

DP-600^{Q&As}

Implementing Analytics Solutions Using Microsoft Fabric

Pass Microsoft DP-600 Exam with 100% Guarantee

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

<https://www.pass2lead.com/dp-600.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 Fabric tenant that contains a data pipeline.

You need to ensure that the pipeline runs every four hours on Mondays and Fridays.

To what should you set Repeat for the schedule?

- A. Daily
- B. By the minute
- C. Weekly
- D. Hourly

Correct Answer: C

Explanation: You should set Repeat for the schedule to Weekly (C). This allows you to specify the pipeline to run on specific days of the week, in this case, every four hours on Mondays and Fridays. References = Scheduling options for data pipelines are available in the Azure Data Factory documentation, which includes details on configuring recurring triggers.

QUESTION 2

You have source data in a folder on a local computer.

You need to create a solution that will use Fabric to populate a data store. The solution must meet the following requirements:

Support the use of dataflows to load and append data to the data store.

Ensure that Delta tables are V-Order optimized and compacted automatically.

Which type of data store should you use?

- A. a lakehouse
- B. an Azure SQL database
- C. a warehouse
- D. a KQL database

Correct Answer: A

Explanation: A lakehouse (A) is the type of data store you should use. It supports dataflows to load and append data and ensures that Delta tables are Z-Order optimized and compacted automatically. References = The capabilities of a lakehouse and its support for Delta tables are described in the lakehouse and Delta table documentation.

QUESTION 3

You need to resolve the issue with the pricing group classification.

How should you complete the T-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

```
CREATE [VIEW] [dbo].[ProductsWithPricingGroup]
AS
SELECT ProductCategory,
ListPrice,
CASE
WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
WHEN (ListPrice > 50 AND ListPrice <= 1000 ) THEN 'medium'
WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
WHEN ListPrice BETWEEN 50 AND 1000 ) THEN 'medium'
END AS PricingGroup
FROM dbo.Products
```

Correct Answer:

```
CREATE VIEW [dbo].[ProductsWithPricingGroup]
AS
SELECT ProductCategory,
ListPrice,
CASE
WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
WHEN (ListPrice > 50 AND ListPrice <= 1000 ) THEN 'medium'
WHEN (ListPrice >= 50 AND ListPrice < 1000 ) THEN 'medium'
WHEN ListPrice BETWEEN 50 AND 1000 ) THEN 'medium'
END AS PricingGroup
FROM dbo.Products
```

You should use CREATE VIEW to make the pricing group logic available for TSQL queries. The CASE statement should be used to determine the pricing group based on the list price. The T-SQL statement should create a view that classifies products into pricing groups based on the list price. The CASE statement is the correct conditional logic to assign each product to the appropriate pricing group. This view will

standardize the pricing group logic across different databases and semantic models.

QUESTION 4

You are the administrator of a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains the following tables:

Table1: A Delta table created by using a shortcut Table2: An external table created by using Spark Table3: A managed table

You plan to connect to Lakehouse1 by using its SQL endpoint. What will you be able to do after connecting to Lakehouse1?

- A. ReadTable3.
- B. Update the data Table3.

C. ReadTable2.

D. Update the data in Table1.

Correct Answer: D

QUESTION 5

You have a Fabric tenant that contains a semantic model. The model contains data about retail stores.

You need to write a DAX query that will be executed by using the XMLA endpoint The query must return a table of stores that have opened since December 1,2023.

How should you complete the DAX expression? 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.

Select and Place:

Values

- DEFINE
- EVALUATE
- FILTER
- SUMMARIZE
- TABLE

Answer Area

```
[ ]  
VAR _SalesSince =  
    DATE ( 2023, 12, 01 )  
[ ]  
FILTER (  
    [ ] ( Store, Store[Name], Store[OpenDate] ),  
    Store[OpenDate] >= _SalesSince  
)
```

Correct Answer:

Values

Answer Area

```
DEFINE  
VAR _SalesSince =  
    DATE ( 2023, 12, 01 )  
EVALUATE  
FILTER (  
    SUMMARIZE ( Store, Store[Name], Store[OpenDate] ),  
    Store[OpenDate] >= _SalesSince  
)
```

The correct order for the DAX expression would be:

```
DEFINE VAR _SalesSince = DATE ( 2023, 12, 01 )  
EVALUATE  
FILTER (  
SUMMARIZE ( Store, Store[Name], Store[OpenDate] ),  
Store[OpenDate] >= _SalesSince )
```

In this DAX query, you're defining a variable `_SalesSince` to hold the date from which you want to filter the stores. `EVALUATE` starts the definition of the query. The `FILTER` function is used to return a table that filters another table or expression. `SUMMARIZE` creates a summary table for the stores, including the `Store[Name]` and `Store[OpenDate]` columns, and the filter expression `Store[OpenDate] >= _SalesSince` ensures only stores opened on or after December 1, 2023, are included in the results.

References =

DAX FILTER Function DAX SUMMARIZE Function

QUESTION 6

You have a Fabric tenant that contains a warehouse.

You use a dataflow to load a new dataset from OneLake to the warehouse.

You need to add a Power Query step to identify the maximum values for the numeric columns.

Which function should you include in the step?

A. Table. MaxN

- B. Table.Max
- C. Table.Range
- D. Table.Profile

Correct Answer: B

Explanation: The Table.Max function should be used in a Power Query step to identify the maximum values for the numeric columns. This function is designed to calculate the maximum value across each column in a table, which suits the requirement of finding maximum values for numeric columns. References = For detailed information on Power Query functions, including Table.Max, please refer to Power Query M function reference.

QUESTION 7

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Nyctaxi_raw. Nyctaxi_raw contains the following columns.

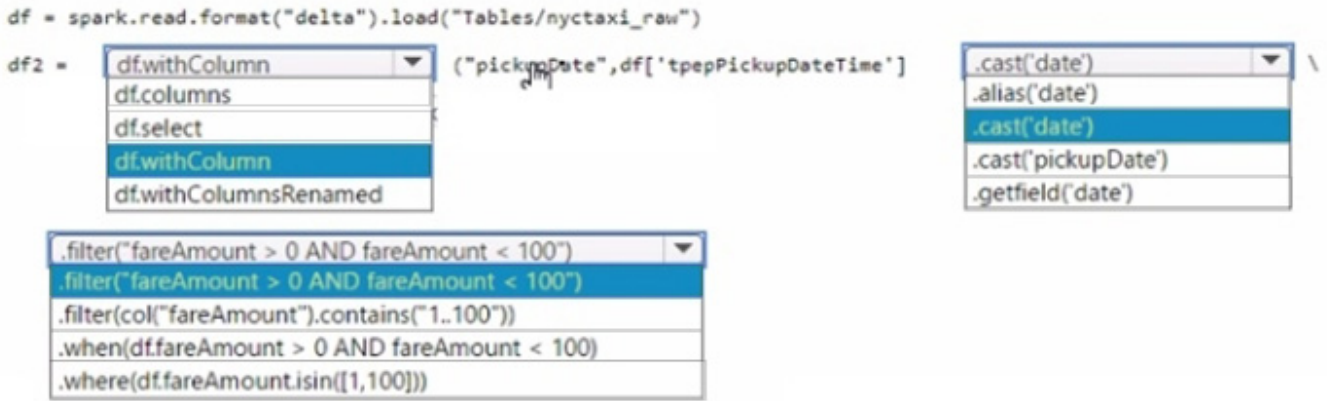
Name	Data type
pickupDateTime	Timestamp
passengerCount	Integer
fareAmount	Double
paymentType	String
tipAmount	Double

You create a Fabric notebook and attach it to lakehouse1.

You need to use PySpark code to transform the data. The solution must meet the following requirements:

Correct Answer:

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")  
df2 = df.withColumn("pickupDate",df['tpepPickupDateTime'].cast("date")).filter("fareAmount > 0 AND fareAmount < 100")
```



The screenshot shows a PySpark IDE with the following code and dropdown menus:

```
df = spark.read.format("delta").load("Tables/nyctaxi_raw")  
df2 = df.withColumn("pickupDate",df['tpepPickupDateTime'].cast("date")).filter("fareAmount > 0 AND fareAmount < 100")
```

The dropdown menus are:

- df.withColumn: df.withColumn, df.columns, df.select, df.withColumn (selected), df.withColumnsRenamed
- .filter("fareAmount > 0 AND fareAmount < 100"): .filter("fareAmount > 0 AND fareAmount < 100") (selected), .filter(col("fareAmount").contains("1..100")), .when(df.fareAmount > 0 AND fareAmount < 100), .where(df.fareAmount.isin([1,100]))
- .cast("date"): .cast("date") (selected), .alias("date"), .cast("pickupDate"), .getfield("date")

Add the pickupDate column: `.withColumn("pickupDate",`

`df["pickupDateTime"].cast("date"))`

Filter the DataFrame: `.filter("fareAmount > 0 AND fareAmount`

In PySpark, you can add a new column to a DataFrame using the `.withColumn` method, where the first argument is the new column name and the second argument is the expression to generate the content of the new column. Here, we use

the `.cast("date")` function to extract only the date part from a timestamp. To filter the DataFrame, you use the `.filter` method with a condition that selects rows where `fareAmount` is greater than 0 and less than 100, thus ensuring only positive

values less than 100 are included.

QUESTION 8

You need to assign permissions for the data store in the AnalyticsPOC workspace. The solution must meet the security requirements.

Which additional permissions should you assign when you share the data store? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

DataEngineers:

Build Reports on the default dataset	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataAnalysts:

Read All Apache Spark	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataScientists:

Read All SQL analytics endpoint data	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

Correct Answer:

DataEngineers:

Build Reports on the default dataset	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataAnalysts:

Read All Apache Spark	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

DataScientists:

Read All SQL analytics endpoint data	▼
Build Reports on the default dataset	
Read All Apache Spark	
Read All SQL analytics endpoint data	

Data Engineers: Read All SQL analytics endpoint data

Data Analysts: Read All Apache Spark

Data Scientists: Read All SQL analytics endpoint data

The permissions for the data store in the AnalyticsPOC workspace should align with the principle of least privilege:

Data Engineers need read and write access but not to datasets or reports.

Data Analysts require read access specifically to the dimensional model objects and the ability to create Power BI reports.

Data Scientists need read access via Spark notebooks. These settings ensure each role has the necessary permissions to fulfill their responsibilities without exceeding their required access level.

QUESTION 9

You have a Fabric tenant that contains a warehouse.

A user discovers that a report that usually takes two minutes to render has been running for 45 minutes and has still not rendered.

You need to identify what is preventing the report query from completing.

Which dynamic management view (DMV) should you use?

- A. sys.dm-exec_requests
- B. sys.dm_exec_sessions
- C. sys.dm_exec_connections
- D. sys.dm_pdw_exec_requests

Correct Answer: D

Explanation: The correct DMV to identify what is preventing the report query from completing is sys.dm_pdw_exec_requests (D). This DMV is specific to Microsoft Analytics Platform System (previously known as SQL Data Warehouse), which is the environment assumed to be used here. It provides information about all queries and load commands currently running or that have recently run. References = You can find more about DMVs in the Microsoft documentation for Analytics Platform System.

QUESTION 10

Which type of data store should you recommend in the AnalyticsPOC workspace?

- A. a data lake
- B. a warehouse
- C. a lakehouse
- D. an external Hive metaStore

Correct Answer: C

Explanation: A lakehouse (C) should be recommended for the AnalyticsPOC workspace. It combines the capabilities of a data warehouse with the flexibility of a data lake. A lakehouse supports semi-structured and unstructured data and allows for T-SQL and Python read access, fulfilling the technical requirements outlined for Litware. References = For further understanding, Microsoft's documentation on the lakehouse architecture provides insights into how it supports various data types and analytical operations.

QUESTION 11

You are analyzing customer purchases in a Fabric notebook by using PySpang You have the following DataFrames:

- transactions: Contains five columns named transaction_id, customer_id, product_id, amount, and date and has 10 million rows, with each row representing a transaction
- customers: Contains customer details in 1,000 rows and three columns named customer_id, name, and country

You need to join the DataFrames on the customer_id column. The solution must minimize data shuffling. You write the following code.

```
from pyspark.sql import functions as F

results =
```

Which code should you run to populate the results DataFrame? A)

```
transactions.join(F.broadcast(customers), transactions.customer_id == customers.customer_id)
```

B)

```
transactions.join(customers, transactions.customer_id == customers.customer_id).distinct()
```

C)

```
transactions.join(customers, transactions.customer_id == customers.customer_id)
```

D)

```
transactions.crossJoin(customers).where(transactions.customer_id == customers.customer_id)
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Correct Answer: A

Explanation: The correct code to populate the results DataFrame with minimal data shuffling is Option A. Using the broadcast function in PySpark is a way to minimize data movement by broadcasting the smaller DataFrame (customers) to each node in the cluster. This is ideal when one DataFrame is much smaller than the other, as in this case with customers. References = You can refer to the official Apache Spark documentation for more details on joins and the broadcast hint.

QUESTION 12

You have a Fabric tenant that contains a new semantic model in OneLake.

You use a Fabric notebook to read the data into a Spark DataFrame.

You need to evaluate the data to calculate the min, max, mean, and standard deviation values for all the string and numeric columns.

Solution: You use the following PySpark expression:

```
df.show()
```

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Explanation: The `df.show()` method also does not meet the goal. It is used to show the contents of the DataFrame, not to compute statistical functions. References = The usage of the `show()` function is documented in the PySpark API documentation.

QUESTION 13

You have a Fabric tenant `tha1` contains a takehouse named `Lakehouse1`. `Lakehouse1` contains a Delta table named `Customer`.

When you query `Customer`, you discover that the query is slow to execute. You suspect that maintenance was NOT performed on the table.

You need to identify whether maintenance tasks were performed on `Customer`.

Solution: You run the following Spark SQL statement:

```
EXPLAIN TABLE customer
```

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Explanation: No, the `EXPLAIN TABLE` statement does not identify whether maintenance tasks were performed on a table. It shows the execution plan for a query. References = The usage and output of the `EXPLAIN` command can be found in the Spark SQL documentation.

QUESTION 14

You need to create a data loading pattern for a Type 1 slowly changing dimension (SCD).

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

NOTE: Each correct answer is worth one point.

- A. Update rows when the non-key attributes have changed.
- B. Insert new rows when the natural key exists in the dimension table, and the non-key attribute values have changed.
- C. Update the effective end date of rows when the non-key attribute values have changed.
- D. Insert new records when the natural key is a new value in the table.

Correct Answer: AD

Explanation: For a Type 1 SCD, you should include actions that update rows when non- key attributes have changed (A), and insert new records when the natural key is a new value in the table (D). A Type 1 SCD does not track historical data, so you always overwrite the old data with the new data for a given key. References = Details on Type 1 slowly changing dimension patterns can be found in data warehousing literature and Microsoft's official documentation.

QUESTION 15

You are creating a semantic model in Microsoft Power BI Desktop.

You plan to make bulk changes to the model by using the Tabular Model Definition Language (TMDL) extension for Microsoft Visual Studio Code.

You need to save the semantic model to a file.

Which file format should you use?

- A. PBIP
- B. PBIX
- C. PBIT
- D. PBIDS

Correct Answer: B

Explanation: When saving a semantic model to a file that can be edited using the Tabular Model Scripting Language (TMSL) extension for Visual Studio Code, the PBIX (Power BI Desktop) file format is the correct choice. The PBIX format contains the report, data model, and queries, and is the primary file format for editing in Power BI Desktop. References = Microsoft's documentation on Power BI file formats and Visual Studio Code provides further clarification on the usage of PBIX files.

[Latest DP-600 Dumps](#)

[DP-600 PDF Dumps](#)

[DP-600 Braindumps](#)