



## **Microsoft**

### **Exam Questions DP-700**

Implementing Data Engineering Solutions Using Microsoft Fabric (beta)

## About ExamBible

### *Your Partner of IT Exam*

## Found in 1998

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

## Our Advances

### \* 99.9% Uptime

All examinations will be up to date.

### \* 24/7 Quality Support

We will provide service round the clock.

### \* 100% Pass Rate

Our guarantee that you will pass the exam.

### \* Unique Gurantee

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

**NEW QUESTION 1**

HOTSPOT - (Topic 1)

You need to create the product dimension.

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

NOTE: Each correct selection is worth one point.

```
SELECT ProductID, ProductNumber, ProductName, ModelName, SubCategoryName, CategoryName
FROM ContosoLake.Products p
    ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID
    ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID
WHERE
```

▼

- FULL JOIN
- INNER JOIN
- LEFT ANTI JOIN
- LEFT OUTER JOIN
- OUTER JOIN

▼

- FULL JOIN
- INNER JOIN
- LEFT ANTI JOIN
- LEFT OUTER JOIN
- OUTER JOIN

▼

- CategoryID = 1;
- CategoryName is not null;
- IsActive = 1;
- IsActive is not null;
- ProductNumber is not null;
- SubCategoryID = 1;
- SubCategoryName is not null;

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Join between Products and ProductSubCategories: Use an INNER JOIN.

The goal is to include only products that are assigned to a subcategory. An INNER JOIN ensures that only matching records (i.e., products with a valid subcategory) are included.

Join between ProductSubCategories and ProductCategories: Use an INNER JOIN.

Similar to the above logic, we want to include only subcategories assigned to a valid product category. An INNER JOIN ensures this condition is met.

WHERE Clause Condition: IsActive = 1

Only active products (where IsActive equals 1) should be included in the gold layer. This filters out inactive products.

**NEW QUESTION 2**

- (Topic 1)

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

What should you do?

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

**Answer:** B

**Explanation:**

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

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

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

**NEW QUESTION 3**

HOTSPOT - (Topic 1)

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.

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	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Bronze Layer: A pipeline Copy activity

The bronze layer is used to store raw, unprocessed data. The requirements specify that no transformations should be applied before landing the data in this layer. Using a pipeline Copy activity ensures minimal development effort, built-in connectors, and the ability to ingest the data directly into the Delta format in the bronze layer.

Silver Layer: A notebook

The silver layer involves extensive data cleansing (deduplication, handling missing values, and standardizing capitalization). A notebook provides the flexibility to implement complex transformations and is well-suited for this task.

**NEW QUESTION 4**

- (Topic 2)

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

What should you do?

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

**Answer:** E

**Explanation:**

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

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

**NEW QUESTION 5**

- (Topic 2)

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**

#### NEW QUESTION 6

- (Topic 3)

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:

An eventstream is the best solution for ingesting data from Azure Event Hub into Fabric, while applying filtering logic such as retaining only the events that have a Neighbourhood value of "Chelsea." Eventstreams in Microsoft Fabric are designed for handling real-time data streams and can apply transformation logic directly on incoming events. In this case, the eventstream can filter events based on the Neighbourhood field before storing the retained events in a Fabric lakehouse. Eventstreams are well-suited for stream processing, such as this case where you need to filter out only specific data (events with a Neighbourhood of "Chelsea") before storing it in the lakehouse.

#### NEW QUESTION 7

- (Topic 3)

You have a Fabric warehouse named DW1 that loads data by using a data pipeline named Pipeline1. Pipeline1 uses a Copy data activity with a dynamic SQL source. Pipeline1 is scheduled to run every 15 minutes.

You discover that Pipeline1 keeps failing.

You need to identify which SQL query was executed when the pipeline failed. What should you do?

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

**Answer: B**

#### Explanation:

The input JSON contains the configuration details and parameters passed to the Copy data activity during execution, including the dynamically generated SQL query.

Viewing the input JSON for the failed pipeline run provides direct insight into what query was executed at the time of failure.

#### NEW QUESTION 8

- (Topic 3)

You have a Fabric warehouse named DW1. DW1 contains a table that stores sales data and is used by multiple sales representatives.

You plan to implement row-level security (RLS).

You need to ensure that the sales representatives can see only their respective data. Which warehouse object do you require to implement RLS?

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

**Answer: D**

#### Explanation:

To implement Row-Level Security (RLS) in a Fabric warehouse, you need to use a function that defines the security logic for filtering the rows of data based on the user's identity or role. This function can be used in conjunction with a security policy to control access to specific rows in a table.

In the case of sales representatives, the function would define the filtering criteria (e.g., based on a column such as SalesRepID or SalesRepName), ensuring that each representative can only see their respective data.

#### NEW QUESTION 9

- (Topic 3)

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

In an external data source, you have data files that are 500 GB each. A new file is added every day.

You need to ingest the data into Lakehouse1 without applying any transformations. The solution must meet the following requirements

Trigger the process when a new file is added.

Provide the highest throughput.

Which type of item should you use to ingest the data?

- A. Event stream
- B. Dataflow Gen2
- C. Streaming dataset
- D. Data pipeline

**Answer:** A

**Explanation:**

To ingest large files (500 GB each) from an external data source into Lakehouse1 with high throughput and to trigger the process when a new file is added, an Eventstream is the best solution.

An Eventstream in Fabric is designed for handling real-time data streams and can efficiently ingest large files as soon as they are added to an external source. It is optimized for high throughput and can be configured to trigger upon detecting new files, allowing for fast and continuous ingestion of data with minimal delay.

**NEW QUESTION 10**

- (Topic 3)

You have a Fabric workspace named Workspacel that contains the following items:

- A Microsoft Power BI report named Reportl
- A Power BI dashboard named Dashboardl
- A semantic model named Modell
- A lakehouse name Lakehouse1

Your company requires that specific governance processes be implemented for the items. Which items can you endorse in Fabric?

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

**Answer:** B

**NEW QUESTION 10**

HOTSPOT - (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse! contains a table named Customer. Customer contains the following data.

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

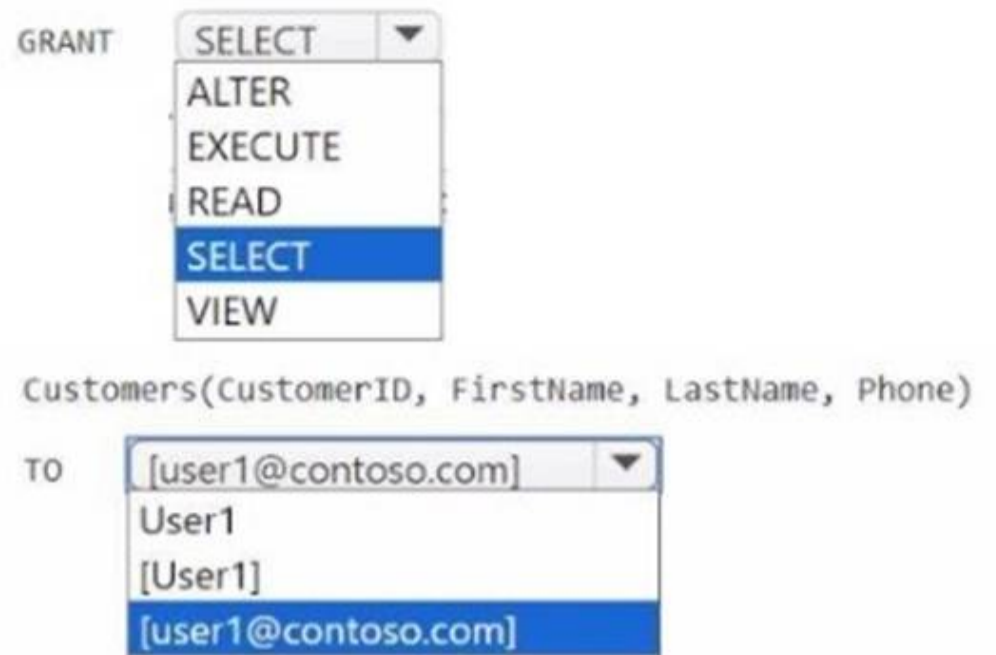
You have an internal Microsoft Entra user named User1 that has an email address of user1@contoso.com.

You need to provide User1 with access to the Customer table. The solution must prevent User1 from accessing the CreditCard column.

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

NOTE: Each correct selection is worth one point.

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area



**NEW QUESTION 15**

- (Topic 3)

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

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

Answer: A

Explanation:

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

**NEW QUESTION 17**

- (Topic 3)

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:** A

**Explanation:**

A lakehouse is the best option for storing semi-structured data when you need to read it using T-SQL, KQL, and Apache Spark. A lakehouse combines the flexibility of a data lake (which can handle semi-structured and unstructured data) with the performance features of a data warehouse. It allows data to be written using Apache Spark and can be queried using different technologies such as T-SQL (for SQL-based querying), KQL (Kusto Query Language for querying), and Apache Spark (for distributed processing). This solution is ideal when dealing with semi-structured data and requiring a versatile querying approach.

**NEW QUESTION 22**

- (Topic 3)

You have an Azure SQL database named DB1.

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

You discover that events are NOT being propagated to EventStreamDBI.

You need to ensure that the events are propagated to EventStreamDBI. What should you do?

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

**Answer:** D

**NEW QUESTION 25**

- (Topic 3)

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

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

You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

Both tables contain millions of rows.

You have the following KQL queryset.

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:**

An outer join will include unmatched rows from both tables, increasing the dataset size and processing time. It does not improve query performance.

**NEW QUESTION 28**

HOTSPOT - (Topic 3)

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.

**Answer Area**

Notebook1:

Notebook2:

Pipeline1:

Model1:

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

**Answer Area**

**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.

**NEW QUESTION 31**

HOTSPOT - (Topic 3)

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**  
 An eventstream

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

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Answer Area

Source1:    
 Apache Spark Structured Streaming  
 An eventstream  
 A data pipeline  
**A streaming dataflow**  
 An eventstream

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

**NEW QUESTION 32**

HOTSPOT - (Topic 3)

You plan to process the following three datasets by using Fabric:

- Dataset1: This dataset will be added to Fabric and will have a unique primary key between the source and the destination. The unique primary key will be an integer and will start from 1 and have an increment of 1.
- Dataset2: This dataset contains semi-structured data that uses bulk data transfer. The dataset must be handled in one process between the source and the destination. The data transformation process will include the use of custom visuals to understand and work with the dataset in development mode.
- Dataset3. This dataset is in a takehouse. The data will be bulk loaded. The data transformation process will include row-based windowing functions during the loading process.

You need to identify which type of item to use for the datasets. The solution must minimize development effort and use built-in functionality, when possible. What should you identify for each dataset? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

### Answer Area

Dataset1:	<input type="text" value="A T-SQL statement"/> A Dataflow Gen2 dataflow A notebook <b>A T-SQL statement</b>
Dataset2:	<input type="text" value="A notebook"/> A Dataflow Gen2 dataflow <b>A notebook</b> A T-SQL statement
Dataset3:	<input type="text" value="A KQL queryset"/> A Dataflow Gen2 dataflow <b>A KQL queryset</b> A T-SQL statement

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

### Answer Area

Dataset1:	<input type="text" value="A T-SQL statement"/> A Dataflow Gen2 dataflow A notebook <b>A T-SQL statement</b>
Dataset2:	<input type="text" value="A notebook"/> A Dataflow Gen2 dataflow <b>A notebook</b> A T-SQL statement
Dataset3:	<input type="text" value="A KQL queryset"/> A Dataflow Gen2 dataflow <b>A KQL queryset</b> A T-SQL statement

### NEW QUESTION 34

- (Topic 3)

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

In an external data source, you have data files that are 500 GB each. A new file is added every day.

You need to ingest the data into Lakehouse1 without applying any transformations. The solution must meet the following requirements

Trigger the process when a new file is added. Provide the highest throughput.

Which type of item should you use to ingest the data?

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

**Answer:** A

**Explanation:**

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

**NEW QUESTION 37**

HOTSPOT - (Topic 3)

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(
```

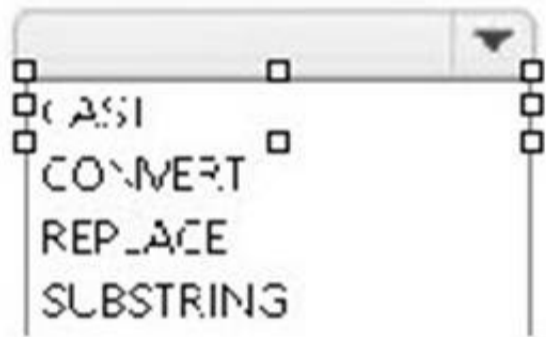
```
SELECT e.EmployeeID
```

```
, e.EmployeeName
```

```
, e.EmployeePosition
```

```
, c.ContractType
```

```
, (date, c.StartDate) as StartDate
```



```
FROM Employee AS e
```

```
Contract AS c on c.EmployeeID = e.EmployeeID
```



```
)
```

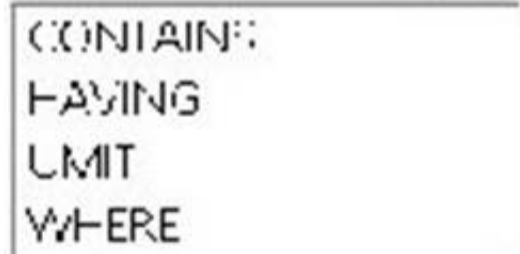
```
SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees
```

```
, ContractType
```

```
FROM result
```

```
GROUP BY ContractType
```

```
COUNT(DISTINCT EmployeeID) > 2
```



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

### Answer Area

WITH result AS(

SELECT e.EmployeeID

, e.EmployeeName

, e.EmployeePosition

, c.ContractType

, (date, c.StartDate) as StartDate



FROM Employee AS e



Contract AS c on c.EmployeeID = e.EmployeeID

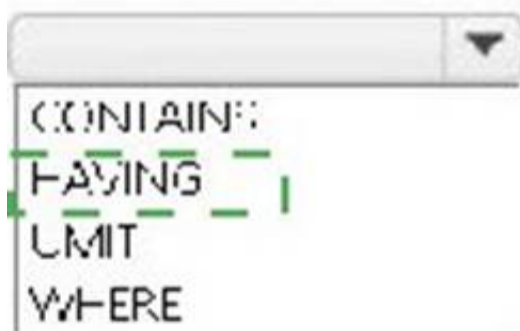
)

SELECT COUNT(DISTINCT EmployeeID) AS TotalEmployees

, ContractType

FROM result

GROUP BY ContractType



COUNT(DISTINCT EmployeeID) > 2

#### NEW QUESTION 39

- (Topic 3)

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:**

To identify in which workspace a specific item runs in Monitoring hub, you should view the Location column. This column indicates the workspace where the item is executed. Since you have multiple workspaces and need to track the execution of items across them, the Location column will show you the exact workspace associated with each item or job execution.

**NEW QUESTION 40**

- (Topic 3)

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

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

**Answer:** A

**Explanation:**

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

**NEW QUESTION 41**

DRAG DROP - (Topic 3)

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**

- Row\_cumsum
- Row\_rank\_dense
- Row\_rank\_min
- Row\_window\_session

**Answer Area**

**Statement1:**

```
TaxiData
| sort by VendorID asc
| extend RunningTotalAmount =  (total_amount, VendorID != prev(VendorID))
```

**Statement2:**

```
TaxiData
| sort by tpep_pickup_datetime asc, payment_type asc
| extend FirstPickupDateTime =  (tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))
```

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Partition the RunningTotalAmount by VendorID. - Row\_cumsum  
 The Row\_cumsum function computes the cumulative sum of a column while optionally restarting the accumulation based on a condition. In this case, it calculates the cumulative sum of total\_amount for each VendorID, restarting when the VendorID changes (VendorID != prev(VendorID)).

```
TaxiData
| sort by VendorID asc
| extend RunningTotalAmount = Row_cumsum(total_amount, VendorID != prev(VendorID))
```

Create a column FirstPickupDateTime that shows the first value of each hour from tpep\_pickup\_datetime, partitioned by payment\_type - Row\_window\_session

```
TaxiData
| sort by tpep_pickup_datetime asc, payment_type asc
| extend FirstPickupDateTime = Row_window_session(tpep_pickup_datetime, 1h, 0m, payment_type != prev(payment_type))
```

**NEW QUESTION 45**

- (Topic 3)

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:**

VACUUM is used to clean up storage by removing files no longer in use by a Delta table. It removes old and unreferenced files from Delta tables. For example, to remove files older than 7 days:

```
VACUUM delta.`/path_to_table` RETAIN 7 HOURS;
```

**NEW QUESTION 49**

DRAG DROP - (Topic 3)

You have a Fabric workspace that contains a warehouse named Warehouse1.

In Warehouse1, you create a table named DimCustomer by running the following statement.

```
CREATE TABLE dbo.DimCustomer (
    CustomerKey VARCHAR(255) NOT NULL,
    Name VARCHAR(255) NOT NULL,
    Email VARCHAR(255) NOT NULL
);
```

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

**Code Segments**

- 0 DROP CONSTRAINT PK\_DimCustomer
- 0 ADD CONSTRAINT PK\_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)
- 0 NOT ENFORCED
- 0 ALTER TABLE dbo.DimCustomer
- 0 ADD CONSTRAINT PK\_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)
- 0 ENFORCED

**Answer Area**

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

### Code Segments

```

0 DROP CONSTRAINT PK_DimCustomer
0
0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY NONCLUSTERED
  (CustomerKey)
0
0 NOT ENFORCED
0
0 ALTER TABLE dbo.DimCustomer
0
0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
  (CustomerKey)
0
0 ENFORCED

```

### Answer Area

```

0
0 ALTER TABLE dbo.DimCustomer
0
0 ADD CONSTRAINT PK_DimCustomer PRIMARY KEY CLUSTERED
  (CustomerKey)
0
0 ENFORCED

```

### NEW QUESTION 51

- (Topic 3)

You have an Azure key vault named KeyVault1 that contains secrets.

You have a Fabric workspace named Workspace!. 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 key vault URL and the name of a linked service.

Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

### NEW QUESTION 55

- (Topic 3)

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: ADE**

### Explanation:

To meet the requirements of ensuring that the data engineer can write data to the Customer table without reading data from the Employee table (which contains Personally Identifiable Information, or PII), you can implement the following steps:

? Share Lakehouse1 with the data engineer.

By sharing Lakehouse1 with the data engineer, you provide the necessary access to the data within the lakehouse. However, this access should be controlled through roles and permissions, which will allow writing to the Customer table but prevent reading from the Employee table.

? Assign the data engineer the Contributor role for Workspace1.

Assigning the Contributor role for Workspace1 grants the data engineer the ability to perform actions such as writing to tables (e.g., the Customer table) within the workspace. This role typically allows users to modify and manage data without necessarily granting them access to view all data (e.g., PII data in the Employee table).

? Migrate the Employee table from Lakehouse1 to Lakehouse2.

To prevent the data engineer from accessing the Employee table (which contains PII), you can migrate the Employee table to a separate lakehouse (Lakehouse2) or workspace

(Workspace2). This separation of sensitive data ensures that the data engineer's access is restricted to the Customer table in Lakehouse1, while the Employee table can be managed separately and protected under different access controls.

### NEW QUESTION 57

HOTSPOT - (Topic 3)

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

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

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

### Answer Area

To find the error details:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

To find the total number of errors:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

### Answer Area

To find the error details:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

To find the total number of errors:

	▼
Data insights	
Data preview	
Details	
Runtime logs	

#### NEW QUESTION 59

- (Topic 3)

You are developing a data pipeline named Pipeline1.

You need to add a Copy data activity that will copy data from a Snowflake data source to a Fabric warehouse.

What should you configure?

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

**Answer:** C

**Explanation:**

When using the Copy data activity in a data pipeline to move data from Snowflake to a Fabric warehouse, the process often involves intermediate staging to handle data efficiently, especially for large datasets or cross-cloud data transfers.

Staging involves temporarily storing data in an intermediate location (e.g., Blob storage or Azure Data Lake) before loading it into the target destination.

For cross-cloud data transfers (e.g., from Snowflake to Fabric), enabling staging ensures data is processed and stored temporarily in an efficient format for transfer.

Staging is especially useful when dealing with large datasets, ensuring the process is optimized and avoids memory limitations.

**NEW QUESTION 63**

- (Topic 3)

You have a Google Cloud Storage (GCS) container named storage1 that contains the files shown in the following table.

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

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

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

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

- A. Stores only
- B. Products only
- C. Stores and Products only
- D. Products, Stores, and Trips
- E. Trips only
- F. Products and Trips only

**Answer:** C

**Explanation:**

When reading data from shortcuts in Fabric (in this case, from a lakehouse like Lakehouse1), the cache for shortcuts helps by storing the data locally for quick access. The last accessed timestamp and the cache expiration rules determine whether data is fetched from the cache or from the source (Google Cloud Storage, in this case).

Products: The ProductFile.parquet was last accessed 12 hours ago. Since the cache has data available for up to 12 hours, it is likely that this data will be retrieved from the cache, as it hasn't been too long since it was last accessed.

Stores: The StoreFile.json was last accessed 4 hours ago, which is within the cache retention period. Therefore, this data will also be retrieved from the cache.

Trips: The TripsFile.csv was last accessed 48 hours ago. Given that it's outside the typical caching window (assuming the cache has a maximum retention period of around 24 hours), it would not be retrieved from the cache. Instead, it will likely require a fresh read from the source.

**NEW QUESTION 64**

- (Topic 3)

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

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

You have a KQL database that contains two tables named Stream and Reference. Stream contains streaming data in the following format.

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

Reference contains reference data in the following format.

Column name	Data type
DeviceId	Int
DeviceName	String

Both tables contain millions of rows. You have the following KQL queryset.

You need to reduce how long it takes to run the KQL queryset. Solution: You add the make\_list() function to the output columns. Does this meet the goal?

01 Stream

02 | extend lat = todecimal(GeoLocation.Latitude), long = todecimal(GeoLocation.Longitude)

03 | join kind=inner Reference on DeviceId

04 | project Timestamp, lat, long, Temperature, DeviceName

05 | filter Temperature >= 10

06 | render scatterchart with (kind = map)

- A. Yes
- B. No

**Answer:** B

**Explanation:**

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

**NEW QUESTION 69**

- (Topic 3)

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:**

Problem Overview:

Table1 has 2,000 small Parquet files (1 MB each).

Query performance suffers when the table contains numerous small files because the query engine must process each file individually, leading to significant overhead.

Solution:

To improve performance, file compaction is necessary to reduce the number of small files and create larger, optimized files.

Commands and Their Roles: OPTIMIZE Command:

- Compacts small Parquet files into larger files to improve query performance.
- It supports optional features like V-Order, which organizes data for efficient scanning.
- VACUUM Command:
  - Removes old, unreferenced data files and metadata from the Delta table.
  - Running VACUUM after OPTIMIZE ensures unnecessary files are cleaned up, reducing storage overhead and improving performance.

**NEW QUESTION 73**

DRAG DROP - (Topic 3)

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 extract, transform, and load (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**

- 0  Change the runtime version.
- 0  Install the libraries.
- 0  Create a pool.
- 0  Create an environment.
- 0  Set the default environment.

**Answer Area**

- 0
- 0
- 0

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

**Actions**

- 0  Change the runtime version.
- 0  Install the libraries.
- 0  Create a pool.
- 0  Create an environment.
- 0  Set the default environment.

**Answer Area**

- 0  Create an environment.
- 0  Install the libraries.
- 0  Set the default environment.

**NEW QUESTION 75**

- (Topic 3)

You have two Fabric workspaces named Workspace1 and Workspace2.

You have a Fabric deployment pipeline named deployPipeline1 that deploys items from Workspace1 to Workspace2. DeployPipeline1 contains all the items in Workspace1.

You recently modified the items in Workspaces1.

The workspaces currently contain the items shown in the following table.

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

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

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

**Answer: D**

**Explanation:**

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

**NEW QUESTION 80**

- (Topic 3)

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:**

To consolidate the underlying Parquet files in Table1 and improve query performance by optimizing the data layout, you should use the OPTIMIZE command in Delta Lake. The OPTIMIZE command coalesces smaller files into larger ones and reorganizes the data for more efficient reads. This is particularly useful when working with large datasets in Delta tables, as it helps reduce the number of files and improves performance for subsequent queries or operations like MERGE.

**NEW QUESTION 85**

- (Topic 3)

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(). parameter

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

**Answer:** B

#### NEW QUESTION 89

- (Topic 3)

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. Product
- F. Stores, and Trips

**Answer:** B

#### Explanation:

When the cache for shortcuts is enabled in Fabric, the data retrieval is governed by the caching behavior, which generally retains data for a specific period after it was last accessed. The data from the shortcuts will be retrieved from the cache if the data is stored in locations that support caching. Here's a breakdown based on the data's location: Products: The ProductFile is stored in Azure Data Lake Storage Gen2 (storage1). Since Azure Data Lake is a supported storage system in Fabric and the file is relatively small (50 MB), this data is most likely cached and can be retrieved from the cache.

Stores: The StoreFile is stored in Amazon S3 (storage2), and even though it is stored in a different cloud provider, Fabric can cache data from Amazon S3 if caching is enabled. This data (25 MB) is likely cached and retrievable.

Trips: The TripsFile is stored in Amazon S3 (storage2) and is significantly larger (2 GB) compared to the other files. While Fabric can cache data from Amazon S3, the larger size of the file (2 GB) may exceed typical cache sizes or retention windows, causing this file to likely be retrieved directly from the source instead of the cache.

#### NEW QUESTION 93

- (Topic 3)

You have a Fabric workspace that contains an eventstream named Eventstream1. Eventstream1 processes data from a thermal sensor by using event stream processing, and then stores the data in a lakehouse.

You need to modify Eventstream1 to include the standard deviation of the temperature. Which transform operator should you include in the Eventstream1 logic?

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

**Answer:** D

#### Explanation:

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

#### NEW QUESTION 98

.....

## Relate Links

**100% Pass Your DP-700 Exam with ExamBible Prep Materials**

<https://www.exambible.com/DP-700-exam/>

## Contact us

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

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