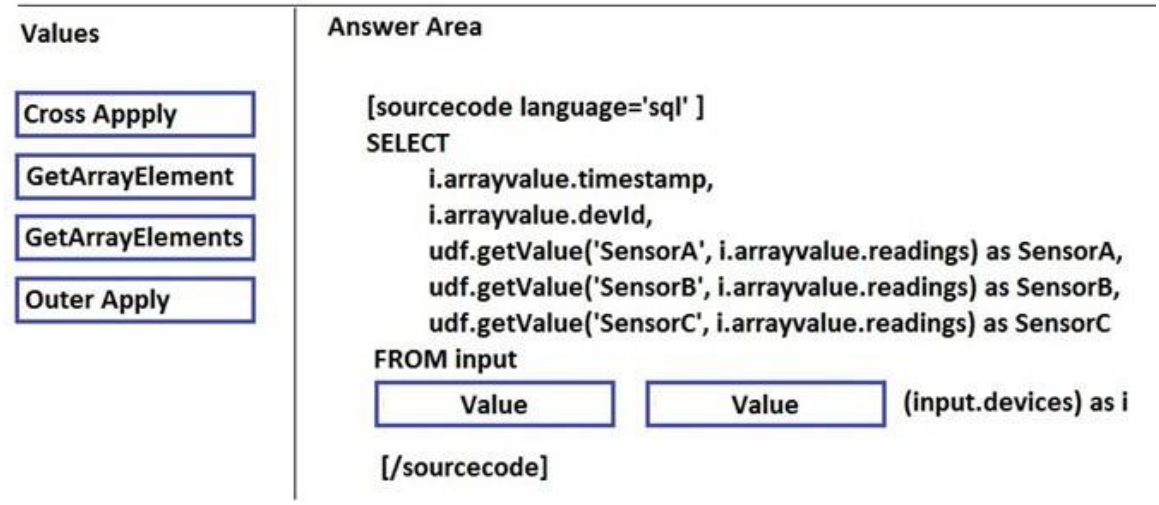
You have IoT devices that produce the following output:

```
sourcecode language='javascript' padlinenumbers='true']
[{
        "devID": "8656787",
        "timestamp": "2017-05-31T10:21:00",
        "readings": [{
            "type": "SensorA",
            "value": 18.965
         }, {
            "type": "SensorB",
            "value": 72.9157
         }, {
            "type": "SensorC",
            "value": 1524.672
         }
      ]
   }
]
[/sourcecode]
```

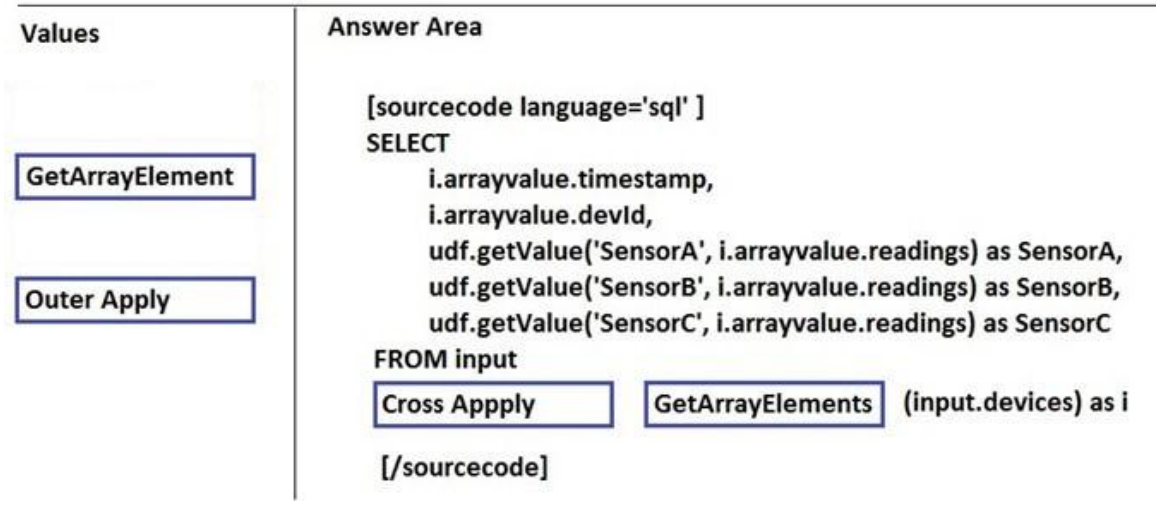
You need to use Microsoft Azure Stream Analytics to convert the output into the tabular format described in the following table:

Timestamp	DevId	SensorA	SensorB	SensorC
2017:05:31T10:21:00Z	8656787	18.965	72.9157	1524.672

How should you complete the Stream Analytics query? (To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.)



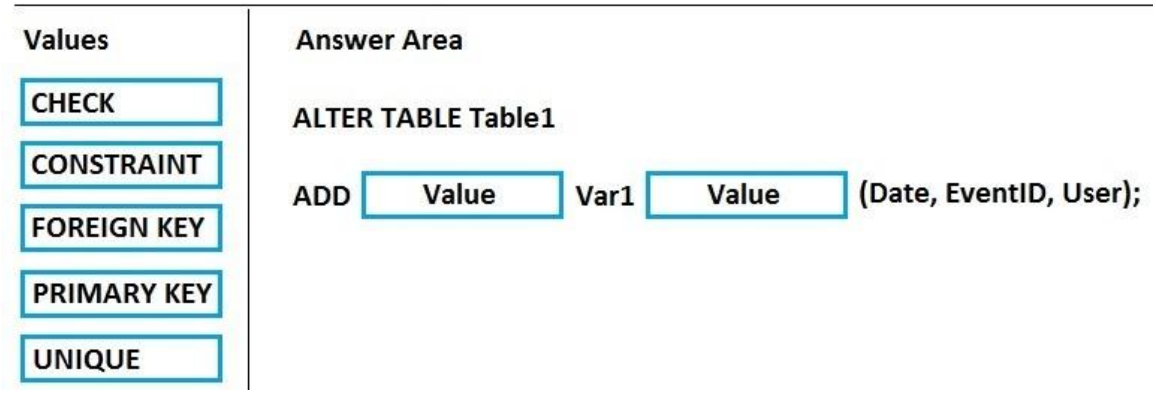
Answer:



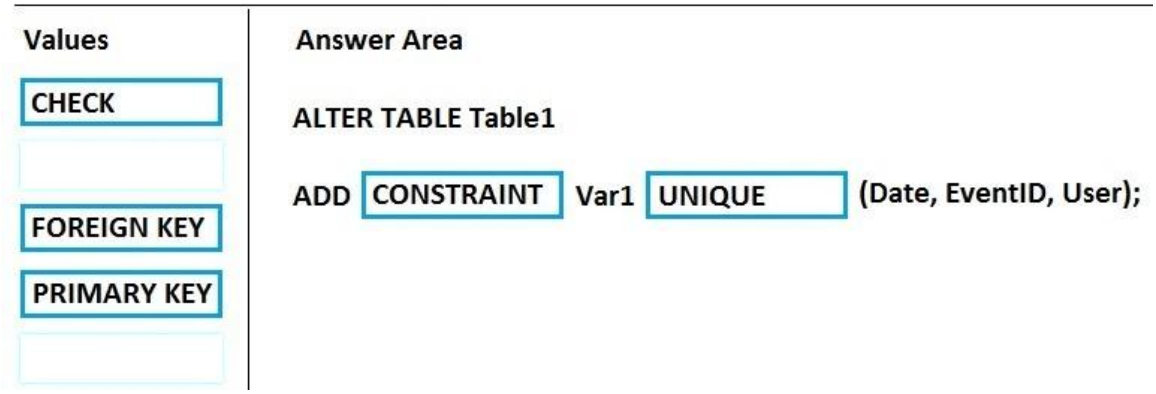
QUESTION 60

Drag and Drop

You use Microsoft Azure Stream Analytics to analyze data from an Azure event hub in real time and send the output to a table named Table1 in an Azure SQL database. Table1 has three columns named Date, EventID, and User. You need to prevent duplicate data from being stored in the database. How should you complete the statement? (To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.)



Answer:



QUESTION 61

.....