

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.leads4pass.com/dp-203.html>

100% Passing Guarantee
100% Money Back Assurance

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

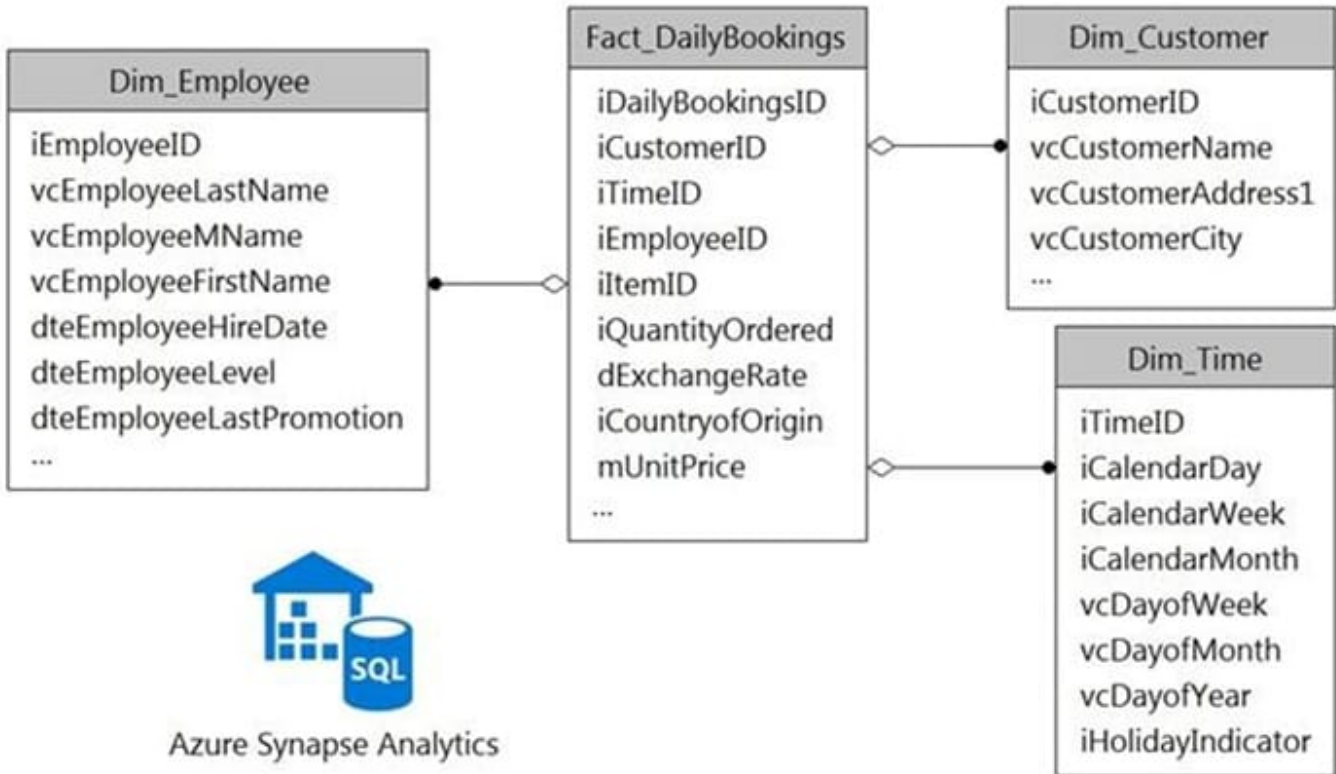
-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers



QUESTION 1

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:	<div style="border: 1px solid gray; padding: 2px;"><div style="background-color: #e0e0e0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Hash distributed</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Round-robin</div><div style="background-color: #c8e6c9; border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Replicated</div></div>
Dim_Employee:	<div style="border: 1px solid gray; padding: 2px;"><div style="background-color: #e0e0e0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Hash distributed</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Round-robin</div><div style="background-color: #c8e6c9; border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Replicated</div></div>
Dim_Time:	<div style="border: 1px solid gray; padding: 2px;"><div style="background-color: #e0e0e0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Hash distributed</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Round-robin</div><div style="background-color: #c8e6c9; border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Replicated</div></div>
Fact_DailyBookings:	<div style="border: 1px solid gray; padding: 2px;"><div style="background-color: #e0e0e0; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><div style="background-color: #c8e6c9; border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Hash distributed</div><div style="border-top: 1px solid gray; border-bottom: 1px solid gray; padding: 2px;">Round-robin</div><div style="border-top: 1px solid gray; padding: 2px;">Replicated</div></div>

QUESTION 2

You have an Azure Data Factory that contains 10 pipelines.

You need to label each pipeline with its main purpose of either ingest, transform, or load. The labels must be available for grouping and filtering when using the monitoring experience in Data Factory.

What should you add to each pipeline?

- A. a resource tag
- B. a correlation ID
- C. a run group ID
- D. an annotation

Correct Answer: D

Annotations are additional, informative tags that you can add to specific factory resources: pipelines, datasets, linked services, and triggers. By adding annotations, you can easily filter and search for specific factory resources.

Reference: <https://www.cathrinewilhelmsen.net/annotations-user-properties-azure-data-factory/>

QUESTION 3

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 designing an Azure Stream Analytics solution that will analyze Twitter data.

You need to count the tweets in each 10-second window. The solution must ensure that each tweet is counted only once.

Solution: You use a session window that uses a timeout size of 10 seconds.

Does this meet the goal?

- A. Yes
- B. No

Correct Answer: B

Instead use a tumbling window. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

Reference: <https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

QUESTION 4

You have an Azure Factory instance named DF1 that contains a pipeline named PL1. PL1 includes a tumbling window trigger.

You create five clones of PL1. You configure each clone pipeline to use a different data source.

You need to ensure that the execution schedules of the clone pipeline match the execution schedule of PL1.

What should you do?

- A. Add a new trigger to each cloned pipeline
- B. Associate each cloned pipeline to an existing trigger.
- C. Create a tumbling window trigger dependency for the trigger of PL1.
- D. Modify the Concurrency setting of each pipeline.

Correct Answer: B

QUESTION 5

You have an Azure Synapse Analytics dedicated SQL pool.

You run `PDW_SHOWSPACEUSED('\\dbo.FactInternetSales\\')`; and get the results shown in the following table.

ROWS	RESERVED_SPACE	DATA_SPACE	INDEX_SPACE	UNUSED_SPACE	PDW_NODE_ID	DISTRIBUTION_ID
694	2776	616	48	2112	1	1
407	2704	576	48	2080	1	2
53	2376	512	16	1848	1	3
58	2376	512	16	1848	1	4
168	2632	528	32	2072	1	5
195	2696	536	32	2128	1	6
5995	3464	1424	32	2008	1	7
0	2232	496	0	1736	1	8
264	2576	544	40	1992	1	9
3008	3016	960	32	2024	1	10
...
1550	2832	752	48	2032	1	50
1238	2832	696	40	2096	1	51
192	2632	528	32	2072	1	52
1127	2768	680	48	2040	1	53
1244	3032	704	64	2264	1	54
409	2632	568	32	2032	1	55
0	2232	496	0	1736	1	56
1437	2832	728	40	2064	1	57
0	2232	496	0	1736	1	58
384	2632	560	32	2040	1	59
225	2768	544	40	2184	1	60

Which statement accurately describes the dbo.FactInternetSales table?

- A. All distributions data.
- B. The table contains less than 10,000 rows.
- C. The table uses round-robin distribution.
- D. The table is skewed.

Correct Answer: D

Data skew means the data is not distributed evenly across the distributions.

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