



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

- (Topic 1)

You need to recommend a solution to prepare the tenant for the PoC.

Which two actions should you recommend performing from the Fabric Admin portal? Each correct answer presents part of the solution.

NOTE: Each correct answer is worth one point.

- A. Enable the Users can try Microsoft Fabric paid features option for specific security groups.
- B. Enable the Allow Azure Active Directory guest users to access Microsoft Fabric option for specific security groups.
- C. Enable the Users can create Fabric items option and exclude specific security groups.
- D. Enable the Users can try Microsoft Fabric paid features option for the entire organization.
- E. Enable the Users can create Fabric items option for specific security groups.

Answer: AE

Explanation:

The PoC is planned to be completed using a Fabric trial capacity, which implies that users involved in the PoC should be able to try paid features. However, this should be limited to specific security groups involved in the PoC to prevent the entire organization from accessing these features before the trial is proven successful (A). The ability for users to create Fabric items should also be enabled for specific security groups to ensure that only the relevant team members participating in the PoC can create items in the Fabric environment (E).

NEW QUESTION 2

HOTSPOT - (Topic 1)

You need to design a semantic model for the customer satisfaction report.

Which data source authentication method and mode should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Authentication method: Service principal authentication
 Basic authentication
 Service principal authentication
 Single sign-on (SSO) authentication

Mode: DirectQuery
 Direct Lake
 DirectQuery
 Import

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

For the semantic model design required for the customer satisfaction report, the choices for data source authentication method and mode should be made based on security and performance considerations as per the case study provided.

Authentication method: The data should be accessed securely, and given that row-level security (RLS) is required for users executing T-SQL queries, you should use an authentication method that supports RLS. Service principal authentication is suitable for automated and secure access to the data, especially when the access needs to be controlled programmatically and is not tied to a specific user's credentials.

Mode: The report needs to show data as soon as it is updated in the data store, and it should only contain data from the current and previous year. DirectQuery mode allows for real-time reporting without importing data into the model, thus meeting the need for up-to-date data. It also allows for RLS to be implemented and enforced at the data source level, providing the necessary security measures.

Based on these considerations, the selections should be:

? Authentication method: Service principal authentication

? Mode: DirectQuery

NEW QUESTION 3

- (Topic 1)

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

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.

NEW QUESTION 4

HOTSPOT - (Topic 1)

You need 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.

Answer 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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

? 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.

NEW QUESTION 5

HOTSPOT - (Topic 1)

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.



Answer Area

```
CREATE  [dbo].[ProductsWithPricingGroup]
AS
SELECT ProductId,
       ProductName,
       ProductCategory,
       ListPrice,
       
       WHEN ListPrice <= 50 THEN 'low'
       
END AS PricingGroup
FROM dbo.Products
```

Answer 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
    
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

C:\Users\Waqas Shahid\Desktop\Mudassir\Untitled.jpg

? You should use CREATE VIEW to make the pricing group logic available for T- SQL 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.

NEW QUESTION 6

- (Topic 1)

What should you recommend using to ingest the customer data into the data store in the AnalyticsPOC workspace?

- A. a stored procedure
- B. a pipeline that contains a KQL activity
- C. a Spark notebook
- D. a dataflow

Answer: D

Explanation:

For ingesting customer data into the data store in the AnalyticsPOC workspace, a dataflow (D) should be recommended. Dataflows are designed within the Power BI service to ingest, cleanse, transform, and load data into the Power BI environment. They allow for the low-code ingestion and transformation of data as needed by Litware's technical requirements. References = You can learn more about dataflows and their use in Power BI environments in Microsoft's Power BI documentation.

NEW QUESTION 7

- (Topic 2)

You have a Fabric tenant that contains 30 CSV files in OneLake. The files are updated daily.

You create a Microsoft Power BI semantic model named Modell that uses the CSV files as a data source. You configure incremental refresh for Model 1 and publish the model to a Premium capacity in the Fabric tenant.

When you initiate a refresh of Model1, the refresh fails after running out of resources. What is a possible cause of the failure?

- A. Query folding is occurring.
- B. Only refresh complete days is selected.
- C. XMLA Endpoint is set to Read Only.
- D. Query folding is NOT occurring.
- E. The data type of the column used to partition the data has changed.

Answer: E

Explanation:

A possible cause for the failure is that query folding is NOT occurring (D). Query folding helps optimize refresh by pushing down the query logic to the source system, reducing the amount of data processed and transferred, hence conserving resources. References = The Power BI documentation on incremental refresh and query folding provides detailed information on this topic.

NEW QUESTION 8

- (Topic 2)

You have a Fabric tenant that contains a warehouse. The warehouse uses row-level security (RLS). You create a Direct Lake semantic model that uses the Delta tables and RLS of the warehouse. When users interact with a report built from the model, which mode will be used by the DAX queries?

- A. DirectQuery

- B. Dual
- C. Direct Lake
- D. Import

Answer: A

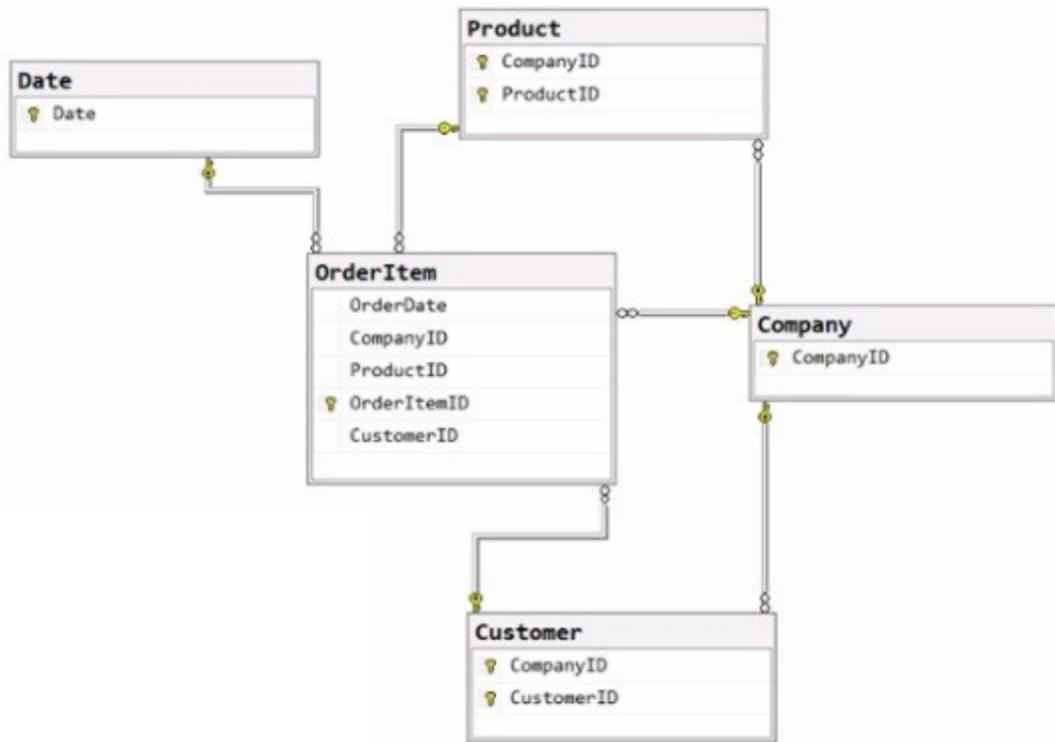
Explanation:

When users interact with a report built from a Direct Lake semantic model that uses row-level security (RLS), the DAX queries will operate in DirectQuery mode (A). This is because the model directly queries the underlying data source without importing data into Power BI. References = The Power BI documentation on DirectQuery provides detailed explanations of how RLS and DAX queries function in this mode.

NEW QUESTION 9

HOTSPOT - (Topic 2)

You have the source data model shown in the following exhibit.



The primary keys of the tables are indicated by a key symbol beside the columns involved in each key. You need to create a dimensional data model that will enable the analysis of order items by date, product, and customer. What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

The relationship between OrderItem and Product must be based on:

- Both the CompanyID and the ProductID columns
- The ProductID column
- Both the CompanyID and the ProductID columns
- A new key that combines the CompanyID and ProductID columns

The Company entity must be:

- Denormalized into the Customer and Product entities
- Omitted
- Denormalized into the Product entity only
- Denormalized into the Customer and Product entities

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

- ? The relationship between OrderItem and Product must be based on: Both the CompanyID and the ProductID columns
- ? The Company entity must be: Denormalized into the Customer and Product entities

In a dimensional model, the relationships are typically based on foreign key constraints between the fact table (OrderItem) and dimension tables (Product, Customer, Date). Since CompanyID is present in both the OrderItem and Product tables, it acts as a foreign key in the relationship. Similarly, ProductID is a foreign key that relates these two tables. To enable analysis by date, product, and customer, the Company entity would need to be denormalized into the Customer and Product entities to ensure that the relevant company information is available within those dimensions for querying and reporting purposes. References =

- ? Dimensional modeling
- ? Star schema design

NEW QUESTION 10

- (Topic 2)

You have a Fabric tenant that contains a warehouse.

Several times a day, the performance of all warehouse queries degrades. You suspect that Fabric is throttling the compute used by the warehouse. What should you use to identify whether throttling is occurring?

- A. the Capacity settings
- B. the Monitoring hub
- C. dynamic management views (DMVs)
- D. the Microsoft Fabric Capacity Metrics app

Answer: B

Explanation:

To identify whether throttling is occurring, you should use the Monitoring hub (B). This provides a centralized place where you can monitor and manage the health, performance, and reliability of your data estate, and see if the compute resources are being throttled. References = The use of the Monitoring hub for performance management and troubleshooting is detailed in the Azure Synapse Analytics documentation.

NEW QUESTION 10

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 14

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

Answer Area

☰ 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.



- 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 16

- (Topic 2)

You have a Fabric workspace named Workspace1 that contains a data flow named Dataflow1. Dataflow1 contains a query that returns the data shown in the following exhibit.



You need to transform the date columns into attribute-value pairs, where columns become rows. You select the VendorID column. Which transformation should you select from the context menu of the VendorID column?

- A. Group by
- B. Unpivot columns
- C. Unpivot other columns
- D. Split column
- E. Remove other columns

Answer: B

Explanation:

The transformation you should select from the context menu of the VendorID column to transform the date columns into attribute-value pairs, where columns become rows, is Unpivot columns (B). This transformation will turn the selected columns into rows with two new columns, one for the attribute (the original column names) and one for the value (the data from the cells). References = Techniques for unpivoting columns are covered in the Power Query documentation, which explains how to use the transformation in data modeling.

NEW QUESTION 17

- (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 22

- (Topic 2)

You have a Fabric tenant that contains a lakehouse. You plan to query sales data files by using the SQL endpoint. The files will be in an Amazon Simple Storage Service (Amazon S3) storage bucket. You need to recommend which file format to use and where to create a shortcut. Which two actions should you include in the recommendation? Each correct answer presents part of the solution. NOTE: Each correct answer is worth one point.

- A. Create a shortcut in the Files section.
- B. Use the Parquet format
- C. Use the CSV format.
- D. Create a shortcut in the Tables section.
- E. Use the delta format.

Answer: BD

Explanation:

You should use the Parquet format (B) for the sales data files because it is optimized for performance with large datasets in analytical processing and create a shortcut in the Tables section (D) to facilitate SQL queries through the lakehouse's SQL endpoint. References = The best practices for working with file formats and shortcuts in a lakehouse environment are covered in the lakehouse and SQL endpoint documentation provided by the cloud data platform services.

NEW QUESTION 24

- (Topic 2)

You have a Fabric tenant that contains a semantic model named Model1. Model1 uses Import mode. Model1 contains a table named Orders. Orders has 100 million rows and the following fields.

Name	Data type	Description
OrderId	Integer	Column imported from the source
OrderDateTime	Date/time	Column imported from the source
Quantity	Integer	Column imported from the source
Price	Decimal	Column imported from the source
TotalSalesAmount	Decimal	Calculated column that multiplies Quantity and Price
TotalQuantity	Integer	Measure

You need to reduce the memory used by Model1 and the time it takes to refresh the model. Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct answer is worth one point.

- A. Split OrderDateTime into separate date and time columns.
- B. Replace TotalQuantity with a calculated column.
- C. Convert Quantity into the Text data type.
- D. Replace TotalSalesAmount with a measure.

Answer: AD

Explanation:

To reduce memory usage and refresh time, splitting the OrderDateTime into separate date and time columns (A) can help optimize the model because date/time data types can be more memory-intensive than separate date and time columns. Moreover, replacing TotalSalesAmount with a measure (D) instead of a calculated column ensures that the calculation is performed at query time, which can reduce the size of the model as the value is not stored but calculated on the fly. References = The best practices for optimizing Power BI models are detailed in the Power BI documentation, which recommends using measures for calculations that don't need to be stored and adjusting data types to improve performance.

NEW QUESTION 28

- (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

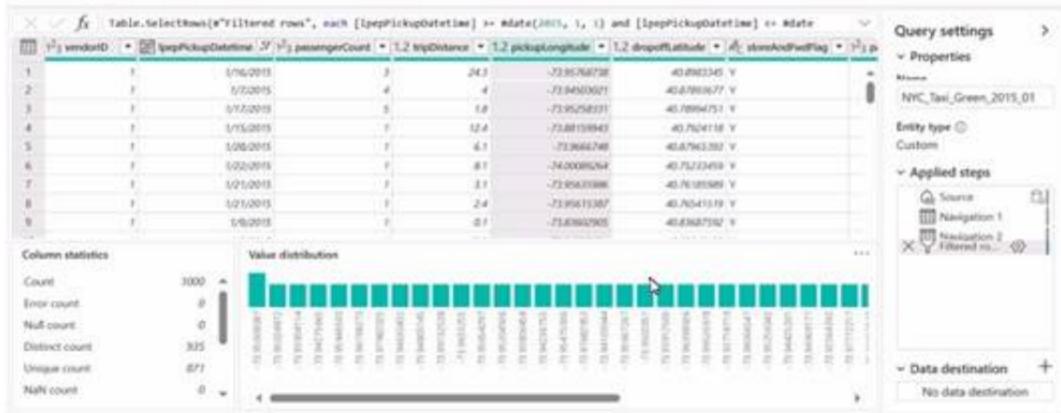
Explanation:

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 29

- (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 33

- (Topic 2)

You have a Microsoft Fabric tenant that contains a dataflow. You are exploring a new semantic model. From Power Query, you need to view column information as shown in the following exhibit.



Which three Data view options should you select? Each correct answer presents part of the solution. NOTE: Each correct answer is worth one point.

- A. Enable column profile
- B. Show column quality details
- C. Show column profile in details pane
- D. Enable details pane
- E. Show column value distribution

Answer: ABE

Explanation:

To view column information like the one shown in the exhibit in Power Query, you need to select the options that enable profiling and display quality and distribution details. These are: A. Enable column profile - This option turns on profiling for each column, showing statistics such as distinct and unique values. B. Show column quality details - It displays the column quality bar on top of each column showing the percentage of valid, error, and empty values. E. Show column value distribution - It enables the histogram display of value distribution for each column, which visualizes how often each value occurs. References: These features and their descriptions are typically found in the Power Query documentation, under the section for data profiling and quality features.

NEW QUESTION 38

- (Topic 2)

You have a Fabric tenant that contains a workspace named Workspace^ Workspacel 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 40

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.



- 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 41

- (Topic 2)

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:

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 43

- (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 47

- (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 50

- (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 54

DRAG DROP - (Topic 2)

You are implementing a medallion architecture in a single Fabric workspace.

You have a lakehouse that contains the Bronze and Silver layers and a warehouse that contains the Gold layer.

You create the items required to populate the layers as shown in the following table.

Layer	Data integration tool
Bronze	Pipelines with Copy activities
Silver	Dataflows
Gold	Stored procedures

You need to ensure that the layers are populated daily in sequential order such that Silver is populated only after Bronze is complete, and Gold is populated only after Silver is complete. The solution must minimize development effort and complexity.

What should you use to execute each set of items? To answer, drag the appropriate options to the correct items. Each option 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.

Execution Methods

- A pipeline Copy activity
- A pipeline Dataflow activity
- A pipeline Stored procedure activity
- A schedule
- A Spark job definition
- An Invoke pipeline activity

Answer Area

Orchestration pipeline:

Bronze layer:

Silver layer:

Gold layer:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

To execute each set of items in sequential order with minimized development effort and complexity, you should use the following options:

? Orchestration pipeline: Use a pipeline with an Invoke pipeline activity. This allows for orchestrating and scheduling the execution of other pipelines, ensuring they run in the correct sequence.

? Bronze layer: Implement a pipeline Copy activity. This aligns with the table indicating that the Bronze layer uses pipelines with Copy activities for data integration.

? Silver layer: Implement a pipeline Dataflow activity. The table specifies that Dataflows are used for the Silver layer.

? Gold layer: Implement a pipeline Stored procedure activity. Stored procedures are specified for the Gold layer according to the table.

NEW QUESTION 55

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.

Answer Area

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 58

- (Topic 2)

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. Tabl
- B. MaxN
- C. Table.Max
- D. Table.Range
- E. Table.Profile

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.

NEW QUESTION 60

- (Topic 2)

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

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.

NEW QUESTION 65

- (Topic 2)

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

Name	Data type	Nullable
ProductID	Integer	No
ProductName	Varchar(30)	No
SalesDate	Datetime2(6)	No
WholesalePrice	Decimal(18, 2)	Yes
Amount	Decimal(18, 2)	Yes

You need to write a T-SQL query that will return data for the year 2023 that displays ProductID and ProductName arxl has a summarized Amount that is higher than 10,000. Which query should you use?

A)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING SUM(Amount) > 10000
```

B)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
GROUP BY ProductID, ProductName
HAVING DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000
```

C)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023' AND SUM(Amount) > 10000
```

D)

```
SELECT ProductID, ProductName, SUM(Amount) AS TotalAmount
FROM Staging.Sales
WHERE DATEPART(YEAR,SaleDate) = '2023'
GROUP BY ProductID, ProductName
HAVING TotalAmount > 10000
```

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

Answer: B

Explanation:

The correct query to use in order to return data for the year 2023 that displays ProductID, ProductName, and has a summarized Amount greater than 10,000 is Option B. The reason is that it uses the GROUP BY clause to organize the data by ProductID and ProductName and then filters the result using the HAVING clause to only include groups where the sum of Amount is greater than 10,000. Additionally, the DATEPART(YEAR, SaleDate) = '2023' part of the HAVING clause ensures that only records from the year 2023 are included. References = For more information, please visit the official documentation on T-SQL queries and the GROUP BY clause at T-SQL GROUP BY.

NEW QUESTION 69

- (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 72

- (Topic 2)

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.

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.

NEW QUESTION 76

.....

Relate Links

100% Pass Your DP-600 Exam with ExamBible 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/>