



Microsoft

Exam Questions DP-600

Implementing Analytics Solutions Using Microsoft Fabric

About ExamBible

[Your Partner of IT Exam](#)

Found in 1998

ExamBible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, ExamBible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1

HOTSPOT - (Topic 1)

You need to create a DAX measure to calculate the average overall satisfaction score.

How should you complete the DAX code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Rolling 12 Overall Satisfaction =

.

VAR NumberOfMonths = 12

VAR LastCurrentDate = MAX ('Date'[Date])

VAR Period = DATESINPERIOD ('Date'[Date], LastCurrentDate, - NumberOfMonths, MONTH)

VAR Result =

CALCULATE (

)

RETURN

Result

Answer Area

Rolling 12 Overall Satisfaction =

.

VAR NumberOfMonths = 12

VAR LastCurrentDate = MAX ('Date'[Date])

VAR Period = DATESINPERIOD ('Date'[Date], LastCurrentDate, - NumberOfMonths, MONTH)

VAR Result =

CALCULATE (

AVERAGE('Survey'[Response Value]),

AVERAGE('Survey'[Response Value]),

AVERAGEA('Question'[Question Text]),

AVERAGEX(VALUES('Survey'[Customer Key]),

NumberOfMonths,

LastCurrentDate,

NumberOfMonths,

Period,

)

RETURN

Result

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? The measure should use the AVERAGE function to calculate the average value.

? It should reference the Response Value column from the 'Survey' table.

? The 'Number of months' should be used to define the period for the average calculation.

To calculate the average overall satisfaction score using DAX, you would need to use the AVERAGE function on the response values related to satisfaction questions. The DATESINPERIOD function will help in calculating the rolling average over the last 12 months.

NEW QUESTION 2

- (Topic 1)

You need to implement the date dimension in the data store. The solution must meet the technical requirements.

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

- A. Populate the date dimension table by using a dataflow.
- B. Populate the date dimension table by using a Stored procedure activity in a pipeline.
- C. Populate the date dimension view by using T-SQL.
- D. Populate the date dimension table by using a Copy activity in a pipeline.

Answer: AB

Explanation:

Both a dataflow (A) and a Stored procedure activity in a pipeline (B) are capable of creating and populating a date dimension table. A dataflow can perform the transformation needed to create the date dimension, and it aligns with the preference for using low-code tools for data ingestion when possible. A Stored procedure could be written to generate the necessary date dimension data and executed within a pipeline, which also adheres to the technical requirements for the PoC.

Your Partner of IT Exam

visit - <https://www.examible.com>

NEW QUESTION 3

HOTSPOT - (Topic 2)

You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains three schemas named schemaA, schemaB. and schemaC. You need to ensure that a user named User1 can truncate tables in schemaA only.

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.

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? GRANT ALTER ON SCHEMA::schemaA TO User1;

The ALTER permission allows a user to modify the schema of an object, and granting ALTER on a schema will allow the user to perform operations like TRUNCATE TABLE on any object within that schema. It is the correct permission to grant to User1 for truncating tables in schemaA.

References =

? GRANT Schema Permissions

? Permissions That Can Be Granted on a Schema

NEW QUESTION 4

- (Topic 2)

You have a Fabric tenant that uses a Microsoft tower BI Premium capacity. You need to enable scale-out for a semantic model. What should you do first?

- A. At the semantic model level, set Large dataset storage format to Off.
- B. At the tenant level, set Create and use Metrics to Enabled.
- C. At the semantic model level, set Large dataset storage format to On.
- D. At the tenant level, set Data Activator to Enabled.

Answer: C

Explanation:

To enable scale-out for a semantic model, you should first set Large dataset storage format to On (C) at the semantic model level. This configuration is necessary to handle larger datasets effectively in a scaled-out environment. References = Guidance on configuring large dataset storage formats for scale-out is available in the Power BI documentation.

NEW QUESTION 5

HOTSPOT - (Topic 2)

You have a Fabric tenant that contains a warehouse named Warehouse1. Warehouse1 contains a fact table named FactSales that has one billion rows. You run the following T- SQL statement.

CREATE TABLE test.FactSales AS CLONE OF Dbo.FactSales;

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
A replica of dbo.Sales is created in the test schema by copying the metadata only.	<input type="radio"/>	<input type="radio"/>
Additional schema changes to dbo.FactSales will also apply to test.FactSales.	<input type="radio"/>	<input type="radio"/>
Additional data changes to dbo.FactSales will also apply to test.FactSales.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? A replica of dbo.Sales is created in the test schema by copying the metadata only.

- No

? Additional schema changes to dbo.FactSales will also apply to test.FactSales. - No

? Additional data changes to dbo.FactSales will also apply to test.FactSales. - Yes

The CREATE TABLE AS CLONE statement creates a copy of an existing table, including its data and any associated data structures, like indexes. Therefore, the statement does not merely copy metadata; it also copies the data. However, subsequent schema changes to the original table do not automatically propagate to the cloned table. Any data changes in the original table after the clone operation will not be reflected in the clone unless explicitly updated.

References =

? CREATE TABLE AS SELECT (CTAS) in SQL Data Warehouse

NEW QUESTION 6

DRAG DROP - (Topic 2)

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.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 7

- (Topic 2)

You have a semantic model named Model 1. Model 1 contains five tables that all use Import mode. Model1 contains a dynamic row-level security (RLS) role named HR. The HR role filters employee data so that HR managers only see the data of the department to which they are assigned.

You publish Model1 to a Fabric tenant and configure RLS role membership. You share the model and related reports to users.

An HR manager reports that the data they see in a report is incomplete. What should you do to validate the data seen by the HR Manager?

- A. Ask the HR manager to open the report in Microsoft Power BI Desktop.
- B. Select Test as role to view the data as the HR role.
- C. Select Test as role to view the report as the HR manager,
- D. Filter the data in the report to match the intended logic of the filter for the HR department.

Answer: B

Explanation:

To validate the data seen by the HR manager, you should use the 'Test as role' feature in Power BI service. This allows you to see the data exactly as it would appear for the HR role, considering the dynamic RLS setup. Here is how you would proceed:

? Navigate to the Power BI service and locate Model1.

? Access the dataset settings for Model1.

? Find the security/RLS settings where you configured the roles.

? Use the 'Test as role' feature to simulate the report viewing experience as the HR role.

? Review the data and the filters applied to ensure that the RLS is functioning correctly.

? If discrepancies are found, adjust the RLS expressions or the role membership as needed.

References: The 'Test as role' feature and its use for validating RLS in Power BI is covered in the Power BI documentation available on Microsoft's official documentation.

NEW QUESTION 8

HOTSPOT - (Topic 2)

You have a data warehouse that contains a table named Stage. Customers. Stage- Customers contains all the customer record updates from a customer relationship management (CRM) system. There can be multiple updates per customer

You need to write a T-SQL query that will return the customer ID, name, postal code, and the last updated time of the most recent row for each customer ID.

How should you complete the code? To answer, select the appropriate options in the answer area,

NOTE Each correct selection is worth one point.

Answer Area

```
WITH CUSTOMERBASE AS (
    SELECT [CustomerID]
    , [CustomerName]
    , [PostalCode]
    , [LastUpdated]
    , X = ROW_NUMBER() OVER (PARTITION BY CustomerID ORDER BY LastUpdated DESC)
    FROM CUSTOMERBASE
    WHERE X = 1
)
```

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

? In the ROW_NUMBER() function, choose OVER (PARTITION BY CustomerID ORDER BY LastUpdated DESC).

? In the WHERE clause, choose WHERE X = 1.

To select the most recent row for each customer ID, you use the ROW_NUMBER() window function partitioned by CustomerID and ordered by LastUpdated in descending order. This will assign a row number of 1 to the most recent update for each customer. By selecting rows where the row number (X) is 1, you get the latest update per customer. References =

- ? Use the OVER clause to aggregate data per partition
 ? Use window functions

NEW QUESTION 9

- (Topic 2)

You have a Fabric tenant that contains a lakehouse named lakehouse1. Lakehouse1 contains an unpartitioned table named Table1.

You plan to copy data to Table1 and partition the table based on a date column in the source data.

You create a Copy activity to copy the data to Table1.

You need to specify the partition column in the Destination settings of the Copy activity. What should you do first?

- A. From the Destination tab, set Mode to Append.
 B. From the Destination tab, select the partition column,
 C. From the Source tab, select Enable partition discovery
 D. From the Destination tab, set Mode to Overwrite.

Answer: B

Explanation:

Before specifying the partition column in the Destination settings of the Copy activity, you should set Mode to Append (A). This will allow the Copy activity to add data to the table while taking the partition column into account. References = The configuration options for Copy activities and partitioning in Azure Data Factory, which are applicable to Fabric dataflows, are outlined in the official Azure Data Factory documentation.

NEW QUESTION 10

HOTSPOT - (Topic 2)

You have a Fabric warehouse that contains a table named Sales.Orders. Sales.Orders contains the following columns.

Name	Data type	Nullable
OrderID	Integer	No
CustomerID	Integer	No
OrderDate	Date	No
Quantity	Integer	Yes
Weight	Decimal(18, 3)	Yes
ListPrice	Decimal(18, 2)	No
SalePrice	Decimal(18, 2)	Yes

You need to write a T-SQL query that will return the following columns.

Name	Description
OrderID	Returns OrderID
CustomerID	Returns CustomerID
PeriodDate	Returns a date representing the first day of the month for OrderDate
DayName	Returns the name of the day for OrderDate, such as Wednesday

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Abstract

SELECT OrderID, CustomerID,

DATEFROMPARTS
DATE_BUCKET
DATEFROMPARTS
DATEPART
DATETRUNC

FROM

DATENAME(
weekday
day
dayofyear
weekday

, OrderDate) AS DayName

FROM Sales.Ord

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

For the PeriodDate that returns the first day of the month for OrderDate, you should use DATEFROMPARTS as it allows you to construct a date from its individual components (year, month, day).

For the `DayName` that returns the name of the day for `OrderDate`, you should use

DATENAME with the weekday date part to get the full name of the weekday. The complete SQL query should look like this:

```
SELECT OrderID, CustomerID,
```

DATEFROMPARTS(YEAR(OrderDate), MONTH(OrderDate), 1) AS PeriodDate, DATENAME(weekday, OrderDate) AS DayName

FROM Sales.Orders

Select DATEFROMPARTS for the PeriodDate and weekday for the DayName in the answer area.

NEW QUESTION 10

- (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 14

HOTSPOT - (Topic 2)

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:

- Add a column named `pickupDate` that will contain only the date portion of `pickupDateTime`.
- Filter the `DataFrame` to include only rows where `fareAmount` is a positive number that is less than 100.

How should you complete the code? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

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

df2 =

- df.withColumn
- df.columns
- df.select
- df.withColumn
- df.withColumnsRenamed

filter("fareAmount > 0 AND fareAmount < 100")

- filter("fareAmount > 0 AND fareAmount < 100")
- filter(col("fareAmount").contains("1,100"))
- when(df.fareAmount > 0 AND fareAmount < 100)
- where(df.fareAmount.isin([1,100]))

cast("date")

- alias("date")
- cast("date")
- cast("pickupDate")
- getField("date")

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

? Add the pickupDate column: .withColumn("pickupDate", df["tpepPickupDateTime"].cast("date"))
 ? Filter the DataFrame: .filter("fareAmount > 0 AND fareAmount < 100")

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.

NEW QUESTION 19

- (Topic 2)

You have a Fabric tenant that contains a semantic model. The model contains 15 tables.

You need to programmatically change each column that ends in the word Key to meet the following requirements:

- Hide the column.
- Set Nullable to False.
- Set Summarize By to None
- Set Available in MDX to False.
- Mark the column as a key column. What should you use?

- A. Microsoft Power BI Desktop
 B. Tabular Editor
 C. ALM Toolkit
 D. DAX Studio

Answer: B

Explanation:

Tabular Editor is an advanced tool for editing Tabular models outside of Power BI Desktop that allows you to script out changes and apply them across multiple columns or tables. To accomplish the task programmatically, you would:

- ? Open the model in Tabular Editor.
- ? Create an Advanced Script using C# to iterate over all tables and their respective columns.
- ? Within the script, check if the column name ends with 'Key'.
- ? For columns that meet the condition, set the properties accordingly: IsHidden = true, IsNullable = false, SummarizeBy = None, IsAvailableInMDX = false.
- ? Additionally, mark the column as a key column.
- ? Save the changes and deploy them back to the Fabric tenant.

References: The ability to batch-edit properties using scripts in Tabular Editor is well- documented in the tool's official documentation and user community resources.

NEW QUESTION 20

DRAG DROP - (Topic 2)

You are creating a dataflow in Fabric to ingest data from an Azure SQL database by using a T-SQL statement.

You need to ensure that any foldable Power Query transformation steps are processed by the Microsoft SQL Server engine.

How should you complete the code? 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.

Values

- EnableFolding
- NativeQuery
- Optimize
- Record
- StopFolding
- Table
- Value

Answer Area

```
let
    Source = Sql.Databases(
        "server.database.windows.net"
    ),
    Database = Source[{"Name": "db"}][Data],
    Query = [ ] (
        Database,
        "SELECT * FROM customer WHERE country IN ('USA', 'UK')",
        null,
        [ ] = true
    )
in
    Query
```

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

You should complete the code as follows:

? Table

? NativeQuery

? EnableFolding

In Power Query, using Table before the SQL statement ensures that the result of the SQL query is treated as a table. NativeQuery allows a native database query to be passed through from Power Query to the source database. The EnableFolding option ensures that any subsequent transformations that can be folded will be sent back and executed at the source database (Microsoft SQL Server engine in this case).

NEW QUESTION 24

HOTSPOT - (Topic 2)

You have a Fabric workspace that uses the default Spark starter pool and runtime version 1,2.

You plan to read a CSV file named Sales.raw.csv in a lakehouse, select columns, and save the data as a Delta table to the managed area of the lakehouse.

Sales_raw.csv contains 12 columns.

You have the following code.

```
from pyspark.sql.functions import year

(spark
 .read
 .format("csv")
 .option("header", 'true')
 .load("Files/sales_raw.csv")
 .select('SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice')
 .withColumn("Year", year("OrderDate"))
 .write
 .partitionBy('Year')
 .saveAsTable("sales")
)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv.	<input type="radio"/>	<input type="radio"/>
Removing the partition will reduce the execution time of the query.	<input type="radio"/>	<input type="radio"/>
Adding inferSchema=true to the options will increase the execution time of the query.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

? The Spark engine will read only the 'SalesOrderNumber', 'OrderDate', 'CustomerName', 'UnitPrice' columns from Sales_raw.csv. - Yes

? Removing the partition will reduce the execution time of the query. - No

? Adding inferSchema=true to the options will increase the execution time of the query. - Yes

The code specifies the selection of certain columns, which means only those columns will be read into the DataFrame. Partitions in Spark are a way to optimize the execution of queries by organizing the data into parts that can be processed in parallel. Removing the partition could potentially increase the execution time because Spark would no longer be

able to process the data in parallel efficiently. The inferSchema option allows Spark to automatically detect the column data types, which can increase the execution time of the initial read operation because it requires Spark to read through the data to infer the schema.

NEW QUESTION 26

- (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.summary()

Does this meet the goal?

- A. Yes
B. No

Answer: A

Explanation:

Yes, the df.summary() method does meet the goal. This method is used to compute specified statistics for numeric and string columns. By default, it provides statistics such as count, mean, stddev, min, and max. References = The PySpark API documentation details the summary() function and the statistics it provides.

NEW QUESTION 29

DRAG DROP - (Topic 2)

You create a semantic model by using Microsoft Power BI Desktop. The model contains one security role named SalesRegionManager and the following tables:

- Sales
- SalesRegion
- Sales Address

You need to modify the model to ensure that users assigned the SalesRegionManager role cannot see a column named Address in Sales Address.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

⋮ Open the model in Power BI Desktop.

⋮ Set Object Level Security to **Default** for SalesRegionManager.

⋮ Set the Hidden property to **True**.

⋮ Open the model in Tabular Editor.

⋮ Select the **Address** column in SalesAddress.

⋮ Set Object Level Security to **None** for SalesRegionManager.

Answer Area

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

To ensure that users assigned the SalesRegionManager role cannot see the Address column in the SalesAddress table, follow these steps in sequence:

- ? Open the model in Tabular Editor.
- ? Select the Address column in SalesAddress.
- ? Set Object Level Security to None for SalesRegionManager.

NEW QUESTION 33

- (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.show()`

Does this meet the goal?

- A. Yes
B. No

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.

NEW QUESTION 38

HOTSPOT - (Topic 2)

You have a Fabric workspace named Workspace1 and an Azure Data Lake Storage Gen2 account named storage!. Workspace1 contains a lakehouse named Lakehouse1.

You need to create a shortcut to storage! in Lakehouse1.

Which connection and endpoint should you specify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Connection:

abfss

abfs

abfss

https

Endpoint:

dfs

blob

dfs

file

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

When creating a shortcut to an Azure Data Lake Storage Gen2 account in a lakehouse, you should use the abfss (Azure Blob File System Secure) connection string and the dfs (Data Lake File System) endpoint. The abfss is used for secure access to Azure Data Lake Storage, and the dfs endpoint indicates that the Data Lake Storage Gen2 capabilities are to be used.

NEW QUESTION 42

- (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

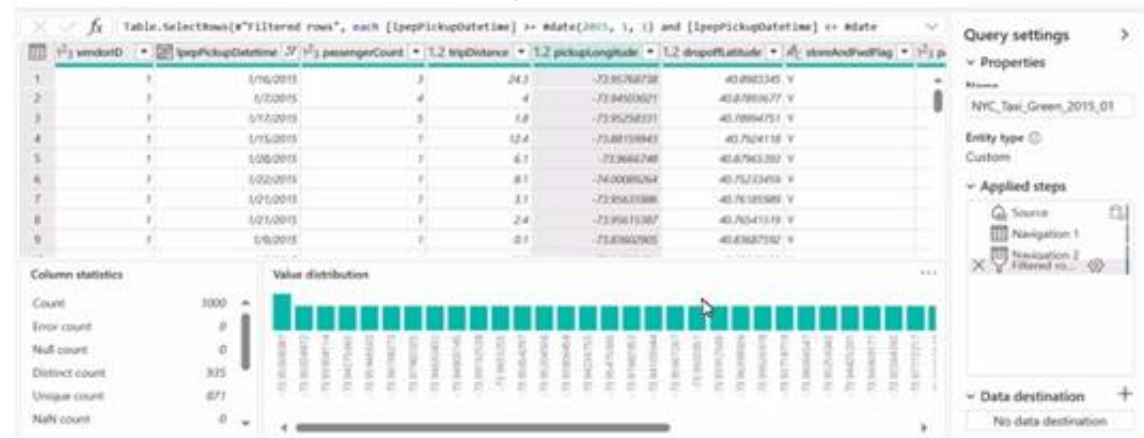
Explanation:

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 45

- (Topic 2)

You have a Fabric workspace named Workspace 1 that contains a dataflow named Dataflow1. Dataflow1 has a query that returns 2,000 rows. You view the query in Power Query as shown in the following exhibit.



What can you identify about the pickupLongitude column?

- A. The column has duplicate values.
 B. All the table rows are profiled.
 C. The column has missing values.
 D. There are 935 values that occur only once.

Answer: B

Explanation:

The pickupLongitude column has duplicate values. This can be inferred because the 'Distinct count' is 935 while the 'Count' is 1000, indicating that there are repeated values within the column. References = Microsoft Power BI documentation on data profiling could provide further insights into understanding and interpreting column statistics like these.

NEW QUESTION 48

- (Topic 2)

You have a Fabric tenant that contains a workspace named Workspace^ Workspace1 is assigned to a Fabric capacity.

You need to recommend a solution to provide users with the ability to create and publish custom Direct Lake semantic models by using external tools. The solution must follow the principle of least privilege.

Which three actions in the Fabric Admin portal should you include in the recommendation? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. From the Tenant settings, set Allow XMLA Endpoints and Analyze in Excel with on- premises datasets to Enabled
 B. From the Tenant settings, set Allow Azure Active Directory guest users to access Microsoft Fabric to Enabled
 C. From the Tenant settings, select Users can edit data models in the Power BI service.
 D. From the Capacity settings, set XMLA Endpoint to Read Write
 E. From the Tenant settings, set Users can create Fabric items to Enabled
 F. From the Tenant settings, enable Publish to Web

Answer: ACD

Explanation:

For users to create and publish custom Direct Lake semantic models using external tools, following the principle of least privilege, the actions to be included are enabling XMLA Endpoints (A), editing data models in Power BI service (C), and setting XMLA Endpoint to Read-Write in the capacity settings (D). References = More information can be found in the Admin portal of the Power BI service documentation, detailing tenant and capacity settings.

NEW QUESTION 51

DRAG DROP - (Topic 2)

You are implementing two dimension tables named Customers and Products in a Fabric warehouse.

You need to use slowly changing dimension (SCD) to manage the versioning of data. The solution must meet the requirements shown in the following table.

Table	Change action
Customers	Create a new version of the row.
Products	Overwrite the existing value in the latest row.

Which type of SCD should you use for each table? To answer, drag the appropriate SCD types to the correct tables. Each SCD type 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.

SCD Types

Type 0

Type 1

Type 2

Type 3

Answer Area

Customers:

Products:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

For the Customers table, where the requirement is to create a new version of the row, you would use:

? Type 2 SCD: This type allows for the creation of a new record each time a change occurs, preserving the history of changes over time.

For the Products table, where the requirement is to overwrite the existing value in the latest row, you would use:

? Type 1 SCD: This type updates the record directly, without preserving historical data.

NEW QUESTION 56

- (Topic 2)

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.

Answer: D

NEW QUESTION 58

HOTSPOT - (Topic 2)

You have a Microsoft Power BI semantic model. You plan to implement calculation groups.

You need to create a calculation item that will change the context from the selected date to month-to-date (MTD).

How should you complete the DAX expression? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

CALCULATE

CALCULATE

GENERATE

MEASURE

SELECTEDMEASURE

COMBINEVALUES

SELECTEDMEASURE

SELECTEDVALUE

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

To create a calculation item that changes the context from the selected date to month-to-date (MTD), the appropriate DAX expression involves using the CALCULATE function to alter the filter context and the DATESMTD function to specify the month-to-date context. The correct completion for the DAX expression would be:

? In the first dropdown, select CALCULATE.

? In the second dropdown, select SELECTEDMEASURE. This would create a DAX expression in the form:

```
CALCULATE( SELECTEDMEASURE(),  
DATESMTD('Date'[DateColumn])  
)
```

NEW QUESTION 60

- (Topic 2)

You have a Fabric tenant that contains a table 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:

REFRESH TABLE customer Does this meet the goal?

A. Yes

B. No

Answer: B

Explanation:

No, the REFRESH TABLE statement does not provide information on whether maintenance tasks were performed. It only updates the metadata of a table to reflect any changes on the data files. References = The use and effects of the REFRESH TABLE command are explained in the Spark SQL documentation.

NEW QUESTION 64

- (Topic 2)

You have a Fabric tenant that contains a warehouse.

You are designing a star schema model that will contain a customer dimension. The customer dimension table will be a Type 2 slowly changing dimension (SCD).

You need to recommend which columns to add to the table. The columns must NOT already exist in the source.

Which three types of columns should you recommend? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

A. an effective end date and time

B. a foreign key

C. a surrogate key

D. a natural key

E. an effective start date and time

Answer: ACE

Explanation:

For a Type 2 slowly changing dimension (SCD), you typically need to add the following types of columns that do not exist in the source system:

? An effective start date and time (E): This column records the date and time from which the data in the row is effective.

? An effective end date and time (A): This column indicates until when the data in the row was effective. It allows you to keep historical records for changes over time.

? A surrogate key (C): A surrogate key is a unique identifier for each row in a table, which is necessary for Type 2 SCDs to differentiate between historical and current records.

References: Best practices for designing slowly changing dimensions in data warehousing solutions, which include Type 2 SCDs, are commonly discussed in data warehousing and business intelligence literature and would be part of the modeling guidance in a Fabric tenant's documentation.

NEW QUESTION 68

- (Topic 2)

You have an Azure Repos Git repository named Repo1 and a Fabric-enabled Microsoft Power BI Premium capacity. The capacity contains two workspaces named Workspace1 and Workspace2. Git integration is enabled at the workspace level.

You plan to use Microsoft Power BI Desktop and Workspace1 to make version-controlled changes to a semantic model stored in Repo1. The changes will be built and deployed to Workspace2 by using Azure Pipelines.

You need to ensure that report and semantic model definitions are saved as individual text files in a folder hierarchy. The solution must minimize development and maintenance effort.

In which file format should you save the changes?

A. PBIP

B. PBIT

C. PBIX

D. PBIDS

Answer: C

Explanation:

When working with Power BI Desktop and Git integration for version control, report and semantic model definitions should be saved in the PBIX format. PBIX is the Power BI Desktop file format that contains definitions for reports, data models, and queries, and it can be easily saved and tracked in a version-controlled environment. The solution should minimize development and maintenance effort, and saving in PBIX format allows for the easiest transition from development to deployment, especially when using Azure Pipelines for CI/CD (continuous integration/continuous deployment) practices.

References: The use of PBIX files with Power BI Desktop and Azure Repos for version control is discussed in Microsoft's official Power BI documentation, particularly in the sections covering Power BI Desktop files and Azure DevOps integration.

NEW QUESTION 69

- (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 73

- (Topic 2)
You are analyzing the data in a Fabric notebook.
You have a Spark DataFrame assigned to a variable named df.
You need to use the Chart view in the notebook to explore the data manually. Which function should you run to make the data available in the Chart view?

- A. displayMTML
- B. show
- C. write
- D. display

Answer: D

Explanation:

The display function is the correct choice to make the data available in the Chart view within a Fabric notebook. This function is used to visualize Spark DataFrames in various formats including charts and graphs directly within the notebook environment. References = Further explanation of the display function can be found in the official documentation on Azure Synapse Analytics notebooks.

NEW QUESTION 76

- (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 80

HOTSPOT - (Topic 2)
You have a Fabric tenant that contains a lakehouse.
You are using a Fabric notebook to save a large DataFrame by using the following code.

```
df.write.partitionBy("year","month","day").mode("overwrite").parquet("Files/SalesOrder")
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.

Statements	Yes	No
The results will form a hierarchy of folders for each partition key.	<input type="radio"/>	<input type="radio"/>
The resulting file partitions can be read in parallel across multiple nodes.	<input type="radio"/>	<input type="radio"/>
The resulting file partitions will use file compression.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? The results will form a hierarchy of folders for each partition key. - Yes

? The resulting file partitions can be read in parallel across multiple nodes. - Yes

? The resulting file partitions will use file compression. - No

Partitioning data by columns such as year, month, and day, as shown in the DataFrame write operation, organizes the output into a directory hierarchy that reflects the partitioning structure. This organization can improve the performance of read operations, as queries that filter by the partitioned columns can scan only the relevant directories. Moreover, partitioning facilitates parallelism because each partition can be processed independently across different nodes in a distributed system like Spark. However, the code snippet provided does not explicitly specify that file compression should be used, so we cannot assume that the output will be compressed without additional context.

References =

? DataFrame write partitionBy

? Apache Spark optimization with partitioning

NEW QUESTION 81

- (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 84

- (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 88

- (Topic 2)

You have a Fabric tenant that contains a lakehouse named lakehouse1. Lakehouse1 contains a table named Table1.

You are creating a new data pipeline.

You plan to copy external data to Table1. The schema of the external data changes regularly.

You need the copy operation to meet the following requirements:

- Replace Table1 with the schema of the external data.

- Replace all the data in Table1 with the rows in the external data.

You add a Copy data activity to the pipeline. What should you do for the Copy data activity?

A. From the Source tab, add additional columns.

B. From the Destination tab, set Table action to Overwrite.

C. From the Settings tab, select Enable staging

D. From the Source tab, select Enable partition discovery

E. From the Source tab, select Recursively

Answer: B

Explanation:

For the Copy data activity, from the Destination tab, setting Table action to Overwrite (B) will ensure that Table1 is replaced with the schema and rows of the external data, meeting the requirements of replacing both the schema and data of the destination table. References = Information about Copy data activity and table actions in Azure Data Factory, which can be applied to data pipelines in Fabric, is available in the Azure Data Factory documentation.

NEW QUESTION 93

HOTSPOT - (Topic 2)

You have a Fabric tenant.

You plan to create a Fabric notebook that will use Spark DataFrames to generate Microsoft Power BI visuals.

You run the following code.

```
from powerbiclient import QuickVisualize, get_dataset_config, Report

PBI_visualize = QuickVisualize(get_dataset_config(df))
PBI_visualize
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
The code embeds an existing Power BI report.	<input type="radio"/>	<input type="radio"/>
The code creates a Power BI report.	<input type="radio"/>	<input type="radio"/>
The code displays a summary of the DataFrame.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? The code embeds an existing Power BI report. - No

? The code creates a Power BI report. - No

? The code displays a summary of the DataFrame. - Yes

The code provided seems to be a snippet from a SQL query or script which is neither creating nor embedding a Power BI report directly. It appears to be setting up a DataFrame for use within a larger context, potentially for visualization in Power BI, but the code itself does not perform the creation or embedding of a report.

Instead, it's likely part of a data processing step that summarizes data.

References =

? Introduction to DataFrames - Spark SQL

? Power BI and Azure Databricks

NEW QUESTION 94

- (Topic 2)

You need to provide Power BI developers with access to the pipeline. The solution must meet the following requirements:

- Ensure that the developers can deploy items to the workspaces for Development and Test.
- Prevent the developers from deploying items to the workspace for Production.
- Follow the principle of least privilege.

Which three levels of access should you assign to the developers? Each correct answer presents part of the solution. NOTE: Each correct answer is worth one point.

- A. Build permission to the production semantic models
- B. Admin access to the deployment pipeline
- C. Viewer access to the Development and Test workspaces
- D. Viewer access to the Production workspace
- E. Contributor access to the Development and Test workspaces
- F. Contributor access to the Production workspace

Answer: BDE

Explanation:

To meet the requirements, developers should have Admin access to the deployment pipeline (B), Contributor access to the Development and Test workspaces (E), and Viewer access to the Production workspace (D). This setup ensures they can perform necessary actions in development and test environments without having the ability to affect production. References = The Power BI documentation on workspace access levels and deployment pipelines provides guidelines on assigning appropriate permissions.

NEW QUESTION 96

- (Topic 2)

You have a Fabric workspace that contains a DirectQuery semantic model. The model queries a data source that has 500 million rows.

You have a Microsoft Power BI report named Report1 that uses the model. Report1 contains visuals on multiple pages.

You need to reduce the query execution time for the visuals on all the pages.

What are two features that you can use? Each correct answer presents a complete solution.

NOTE: Each correct answer is worth one point.

- A. user-defined aggregations
- B. automatic aggregation
- C. query caching
- D. OneLake integration

Answer: AB

Explanation:

User-defined aggregations (A) and query caching (C) are two features that can help reduce query execution time. User-defined aggregations allow precalculation of large datasets, and query caching stores the results of queries temporarily to speed up future queries. References = Microsoft Power BI documentation on performance optimization offers in-depth knowledge on these features.

NEW QUESTION 97

.....

Relate Links

100% Pass Your DP-600 Exam with Examible Prep Materials

<https://www.exambible.com/DP-600-exam/>

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - <https://www.exambible.com/>