70-776 Q&As

Perform Big Data Engineering on Microsoft Cloud Services

Pass Microsoft 70-776 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass4itsure.com/70-776.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

QUESTION 1

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

Correct Answer: B

QUESTION 2

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;
```


A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: A

QUESTION 3

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

Correct Answer: C

QUESTION 4

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

Correct Answer: B

QUESTION 5

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

Correct Answer: A

QUESTION 6

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.

Correct Answer: A

QUESTION 7

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

Correct Answer: D

QUESTION 8

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.

NOTE: Each correct selection is worth one point.

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.

Correct Answer: DF

QUESTION 9

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

Correct Answer: A

QUESTION 10

HOTSPOT

You are designing a fact table that has 100 million rows and 1,800 partitions. The partitions are defined based on a column named OrderDayKey. The fact table will contain:

Data from the last five years

A clustered columnstore index

A column named YearMonthKey that stores the year and the month

Multiple transformations will be performed on the fact table during the loading process. The fact table will be hash distributed on a column named OrderId.

You plan to load the data to a staging table and to perform transformations on the staging table. You will then load the data from the staging table to the final fact table.

You need to design a solution to load the data to the fact table. The solution must minimize how long it takes to perform the following tasks:

Load the staging table.

Transfer the data from the staging table to the fact table.

Remove data that is older than five years.

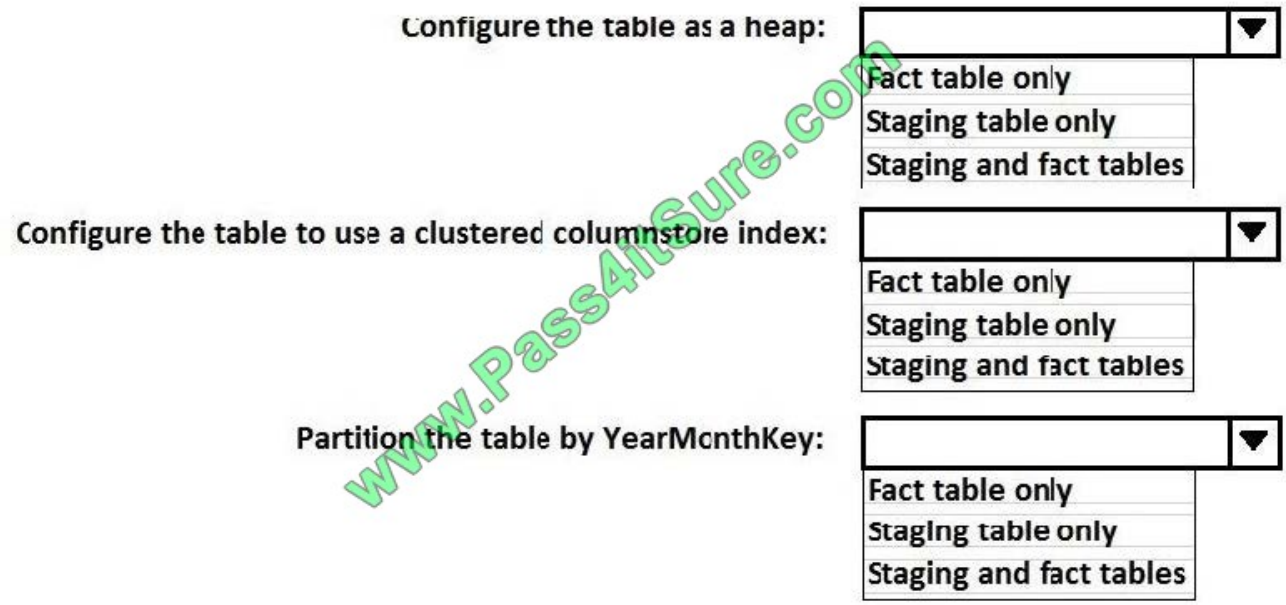
Query the data in the fact table

How should you configure the tables? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

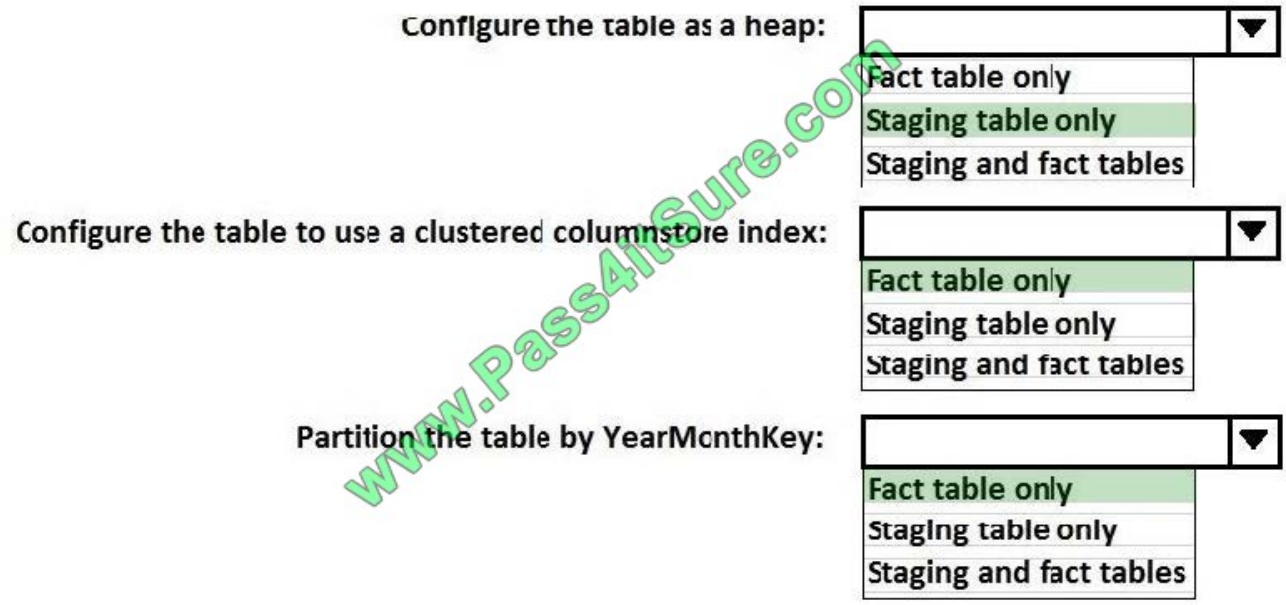
Hot Area:

Answer Area



Correct Answer:

Answer Area



QUESTION 11

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

Correct Answer: C

QUESTION 12

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.

Correct Answer: C