IT认证电子书

质　量　更　高　　服　务　更　好

半年免费升级服务

http://www.itrenzheng.com

Exam : **DP-700**

Title : Implementing Data Engineering Solutions Using Microsoft Fabric

Version : DEMO

1. 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.

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:

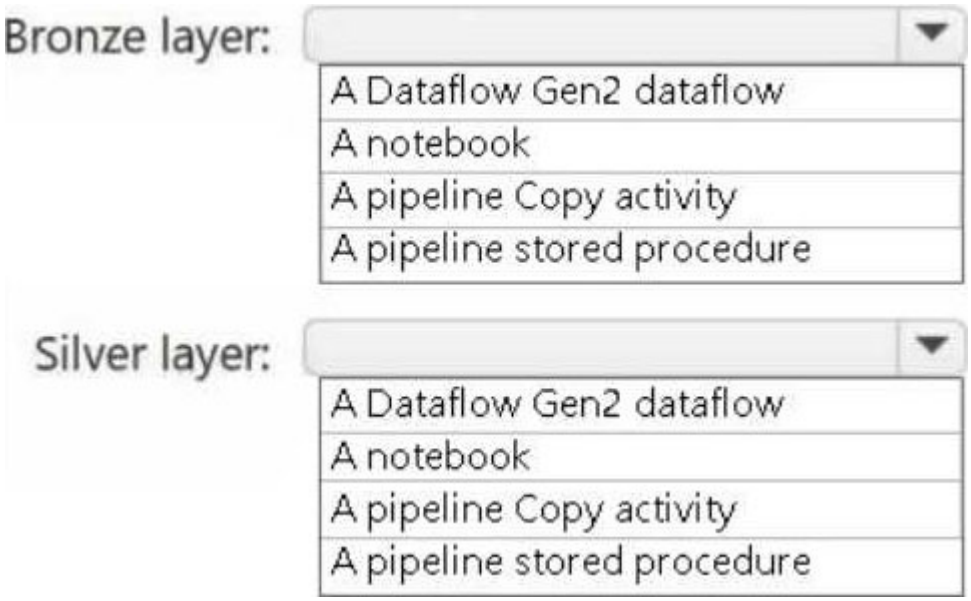
Data Analysts' Access Requirements must only have read access to the Delta tables in the gold layer and not have access to the bronze and silver layers.

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.

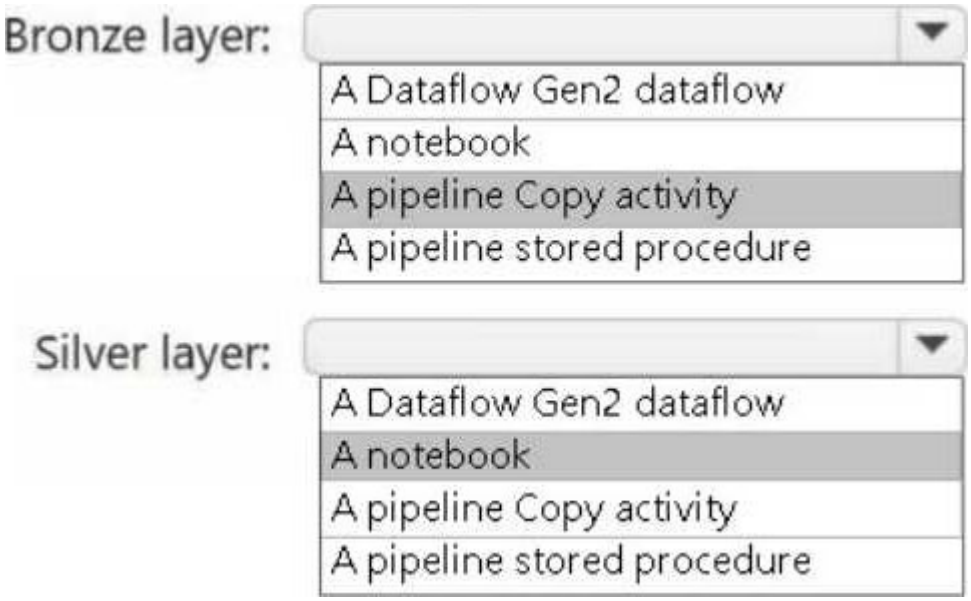
2.HOTSPOT

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:
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.

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.

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, CategoryName

FROM ContosoLake.Products p

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

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

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

Answer:

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

FROM ContosoLake.Products p

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

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

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

Explanation:

Join between Products and ProductSubCategories:

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

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

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: AB
Explanation:
MAR1 has seven entities, each accessible via a different API endpoint. A ForEach activity is required to iterate over these endpoints to fetch data from each one. It enables dynamic execution of API calls for each entity.
The Copy data activity is the primary mechanism to extract data from REST APIs and load it into the bronze layer in Delta format. It supports native connectors for REST APIs and Delta, minimizing development effort.