Microsoft

Exam Questions DP-700

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

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

```
SELECT ProductID, ProductNumber, ProductName, ModelName, SubCategoryName, CategoryName

FROM ContosoLake.Products p
```

[dropdown]
```
FULL JOIN
INNER JOIN
LEFT ANTI JOIN
LEFT OUTER JOIN
OUTER JOIN
```
`ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID`

[dropdown]
```
FULL JOIN
INNER JOIN
LEFT ANTI JOIN
LEFT OUTER JOIN
OUTER JOIN
```
`ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID`

WHERE [dropdown]
```
CategoryID = 1;
CategoryName is not null;
IsActive = 1;
IsActive is not null;
ProductNumber is not null;
SubCategoryID = 1;
SubCategoryName is not null;
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Join between Products and ProductSubCategories: Use an INNER JOIN.
The goal is to include only products that are assigned to a subcategory. An INNER JOIN ensures that only matching records (i.e., products with a valid subcategory) are included.
Join between ProductSubCategories and ProductCategories: Use an INNER JOIN.
Similar to the above logic, we want to include only subcategories assigned to a valid product category. An INNER JOIN ensures this condition is met.
WHERE Clause Condition: IsActive = 1
Only active products (where IsActive equals 1) should be included in the gold layer. This filters out inactive products.

NEW QUESTION 2
- (Topic 1)
You need to populate the MAR1 data in the bronze layer.
Which two types of activities should you include in the pipeline? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. ForEach
B. Copy data
C. WebHook
D. Stored procedure

Answer: AB

Explanation:
MAR1 has seven entities, each accessible via a different API endpoint. A ForEach activity is required to iterate over these endpoints to fetch data from each one. It enables dynamic execution of API calls for each entity.
The Copy data activity is the primary mechanism to extract data from REST APIs and load it into the bronze layer in Delta format. It supports native connectors for REST APIs and Delta, minimizing development effort.
You need to schedule the population of the medallion layers to meet the technical requirements.
What should you do?
* A. Schedule a data pipeline that calls other data pipelines.
* B. Schedule a notebook.
* C. Schedule an Apache Spark job.
* D. Schedule multiple data pipelines.

* Answer: A

The technical requirements specify that:

Medallion layers must be fully populated sequentially (bronze silver gold). Each layer must be populated before the next.

If any step fails, the process must notify the data engineers. Data imports should run simultaneously when possible.

Why Use a Data Pipeline That Calls Other Data Pipelines?

A data pipeline provides a modular and reusable approach to orchestrating the sequential population of medallion layers.

By calling other pipelines, each pipeline can focus on populating a specific layer (bronze, silver, or gold), simplifying development and maintenance.

A parent pipeline can handle:

- Sequential execution of child pipelines.
- Error handling to send email notifications upon failures.
- Parallel execution of tasks where possible (e.g., simultaneous imports into the bronze layer).

NEW QUESTION 3
- (Topic 1)

You need to ensure that usage of the data in the Amazon S3 bucket meets the technical requirements.

What should you do?

A. Create a workspace identity and enable high concurrency for the notebooks.
B. Create a shortcut and ensure that caching is disabled for the workspace.
C. Create a workspace identity and use the identity in a data pipeline.
D. Create a shortcut and ensure that caching is enabled for the workspace.

Answer: B

Explanation:

To ensure that the usage of the data in the Amazon S3 bucket meets the technical requirements, we must address two key points:

Minimize egress costs associated with cross-cloud data access: Using a shortcut ensures that Fabric does not replicate the data from the S3 bucket into the lakehouse but rather provides direct access to the data in its original location. This minimizes cross-cloud data transfer and avoids additional egress costs.

Prevent saving a copy of the raw data in the lakehouses: Disabling caching ensures that the raw data is not copied or persisted in the Fabric workspace. The data is accessed on- demand directly from the Amazon S3 bucket.

NEW QUESTION 4
- (Topic 2)

You need to resolve the sales data issue. The solution must minimize the amount of data transferred.

What should you do?

A. Spilt the dataflow into two dataflows.
B. Configure scheduled refresh for the dataflow.
C. Configure incremental refresh for the dataflo
D. Set Store rows from the past to 1 Month.
E. Configure incremental refresh for the dataflo
F. Set Refresh rows from the past to 1 Year.
G. Configure incremental refresh for the dataflo
H. Set Refresh rows from the past to 1 Month.

Answer: E

Explanation:

The sales data issue can be resolved by configuring incremental refresh for the dataflow. Incremental refresh allows for only the new or changed data to be processed, minimizing the amount of data transferred and improving performance.

The solution specifies that data older than one month never changes, so setting the refresh period to 1 Month is appropriate. This ensures that only the most recent month of data will be refreshed, reducing unnecessary data transfers.

NEW QUESTION 5
HOTSPOT - (Topic 2)

You need to troubleshoot the ad-hoc query issue.

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

NOTE: Each correct selection is worth one point.

```
SELECT last_run_start_time, last_run_command

FROM    [ ▼ ]
        queryinsights.exec_requests_history
        queryinsights.exec_sessions_history
        queryinsights.frequently_run_queries
        queryinsights.long_running_queries

WHERE last_run_total_elapsed_time_ms > 7200000

AND     [ ▼ ]
        max_run_total_elapsed_time_ms > 7200000
        median_total_elapsed_time_ms > 7200000
        number_of_canceled_runs > 1
        number_of_failed_runs > 1
        number_of_runs > 1
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
SELECT last_run_start_time, last_run_command: These fields will help identify the execution details of the long-running queries.
FROM queryinsights.long_running_queries: The correct solution is to check the long- running queries using the queryinsights.long_running_queries view, which provides insights into queries that take longer than expected to execute.
WHERE last_run_total_elapsed_time_ms > 7200000: This condition filters queries that took more than 2 hours to complete (7200000 milliseconds), which is relevant to the issue described.
AND number_of_failed_runs > 1: This condition is key for identifying queries that have failed more than once, helping to isolate the problematic queries that cause failures and need attention.

NEW QUESTION 6
HOTSPOT - (Topic 3)
You are building a data orchestration pattern by using a Fabric data pipeline named Dynamic Data Copy as shown in the exhibit. (Click the Exhibit tab.)
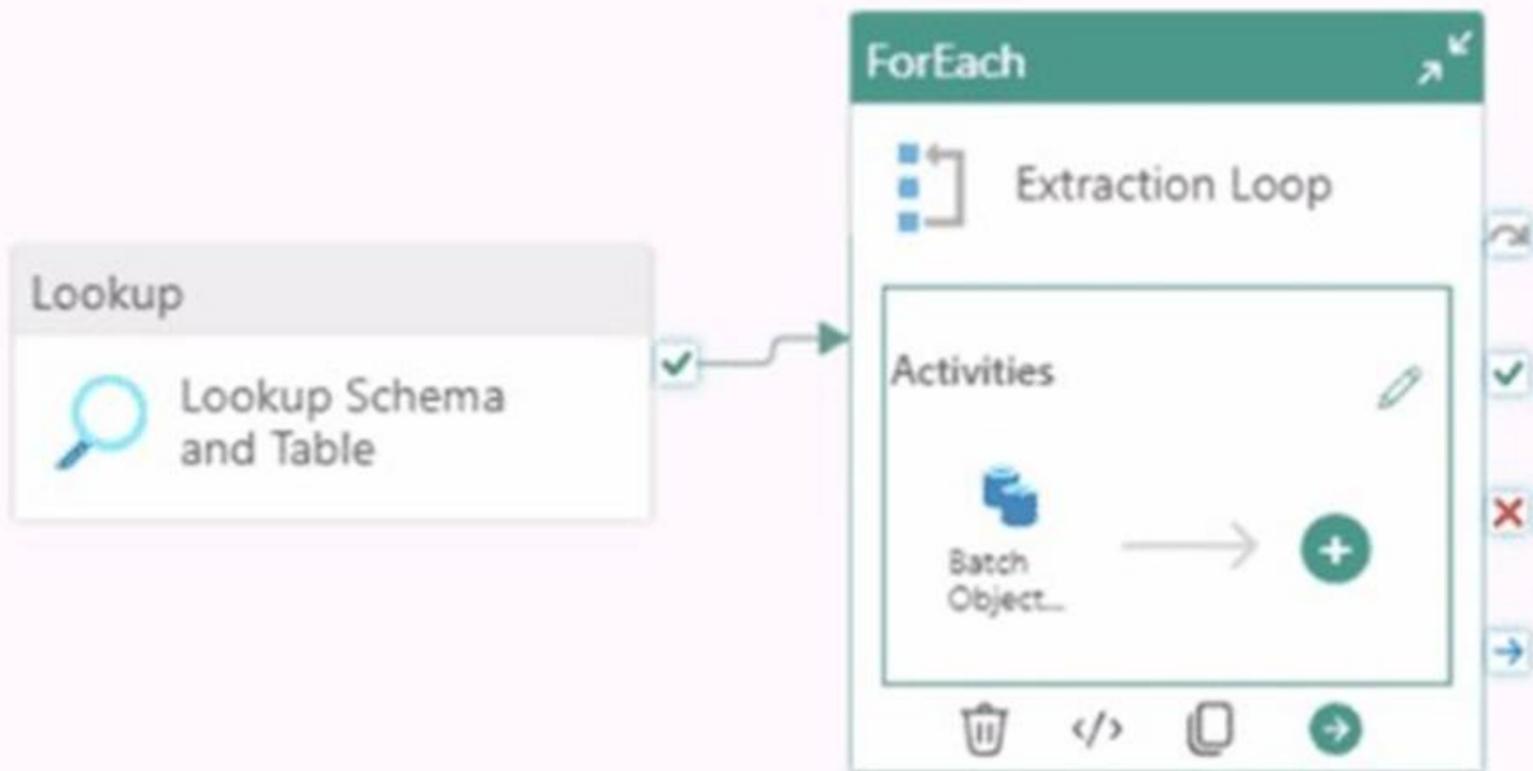
Dynamic Data Copy does NOT use parametrization.
You need to configure the ForEach activity to receive the list of tables to be copied. How should you complete the pipeline expression? To answer, select the appropriate
options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

@activity(' [Lookup Schema and Table ▼] '). [output.value ▼]

| Batch Object Copy |
| Dynamic Data Copy |
| Extraction Loop |
| **Lookup Schema and Table** |

| output |
| output.count |
| output.pipelineReturnValue |
| **output.value** |

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

@activity('

Lookup Schema and Table ▼
Batch Object Copy
Dynamic Data Copy
Extraction Loop
Lookup Schema and Table

').

output.value ▼
output
output.count
output.pipelineReturnValue
output.value

NEW QUESTION 7
- (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1.
While monitoring Warehouse1, you discover that query performance has degraded during the last 60 minutes.
You need to isolate all the queries that were run during the last 60 minutes. The results must include the username of the users that submitted the queries and the query statements. What should you use?

A. the Microsoft Fabric Capacity Metrics app
B. views from the queryinsights schema
C. Query activity
D. the sys.dm_exec_requests dynamic management view

Answer: B

NEW QUESTION 8
- (Topic 3)
You have a Fabric warehouse named DW1 that loads data by using a data pipeline named Pipeline1. Pipeline1 uses a Copy data activity with a dynamic SQL source. Pipeline1 is scheduled to run every 15 minutes.
You discover that Pipeline1 keeps failing.
You need to identify which SQL query was executed when the pipeline failed. What should you do?

A. From Monitoring hub, select the latest failed run of Pipeline1, and then view the output JSON.
B. From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.
C. From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemReadFailed.
D. From Real-time hub, select Fabric events, and then review the details of Microsof
E. Fabric.ItemUpdateFailed.

Answer: B

Explanation:
The input JSON contains the configuration details and parameters passed to the Copy data activity during execution, including the dynamically generated SQL query.
Viewing the input JSON for the failed pipeline run provides direct insight into what query was executed at the time of failure.

NEW QUESTION 9
- (Topic 3)
You have a Fabric workspace that contains a lakehouse named Lakehouse1. Data is ingested into Lakehouse1 as one flat table. The table contains the following columns.

Name	Description
TransactionID	Contains a unique ID for each transaction
Date	Contains the date of a transaction
ProductID	Contains a unique ID for each product
ProductColor	Contains a descriptive attribute that describes the color of each product
ProductName	Contains a unique name for each product
SalesAmount	Contains the sales amount of a transaction

You plan to load the data into a dimensional model and implement a star schema. From the original flat table, you create two tables named FactSales and DimProduct. You will track changes in DimProduct.
You need to prepare the data.
Which three columns should you include in the DimProduct table? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Date
B. ProductName
C. ProductColor
D. TransactionID

E. SalesAmount
F. ProductID

Answer: BCF

Explanation:
In a star schema, the DimProduct table serves as a dimension table that contains descriptive attributes about products. It will provide context for the FactSales table, which contains transactional data. The following columns should be included in the DimProduct table:
? ProductName: The ProductName is an important descriptive attribute of the product, which is needed for analysis and reporting in a dimensional model.
? ProductColor: ProductColor is another descriptive attribute of the product. In a star schema, it makes sense to include attributes like color in the dimension table to help categorize products in the analysis.
? ProductID: ProductID is the primary key for the DimProduct table, which will be used to join the FactSales table to the product dimension. It's essential for uniquely identifying each product in the model.

NEW QUESTION 10
- (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1. Data is loaded daily into Warehouse1 by using data pipelines and stored procedures.
You discover that the daily data load takes longer than expected.
You need to monitor Warehouse1 to identify the names of users that are actively running queries.
Which view should you use?

A. sys.dm_exec_connections
B. sys.dm_exec_requests
C. queryinsights.long_running_queries
D. queryinsights.frequently_run_queries
E. sys.dm_exec_sessions

Answer: E

Explanation:
sys.dm_exec_sessions provides real-time information about all active sessions, including the user, session ID, and status of the session. You can filter on session status to see users actively running queries.

NEW QUESTION 10
- (Topic 3)
You have a Fabric warehouse named DW1. DW1 contains a table that stores sales data and is used by multiple sales representatives.
You plan to implement row-level security (RLS).
You need to ensure that the sales representatives can see only their respective data. Which warehouse object do you require to implement RLS?

A. ISTORED PROCEDURE
B. CONSTRAINT
C. SCHEMA
D. FUNCTION

Answer: D

Explanation:
To implement Row-Level Security (RLS) in a Fabric warehouse, you need to use a function that defines the security logic for filtering the rows of data based on the user's identity or role. This function can be used in conjunction with a security policy to control access to specific rows in a table.
In the case of sales representatives, the function would define the filtering criteria (e.g., based on a column such as SalesRepID or SalesRepName), ensuring that each representative can only see their respective data.

NEW QUESTION 14
- (Topic 3)
You have a Fabric workspace named Workspacel that contains the following items:
• A Microsoft Power BI report named Reportl
• A Power BI dashboard named Dashboardl
• A semantic model named Modell
• A lakehouse name Lakehouse1
Your company requires that specific governance processes be implemented for the items. Which items can you endorse in Fabric?

A. Lakehouse1, Modell, and Dashboard1 only
B. Lakehouse1, Modell, Report1 and Dashboard1
C. Report1 and Dashboard1 only
D. Model1, Report1, and Dashboard1 only
E. Lakehouse1, Model1, and Report1 only

Answer: B

NEW QUESTION 18
HOTSPOT - (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse! contains a table named Customer. Customer contains the following data.

CustomerID	FirstName	LastName	Phone	CreditCard
1	John	Doe	555-123-4567	1234567812345670
2	Jane	Smith	555-987-6543	8765432187654320
3	Michael	Johnson	555-555-5555	1234987654321230
4	Emily	Davis	555-222-3333	4321123456789870
5	David	Brown	555-444-5555	5678123498761230

You have an internal Microsoft Entra user named User1 that has an email address of user1@contoso.com.
You need to provide User1 with access to the Customer table. The solution must prevent User1 from accessing the CreditCard column.
How should you complete the statement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

GRANT [SELECT ▼]
ALTER
EXECUTE
READ
SELECT
VIEW

Customers(CustomerID, FirstName, LastName, Phone)

TO [[user1@contoso.com] ▼]
User1
[User1]
[user1@contoso.com]

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

GRANT SELECT ▼
ALTER
EXECUTE
READ
SELECT
VIEW

Customers(CustomerID, FirstName, LastName, Phone)

TO [user1@contoso.com] ▼
User1
[User1]
[user1@contoso.com]

NEW QUESTION 22
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse1.
You plan to deploy Warehouse1 to a new workspace named Workspace2.
As part of the deployment process, you need to verify whether Warehouse1 contains invalid references. The solution must minimize development effort.
What should you use?

A. a database project
B. a deployment pipeline
C. a Python script
D. a T-SQL script

Answer: C

Explanation:
A deployment pipeline in Fabric allows you to deploy assets like warehouses, datasets, and reports between different workspaces (such as from Workspace1 to Workspace2). One of
the key features of a deployment pipeline is the ability to check for invalid references before deployment. This can help identify issues with assets, such as broken links or dependencies, ensuring the deployment is successful without introducing errors. This is the most efficient way to verify references and manage the deployment with minimal development effort.

NEW QUESTION 26
- (Topic 3)
You have a Fabric workspace that contains a Real-Time Intelligence solution and an eventhouse.
Users report that from OneLake file explorer, they cannot see the data from the eventhouse.
You enable OneLake availability for the eventhouse. What will be copied to OneLake?

A. only data added to new databases that are added to the eventhouse
B. only the existing data in the eventhouse
C. no data
D. both new data and existing data in the eventhouse
E. only new data added to the eventhouse

Answer: D

Explanation:
When you enable OneLake availability for an eventhouse, both new and existing data in the eventhouse will be copied to OneLake. This feature ensures that data, whether newly ingested or already present, becomes available for access through OneLake, making it easier for users to interact with and explore the data directly from OneLake file explorer.

NEW QUESTION 31
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a notebook named Notebook1.
In Workspace1, you create a new notebook named Notebook2.
You need to ensure that you can attach Notebook2 to the same Apache Spark session as Notebook1.
What should you do?

A. Enable high concurrency for notebooks.
B. Enable dynamic allocation for the Spark pool.
C. Change the runtime version.
D. Increase the number of executors.

Answer: A

Explanation:

To ensure that Notebook2 can attach to the same Apache Spark session as Notebook1, you need to enable high concurrency for notebooks. High concurrency allows multiple notebooks to share a Spark session, enabling them to run within the same Spark context and thus share resources like cached data, session state, and compute capabilities. This is particularly useful when you need notebooks to run in sequence or together while leveraging shared resources.

NEW QUESTION 34
- (Topic 3)
You are implementing a medallion architecture in a Fabric lakehouse.
You plan to create a dimension table that will contain the following columns:
• ID
• CustomerCode
• CustomerName
• CustomerAddress
• CustomerLocation
• ValidFrom
• ValidTo
You need to ensure that the table supports the analysis of historical sales data by customer location at the time of each sale Which type of slowly changing dimension (SCD) should you use?

A. Type 2
B. Type 0
C. Type 1
D. Type 3

Answer: A

NEW QUESTION 38
HOTSPOT - (Topic 3)
You have a Fabric workspace that contains two lakehouses named Lakehouse1 and Lakehouse2. Lakehouse1 contains staging data in a Delta table named Orderlines. Lakehouse2 contains a Type 2 slowly changing dimension (SCD) dimension table named Dim_Customer.
You need to build a query that will combine data from Orderlines and Dim_Customer to create a new fact table named Fact_Orders. The new table must meet the following requirements:
Enable the analysis of customer orders based on historical attributes. Enable the analysis of customer orders based on the current attributes.
How should you complete the statement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

```
SELECT
  OrderLineID order_line_id
  ,OrderDate order_date
  ,c.customer_key
  ,c.customer_id
  ,Quantity order_quantity
  ,unitprice unit_price
  ,TaxRate tax_rate
FROM
  Lakehouse1.orderlines o
INNER JOIN
  Lakehouse2.dim_customer c
    ON o.customerid = c.customer_id
```

AND [▼]

```
c.is_current = 1
o.OrderDate >= valid_to_datetime
o.OrderDate >= valid_from_catetime
```

AND [▼]

```
c.is_current = 1
o.OrderDate < valid_to_datetime
o.OrderDate <= valid_from_catetime
```

NEW QUESTION 40
- (Topic 3)
You have an Azure SQL database named DB1.
In a Fabric workspace, you deploy an eventstream named EventStreamDBI to stream record changes from DB1 into a lakehouse.
You discover that events are NOT being propagated to EventStreamDBI.
You need to ensure that the events are propagated to EventStreamDBI. What should you do?

A. Create a read-only replica of DB1.
B. Create an Azure Stream Analytics job.
C. Enable Extended Events for DB1.
D. Enable change data capture (CDC) for DB1.

Answer: D

NEW QUESTION 41
- (Topic 3)
You need to develop an orchestration solution in fabric that will load each item one after the other. The solution must be scheduled to run every 15 minutes. Which type of item should you use?

A. warehouse
B. data pipeline
C. Dataflow Gen2 dataflow
D. notebook

Answer: B

NEW QUESTION 42

DRAG DROP - (Topic 3)
You are building a data loading pattern by using a Fabric data pipeline. The source is an Azure SQL database that contains 25 tables. The destination is a lakehouse.
In a warehouse, you create a control table named Control.Object as shown in the exhibit. (Click the Exhibit tab.)
You need to build a data pipeline that will support the dynamic ingestion of the tables listed in the control table by using a single execution.
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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 44
- (Topic 3)
You have a Fabric workspace that contains a data pipeline named Pipeline! as shown in the exhibit.

What will occur the next time Pipelinel tuns?

A. Both activities will run simultaneously.
B. Both activities will be skipped.
C. Execute procedurel will run and Copy_kdi will be skipped.
D. Copy.kdi will run and Execute procedurel will be skipped.
E. Execute procedure1 will run first, and then Copy_kdi will run.
F. Copy.kdi will run first, and then Execute procedurel will run.

Answer: A

NEW QUESTION 47
- (Topic 3)
Your company has a sales department that uses two Fabric workspaces named Workspace1 and Workspace2.
The company decides to implement a domain strategy to organize the workspaces. You need to ensure that a user can perform the following tasks:
Create a new domain for the sales department.
Create two subdomains: one for the east region and one for the west region. Assign Workspace1 to the east region subdomain.
Assign Workspace2 to the west region subdomain. The solution must follow the principle of least privilege. Which role should you assign to the user?

A. workspace Admin
B. domain admin
C. domain contributor
D. Fabric admin

Answer: B

Explanation:
To implement a domain strategy and manage subdomains within Fabric, the domain admin role is the appropriate role for the user. A domain admin has the permissions necessary to:
? Create a new domain (for the sales department).
? Create subdomains (for the east and west regions).
? Assign workspaces (such as Workspace1 and Workspace2) to the appropriate subdomains.
The domain admin role allows for managing the structure and organization of workspaces in the context of domains and subdomains while maintaining the principle of least privilege by limiting the user's access to managing the domain structure specifically.

NEW QUESTION 52
- (Topic 3)
You have a Fabric workspace named Workspace1 that contains a warehouse named DW1 and a data pipeline named Pipeline1.
You plan to add a user named User3 to Workspace1.
You need to ensure that User3 can perform the following actions: View all the items in Workspace1.
Update the tables in DW1.
The solution must follow the principle of least privilege.
You already assigned the appropriate object-level permissions to DW1. Which workspace role should you assign to User3?

A. Admin
B. Member
C. Viewer
D. Contributor

Answer: D

Explanation:
To ensure User3 can view all items in Workspace1 and update the tables in DW1, the most appropriate workspace role to assign is the Contributor role. This role allows User3 to: View all items in Workspace1: The Contributor role provides the ability to view all objects within the workspace, such as data pipelines, warehouses, and other resources.
Update the tables in DW1: The Contributor role allows User3 to modify or update resources within the workspace, including the tables in DW1, assuming that appropriate object-level permissions are set for the warehouse.
This role adheres to the principle of least privilege, as it provides the necessary permissions without granting broader administrative rights.

NEW QUESTION 54
- (Topic 3)
You have a Fabric workspace that contains a semantic model named Modell. You need to monitor the refresh history of Model 1 and visualize the refresh history in a chart. What should you use?

A. the refresh history from the settings of Model1.
B. a notebook
C. a Dataflow Gen2 dataflow
D. a data pipeline

Answer: A

NEW QUESTION 55
HOTSPOT - (Topic 3)
You plan to process the following three datasets by using Fabric:
• Dataset1: This dataset will be added to Fabric and will have a unique primary key between the source and the destination. The unique primary key will be an integer and will start from 1 and have an increment of 1.
• Dataset2: This dataset contains semi-structured data that uses bulk data transfer. The dataset must be handled in one process between the source and the destination. The data transformation process will include the use of custom visuals to understand and work with the dataset in development mode.
• Dataset3. This dataset is in a takehouse. The data will be bulk loaded. The data transformation process will include row-based windowing functions during the loading process.
You need to identify which type of item to use for the datasets. The solution must minimize development effort and use built-in functionality, when possible. What

should you identify for each dataset? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

Dataset1: [A T-SQL statement ▼]
A Dataflow Gen2 dataflow
A notebook
A T-SQL statement

Dataset2: [A notebook ▼]
A Dataflow Gen2 dataflow
A notebook
A T-SQL statement

Dataset3: [A KQL queryset ▼]
A Dataflow Gen2 dataflow
A KQL queryset
A T-SQL statement

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Dataset1: [A T-SQL statement ▼]
A Dataflow Gen2 dataflow
A notebook
A T-SQL statement

Dataset2: [A notebook ▼]
A Dataflow Gen2 dataflow
A notebook
A T-SQL statement

Dataset3: [A KQL queryset ▼]
A Dataflow Gen2 dataflow
A KQL queryset
A T-SQL statement

NEW QUESTION 56
- (Topic 3)
You have a Fabric workspace that contains a lakehouse named Lakehouse1.
In an external data source, you have data files that are 500 GB each. A new file is added every day.
You need to ingest the data into Lakehouse1 without applying any transformations. The solution must meet the following requirements
Trigger the process when a new file is added. Provide the highest throughput.
Which type of item should you use to ingest the data?

A. Data pipeline
B. Environment
C. KQL queryset
D. Dataflow Gen2

Answer: A

Explanation:
To efficiently ingest large data files (500 GB each) into Lakehouse1 with high throughput and trigger the process when a new file is added, a Data pipeline is the most suitable solution. Data pipelines in Fabric are ideal for orchestrating data movement and can be configured to automatically trigger based on file arrivals or other events. This solution meets both requirements: ingesting the data without transformations (since you just need to copy the data) and triggering the process when new files are added.

NEW QUESTION 61
- (Topic 3)
You have a Fabric workspace named Workspace1. You plan to integrate Workspace1 with Azure DevOps.
You will use a Fabric deployment pipeline named deployPipeline1 to deploy items from Workspace1 to higher environment workspaces as part of a medallion architecture. You will run deployPipeline1 by using an API call from an Azure DevOps pipeline.
You need to configure API authentication between Azure DevOps and Fabric. Which type of authentication should you use?

A. service principal
B. Microsoft Entra username and password
C. managed private endpoint
D. workspace identity

Answer: A

Explanation:
When integrating Azure DevOps with Fabric (Workspace1), using a service principal is the recommended authentication method. A service principal provides a way for applications (such as an Azure DevOps pipeline) to authenticate and interact with resources securely. It allows Azure DevOps to authenticate API calls to Fabric without requiring direct user credentials. This method is ideal for automating tasks such as deploying items through a Fabric deployment pipeline.

NEW QUESTION 62
DRAG DROP - (Topic 3)
You have two Fabric notebooks named Load_Salesperson and Load_Orders that read data from Parquet files in a lakehouse. Load_Salesperson writes to a Delta table named dim_salesperson. Load.Orders writes to a Delta table named fact_orders and is dependent on the successful execution of Load_Salesperson.
You need to implement a pattern to dynamically execute Load_Salesperson and Load_Orders in the appropriate order by using a notebook.
How should you complete the code? To answer, drag the appropriate values 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	Answer Area
activities	name : Load_Salesperson ,
broadcast	"path": "Load_Salesperson",
dependencies	"timeoutPerCellInSeconds": 300,
execute	},
notebooks	{
runMultiple	"name": "Load_Orders",
	"path": "Load_Orders",
	"timeoutPerCellInSeconds": 600,
	" [_____] ": ["Load_Salesperson"]
	}
	],
	"timeoutInSeconds": 43200
	}
	mssparkutils.notebook. [_____] (DAG)

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Values

- ⠿ activities
- ⠿ broadcast
- ⠿ dependencies
- ⠿ execute
- ⠿ notebooks
- ⠿ runMultiple

Answer Area

```
        name : Load_Salesperson ,
        "path": "Load_Salesperson",
        "timeoutPerCellInSeconds": 300,
    },
    {
        "name": "Load_Orders",
        "path": "Load_Orders",
        "timeoutPerCellInSeconds": 600,
        " ⠿ dependencies    ": ["Load_Salesperson"]
    }
    ],
        "timeoutInSeconds": 43200
    }
mssparkutils.notebook. ⠿ runMultiple    (DAG)
```

NEW QUESTION 66
- (Topic 3)
You are developing a data pipeline named Pipeline1.
You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse. Which option from the Settings tab of the Copy data activity must you configure?

A. Enable logging
B. Fault tolerance
C. Enable staging
D. Degree of copy parallelism

Answer: C

NEW QUESTION 69
- (Topic 3)
You have a Fabric notebook named Notebook1 that has been executing successfully for the last week.
During the last run, Notebook1executed nine jobs. You need to view the jobs in a timeline chart. What should you use?

A. Real-Time hub
B. Monitoring hub
C. the job history from the application run
D. Spark History Server
E. the run series from the details of the application run

Answer: E

Explanation:
The run series from the details of the application run is the most detailed and relevant feature for visualizing job execution in a timeline format, making it the correct choice for this scenario. It provides an intuitive way to analyze job execution patterns and improve the efficiency of the notebook.

NEW QUESTION 70
- (Topic 3)
You have a Fabric workspace that contains a warehouse named DW1. DW1 is loaded by using a notebook named Notebook1.
You need to identify which version of Delta was used when Notebook1 was executed. What should you use?

A. Real-Time hub
B. OneLake data hub
C. the Admin monitoring workspace
D. Fabric Monitor
E. the Microsoft Fabric Capacity Metrics app

Answer: C

Explanation:
To identify the version of Delta used when Notebook1 was executed, you should use the Admin monitoring workspace. The Admin monitoring workspace allows you to track and
monitor detailed information about the execution of notebooks and jobs, including the underlying versions of Delta or other technologies used. It provides insights into execution details, including versions and configurations used during job runs, making it the most appropriate choice for identifying the Delta version used during the execution of Notebook1.

NEW QUESTION 71
- (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1.

You have an on-premises Microsoft SQL Server database named Database1 that is accessed by using an on-premises data gateway.
You need to copy data from Database1 to Warehouse1. Which item should you use?

A. an Apache Spark job definition
B. a data pipeline
C. a Dataflow Gen1 dataflow
D. an eventstream

Answer: B

Explanation:
To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Fabric, a data pipeline is the most appropriate tool. A data pipeline in Fabric is designed to move data between various data sources and destinations, including on-premises databases like SQL Server, and cloud-based storage like Fabric warehouses. The data pipeline can handle the connection through an on-premises data gateway, which is required to access on-premises data. This solution facilitates the orchestration of data movement and transformations if needed.

NEW QUESTION 72
DRAG DROP - (Topic 3)
You have a Fabric workspace that contains a warehouse named Warehouse1.
In Warehouse1, you create a table named DimCustomer by running the following statement.

```
CREATE TABLE dbo.DimCustomer (

    CustomerKey VARCHAR(255) NOT NULL,

    Name VARCHAR(255) NOT NULL,

    Email VARCHAR(255) NOT NULL

);
```

You need to set the Customerkey column as a primary key of the DimCustomer table. Which three code segments should you run in sequence? To answer, move the appropriate
code segments from the list of code segments to the answer area and arrange them in the correct order.

Code Segments

```
DROP CONSTRAINT PK_DimCustomer

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED
(CustomerKey)

NOT ENFORCED

ALTER TABLE dbo.DimCustomer

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
(CustomerKey)

ENFORCED
```

Answer Area

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Code Segments

DROP CONSTRAINT PK_DimCustomer

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED
(CustomerKey)

NOT ENFORCED

ALTER TABLE dbo.DimCustomer

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
(CustomerKey)

ENFORCED

Answer Area

ALTER TABLE dbo.DimCustomer

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
(CustomerKey)

ENFORCED

NEW QUESTION 76

- (Topic 3)

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some
question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have a Fabric eventstream that loads data into a table named Bike_Location in a KQL database. The table contains the following columns:

BikepointID Street Neighbourhood No_Bikes No_Empty_Docks Timestamp

You need to apply transformation and filter logic to prepare the data for consumption. The solution must return data for a neighbourhood named Sands End when No_Bikes is at least 15. The results must be ordered by No_Bikes in ascending order.

Solution: You use the following code segment:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| order by No_Bikes
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?

A. Yes

B. no

Answer: B

Explanation:

This code does not meet the goal because it uses order by, which is not valid in KQL. The correct term in KQL is sort by.
Correct code should look like:

```
bike_location
| filter Neighbourhood == "Sands End" and No_Bikes >= 15
| sort by No_Bikes asc
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

NEW QUESTION 79

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains an eventstream named EventStream1. You discover that an EventStream1 transformation fails.

You need to find the following error information: The error details, including the occurrence time The total number of errors

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

Answer Area

To find the error details:

Data insights
Data preview
Details
Runtime logs

To find the total number of errors:

Data insights
Data preview
Details
Runtime logs

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

To find the error details:

Data insights
Data preview
Details
Runtime logs

To find the total number of errors:

Data insights
Data preview
Details
Runtime logs

NEW QUESTION 84
- (Topic 3)
You are developing a data pipeline named Pipeline1.
You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse.
What should you configure?

A. Degree of copy parallelism

B. Fault tolerance
C. Enable staging
D. Enable logging

Answer: C

Explanation:
When using the Copy data activity in a data pipeline to move data from Snowflake to a Fabric warehouse, the process often involves intermediate staging to handle data efficiently, especially for large datasets or cross-cloud data transfers.
Staging involves temporarily storing data in an intermediate location (e.g., Blob storage or Azure Data Lake) before loading it into the target destination.
For cross-cloud data transfers (e.g., from Snowflake to Fabric), enabling staging ensures data is processed and stored temporarily in an efficient format for transfer.
Staging is especially useful when dealing with large datasets, ensuring the process is optimized and avoids memory limitations.

NEW QUESTION 85
- (Topic 3)
You have a Fabric workspace named Workspace1. Your company acquires GitHub licenses.
You need to configure source control for Workpace1 to use GitHub. The solution must follow the principle of least privilege. Which permissions do you require to ensure that you can commit code to GitHub?

A. Actions (Read and write) and Contents (Read and write)
B. Actions (Read and write) only
C. Contents (Read and write) only
D. Contents (Read) and Commit statuses (Read and write)

Answer: C

NEW QUESTION 90
HOTSPOT - (Topic 3)
You have a Fabric workspace named Workspace1_DEV that contains the following items: 10 reports
Four notebooks Three lakehouses Two data pipelines
Two Dataflow Gen1 dataflows Three Dataflow Gen2 dataflows
Five semantic models that each has a scheduled refresh policy
You create a deployment pipeline named Pipeline1 to move items from Workspace1_DEV to a new workspace named Workspace1_TEST.
You deploy all the items from Workspace1_DEV to Workspace1_TEST.
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
Data from the semantic models will be deployed to the target stage.	○	○
The Dataflow Gen1 dataflows will be deployed to the target stage.	○	○
The scheduled refresh policies will be deployed to the target stage.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	○	○
The Dataflow Gen1 dataflows will be deployed to the target stage.	○	○
The scheduled refresh policies will be deployed to the target stage.	○	○

NEW QUESTION 94
- (Topic 3)
You have a Fabric workspace that contains a semantic model named Model1. You need to dynamically execute and monitor the refresh progress of Model1. What should you use?

A. dynamic management views in Microsoft SQL Server Management Studio
B. Monitoring hub
C. dynamic management views in Azure Data Studio
D. a semantic link in a notebook

Answer: D

Explanation:
Semantic models in Microsoft Fabric are part of Power BI datasets and require refreshes to stay updated with the latest data.
Dynamically executing and monitoring the refresh progress requires a tool or approach that integrates with Fabric's capabilities for semantic models.

NEW QUESTION 97
- (Topic 3)
You have a Google Cloud Storage (GCS) container named storage1 that contains the files shown in the following table.

Name	Size
ProductFile.parquet	8 MB
StoreFile.json	500 MB
TripsFile.csv	99 MB

You have a Fabric workspace named Workspace1 that has the cache for shortcuts enabled. Workspace1 contains a lakehouse named Lakehouse1. Lakehouse1 has the shortcuts shown in the following table.

Name	Source	Last accessed
Products	ProductFile	12 hours ago
Stores	StoreFile	4 hours ago
Trips	TripsFile	48 hours ago

You need to read data from all the shortcuts. Which shortcuts will retrieve data from the cache?

A. Stores only
B. Products only
C. Stores and Products only
D. Products, Stores, and Trips

E. Trips only
F. Products and Trips only

Answer: C

Explanation:
When reading data from shortcuts in Fabric (in this case, from a lakehouse like Lakehouse1), the cache for shortcuts helps by storing the data locally for quick access. The last accessed timestamp and the cache expiration rules determine whether data is fetched from the cache or from the source (Google Cloud Storage, in this case).
Products: The ProductFile.parquet was last accessed 12 hours ago. Since the cache has data available for up to 12 hours, it is likely that this data will be retrieved from the cache, as it hasn't been too long since it was last accessed.
Stores: The StoreFile.json was last accessed 4 hours ago, which is within the cache retention period. Therefore, this data will also be retrieved from the cache.
Trips: The TripsFile.csv was last accessed 48 hours ago. Given that it's outside the typical caching window (assuming the cache has a maximum retention period of around 24 hours), it would not be retrieved from the cache. Instead, it will likely require a fresh read from the source.

NEW QUESTION 102
- (Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

Column name	Data type
Timestamp	Datetime
GeoLocation	Dynamic
Temperature	Decimal
DeviceId	Int

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

Both tables contain millions of rows. You have the following KQL queryset.
You need to reduce how long it takes to run the KQL queryset. Solution: You add the make_list() function to the output columns. Does this meet the goal?

```
01  Stream
02  | extend lat = todecimal(GeoLocation.Latitude), long = todecimal(GeoLocation.Longitude)
03  | join kind=inner Reference on DeviceId
04  | project Timestamp, lat, long, Temperature, DeviceName
05  | filter Temperature >= 10
06  | render scatterchart with (kind = map)
```

A. Yes
B. No

Answer: B

Explanation:
Adding an aggregation like make_list() would require additional processing and memory, which could make the query slower.

NEW QUESTION 106
HOTSPOT - (Topic 3)
Your company has three newly created data engineering teams named Team1, Team2, and Team3 that plan to use Fabric. The teams have the following personas:
• Team1 consists of members who currently use Microsoft Power BI. The team wants to transform data by using by a low-code approach.

• Team2 consists of members that have a background in Python programming. The team wants to use PySpark code to transform data.
• Team3 consists of members who currently use Azure Data Factory. The team wants to move data between source and sink environments by using the least amount of effort.
You need to recommend tools for the teams based on their current personas.
What should you recommend for each team? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

Team1: Dataflow Gen2 dataflows ▼
Data pipelines
Notebooks
Dataflow Gen2 dataflows

Team2: Notebooks ▼
Data pipelines
Notebooks
Dataflow Gen2 dataflows

Team3: Data pipelines ▼
Data pipelines
Notebooks
Dataflow Gen2 dataflows

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Team1: Dataflow Gen2 dataflows ▼
Data pipelines
Notebooks
Dataflow Gen2 dataflows

Team2: Notebooks ▼
Data pipelines
Notebooks
Dataflow Gen2 dataflows

Team3: Data pipelines ▼
Data pipelines
Notebooks
Dataflow Gen2 dataflows

NEW QUESTION 107
- (Topic 3)
You have two Fabric workspaces named Workspace1 and Workspace2.
You have a Fabric deployment pipeline named deployPipeline1 that deploys items from Workspace1 to Workspace2. DeployPipeline1 contains all the items in Workspace1.
You recently modified the items in Workspaces1.
The workspaces currently contain the items shown in the following table.

Workspace	Items
Workspace1	Model1 Notebook1 Report1 Lakehouse1 Pipeline1
Workspace2	Model1 Notebook2 Report1 Lakehouse2

Items in Workspace1 that have the same name as items in Workspace2 are currently paired.
You need to ensure that the items in Workspace1 overwrite the corresponding items in Workspace2. The solution must minimize effort.
What should you do?

A. Delete all the items in Workspace2, and then run deployPipeline1.
B. Rename each item in Workspace2 to have the same name as the items in Workspace1.
C. Back up the items in Workspace2, and then run deployPipeline1.
D. Run deployPipeline1 without modifying the items in Workspace2.

Answer: D

Explanation:
When running a deployment pipeline in Fabric, if the items in Workspace1 are paired with the corresponding items in Workspace2 (based on the same name), the deployment pipeline will automatically overwrite the existing items in Workspace2 with the modified items from Workspace1. There's no need to delete, rename, or back up items manually unless you need to keep versions. By simply running deployPipeline1, the pipeline will handle overwriting the existing items in Workspace2 based on the pairing, ensuring the latest version of the items is deployed with minimal effort.

NEW QUESTION 110
- (Topic 3)
You have a Fabric workspace named Workspace1.
You plan to configure Git integration for Workspacel by using an Azure DevOps Git repository. An Azure DevOps admin creates the required artifacts to support the integration of Workspacel Which details do you require to perform the integration?

A. the project, Git repository, branch, and Git folder
B. the organization, projec
C. Git repository, and branch
D. the Git repository URL and the Git folder
E. the personal access token (PAT) for Git authentication and the Git repository URL

Answer: B

NEW QUESTION 113
HOTSPOT - (Topic 3)
You have a table in a Fabric lakehouse that contains the following data.

SalesOrderNumber	OrderDate	CustomerName	Email
SO49172	2021-01-01	Brian Howard	brian23@adventure-works.com
SO49173	2021-01-01	Linda Alvarez	linda19@adventure-works.com
SO49174	2021-01-01	Gina Hernandez	gina4@adventure-works.com
SO49178	2021-01-01	Beth Ruiz	beth4@adventure-works.com
SO49179	2021-01-01	Evan Ward	evan13@adventure-works.com

You have a notebook that contains the following code segment.

```
01 df = df.withColumn("CustomerName", when((col("CustomerName").isNull() | (col("CustomerName")=="")),lit("Unknown")).otherwise(col("CustomerName")))
02 df = df.withColumn("Username",split(col("Email"), "@").getItem(1))
03 df = df.dropDuplicates(["OrderDate"]).select(col("OrderDate"), year("OrderDate").alias("Year"), ("CustomerName"), ("Username"))
04 display(df.head(10))
```

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
Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value.	○	○
Line 02 will extract the value before the @ character and generate a new column named Username.	○	○
Line 03 will extract the year value from the OrderDate column and keep only the first occurrence for each year.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements	Yes	No
Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value.	⊙	○
Line 02 will extract the value before the @ character and generate a new column named Username.	○	⊙
Line 03 will extract the year value from the OrderDate column and keep only the first occurrence for each year.	○	⊙

NEW QUESTION 116
- (Topic 3)
Exhibit.
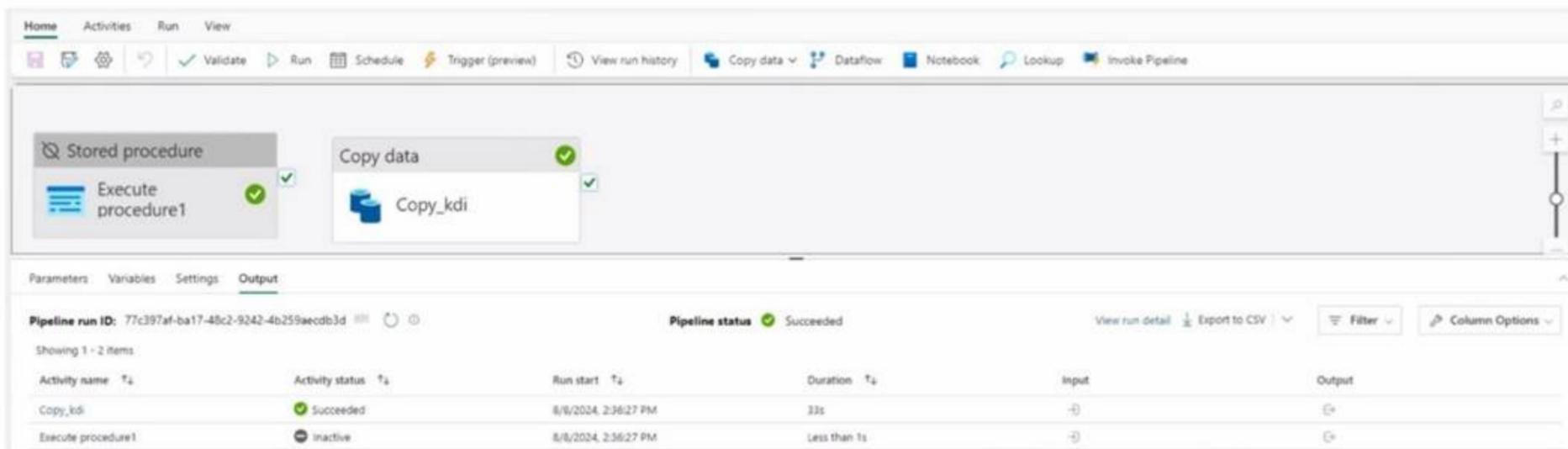
You have a Fabric workspace that contains a write-intensive warehouse named DW1. DW1 stores staging tables that are used to load a dimensional model. The tables are often read once, dropped, and then recreated to process new data.
You need to minimize the load time of DW1. What should you do?

A. Disable V-Order.
B. Drop statistics.
C. Enable V-O-der.
D. Create statistics.

Answer: C

NEW QUESTION 120
- (Topic 3)
You have a Fabric workspace that contains a lakehouse named Lakehousel.
You plan to create a data pipeline named Pipeline! to ingest data into Lakehousel. You will use a parameter named paraml to pass an external value into Pipeline1!. The paraml parameter has a data type of int
You need to ensure that the pipeline expression returns param1 as an int value. How should you specify the parameter value?

A. "@pipeline(). parameter
B. paraml"
C. "@{pipeline().parameters.paraml}"
D. "@{pipeline().parameters.[paraml]}"
E. "@{pipeline().parameters.paraml}-

Answer: B

NEW QUESTION 122
- (Topic 3)
You have a Fabric workspace that contains an eventstream named Eventstream1. Eventstream1 processes data from a thermal sensor by using event stream processing, and then stores the data in a lakehouse.
You need to modify Eventstream1 to include the standard deviation of the temperature. Which transform operator should you include in the Eventstream1 logic?

A. Expand
B. Group by
C. Union
D. Aggregate

Answer: D

Explanation:
To compute the standard deviation of the temperature from the thermal sensor data, you would use the Aggregate transform operator in Eventstream1. The Aggregate operator allows you to apply functions like sum, average, count, and statistical functions like standard deviation across a group of rows or events. This operator is ideal for operations that require summarizing or computing statistics over a dataset, such as calculating the standard deviation.

NEW QUESTION 124
......

Relate Links

100% Pass Your DP-700 Exam with Exambible Prep Materials

https://www.exambible.com/DP-700-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/