



DP-203^{Q&As}

Data Engineering on Microsoft Azure

Pass Microsoft DP-203 Exam with 100% Guarantee

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

<https://www.geekcert.com/dp-203.html>

100% Passing Guarantee
100% Money Back Assurance

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

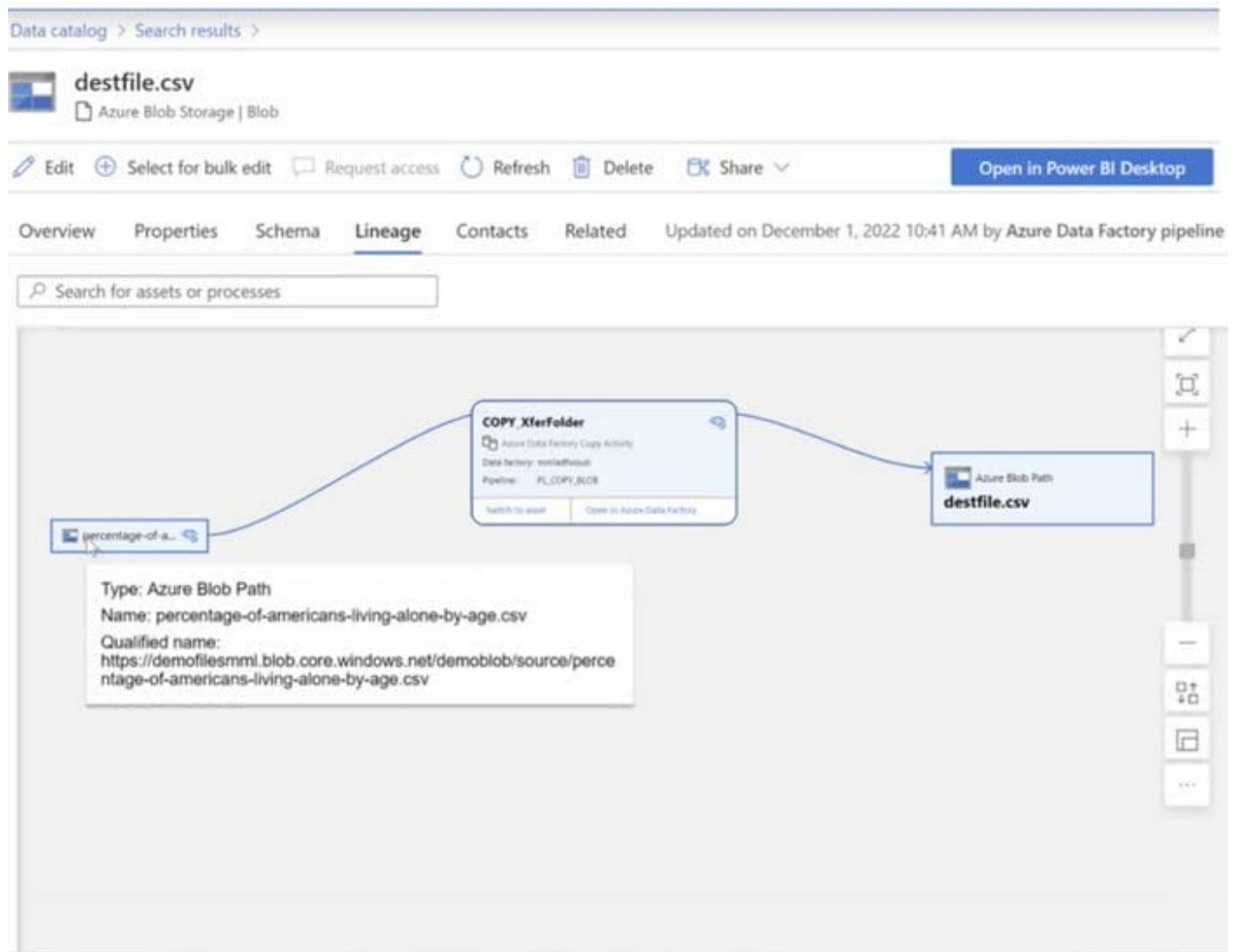
- ⚙️ **Instant Download** After Purchase
- ⚙️ **100% Money Back** Guarantee
- ⚙️ **365 Days** Free Update
- ⚙️ **800,000+** Satisfied Customers





QUESTION 1

You have a Microsoft Purview account. The Lineage view of a CSV file is shown in the following exhibit.



How is the data for the lineage populated?

- A. manually
- B. by scanning data stores
- C. by executing a Data Factory pipeline

Correct Answer: C

From the exhibit we see Copy_XferFolder (and even: From Data Factory).

The following example is a typical use case of data moving across multiple systems, where the Data Catalog would connect to each of the systems for lineage.

1.



Data Factory copies data from on-prem/raw zone to a landing zone in the cloud.

2.

Etc.

Reference: <https://learn.microsoft.com/en-us/azure/purview/concept-data-lineage>

QUESTION 2

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 an Azure Storage account that contains 100 GB of files. The files contain rows of text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is more than 1 MB.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Instead convert the files to compressed delimited text files.

Reference: <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

QUESTION 3

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 b worth one point.

Hot Area:



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

Correct 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

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.

Incorrect Answers:

Event Hubs Dedicated: Event Hubs clusters offer single-tenant deployments for customers with the most demanding streaming needs. This single-tenant offering has a guaranteed 99.99% SLA and is available only on our Dedicated pricing

tier.

Auto-Inflate: The Auto-inflate feature of Event Hubs automatically scales up by increasing the number of TUs, to meet usage needs. 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).

Incorrect Answers:

Azure Databricks: An Azure administrator with the proper permissions can configure Azure Active Directory conditional access to control where and when users are permitted to sign in to Azure Databricks.

Azure Storage supports using Azure Active Directory (Azure AD) to authorize requests to blob data. You can scope access to Azure blob resources at the following levels, beginning with the narrowest scope:

- An individual container. At this scope, a role assignment applies to all of the blobs in the container, as well as container properties and metadata.

- The storage account. At this scope, a role assignment applies to all containers and their blobs.

-

The resource group. At this scope, a role assignment applies to all of the containers in all of the storage accounts in the resource group.

-

The subscription. At this scope, a role assignment applies to all of the containers in all of the storage accounts in all of the resource groups in the subscription.

- A management group.

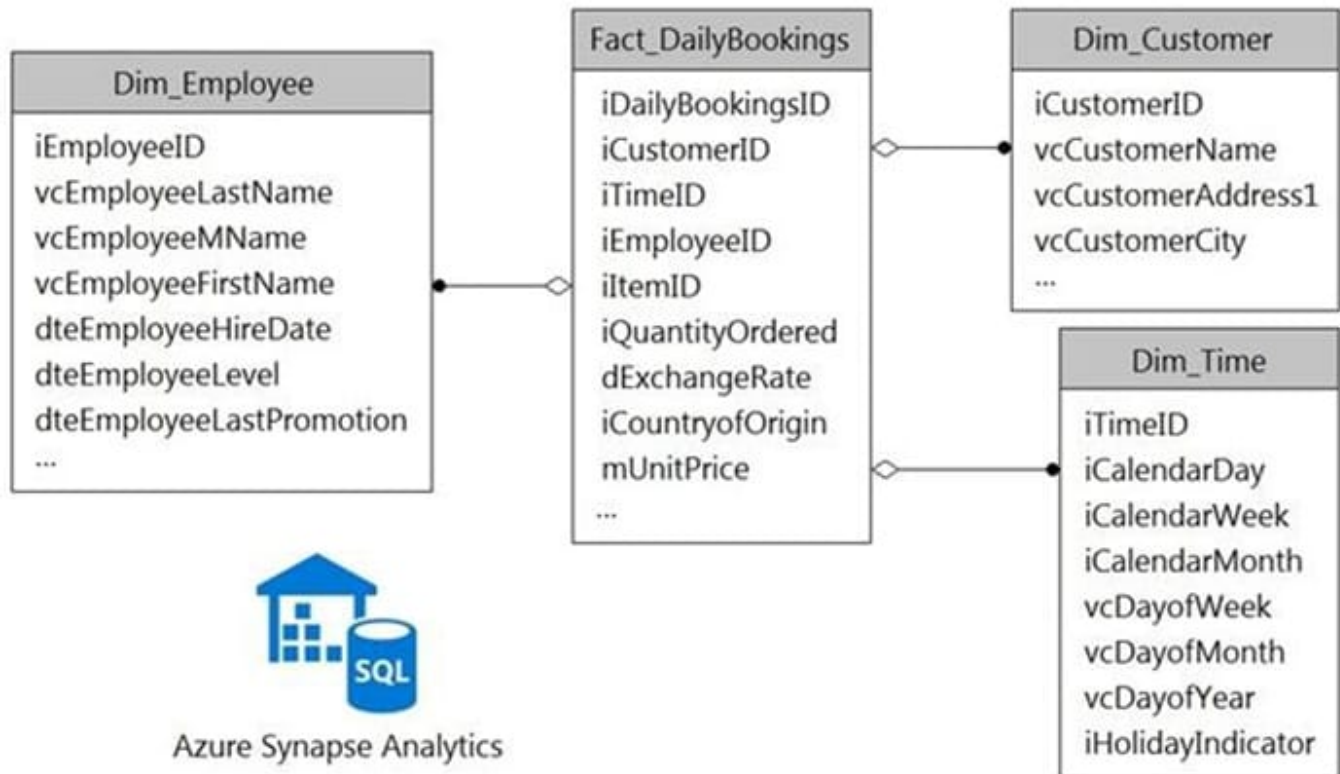
Reference: <https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features> <https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control>

QUESTION 4



HOTSPOT

You have a data model that you plan to implement in a data warehouse in Azure Synapse Analytics as shown in the following exhibit.



All the dimension tables will be less than 2 GB after compression, and the fact table will be approximately 6 TB. Which type of table should you use for each table? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

Dim_Customer:

	▼
Hash distributed	
Round-robin	
Replicated	

Dim_Employee:

	▼
Hash distributed	
Round-robin	
Replicated	

Dim_Time:

	▼
Hash distributed	
Round-robin	
Replicated	

Fact_DailyBookings:

	▼
Hash distributed	
Round-robin	
Replicated	

Correct Answer:



Answer Area

Dim_Customer:

	▼
Hash distributed	
Round-robin	
Replicated	

Dim_Employee:

	▼
Hash distributed	
Round-robin	
Replicated	

Dim_Time:

	▼
Hash distributed	
Round-robin	
Replicated	

Fact_DailyBookings:

	▼
Hash distributed	
Round-robin	
Replicated	

**QUESTION 5**

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 an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1.

You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files in container1 into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.

You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.

Solution: In an Azure Synapse Analytics pipeline, you use a data flow that contains a Derived Column transformation.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

Use the derived column transformation to generate new columns in your data flow or to modify existing fields.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/data-flow-derived-column>

[Latest DP-203 Dumps](#)

[DP-203 PDF Dumps](#)

[DP-203 Practice Test](#)