



# DP-203<sup>Q&As</sup>

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.passapply.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 an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1. You need to verify whether the size of the transaction log file for each distribution of DW1 is smaller than 160 GB. What should you do?

- A. On the master database, execute a query against the sys.dm\_pdw\_nodes\_os\_performance\_counters dynamic management view.
- B. From Azure Monitor in the Azure portal, execute a query against the logs of DW1.
- C. On DW1, execute a query against the sys.database\_files dynamic management view.
- D. Execute a query against the logs of DW1 by using the Get-AzOperationalInsightSearchResult PowerShell cmdlet.

Correct Answer: A

The following query returns the transaction log size on each distribution. If one of the log files is reaching 160 GB, you should consider scaling up your instance or limiting your transaction size.

```
-- Transaction log size SELECT instance_name as distribution_db, cntr_value*1.0/1048576 as  
log_file_size_used_GB, pdw_node_id FROM sys.dm_pdw_nodes_os_performance_counters WHERE instance_name  
like '\\Distribution_%\\' AND counter_name = '\\Log File(s) Used Size (KB)\\'
```

References: <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-managemonitor>

---

### QUESTION 2

You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a fact table named Table1.

You need to identify the extent of the data skew in Table1.

What should you do in Synapse Studio?

- A. Connect to Pool1 and run DBCC PDW\_SHOWSPACEUSED.
- B. Connect to the built-in pool and run DBCC PDW\_SHOWSPACEUSED.
- C. Connect to the built-in pool and run DBCC CHECKALLOC.
- D. Connect to the built-in pool and query sys.dm\_pdw\_sys\_info.

Correct Answer: A

sys.dm\_pdw\_sys\_info actually provides a set of appliance-level counters that reflect overall activity on the appliance. DBCC PDW\_SHOWSPACEUSED should be used instead since it displays the number of rows, disk space reserved, and disk space used for a specific table, or for all tables in an Azure Synapse Analytics or Analytics Platform System (PDW) database.

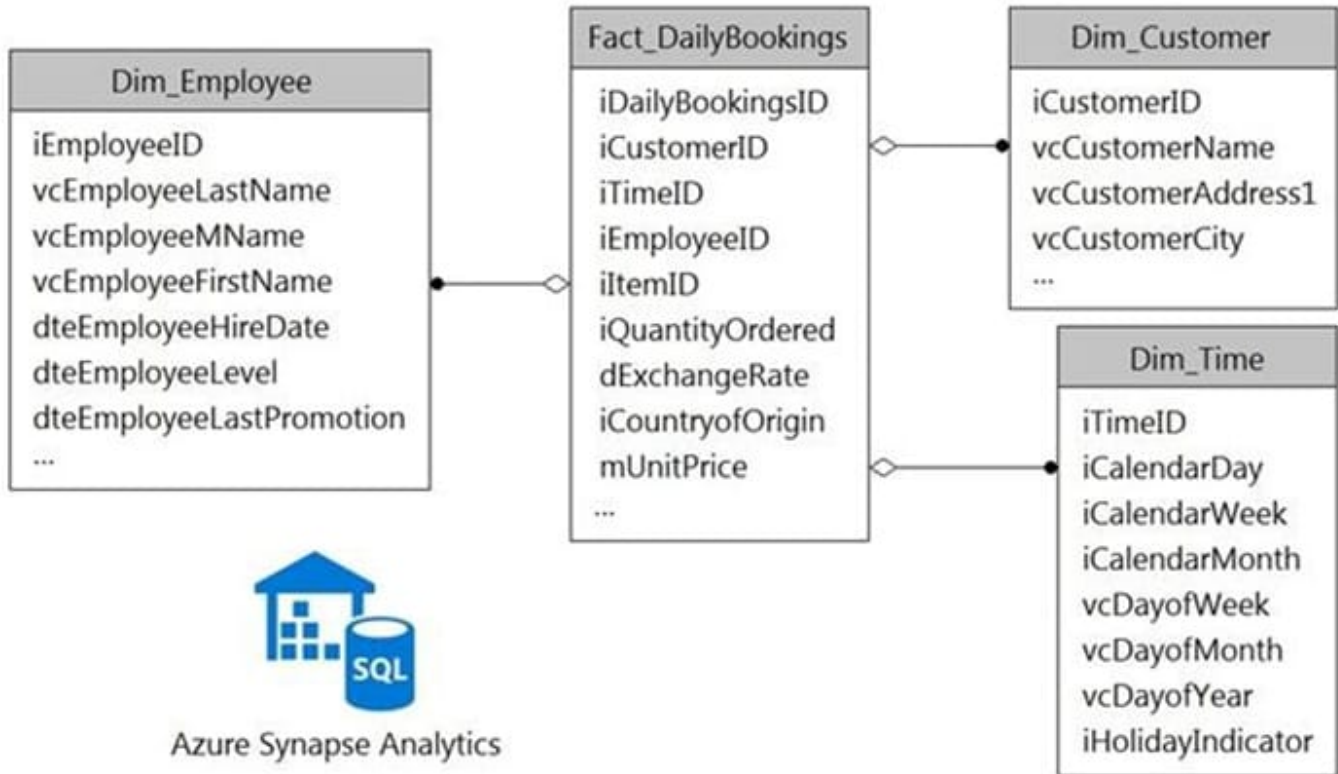
---

### QUESTION 3



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:	<input type="text"/> Hash distributed Round-robin Replicated
Dim_Employee:	<input type="text"/> Hash distributed Round-robin Replicated
Dim_Time:	<input type="text"/> Hash distributed Round-robin Replicated
Fact_DailyBookings:	<input type="text"/> 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;"><span></span><span>▼</span></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;"><span></span><span>▼</span></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;"><span></span><span>▼</span></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;"><span></span><span>▼</span></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 4

##### HOTSPOT

You store files in an Azure Data Lake Storage Gen2 container. The container has the storage policy shown in the following exhibit.

```
{
  "rules": [
    {
      "enabled": true,
      "name": "contosorule",
      "type": "Lifecycle",
      "definition": {
        "actions": {
          "version": {
            "delete": {
              "daysAfterCreationGreaterThan": 60
            }
          },
          "baseBlob": {
            "tierToCool": {
              "daysAfterModificationGreaterThan":
30
            },
          },
        },
      },
      "filters": {
        "blobTypes": [
          "blockBlob"
        ],
        "prefixMatch": [
          "container1/contoso"
        ]
      }
    }
  ]
}
```

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

NOTE: Each correct selection is worth one point.

Hot Area:





Segment the data stream into distinct time Segments that repeat but do not overlap:

Deleted from the container
Moved to archive storage
Moved to cool storage
Moved to cool storage

The storage policy applies to [answer choice]

Container1/contoso1.csv
Container1/docs/contoso.json
Container1/mycontoso/contoso.csv

Correct Answer:

Segment the data stream into distinct time Segments that repeat but do not overlap:

Deleted from the container
Moved to archive storage
Moved to cool storage
Moved to cool storage

The storage policy applies to [answer choice]

Container1/contoso1.csv
Container1/docs/contoso.json
Container1/mycontoso/contoso.csv

### QUESTION 5

You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

1.  
A workload for data engineers who will use Python and SQL.
2.  
A workload for jobs that will run notebooks that use Python, Scala, and SOL.
3.  
A workload that data scientists will use to perform ad hoc analysis in Scala and R.

The enterprise architecture team at your company identifies the following standards for Databricks environments:

- 1.



The data engineers must share a cluster.

2.

The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

3.

All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

We need a High Concurrency cluster for the data engineers and the jobs.

Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

Reference:

<https://docs.azuredatabricks.net/clusters/configure.html>

[DP-203 PDF Dumps](#)

[DP-203 Practice Test](#)

[DP-203 Brainsdumps](#)