

DP -700

**Microsoft Certified:
Fabric Data Engineer
Associate - Preparation
Guide**

Disclaimer: This content is for educational purposes. Accuracy is attempted but not guaranteed.
No job outcome is promised.

Question 1

You are building a data loading pattern for Fabric notebook workloads.

You have the following code segment:

```
def loading_pattern_sample(df_source):
    try:
        deltaTable = DeltaTable.forName(spark, target_table)
    except Exception:
        try:
            df_source.write.format('delta').mode('overwrite').saveAsTable(f'{target_table}')
        except Exception as e:
            print(f':Load for table {target_table} failed with error: {str(e)}')
            raise
    return

    try:
        change_detection_columns = [col for col in df_source.columns if col not in candidate_key]

        match_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate_key])
        update_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change_detection_columns])

        update_expr = {col: f'source.{col}' for col in df_source.columns}

        merge_operation = deltaTable.alias('target').merge(
            source=df_source.alias('source'),
            condition=match_condition
        ).whenMatchedUpdate(
            condition=update_condition,
            set=update_expr
        ).whenNotMatchedInsertAll()

        merge_operation.execute()
    except Exception as e:
        print(f'Insert operation for table {target_table} failed with error: {str(e)}')
    return
```

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

The target table will always be overwritten.

Yes No

The merge operation will always run.

The loading pattern supports both full and incremental loading requirements.

Statements	Yes	No
The target table will always be overwritten.	<input type="radio"/>	<input checked="" type="checkbox"/>
The merge operation will always run.	<input type="radio"/>	<input checked="" type="checkbox"/>
The loading pattern supports both full and incremental loading requirements.	<input checked="" type="checkbox"/>	<input type="radio"/>

Explanation:

"The target table will always be overwritten."

Selected Answer: No

In many data loading strategies, especially when using **incremental loads** or **merge operations**, the target table is **not always overwritten**.

Instead, new data is **appended**, **updated**, or **merged** based on keys.

Overwriting usually happens in **full refresh scenarios**, which is not always the case.

"The merge operation will always run."

Selected Answer: No

The **merge operation** (such as SQL MERGE or Delta Lake MERGE INTO) only runs if **certain conditions are met**, such as the presence of **new or changed data**.

If there is no data to update or merge, it **may not execute**.

Thus, it's correct to say that it **does not always run**.

"The loading pattern supports both full and incremental loading requirements."

Selected Answer: Yes

A **well-designed data pipeline** often supports both **full** and **incremental loads**.

- **Full loads** replace the entire dataset.
- **Incremental loads** append or update only changed records.
Since this is a **common practice**, selecting "**Yes**" is correct.

Question 2

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.

Answer Area

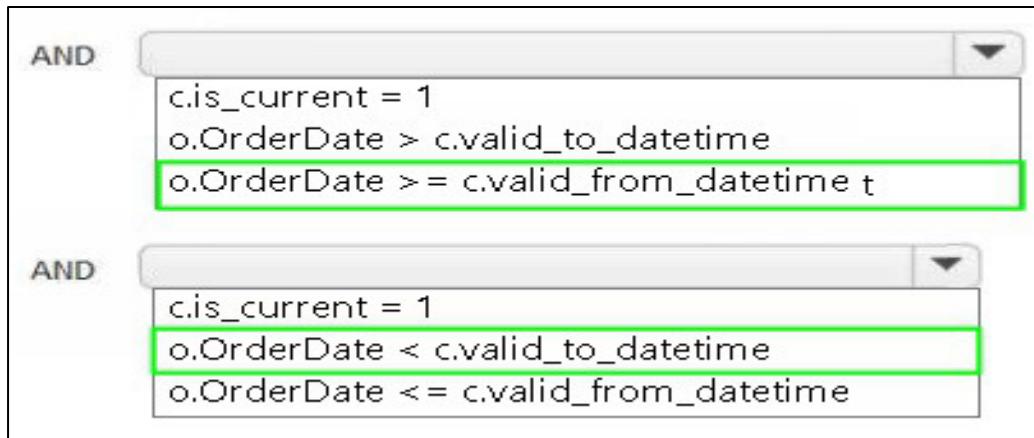
```
SELECT  
    OrderLineID order_line_id  
    ,OrderDate order_date  
    ,c.customer_key  
    ,c.customer_id  
    ,Quantity order_quantity  
    ,UnitPrice unit_price  
    ,TaxRate tax_rate  
FROM  
    Lakehouse1.orderlines o  
INNER JOIN  
    Lakehouse2.dim_customer c  
    ON o.customerid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

AND

```
c.is_current = 1  
o.OrderDate > c.valid_to_datetime  
o.OrderDate >= c.valid_from_datetime t
```

AND

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



Explanation:

1. **o.OrderDate >= c.valid_from_datetime**

This ensures that the **OrderDate** falls on or after the start of the **valid period**.

This is essential to **capture all orders** that are valid based on the entity's timeline.

2. **o.OrderDate < c.valid_to_datetime**

This ensures that the **OrderDate** is strictly **before the valid end date**.

This prevents fetching **orders that occur after the entity has expired**.

Question 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. Eventstream
- B. Dataflow Gen2
- C. Streaming dataset
- D. Data pipeline

Answer: D

Explanation:

D. Data pipeline

A data pipeline is designed to handle large-scale data ingestion and movement efficiently.

It can be configured to automatically trigger the ingestion process when a new file is added to the external data source, ensuring that the data is ingested into Lakehouse1 as soon as it becomes available.

Data pipelines are optimized for high throughput, making them suitable for handling large files (like the 500 GB files mentioned) and ensuring the process is both fast and efficient.

Question 4

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:**A. Data pipeline**

Data pipelines in Fabric are designed for high-throughput data ingestion and can be triggered automatically when new files are added to the external data source.

They are optimized for moving large volumes of data efficiently and can handle the ingestion of 500 GB files without applying transformations.

Question 5

You have a Fabric workspace that contains an eventhouse and a KQL database named Database1.

Database1 has the following:

- A table named Table1
- A table named Table2
- An update policy named Policy1

Policy1 sends data from Table1 to Table2.

The following is a sample of the data in Table2.

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2024-05-18 12:45:17.16524	81416f30- 60a2-4e75- 9b19- 2a84ea059735	[{ "index": 0, "eventid": "719afca0- be30-4559-bb5e- 59feade642f6" }]
2024-05-18 12:45:21.76423	bb664e1e- 02aa-4e17- 8c8a- 116cd4458d52	[{ "index": 0, "eventid": "782222b2- fbcb-43c0-82d6-ecd49a99dbf5" }]
2024-05-18 12:45:23.98642	717bfe7d- 0e5d-498f- 9f21- e60aaaf258056	[{ "index": 0, "eventid": "d5730286- 0da4-41f8-8e59-f75e209310a9" }]

Recently, the following actions were performed on Table1:

- An additional element named temperature was added to the StreamData column.
- The data type of the Timestamp column was changed to date.
- The data type of the Deviceld column was changed to string.

You plan to load additional records to Table2.

Which two records will load from Table1 to Table2?

Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A.

Timestamp (datetime)	Deviceld (guid)	StreamData (dynamic)
2024-05-18	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 40, "eventid": "729afca2-be30-4559-bb5e-59feade642f3", "temperature": 32 }]

B.

Timestamp (datetime)	Deviceld (guid)	StreamData (dynamic)
2024-05-21	81416f30	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-5werade642f6", "temperature": 27 }]

C.

Timestamp (datetime)	Deviceld (guid)	StreamData (dynamic)
2024-05-23	81416f3060a24e759b192a84ea05973532dhdyte3	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6" }]

D.

Timestamp (datetime)	Deviceld (guid)	StreamData (dynamic)
2024-05-24	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6" }]

Answer: B, D

Explanation:

Record B loads because it conforms to the updated schema (string Deviceld, StreamData with temperature).

Record D loads because it conforms to the original schema (GUID Deviceld, no temperature in StreamData).

Question 6

You have a Fabric workspace.

You are debugging a statement and discover the following issues:

- Sometimes, the statement fails to return all the expected rows.
- The PurchaseDate output column is NOT in the expected format of mmm dd, yy.

You need to resolve the issues.

The solution must ensure that the data types of the results are retained.

The results can contain blank cells.

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

SELECT

item_id as ItemId

,convert(varchar(20), item_name)
,convert(varchar(max), item_name)
,try_cast(item_name as varchar(20))

,item_description as ItemDescription

,convert(varchar, purchase_date, 7)
,convert(varchar, purchase_date, 109)
,convert(varchar, purchase_date, 112)

FROM

Table1

WHERE

item_type = @itemtype_parameter

```
SELECT  
    item_id as ItemId  
    ,convert(varchar(20), item_name) as ItemName  
    ,convert(varchar(max), item_name)  
    ,try_cast(item_name as varchar(20))  
  
    ,item_description as ItemDescription  
    ,convert(varchar, purchase_date, 7) as PurchaseDate  
    ,convert(varchar, purchase_date, 109)  
    ,convert(varchar, purchase_date, 112)  
  
FROM  
    Table1  
  
WHERE  
    item_type = @itemtype_parameter
```

Explanation:

1. **try_cast(item_name as varchar(20))**

- Function: TRY_CAST() is a safer alternative to CAST(), introduced in SQL Server 2012.
- Purpose: Attempts to convert item_name into a VARCHAR(20).
- If conversion fails, it returns NULL instead of throwing an error — which helps in debugging inconsistent or unexpected data types.

2. **convert(varchar, purchase_date, 7)**

- Function: CONVERT() is used to transform purchase_date into a string (VARCHAR).
- Purpose: Converts purchase_date to a specific date format.
- Format Code 7 (British/French Format):
 - Format code 7 corresponds to dd/mm/yy (e.g., 05/02/25 for February 5, 2025).
 - Useful when presenting dates in a human-readable format.

Question 7

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:

Enable Staging:

When copying data from a Snowflake data source to a Fabric warehouse, enabling staging can significantly improve the efficiency and reliability of the data transfer process.

Staging involves temporarily storing the data in an intermediate location before loading it into the final destination.

This approach helps in handling large datasets and complex transformations, ensuring that the data is transferred smoothly without interruptions.

It also allows for more manageable and optimized data movement, particularly when dealing with different data storage systems like Snowflake and Fabric.



Question 8

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**.

```
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)
```

You need to reduce how long it takes to run the KQL queryset.

Solution: You change the join type to kind=outer.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

No. An outer join can be more computationally intensive than an inner join because it needs to process all rows from both tables and include rows that don't have matching entries.

Question 9

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.

```
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)
```

You need to reduce how long it takes to run the KQL queryset.

Solution: You change project to extend.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

No. The project operator is used to select specific columns, which can help reduce data volume and optimize performance.

In contrast, extend is used to add new calculated columns to the result set.

They serve different purposes, and changing from project to extend would not improve performance—in fact, it could add overhead.

Question 10

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**.

```
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)
```

Solution: You move the filter to **line 02**.

Does this meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Yes. By applying the where clause early in the query, you reduce the number of rows processed in subsequent operations, which significantly improves performance.

Question 11

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.

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.

Explanation:

Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value.

Yes

This line describes a data cleaning operation where missing or empty entries in a column are filled with a default value, "Unknown". This is a common transformation in tools like Power Query (M Language), SQL, or Python (e.g., `fillna("Unknown")` in pandas). It is valid and accurate, assuming the syntax is correct.

Line 02 will extract the value before the @ character and generate a new column named Username.

No

This sounds like it's trying to split email addresses into usernames. However, without seeing the actual syntax of Line 02, we cannot assume that it correctly: Splits the string using '@', Extracts the first part, Assigns it to a new column. Often, this mistake happens if the line tries to modify the column in-place or fails to assign a new column. So unless it explicitly creates a new column (e.g., `Username = Text.BeforeDelimiter>Email, "@"`), the claim is inaccurate.

Line 03 will extract the year value from the OrderDate column and keep only the first occurrence for each year.

No.

Extracting the year from a date column is easy and common. However, keeping only the first occurrence for each year suggests some kind of grouping and filtering step. Unless Line 03 explicitly includes logic like Group By Year and selecting the first row, it doesn't do what this statement claims. So this statement assumes extra behavior not guaranteed by merely extracting the year.

Question 12

You have a Fabric workspace that contains an eventhouse named **Eventhouse1**.

In Eventhouse1, you plan to create a table named **DeviceStreamData** in a KQL database. The table will contain data based on the following sample.

Timestamp	DeviceID	StreamData
2024-05-18 12:45:17.16524	81416f30-60a2-4e75-9b19-2a84ea059735	[{ "index": 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6", "isActive": false, "latitude": 5.390012, "longitude": -40.100235, "tags": ["tempor"] }]
2024-05-18 12:45:21.76423	bb664e1e-02aa-4e17-8c8a-116cd4458d52	[{ "index": 0, "eventid": "782222b2-fbcb-43c0-82d6-ecd49a99dbf5", "isActive": true, "latitude": -56.153786, "longitude": 130.870907, "tags": ["adipisicing"] }]
2024-05-18 12:45:23.98642	717bfe7d-0e5d-498f-9f21-e60aad258056	[{ "index": 0, "eventid": "d5730286-0da4-41f8-8e59-f75e209310a9", "isActive": true, "latitude": -21.39289, "longitude": 123.959442, "tags": ["ad"] }]
2024-05-18 12:45:25.39523	1a390e71-4faf-4df5-a479-2238d84001f7	[{ "index": 0, "eventid": "9572e141-8692-4d16-89e8-002f9b7e22b6", "isActive": true, "latitude": -84.926214, "longitude": -11.499007, "tags": ["ex"] }]
2024-05-18 12:45:27.43343	2f0ba7d0-6dff-4081-bf7f-d0d39d4c3260	[{ "index": 0, "eventid": "08a42b87-ce84-4bb2-99f0-4fb5c75ff63f", "isActive": true, "latitude": -49.909339, "longitude": -177.505775, "tags": ["laboris"] }]

You need to use a KQL query to develop the solution for Eventhouse1.

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	Answer Area
StreamData:long)	1.
.create table EventStreamData (2.
StreamData:dynamic)	3.
.create function EventStreamData (
TimeStamp:datetime, DeviceId:string	

Answer Area
1. .create table EventStreamData (
2. TimeStamp:datetime, DeviceId:string
3. StreamData:dynamic)

Explanation:

1. **create table DeviceStreamData (**

This is the first step. It initializes the creation of a new table in the KQL database called DeviceStreamData. The create table command begins the definition of the table's schema.

2. **TimeStamp:datetime, Deviceld:string,**

This line defines two essential columns for the table:

- TimeStamp:datetime: Holds the event timestamp, which is critical for time-series data operations.
- Deviceld:string: Stores the device identifier for grouping and filtering operations.

3. **StreamData:dynamic)**

This completes the table schema by adding a column named StreamData with a dynamic data type. The dynamic type is ideal for storing semi-structured or JSON-like data. The closing parenthesis) ends the schema definition.

Final Correct Order:

1. create table DeviceStreamData (
2. TimeStamp:datetime, Deviceld:string,
3. StreamData:dynamic)

Question 13

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. a data pipeline
- B. an Apache Spark job definition
- C. a streaming dataflow
- D. a notebook

Answer: A

Explanation:

A. a data pipeline.

A data pipeline is designed to copy and transform data between sources and destinations, such as transferring data from an on-premises Microsoft SQL Server database to a Fabric warehouse like Warehouse1. It can seamlessly leverage the on-premises data gateway for connectivity and ensure efficient movement of data.

Question 14

You have a Fabric warehouse named DW1 that contains a Type 2 slowly changing dimension (SCD) dimension table named DimCustomer. DimCustomer contains 100 columns.

You need to identify incoming changes to the table and update the records when there is a change. The solution must minimize resource consumption.

What should you use to identify changes to attributes?

- A. a hash function to compare the attributes in the source table.
- B. a direct attributes comparison across the attributes in the DimCustomer table.
- C. a direct attributes comparison for the attributes in the source table.
- D. a hash function to compare the attributes in the DimCustomer table.

Answer: A

Explanation:

A. a hash function to compare the attributes in the source table.

Using a hash function is an efficient way to identify changes, as it minimizes resource consumption. By generating and comparing hash values for attributes, you can quickly detect differences between the source table and the target table without comparing each attribute directly, which can be resource-intensive.

Question 15

You have an Azure SQL database named DB1.

In a Fabric workspace, you deploy an eventstream named EventStreamDB1 to stream record changes from DB1 into a lakehouse.

You discover that events are NOT being propagated to EventStreamDB1.

You need to ensure that the events are propagated to EventStreamDB1.

What should you do?

- D. Enable change data capture (CDC) for DB1.
- B. Create an Azure Stream Analytics job.
- C. Enable Extended Events for DB1.
- D. Enable change data capture (CDC) for DB1.

Answer: D

Explanation:

D: Enable change data capture (CDC) for DB1.

Change Data Capture (CDC) is a feature used to track changes (inserts, updates, and deletes) in a database table and make those changes available in a way that they can be consumed by other systems or processes, such as EventStreamDB1. When events are not being propagated, it typically means that the system responsible for capturing changes (in this case, CDC) is not enabled or configured.



Question 16

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included in the time provided. To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study. At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview -

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure -

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems. The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data. The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment. Fabric -

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items. Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website. The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint. Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure -

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription. The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric. The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes -

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation. Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization. Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers. Data imports must run simultaneously, when possible. The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.

- In the event of a connectivity error, the ingestion processes must attempt the connection again.
- Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA. Semantic models, reports, and dataflows must be stored in WorkspaceB. Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1. Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements. Data Security -

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must only have read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

Question:

You need to recommend a solution to resolve the MAR1 connectivity issues. The solution must minimize development effort.

What should you recommend?

- A. Add a ForEach activity to the data pipeline.
- B. Configure retries for the Copy data activity.
- C. Call a notebook from the data pipeline.
- D. Configure Fault tolerance for the Copy data activity.

Answer: B

Explanation:

B. Configure retries for the Copy data activity.

Configuring retries for the Copy data activity is a straightforward solution that minimizes development effort while addressing connectivity issues. By enabling retries, the pipeline can automatically attempt to reconnect and complete the operation without requiring additional complex configurations or manual intervention.

Question 17

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview -

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure -

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment. Fabric -

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure -

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes -

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers:

- Bronze
- Silver
- Gold

There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including:

- Deduplication
- The handling of missing values
- The standardizing of capitalization

Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA.

Semantic models, reports, and dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation

In the POS1 product data, ProductID values are unique.

The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements. Data Security -

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must only have read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

Question:

You need to recommend a solution for handling old files. The solution must meet the technical requirements.

What should you include in the recommendation?

- A. a data pipeline that includes a Copy data activity
- B. a data pipeline that includes a Delete data activity
- C. a notebook that runs the VACUUM command
- D. a notebook that runs the OPTIMIZE command

Answer: C

Explanation:

The correct answer is C: a notebook that runs the VACUUM command.

The VACUUM command is typically used to handle old files in a data lake or Delta table environment. It removes files that are no longer referenced by the current state of the table. This is essential for cleaning up outdated files and optimizing storage usage, while still preserving the technical requirements that ensure data integrity and compliance with retention policies.



Question 18

You have a KQL database that contains a table named Readings.

You need to build a KQL query to compare the MeterReading value of each row to the previous row based on the Timestamp value.

A sample of the expected output is shown in the following table.

City	Area	MeterReading	Timestamp	PrevMeterReading	PrevTimestamp
Kansas	Area1	1500	2024-07-30 10:00:00		
Kansas	Area2	1520	2024-07-30 11:00:00	1500	2024-07-30 10:00:00
Kansas	Area1	1550	2024-07-30 12:00:00	1520	2024-07-30 11:00:00
Kansas	Area2	1580	2024-07-30 13:00:00	1550	2024-07-30 12:00:00

How should you complete the query? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
<code>evaluate</code>	Readings
<code>extend</code>	<code> filter City == "Kansas"</code>
<code>lookup</code>	<code>[] by Timestamp</code>
<code>project</code>	<code>[] PrevMeterReading = prev(MeterReading),</code>
<code>sort</code>	<code>[] PrevTimestamp = prev(Timestamp)</code>
<code>summarize</code>	<code>[] City, Area, MeterReading, Timestamp, PrevMeterReading, PrevTimeStamp</code>
<code>take</code>	

Answer Area
Readings
<code> filter City == "Kansas"</code>
<code>sort [] by Timestamp</code>
<code>extend [] PrevMeterReading = prev(MeterReading),</code>
<code>PrevTimestamp = prev(Timestamp)</code>
<code>project [] city, Area, MeterReading, Timestamp, PrevMeterReading, PrevTimeStamp</code>

Explanation:

sort by Timestamp.

This line sorts the filtered results (the Kansas readings) based on the values in the Timestamp column.

extend PrevMeterReading = prev(MeterReading), PrevTimestamp = prev(Timestamp):

This is where the magic of looking at previous rows happens. The extend operator adds new columns to your result set.

project City, Area, MeterReading, Timestamp, PrevMeterReading, PrevTimestamp:

The project operator selects which columns to keep in the final output and in what order they should appear.

Question 19

You need to recommend a Fabric streaming solution that will use the sources shown in the following table.

Name	Message size	Description
Source1	10 MB	Contains semi-structured data that has a bigint column in the messages
Source2	25 MB	Contains structured data that has 19 columns
Source3	5 MB	Contains unstructured data that has images in the messages

The solution must minimize development effort.

What should you include in the recommendation for each source? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Source1:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline
- A streaming dataflow

Source2:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline
- A streaming dataflow

Source3:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline
- A streaming dataflow

Answer Area

Source1:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline**
- A streaming dataflow

Source2:

- Apache Spark Structured Streaming**
- An eventstream
- A data pipeline
- A streaming dataflow

Source3:

- Apache Spark Structured Streaming
- An eventstream
- A data pipeline
- A streaming dataflow**

Explanation:

Source1: A data pipeline.

Data pipelines are used for orchestrating and scheduling data movement and transformation tasks.

They support:

- Moving data between systems
- Running ETL jobs
- Integrating batch workloads

Source2: Apache Spark Structured Streaming.

Apache Spark Structured Streaming is designed for real-time stream processing with high scalability.

It's ideal when:

- You need custom logic or transformations
 - You're processing event-based or IoT data in real time
- Often used in notebooks in Microsoft Fabric for advanced streaming logic.

Source3: A streaming dataflow.

Streaming dataflows in Microsoft Fabric provide a no-code/low-code interface for ingesting and transforming streaming data.

Great for:

- Lightweight streaming pipelines
- Visual design of data transformations
- Power users or analysts who want to work with real-time data



Question 20

You are building a data loading pattern for Fabric notebook workloads.

You have the following code segment.

```
def loading_pattern_sample(df_source):
    try:
        deltaTable = DeltaTable.forName(spark, target_table)
    except Exception:
        try:
            df_source.write.format('delta').mode('overwrite').saveAsTable(f'{target_table}')
        except Exception as e:
            print(f':Load for table {target_table} failed with error: {str(e)}')
            raise
    return

try:
    change_detection_columns = [col for col in df_source.columns if col not in candidate_key]

    match_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate_key])
    update_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change_detection_columns])

    update_expr = {col: f'source.{col}' for col in df_source.columns}

    merge_operation = deltaTable.alias('target').merge(
        source=df_source.alias('source'),
        condition=match_condition
    ).whenMatchedUpdate(
        condition=update_condition,
        set=update_expr
    ).whenNotMatchedInsertAll()

    merge_operation.execute()
except Exception as e:
    print(f'Insert operation for table {target_table} failed with error: {str(e)}')
return
```

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

The target table will always be overwritten.

Yes

No

The merge operation will always run.

The code supports incremental loading and the initial table creation.

Answer Area

Statements

The target table will always be overwritten.

Yes

No

The merge operation will always run.

The code supports incremental loading and the initial table creation.

Question 21

Case Study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study: To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1. The company has a data engineering team that uses Python for data processing.

Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files into a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes. In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment – Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process. The company’s sales team reports that during the last month, the sales data has not been up-to-date when they arrive at work in the morning.

Requirements

Planned Changes

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Version Control

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items.

Additional Azure resources must NOT be provisioned.

Data Requirements

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

Question

You need to implement the solution for the book reviews.

Which should you do?

- A. Create a Dataflow Gen2 dataflow
- B. Create a shortcut
- C. Enable external data sharing
- D. Create a data pipeline

Answer: B

Explanation

B. Create a shortcut.

Creating a shortcut in the lakehouse allows you to link to external data sources without making a copy of the data. This means you can make the book reviews available in the lakehouse by creating a shortcut to the location where the book reviews are stored. The data remains in its original location but is accessible from the lakehouse, meeting the requirement of not duplicating the data.



Question: 22

Case Study – Litware, Inc.

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study: To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All-Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment

Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1. The company has a data engineering team that uses Python for data processing.

Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company’s sales team reports that during the last month, the sales data has **NOT** been up-to-date when they arrive at work in the morning.

Planned Changes

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Version Control

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

Data Requirements

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

Question

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

What should you do?

- A. Split the dataflow into two dataflows.
- B. Configure scheduled refresh for the dataflow.
- C. Configure incremental refresh for the dataflow. Set *Store rows from the past to 1 Month*.
- D. Configure incremental refresh for the dataflow. Set *Refresh rows from the past to 1 Year*.
- E. Configure incremental refresh for the dataflow. Set *Refresh rows from the past to 1 Month*.

Answer: E

Explanation

E. Configure incremental refresh for the dataflow. Set Refresh rows from the past to 1 Month.

This approach ensures minimal data transfer while keeping the refresh scope limited to the most recent and relevant data (1 month), which is aligned with the requirement to minimize data transfer.

Question 23

Case Study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study: To display the first question in this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Overview: Company Overview

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview: IT Structure

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

- The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.
- The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment: Fabric

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment: Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment: Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment: Azure

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription. The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment: User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. However, the team experiences transient connectivity errors, which causes the data exports to fail.

Requirements: Planned Changes

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation. Contoso plans to use Azure Repos for source control in Fabric.

Requirements: Technical Requirements

- The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold.
- There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, handling of missing values, and standardizing of capitalization.
- Each layer must be fully populated before moving to the next layer.
- If any step in populating the lakehouses fails, an email must be sent to the data engineers.
- Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA.

Semantic models, reports, and dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements: Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements: Data Security

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must have only read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

HOTSPOT Question

You need to recommend a method to populate the POS1 data to the **lakehouse medallion layers**.

What should you recommend for each layer?

To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Bronze layer:

- A Dataflow Gen2 dataflow
- A notebook
- A pipeline Copy activity
- A pipeline stored procedure

Silver layer:

- A Dataflow Gen2 dataflow
- A notebook
- A pipeline Copy activity
- A pipeline stored procedure

Bronze layer:	<ul style="list-style-type: none">A Dataflow Gen2 dataflowA notebookA pipeline Copy activityA pipeline stored procedure
Silver layer:	<ul style="list-style-type: none">A Dataflow Gen2 dataflowA notebookA pipeline Copy activityA pipeline stored procedure

Answer Area

1. **Bronze Layer:**

A pipeline Copy activity

- The **Bronze Layer** is typically the raw data ingestion layer in a medallion architecture.
- A **Copy activity in a pipeline** is commonly used in Azure Data Factory (ADF) or Synapse Pipelines to ingest and store raw data into the Bronze Layer (such as a Data Lake or Delta Lake).
- This choice ensures **efficient and scalable data ingestion** from various sources.

2. **Silver Layer:**

A notebook

- The **Silver Layer** is used for **data transformation and cleansing**.
- A **notebook** (such as an Azure Databricks or Synapse notebook) is often used to **apply transformations, perform data validation, and enrich the raw ing**

Question 24

Case Study

Note:

Case studies are not timed separately. You can use as much exam time as needed to complete each case. However, there may be additional case studies and sections in this exam. You must manage your time accordingly.

To answer the questions in a case study, refer to the information provided. Case studies may include exhibits and other resources that explain the scenario. Each question is independent of the others in the same case.

At the end of this case study, a review screen will appear. It allows you to review your answers and make changes before proceeding to the next section. After beginning a new section, you cannot return to this case study.

To Start the Case Study

Click the Next button to display the first question. Use the buttons in the left pane to explore the full content of the case study. These buttons display information such as:

- Business Requirements
- Existing Environment
- Problem Statements

If the case study has an All Information tab, note that the content is identical to that on the other tabs. When you're ready to answer questions, click the Question button.

Overview

Company Overview

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by migrating to Microsoft Fabric. They plan to begin using Fabric for marketing analytics.

IT Structure

Contoso has two technical teams:

- Data Engineers: Responsible for ingestion, transformation, and loading. They prefer Python or SQL.
- Data Analysts: Responsible for queries, semantic models, and reports. They are proficient in Power Query and T-SQL.

Existing Fabric Environment

- Fabric capacity: F64, named Cap1.
- All users are allowed to create Fabric items.
- Two workspaces: WorkspaceA and WorkspaceB, both using Pro license mode.

Source Systems

- POS1: A point-of-sale system using SQL Server on Azure VMs (same Entra tenant). The VM is in a private network, with public access blocked.
- MAR1: A SaaS marketing app with 7 entities. It provides REST APIs for exporting data related to:
 - Email open rates
 - Interaction rates
 - Website interactions
- Historical MAR1 data (from prior years) is stored in Parquet files in Amazon S3.
 - 12 files total, ranging from 300 MB to 900 MB.

Product Data (POS1)

Three related tables:

- Products
- ProductCategories
- ProductSubcategories

Product → Subcategory → Category relationship exists.

Azure Environment

- Microsoft Entra tenant with mail-enabled groups:
 - DataAnalysts
 - DataEngineers
- Azure subscription is active.
- An Azure DevOps organization exists and is used for Fabric-related repositories.

User Problems

- The VP of Marketing needs analytics on email content effectiveness.
 - Current manual analysis takes one week.
 - Goal: Reduce to under one day using Fabric.
- Data Engineers successfully exported data from MAR1 but face transient connectivity issues causing export failures.

Requirements

Planned Changes

- Lakehouse1: Will store raw and cleansed source data.
- Lakehouse2: Will serve dimensional models for analytics.
- ~~Azure Repos will be used for source control in Fabric.~~

Technical Requirements

- Use medallion architecture:
 - Bronze → Silver → Gold layers
 - Extensive cleansing of MAR1 data in silver layer:
 - Deduplication
 - Handling missing values
 - Standardizing capitalization
- Each layer must be fully populated before progressing.
- If failure occurs, an email notification must be sent to data engineers.
- Parallel imports when possible.
- Amazon S3 data usage must:
 - Minimize egress costs
 - Prevent raw data duplication in lakehouses

Data Ingestion Requirements

- All ingestion items must be source-controlled with other workspace items.
- Data must land in bronze layer of Lakehouse1 in Delta format.
- Only file format conversion is allowed before data lands in bronze.
- Use built-in connections to minimize development effort.
- Retry on connectivity failure.
- Store:
 - Lakehouses, pipelines, notebooks in WorkspaceA
 - Semantic models, reports, dataflows in WorkspaceB
- Weekly cleanup of unreferenced Delta table files.

Data Transformation

- In POS1, ProductID is unique.
- Gold layer must include only active products (IsActive = 1).
- Exclude categories/subcategories not assigned to any product.

Data Security

- Data Engineers: Read/Write access to all lakehouses and files.
- Data Analysts:
 - Read access only to Delta tables in the Gold layer
 - No access to Bronze and Silver layers
- Data Engineers must be able to commit to source control in WorkspaceA.

Question

You need to ensure that usage of the data in the Amazon S3 bucket meets the technical requirements. What should you do?

Options:

- 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: D**

Explanation:

Enabling caching for the workspace will help minimize egress costs by reducing the volume of data transferred across clouds. Creating a shortcut ensures that the raw data is not duplicated in the lakehouse.

This approach satisfies both of the technical requirements:

- No duplication of raw data
- Reduced cross-cloud transfer costs

Question 25

HOTSPOT - Case Study

Case Study:

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study:

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Company Overview:

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

IT Structure:

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Fabric:

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Source Systems:

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to:

Product Data:

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Azure:

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Planned Changes

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Technical Requirements

- The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold.
- Extensive data cleansing will be required to populate the MAR1 data in the silver layer, including:
 - Deduplication
 - Handling of missing values
- Standardizing capitalization

- Each layer must be fully populated before moving on to the next layer.
- If any step in populating the lakehouses fails, an email must be sent to the data engineers.
- Data imports must run simultaneously, when possible.

Amazon S3 Data Requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Data Ingestion Requirements:

- Must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in Delta format.
- No changes other than file format changes must be implemented before landing in bronze layer.
- Development effort must be minimized and a built-in connection must be used to import source data.
- In the event of a connectivity error, ingestion processes must attempt the connection again.

Storage & Structure Requirements:

- Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA.
- Semantic models, reports, and dataflows must be stored in WorkspaceB.
- Once a week, old files that are no longer referenced by a Delta table log must be removed.

Data Transformation

In the POS1 product data, ProductID values are unique.

The product dimension in the gold layer must include:

- Only active products from product list
 - Active products identified by IsActive = 1
- Exclude product categories and subcategories not assigned to any product (they are not analytically relevant)

Data Security

- Data Engineers must have read and write access to all lakehouses, including underlying files.
- Data Analysts must have read access only to Delta tables in the gold layer.
- Data Analysts must not have access to bronze or silver layers.
- Data Engineers must be able to commit changes to source control in WorkspaceA.

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.

Answer Area

```
SELECT ProductID, ProductNumber,ProductName, ModelName, SubCategoryName, CategoryName
FROM ContosoLake.Products p
    JOIN ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID
    JOIN ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID
WHERE CategoryID = 1;
      CategoryName is not null;
      IsActive = 1;
      IsActive is not null;
      ProductNumber is not null;
      SubCategoryID = 1;
      SubCategoryName is not null;
```

Answer Area

```
SELECT ProductID, ProductNumber,ProductName, ModelName, SubCategoryName, CategoryName
FROM ContosoLake.Products p
    JOIN ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID
    JOIN ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID
WHERE CategoryID = 1;
      CategoryName is not null;
      IsActive = 1;
      IsActive is not null;
      ProductNumber is not null;
      SubCategoryID = 1;
      SubCategoryName is not null;
```

Based on the provided case study, here's how to complete the Apache Spark SQL query to create the **product dimension** while meeting the requirements:

Key Points:

1. Joins Selection:

- The **first join** with `ProductSubCategories` should be a **LEFT OUTER JOIN** because some subcategories may not have any assigned products, and we want to keep all products.
- The **second join** with `ProductCategories` should be an **INNER JOIN** to exclude irrelevant categories that are not assigned to any subcategory.

2. WHERE Clause:

- Only **active products** should be included, which is filtered by `IsActive = 1`.

Final Answer:

- Joins:
 - First Join: LEFT OUTER JOIN
 - Second Join: INNER JOIN
- WHERE Clause: `IsActive = 1`

Explanation:

These selections ensure that:

- All products are retained, even if they are not assigned to a subcategory.
- Only valid categories and subcategories assigned to products are included.
- Only active products are considered.
- The first join should be a LEFT OUTER JOIN to ensure that all products are retained, even if they are not assigned to a subcategory.
- The second join should be an INNER JOIN to exclude categories and subcategories that are not linked to any product, as they are not analytically relevant.
- Only active products, identified by `IsActive = 1`, should be included in the product dimension in the gold layer.

Question 26

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview -

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure -

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment. Fabric -

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

Products -

ProductCategories -

ProductSubcategories -

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure -

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

DataAnalysts: Contains the data analysts

DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes -

Contoso plans to create the following two lakehouses:

Lakehouse1: Will store both raw and cleansed data from the sources

Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization.

Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

Minimize egress costs associated with cross-cloud data access.

Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

The items must be source controlled alongside other workspace items.

Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.

No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.

Development effort must be minimized and a built-in connection must be used to import the source data.

In the event of a connectivity error, the ingestion processes must attempt the connection again. Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA. Semantic models, reports, and dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements. Data Security -

Security in Fabric must meet the following requirements:

The data engineers must have read and write access to all the lakehouses, including the underlying files.

The data analysts must only have read access to the Delta tables in the gold layer.

The data analysts must NOT have access to the data in the bronze and silver layers.

The data engineers must be able to commit changes to source control in WorkspaceA.

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:

ForEach: This activity allows you to iterate over a collection of items and execute activities for each item. In this context, it can be used to process multiple datasets or files within the bronze layer, ensuring that each file is appropriately handled and transformed.

Copy Data: This activity is fundamental in pipelines for data movement. It enables you to copy data from a source to a destination, such as moving data from a staging area to the bronze layer. The Copy Data activity can read the MAR1 data from its source and write it to the bronze layer, ensuring the data is properly ingested.

Question: 27

HOTSPOT -

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 contains the following tables and columns.

Table name	Column name	Data type
Employee	EmployeeID	Int
Employee	EmployeeName	Varchar(128)
Employee	EmployeePosition	Varchar(64)
Contract	EmployeeID	Int
Contract	ContractType	Varchar(64)
Contract	StartDate	Datetime2
Contract	EndDate	Datetime2

You need to denormalize the tables and include the ContractType and StartDate columns in the Employee table.

The solution must meet the following requirements:

- Ensure that the StartDate column is of the date data type.
- Ensure that all the rows from the Employee table are preserved and include any matching rows from the Contract table.
- Ensure that the result set displays the total number of employees per contract type for all the contract types that have more than two employees.

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

```
WITH result AS(
    SELECT e.EmployeeID
        , e.EmployeeName
        , e.EmployeePosition
        , c.ContractType
        ,  (date, c.startDate) as startDate
        , 
            CAST
            CONVERT
            REPLACE
            SUBSTRING
)
FROM Employee AS e
     Contract AS c on c.EmployeeID = e.EmployeeID
    CROSS JOIN
    INNER JOIN
    LEFT OUTER JOIN
    RIGHT OUTER JOIN
)
SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees
    , ContractType
    FROM result
    GROUP BY ContractType
     COUNT(DISTINCT EmployeeID) > 2
    CONTAINS
    HAVING
    UMIT
    WHERE
```

Answer Area

```
WITH result AS(
    SELECT e.EmployeeID
        , e.EmployeeName
        , e.EmployeePosition
        , c.ContractType
        , CAST((date, c.StartDate) AS StartDate)
        , CONVERT(date, c.StartDate) AS StartDate
        , REPLACE(STR(c.StartDate), '.', '-') AS StartDate
        , SUBSTRING(c.StartDate, 1, 4) + '-' + SUBSTRING(c.StartDate, 5, 2) + '-' + SUBSTRING(c.StartDate, 7, 2) AS StartDate
    FROM Employee AS e
        LEFT OUTER JOIN Contract AS c ON c.EmployeeID = e.EmployeeID
)
SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees
    , ContractType
FROM result
GROUP BY ContractType
HAVING COUNT(DISTINCT EmployeeID) > 2
```

Explanation:

1. **CONVERT(date, c.StartDate) AS StartDate**
The CONVERT function is used to explicitly convert data types in SQL Server.
In this case, it converts c.StartDate to date format, which is appropriate.
2. **LEFT OUTER JOIN** between Employee and Contract tables
A LEFT OUTER JOIN ensures all employees are included, even if they do not have a corresponding contract.
If some employees do not have contracts, this join type ensures they are still listed with NULL contract values.
3. **HAVING COUNT(DISTINCT EmployeeID) > 2**
HAVING is used because COUNT(DISTINCT EmployeeID) is an aggregate function, and aggregate functions cannot be used in WHERE.
HAVING filters groups after aggregation.

Question 28

HOTSPOT -

You have an Azure Event Hubs data source that contains weather data.

You ingest the data from the data source by using an eventstream named Eventstream1. Eventstream1 uses a lakehouse as the destination.

You need to batch ingest only rows from the data source where the City attribute has a value of Kansas. The filter must be added before the destination. The solution must minimize development effort.

What should you use for the data processor and filtering?

To answer, select the appropriate options in the answer area.

Answer Area

Data processor:

- A data pipeline
- A Dataflow Gen2 dataflow
- An eventstream with a custom endpoint
- An eventstream with an external data source

Filtering:

- A Filter activity in a data pipeline
- A filter in a Dataflow Gen2 dataflow
- A KQL statement
- An eventstream processor

Answer Area

Data processor:

- A data pipeline
- A Dataflow Gen2 dataflow
- An eventstream with a custom endpoint
- An eventstream with an external data source

Filtering:

- A Filter activity in a data pipeline
- A filter in a Dataflow Gen2 dataflow
- A KQL statement
- An eventstream processor

Explanation:

1. Data Processor: *An eventstream with an external data source*
Eventstream refers to real-time streaming data processing.
Selecting "An eventstream with an external data source" means data is coming from an external system such as IoT devices, logs, or real-time telemetry.
This is appropriate when dealing with real-time ingestion from sources like Azure Event Hubs, IoT Hub, or Kafka.
2. Filtering: *An eventstream processor*
Filtering in streaming systems typically happens during real-time data ingestion to remove irrelevant or unnecessary events before further processing.
An eventstream processor can be used to apply transformations, filtering, and aggregations dynamically.
This ensures that only relevant data moves forward in the pipeline.

Question 29

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: B**Explanation:**

The Group by transform operator contains the Standard deviation aggregation.
The Aggregate transform operator only contains Average, Max, Min, and Sum aggregation.

Question 30

You have an Azure event hub. Each event contains the following fields:
BikepointID -
Street -
Neighbourhood -
Latitude -
Longitude -
No_Bikes -
No_Empty_Docks -

You need to ingest the events. The solution must only retain events that have a Neighbourhood value of Chelsea, and then store the retained events in a Fabric lakehouse. What should you use?

- A. a KQL queryset
- B. an eventstream
- C. a streaming dataset
- D. Apache Spark Structured Streaming

Answer: B

Explanation:

B. an eventstream.

An eventstream is specifically designed for processing and managing events in real-time.

It allows you to filter, transform, and route events efficiently.

In this scenario, you can configure the eventstream to retain only the events where the Neighbourhood value is "Chelsea" and then store the filtered events in a Fabric lakehouse.

This approach ensures that only the relevant events are ingested, adhering to the requirement to retain only specific events based on the Neighbourhood value.



Question 31

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**.

```
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)
```

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?

- A. Yes
- B. No

Answer: B

Explanation:

No. The `make_list()` function aggregates values into a list, which can be useful for grouping or summary analysis, but it does not inherently reduce query execution time. In fact, aggregation functions like `make_list()` may increase the processing overhead, especially over millions of rows, which could negatively impact performance.

Question 32

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment. Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

Existing Environment. Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment. Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment. Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment. Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company’s sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.

Requirements. Planned Changes -

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Requirements. Version Control -

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Requirements. Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

Requirements. Data Requirements -

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

You need to create a workflow for the new book cover images.

Which two components should you include in the workflow?

Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. a time-based schedule
- B. a streaming dataflow
- C. a blob storage action
- D. a data pipeline
- E. a notebook that uses Apache Spark Structured Streaming
- F. a reflex item

Answer: C, D

Explanation:

C. A blob storage action – This is essential for storing and managing the book cover images.

D. A data pipeline – This helps in processing and transferring the images efficiently.

Question 33

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment. Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

Existing Environment. Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment. Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment. Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment. Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected.

Upon further investigation, the data engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company’s sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.

Requirements. Planned Changes -

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Requirements. Version Control -

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Requirements. Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

Requirements. Data Requirements -

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

Question:

What should you recommend that the data engineering team use to ingest the SEO data?

- A. a streaming dataflow
- B. a streaming dataset
- C. a notebook that uses Apache Spark Structured Streaming
- D. an eventstream

Answer: D

Explanation:

D. an eventstream.

Microsoft Fabric Eventstream is a modern tool designed specifically for real-time data ingestion and processing scenarios in Fabric. It supports:

- Near-real-time data capture from various sources (IoT hubs, Event Hubs, etc.)
- Integration with lakehouses, data warehouses, KQL databases, and Power BI
- Processing and routing of streaming data to multiple destinations

This makes **Eventstream** the ideal choice for NRT SEO data ingestion in a Fabric environment.



Question 34

HOTSPOT -

You have a Fabric warehouse named DW1 that contains four staging tables named ProductCategory, ProductSubcategory, Product, and SalesOrder. ProductCategory, ProductSubcategory, and Product are used often in analytical queries.

You need to implement a star schema for DW1. The solution must minimize development effort.

Which design approach should you use?

To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

ProductCategory, ProductSubcategory
and Product must be:

- Added to the model as individual tables
- Denormalized by being added to the SalesOrder table
- Denormalized into a single product dimension table

The joining key must be:

- The product name and the date
- The unique system generated identifier
- The product category name

Answer Area

ProductCategory, ProductSubcategory
and Product must be:

- Added to the model as individual tables
- Denormalized by being added to the SalesOrder table
- Denormalized into a single product dimension table

The joining key must be:

- The product name and the date
- The unique system generated identifier
- The product category name

Explanation:

Denormalized into a single product dimension table.

In dimensional modeling, especially when designing a star schema, it's common to denormalize hierarchies like Product > Subcategory > Category into one dimension table (e.g., DimProduct).

This simplifies relationships, speeds up queries, and is optimal for analytical workloads.

Having one dimension (e.g., DimProduct) containing all relevant attributes makes slicing and dicing in reports easier.

The unique system generated identifier.

The best practice for joining dimension and fact tables is to use a surrogate key or a system-generated unique identifier (such as ProductID).

This ensures efficiency, uniqueness, and referential integrity between the fact (SalesOrder) and dimension (Product) tables.

Question 35

HOTSPOT -

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 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:

Data pipelines
Notebooks
Dataflow Gen2 dataflow

Team2:

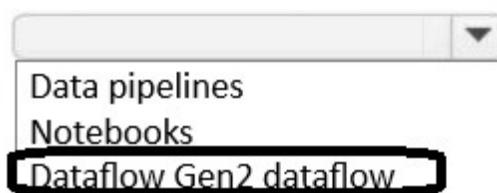
Data pipelines
Notebooks
Dataflow Gen2 dataflow

Team3:

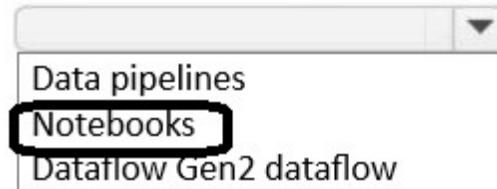
Data pipelines
Notebooks
Dataflow Gen2 dataflow

Answer Area

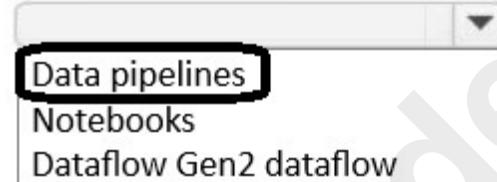
Team1:



Team2:



Team3:



Explanation:

Dataflow Gen2 dataflow.

Team1 has selected Dataflow Gen2 dataflow, which refers to Azure Dataflow Gen2, a data transformation tool used in Azure Data Factory. It is suitable for visually designing data transformations at scale using a code-free environment.

Notebooks.

Team2 has chosen Notebooks, which are typically used in tools like Azure Synapse or Databricks. Notebooks support interactive data analysis, machine learning, and visualizations, written in languages like Python, Scala, or SQL.

Data pipelines.

Team3 has picked Data pipelines, which are end-to-end workflows that move and transform data across different environments. In Azure, this refers to Azure Data Factory pipelines that help in orchestrating and automating data movement and data transformation tasks.

Question 36

HOTSPOT -

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 lakehouse. 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:

- Dataflow Gen2 dataflow
- A notebook
- A T-SQL statement

Dataset2:

- Dataflow Gen2 dataflow
- A notebook
- A T-SQL statement

Dataset3:

- Dataflow Gen2 dataflow
- A KQL queryset
- A T-SQL statement

Answer Area

Dataset1:

- Dataflow Gen2 dataflow
- A notebook
- A T-SQL statement

Dataset2:

- Dataflow Gen2 dataflow
- A notebook
- A T-SQL statement

Dataset3:

- Dataflow Gen2 dataflow
- A KQL queryset
- A T-SQL statement

Explanation:

Dataset1: Dataflow Gen2 dataflow

Dataflow Gen2 is used to ingest, transform, and load (ETL) data using a visual, low-code experience. Pulling in data from various sources (e.g., databases, files), performing lightweight or medium-complexity transformations. Common in Microsoft Fabric and Power BI for reusable, refreshable logic.

Dataset2: A notebook

Notebooks support rich, code-driven data processing using languages like PySpark, Scala, or SQL in a Spark environment. Use a notebook when you need custom transformation logic, are doing advanced analytics, machine learning, or structured streaming. Suitable for big data and complex logic where visual dataflows aren't sufficient.

Dataset3: A T-SQL statement

T-SQL (Transact-SQL) is the language for querying SQL-based engines like Azure SQL Database or Lakehouse SQL Endpoints. Ideal for querying structured data and creating views or computed tables in a Lakehouse or Warehouse. T-SQL statements are highly performant for structured, relational data operations.

Question 37

HOTSPOT

You have a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains a table named Status_Target that has the following columns:

- Key
- Status
- LastModified

The data source contains a table named Status_Source that has the same columns as Status_Target. Status_Source is used to populate Status_Target.

In a notebook named Notebook1, you load Status_Source to a DataFrame named sourceDF and Status_Target to a DataFrame named targetDF.

You need to implement an incremental loading pattern by using Notebook1. The solution must meet the following requirements:

- For all the matching records that have the same value of Key, update the value of LastModified in Status_Target to the value of LastModified in Status_Source.
- Insert all the records that exist in Status_Source that do NOT exist in Status_Target.
- Set the value of Status in Status_Target to inactive for all the records that were last modified more than seven days ago and that do NOT exist in Status_Source.

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

```
...
(targetDF
    .merge(sourceDF, "sourceDF.Key" = "targetDF.Key")
    .whenMatchedInsert(
        .whenMatchedUpdate(
            .whenNotMatchedBySourceInsert(
                .whenNotMatchedBySourceUpdate(
                    .whenNotMatchedInsert(
                        .whenNotMatchedUpdate(
                            set = {"targetDF.LastModified": "sourceDF.LastModified"}
                        )
                    ).whenMatchedInsert(
                        .whenMatchedUpdate(
                            .whenNotMatchedBySourceInsert(
                                .whenNotMatchedBySourceUpdate(
                                    .whenNotMatchedInsert(
                                        .whenNotMatchedUpdate(
                                            values = {
                                                "targetDF.Key": "sourceDF.Key",
                                                "targetDF.LastModified": "sourceDF.LastModified",
                                                "targetDF.Status": "sourceDF.Status"
                                            }
                                        )
                                    ).whenMatchedInsert(
                                        .whenMatchedUpdate(
                                            .whenNotMatchedBySourceInsert(
                                                .whenNotMatchedBySourceUpdate(
                                                    .whenNotMatchedInsert(
                                                        .whenNotMatchedUpdate(
                                                            condition="targetDF.LastModified > (current_date() - INTERVAL '7' DAY)",
                                                            set = {"targetDF.Status": "'inactive'"}
                                                        )
                                                    ).execute()
                                                )
                                            )
                                        )
                                    ...
                                )
                            )
                        )
                    )
                )
            )
        )
    )
)
...
}
```

Answer Area

```
...
(targetDF
    .merge(sourceDF, "sourceDF.Key" = "targetDF.Key")
    .whenMatchedInsert()
    .whenMatchedUpdate()
    .whenNotMatchedBySourceInsert()
    .whenNotMatchedBySourceUpdate()
    .whenNotMatchedInsert()
    .whenNotMatchedUpdate()

    set = {"targetDF.LastModified": "sourceDF.LastModified"})

)

.values = {
    "targetDF.Key": "sourceDF.Key",
    "targetDF.LastModified": "sourceDF.LastModified",
    "targetDF.Status": "sourceDF.Status"
}

)

.condition="targetDF.LastModified > (current_date() - INTERVAL '7' DAY)",
.set = {"targetDF.Status": "'inactive'"}

.execute()

...
...
```

Explanation:

1. whenMatchedUpdate()

Selected for when existing records match between sourceDF and targetDF based on Key.

Action: Update the LastModified field in the target table with the one from the source table.

```
.whenMatchedUpdate()
```

```
    set = { "targetDF.LastModified": "sourceDF.LastModified" }  
)
```

Meaning: If a record with the same Key exists, update its LastModified value to the new one.

2. whenNotMatchedInsert()

Selected for when the record does NOT exist in targetDF.

Action: Insert a new record with fields from the sourceDF.

```
.whenNotMatchedInsert()
```

```
    values = {  
        "targetDF.Key": "sourceDF.Key",  
        "targetDF.LastModified": "sourceDF.LastModified",  
        "targetDF.Status": "sourceDF.Status"  
    }  
)
```

Meaning: If a Key is found in sourceDF but not in targetDF, insert the full new row (Key, LastModified, Status).

3. whenNotMatchedBySourceUpdate()

Selected for when a record exists in targetDF but not in sourceDF.

Action: Mark the existing record as "inactive" if it hasn't been updated recently.

```
.whenNotMatchedBySourceUpdate()
```

```
    condition = "targetDF.LastModified < (current_date() - INTERVAL '7' DAY)",  
    set = { "targetDF.Status": "'inactive'" }  
)
```

Meaning: If a record is missing from the incoming sourceDF AND it was last modified more than 7 days ago, set its Status to "inactive".

Question 38

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.)

	schema_name	table_name
1	Warehouse	ColdRoomTemperatures
2	Warehouse	Colors
3	Warehouse	PackageTypes
4	Warehouse	StockGroups
5	Warehouse	StockItems
6	dbo	BuildVersion
7	dbo	ErrorLog
8	Application	SystemParameters
9	Purchasing	PurchaseOrderLines
10	Purchasing	PurchaseOrders

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.

Actions	Answer Area
Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.	1.
Add a Get metadata activity to query Control.Object and generate a list of schemas and tables to copy.	2.
Add an Until activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.	3.
Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.	
Add a Copy data activity as an inner activity to the iterator activity.	

Answer Area
1. Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.
2. Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.
3. Add a Copy data activity as an inner activity to the iterator activity.

Explanation:

- 1. Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.**

The Lookup activity is used to retrieve metadata — in this case, information from **Control.Object** (likely a control table or database containing a list of source schemas and table names).

This activity returns a list of tables and schemas that need to be copied.

Think of Lookup as fetching a dynamic list. Instead of hardcoding table names, you retrieve them automatically.

- 2. Add a ForEach activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables.**

The ForEach activity is used to loop through the list generated by the Lookup activity.

For each table/schema combination, you perform operations — in this case, copying data into the lakehouse (into **Delta tables**, which are optimized tables supporting **ACID transactions** and fast querying).

ForEach loops allow you to automate operations over multiple tables without manually repeating the same logic for each table.

- 3. Add a Copy data activity as an inner activity to the iterator activity.**

Inside the ForEach loop, the Copy Data activity will **copy the actual data** from the source system into the Delta table in the lakehouse.

Each table in the list will be copied one by one as the loop runs.

This is the main operation that **moves the data**.

Each iteration uses parameters (table name, schema, etc.) from the current item being looped over.

Final Sequence:

- 1. Add a Lookup activity**
- 2. Add a ForEach activity**
- 3. Add a Copy data activity inside the ForEach**

Question 39

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

Explanation:

A. Type 2.

Type 2 slowly changing dimensions allow you to keep a full history of changes over time. In your case, since the goal is to analyze historical sales data by customer location at the time of each sale, you'll need to preserve every past change to the customer's location. Type 2 achieves this by creating a new row in the dimension table for each change, including the date range for when the record is valid.

Question 40

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

EventStream1 outputs events to a table named Table1 in a lakehouse. The streaming data is sourced from motorway sensors and represents the speed of cars.

You need to add a transformation to EventStream1 to average the car speeds. The speeds must be grouped by non-overlapping and contiguous time intervals of one minute. Each event must belong to exactly one window.

Which windowing function should you use?

- A. sliding**
- B. hopping
- C. tumbling**
- D. session

Answer: C

Explanation:

C. Tumbling.

Tumbling windows divide the data stream into fixed, non-overlapping, and contiguous time intervals, such as one-minute windows in this case. Each event belongs to exactly one window, making tumbling windows ideal for calculating averages or other aggregate metrics over defined intervals of time.

Question 41

You have an Azure event hub. Each event contains the following fields:

- BikepointID
- Street
- Neighbourhood
- Latitude
- Longitude
- No_Bikes
- No_Empty_Docks

You need to ingest the events. The solution must only retain events that have a Neighbourhood value of Chelsea, and then store the retained events in a Fabric lakehouse.

Data retention in case of failure is required to be two days.

What should you use?

- A.an eventstream
- B.Apache Spark Structured Streaming
- C.a streaming dataset
- D.a KQL queryset

Answer: A



Question 42

Case Study – This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study: To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview –

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment – Fabric Environment

Litware has a Fabric workspace named **Workspace1**. High concurrency is enabled for Workspace1. The company has a data engineering team that uses Python for data processing.

Existing Environment – Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: **bronze**, **silver**, and **gold**. The sales data is ingested from the ERP system as Parquet files that land in the **Files** folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has **V-Order disabled**.

Litware has image files of book covers in **Azure Blob Storage**. The files are loaded into the **Files** folder.

Existing Environment – Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a **Dataflow Gen2** dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named **AuthorSales** stores the sales data that relates to each author. The table contains a column named **AuthorEmail**. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment – Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment – Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company’s sales team reports that during the last month, the sales data has **not** been up-to-date when they arrive at work in the morning.

Requirements – Planned Changes

Litware recently signed a contract to receive **book reviews**. The provider of the reviews exposes the data in **Amazon Simple Storage Service (Amazon S3)** buckets.

Litware plans to manage **Search Engine Optimization (SEO)** for the authors. The SEO data will be streamed from a **REST API**.

Requirements – Version Control

Litware plans to implement a version control solution in Fabric that will use **GitHub integration** and follow the **principle of least privilege**.

Requirements – Governance Requirements

To control data platform costs, the data platform must use **only Fabric services and items**. Additional Azure resources must **not** be provisioned.

Requirements – Data Requirements

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT)
- Make the book reviews available in the lakehouse without making a copy of the data
- When a new book cover image arrives in the Files folder, process the image as soon as possible

You need to create a workflow for the new book cover images.

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

NOTE: Each correct selection is worth one point.

- A. an activator item
- B. a data pipeline
- C. a blob storage action
- D. a time-based schedule
- E. a streaming dataflow
- F. a notebook that uses Apache Spark Structured Streaming

Answer: B C



Question: 43

Case Study –

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study –

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview –

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment – Fabric Environment

Litware has a Fabric workspace named **Workspace1**. High concurrency is enabled for Workspace1. The company has a data engineering team that uses Python for data processing.

Existing Environment – Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment – Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment – Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment – Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company’s sales team reports that during the last month, the sales data has **NOT** been up-to-date when they arrive at work in the morning.

Requirements – Planned Changes

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Requirements – Version Control

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Requirements – Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

Requirements – Data Requirements

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.
- 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.



Answer Area

```
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
```

Answer Area

```
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
```

Explanation:

queryinsights.frequently_run_queries

number_of_failed_runs > 1

Only this table has the fields specified in the SELECT and WHERE statements.

The data engineering team wants to debug the issue and find queries that cause more than one failure.

Troubleshooting Steps

1. Dataset for Debugging:

The correct dataset to troubleshoot query performance issues is

`queryinsights.frequently_run_queries`, as it provides information about queries that are frequently executed, including their performance metrics.

- `queryinsights.exec_requests_history` and `queryinsights.exec_sessions_history` are not specific to recurring query failures or performance trends.
- `queryinsights.long_running_queries` focuses on queries with extended durations but doesn't help directly with identifying frequent failures.

2. Filtering Conditions:

To identify problematic queries that fail repeatedly, you should filter on `number_of_failed_runs`

`> 1`.

- This ensures that only queries with multiple failures are flagged.
- Other metrics like `max_run_total_elapsed_time_ms > 7200000` focus on duration, not on failure counts, and would not address the primary issue of identifying frequent failures.

Question 44

Case Study –

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study –

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview –

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure –

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment. Fabric –

Contoso has an F64 capacity named **Cap1**. All Fabric users are allowed to create items.

Contoso has two workspaces named **WorkspaceA** and **WorkspaceB** that currently use **Pro license mode**.

Existing Environment. Source Systems –

Contoso has a point of sale (POS) system named **POS1** that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named **MAR1**. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an **Amazon Simple Storage Service (Amazon S3)** bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data –

POS1 contains a product list and related data. The data comes from the following three tables:

- **Products**
- **ProductCategories**
- **ProductSubcategories**

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure –

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- **DataAnalysts**: Contains the data analysts
- **DataEngineers**: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems –

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes –

Contoso plans to create the following two lakehouses:

- **Lakehouse1**: Will store both raw and cleansed data from the sources
- **Lakehouse2**: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements –

The new lakehouses must follow a **medallion architecture** by using the following three layers: **bronze, silver, and gold**. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization.

Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in **WorkspaceA**. Semantic models, reports, and dataflows must be stored in **WorkspaceB**.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation –

In the POS1 product data, **ProductID** values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an **IsActive** value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements. Data Security –

Security in Fabric must meet the following requirements:

- The data engineers must have **read and write access** to all the lakehouses, including the underlying files.
- The data analysts must only have **read access** to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

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

Explanation:

Schedule a data pipeline that calls other data pipelines: This approach allows you to orchestrate and manage the population of medallion layers efficiently. By scheduling a main data pipeline that calls other data pipelines, you can ensure that each step in the data processing workflow is executed in the correct sequence. This method provides better modularity and manageability, as each sub-pipeline can focus on a specific layer or task within the medallion architecture.



Question 45

You have a Fabric eventhouse that contains a KQL database. The database contains a table named **TaxiData**. The following is a sample of the data in **TaxiData**.

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	PULocationID	DOLocationID	payment_type	total_amount
2	2022-06-06T11:08:32Z	2022-06-06T11:22:17Z	1	0.17	231	50	2	7.12
2	2022-06-06T11:12:05Z	2022-06-06T11:20:43Z	1	1.02	161	163	1	10.56
1	2022-06-06T11:15:00Z	2022-06-06T11:25:32Z	1	1.07	142	230	2	17.12
2	2022-06-06T11:29:54Z	2022-06-06T11:49:34Z	2	2.07	162	236	2	12.01
1	2022-06-06T11:50:50Z	2022-06-06T12:07:24Z	2	2.65	140	142	1	7.89

You need to build two KQL queries. The solution must meet the following requirements:

- One of the queries must partition RunningTotalAmount by VendorID.
- The other query must create a column named FirstPickupDateTime that shows the first value of each hour from tpep_pickup_datetime partitioned by payment_type.

How should you complete each query?

To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
<input type="checkbox"/> Row_cumsum	Statement1: TaxiData sort by VendorID asc extend RunningTotalAmount = <input type="text"/> (total_amount, VendorID != prev(VendorID))
<input type="checkbox"/> Row_rank_dense	
<input type="checkbox"/> Row_rank_min	
<input type="checkbox"/> Row_window_session	Statement2: TaxiData sort by tpep_pickup_datetime asc, payment_type asc extend FirstPickupDateTime = <input type="text"/> (tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))

Answer:	
Values <input type="checkbox"/> Row_cumsum <input type="checkbox"/> Row_rank_dense <input type="checkbox"/> Row_rank_min <input type="checkbox"/> Row_window_session	Answer Area Statement1: TaxiData sort by VendorID asc extend RunningTotalAmount = <input type="checkbox"/> Row_cumsum (total_amount, VendorID != prev(VendorID)) Statement2: TaxiData sort by tpep_pickup_datetime asc, payment_type asc extend FirstPickupDateTime = <input type="checkbox"/> Row_window_session (tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))

Explanation:

- row_cumsum: Computes the cumulative sum of a column, resetting when a condition is met.
- row_window_session: Groups records into sessions based on time intervals or other conditions.

Question 46

You are processing streaming data from an external data provider.
You have the following code segment.

```
datatable (Location:string, Company:string, UnitsSold:long)
[
    "New York", "Contoso", 300,
    "New York", "Litware", 1000,
    "New York", "Relecloud", 300,
    "New York", "Fabrikam", 200,
    "Seattle", "Contoso", 300,
    "Seattle", "Litware", 100,
    "Seattle", "Fabrikam", 100,
    "San Francisco", "Relecloud", 500,
    "San Francisco", "Litware", 500,
    "Washington DC", "Litware", 300,
    "Washington DC", "Contoso", 400
]
| sort by Location desc, UnitsSold desc
| extend Rank=row_rank_dense(UnitsSold, prev(Location) != Location)
```

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

Litware from New York will be displayed at the top of the result set.

Yes No

Fabrikam in Seattle will have value = 2 in the Rank column.

Yes No

Litware in San Francisco will have the same value in the Rank column as Litware in New York.

Yes No

Statements

Litware from New York will be displayed at the top of the result set.

Yes No

Fabrikam in Seattle will have value = 2 in the Rank column.

Yes No

Litware in San Francisco will have the same value in the Rank column as Litware in New York.

Yes No

Explanation -

1. "Litware from New York will be displayed at the top of the result set."

- The result is sorted by `Location` in descending order, so `Washington DC` appears at the top of the result set, not New York.
- Answer: No

2. "Fabrikam in Seattle will have value = 2 in the Rank column."

- For `Seattle`, `Litware` is ranked 1 (`1000 UnitsSold`), and `Fabrikam` is ranked 2 (`100 UnitsSold`).
- Answer: Yes

3. "Litware in San Francisco will have the same value in the Rank column as Litware in New York."

- For `San Francisco`, `Litware` has `UnitsSold = 500` and is ranked 1 (tied with `Relecloud`).
- For `New York`, `Litware` has `UnitsSold = 1000` and is ranked 1.
- Both have the same rank (1) within their respective locations.
- Answer: Yes

Question 47

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

Lakehouse1 contains a Delta table named Table1. You analyze Table1 and discover that Table1 contains 2,000 Parquet files of 1 MB each. You need to minimize how long it takes to query Table1.

What should you do?

- A. Disable V-Order and run the OPTIMIZE command.
- B. Disable V-Order and run the VACUUM command.
- C. Run the OPTIMIZE and VACUUM commands.

Answer: C

Explanation:

C. Run the OPTIMIZE and VACUUM commands.

OPTIMIZE Command: Running the OPTIMIZE command on a Delta table helps to combine smaller files into larger ones, which can significantly improve query performance. This process, known as compaction, reduces the number of Parquet files that need to be read during a query, thereby decreasing query latency. In your case, with 2,000 Parquet files of 1 MB each, running OPTIMIZE will consolidate these files into fewer, larger files, making queries faster and more efficient.

VACUUM Command: The VACUUM command cleans up old versions of data files that are no longer needed, which helps to free up storage space and maintain the performance of the Delta table. After running OPTIMIZE, it's a good practice to run VACUUM to remove any obsolete files and further streamline the data storage.

Question 48

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: This view provides detailed information about all active user connections to the SQL server. It includes information about the user, session ID, login time, and more. By querying this view, you can identify which users are currently connected and actively running queries. Use sys.dm_exec_sessions. This view has info about all active user sessions, including user names, session IDs and status.

Question 49

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table in a lakehouse. You need to remove files that are older than seven days and are no longer in use.

Which command should you run?

- A. VACUUM
- B. COMPUTE
- C. OPTIMIZE
- D. CLONE

Answer: A

Explanation:

The VACUUM command is used to clean up old files that are no longer in use, which fits the requirement of removing files that are older than seven days. This command is typically used in data lake environments to delete files that are no longer needed by the system, ensuring that storage is efficiently managed.

Question 50

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 Microsoft.Fabric.ItemUpdateFailed.

Answer: B

Explanation:

B. From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON. Monitoring hub: The Monitoring hub provides detailed logs and information about the execution of your data pipelines. By selecting the latest failed run of Pipeline1, you can access the execution details and diagnose the issue.

View the input JSON: The input JSON contains the parameters, configurations, and the dynamic SQL query used for the Copy data activity. By examining the input JSON, you can identify the specific SQL query that was executed at the time the pipeline failed. This information is crucial for troubleshooting the issue and understanding why the pipeline keeps failing.

Question: 51

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview -

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure -

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment. Fabric -

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure -

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes -

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization.

Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA. Semantic models, reports, and dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements. Data Security -

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must only have read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

Question

You need to ensure that the data analysts can access the gold layer lakehouse.

What should you do?

- A.** Add the DataAnalyst group to the Viewer role for WorkspaceA.
- B.** Share the lakehouse with the DataAnalysts group and grant the Build reports on the default semantic model permission.
- C.** Share the lakehouse with the DataAnalysts group and grant the Read all SQL Endpoint data permission.
- D.** Share the lakehouse with the DataAnalysts group and grant the Read all Apache Spark permission.

Answer: C

Explanation:

C: Share the lakehouse with the DataAnalysts group and grant the Read all data permission.

This approach ensures that data analysts have the necessary read access to the Delta tables in the gold layer, aligning with the requirement that they should not have access to data in the bronze and silver layers.

By granting Read all SQL Endpoint data permission, the analysts get the necessary and sufficient access to query the gold layer data while adhering to the principle of least privilege.

Question:52

You have a Fabric workspace.

You have semi-structured data.

You need to read the data by using T-SQL, KQL, and Apache Spark. The data will only be written by using Spark.

What should you use to store the data?

- A. a lakehouse
- B. an eventhouse
- C. a datamart
- D. a warehouse

Answer: B

Explanation:

B. an eventhouse.

An eventhouse suggests the focus is on storing event-based or streaming data. While the data itself is semi-structured and written using Apache Spark, the choice of an eventhouse aligns with scenarios where real-time ingestion and analysis of data streams are required.

Additionally, an eventhouse is optimized for applications handling high-frequency data events, making it suitable for Spark-based write operations and enabling the integration of T-SQL, KQL, and Spark query capabilities.

Question:53

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. a Dataflow Gen1 dataflow
- B. a data pipeline
- C. a KQL queryset
- D. a notebook

Answer: B

Explanation:

B: a data pipeline.

A data pipeline is the most suitable tool for moving data between different sources and destinations. In this case, you need to copy data from your on-premises Microsoft SQL Server database (Database1) to your Fabric warehouse (Warehouse1). A data pipeline can efficiently handle this task by allowing you to define and manage the data transfer process.

Question:54

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:

B: a data pipeline.

A data pipeline is specifically designed for orchestrating and automating data movement tasks between different sources and destinations. Here's why a data pipeline is the best choice for copying data from your on-premises Microsoft SQL Server database (Database1) to your Fabric warehouse (Warehouse1):

Data pipelines in Microsoft Fabric are designed to facilitate the movement and transformation of data between various sources and destinations. In this scenario, a data pipeline can be configured to copy data from the on-premises SQL Server database to the Fabric warehouse, utilizing the on-premises data gateway for secure connectivity.

Question:55

You have a Fabric F32 capacity that contains a workspace. The workspace contains a warehouse named DW1 that is modelled by using MD5 hash surrogate keys.

DW1 contains a single fact table that has grown from 200 million rows to 500 million rows during the past year.

You have Microsoft Power BI reports that are based on Direct Lake. The reports show year-over-year values.

Users report that the performance of some of the reports has degraded over time, and some visuals show errors.

You need to resolve the performance issues. The solution must meet the following requirements:

- Provide the best query performance
- Minimize operational costs

Which should you do?

- A. Change the MD5 hash to SHA256
- B. Increase the capacity
- C. Enable V-Order
- D. Modify the surrogate keys to use a different data type
- E. Create views

Answer: C

Explanation:

C. Enable V-Order.

V-Order is a feature that optimizes query performance by enabling faster data retrieval, especially for large datasets. It organizes the data in a compressed format that improves storage efficiency and query speed, which directly addresses the issue of degraded performance in reports. Additionally, enabling V-Order minimizes operational costs because it reduces the amount of storage used and accelerates query execution, avoiding the need for expensive resource scaling (like increasing capacity).

Question: 56

You have a Fabric workspace that contains a warehouse named DW1.

DW1 contains the following tables and columns.

You need to create an output that presents the summarized values of all the order quantities by year and product.

The results must include a summary of the order quantities at the year level for all the products.

How should you complete the code?

To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Table name	Column name	Description
SalesOrderDetail	ProductID	Contains the product ID of the ordered product
SalesOrderDetail	ModifiedDate	Contains the date of an order
SalesOrderDetail	OrderQty	Contains the order quantity
Product	ProductID	Contains the unique ID of a product
Product	Name	Contains a product name

Answer Area

```
SELECT CAST(YEAR(so.ModifiedDate) AS OrderDate)
SELECT CONVERT(VARCHAR(4), so.ModifiedDate, 120)
SELECT YEAR(so.ModifiedDate)
, P.Name AS ProductName
, SUM(so.OrderQty) AS OrderQty
FROM [dbo].[SalesOrderDetail] so
INNER JOIN [dbo].[Product] P
ON P.ProductID = SO.ProductID
GROUP BY
CUBE(YEAR(SO.ModifiedDate), P.Name)
GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
ROLLUP(YEAR(SO.ModifiedDate), P.Name)
YEAR(SO.ModifiedDate), P.Name
ORDER BY OrderDate
```

Answer:

Box1 -> SELECT YEAR || Box2 -> ROLLUP(YEAR(SO.ModifiedDATE), P.Name)

Explanation:

The screenshot shows a SQL query in the SSMS query editor. The code is as follows:

```
SELECT CAST
      SELECT CONVERT
      SELECT YEAR
      ,P.Name AS ProductName
      ,SUM(SO.OrderQty) AS OrderQty
   FROM [dbo].[SalesOrderDetail] SO
  INNER JOIN [dbo].[Product] P
    ON P.ProductID = SO.ProductID
 GROUP BY
      CUBE(YEAR(SO.ModifiedDate), P.Name)
      GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
      ROLLUP(YEAR(SO.ModifiedDate), P.Name)
      YEAR(SO.ModifiedDate), P.Name
 ORDER BY OrderDate
```

The highlighted sections are:

- Box1: `SELECT YEAR`
- Box2: `ROLLUP(YEAR(SO.ModifiedDate), P.Name)`

To address the requirements for summarizing order quantities by year and product, with a summary at the year level for all products, let's analyze the SQL solution:

Explanation of Choices:

1. `SELECT YEAR` : Extracts the year from the `ModifiedDate` column, which is essential for grouping data by year.
2. `ROLLUP(YEAR(SO.ModifiedDate), P.Name)` :
 - Adds subtotals at each hierarchical level of grouping (e.g., yearly totals across all products).
 - Ensures the result includes summaries for both each product by year and all products within a year.

Completed Query:

The screenshot shows the completed query in the SSMS results pane. The query is:

```
sql
SELECT YEAR(SO.ModifiedDate) AS OrderDate,
      P.Name AS ProductName,
      SUM(SO.OrderQty) AS OrderQty
   FROM [dbo].[SalesOrderDetail] SO
  INNER JOIN [dbo].[Product] P
    ON P.ProductID = SO.ProductID
 GROUP BY ROLLUP(YEAR(SO.ModifiedDate), P.Name)
 ORDER BY OrderDate;
```

Key Details:

The use of ROLLUP ensures compliance with the requirement for summarized values at different grouping levels. -- `SUM(SO.OrderQty)` calculates the total order quantities.

Question 57

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.

Answer Choices:

- A. Date
- B. ProductName
- C. ProductColor
- D. TransactionID
- E. SalesAmount
- F. ProductID

Answer: B, C, F

Explanation:

- **B. ProductName:** This attribute describes the product and is crucial for understanding and analyzing the data related to each product.
- **C. ProductColor:** This attribute provides additional information about the product, which can be useful for analysis, reporting, and segmentation.
- **F. ProductID:** This is the unique identifier for each product and serves as the primary key for the DimProduct table. It's essential for establishing the relationship between the FactSales table and the DimProduct table.

Question: 58

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:

A. Enable high concurrency for notebooks: High concurrency allows multiple notebooks to share the same Apache Spark session. This setting ensures that different notebooks can run simultaneously within the same session, facilitating collaboration and efficient resource usage.

Question: 59

You have a Fabric workspace named **Workspace1** that contains a lakehouse named **Lakehouse1**.

Lakehouse1 contains the following tables:

- Orders
- Customer
- Employee

The **Employee** table contains **Personally Identifiable Information (PII)**.

A data engineer is building a workflow that requires writing data to the **Customer** table, however, the user does NOT have the elevated permissions required to view the contents of the **Employee** table.

You need to ensure that the data engineer can write data to the **Customer** table without reading data from the **Employee** table.

Which three actions should you perform?

Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Share **Lakehouse1** with the data engineer
- B. Assign the data engineer the Contributor role for **Workspace2**
- C. Assign the data engineer the Viewer role for **Workspace2**
- D. Assign the data engineer the Contributor role for **Workspace1**
- E. Migrate the **Employee** table from **Lakehouse1** to **Lakehouse2**
- F. Create a new workspace named **Workspace2** that contains a new lakehouse named **Lakehouse2**
- G. Assign the data engineer the Viewer role for **Workspace1**

Answer: D, E, F

Explanation:

- **D. Assign the data engineer the Contributor role for Workspace1:**
Assigning the Contributor role to the data engineer for Workspace1 grants them the necessary permissions to write data to the Customer table in Lakehouse1. However, since the data engineer does not have elevated permissions to view the Employee table, they won't be able to access its content.
- **E. Migrate the Employee table from Lakehouse1 to Lakehouse2:**
Moving the Employee table, which contains PII, to a separate Lakehouse2 helps ensure that the data engineer cannot accidentally or intentionally access it. This action keeps sensitive data segregated from the data engineer's operational environment.
- **F. Create a new workspace named Workspace2 that contains a new lakehouse named Lakehouse2:**
By creating a new workspace and lakehouse for the Employee table, you further isolate the sensitive data. The data engineer can still perform their tasks in Workspace1 without accessing Workspace2, ensuring secure data handling and compliance with privacy requirements.

Question: 60

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. STORED PROCEDURE
- B. CONSTRAINT
- C. SCHEMA
- D. FUNCTION

Answer: D

Explanation:

To implement Row-Level Security (RLS) in a Fabric warehouse like DW1, you need to use a FUNCTION to define the filtering logic. Specifically, a user-defined function (UDF) is created and associated with the RLS policy to determine which rows each user can access.

Question: 61

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	Answer Area
## 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:

Code Segments	Answer Area
## 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	

Explanation:

ALTER TABLE dbo.DimCustomer.

This is necessary to modify the structure of an existing table.

Since adding or dropping a primary key constraint requires modifying a table, this statement is correct.

ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)

This statement is used to define a primary key on the CustomerKey column.

It specifies a **NONCLUSTERED primary key**, meaning the physical ordering of data is not changed, and a separate index structure is created.

This selection aligns with the requirement of having a **nonclustered primary key**.

NOT ENFORCED

In some data warehousing scenarios, constraints might not be enforced to allow better query performance and faster data ingestion. If the system does not enforce referential integrity (e.g., in Azure Synapse Analytics), this would be applicable.

Question: 62

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 (SSMS)
- B.Monitoring hub
- C.dynamic management views in Azure Data Studio
- D.a semantic link in a notebook

Answer: D**Explanation:**

D. a semantic link in a notebook.

Semantic link in a notebook: This approach allows you to dynamically execute operations and monitor the refresh progress of the semantic model (Model1) within the interactive and flexible environment of a notebook. By using a semantic link, you can write custom scripts to trigger the refresh process and track its progress in real-time.

Question: 63

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  
| sort by No_Bikes  
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp  
| project BikepointID, street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?

- A.Yes
- B.No

Answer: B

Explanation:

The answer is B. No because the "sort by" is sorting values in descending order (default behavior --> One should add "asc" to sort values as required. The double "project" at the end does not affect the final result.

Question: 64

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:

The default sorting order in KQL is descending (desc), not ascending (asc).

The solution does not explicitly specify asc in the order by clause, so the results will be sorted in descending order by default.

The requirement is to sort the data by No_Bikes in ascending order, which is not achieved without explicitly specifying asc.

Why other answers are not correct:

A. Yes: This would be incorrect because the solution fails to meet the requirement of sorting in ascending order due to the default descending behavior in KQL.

Important Tip:

Always explicitly specify the sorting order (asc or desc) in KQL to avoid confusion, especially since its default behavior differs from SQL.

Question: 65

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
| sort by No_Bikes asc
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp
```

Does this meet the goal?

A.Yes

B.No

Answer: A

Explanation:

The sort and order operators are equivalent.

The provided code segment correctly filters the data for the neighborhood "Sands End" where the number of bikes (No_Bikes) is at least 15. It then explicitly sorts the results by No_Bikes in ascending order using sort by No_Bikes asc and projects the required columns (BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp). This meets all the stated goals of the problem.

Why other answers are not correct:

B. No: This would be incorrect because the solution explicitly specifies asc in the sort by clause, ensuring the data is ordered by No_Bikes in ascending order as required.

Important Tip:

Always ensure that the sorting order is explicitly specified in KQL to match the requirements, as the default behavior might differ from other query languages like SQL.

Question: 66

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:

```
SELECT BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp  
FROM bike_location  
WHERE neighbourhood = 'Sands End'  
AND no_bikes >= 15  
ORDER BY no_bikes
```

Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

The provided solution uses SQL syntax (SELECT, FROM, WHERE, ORDER BY), but the scenario specifies that the data is in a KQL (Kusto Query Language) database.

KQL and SQL have different syntax and functions.

The correct KQL syntax should be used to filter and sort the data in a KQL database.

Why other answers are not correct:

A. Yes: This would be incorrect because the solution uses SQL syntax instead of KQL, which is not applicable in this context.

Important Tip:

Always use the appropriate query language for the database you are working with.

In this case, KQL should be used instead of SQL to interact with the KQL.

Question: 67

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment. Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

Existing Environment. Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment. Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment. Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment. Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against engineering team receives the following error message when the reports fail to load:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company’s sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.

Requirements. Planned Changes -

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Requirements. Version Control -

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Requirements. Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

Requirements. Data Requirements -

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

You need to ensure that processes for the bronze and silver layers run in isolation.

How should you configure the Apache Spark settings?

- A. Disable high concurrency.
- B. Create a custom pool.
- C. Modify the number of executors.
- D. Set the default environment.

Answer: B

Explanation:

B. Create a custom pool.

While disabling high concurrency (Option A) might seem like it isolates processes, it's not the recommended approach for managing isolation in layered architectures like bronze and silver. By creating a custom pool (Option B), you can allocate dedicated resources to each layer, ensuring they run independently without interfering with one another. Custom pools give you fine-grained control over resource allocation, making them the ideal solution for this scenario.

Question: 68

Case Study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment – Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

Existing Environment – Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold.

The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse.

Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment – Sales Data

- Month-end sales data is processed on the first calendar day of each month.
- Data that is older than one month never changes.
- In the source system, the sales data refreshes every six hours starting at midnight each day.
- The sales data is captured in a Dataflow Gen1 dataflow.
- When the dataflow runs, new and historical data is captured.

The dataflow captures the following fields:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail.

Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment – Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment – Performance Issues

Business users perform ad-hoc queries against the warehouse.

The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected.

Upon further investigation, the data engineering team receives the following error message:

“The SQL query failed while running.”

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity, which slows the data ingestion process.

The sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.

Requirements – Planned Changes

- Litware recently signed a contract to receive book reviews. The provider exposes the data in Amazon S3 buckets.
- Litware plans to manage Search Engine Optimization (SEO) for the authors.
- The SEO data will be streamed from a REST API.

Requirements – Version Control

- Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Requirements – Governance Requirements

- To control data platform costs, the data platform must use only Fabric services and items.
- Additional Azure resources must NOT be provisioned.

Requirements – Data Requirements

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

QUESTION

You need to ensure that the authors can see only their respective sales data.

How should you complete the statement?

To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
AuthorSales	CREATE FUNCTION dbo.tvf_rlspredicate(@Author AS varchar(50))
AuthorEmail	RETURNS TABLE
AuthorSales.AuthorEmail	WITH [REDACTED]
BLOCK	AS
FILTER	RETURN SELECT 1 AS tvf_rlspredicate_result
INLINE	WHERE @Author = [REDACTED]
SCHEMABINDING	GO
USER_NAME()	CREATE SECURITY POLICY RLSfilter
	ADD FILTER PREDICATE Security.tvf_rlspredicate(AuthorEmail)
	ON [REDACTED]
	WITH (STATE = ON)

Answer Area

```

CREATE FUNCTION dbo.tvf_rlspredicate(@Author AS varchar(50))
RETURNS TABLE
WITH SCHEMABINDING
AS
RETURN SELECT 1 AS tvf_rlspredicate_result
WHERE @Author = USER_NAME()
GO

CREATE SECURITY POLICY RLSfilter
ADD FILTER PREDICATE Security.tvf_rlspredicate(AuthorEmail)
ON AuthorSales
WITH (STATE = ON)

```

Explanation:

- **SCHEMABINDING:** Ensures the function is bound to the schema of the referenced objects. Required for RLS (Row-Level Security) functions.
- **USER_NAME():** Returns the SQL Server login or database user name.
- **AuthorSales:** The RLS policy is applied to the AuthorSales table.

Question: 69

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 an Azure key vault named KeyVault1 that contains secrets.

You have a Fabric workspace named Workspace1. Workspace contains a notebook named Notebook1 that performs the following tasks:

- Loads stage data to the target tables in a lakehouse
- Triggers the refresh of a semantic model

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes.

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

Solution:

You use the following code segment:

Use notebookutils.credentials.getSecret and specify the key vault URL and key vault secret.

Does this meet the goal?

- A. Yes
B. No

Answer:

B

Explanation:

The method notebookutils.credentials.getSecret() in Microsoft Fabric does not accept a Key Vault URL.

Instead, it requires the name of a linked service (which securely points to the Key Vault) and the name of the secret.

Question: 70

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 an Azure key vault named KeyVault1 that contains secrets.

You have a Fabric workspace named Workspace1. Workspace contains a notebook named Notebook1 that performs the following tasks:

- Loads stage data to the target tables in a lakehouse
- Triggers the refresh of a semantic model

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes.

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

Solution:

You use the following code segment:

Use notebookutils.credentials.putSecret and specify the key vault URL and key vault secret.

Does this meet the goal?

- A. Yes
B. No

Answer:

B

Explanation:

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

The function notebookutils.credentials.putSecret is used to store a secret into a secret scope not retrieve it

Question: 71

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: B

Explanation:

Member: This role allows users to view and interact with all the items in the workspace. When combined with the already assigned object-level permissions to DW1, it ensures that User3 can update the tables in DW1.

Question 72

You have a Fabric capacity that contains a workspace named Workspace1. Workspace1 contains a lakehouse named Lakehouse1, a data pipeline, a notebook, and several Microsoft Power BI reports.

A user named User1 wants to use SQL to analyze the data in Lakehouse1.

You need to configure access for User1. The solution must meet the following requirements:

- Provide User1 with read access to the table data in Lakehouse1.
- Prevent User1 from using Apache Spark to query the underlying files in Lakehouse1.
- Prevent User1 from accessing other items in Workspace1.

What should you do?

- A. Share Lakehouse1 with User1 directly and select Read all SQL endpoint data.
- B. Assign User1 the Viewer role for Workspace1. Share Lakehouse1 with User1 and select Read all SQL endpoint data.
- C. Share Lakehouse1 with User1 directly and select Build reports on the default semantic model.
- D. Assign User1 the Member role for Workspace1. Share Lakehouse1 with User1 and select Read all SQL endpoint data.

Answer: A

Explanation:

A. Share Lakehouse1 with User1 directly and select Read all SQL endpoint data.

Share Lakehouse1 with User1 directly and select Read all SQL endpoint data: This approach grants User1 read access specifically to the table data in Lakehouse1 through the SQL endpoint, without giving them broader permissions in Workspace1 or access to other items. By directly sharing Lakehouse1 and selecting the "Read all SQL endpoint data" option, you ensure User1 can use SQL to analyze the data while preventing them from using Apache Spark to query the underlying files.

Question: 73

You are implementing the following data entities in a Fabric environment:

- **Entity1:** Available in a lakehouse and contains data that will be used as a core organization entity
- **Entity2:** Available in a semantic model and contains data that meets organizational standards
- **Entity3:** Available in a Microsoft Power BI report and contains data that is ready for sharing and reuse
- **Entity4:** Available in a Power BI dashboard and contains approved data for executive-level decision making

Your company requires that specific governance processes be implemented for the data. You need to apply endorsement badges to the entities based on each entity's use case.

Which badge should you apply to each entity?

To answer, drag the appropriate badges to the correct entities.

Each badge may be used once, more than once, or not at all.

NOTE: Each correct selection is worth one point.



Badges

■■■ Certified

■■■ Master data

■■■ Promoted

■■■ Cannot be endorsed

Answer Area

Entity1:

Entity2:

Entity3:

Entity4:

Answer:

Badges

■■■ Certified

■■■ Master data

■■■ Promoted

■■■ Cannot be endorsed

Answer Area

Entity1: ■■■ Master data

Entity2: ■■■ Certified

Entity3: ■■■ Promoted

Entity4: ■■■ Cannot be endorsed

Explanation of Badge Assignments:

1. Master Data

Refers to authoritative data that is central to business operations, often stored in a master data management system.

This is typically well-maintained and used across multiple departments.

Assigned to Entity1, as it represents centralized and validated business data.

2. Certified

Indicates that an entity (such as a dataset or report) is officially validated by an authority in the organization.

Typically used for trusted and critical business data.

Assigned to Entity2 because this entity meets the highest quality standards.

3. Promoted

Indicates that an entity is recommended for use but is not fully certified.

This badge is usually given when an item is considered useful but has not gone through a formal approval process.

Assigned to Entity3, which signifies that it is endorsed for use but not yet fully certified.

4. Cannot be Endorsed

Indicates that an entity does not qualify for endorsement (either promoted or certified).

This could be due to low-quality data, lack of validation, or experimental datasets.

Assigned to Entity4, meaning it has not met the standards for endorsement.

Question: 74

You have three users named User1, User2, and User3.

You have the Fabric workspaces shown in the following table.

Name	Workspace admin
Workspace1	User1
Workspace2	User2

You have a security group named Group1 that contains User1 and User3.

The Fabric admin creates the domains shown in the following table.

Name	Domain admin
Domain1	User1
Domain2	User2

User1 creates a new workspace named Workspace3.

You add Group1 to the default domain of Domain1.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Statements	Yes	No
User3 has Viewer role access to Workspace3.	<input type="radio"/>	<input type="radio"/>
User3 has Domain contributor access to Domain1.	<input type="radio"/>	<input type="radio"/>
User2 has Contributor role access to Workspace3.	<input type="radio"/>	<input type="radio"/>

Answer Area		
Statements	Yes	No
User3 has Viewer role access to Workspace3.	<input checked="" type="checkbox"/>	<input type="radio"/>
User3 has Domain contributor access to Domain1.	<input checked="" type="checkbox"/>	<input type="radio"/>
User2 has Contributor role access to Workspace3.	<input type="radio"/>	<input checked="" type="checkbox"/>

Explanation:

User3 has Viewer role access to Workspace3.

The "Yes" option is selected, meaning User3 does have Viewer access to Workspace3.

The Viewer role allows read-only access to the workspace but does not permit modifications.

User3 has Domain Contributor access to Domain1.

The "Yes" option is selected, meaning User3 has Domain Contributor permissions in Domain1.

The Domain Contributor role typically allows managing content within a domain but does not grant full admin

rights.

User2 has Contributor role access to Workspace3.

The "No" option is selected, meaning User2 does NOT have Contributor access to Workspace3.

The Contributor role would allow editing content in the workspace, but since "No" is selected, User2 lacks these permissions.

Question: 75

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?

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

Answer: D

Explanation:

D. Run deployPipeline1 without modifying the items in Workspace2.

When items in Workspace1 and Workspace2 are paired and you run the deployment pipeline (deployPipeline1), the pipeline will automatically update the paired items in Workspace2 with the changes made in Workspace1. This means that the modifications in Workspace1 will overwrite the corresponding items in Workspace2 without requiring any additional steps.

Question: 76

You have a Fabric workspace named Workspace1 that contains a data pipeline named Pipeline1 and a lakehouse named Lakehouse1.

You have a deployment pipeline named deployPipeline1 that deploys Workspace1 to Workspace2.

You restructure Workspace1 by adding a folder named Folder1 and moving Pipeline1 to Folder1.

You use deployPipeline1 to deploy Workspace1 to Workspace2.

What occurs to Workspace2?

- A. Folder1 is created, Pipeline1 moves to Folder1, and Lakehouse1 is deployed.
- B. Only Pipeline1 and Lakehouse1 are deployed.
- C. Folder1 is created, and Pipeline1 and Lakehouse1 move to Folder1.
- D. Only Folder1 is created and Pipeline1 moves to Folder1.

Answer: A

Explanation:

A. Folder1 is created, Pipeline1 moves to Folder1, and Lakehouse1 is deployed.

- **Folder1 is created:** The deployment pipeline will replicate the structure of Workspace1 in Workspace2, including the creation of Folder1.
- **Pipeline1 moves to Folder1:** Since Pipeline1 was moved to Folder1 in Workspace1, it will be deployed to Folder1 in Workspace2.
- **Lakehouse1 is deployed:** Lakehouse1 is part of Workspace1 and will be deployed to Workspace2 as part of the deployment process.

Question: 77

Your company has a team of developers. The team creates Python libraries of reusable code that is used to transform data.

You create a Fabric workspace name **Workspace1** that will be used to develop ETL solutions by using notebooks.

You need to ensure that the libraries are available by default to new notebooks in Workspace1.

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
■ Change the runtime version.	
■ Install the libraries.	
■ Create a pool.	
■ Create an environment.	
■ Set the default environment.	

Actions	Answer Area
■ Change the runtime version.	
■ Install the libraries.	
■ Create a pool.	
■ Create an environment.	
■ Set the default environment.	

Correct Order:

1. **Create an environment**
2. **Install the libraries**
3. **Set the default environment**

Explanation:

- **Create an environment:** Matches with setting up a new runtime environment.
- **Install the libraries:** Matches with configuring that environment with required packages.
- **Set the default environment:** Ensures all new notebooks in Workspace1 will use this environment automatically.

Question: 78

You have a Fabric workspace that contains a lakehouse and a notebook named Notebook1. Notebook1 reads data into a DataFrame from a table named Table1 and applies transformation logic.

The data from the DataFrame is then written to a new Delta table named Table2 by using a merge operation.

You need to consolidate the underlying Parquet files in Table1.

Which command should you run?

- A. VACUUM
- B. BROADCAST
- C. OPTIMIZE
- D. CACHE

Answer: C**Explanation:**

OPTIMIZE: This command is used to compact small files into larger ones and optimize the layout of data in a Delta table.

By running the **OPTIMIZE** command on Table1, you can consolidate the Parquet files and improve the performance of read and write operations.

Question: 79

You have five Fabric workspaces.

You are monitoring the execution of items by using Monitoring hub.

You need to identify in which workspace a specific item runs.

Which column should you view in Monitoring hub?

- A. Start time
- B. Capacity
- C. Activity name
- D. Submitter
- E. Item type
- F. Job type
- G. Location

Answer: G**Explanation:**

Location: This column displays the **workspace** where the item is being executed, helping you pinpoint the exact workspace of the item.

Question: 80

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: D

Explanation:

D. Fabric Monitor:

This tool provides detailed monitoring and logging capabilities for various components within a Fabric workspace, including notebooks and data processing tasks.

By using **Fabric Monitor**, you can track and analyze the execution details of Notebook1, including the version of Delta used during its execution.



Question: 81

You have a Fabric workspace named Workspace1 that contains a lakehouse named Lakehouse1. Workspace1 contains the following items:

- A Dataflow Gen2 dataflow that copies data from an on-premises Microsoft SQL Server database to Lakehouse1
- A notebook that transforms files and loads the data to Lakehouse1
- A data pipeline that loads a CSV file to Lakehouse1

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

Answer: D

Explanation:

D. data pipeline.

A data pipeline is designed for orchestrating and scheduling workflows in Fabric. It enables you to load items sequentially (one after the other) and can be set to run on a defined schedule, such as every 15 minutes. This makes it the ideal choice for your requirement.

Question: 82

You are building a Fabric notebook named MasterNotebook1 in a workspace. MasterNotebook1 contains the following code.

```
DAG = {
    "activities": [
        { "name": "execute_notebook_1",
          "path": "notebook_01",
          "timeoutPerCellInSeconds": 600,
          "args": {
              "input_value": "999"
            },
          "retry": 1,
          "retryIntervalInSeconds": 30
        },
        { "name": "execute_notebook_2",
          "path": "notebook_02",
          "timeoutPerCellInSeconds": 400,
          "args": {
              "input_value": "888"
            },
          "retry": 1,
          "retryIntervalInSeconds": 30
        },
        { "name": "execute_notebook_3",
          "path": "notebook_03",
          "timeoutPerCellInSeconds": 600,
          "args": {
              "input_value": "777"
            },
          "retry": 1,
          "retryIntervalInSeconds": 30
        },
        { "name": "execute_notebook_3",
          "path": "notebook_03",
          "timeoutPerCellInSeconds": 600,
          "args": {
              "input_value": "777"
            },
          "retry": 1,
          "retryIntervalInSeconds": 30
        }
    ],
    "timeoutInSeconds": 43200,
    "concurrency": 0
}

mssqlutils.notebook.runMultiple(DAG, {"displayDAGViaGraphviz": True})
```

You need to ensure that the notebooks are executed in the following sequence:

1. Notebook_03
2. Notebook_01
3. Notebook_02

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Move the declaration of Notebook_02 to the bottom of the Directed Acyclic Graph (DAG) definition.
- B. Add dependencies to the execution of Notebook_03.
- C. Split the Directed Acyclic Graph (DAG) definition into three separate definitions.
- D. Add dependencies to the execution of Notebook_02.
- E. Change the concurrency to 3.
- F. Move the declaration of Notebook_03 to the top of the Directed Acyclic Graph (DAG) definition.

Answer: D F

Explanation:

D. Add dependencies to the execution of Notebook_02: Adding dependencies ensures that Notebook_02 runs in the proper order within the Directed Acyclic Graph (DAG), maintaining task execution logic and avoiding conflicts.

F. Move the declaration of Notebook_03 to the top of the Directed Acyclic Graph (DAG) definition: Reordering the declaration of Notebook_03 ensures it aligns with the intended flow of execution within the DAG, helping maintain the logical sequencing of tasks.

Question: 83

You have a Fabric workspace that contains a data pipeline named Pipeline1 as shown in the exhibit. (Click the Exhibit tab.)

The screenshot shows the Azure Data Factory interface for Pipeline1. At the top, there are tabs for Home, Activities, Run, View, Validate, Run, Schedule, Trigger (preview), View run history, Copy data, Dataflow, Notebook, Lookup, and Invoke Pipeline. Below the tabs, the pipeline run history is displayed. The pipeline run ID is 77c397af-ba17-48c2-9242-4b259aecdb3d. The status is Succeeded. The table below shows two activities:

Activity name	Activity status	Run start	Duration	Input	Output
Copy_kdi	Succeeded	8/8/2024, 2:36:27 PM	33s	[Icon]	[Icon]
Execute procedure1	Inactive	8/8/2024, 2:36:27 PM	Less than 1s	[Icon]	[Icon]

What will occur the next time Pipeline1 runs?

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

Answer: D

Explanation:

D. Copy_kdi will run and Execute procedure1 will be skipped.

Copy_kdi is set to run, and Execute procedure1 will be skipped under certain conditions or because of pipeline configuration, perhaps based on dependencies, conditions, or a failure in a prior step.

Question 84

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to

complete each case. However, there may be additional case studies and sections on this exam. You must manage

your time to ensure that you are able to complete all questions included in this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the

case study. Case studies might contain exhibits and other resources that provide more information about the

scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to

make changes before you move to the next section of the exam. After you begin a new section, you cannot return

to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore

the content of the case study before you answer the questions. Clicking these buttons displays information such

as business requirements, existing environment, and problem statements. If the case study has an All Information

tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you

are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview -

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure -

The company's IT department has a team of data analysts and a team of data engineers that use analytics

systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power

Query and T-SQL.

Existing Environment. Fabric -

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website. The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint. Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure -

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It

typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes -

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization. Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA. Semantic models, reports, and

dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only

active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant

and must be omitted from the product dimension in the gold layer.

Requirements. Data Security -

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must only have read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

You need to ensure that WorkspaceA can be configured for source control.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. From Tenant setting, set Users can synchronize workspace items with their Git repositories to Enabled.
- B. From Tenant setting, set Users can sync workspace items with GitHub repositories to Enabled.
- C. Configure WorkspaceA to use a Premium Per User (PPU) license.
- D. Assign WorkspaceA to Cap1.

Answer: A D

Explanation:

A. From Tenant setting, set Users can synchronize workspace items with their Git repositories to Enabled.

This setting enables Git integration for workspaces at the tenant level.

Without enabling this, users cannot link a workspace to a Git repo.

D. Assign WorkspaceA to Cap1.

Cap1 refers to a deployment pipeline capacity, typically a Premium capacity.

Workspaces must be assigned to a capacity that supports Git integration, not just shared or PPU.

Git integration requires Premium capacity (P SKUs or Embedded A SKUs), not just PPU.



Question 55**HOTSPOT**

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 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 .

You need to provide User1 with access to the Customer table. The solution must prevent User1 from accessing the

CreditCard column. NOTE: Each correct selection is worth one point.

Answer Area

GRANT ON

ALTER
EXECUTE
READ
SELECT
VIEW

Customers(CustomerID, FirstName, LastName, Phone)

TO

User1
[User1]
[user1@contoso.com]

Answer Area

The screenshot shows a SQL query window with the following text:

```
GRANT SELECT ON Customers(CustomerID, FirstName, LastName, Phone) TO [User1]
```

The word "SELECT" is highlighted with a red rectangular box. The database name "Customers" and the columns "CustomerID", "FirstName", "LastName", and "Phone" are also highlighted with red boxes.

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

Explanation:

GRANT: SELECT.

The SELECT permission allows a user to query data from a table or view. In this case, the permission is being granted on the Customers table.

This includes the columns: CustomerID, FirstName, LastName, Phone.

TO: [User1].

This is the correct syntax for referencing a database principal (like a user or database role) when:

The user name contains special characters

Or to avoid confusion with reserved keywords

The square brackets [] are T-SQL delimiters to ensure the identifier is treated literally.

Question 86

You have a Fabric deployment pipeline that uses three workspaces named Dev, Test, and Prod.

You need to deploy an Eventhouse as part of the deployment process.

What should you use to add the Eventhouse to the deployment process?

- A. an Azure DevOps pipeline
- B. an eventstream
- C. GitHub Actions

Answer: A

Question 87

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. TRIGGER
- B. SCHEMA
- C. FUNCTION
- D. DATABASE ROLE

Answer: C

Question 88

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. SECURITY POLICY
- B. TABLE
- C. TRIGGER
- D. STORED PROCEDURE

Answer: A

Question 89

You have a Fabric F32 capacity that contains a workspace. The workspace contains a warehouse named DW1 that is modelled by using MD5 hash surrogate keys.

DW1 contains a single fact table that has grown from 200 million rows to 500 million rows during the past year.

You have Microsoft Power BI reports that are based on Direct Lake. The reports show year-over-year values.

Users report that the performance of some of the reports has degraded over time and some visuals show errors.

You need to resolve the performance issues. The solution must meet the following requirements:

- Provide the best query performance.
- Minimize operational costs.

Which should you do?

- A. Create views.
- B. Modify the surrogate keys to use a different data type.
- C. Change the MD5 hash to SHA256.
- D. Increase the capacity.
- E. Disable V-Order on the warehouse.

Answer: B

Question 90

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 and provide detailed information about the invalid references.

What should you use?

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

Answer: B

Question: 91

You have a Fabric workspace named Workspace1_DEV that contains the following items:

- 10 reports
- 4 notebooks
- 3 lakehouses
- 2 data pipelines
- 2 Dataflow Gen1 dataflows
- 3 Dataflow Gen2 dataflows
- 5 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.

Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input type="radio"/>
Answer: No/Yes/No		
Explanation:		
Statements	Yes	No
Data from the semantic models will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="checkbox"/>
The Dataflow Gen1 dataflows will be deployed to the target stage.	<input checked="" type="checkbox"/>	<input type="radio"/>
The scheduled refresh policies will be deployed to the target stage.	<input type="radio"/>	<input checked="" type="checkbox"/>

Each correct selection is worth one point.

1. Data from the semantic models will be deployed to the target stage.

Answer: No

Semantic models are only deployed to the target stage in the form of **metadata**. The deployment process does not copy actual data; instead, only the structural and configuration metadata (e.g., model schema and measures) is deployed. The target stage will require a refresh to fetch the data into the semantic models.

Reference: Microsoft Learn - Item Properties Copied During Deployment

2. The Dataflow Gen1 dataflows will be deployed to the target stage.

Answer: Yes

Dataflow Gen1 objects are included in the deployment pipeline and are fully deployed to the target stage, including their configurations. This ensures that Dataflow Gen1 pipelines can run in the target environment. The deployment process supports this functionality without requiring a manual configuration.

3. The scheduled refresh policies will be deployed to the target stage.

Answer: No

The deployment process does not copy or deploy **refresh schedules** for datasets, semantic models, or other items. Although metadata for the items is deployed, refresh schedules must be manually recreated or configured in the target stage. This limitation is highlighted in Microsoft's documentation.

Question: 92

You have a Fabric deployment pipeline that uses three workspaces named Dev, Test, and Prod. You need to deploy an eventhouse as part of the deployment process.

What should you use to add the eventhouse to the deployment process?

- A. GitHub Actions
- B. a deployment pipeline
- C. an Azure DevOps pipeline

Answer: B

Explanation:

B. a deployment pipeline

Deployment Pipeline: In Microsoft Fabric, a deployment pipeline is specifically designed for managing and deploying resources across different environments (Dev, Test, and Prod). It allows you to automate the deployment process, ensuring consistency and efficiency. By using a deployment pipeline, you can easily include the eventhouse in your deployment process and manage its promotion through the different stages (Dev, Test, Prod).

Question: 93

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: B

Explanation:

B. a deployment pipeline

Microsoft Fabric's deployment pipelines provide a built-in mechanism to manage and validate the deployment of artifacts like warehouses. When you use a deployment pipeline to move Warehouse1 from one workspace (Workspace1) to another (Workspace2), the pipeline automatically checks for issues such as invalid references or missing dependencies during the deployment process.

Question: 94

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:

D. both new data and existing data in the eventhouse

When you enable OneLake availability for the eventhouse, all existing data in the eventhouse is copied to OneLake, ensuring that users have access to historical data. Additionally, any new data added to the eventhouse after enabling OneLake availability will also be synchronized and accessible through OneLake. This ensures seamless integration of past and future data for users leveraging OneLake file explorer.

Question: 95

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:

A. service principal

A service principal is a security identity used by applications, services, and automation tools to access specific Azure resources. It provides a secure way to authenticate and authorize API calls between Azure DevOps and Fabric. By using a service principal, you can grant the necessary permissions to deployPipeline1 to interact with the Fabric workspace (Workspace1) and deploy items to higher environments. This approach ensures secure and managed access without relying on individual user credentials.

Question: 96

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:

C. Stores and Products only.

When the cache for shortcuts is enabled in a Fabric workspace, it allows for faster access to the data by caching the files locally. However, the effectiveness of this caching depends on whether the cache was enabled before the files were added to the storage or if the shortcuts were already pointing to those files.

Question 97

You have a Fabric workspace named Workspace1 that contains an Apache Spark job definition named Job1.

You have an Azure SQL database named Source1 that has public internet access disabled.

You need to ensure that Job1 can access the data in Source1.

What should you create?

- A. an on-premises data gateway
- B. a managed private endpoint
- C. an integration runtime
- D. a data management gateway

Answer: B

Explanation:

B. a managed private endpoint

Managed Private Endpoint: This allows secure and private communication between Azure services without exposing data to the public internet. By creating a managed private endpoint, you can establish a direct connection between the Apache Spark job in Workspace1 and the Azure SQL database (Source1) while keeping public internet access disabled. This approach ensures that data transfer happens securely within the Azure network.

To ensure that Job1 can access the data in Source1, you need to create a managed private endpoint. This will allow the Spark job to securely connect to the Azure SQL database without requiring public internet access.

Question 98

You have an Azure Data Lake Storage Gen2 account named storage1 and an Amazon S3 bucket named storage2.

You have the Delta Parquet files shown in the following table.

Name	Stored in	Size	Description
ProductFile	storage1	50 MB	Contains a list of products and their details
TripsFile	storage2	2 GB	Contains one month's worth of taxi trip data
StoreFile	storage2	25 MB	Contains a list of stores and their addresses

You have a Fabric workspace named Workspace1 that has the cache for shortcuts enabled.

Workspace1 contains a lakehouse named Lakehouse1.

Lakehouse1 has the following shortcuts:

- A shortcut to ProductFile aliased as Products
- A shortcut to StoreFile aliased as Stores
- A shortcut to TripsFile aliased as Trips

The data from which shortcuts will be retrieved from the cache?

- A. Trips and Stores only
- B. Products and Store only
- C. Stores only
- D. Products only
- E. Products, Stores, and Trips

Answer: B

Explanation:

B. Products and Stores only.

When the cache for shortcuts is enabled in a Fabric workspace, it allows for faster access to the data by caching the files locally. This means that data accessed through the cached shortcuts is retrieved from the local cache instead of the original storage locations, which improves performance.

Question: 99**HOTSPOT**

You have a Fabric workspace named Workspace1 that contains the items shown in the following table.

Name	Type
Notebook1	Notebook
Notebook2	Notebook
Lakehouse1	Lakehouse
Pipeline1	Data pipeline
Model1	Semantic model

For Model1, the Keep your Direct Lake data up to date option is disabled.

You need to configure the execution of the items to meet the following requirements:

- Notebook1 must execute every weekday at 8:00 AM.
- Notebook2 must execute when a file is saved to an Azure Blob Storage container.
- Model1 must refresh when Notebook1 has executed successfully.

How should you orchestrate each item?

To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Notebook1:	Add Notebook1 to an Apache Spark job definition. Add Notebook1 to Pipeline1. From Real-Time hub, configure the execution of Notebook1.
Notebook2:	Add Notebook2 to an Apache Spark job definition. Add Notebook2 to Pipeline1. From Real-Time hub, configure the execution of Notebook2.
Pipeline1:	Add Pipeline1 to an Apache Spark job definition. Configure the execution of Pipeline1 by using a schedule. From Real-Time hub, configure the execution of Pipeline1.
Model1:	Add Model1 to Pipeline1. From Real-Time hub, configure Model1 to refresh. Set Keep your Direct Lake data up to date to On.

Answer:

- Box1 → 2
- Box2 → 3
- Box3 → 2
- Box4 → 1

Explanation:

Notebook1:

Add Notebook1 to an Apache Spark job definition.
Add Notebook1 to Pipeline1.
From Real-Time hub, configure the execution of Notebook1.

Notebook2:

Add Notebook2 to an Apache Spark job definition.
Add Notebook2 to Pipeline1.
From Real-Time hub, configure the execution of Notebook2.

Pipeline1:

Add Pipeline1 to an Apache Spark job definition.
Configure the execution of Pipeline1 by using a schedule.
From Real-Time hub, configure the execution of Pipeline1.

Model1:

Add Model1 to Pipeline1.
From Real-Time hub, configure Model1 to refresh.
Set Keep your Direct Lake data up to date to On.

1. Notebook1:

- Selection: Add Notebook1 to Pipeline1.
- Justification: To schedule Notebook1 for execution every weekday at 8:00 AM, adding it to a pipeline that allows scheduled execution is the correct choice.

2. Notebook2:

- Selection: From Real-Time hub, configure the execution of Notebook2.
- Justification: The requirement states that Notebook2 should execute when a file is saved to an Azure Blob Storage container. The Real-Time hub provides event-driven execution, making this the correct option.

3. Pipeline1:

- Selection: Configure the execution of Pipeline1 by using a schedule.
- Justification: Pipeline1 likely orchestrates multiple tasks, and scheduling its execution ensures that it runs automatically according to a predefined timetable.

4. Model1:

- Selection: Add Model1 to Pipeline1.
- Justification: Since Model1 needs to refresh after Notebook1 executes successfully, including it in the same pipeline ensures proper orchestration and dependencies.

Question 100

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: D

Explanation:

Fabric Admin:

Possesses the highest level of permissions within the Fabric environment, enabling the creation of domains and subdomains, as well as the assignment of resources to those subdomains.

Question: 101

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 an Azure Key Vault named KeyVault1 that contains secrets.

You have a Fabric workspace named Workspace1.

Workspace contains a notebook named Notebook1 that performs the following tasks:

- Loads stage data to the target tables in a lakehouse
- Triggers the refresh of a semantic model

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes.

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

Solution:

You use the following code segment:

Use notebookutils.credentials.getSecret and specify the **Key Vault URL** and the **name of a linked service**.

Does this meet the goal?

- A. Yes
- B. No

Answer:

A

Explanation:

A. Yes

The notebookutils.credentials.getSecret function is designed to retrieve secrets from Azure Key Vault in a Fabric environment.

By specifying the Key Vault URL and the name of a linked service, you can successfully access the registered application ID and secret stored in KeyVault1.

This method ensures secure retrieval and meets the goal for generating the authentication token.

Question: 102

You have two Fabric notebooks named:

- Load_Salesperson
- Load_Orders

They 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.

❖ Task:

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 to the correct targets.

Each value may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
activities	DAG = { " " : [{ "name": "Load_Salesperson", "path": "Load_Salesperson", "timeoutPerCellInSeconds": 300, }, { "name": "Load_Orders", "path": "Load_Orders", "timeoutPerCellInSeconds": 600, }], "timeoutInSeconds": 43200 }
broadcast	
dependencies	
execute	
notebooks	
runMultiple	mssparkutils.notebook. (DAG)

Answer Area

```
DAG = {
    "activities": [
        {
            "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)
```

Explanation:

- activities: to list the activities.
- dependencies: to define the execution order.
- runMultiple: because two notebooks must be executed in sequence.

Question: 103

You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse2.

A team of data analysts has Viewer role access to Workspace1.

You create a table by running the following statement.

```
CREATE TABLE [warehouse2].[dbo].[CreditCard]
(
    CreditCard varchar(20) NOT NULL
    ,CreditCardType varchar(10) NOT NULL
)
GO
```

You need to ensure that the team can view only the first two characters and the last four characters of the CreditCard attribute.

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

```
TABLE dbo.CreditCard
ALTER
CREATE
DEFAULT
DROP
EMAIL
PARTIAL
REPLACE
UPDATE

COLUMN [CreditCard]
ALTER
CREATE
DEFAULT
DROP
EMAIL
PARTIAL
REPLACE
UPDATE

ADD MASKED

WITH (FUNCTION = ' (2, "XXXXXXXXXX", 4) ')
ALTER
CREATE
DEFAULT
DROP
EMAIL
PARTIAL
REPLACE
UPDATE
```

Answer Area

```
TABLE dbo.CreditCard
ALTER
CREATE
DEFAULT
DROP
EMAIL
PARTIAL
REPLACE
UPDATE

COLUMN [CreditCard]
ALTER
CREATE
DEFAULT
DROP
EMAIL
PARTIAL
REPLACE
UPDATE

ADD MASKED

WITH (FUNCTION = '(
2,"XXXXXXXXXX",4)')

ALTER
CREATE
DEFAULT
DROP
EMAIL
PARTIAL
REPLACE
UPDATE
```

Explanation:

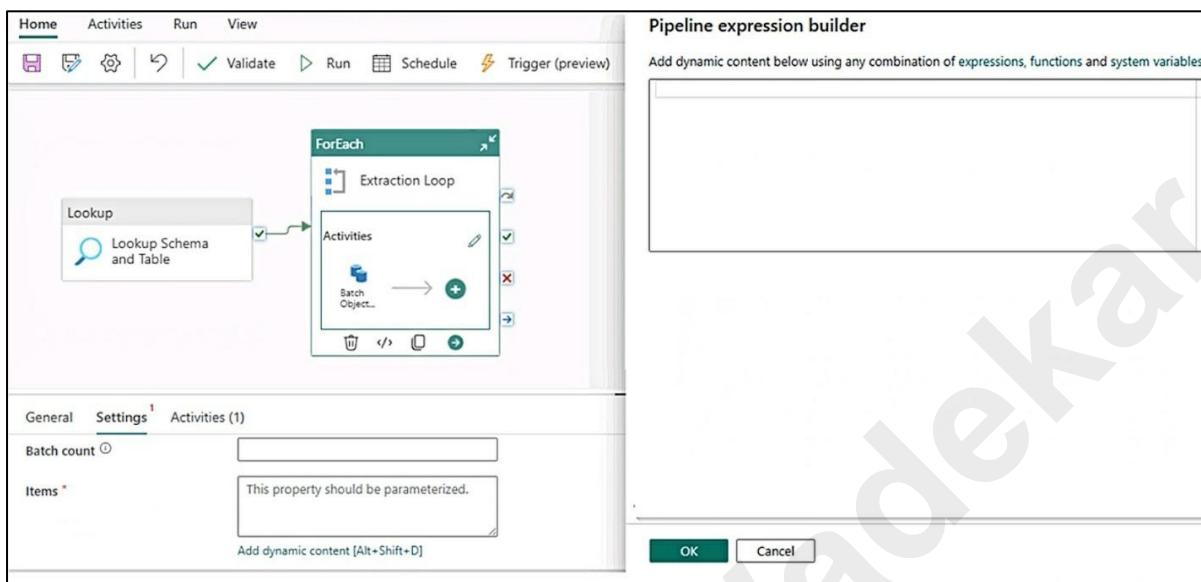
ALTER TABLE - to modify the table

ALTER COLUMN - to modify the column

PARTIAL(prefix padding, padding string, suffix padding) - to expose first and last n characters, adding custom padding 'xxx' of a text field

Question: 104

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('
    Batch Object Copy
    Dynamic Data Copy
    Extraction Loop
    Lookup Schema and Table
').
```

output
output.count
output.pipelineReturnValue
output.value

Answer:

Answer Area

```
@activity('
    Batch Object Copy
    Dynamic Data Copy
    Extraction Loop
    Lookup Schema and Table
').
```

output
output.count
output.pipelineReturnValue
output.value

Explanation:

- **Lookup Schema and Table:**
A Lookup activity is used to retrieve a list of items (like table names and schemas) from a data source, commonly used to feed dynamic structures into ForEach activities.
- **output.value:**
The output from the Lookup activity contains a value property that holds the retrieved data (typically an array). The ForEach activity should iterate over this @activity('Lookup Activity Name').output.value.

Question 105:

You have a Fabric workspace that contains a warehouse named **Warehouse1**. Warehouse1 contains a table named **DimCustomers**. DimCustomers contains the following columns:

- CustomerName
- CustomerID
- BirthDate
- EmailAddress

You need to configure security to meet the following requirements:

- **BirthDate** in **DimCustomer** must be masked and display **1900-01-01**.
- **EmailAddress** in **DimCustomer** must be masked and display only the **first leading character and the last five characters**.

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

ALTER TABLE DimCustomer

ALTER COLUMN BirthDate

ADD MASKED WITH (FUNCTION =

```
'default()'  
'partial(1900-01-01)'  
'random(1900-01-01, 1900-01-01)'
```

ALTER TABLE DimCustomer

ALTER COLUMN EmailAddress

ADD MASKED WITH (FUNCTION =

```
'default()'  
'email()'  
'partial(1,'@',5)'  
'random(1,'@',5)'
```



Answer Area

```
ALTER TABLE DimCustomer
```

```
ALTER COLUMN BirthDate
```

```
ADD MASKED WITH (FUNCTION =
```

```
'default()'  
'partial(1900-01-01)'  
'random(1900-01-01, 1900-01-01)'
```

```
ALTER TABLE DimCustomer
```

```
ALTER COLUMN EmailAddress
```

```
ADD MASKED WITH (FUNCTION =
```

```
'default()'  
'email()'  
'partial(1,"@",5)'  
'random(1,"@",5)'
```

Explanation:

- **default()** is used for **BirthDate** to mask it as a fixed default value based on type — for **date**, this results in 1900-01-01.
- **partial(1,"@",5)** is the correct syntax for masking **EmailAddress**:
 - 1 = Show first 1 character
 - "@" = Pad middle characters with @
 - 5 = Show last 5 characters

Question: 106

You have a Fabric workspace named Workspace1 that contains the following items:

- A Microsoft Power BI report named Report1
- A Power BI dashboard named Dashboard1
- A semantic model named Model1
- 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, Model1, and Dashboard1 only
- B. Lakehouse1, Model1, Report1 and Dashboard1
- C. Report1 and Dashboard1 only
- D. Model1, Report1, and Dashboard1 only
- E. Lakehouse1, Model1, and Report1 only

Answer: B

Explanation:

B. Lakehouse1, Model1, Report1, and Dashboard1 is the correct answer.

Lakehouse Yes Can be endorsed to show it's a trusted data source

Model (Semantic Model) Yes Frequently used in Power BI, key for analytics and reporting

Report Yes Reports can be promoted/certified to guide users toward reliable content

Dashboard Yes Dashboards can also be endorsed for visibility and trust

Question 107

You have a Fabric workspace named Workspace1.

Your company acquires GitHub licenses.

You need to configure source control for Workspace1 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

Explanation:

C. Contents (Read and write) only.

To commit code to GitHub while adhering to the principle of least privilege, you need permissions limited to

Contents (Read and write) to access and update the repository's content. This ensures you can perform the

required actions without granting unnecessary permissions like Actions, which are not needed for committing code.

Question 108

You have a Fabric workspace named Workspace1.

You plan to configure Git integration for Workspace1 by using an Azure DevOps Git repository.

An Azure DevOps admin creates the required artifacts to support the integration of Workspace1.

Which details do you require to perform the integration?

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

Answer: A

Explanation:

A. the organization, project, Git repository, and branch.

To configure Git integration for a Microsoft Fabric workspace with an Azure DevOps Git repository, you need

to provide details about the organization, project, Git repository, and branch. These details ensure that the

workspace is correctly linked to the desired repository and branch for version control and collaboration.

Question 109

You have a Fabric workspace that contains a lakehouse and a semantic model named Model1.

You use a notebook named Notebook1 to ingest and transform data from an external data source.

You need to execute Notebook1 as part of a data pipeline named Pipeline1. The process must meet the following requirements:

- Run daily at 07:00 AM UTC.
- Attempt to retry Notebook1 twice if the notebook fails.
- After Notebook1 executes successfully, refresh Model1.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Place the Semantic model refresh activity after the Notebook activity and link the activities by using the On success condition.
- B. From the Schedule settings of Pipeline1, set the time zone to UTC.
- C. Set the Retry setting of the Notebook activity to 2.
- D. From the Schedule settings of Notebook1, set the time zone to UTC.
- E. Set the Retry setting of the Semantic model refresh activity to 2.
- F. Place the Semantic model refresh activity after the Notebook activity and link the activities by using an On completion condition.

Answer: ABC

Explanation:

- A. Place the Semantic model refresh activity after the Notebook activity and link the activities by using the On success condition.
- B. From the Schedule settings of Pipeline1, set the time zone to UTC.
- C. Set the Retry setting of the Notebook activity to 2.

- **A - On Success condition:** Ensures proper sequencing. The semantic model refresh activity will only run if the Notebook activity is successful.
- **B - Time zone setting for Pipeline1:** By configuring the time zone to UTC, the scheduling of the pipeline becomes consistent and clear across global teams or systems.
- **C - Retry setting for Notebook:** Setting a retry count helps ensure robustness, as transient failures can automatically trigger retries to avoid manual intervention.

Question: 110

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

You plan to create a data pipeline named Pipeline1 to ingest data into Lakehouse1. You will use a parameter named

param1 to pass an external value into Pipeline1. The param1 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().parameters.param1"
- B. "@ pipeline().parameters.param1 "
- C. "@ pipeline().parameters.[param1] "
- D. "@@ pipeline().parameters.param1 "

Answer: B

Explanation:

B. "@ pipeline().parameters.param1 ".

The syntax @ pipeline().parameters.param1 is used in data pipelines to dynamically reference the parameter

param1. This ensures that the parameter value is correctly evaluated as an integer during pipeline execution.

The curly braces are essential for indicating dynamic expression evaluation.

Question 111

You have a Fabric notebook named **Notebook1** that has been executing successfully for the last week.

During the last run, Notebook1 executed 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:

E. the run series from the details of the application run.

The run series from the details of the application run: This option allows you to view a detailed timeline of the jobs that were executed during the last run of Notebook1. The run series provides a chronological view of all the jobs, including their start and end times, which enables you to visualize the execution timeline effectively.

Question 112

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:

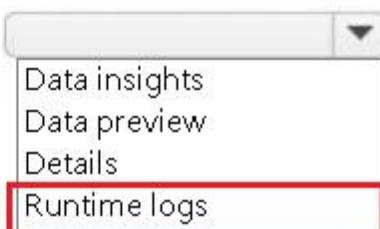
Data insights
 Data preview
 Details
 Runtime logs

To find the total number of errors:

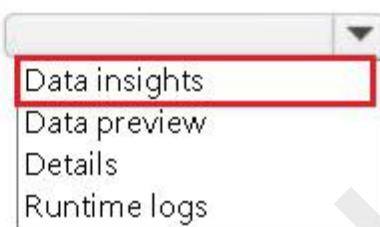
Data insights
 Data preview
 Details
 Runtime logs

Answer Area

To find the error details:



To find the total number of errors:



- The error details, including the occurrence time → Runtime logs
- The total number of errors → Data insights

Explanation:

- Runtime logs provide detailed error messages and timestamps when the error occurred.
- Data insights summarize metrics such as the total number of errors, throughput, and performance statistics.

Question 113

Case Study –

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study –

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview –

Litware, Inc. is a publishing company that has an online bookstore and several retail bookstores worldwide. Litware also manages an online advertising business for the authors it represents.

Existing Environment. Fabric Environment

Litware has a Fabric workspace named Workspace1. High concurrency is enabled for Workspace1.

The company has a data engineering team that uses Python for data processing.

Existing Environment. Data Processing

The retail bookstores send sales data at the end of each business day, while the online bookstore constantly provides logs and sales data to a central enterprise resource planning (ERP) system.

Litware implements a medallion architecture by using the following three layers: bronze, silver, and gold. The sales data is ingested from the ERP system as Parquet files that land in the Files folder in a lakehouse. Notebooks are used to transform the files in a Delta table for the bronze and silver layers. The gold layer is in a warehouse that has V-Order disabled.

Litware has image files of book covers in Azure Blob Storage. The files are loaded into the Files folder.

Existing Environment. Sales Data

Month-end sales data is processed on the first calendar day of each month. Data that is older than one month never changes.

In the source system, the sales data refreshes every six hours starting at midnight each day.

The sales data is captured in a Dataflow Gen1 dataflow. When the dataflow runs, new and historical data is captured. The dataflow captures the following fields of the source:

- Sales Date
- Author
- Price
- Units
- SKU

A table named AuthorSales stores the sales data that relates to each author. The table contains a column named AuthorEmail. Authors authenticate to a guest Fabric tenant by using their email address.

Existing Environment. Security Groups

Litware has the following security groups:

- Sales
- Fabric Admins
- Streaming Admins

Existing Environment. Performance Issues

Business users perform ad-hoc queries against the warehouse. The business users indicate that reports against the warehouse sometimes run for two hours and fail to load as expected. Upon further investigation, the data engineering team receives the following error message when the reports fail to load: "The SQL query failed while running."

The data engineering team wants to debug the issue and find queries that cause more than one failure.

When the authors have new book releases, there is often an increase in sales activity. This increase slows the data ingestion process.

The company's sales team reports that during the last month, the sales data has NOT been up-to-date when they arrive at work in the morning.

Requirements. Planned Changes –

Litware recently signed a contract to receive book reviews. The provider of the reviews exposes the data in Amazon Simple Storage Service (Amazon S3) buckets.

Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.

Requirements. Version Control –

Litware plans to implement a version control solution in Fabric that will use GitHub integration and follow the principle of least privilege.

Requirements. Governance Requirements

To control data platform costs, the data platform must use only Fabric services and items. Additional Azure resources must NOT be provisioned.

Requirements. Data Requirements –

Litware identifies the following data requirements:

- Process the SEO data in near-real-time (NRT).
- Make the book reviews available in the lakehouse without making a copy of the data.
- When a new book cover image arrives in the Files folder, process the image as soon as possible.

What should you do to optimize the query experience for the business users?

- A. Enable V-Order.
- B. Create and update statistics.
- C. Run the VACUUM command.
- D. Introduce primary keys.

Answer: B

Explanation:

- B. Create and update statistics.

Creating and updating statistics helps optimize query performance by providing the query engine with accurate information about the data distribution. This allows the engine to generate efficient query execution plans, ultimately improving the query experience for business users.

Question: 114

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.

Answer: B. views from the queryinsights schema

Explanation:

The **queryinsights schema** in Microsoft Fabric provides system views like `queryinsights.query_text`, `queryinsights.query_stats`, and `queryinsights.long_running_queries`. These views allow you to retrieve rich diagnostic data about query executions, including **query text**, **duration**, **start time**, and **usernames** of those who executed the queries. This makes it the ideal tool for isolating performance issues within a given time window.

Question 115

You have a Fabric workspace that contains a semantic model named **Model1**. You need to monitor the refresh history of Model1 and visualize the refresh history in a chart.

Answer: B. a notebook

Explanation:

While you can view basic refresh history from the semantic model settings, only a notebook allows you to programmatically query and visualize refresh history logs or telemetry. With a notebook, you can extract refresh metadata using the appropriate APIs or queries and then use built-in visualization libraries (like matplotlib or seaborn) to chart refresh trends, durations, and success/failure rates.

Question 116

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.

Answer: D. Disable V-Order

Explanation:

V-Order improves query performance for **read-heavy analytical workloads** by optimizing data file layout. However, it introduces additional overhead during **write operations** because data must be sorted and reorganized into the V-Order format. Since DW1 is **write-intensive**, with staging tables being frequently dropped and recreated, **disabling V-Order** will reduce write latency and improve overall load performance.

Question 117

Case Study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview. Company Overview -

Contoso, Ltd. is an online retail company that wants to modernize its analytics platform by moving to Fabric. The company plans to begin using Fabric for marketing analytics.

Overview. IT Structure -

The company's IT department has a team of data analysts and a team of data engineers that use analytics systems.

The data engineers perform the ingestion, transformation, and loading of data. They prefer to use Python or SQL to transform the data.

The data analysts query data and create semantic models and reports. They are qualified to write queries in Power Query and T-SQL.

Existing Environment. Fabric -

Contoso has an F64 capacity named Cap1. All Fabric users are allowed to create items.

Contoso has two workspaces named WorkspaceA and WorkspaceB that currently use Pro license mode.

Existing Environment. Source Systems

Contoso has a point of sale (POS) system named POS1 that uses an instance of SQL Server on Azure Virtual Machines in the same Microsoft Entra tenant as Fabric. The host virtual machine is on a private virtual network that has public access blocked. POS1 contains all the sales transactions that were processed on the company's website.

The company has a software as a service (SaaS) online marketing app named MAR1. MAR1 has seven entities. The entities contain data that relates to email open rates and interaction rates, as well as website interactions. The data can be exported from MAR1 by calling REST APIs. Each entity has a different endpoint.

Contoso has been using MAR1 for one year. Data from prior years is stored in Parquet files in an Amazon Simple Storage Service (Amazon S3) bucket. There are 12 files that range in size from 300 MB to 900 MB and relate to email interactions.

Existing Environment. Product Data

POS1 contains a product list and related data. The data comes from the following three tables:

- Products
- ProductCategories
- ProductSubcategories

In the data, products are related to product subcategories, and subcategories are related to product categories.

Existing Environment. Azure -

Contoso has a Microsoft Entra tenant that has the following mail-enabled security groups:

- DataAnalysts: Contains the data analysts
- DataEngineers: Contains the data engineers

Contoso has an Azure subscription.

The company has an existing Azure DevOps organization and creates a new project for repositories that relate to Fabric.

Existing Environment. User Problems

The VP of marketing at Contoso requires analysis on the effectiveness of different types of email content. It typically takes a week to manually compile and analyze the data. Contoso wants to reduce the time to less than one day by using Fabric.

The data engineering team has successfully exported data from MAR1. The team experiences transient connectivity errors, which causes the data exports to fail.

Requirements. Planned Changes -

Contoso plans to create the following two lakehouses:

- Lakehouse1: Will store both raw and cleansed data from the sources
- Lakehouse2: Will serve data in a dimensional model to users for analytical queries

Additional items will be added to facilitate data ingestion and transformation.

Contoso plans to use Azure Repos for source control in Fabric.

Requirements. Technical Requirements

The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold. There will be extensive data cleansing required to populate the MAR1 data in the silver layer, including deduplication, the handling of missing values, and the standardizing of capitalization.

Each layer must be fully populated before moving on to the next layer. If any step in populating the lakehouses fails, an email must be sent to the data engineers.

Data imports must run simultaneously, when possible.

The use of email data from the Amazon S3 bucket must meet the following requirements:

- Minimize egress costs associated with cross-cloud data access.
- Prevent saving a copy of the raw data in the lakehouses.

Items that relate to data ingestion must meet the following requirements:

- The items must be source controlled alongside other workspace items.
- Ingested data must land in the bronze layer of Lakehouse1 in the Delta format.
- No changes other than changes to the file formats must be implemented before the data lands in the bronze layer.
- Development effort must be minimized and a built-in connection must be used to import the source data.
- In the event of a connectivity error, the ingestion processes must attempt the connection again.

Lakehouses, data pipelines, and notebooks must be stored in WorkspaceA. Semantic models, reports, and dataflows must be stored in WorkspaceB.

Once a week, old files that are no longer referenced by a Delta table log must be removed.

Requirements. Data Transformation

In the POS1 product data, ProductID values are unique. The product dimension in the gold layer must include only active products from product list. Active products are identified by an IsActive value of 1.

Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted from the product dimension in the gold layer.

Requirements. Data Security -

Security in Fabric must meet the following requirements:

- The data engineers must have read and write access to all the lakehouses, including the underlying files.
- The data analysts must only have read access to the Delta tables in the gold layer.
- The data analysts must NOT have access to the data in the bronze and silver layers.
- The data engineers must be able to commit changes to source control in WorkspaceA.

You need to ensure that the data engineers are notified if any step in populating the lakehouses fails. The solution must meet the technical requirements and minimize development effort. 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 identify the failure:

- A Fail activity
- An If condition activity
- An On failure dependency condition
- An On completion dependency condition

To send the notification:

- A Teams activity
- An Invoke pipeline activity
- An Office365Outlook activity

Answer Area

To identify the failure:

- A Fail activity
- An If condition activity
- An On failure dependency condition**
- An On completion dependency condition

To send the notification:

- A Teams activity
- An Invoke pipeline activity
- An Office365Outlook activity**

Question 118

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. KQL queryset
- B. Streaming dataset
- C. Notebook
- D. Dataflow Gen2

Answer: C

