

## Exam Questions DP-600

Implementing Analytics Solutions Using Microsoft Fabric

<https://www.2passeasy.com/dumps/DP-600/>



### NEW QUESTION 1

- (Topic 1)

You need to ensure the data loading activities in the AnalyticsPOC workspace are executed in the appropriate sequence. The solution must meet the technical requirements.

What should you do?

- A. Create a pipeline that has dependencies between activities and schedule the pipeline.
- B. Create and schedule a Spark job definition.
- C. Create a dataflow that has multiple steps and schedule the dataflow.
- D. Create and schedule a Spark notebook.

**Answer:** A

#### Explanation:

To meet the technical requirement that data loading activities must ensure the raw and cleansed data is updated completely before populating the dimensional model, you would need a mechanism that allows for ordered execution. A pipeline in Microsoft Fabric with dependencies set between activities can ensure that activities are executed in a specific sequence. Once set up, the pipeline can be scheduled to run at the required intervals (hourly or daily depending on the data source).

### NEW QUESTION 2

- (Topic 2)

You have a Fabric tenant that contains a lakehouse. You plan to use a visual query to merge two tables.

You need to ensure that the query returns all the rows that are present in both tables. Which type of join should you use?

- A. left outer
- B. right anti
- C. full outer
- D. left anti
- E. right outer
- F. inner

**Answer:** C

#### Explanation:

When you need to return all rows that are present in both tables, you use a full outer join. This type of join combines the results of both left and right outer joins and returns all rows from both tables, with matching rows from both sides where available. If there is no match, the result is NULL on the side of the join where there is no match. References: Information about joins and their use in querying data in a lakehouse can be typically found in the SQL and data processing documentation of the Fabric tenant or lakehouse solutions.

### NEW QUESTION 3

- (Topic 2)

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.explain()`

Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

#### Explanation:

The `df.explain()` method does not meet the goal of evaluating data to calculate statistical functions. It is used to display the physical plan that Spark will execute. References = The correct usage of the `explain()` function can be found in the PySpark documentation.

### NEW QUESTION 4

- (Topic 2)

You have a Fabric tenant that contains a lakehouse named Lakehouse1. Lakehouse1 contains a subfolder named Subfolder1 that contains CSV files. You need to convert the CSV files into the delta format that has V-Order optimization enabled. What should you do from Lakehouse explorer?

- A. Use the Load to Tables feature.
- B. Create a new shortcut in the Files section.
- C. Create a new shortcut in the Tables section.
- D. Use the Optimize feature.

**Answer:** D

#### Explanation:

To convert CSV files into the delta format with Z-Order optimization enabled, you should use the Optimize feature (D) from Lakehouse Explorer. This will allow you to optimize the file organization for the most efficient querying. References = The process for converting and optimizing file formats within a lakehouse is discussed in the lakehouse management documentation.

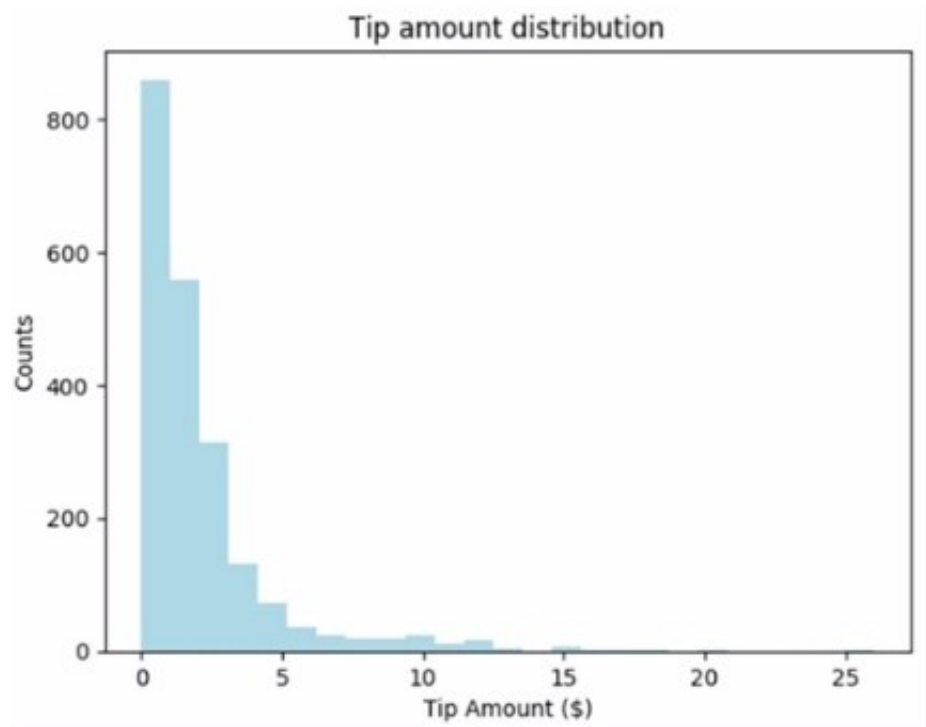
### NEW QUESTION 5

- (Topic 2)

You have a Fabric notebook that has the Python code and output shown in the following exhibit.

```
# Look at a histogram of tips by count by using Matplotlib

ax1 = sampled_taxi_pd_df['tipAmount'].plot(kind='hist', bins=25, facecolor='lightblue')
ax1.set_title('Tip amount distribution')
ax1.set_xlabel('Tip Amount ($)')
ax1.set_ylabel('Counts')
plt.suptitle('')
plt.show()
```



Which type of analytics are you performing?

- A. predictive
- B. descriptive
- C. prescriptive
- D. diagnostic

**Answer: B**

**Explanation:**

The Python code and output shown in the exhibit display a histogram, which is a representation of the distribution of data. This kind of analysis is descriptive analytics, which is used to describe or summarize the features of a dataset. Descriptive analytics answers the question of "what has happened" by providing insight into past data through tools such as mean, median, mode, standard deviation, and graphical representations like histograms.

References: Descriptive analytics and the use of histograms as a way to visualize data distribution are basic concepts in data analysis, often covered in introductory analytics and Python programming resources.

**NEW QUESTION 6**

- (Topic 2)

You have a Fabric tenant named Tenant1 that contains a workspace named WS1. WS1 uses a capacity named C1 and contains a dataset named DS1. You need to ensure read- write access to DS1 is available by using the XMLA endpoint. What should be modified first?

- A. the DS1 settings
- B. the WS1 settings
- C. the C1 settings
- D. the Tenant1 settings

**Answer: C**

**Explanation:**

To ensure read-write access to DS1 is available by using the XMLA endpoint, the C1 settings (which refer to the capacity settings) should be modified first. XMLA endpoint configuration is a capacity feature, not specific to individual datasets or workspaces. References = The configuration of XMLA endpoints in Power BI capacities is detailed in the Power BI documentation on dataset management.

**NEW QUESTION 7**

- (Topic 2)

You have a Fabric tenant that contains a semantic model. The model uses Direct Lake mode.

You suspect that some DAX queries load unnecessary columns into memory. You need to identify the frequently used columns that are loaded into memory.

What are two ways to achieve the goal? Each correct answer presents a complete solution. NOTE: Each correct answer is worth one point.

- A. Use the Analyze in Excel feature.
- B. Use the Vertipaq Analyzer tool.
- C. Query the \$system.discovered\_STORAGE\_TABLE\_COLUMN-IN\_SEGMENTS dynamic management view (DMV).
- D. Query the discover\_hehory6Rant dynamic management view (DMV).

**Answer: BC**

**Explanation:**

The Vertipaq Analyzer tool (B) and querying the

\$system.discovered\_STORAGE\_TABLE\_COLUMNS\_IN\_SEGMENTS dynamic management view (DMV) (C) can help identify which columns are frequently loaded into memory. Both methods provide insights into the storage and retrieval aspects of the semantic model. References = The Power BI documentation on

Vertipaq Analyzer and DMV queries offers detailed guidance on how to use these tools for performance analysis.

NEW QUESTION 8

- (Topic 2)

You have a Fabric tenant that contains a machine learning model registered in a Fabric workspace. You need to use the model to generate predictions by using the predict function in a fabric notebook. Which two languages can you use to perform model scoring? Each correct answer presents a complete solution. NOTE: Each correct answer is worth one point.

- A. T-SQL
- B. DAX EC.
- C. Spark SQL
- D. PySpark

Answer: CD

Explanation:

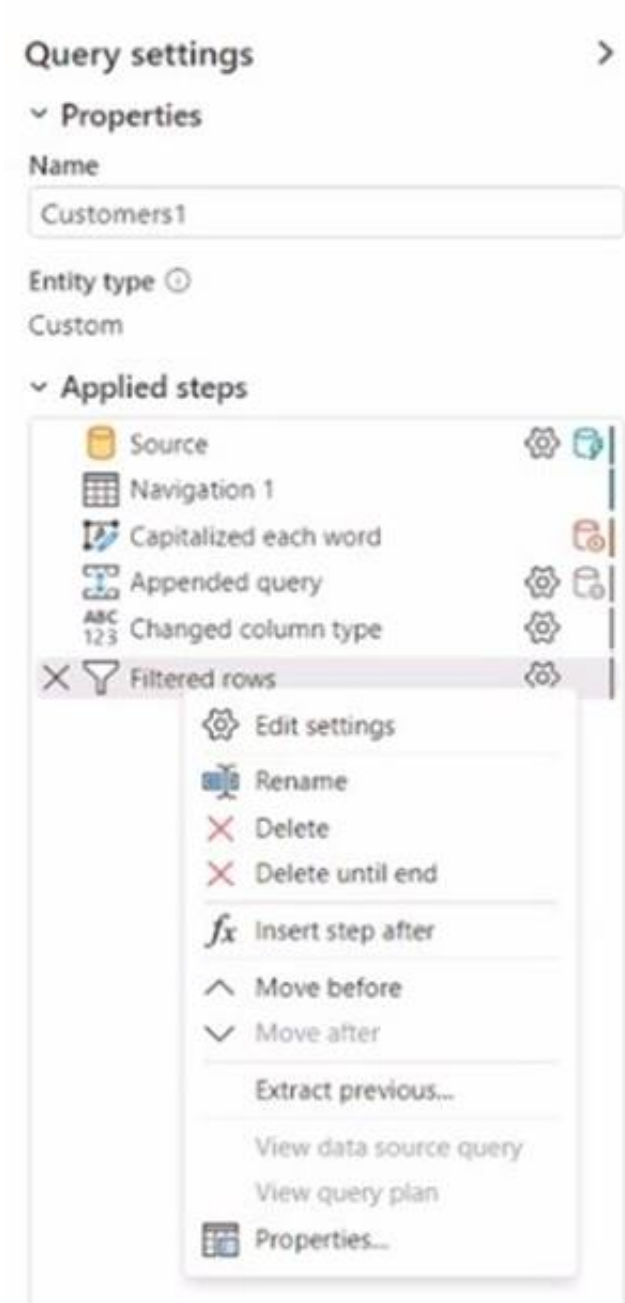
The two languages you can use to perform model scoring in a Fabric notebook using the predict function are Spark SQL (option C) and PySpark (option D). These are both part of the Apache Spark ecosystem and are supported for machine learning tasks in a Fabric environment. References = You can find more information about model scoring and supported languages in the context of Fabric notebooks in the official documentation on Azure Synapse Analytics.

NEW QUESTION 9

HOTSPOT - (Topic 2)

You have a Fabric tenant that contains two lakehouses.

You are building a dataflow that will combine data from the lakehouses. The applied steps from one of the queries in the dataflow is shown in the following exhibit.



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.

Answer Area

[Answer choice] of the transformation steps in the query will fold.

Some

All

None

Some

The Added custom step will be performed in [answer choice].

the Microsoft Power Query engine

each lakehouse's query engine

the Microsoft Power Query engine

the source lakehouse query engine

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Folding in Power Query refers to operations that can be translated into source queries. In this case, "some" of the steps can be folded, which means that some

transformations will be executed at the data source level. The steps that cannot be folded will be executed within the Power Query engine. Custom steps, especially those that are not standard query operations, are usually executed within Power Query engine rather than being pushed down to the source system.

References =

? Query folding in Power Query

? Power Query M formula language

#### NEW QUESTION 10

- (Topic 2)

You are analyzing customer purchases in a Fabric notebook by using PySpark. 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

**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.

#### NEW QUESTION 10

- (Topic 2)

You have a Fabric tenant that contains a lakehouse named Lakehouse1.

You need to prevent new tables added to Lakehouse1 from being added automatically to the default semantic model of the lakehouse.

What should you configure? (5)

A. the semantic model settings

B. the Lakehouse1 settings

C. the workspace settings

D. the SQL analytics endpoint settings

**Answer: A**

#### Explanation:

To prevent new tables added to Lakehouse1 from being automatically added to the default semantic model, you should configure the semantic model settings. There should be an option within the settings of the semantic model to include or exclude new tables by default. By adjusting these settings, you can control the automatic inclusion of new tables.

References: The management of semantic models and their settings would be covered under the documentation for the semantic layer or modeling features of the Fabric tenant's lakehouse solution.

#### NEW QUESTION 14

- (Topic 2)

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

**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.



#### NEW QUESTION 18

- (Topic 2)

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

**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.

#### NEW QUESTION 22

- (Topic 2)

You have a Microsoft Power BI semantic model.

You need to identify any surrogate key columns in the model that have the Summarize By property set to a value other than to None. The solution must minimize effort.

What should you use?

- A. DAX Formatter in DAX Studio
- B. Model view in Microsoft Power BI Desktop
- C. Model explorer in Microsoft Power BI Desktop
- D. Best Practice Analyzer in Tabular Editor

**Answer: D**

#### Explanation:

To identify surrogate key columns with the "Summarize By" property set to a value other than "None," the Best Practice Analyzer in Tabular Editor is the most efficient tool. The Best Practice Analyzer can analyze the entire model and provide a report on all columns that do not meet a specified best practice, such as having the "Summarize By" property set correctly for surrogate key columns. Here's how you would proceed:

? Open your Power BI model in Tabular Editor.

? Go to the Advanced Scripting window.

? Write or use an existing script that checks the "Summarize By" property of each column.

? Execute the script to get a report on the surrogate key columns that do not have their "Summarize By" property set to "None".

? You can then review and adjust the properties of the columns directly within the Tabular Editor.

References: The functionality of the Best Practice Analyzer in Tabular Editor is documented in the community and learning resources for Power BI.

#### NEW QUESTION 27

- (Topic 2)

You have a Fabric tenant.

You are creating a Fabric Data Factory pipeline.

You have a stored procedure that returns the number of active customers and their average sales for the current month.

You need to add an activity that will execute the stored procedure in a warehouse. The returned values must be available to the downstream activities of the pipeline.

Which type of activity should you add?

- A. Stored procedure
- B. Get metadata
- C. Lookup
- D. Copy data

**Answer: C**

#### Explanation:

In a Fabric Data Factory pipeline, to execute a stored procedure and make the returned values available for downstream activities, the Lookup activity is used. This activity can retrieve a dataset from a data store and pass it on for further processing. Here's how you would use the Lookup activity in this context:

? Add a Lookup activity to your pipeline.

? Configure the Lookup activity to use the stored procedure by providing the necessary SQL statement or stored procedure name.

? In the settings, specify that the activity should use the stored procedure mode.

? Once the stored procedure executes, the Lookup activity will capture the results and make them available in the pipeline's memory.

? Downstream activities can then reference the output of the Lookup activity. References: The functionality and use of Lookup activity within Azure Data Factory is documented in Microsoft's official documentation for Azure Data Factory, under the section for pipeline activities.

#### NEW QUESTION 30

- (Topic 2)

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

**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.

**NEW QUESTION 31**

.....

## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-600 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-600 Product From:

<https://www.2passeasy.com/dumps/DP-600/>

## Money Back Guarantee

### DP-600 Practice Exam Features:

- \* DP-600 Questions and Answers Updated Frequently
- \* DP-600 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-600 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-600 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year