

DP-203 Free Brain Dumps & Frequent DP-203 Update

Actions	Answer Area
Switch the partition containing the stale data from SalesFact to SalesFact_Work.	
Truncate the partition containing the stale data.	
Drop the SalesFact_Work table.	
Create an empty table named SalesFact_Work that has the same schema as SalesFact.	
Execute a DELETE statement where the value in the Date column is more than 36 months ago.	
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).	

BONUS!!! Download part of Test4Engine DP-203 dumps for free: https://drive.google.com/open?id=1rFH1q2Wv_xdH_taeT8grw4IMhGj56eLk

We have prepared our Microsoft DP-203 Training Materials for you. They are professional practice material under warranty. Accompanied with acceptable prices for your reference, all our materials with three versions are compiled by professional experts in this area more than ten years long.

To prepare for the DP-203 Exam, candidates can leverage a variety of resources offered by Microsoft, such as official study guides, online training courses, and practice tests. Microsoft also recommends that candidates have hands-on experience with Azure-based data solutions before taking the exam. This can be achieved through practical work experience or by using Azure's free trial environment to gain hands-on experience with Azure-based data services.

>> DP-203 Free Brain Dumps <<

Free PDF Quiz 2026 Microsoft DP-203: First-grade Data Engineering on Microsoft Azure Free Brain Dumps

The meaning of qualifying examinations is, in some ways, to prove the candidate's ability to obtain qualifications that show your ability in various fields of expertise. If you choose our DP-203 learning guide materials, you can create more unlimited value in the limited study time, learn more knowledge, and take the DP-203 Exam that you can take. Through qualifying examinations, this is our DP-203 real questions and the common goal of every user, we are trustworthy helpers. The acquisition of DP-203 qualification certificates can better meet the needs of users' career development.

The DP-203 exam is a challenging certification that demands a deep understanding of data engineering concepts and Azure services. To successfully pass the exam, you must have a strong understanding of data engineering principles, as well as a thorough knowledge of Azure services and tools. Data Engineering on Microsoft Azure certification is a valuable asset to have for professionals who work with Azure data services, and it can help you advance your career in data engineering and analytics. With the DP-203 certification, you can demonstrate your expertise in designing and implementing data solutions on Azure and showcase your ability to work with data at scale.

Microsoft DP-203 (Data Engineering on Microsoft Azure) Exam is a certification exam that tests the skills and knowledge of candidates in designing and implementing data solutions on Microsoft Azure. DP-203 exam is designed for data engineers who work with data storage, processing, and analysis on Azure. DP-203 exam covers a range of topics such as data ingestion, transformation, storage, and processing using Azure services like Azure Data Factory, Azure Databricks, Azure Stream Analytics, and more.

Microsoft Data Engineering on Microsoft Azure Sample Questions (Q220-Q225):

NEW QUESTION # 220

You are designing a fact table named FactPurchase in an Azure Synapse Analytics dedicated SQL pool. The table contains purchases from suppliers for a retail store. FactPurchase will contain the following columns.

Name	Data type	Nullable
PurchaseKey	Bigint	No
DateKey	Int	No
SupplierKey	Int	No
StockItemKey	Int	No
PurchaseOrderID	Int	Yes
OrderedQuantity	Int	No
OrderedOuters	Int	No
ReceivedOuters	Int	No
Package	Nvarchar(50)	No
IsOrderFinalized	Bit	No
LineageKey	Int	No

FactPurchase will have 1 million rows of data added daily and will contain three years of data.

Transact-SQL queries similar to the following query will be executed daily.

```
SELECT  
SupplierKey, StockItemKey, COUNT(*)
```

```
FROM FactPurchase
```

```
WHERE DateKey >= 20210101
```

```
AND DateKey <= 20210131
```

```
GROUP By SupplierKey, StockItemKey
```

Which table distribution will minimize query times?

- A. hash-distributed on DateKey
- B. replicated
- C. round-robin
- D. hash-distributed on PurchaseKey

Answer: D

Explanation:

Hash-distributed tables improve query performance on large fact tables, and are the focus of this article.

Round-robin tables are useful for improving loading speed.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute>

NEW QUESTION # 221

You are designing an application that will use an Azure Data Lake Storage Gen 2 account to store petabytes of license plate photos from toll booths. The account will use zone-redundant storage (ZRS).

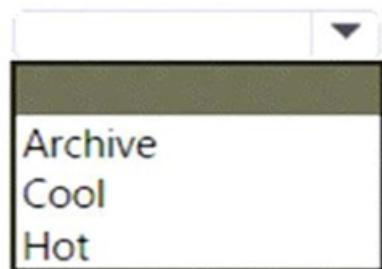
You identify the following usage patterns:

* The data will be accessed several times a day during the first 30 days after the data is created. The data must meet an availability SU of 99.9%.

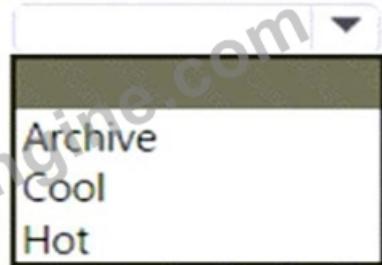
* After 90 days, the data will be accessed infrequently but must be available within 30 seconds.

* After 365 days, the data will be accessed infrequently but must be available within five minutes.

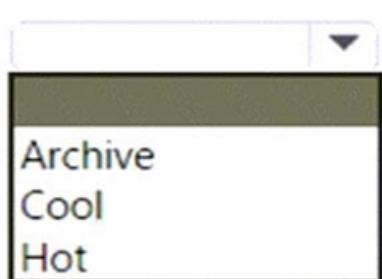
First 30 days:



After 90 days:



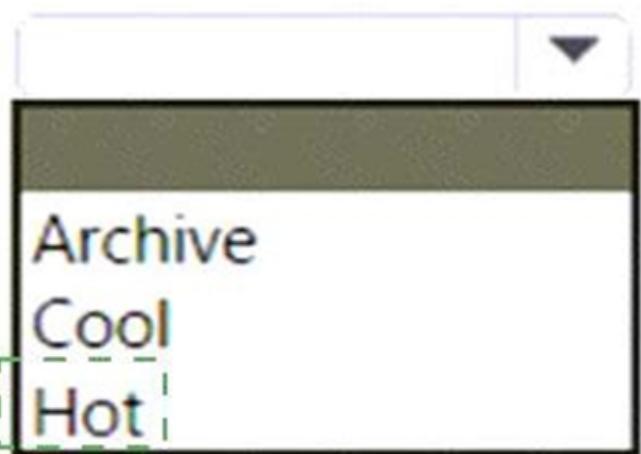
After 365 days:



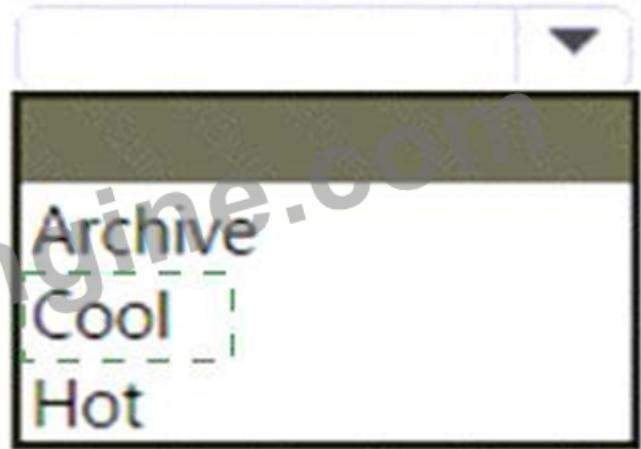
Answer:

Explanation:

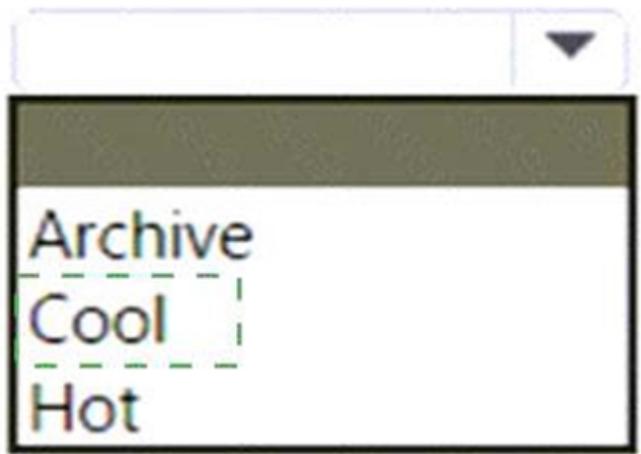
First 30 days:



After 90 days:



After 365 days:



Explanation:

Box 1: Hot

The data will be accessed several times a day during the first 30 days after the data is created. The data must meet an availability SLA of 99.9%.

Box 2: Cool

After 90 days, the data will be accessed infrequently but must be available within 30 seconds.

Data in the Cool tier should be stored for a minimum of 30 days.

When your data is stored in an online access tier (either Hot or Cool), users can access it immediately. The Hot tier is the best choice for data that is in active use, while the Cool tier is ideal for data that is accessed less frequently, but that still must be available for reading and writing.

Box 3: Cool

After 365 days, the data will be accessed infrequently but must be available within five minutes.

Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/access-tiers-overview>

NEW QUESTION # 222

You need to design an analytical storage solution for the transactional data. The solution must meet the sales transaction dataset requirements.

What should you include in the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Microsoft

Table type to store retail store data:

Table type to store promotional data:



Hash
Replicated
Round-robin

Hash
Replicated
Round-robin

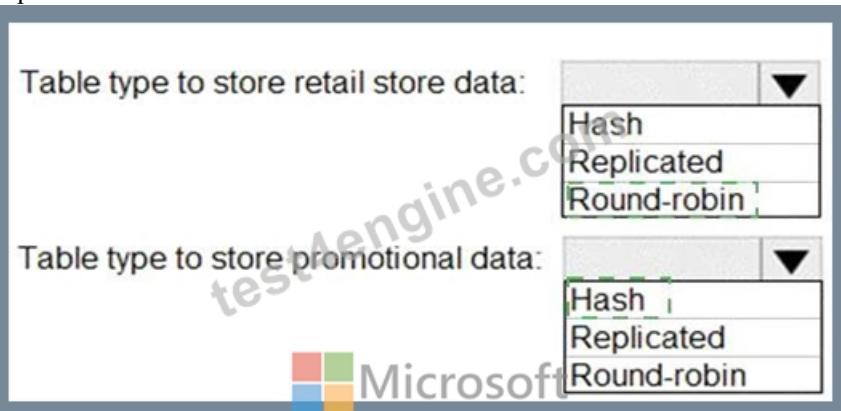
Answer:

Explanation:

Microsoft

Table type to store retail store data:

Table type to store promotional data:



Hash
Replicated
Round-robin

Hash
Replicated
Round-robin

Explanation:

Microsoft

Table type to store retail store data:

Table type to store promotional data:



Hash
Replicated
Round-robin

Hash
Replicated
Round-robin

Box 1: Round-robin

Round-robin tables are useful for improving loading speed.

Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month.

Box 2: Hash

Hash-distributed tables improve query performance on large fact tables.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute>

NEW QUESTION # 223

You are designing 2 solution that will use tables in Delta Lake on Azure Databricks.

You need to minimize how long it takes to perform the following:

- *Queries against non-partitioned tables
- * Joins on non-partitioned columns

Which two options should you include in the solution? Each correct answer presents part of the solution.

(Choose Correct Answer and Give Explanation and References to Support the answers based from Data Engineering on Microsoft Azure)

- **A. Apache Spark caching**
- B. dynamic file pruning (DFP)
- C. the clone command
- **D. Z-Ordering**

Answer: A,D

Explanation:

Explanation

According to the information I found on the web, two options that you should include in the solution to minimize how long it takes to perform queries and joins on non-partitioned tables are:

Z-Ordering: This is a technique to colocate related information in the same set of files. This co-locality is automatically used by Delta Lake in data-skipping algorithms. This behavior dramatically reduces the amount of data that Delta Lake on Azure Databricks needs to read.

Apache Spark caching: This is a feature that allows you to cache data in memory or on disk for faster access. Caching can improve the performance of repeated queries and joins on the same data. You can cache Delta tables using the CACHE TABLE or CACHE LAZY commands.

To minimize the time it takes to perform queries against non-partitioned tables and joins on non-partitioned columns in Delta Lake on Azure Databricks, the following options should be included in the solution:

A: Z-Ordering: Z-Ordering improves query performance by co-locating data that share the same column values in the same physical partitions. This reduces the need for shuffling data across nodes during query execution. By using Z-Ordering, you can avoid full table scans and reduce the amount of data processed.

B: Apache Spark caching: Caching data in memory can improve query performance by reducing the amount of data read from disk. This helps to speed up subsequent queries that need to access the same data. When you cache a table, the data is read from the data source and stored in memory. Subsequent queries can then read the data from memory, which is much faster than reading it from disk.

References:

Delta Lake on Databricks: <https://docs.databricks.com/delta/index.html>

Best Practices for Delta Lake on

Databricks: <https://databricks.com/blog/2020/05/14/best-practices-for-delta-lake-on-databricks.html>

NEW QUESTION # 224

You have a Microsoft SQL Server database that uses a third normal form schema.

You plan to migrate the data in the database to a star schema in an Azure Synapse Analytics dedicated SQL pool.

You need to design the dimension tables. The solution must optimize read operations.

What should you include in the solution? to answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Transform data for the dimension tables by:

Maintaining to a third normal form
Normalizing to a fourth normal form
Denormalizing to a second normal form

For the primary key columns in the dimension tables, use:

New IDENTITY columns
A new computed column
The business key column from the source sys

Answer:

Explanation:

Maintaining to a third normal form
Normalizing to a fourth normal form
Denormalizing to a second normal form

For the primary key columns in the dimension tables, use:

New IDENTITY columns
A new computed column
The business key column from the source sys

Explanation:

Text, table Description automatically generated

Transform data for the dimension tables by: 

Maintaining to a third normal form
Normalizing to a fourth normal form
Denormalizing to a second normal form

For the primary key columns in the dimension tables, use:

New IDENTITY columns
A new computed column
The business key column from the source sys

Box 1: Denormalize to a second normal form

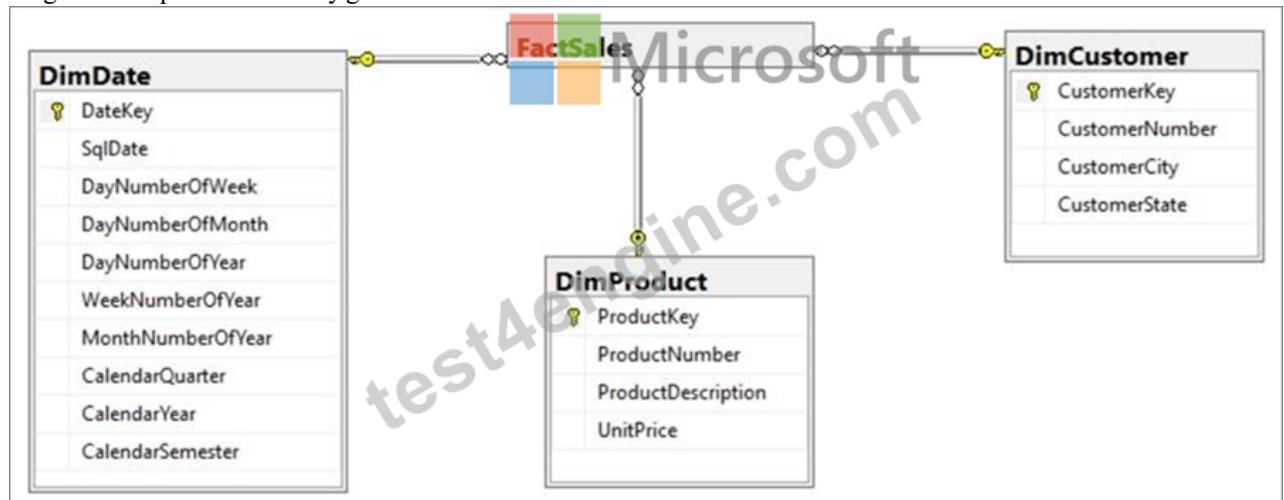
Denormalization is the process of transforming higher normal forms to lower normal forms via storing the join of higher normal form relations as a base relation. Denormalization increases the performance in data retrieval at cost of bringing update anomalies to a database.

Box 2: New identity columns

The collapsing relations strategy can be used in this step to collapse classification entities into component entities to obtain at dimension tables with single-part keys that connect directly to the fact table. The single-part key is a surrogate key generated to ensure it remains unique over time.

Example:

Diagram Description automatically generated



Note: A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data.

Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

<https://www.mssqltips.com/sqlservertip/5614/explore-the-role-of-normal-forms-in-dimensional-modeling/>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity>

NEW QUESTION # 225

.....

Frequent DP-203 Update: https://www.test4engine.com/DP-203_exam-latest-braindumps.html

BTW, DOWNLOAD part of Test4Engine DP-203 dumps from Cloud Storage: https://drive.google.com/open?id=1rFH1q2Wv_xdH_taeT8grw4IMhGj56eLk