ATMIC NETWORKS
YOUR GUIDE TO GLOBAL CERTIFIATION

"Please note that these files may not be up to date. However, the questions will help you understand the exam format and typical question patterns."

www.atmicnetworks .com

Warning: Keep connected with our support team for latest updates

Topic 1, Contoso, Ltd

Case Study

Overview

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: 1

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:

Data Analysts' Access Requirements must only have read access to the Delta tables in the gold layer and not have access to the bronze and silver layers.

The gold layer data is typically queried via SQL Endpoints. Granting the Read all SQL Endpoint data permission allows data analysts to query the data using familiar SQL-based tools while restricting access to the underlying files.

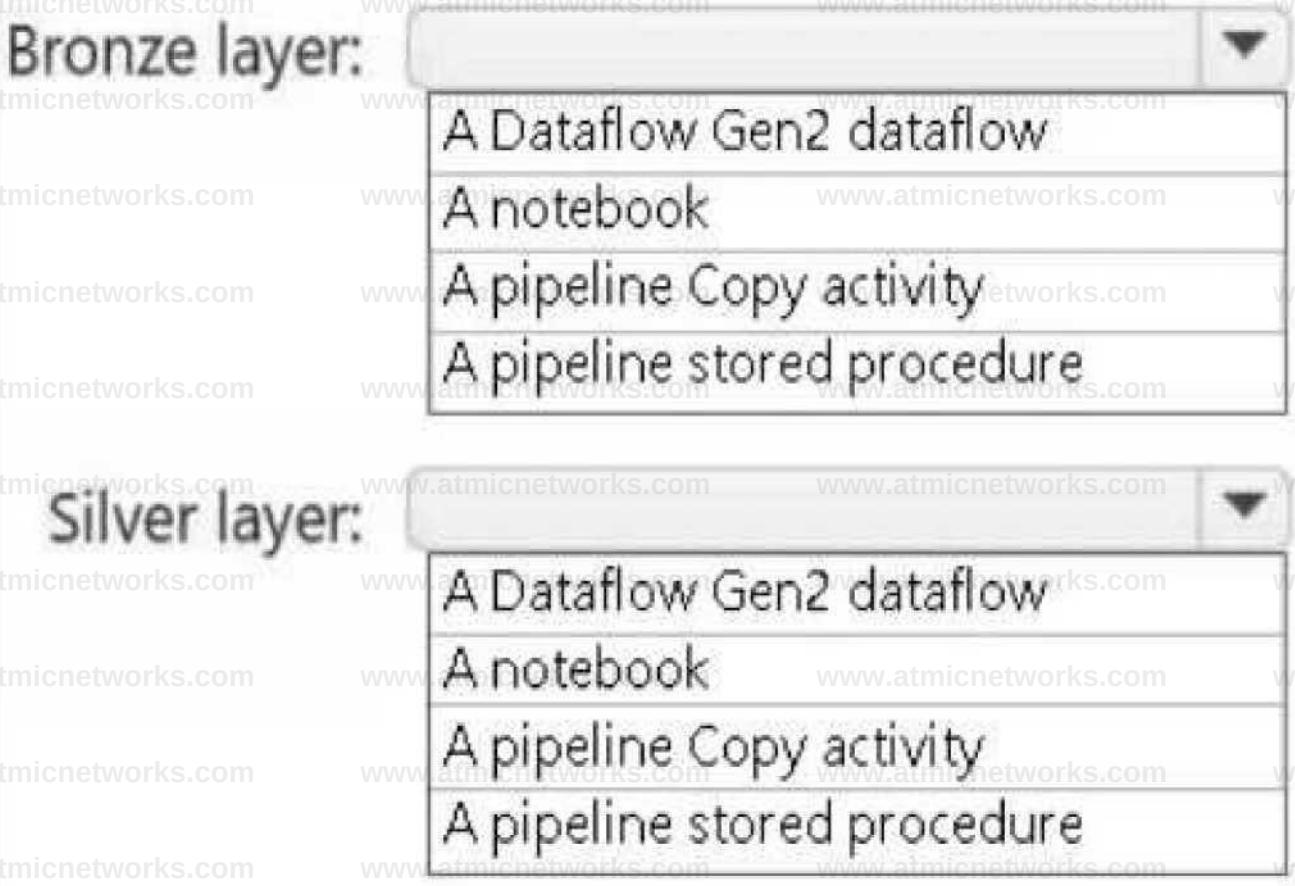
Question: 2

HOTSPOT

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

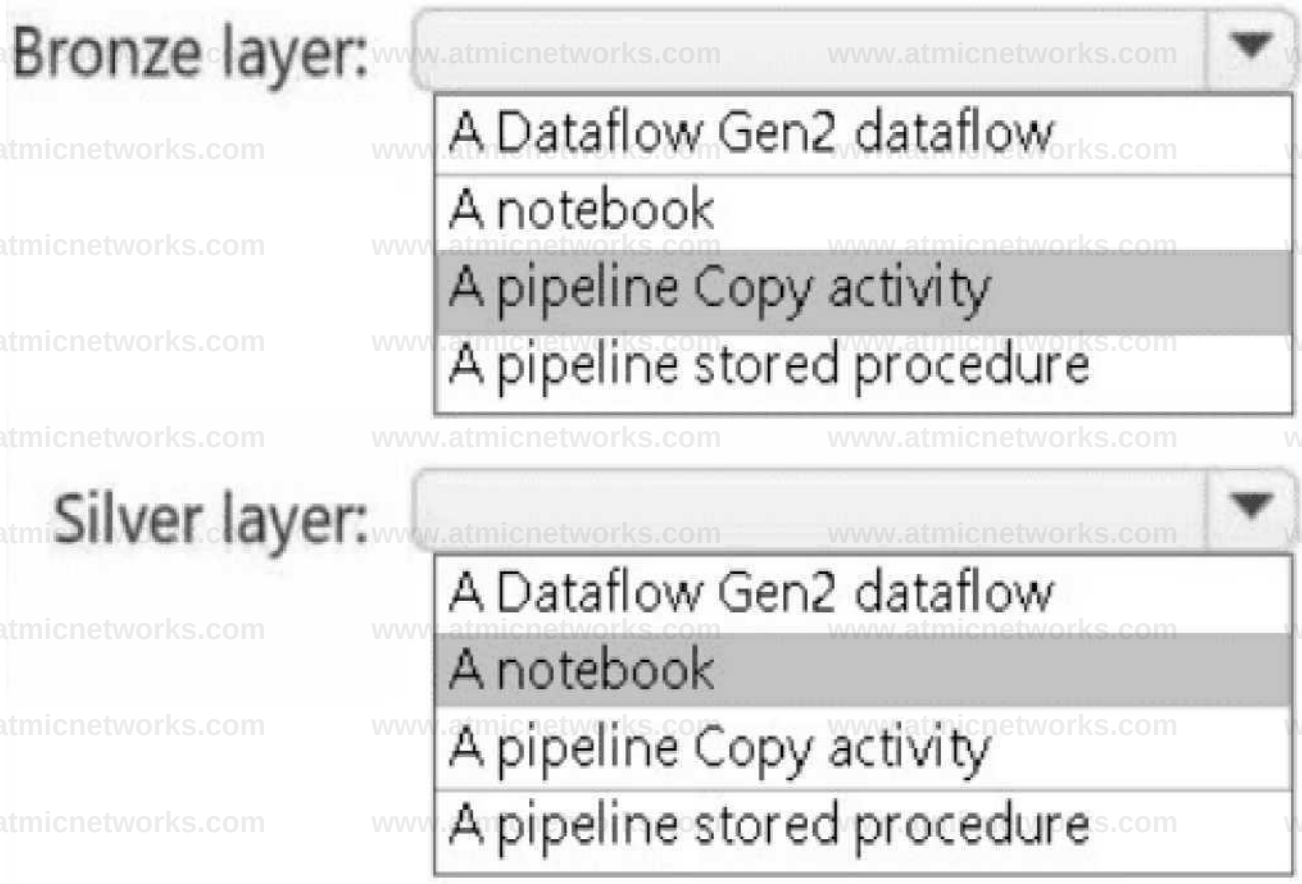
What should you recommend for each layer? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:

Explanation:



A screenshot of a computer Description automatically generated

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.

Question: 3

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.

Question: 4

HOTSPOT

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, Categorykame
FROM ContosoLake.Products p

  T ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID
  FULL JOIN
  TNNERJOTN
  LEFT ANTI JOIN
  LEFT OUTER JOIN
  OUTER JOIN
  VI ContosoLake.Productcategories c ON c. CategoryID = s.CategorylD
  FULL JOIN
  INNER JOIN
  LEFT ANTI JOIN
  LEFT OUTER JOIN
  OUTER JOIN

WHERE
        CategoryID = 1;
        CategoryName is not null;
        IsActive = 1;
        isActive is not null;
        ProductNumber is not null;
        SubCategorylD = 1;
        SubCategcryName is not null;
```

Answer:

Explanation:

```
SELECT ProductID, ProductNumber^ ProductName, ModelName, SubCategoryNamej CategoryName

FROM Contosolake.Products p
                        Contosolake.ProductSubCategories s      ON p.SubCategory10 = s.SubCategoryID
  FULL JOIN
   INNER JOIN
   LEFT ANTI JOIN
   LEFT OUTER JOIN
   OUTER JOIN

                        Contosolake.Productcategories c ON      C.CategorylD = s.CategorylD
   FULL JOIN
   INNER JOIN
   LEFT ANTI JOIN
   LEFT OUTER JOIN
   OUTER JOIN

WHERE                                   ▼
          CategorylD = 1;
          CategoryName is not null;
          IsActive = 1;
          IsActive is not null;
          ProductNumber is not null;
          SubCategorylD = 1;
          SubCategoryName is not null;
```

A screenshot of a computer Description automatically generated

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.

Question: 5

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: A,B

Explanation:

MAR1 has seven entities, each accessible via a different API endpoint. A ForEach activity is required to iterate over these endpoints to fetch data from each one. It enables dynamic execution of API calls for each entity.

The Copy data activity is the primary mechanism to extract data from REST APIs and load it into the bronze layer in Delta format. It supports native connectors for REST APIs and Delta, minimizing development effort.

Question: 6

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:

The technical requirements specify that:
Medallion layers must be fully populated sequentially (bronze → silver → gold). Each layer must be populated before the next.
If any step fails, the process must notify the data engineers.
Data imports should run simultaneously when possible.

Why Use a Data Pipeline That Calls Other Data Pipelines?

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

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

A parent pipeline can handle:

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

Question: 7

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. Configure Fault tolerance for the Copy data activity.
D. Call a notebook from the data pipeline.

Answer: B

Explanation:

Question: 8

HOTSPOT

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 completion dependency condition

To send the notification:
- A Teams activity
- An Invoke pipeline activity
- An Office3650ut look activity

Answer:

Explanation:

Answer Area

To identify the failure: An On failure dependency condition

To send the notification: An Office36SOutlook activity

Question: 9

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

Answer: B

Explanation:

Question: 10

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. Assign WorkspaceA to Capl.

B. From Tenant setting, set Users can synchronize workspace items with their Git repositories to Enabled

C. Configure WorkspaceA to use a Premium Per User (PPU) license

D. From Tenant setting, set Users can sync workspace items with GitHub repositories to Enabled Answer: A, B

Explanation:

Topic 2, Litware, Inc

Case Study

Overview

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: 11

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:

The requirement specifies that Litware plans to make the book reviews available in the lakehouse without making a copy of the data. In this case, creating a shortcut in Fabric is the most appropriate solution. A shortcut is a reference to the external data, and it allows Litware to access the book reviews stored in Amazon S3 without duplicating the data

into the lakehouse.

Question: 12

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 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:

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.

Question: 13

HOTSPOT

HOTSPOT

You need to troubleshoot the ad-hoc query issue.

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

NOTE: Each correct selection is worth one point.

```
SELECT last_run_start_time, last_run_command

FROM _____▼
        queryi nsi ghts.exec ^requests _hi story queryi nsi
        ghts.exec_sessions_hi story
        queryinsights.frequently_run_queries queryi nsi
        ghts.long running queri es

WHERE last_run_total_elapsed_time_ms > 7200000

AND f_____v |
```

max_run total elapsed_time_ms > 72 00000 median_total_elapsed_time_ms > 7200000 number_o f_c anceI ed^runs > 1 number_of_failed_runs > 1 number_of_runs > 1

Answer:

Explanation:

Answer Area

SELECT lastrunstarttimej 1astrun command

FROM queryinsights.exec_requests_history queryinsights.exec_sessions_history queryinsightsdrequently_run_queries queryinsightsJong_running_queries

WHERE last_run_total_elaosec_time_ms > 7206008

AND _____

max_run_total_elapsed_time_ms > 72 00000
median_total_elapsed_time_ms > 72 00000
number_of_canceled_runs > 1
number_of_failed_runs > 1
number of runs > 1

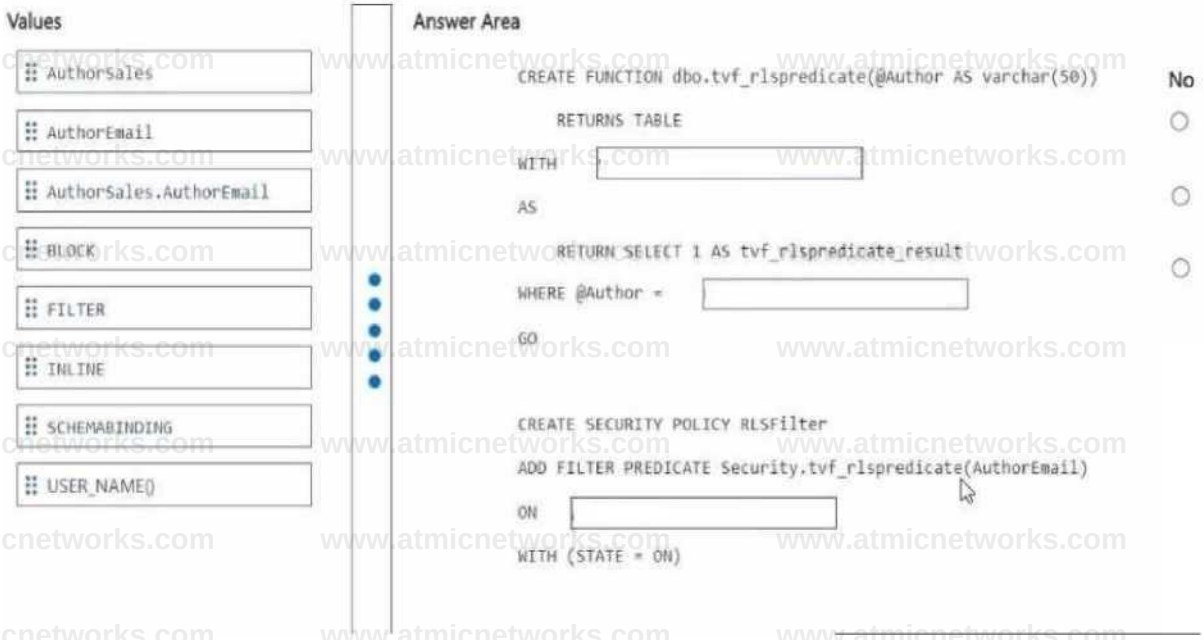
A screenshot of a computer Description automatically generated

SELECT last_run_start_time, last_run_command: These fields will help identify the execution details of the long-running queries.
FROM queryinsights.long_running_queries: The correct solution is to check the long-running queries using the queryinsights.long_running_queries view, which provides insights into queries that take longer than expected to execute.

WHERE last_run_total_elapsed_time_ms > 7200000: This condition filters queries that took more than 2 hours to complete (7200000 milliseconds), which is relevant to the issue described.
AND number_of_failed_runs > 1: This condition is key for identifying queries that have failed more than once, helping to isolate the problematic queries that cause failures and need attention.

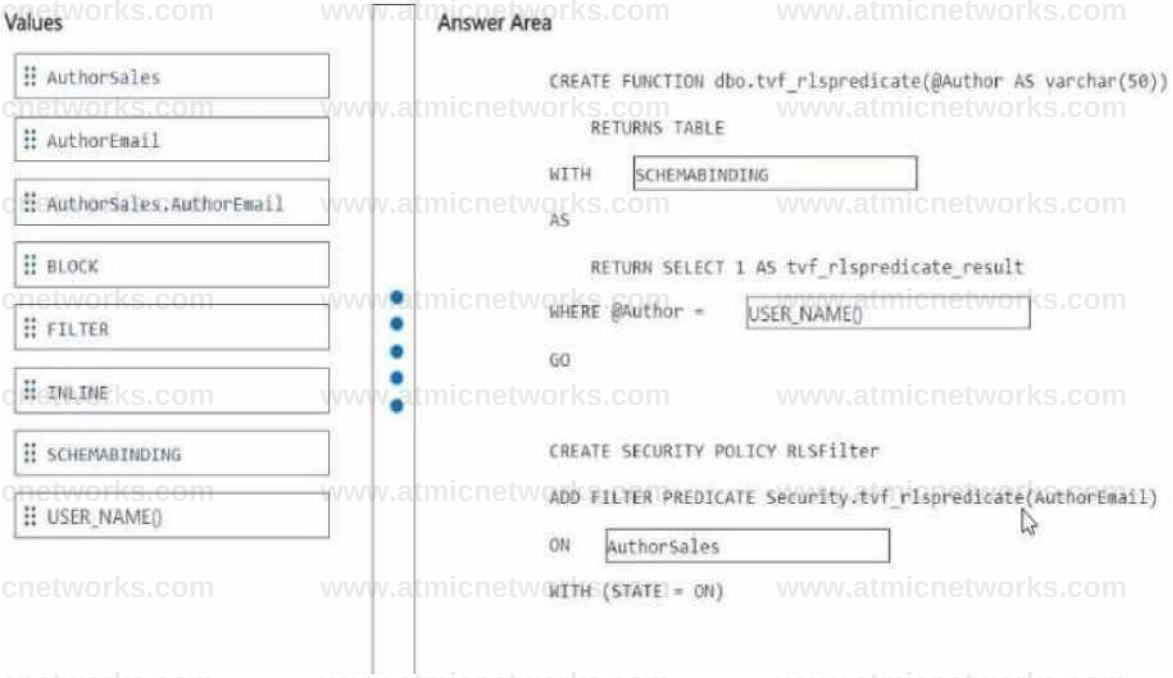
Question: 14
DRAG DROP
You need to ensure that the authors can see only their respective sales data.
How should you complete the statement? To answer, drag the appropriate values the correct targets.
Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content

NOTE: Each correct selection is worth one point.



Answer:

Explanation:



Question: 15

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:

Question: 16

You need to ensure that processes for the bronze and silver layers run in isolation How should you configure the Apache

Spark settings?

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

Answer: C

Explanation:

Question: 17

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

Answer: E, F

Explanation:

Topic 3, Misc. Questions Set

Question: 18

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 semistructured 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.

Question: 19

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:

To copy data from an on-premises Microsoft SQL Server database (Database1) to a warehouse (Warehouse1) in Microsoft Fabric, the best option is to use a data pipeline. A data pipeline in Fabric allows for the orchestration of data movement, from source to destination, using connectors, transformations, and scheduled workflows. Since the data is being transferred from an on-premises database and requires the use of a data gateway, a data pipeline provides the appropriate framework to facilitate this data movement efficiently and reliably.

Question: 20

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

You have an on-premises Microsoft SQL Server database named Database1 that is accessed by using an on-premises data gateway.

You need to copy data from Database1 to Warehouse1.

Which item should you use?

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

Answer: B

Explanation:

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

Question: 21

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

C. Modify the surrogate keys to use a different data type.

D. Create views.

Answer: C

Explanation:

In this case, the key issue causing performance degradation likely stems from the use of MD5 hash surrogate keys. MD5 hashes are 128-bit values, which can be inefficient for large datasets like the 500 million rows in your fact table. Using a more efficient data type for surrogate keys (such as integer or bigint) would reduce the storage and processing overhead, leading to better query performance. This approach will improve performance while minimizing operational costs because it reduces the complexity of querying and indexing, as smaller data types are generally faster and more efficient to process.

Question: 22

HOTSPOT

HOTSPOT

You have a Fabric workspace that contains a warehouse named DW1. DW1 contains the following tables and columns.

Table name	Column name	Description
SalesOrderDetail	Product! D	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

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.

Answer Area

▼ (SO.HodifiedDate) AS OrderDate

SELECT CAST

SELECT CCNVER"

SELECT YEAR

jP.Hdiie AS PruductNsne

,SiM(S(i. nrrlerOtyi AS nrderQty

FROM dboJ-LSalp^rciPr^taHJ so

IMO 30IH [dbo].[Product] P

UH P.PnoductLD - SU.ProductIV

GROUP BY

CL BE;YEAR(SC ,M O C i fi C C D ate). P. Na m c)
(RO. JP N(i S S H?l Aj >(SO.S1cdifMdl :atei$_f$ P N^me). ('l AH O() Mod led)*-*)))
RO LL _IP ;YEAR(SC ,M DC itiec D ate), P.Mam e)
'■EAR I SO Mod 'iedOa.^ P.Manic

ORDER BY orderoate

Answer:

Explanation:

Answer Area

▼ (so.ModifiedDate) AS order-Date

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

Question: 23

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 □f each product
ProductName	Contains a unique name for each product
SalesAmount	Contains the sales amount of a transaction

You plan to load the data into a dimensional model and implement a star schema. From the original flat table, you create two tables named FactSales and DimProduct. You will track changes in DimProduct.

You need to prepare the data.

Which three columns should you include in the DimProduct table? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

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

Answer: B,C,F

Explanation:

In a star schema, the DimProduct table serves as a dimension table that contains descriptive attributes about products. It will provide context for the FactSales table, which contains transactional data. The following columns should be included in the DimProduct table:

ProductName: The ProductName is an important descriptive attribute of the product, which is needed for analysis and reporting in a dimensional model.

ProductColor: ProductColor is another descriptive attribute of the product. In a star schema, it makes sense to include attributes like color in the dimension table to help categorize products in the analysis.

ProductID: ProductID is the primary key for the DimProduct table, which will be used to join the FactSales table to the product dimension. It's essential for uniquely identifying each product in the model.

Question: 24

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.

Question: 25

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: A,D,E 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.

Question: 26

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

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

Answer: D

Explanation:

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

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

Question: 27

HOTSPOT

HOTSPOT

You have a Fabric workspace named Workspace1_DEV that contains the following items:

10 reports
Four notebooks
Three lakehouses
Two data pipelines
Two Dataflow Gen1 dataflows
Three Dataflow Gen2 dataflows
Five semantic models that each has a scheduled refresh policy

You create a deployment pipeline named Pipeline1 to move items from Workspace1_DEV to a new workspace named Workspace1_TEST.

You deploy all the items from Workspace1_DEV to Workspace1_TEST.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

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

Answer:

Explanation:

Answer Area

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

Question: 28

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: C

Explanation:

A deployment pipeline in Fabric is designed to automate the process of deploying assets (such as reports, datasets, eventhouses, and other objects) between environments like Dev, Test, and Prod.

Since you need to deploy an eventhouse as part of the deployment process, a deployment pipeline is the appropriate tool to move this asset through the different stages of your environment.

Question: 29

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

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

Answer: C

Explanation:

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

Question: 30

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: E

Explanation:

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

Question: 31

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.

Question: 32

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.

Question: 33

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:

To allow Job1 in Workspace1 to access an Azure SQL database (Source1) with public internet access disabled, you need to create a managed private endpoint. A managed private endpoint is a secure, private connection that enables services like Fabric (or other Azure services) to access resources such as databases, storage accounts, or other services within a virtual network (VNet) without requiring public internet access. This approach maintains the security and integrity of your data while enabling access to the Azure SQL database.

Question: 34

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	storage 1	50 MB	Contains a list of products and their details
TripsFile	storage2	2 GB	Contains one month's worth of taxi trip data
Store File	storage?	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:

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.

Question: 35

HOTSPOT

You have a Fabric workspace named Workspace1 that contains the items shown in the following table.

Name	Type |
Notebook!	Notebook
Notebooks	Notebook
Lakehouse 1	Lakehouse
Pipeline!	Data pipeline
Modell	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

Notebook!:

Acid Notebook! to an Apache Spark job definition.
Add Notebook! to P pel i ne!.
From Peal-Time bub. configu-e the execution of Notebook!

Notebook?:

Add Notebook? to an Apache Spcik job definition.
Add Notebook? Lo P pdine!.
From ^al-Time bub. configu'e the execution of Notebook?

Pipeline!:

Add P pel i ne! :c an Apache Spark joo definition. Configure the execution of Pi pel ne1 by using a schedule From Peal-Time bub. configu-e the execution cf Pipeline'.

Modell:

Add Weed! lo Pipel r e1
From ^eal-Time hub. configu'e Modell to refresh.
Set Keen your D rect La<e data up to cate to On.

Answer:

Explanation:

Answer Area

Notebook!:

- Add Notebook! to an Apache Spark job definition.
- Add Notebook! to Pipeline!.
- From Real-Time hub, configure the execution of Notebook!.

Notebooks

- Add Notebook2 to an Apache Spark job definition.
- Add Notebook2 to Pipeline!.
- From Real-Time hub, configure the execution of Notebook2.

Pipeline!:

- Add Pipeline! to an Apache Spark job definition.
- Configure the execution of Pipeline! by using a schedule.
- From Real-Time hub, configure the execution of Pipeline!.

Model!:

- Add Model! to Pipeline!.
- From Real-Time hub, configure Model! to refresh.
- Set Keep your Direct Lake data up to date to On.

Question: 36

Your company has a sales department that uses two Fabric workspaces named Workspace1 and Workspace2.

The company decides to implement a domain strategy to organize the workspaces.

You need to ensure that a user can perform the following tasks:

Create a new domain for the sales department.

Create two subdomains: one for the east region and one for the west region.

Assign Workspace1 to the east region subdomain.

Assign Workspace2 to the west region subdomain.

The solution must follow the principle of least privilege.

Which role should you assign to the user?

A. workspace Admin

B. domain admin

C. domain contributor

D. Fabric admin

Answer: B

Explanation:

To implement a domain strategy and manage subdomains within Fabric, the domain admin role is the appropriate role for the user. A domain admin has the permissions necessary to: Create a new domain (for the sales department).
Create subdomains (for the east and west regions).
Assign workspaces (such as Workspace1 and Workspace2) to the appropriate subdomains.

The domain admin role allows for managing the structure and organization of workspaces in the context of domains and subdomains while maintaining the principle of least privilege by limiting the user's access to managing the domain structure specifically.

Question: 37

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

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

Answer: D

Explanation:

To ensure User3 can view all items in Workspace1 and update the tables in DW1, the most appropriate workspace role to assign is the Contributor role. This role allows User3 to:
View all items in Workspace1: The Contributor role provides the ability to view all objects within the workspace, such as data pipelines, warehouses, and other resources.
Update the tables in DW1: The Contributor role allows User3 to modify or update resources within the workspace, including the tables in DW1, assuming that appropriate object-level permissions are set for the warehouse.

This role adheres to the principle of least privilege, as it provides the necessary permissions without granting broader administrative rights.

Question: 38

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: B

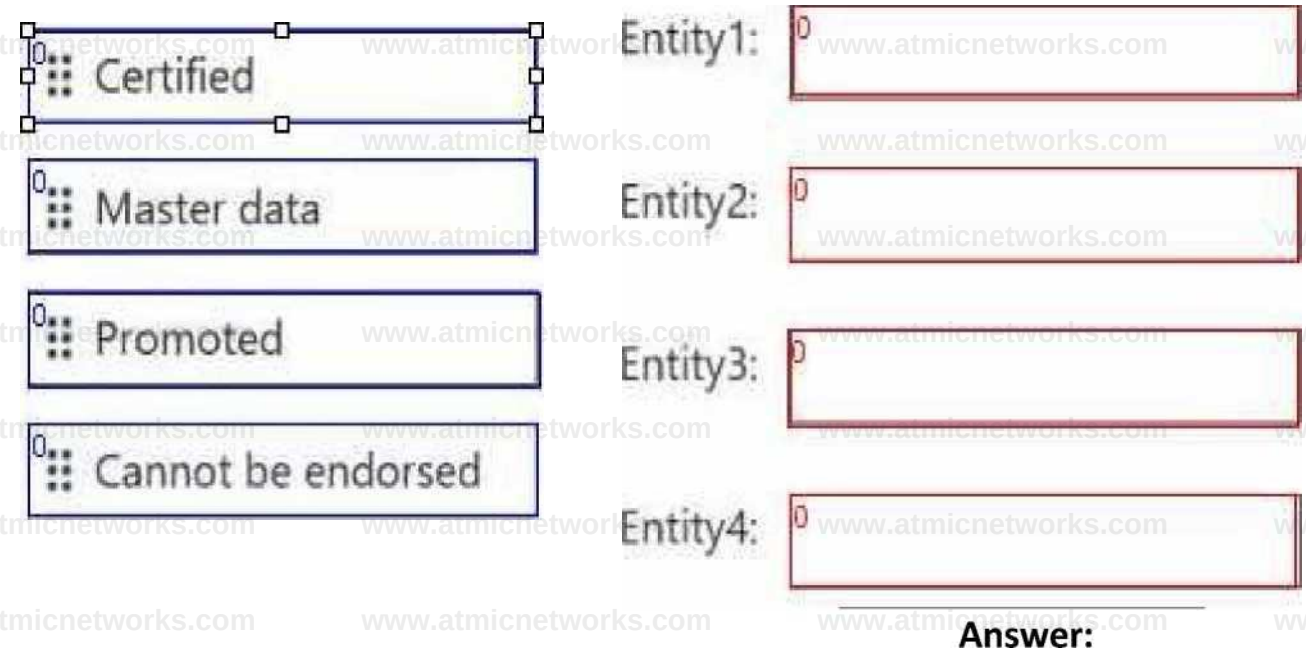
Explanation:

To meet the specified requirements for User1, the solution must ensure:
Read access to the table data in Lakehouse1: User1 needs permission to access the data within Lakehouse1. By sharing Lakehouse1 with User1 and selecting the Read all SQL endpoint data option, User1 will be able to query the data via SQL endpoints.
Prevent Apache Spark usage: By sharing the lakehouse directly and selecting the SQL endpoint data option, you specifically enable SQL-based access to the data, preventing User1 from using Apache Spark to query the data.
Prevent access to other items in Workspace1: Assigning User1 the Viewer role for Workspace1 ensures that User1 can only view the shared items (in this case, Lakehouse1), without accessing other resources such as notebooks, pipelines, or Power BI reports within Workspace1.

This approach provides the appropriate level of access while restricting User1 to only the required resources and preventing access to other workspace assets.

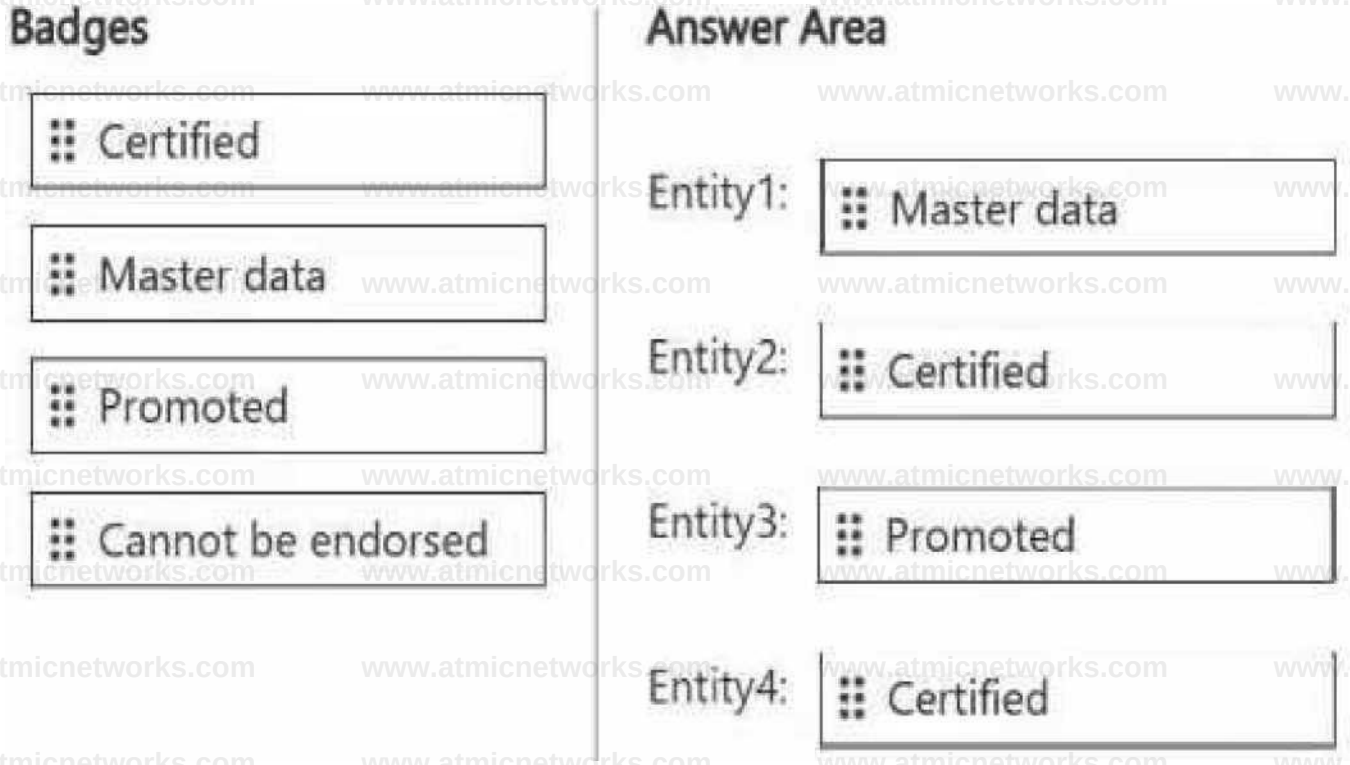
Question: 39

DRAG DROP
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 the correct entities. Each badge 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.



Answer:

Explanation:



Question: 40

HOTSPOT

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
Domainl	Useri
Domain2	UserZ

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.

Answer Area

Statements	Yes	No
User3 has Viewer role access to Workspaces*	☐	☐
User3 has Domain contributor access to Domainl.	☐	☐
User2 has Contributor role access to Workspaces.	☐	☐

Answer:

Explanation:

Answer Area

Statements	Yes	No
User3 has Viewer role access to Workspace3.	○	○
User3 has Domain contributor access to Domain1.	○	○
User2 has Contributor role access to Workspace3.	○	○

Question: 41

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
Workspace!	Modell Notebook! Report! Lakehouse 1 Pipeline!
Workspace?	Modell Notebook? Report! Lakehouse?

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.

Question: 42

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:

When you restructure Workspace1 by adding a new folder (Folder1) and moving Pipeline1 into it, deployPipeline1 will deploy the entire structure of Workspace1 to Workspace2, preserving the changes made in Workspace1. This includes:
Folder1 will be created in Workspace2, mirroring the structure in Workspace1.
Pipeline1 will be moved into Folder1 in Workspace2, maintaining the same folder structure.
Lakehouse1 will be deployed to Workspace2 as it exists in Workspace1.

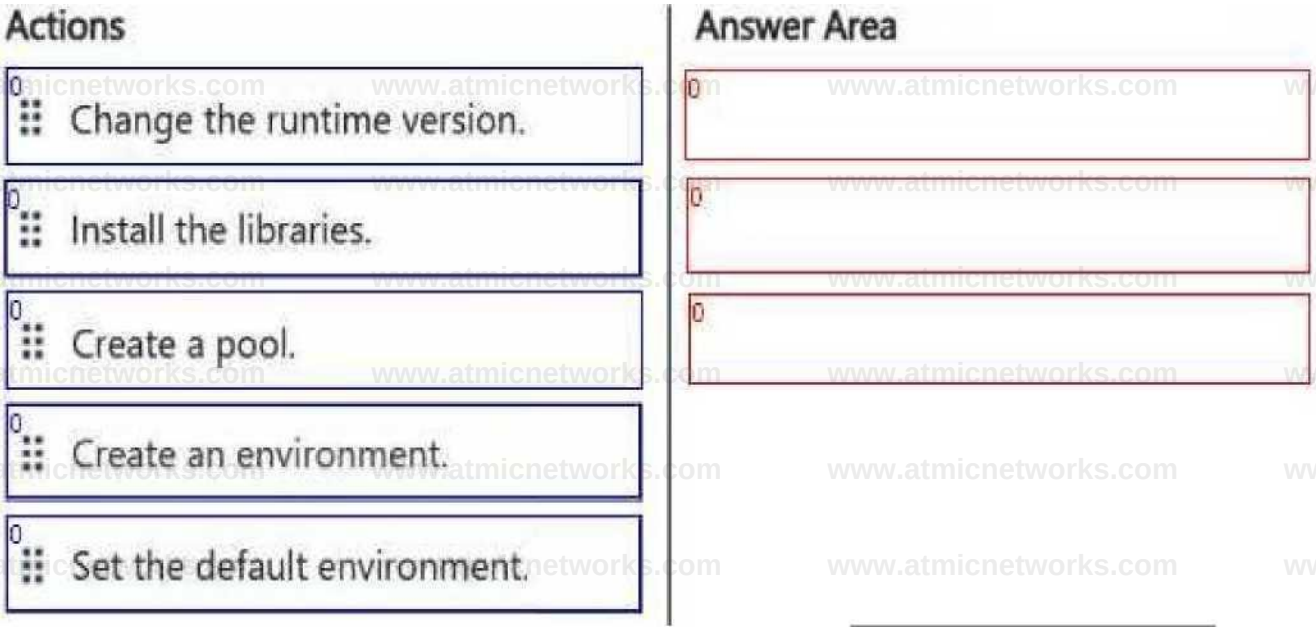
Question: 43

DRAG DROP

Your company has a team of developers. The team creates Python libraries of reusable code that is used to transform data.

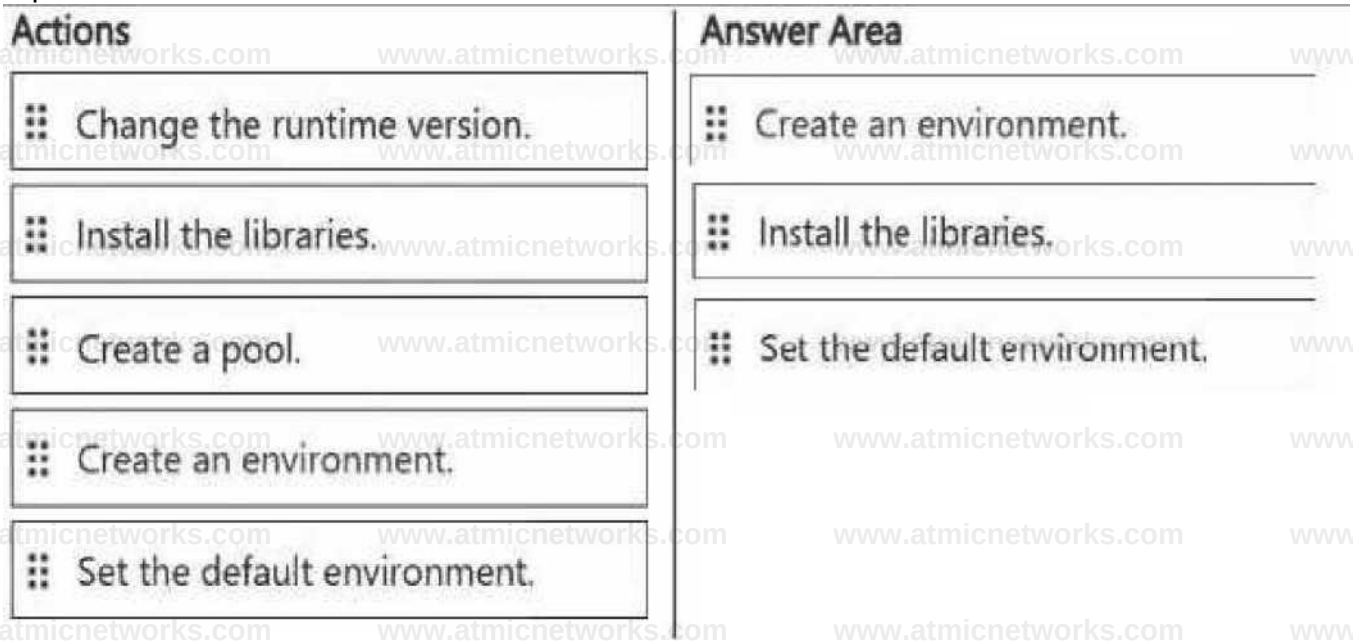
You create a Fabric workspace name Workspace1 that will be used to develop extract, transform, and load (ETL) solutions by using notebooks.

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.



Answer:

Explanation:



Question: 44

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.

Question: 45

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.

Question: 46

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:

To identify the version of Delta used when Notebook1 was executed, you should use the Admin monitoring workspace. The Admin monitoring workspace allows you to track and monitor detailed information about the execution of notebooks and

jobs, including the underlying versions of Delta or other technologies used. It provides insights into execution details, including versions and configurations used during job runs, making it the most appropriate choice for identifying the Delta version used during the execution of Notebook1.

Question: 47

DRAG DROP

You have a Fabric .

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

```
CREATE TABLE dbo.DimCuStomer (
    CustomerKey VARCHAR(255) NOT NULL,
    Name VAHCHAR(255) NOT HULL.
    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.

Answer:

Explanation:

Code Segments

;; DROP CONSTRAINT PKDimCUStomer

.. ADO CONSTRAINT PK_DimCUStomer PRIMARY KEY NONCLUSTERED

** (CustomerKey)

;: NOT ENFORCED

ALTER TABLE dbo.DimCustomer

., ADD CONSTRAINT PKDimCUStomer PRIMARY KEY CLUSTERED

" (CustomerKey)

;• ENFORCED

Answer Area

ALTER TABLE dbo.DimEustouer

ADD CONSTRAINT PK_DinCustomer PRIMARY KEY CLUSTERED

(CustomerKey)

:: ENFORCED

Question: 48

You have a Fabric workspace that contains a semantic model named Model1.

You need to dynamically execute and monitor the refresh progress of Model1. What should you use?

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

Answer: B

Explanation:

Semantic models in Microsoft Fabric are part of Power BI datasets and require refreshes to stay updated with the latest data.

Dynamically executing and monitoring the refresh progress requires a tool or approach that integrates with Fabric's capabilities for semantic models.

Question: 49

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 NoBikes >= 15
| sort by NoBikes
| project BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp | project
BikepointID, Street, Neighbourhood, NoBikes, NoEmptyDocks, Timestamp
```

Does this meet the goal?

A. Yes
B. no

Answer: B

Explanation:

This code does not meet the goal because it uses sort by without specifying the order, which defaults to ascending, but explicitly mentioning asc improves clarity.

Correct code should look like:

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

Question: 50

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:

```
bikelocation
| filter Neighbourhood == "sands End" and no Bikes >= 15
| order by NoBikes
| project BikepointID, Street, Neighbourhood, NoBikes, NoEmptyDocks, Timestamp
```

Does this meet the goal?

A. Yes
B. no

Answer: B

Explanation:

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

Correct code should look like:

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

Question: 51

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:

```
bikelocation
| filter Neighbourhood == “Sands End" and NoBikes >= 15
| sort by No Bikes asc
| project BikepointID, Street, Neighbourhood, NoBikes, NoEmptyDocks, Timestamp
```

Does this meet the goal?

A. Yes
B. no

Answer: A

Explanation:

Filter Condition: It correctly filters rows where Neighbourhood is "Sands End" and No_Bikes is greater than or equal to 15.
Sorting: The sorting is explicitly done by No_Bikes in ascending order using sort by No_Bikes asc. Projection: It projects the required columns (BikepointID, Street, Neighbourhood, No_Bikes, No_Empty_Docks, Timestamp), which minimizes the data returned for consumption.

Question: 52

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:

This code does not meet the goal because this is an SQL-like query and cannot be executed in KQL, which is required for the database.

Correct code should look like:

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

Question: 53

HOTSPOT

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	End Date	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

WIT:- rttull ASI'

SELECT £. Employee!!'

J ti. FirlplnyPP'N^r^

j e.EmployeePosit Lon

j c.CMtrattrype

j I T n **(date* c.startujtc) as startiMtc**

n tASI **g**
n n
CONVERT
REP-ACE
SUBSTRING

FROM Employee AS e

contract AS c on c.tnploycciu - e. Employcel J

CROSS JON
INNIR OIN
LEFT OUTER JON RIGHTCUTERJOIN

)

=FIFCT COLI|rr(OT^NCT FnpInyppTR) £S Total Fap lOye^S

j ContractType

HWfl result

CHOj' EV contract ypc

cCUNT(DISTINCT EmployeeID) > 2

CONIAIN:;
HAVING
EMIT
Y/RERE

Answer:

Explanation:

Answer Area

```
WITH result AS(
SELECT e.EmployeeID
    , e.Employeename
    , e.Employeeposition
    , c.ContractType
    ,                 ^ (date, c.startDate) as startDate
```

r^^^^“^^^~^^n CAST CONVERT REPLACE SUBSTRING

```
FROM Employee AS e
                ^ Contract AS c on e.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 LIMIT WHERE

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

Question: 54

HOTSPOT

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.

NOTE: Each correct selection is worth one point.

Answer Area

Data processor:

A data pipeline

A Dataflow Gen2 dataflow

An eventstream with a custom endpoint

An eventstream with an external data source

iltering:

A Filter activity in a data pipeline

A filter in a Dataflow Gen2 dataflow

A KQL statement

An eventstream processor

Answer:

Explanation:

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

Question: 55

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.

Question: 56

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.

Question: 57

HOTSPOT

HOTSPOT

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:
            dfsource.write.format('delta').mode('overwrite').saveAsTableCfftargettable}")
        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 changedetectioncolumns])

        update expr = {col: f source.{col}' for col in df source.columns}

        mergeoperation = deltaTable.alias('target') ,merge(
            source=df_source.alias('source'), condition=match_condition
        ) .whenMatchedllpdate(
            condition=update_condition,
            set=update_expr
        ).whenNotMatchedlnsertAl1()

        mergeoperation .execute()
    except Exception as e:
            print(f'Insert operation for table {targettable} 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	Yes	No
The target table will always be overwritten.	○	○
The merge operation will always run.	○	○
The loading pattern supports both full and incremental loading requirements.	○	○

Answer:

Explanation:

Answer Area

Statements	Yes	No
The target table will always be overwritten.		
The merge operation will always run	O	
The loading pattern supports both full and incremental loading requirements.		

Question: 58

HOTSPOT

HOTSPOT

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

orcerLine-j order_li^e_dd
.ClP^crDatc prdi?r_dc^?
,E .< jstt»r ^ty
p[L< ■.h:«'" id
j Qua' 7 i ty orc er_<nanti ty ,unitpnte unit_price , ex late taj;_rate
RO*
LakenouseJ .c-der l_iei □

INNER JOIN
Like iou ^ u2.di m_c u stomt^x L
ON o.custonie'id c.cbstomerjd

AND
l .i ■_: ..ii t nl = 1 ixCrderCate > eval id o? date dm? o.CrderDate >- e.vil c_frcr_catatiT

AND _____
c,ii_-rrt:rit - 1
■ >.('i I IHI Rain < ■ '.'.did- cjldkliinH
o.CrdcrC.atc <=c?dc frer cat:tir?

Answer:

Explanation:

Answer Area

```
SELECT
   SrderLinelD order. lirte_id
   jOrderDate order date
   ^•customer_key
   ,c.customer id
   ^Quantity orderquarttity
   jUnitPrice unit_price
   ^TaxRate tax_rate
FROM
   Lakehem sei.orderlines o
INNER JOIN
   Lakehouses.dimcu stoner c
```

ON o.customerid = c.customer id

AND _____

cis.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

Question: 59

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.

Question: 60

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.

Question: 61

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	81416f30-6032-4075-9bl9-238403059735	{ "index”: 0, "eventid": "719afca0-be30-4559-bb5e-59feade642f6"]

2024-05-18 12:45:21.76423	bb664010-02aa-4el7-8C83-116cd4458d52	["index": 0, "eventid": "782222b2-fbcb-43c0-82d6-ecd49a99dbf5")]
2024-05-18 12:45:23.98642	717bfe7d-0e5d-498f-9f21-e60aaf258056	"index": 0, "eventid": "d5730286-0da4-41f8-8e59-f7Se209310a9" }]

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 DeviceId 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)	DeviceId (guid)	StreamData (dynamic)
2024-05-18	81416f30-60a2-4e75-9bl9-2a84ea059735	{ "index": 40, "eventid": "729afca2-be30-4559-bb5e-59feade642f3", "temperature": 32 }]

B)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)

2024-05-21	81416f30	[
		{
		"index": 0,
		"eventid": "719afca0-be30-4559-bb5e-5werade642f6", "temperature": 27
		}

C)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)
2014-05-23	81416-f3060a24e759bi92a84eB05973532dhdyte3	("index": 0, "eventid": "71Mfca0-be30-*559-bb5e-59feade642f6')]

D)

Timestamp (datetime)	DeviceId (guid)	StreamData (dynamic)

2024-05-24	81416f30-60a2-4e75-9bl9-2a84ea059735	[
		{
		"index": 0,
		"eventid": "7i9afca0-be30-4559-bb5e-59feade642f6"
		}
		]

A. Option A B. Option B

C. Option c

D. Option D

Answer: B,D

Explanation:

Changes to Table1 Structure:
StreamData column: An additional temperature element was added.
Timestamp column: Data type changed from datetime to date.
DeviceId column: Data type changed from guid to string.

Impact of Changes:
Only records that comply with Table2's structure will load.
Records that deviate from Table2's column data types or structure will be rejected.

Record B:
Timestamp: Matches Table2 (datetime format).
DeviceId: Matches Table2 (guid format).
StreamData: Contains only the index and eventid, which matches Table2.

Accepted because it fully matches Table2's structure and data types.

Record D:
Timestamp: Matches Table2 (datetime format).
DeviceId: Matches Table2 (guid format).
StreamData: Matches Table2's structure.

Accepted because it fully matches Table2's structure and data types.

Question: 62

HOTSPOT

HOTSPOT

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.

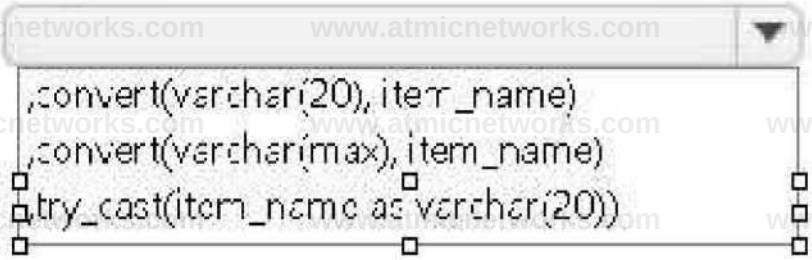
You need to resolve the issues. The solution must ensure that the data types of the results are retained. The results can contain blank cells.

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

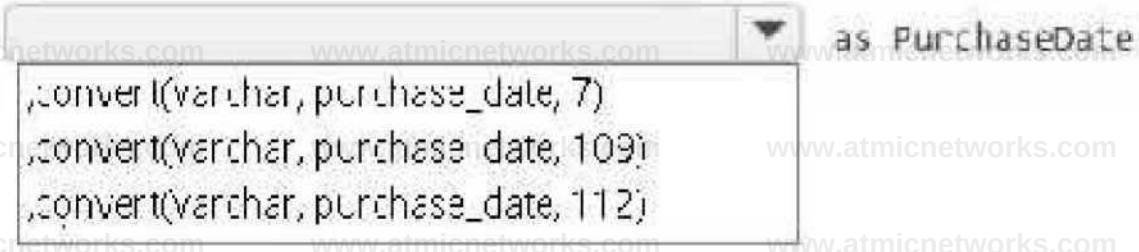
NOTE: Each correct selection is worth one point.

Answer Area

```
SELECT
  item_id as ItemId
```




```
,item_description as ItemDescription
```




```
FROM
  Table1
WHERE
  item_type = @itemtype_parameter
```

Answer:

Explanation:

Answer Area

SELECT

item id as ItemId

[▼] as itemName

- ,convert(varchar(2 0), i tem_nam e)
- , c o nve r t(va r c h a r (m ax), i te m _n am
- e) ,try_cast(item_name as varchar(20))

,item description as ItemDescription

[▼] ^ as PurchaseDate

- ,convert(varchar, purchase_date, 7)
- ,convert(varchar, purchase_date, 109)
- ,convert(varchar, purchase-date, 112)

FROM

Table1

WHERE

item type = ^itemtype_parameter

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

Question: 63

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.

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 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
Device Id	Int

Reference contains reference data in the following format.

Column name	Data type
Device Id	Int
Device Na me	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 oeviceId
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.

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 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) join
03 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:

Using extend retains all columns in the table, potentially increasing the size of the output unnecessarily. project is more efficient because it selects only the required columns.

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 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 move the filter to line 02.

Does this meet the goal?

A. Yes

B. No

Answer: A

Explanation:

Moving the filter to line 02: Filtering the Stream table before performing the join operation reduces the number of rows that need to be processed during the join. This is an effective optimization technique for queries involving large datasets.

Question: 67

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
Geo Location	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:

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

slower.

Question: 68

DRAG DROP

DRAG DROP

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.

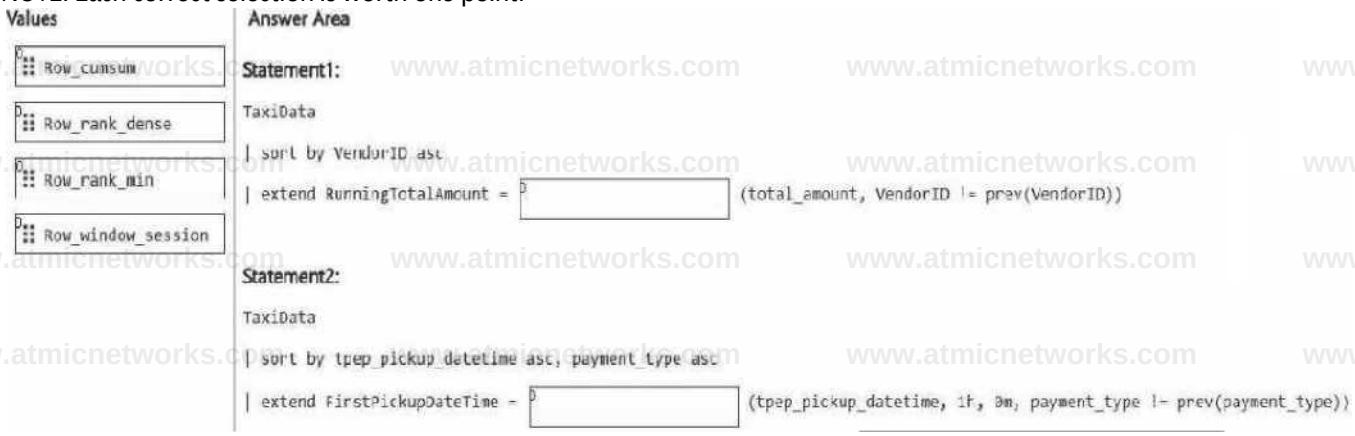
VendorlD	tpep_pickup_datetime	tpep.dropoff.datetime	passenger.count	trip distance	PULocationlD	DOlocationlD	payment, type	total.amount
2	2022-06 06711.08:322	2022-06-06111:22:172	1	0.17	231	SO	2	7.12
2	2022-06-06T11 12:05Z	2022-06-06111:20:432	1	1.02	161	163	1	10.56
1	2022-06 06711:15:002	2022-06-06711:25:322	1	1.07	142	230	2	17.12
2	2022-06 06111 29:542	2022-06-06711:49:342	2	2.07	162	236	2	12.01
1	2022-06-06T11:50:50Z	2022-06-06712:07:242	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.

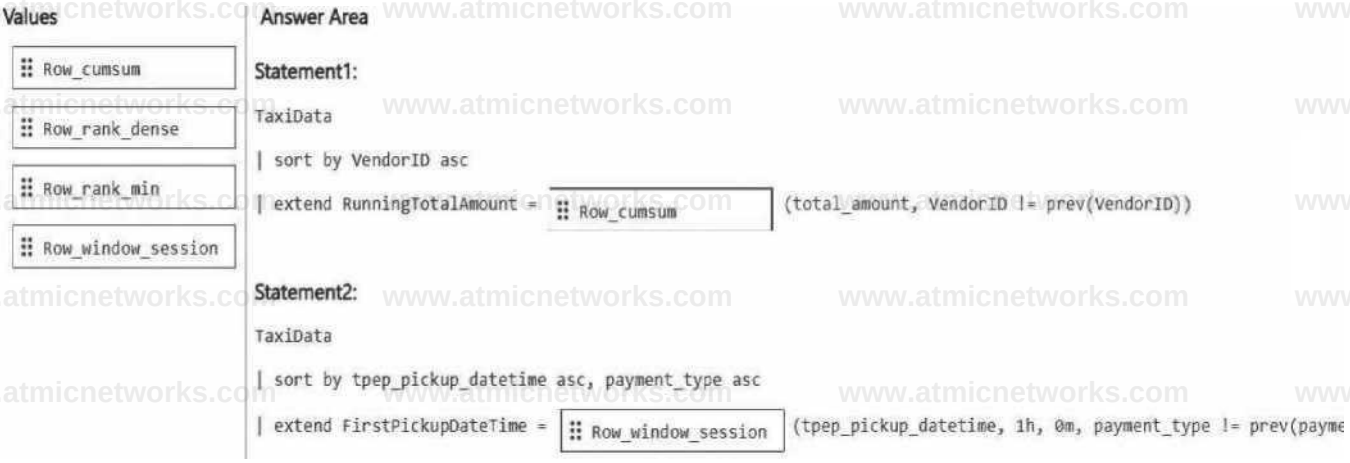
How should you complete each query? To answer, drag the appropriate values the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.



Answer:

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 = Rowcumsun^totalamount, 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 FirstPickupOateTime - Row_window_session(tpep_pickup_datetime, Ih, 0m, paymenttype != prev(payment type))
```

Question: 69

HOTSPOT

HOTSPOT

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	Yes	No
Litware from New York will be displayed at the top of the result set.	○	○
Fabrikam in Seattle will have value = 2 in the Rank column.	◉	○
Litware in San Francisco will have the same value in the Rank column as Litware in New York.		

Answer:

Explanation:

Answer Area

Statements	Yes	No
Litware from New York will be displayed at the top of the result set.		
Fabrikam in Seattle will have value = 2 in the Rank column.		

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

Litware from New York will be displayed at the top of the result set – Yes
The data is sorted first by Location in descending order and then by UnitsSold in descending order.
Since "New York" is alphabetically the last Location, it will appear first in the result set. Within "New York", Litware has the highest UnitsSold (1000), so it will be displayed at the top.

Fabrikam in Seattle will have value = 2 in the Rank column – No
The row_rank_dense function assigns dense ranks based on UnitsSold within each location. In "Seattle":
Contoso has UnitsSold = 300 → Rank 1
Litware has UnitsSold = 100 → Rank 2
Fabrikam also has UnitsSold = 100, so it shares the same rank (2) as Litware.

Litware in San Francisco will have the same value in the Rank column as Litware in New York – No
The rank is calculated separately for each location. In "San Francisco":
Both Relecloud and Litware have UnitsSold = 500, so they share the same rank (1).
In "New York", Litware has the highest UnitsSold = 1000 → Rank 1.

Since ranks are calculated independently for each location, Litware in San Francisco does not share the same rank as Litware in New York.

Question: 70

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.

Question: 71

- ou have a Fabric workspace that contains a warehouse named Warehouse1. Data is loaded daily into Warehouse1 by using data pipelines and stored procedures.

You discover that the daily data load takes longer than expected.

You need to monitor Warehouse1 to identify the names of users that are actively running queries. Which view should you use?

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

Answer: E

Explanation:

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

Question: 72

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:

Y ACUUM 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;

Question: 73

Y ou 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.

Y ou 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:

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.

Question: 74

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

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

Answer: E

Explanation:

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

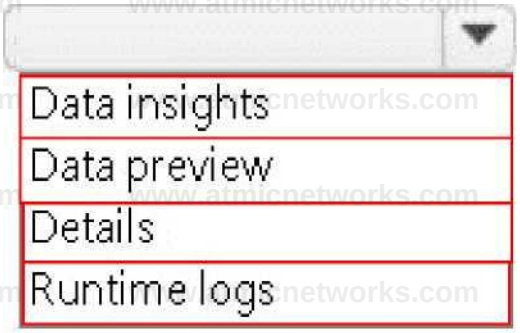
Question: 75

HOTSPOT

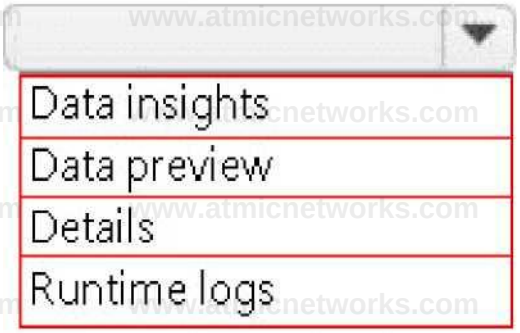
HOTSPOT
You have a Fabric workspace that contains an eventstream named EventStream1.
You discover that an EventStream1 transformation fails.
You need to find the following error information:
The error details, including the occurrence time
The total number of errors
What should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

To find the erro: details;



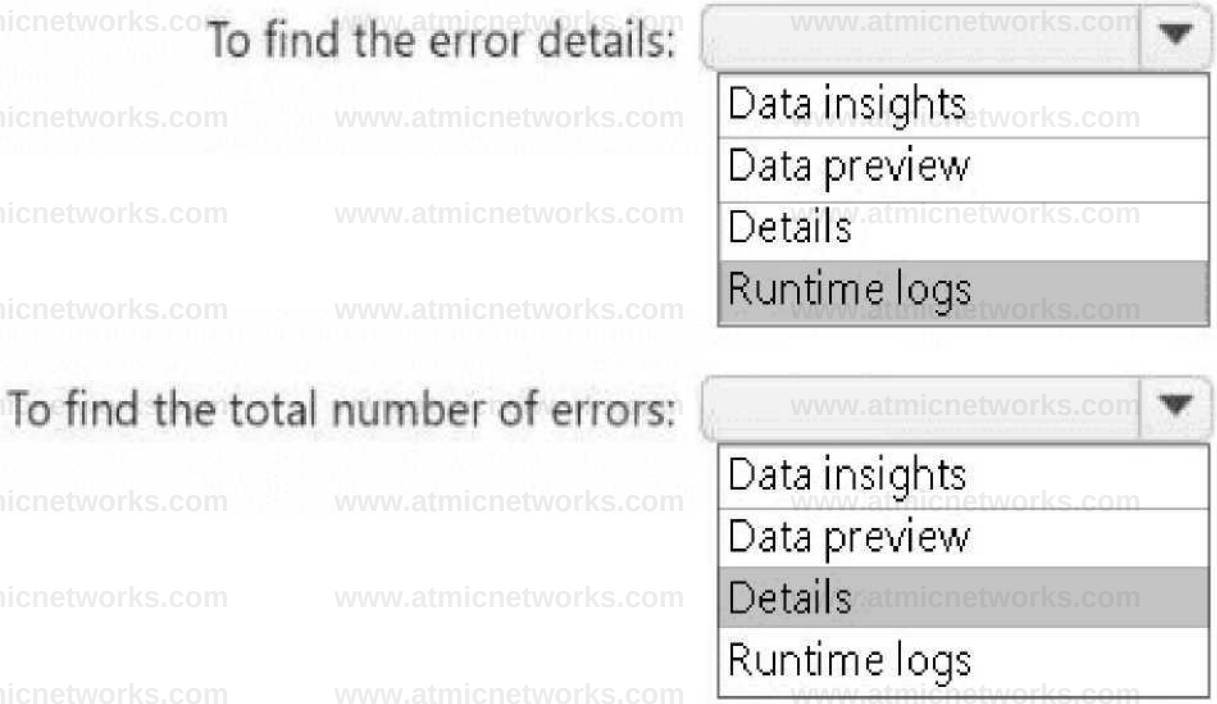
To find the total number of errors:



Explanation:

Answer:

Answer Area



Question: 76

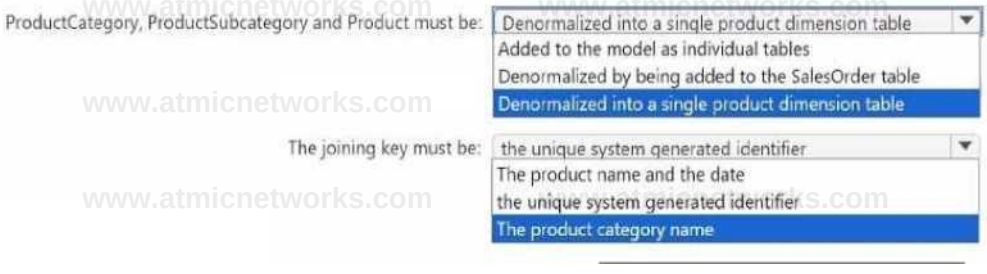
HOTSPOT

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



Answer:

Explanation:

Answer Area

PioductCategory, ProductSubcategory and Product must be: Denormalized into a single product dimension table

The joining key must be: the unique system generated identifier

Question: 77

DRAG DROP

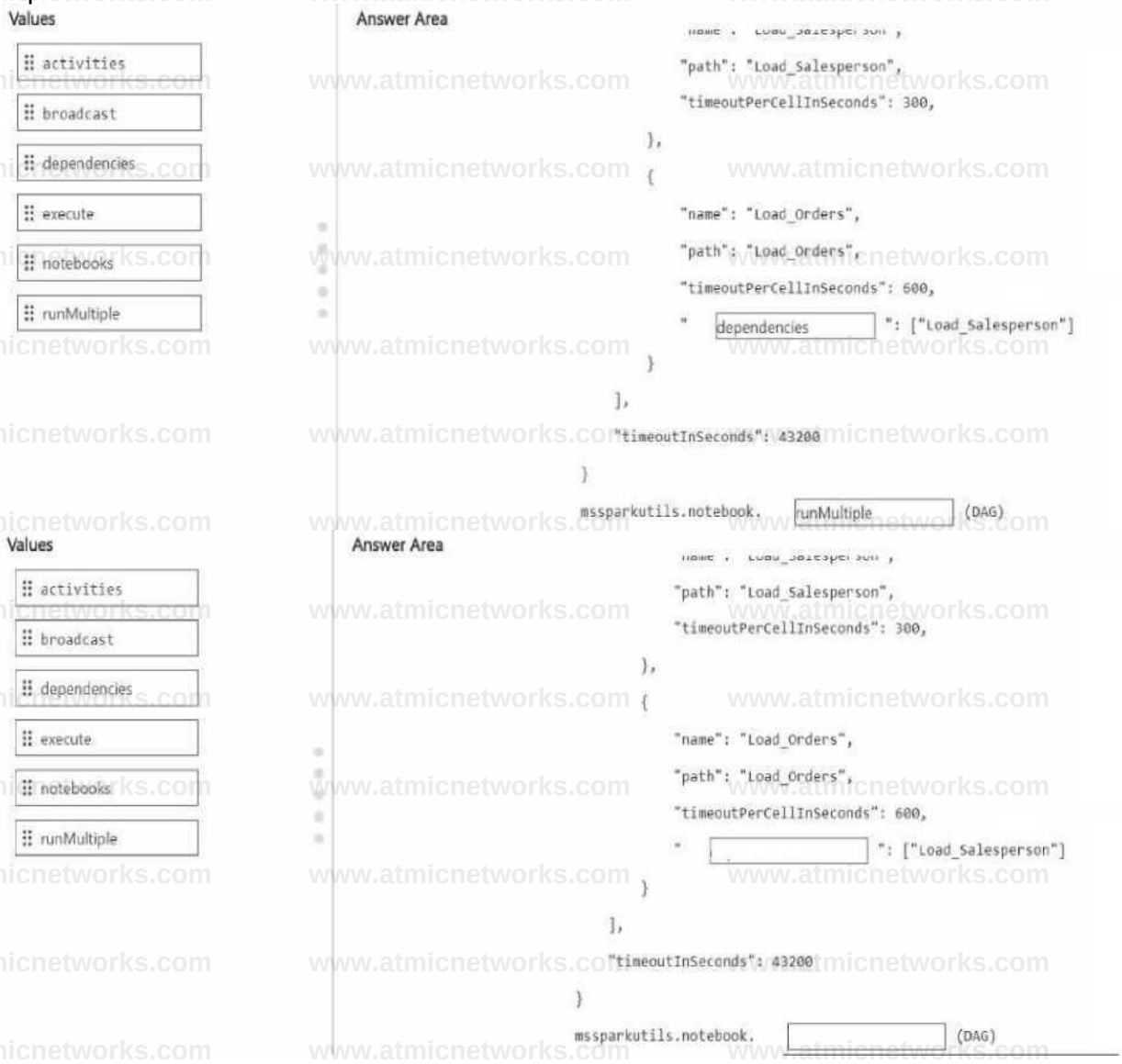
You have two Fabric notebooks named Load_Salesperson and Load_Orders that read data from Parquet files in a lakehouse. Load_Salesperson writes to a Delta table named dim_salesperson. Load.Orders writes to a Delta table named fact_orders and is dependent on the successful execution of Load_Salesperson.

You need to implement a pattern to dynamically execute Load_Salesperson and Load_Orders in the appropriate order by using a notebook.

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

NOTE: Each correct selection is worth one point.

Explanation:



Answer:

Question: 78

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 by a low-code approach.
- Team2 consists of members that have a background in Python programming. The team wants to use PySpark code to transform data.
- Team3 consists of members who currently use Azure Data Factory. The team wants to move data between source and sink environments by using the least amount of effort.

You need to recommend tools for the teams based on their current personas.

What should you recommend for each team? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Teaml |. Dataflow Gen2 dataflows * j Data pipelines Notebooks

Team2: Notebooks
I Data pipelines

Dataflow Gen2 dataflows

Team3: Data pipelines

Notebooks

Dataflow Gen2 dataflows

Answer:

Explanation:

Answer Area

Teaml; [Dataflow Gen2 dataflows ^j

Team2: Notebooks

Team3: Data pipelines

Question: 79

HOTSPOT

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.

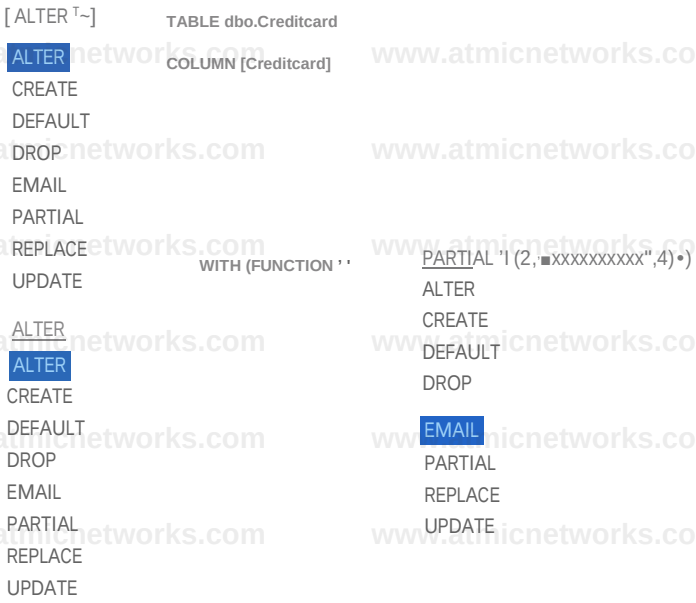
```
CREATE TABLE [warehouses].[dbo].[creditcard]
    Creditcard varchar(20) MOT HULL
    .creditCardType varchar(10) NOT NULL)
GO
```

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.

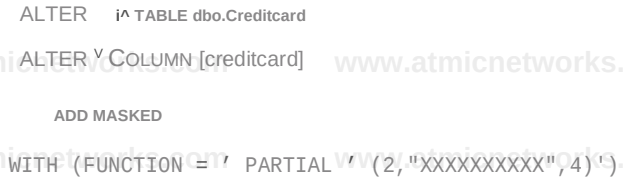
Answer Area



Answer:

Explanation:

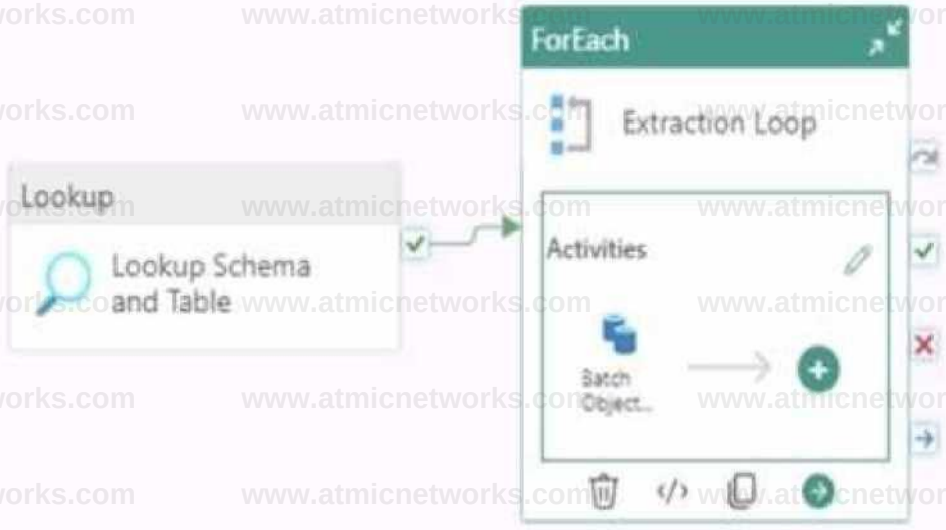
Answer Area



Question: 80

HOTSPOT

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

8activity(' Lookup Schema and Table *).

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

output.value
output
output.count
output.pipelineReturnValue
outputvalue

Answer:

Explanation:

Answer Area

^activity(' Lookup Schema and Table '). output.value

Question: 81

HOTSPOT

You have a Fabric workspace that contains a warehouse named Warehouse!. Warehousel contains a table named DimCustomers. DimCustomers contains the following columns:

- CustomerName
- CustomerlD
- BirthDate
- Email

You need to configure security to meet the following requirements:

- BirthDate in DimCustomer must be masked and display 1900-01-01.
- Email 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 =   'random (1/@75) default!)'
                               emailf)'
                              ■paitial(1/@\5y_
```

Answer

Explanation:

Answer Area

```
ALTER TABLE DimCustomer
ALTER COLUMN BirthDate
ADD MASKED WITH (FUNCTION =   [ 'default!)-

ALTER TABLE DimCustomer
ALTER COLUMN EmailAddress
ADD MASKED WITH (FUNCTION -   'random (1, "@", 5)'
```

.dr)

Question: 82

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

- A Microsoft Power Bl report named Reportl
- A Power Bl 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

Explanation:

Question: 83

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

While monitoring Warehouse1, you discover that query performance has degraded during the last 60 minutes.

You need to isolate all the queries that were run during the last 60 minutes. The results must include the username of the users that submitted the queries and the query statements. What should you use?

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

Answer: C

Explanation:

Question: 84

You have a Fabric workspace that contains a semantic model named Modell. You need to monitor the refresh history of Model 1 and visualize the refresh history in a chart. What should you use?

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

Answer: B

Explanation:

Question: 85

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 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

DatasetI; [A T -SQL statement
A Dataflow Gen2 dataflow
A notebook
A T-SQL statement

Dataset2: A notebook
A Dataflow Gen2 dataflow
A notebook
A T-SQL statement

Dataset3: A KQL queryset
A Dataflow Gen2 dataflow
A KQL queryset
' A T-SQL statement

Answer:

Explanation:

Answer Area

DatasetI: A T-SQL statement

Dataset2: A notebook

Dataset3: A KQL queryset

Question: 86

HOTSPOT

You have a Fabric workspace that contains a lakehouse named Lakehousel. Lakehousel 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 name Notebook!, 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 Notebook-!. 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
,nerge(sourceDF, "sourceDF.Key" " "targetDF.Key")
.whenMatchedUpdatet _____ if
.whenMatchedInsert(
) whenNotMatchedBySourceInsert(,whenNotMatchedBySourceUpdate(.whenNotMatchedInsert(
whenNotMatchedIlpdate(
)
,whenNotMatchedInsert(_____ ')
.whenMatchedInsert(whenMatchedUpdate(whenNotMatchedBySourceInsert(
whenNotMatchedBySourceUpdatef _____
] ,whenNotMatchedUpdate(_____ }
) .whenNotMatchedBySourceUpdate(_____ y]
.when Matched 1 nsertf
. M k JI. J / ent_date() - INTERVAL '7' DAY)",
whenMatchedUpdatef
whenNotMatchedBySourceInsert(
.whenNotMatchedInsertf . _whenNotMatchedUpdate(

Answer:

Explanation:

Answer Area

(targetDF
•merge (sourceDF, "sourceDF.Key" = "targetDF.Key")
whenMatchedUpdatef . *
set = {"targetDF.LastModified": "sourceDF.LastModified")

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

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

.execute()

)
```

Question: 87

DRAG DROP

You are building a data loading pattern by using a Fabric data pipeline. The source is an Azure SQL database that contains 25 tables. The destination is a lakehouse.

In a warehouse, you create a control table named Control.Object as shown in the exhibit. (Click the Exhibit tab.)

You need to build a data pipeline that will support the dynamic ingestion of the tables listed in the control table by using a single execution.

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

Actions

Answer Area

,, Add a Get metadata activity to query Control,Object and generate a list of
" schemas and tables to copy.

Add an Until activity to iterate over the list of tables and copy the source data to the lakehouse Delta tables

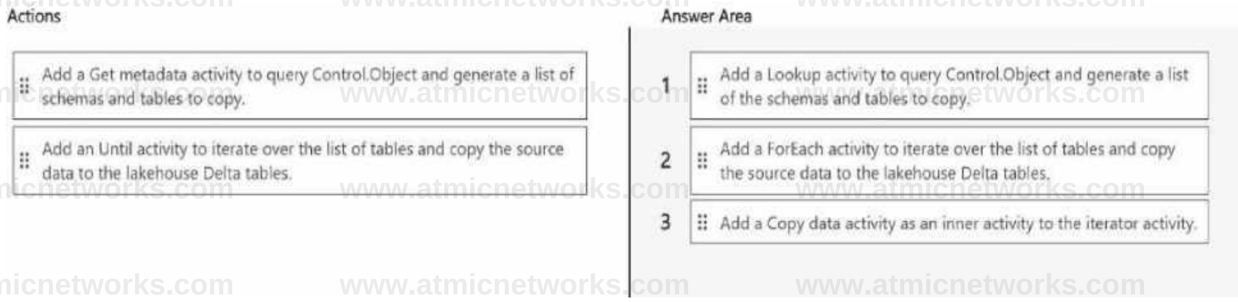
Add a Lookup activity to query Control.Object and generate a list of the schemas and tables to copy.

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

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

Answer:

Explanation:



Question: 88

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:

Question: 89

You have a Fabric workspace that contains an eventstream named EventStreaml. EventStreaml outputs events to a table named Tablel in a lakehouse. The streaming data is souiced 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:

Question: 90

HOTSPOT

You have a table in a Fabric lakehouse that contains the following data.

SalesOrderNumber	OrderDate	CustomerName	Email
SO49172	2021-01 01	Brian Howard	bnan23@adventure-works.com
SO49173	2021-01 01	Linda Alvarez	linda19@adventuie-works.com
SO49174	2021-01-	Gina Hernandez	gina4@adventure woiks.com
SO49178	2021-01 01	Beth Ruiz	bet h4 @ ad ve n tu re - works.com
SO49179	2021-01 01	Evan Ward	evan13@adventuie-works.com

You have a notebook that contains the following code segment.

```
01 df = df.withColunmCCustomerNaiae", when((col("CustcnerNaineM).isNull{) | (col("Cu5to^^e^^^allle")==",k)),lit(,Unknownk)).othe^wise(col("Customer^Ja^le,")))
02 df = df.withcolwinCusernaBe'^splitCcolCEmail"), "^J-getitemfi))
03 df - df.drDpDuplicates([',orderDate"]).select(col("orderDaLe")) year(''OrderDate") .alias ("Year"), ("Customername")( (''username"))
04 d Up layoff* 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 Oidei Date column and keep only the first occurrence for each year	○	

Answer:

Explanation:

Answer Area

Statements	Yes	No
Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value.	©	0
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.		©

Question: 91

You have a Fabric workspace named Workspace1. Your company acquires GitHub licenses.

You need to configure source control for Workpace1 to use GitHub. The solution must follow the principle of least privilege.

Which permissions do you require to ensure that you can commit code to GitHub?

A. Actions (Read and write) and Contents (Read and write)

B. Actions (Read and write) only

C. Contents (Read and write) only

D. Contents (Read) and Commit statuses (Read and write)

Answer: C

Explanation:

Question: 92

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

You have a Fabric workspace named Workspace-!. Workspace! contains a notebook named Notebookl 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 Notebookl that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVaultl 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: A

Explanation:

Question: 93

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

You have a Fabric workspace named Workspace!. Workspace! contains a notebook named Notebookl 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 Notebookl that will use the Fabric API to monitor the semantic model refreshes. You need to retrieve the registered application ID and secret from KeyVaultl 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

Explanation:

Question: 94

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

A. warehouse

B. data pipeline

C. Dataflow Gen2 dataflow

D. notebook

Answer: B

Explanation:

Question: 95

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

You plan to create a data pipeline named Pipeline! to ingest data into Lakehousel. You will use a parameter named paraml to pass an external value into Pipeline1!. The paraml parameter has a data type of int

You need to ensure that the pipeline expression returns param1 as an int value.

How should you specify the parameter value?

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

Answer: B

Explanation:

Question: 96

You have a Fabric workspace that contains a takehouse 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. Set the Retry setting of the Notebook activity to 2.
B. Place the Semantic model refresh activity after the Notebook activity and link the activities by using an On completion condition.
C. Place the Semantic model refresh activity after the Notebook activity and link the activities by using the On success condition.
D. From the Schedule settings of Notebook1, set the time zone to UTC.
E. From the Schedule settings of Pipeline1, set the time zone to UTC.
F. Set the Retry setting of the Semantic model refresh activity to 2.

Answer: A, C, E

Explanation:

Question: 97

HOTSPOT

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

Customer! D	First Name	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	Biown	555-444 5555	5678123498761230

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

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

CreditCard column.

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

NOTE: Each correct selection is worth one point.

Answer Area

GRANT **SELECT** i^v

ALTER EXECUTE

i READ

SELECT

VIEW

Customers(CustomerlD, FirstName, LastNane, Phone) TO [[user 1

@contoso.comj

User! [Used]

[user 1 @contoso.com]

Answer:

Explanation:

Answer Area

GRANT [SELECT on

Custoners(CustomerlO, FirstName, LastName, Phone)

TO [user1@contoso.com] ^

Question: 98 Exhibit.

You have a Fabric workspace that contains a write-intensive warehouse named DW1. DW1 stores staging tables that are used to load a dimensional model. The tables are often read once, dropped, and then recreated to process new data.

You need to minimize the load time of DW1.

What should you do?

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

Answer: A

Explanation:

Question: 99

You have a Fabric workspace that contains a data pipeline named Pipeline! as shown in the exhibit.

(Click the Exhibit tab.) What will occur the next time Pipelinel tuns?

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

Answer: B

Explanation:

Question: 100

You are building a Fabric notebook named MasterNotebookl in a workspace. MasterNotebookl contains the following code.

```
DAG = {
    "activities": [
        { "name": "executenotebookl",
            "path": "notebook 01",
            "tineoutPerCelllnSeconds"; 600,
            "args": (
                        "input_value": "999"
                b
            "retry": 1,
            "retrylntervalinseconds": 30

        ( "name": "execute notebook^?",
            "path": "notebook_02",
            "tineoutPercelllnseconds": 400,
            "args": {
                        "inputvalue": "888"
                    b
            "retry": 1,
            "retryinterval Inseconds": 10
        b
        { "name": "execute_notebook_3",
            "path": "notebook 03",
            "timeoutPerCelllnSeconds": 600,
            "args": (
                        "input value": "777"
                    b
            "retry": 1,
            "retrylntervalinseconds": 30

    "timeout Inseconds": 43200,
    "concurrency": 0
```

You need to ensure that the notebooks are executed in the following sequence:

1. Notebook_03
2. Notebook.Ol
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. Split the Directed Acyclic Graph (DAG) definition into three separate definitions.
B. Change the concurrency to 3.
C. Move the declaration of Notebook_03 to the top of the Directed Acyclic Graph (DAG) definition.

D. Move the declaration of Notebook_02 to the bottom of the Directed Acyclic Graph (DAG) definition.

E. Add dependencies to the execution of Note boo k_02.

F. Add dependencies to the execution of Notebook_03.

Answer: C, E

Explanation:

Question: 101

HOTSPOT

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

Name	Message size	Description
Source 1	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

Source! A streaming dataflow
Apache Spark Structured Streaming
An eventstream
A data pipeline

A streaming dataflow
An eventstream

Source?: Adata pipeline
Apache Spark Structured Streaming
An eventstream

A data pipeline
A streaming dataflow

Source3: An eventstream
Apache Spark Structured Streaming

An eventstream
A data pipeline
A streaming dataflow

Answer:

Explanation:

Answer Area

Source! A streaming dataflow

Source? A data pipeline

Souice3. An eventstream

Question: 102

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

Explanation:

Question: 103

You have a Fabric warehouse named DW1 that contains a Type 2 slowly changing dimension (SCD) dimension table named DimCustomer. DimCustomer contains 100 columns and 20 million rows. The columns are of various data types, including int, varchar, date, and varbinary.

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

Answer: D

Explanation:

Question: 104

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

Answer: B

Explanation:

Question: 105

DRAG DROP

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.

Code Segments

Answer Area

Answer:

Explanation:

Answer Area

:: .create table EventStreaaData (

2 :: TimeStamp:(idletime, DeviceId;string

3 ;; Stream Data ;k> nq)

Question: 106

DRAG DROP

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

You need to build a KQL query to compare the Meter-Reading value of each row to the previous row base on the ilmestamp value

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

City	Area	MeterReading	Timestamp	PrevMeterReading	PrevTimestamp
tor insets	Areal	1500	2024-0 7 30 10'00:00		
| Kansas	Area?	1520	2024-0730 11:00:00	1500	2024-0730 10:00:00

Values

- | evaluate
- | extend
- | lookup
- | project
- | sort

Answer Area

Readings

| filter City «* "Kansas"

Value by riaestanp

j Value | PrevMeter Read Ing = prev(HeterReading),

PrevTiaestaap - prev(Tinestanp)

Value City, Area, HeterReading, Tinestamp, PrevMeterReading, PrevT intestamp

Answer:

Explanation:

Answer Area

```
Readings
| filter City "Kansas"
| sort          by Timestamp
| extend        PrevHeterReading = prev(HeterReading),
                PrevTimestamp = prev(Timestamp)
| project       City, Area, MeterReading, Timestamp, PrevNeterReading,
                PrevTlmestamp
```