



*Up-to-date Questions and Answers from authentic resources to improve knowledge and pass the exam at very first attempt. ----- Guaranteed.*



DP-203 Dumps  
DP-203 Braindumps  
DP-203 Real Questions  
DP-203 Practice Test  
DP-203 Actual Questions



**Microsoft**

# DP-203

*Data Engineering on Microsoft Azure*



<https://killexams.com/pass4sure/exam-detail/DP-203>

Question: 92

HOTSPOT

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.

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Answer:

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Explanation:

Graphical user

interface, text, application, table

Description automatically generated

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.

### Question: 93

You have an Azure data factory.

You need to examine the pipeline failures from the last 180 flays.

What should you use?

- A. the Activity tog blade for the Data Factory resource
- B. Azure Data Factory activity runs in Azure Monitor
- C. Pipeline runs in the Azure Data Factory user experience
- D. the Resource health blade for the Data Factory resource

### Answer: B

Explanation:

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor>

### Question: 94

HOTSPOT

You build an Azure Data Factory pipeline to move data from an Azure Data Lake Storage Gen2 container to a database in an Azure Synapse Analytics dedicated SQL pool.

Data in the container is stored in the following folder structure.

/in/{YYYY}/{MM}/{DD}/{HH}/{mm}

The earliest folder is /in/2021/01/01/00/00. The latest folder is /in/2021/01/15/01/45.

You need to configure a pipeline trigger to meet the following requirements:

Existing data must be loaded.

Data must be loaded every 30 minutes.

Late-arriving data of up to two minutes must be included in the load for the time at which the data should have arrived.

How should you configure the pipeline trigger? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Type: 

▼

Event

On-demand

Schedule

Tumbling window

Additional properties: 

▼

Prefix: /in/, Event: Blob created

Recurrence: 30 minutes, Start time: 2021-01-01T00:00

Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes

Recurrence: 32 minutes, Start time: 2021-01-15T01:45

Answer:

Type: 

▼

Event

On-demand

Schedule

Tumbling window

Additional properties: 

▼

Prefix: /in/, Event: Blob created

Recurrence: 30 minutes, Start time: 2021-01-01T00:00

Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes

Recurrence: 32 minutes, Start time: 2021-01-15T01:45

Explanation:

Box 1: Tumbling window

To be able to use the Delay parameter we select Tumbling window.

Box 2:

Recurrence: 30 minutes, not 32 minutes

Delay: 2 minutes.

The amount of time to delay the start of data processing for the window. The pipeline run is started after the expected execution time plus the amount of delay. The delay defines how long the trigger waits past the due time before triggering a new run. The delay doesn't alter the window startTime.

Question: 95

HOTSPOT

You need to design a data ingestion and storage solution for the Twitter feeds. The solution must meet the customer sentiment analytics 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.

To increase the throughput of ingesting the Twitter feeds:

Configure Event Hubs partitions.  
Enable Auto-Inflate in Event Hubs.  
Use Event Hubs Dedicated.

To store the Twitter feed data, use:

An Azure Data Lake Storage Gen2 account  
An Azure Databricks high concurrency cluster  
An Azure General-purpose v2 storage account in the Premium tier

Answer:

To increase the throughput of ingesting the Twitter feeds:

Configure Event Hubs partitions.  
Enable Auto-Inflate in Event Hubs.  
Use Event Hubs Dedicated.

To store the Twitter feed data, use:

An Azure Data Lake Storage Gen2 account  
An Azure Databricks high concurrency cluster  
An Azure General-purpose v2 storage account in the Premium tier

Explanation:

Graphical user interface, text

Description automatically generated

Box 1: Configure Event Hubs partitions

Scenario: Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Event Hubs is designed to help with processing of large volumes of events. Event Hubs throughput is scaled by using partitions and throughput-unit allocations.

Event Hubs traffic is controlled by TUs (standard tier). Auto-inflate enables you to start small with the minimum required TUs you choose. The feature then scales automatically to the maximum limit of TUs you need, depending on the increase in your traffic.

#### Box 2: An Azure Data Lake Storage Gen2 account

Scenario: Ensure that the data store supports Azure AD-based access control down to the object level.

Azure Data Lake Storage Gen2 implements an access control model that supports both Azure role-based access control (Azure RBAC) and POSIX-like access control lists (ACLs).

#### Question: 96

You have an Azure Stream Analytics query. The query returns a result set that contains 10,000 distinct values for a column named clusterID.

You monitor the Stream Analytics job and discover high latency.

You need to reduce the latency.

Which two actions should you perform? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Add a pass-through query.
- B. Add a temporal analytic function.
- C. Scale out the query by using PARTITION BY.
- D. Convert the query to a reference query.
- E. Increase the number of streaming units.

**Answer:** C,E

Explanation:

C: Scaling a Stream Analytics job takes advantage of partitions in the input or output. Partitioning lets you divide data into subsets based on a partition key. A process that consumes the data (such as a Streaming Analytics job) can consume and write different partitions in parallel, which increases throughput.

E: Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job. This capacity lets you focus on the query logic and abstracts the need to manage the hardware to run your Stream Analytics job in a timely manner.

References:

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

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption>

#### Question: 97

HOTSPOT



You have an Azure subscription.

You need to deploy an Azure Data Lake Storage Gen2 Premium account.

The solution must meet the following requirements:

- Blobs that are older than 365 days must be deleted.
- Administrator efforts must be minimized.
- Costs must be minimized

What should you use? To answer, select the appropriate options in the answer area. NOTE Each correct selection is worth one point.

#### Answer Area

To minimize costs:

- Locally-redundant storage (LRS)
- The Archive access tier
- The Cool access tier
- Zone-redundant storage (ZRS)

To delete blobs:

- Azure Automation runbooks
- Azure Storage lifecycle management
- Soft delete

These are the selections for 'To delete blobs.'

#### Answer: Answer Area

To minimize costs:

- Locally-redundant storage (LRS)
- The Archive access tier
- The Cool access tier
- Zone-redundant storage (ZRS)

To delete blobs:

- Azure Automation runbooks
- Azure Storage lifecycle management
- Soft delete

These are the selections for 'To delete blobs.'

Explanation:

<https://learn.microsoft.com/en-us/azure/storage/blobs/premium-tier-for-data-lake-storage>

Question: 98

DRAG DROP

You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool.

The solution must meet the customer sentiment analytics requirements.

Which three Transaction-SQL DDL commands should you run in sequence? To answer, move the appropriate commands from the list of commands to the answer area and arrange them in the correct order. NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Commands

CREATE EXTERNAL DATA SOURCE
CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL TABLE
CREATE EXTERNAL TABLE AS SELECT
CREATE DATABASE SCOPED CREDENTIAL

Answer Area

Answer:

Commands

CREATE EXTERNAL DATA SOURCE
CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL TABLE
CREATE EXTERNAL TABLE AS SELECT
CREATE DATABASE SCOPED CREDENTIAL

Answer Area

CREATE EXTERNAL DATA SOURCE
CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL TABLE AS SELECT

Explanation:

Scenario: Allow Contoso users to use PolyBase in an Azure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own Azure AD credentials.

Box 1: CREATE EXTERNAL DATA SOURCE

External data sources are used to connect to storage accounts.

Box 2: CREATE EXTERNAL FILE FORMAT

CREATE EXTERNAL FILE FORMAT creates an external file format object that defines external data stored in Azure Blob Storage or Azure Data Lake Storage. Creating an external file format is a prerequisite for creating an external table.

Box 3: CREATE EXTERNAL TABLE AS SELECT

When used in conjunction with the CREATE TABLE AS SELECT statement, selecting from an external table imports data into a table within the SQL pool. In addition to the COPY statement, external tables are useful for loading data.

Question: 99

DRAG DROP



You have the following table named Employees.

first_name	last_name	hire_date	employee_type
Jane	Doe	2019-08-23	new
Ben	Smith	2017-12-15	Standard

You need to calculate the employee\_type value based on the hire\_date value.

How should you complete the Transact-SQL 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. NOTE: Each correct selection is worth one point.

Values

Answer Area

- CASE
- ELSE
- OVER
- PARTITION BY
- ROW\_NUMBER

```
SELECT
    *,
    [ ]
    WHEN hire_date >= '2019-01-01' THEN 'New'
    [ ] 'Standard'
END AS employee_type
FROM
    employees
```

Answer:  
Values

Answer Area

- CASE
- ELSE
- OVER
- PARTITION BY
- ROW\_NUMBER

```
SELECT
    *,
    CASE
        WHEN hire_date >= '2019-01-01' THEN 'New'
        ELSE 'Standard'
    END AS employee_type
FROM
    employees
```

Explanation:  
Graphical user

interface, text, application

Description automatically generated

#### Box 1: CASE

CASE evaluates a list of conditions and returns one of multiple possible result expressions.

CASE can be used in any statement or clause that allows a valid expression. For example, you can use CASE in statements such as SELECT, UPDATE, DELETE and SET, and in clauses such as select\_list, IN, WHERE, ORDER BY, and HAVING.

Syntax: Simple CASE expression:

CASE input\_expression

WHEN when\_expression THEN result\_expression [ ...n ] [ ELSE else\_result\_expression ] END

#### Box 2: ELSE

Question: 100

#### HOTSPOT

You are building a database in an Azure Synapse Analytics serverless SQL pool.

You have data stored in Parquet files in an Azure Data Lake Storage Gen2 container.

Records are structured as shown in the following sample.

```
{  
  "id": 123,  
  "address_housenumber": "19c",  
  "address_line": "Memory Lane",  
  "applicant1_name": "Jane",  
  "applicant2_name": "Dev"  
}
```

The records contain two applicants at most.

You need to build a table that includes only the address fields.

How should you complete the Transact-SQL statement? To answer, select the appropriate options in the answer area.  
NOTE: Each correct selection is worth one point.

## Answer Area

```
applications
CREATE EXTERNAL TABLE
CREATE TABLE
CREATE VIEW
WITH (
    LOCATION = 'applications/',
    DATA_SOURCE = applications_ds,
    FILE_FORMAT = applications_file_format
)
AS
SELECT id, [address_housenumber] as addresshousenumber, [address_line1] as addressline1
FROM
(BULK 'https://contosol.dfs.core.windows.net/applications/year=*/*.parquet',
CROSS APPLY
OPENJSON
OPENROWSET
FORMAT=' PARQUET') AS [r]
GO
```

## Answer:

### Answer Area

```
applications
CREATE EXTERNAL TABLE
CREATE TABLE
CREATE VIEW
WITH (
    LOCATION = 'applications/',
    DATA_SOURCE = applications_ds,
    FILE_FORMAT = applications_file_format
)
AS
SELECT id, [address_housenumber] as addresshousenumber, [address_line1] as addressline1
FROM
(BULK 'https://contosol.dfs.core.windows.net/applications/year=*/*.parquet',
CROSS APPLY
OPENJSON
OPENROWSET
FORMAT=' PARQUET') AS [r]
GO
```

Explanation:

### Box 1: CREATE EXTERNAL TABLE

An external table points to data located in Hadoop, Azure Storage blob, or Azure Data Lake Storage. External tables are used to read data from files or write data to files in Azure Storage. With Synapse SQL, you can use external tables to read external data using dedicated SQL pool or serverless SQL pool.

Syntax:

```
CREATE EXTERNAL TABLE { database_name.schema_name.table_name | schema_name.table_name | table_name
} ( <column_definition> [ ,...n ] )
```

WITH (

LOCATION = 'folder\_or\_filepath',

DATA\_SOURCE = external\_data\_source\_name, FILE\_FORMAT = external\_file\_format\_name

## Box 2. OPENROWSET

When using serverless SQL pool, CETAS is used to create an external table and export query results to Azure Storage Blob or Azure Data Lake Storage Gen2.

Example:

AS

SELECT decennialTime, stateName, SUM(population) AS population

FROM

OPENROWSET(BULK

'https://azureopendatastorage.blob.core.windows.net/censusdatacontainer/release/us\_pop

ulation\_county/year=\*/\*.parquet',

FORMAT='PARQUET') AS [r]

GROUP BY decennialTime, stateName

GO



# SAMPLE QUESTIONS

*These questions are for demo purpose only. **Full version** is up to date and contains actual questions and answers.*

*Killexams.com is an online platform that offers a wide range of services related to certification exam preparation. The platform provides actual questions, exam dumps, and practice tests to help individuals prepare for various certification exams with confidence. Here are some key features and services offered by Killexams.com:*

**Actual Exam Questions:** *Killexams.com provides actual exam questions that are experienced in test centers. These questions are updated regularly to ensure they are up-to-date and relevant to the latest exam syllabus. By studying these actual questions, candidates can familiarize themselves with the content and format of the real exam.*

**Exam Dumps:** *Killexams.com offers exam dumps in PDF format. These dumps contain a comprehensive collection of questions and answers that cover the exam topics. By using these dumps, candidates can enhance their knowledge and improve their chances of success in the certification exam.*

**Practice Tests:** *Killexams.com provides practice tests through their desktop VCE exam simulator and online test engine. These practice tests simulate the real exam environment and help candidates assess their readiness for the actual exam. The practice tests cover a wide range of questions and enable candidates to identify their strengths and weaknesses.*

**Guaranteed Success:** *Killexams.com offers a success guarantee with their exam dumps. They claim that by using their materials, candidates will pass their exams on the first attempt or they will refund the purchase price. This guarantee provides assurance and confidence to individuals preparing for certification exams.*

**Updated Content:** *Killexams.com regularly updates its question bank and exam dumps to ensure that they are current and reflect the latest changes in the exam syllabus. This helps candidates stay up-to-date with the exam content and increases their chances of success.*

**Technical Support:** *Killexams.com provides free 24x7 technical support to assist candidates with any queries or issues they may encounter while using their services. Their certified experts are available to provide guidance and help candidates throughout their exam preparation journey.*

For More exams visit <https://killexams.com/vendors-exam-list>  
Kill your exam at First Attempt....Guaranteed!