Q20

# RAPPEL

**🧾 Section “Tables” dans un Lakehouse (ou Warehouse) – Version Révision**

* La section **“Tables”** contient des **Managed Tables** :
  + Ce sont des **fichiers Delta** (Parquet + \_delta\_log), organisés comme une **base de données**.
  + On peut faire des requêtes SQL ou Spark dessus.
  + Gérées automatiquement par Fabric (créées, maintenues, versionnées).
* 📌 **Fonctionne comme une base relationnelle**, mais repose sur du **stockage fichier structuré**.
* ❌ On ne peut **pas créer de Shortcut dans la section “Tables”**.
  + Les **Shortcuts** (vers d’autres Lakehouses, OneLake, S3, etc.) se créent uniquement dans la section **“Files”**.

**🧠 À retenir :**

✔️ **“Tables” = tables Delta gérées** (pas juste des vues sur des fichiers)  
❌ **Shortcuts ≠ Tables** → ils ne peuvent pas apparaître dans “Tables”

**🔗 Shortcuts dans Microsoft Fabric (Lakehouse)**

* Un **Shortcut** est un **alias** (ou un lien symbolique) **vers une source externe** :
  + Un autre **Lakehouse** (même ou autre workspace),
  + Un stockage OneLake,
  + Un **compte ADLS**,
  + Un bucket **Amazon S3**.
* 📂 Les Shortcuts apparaissent uniquement dans la section **“Files”** d’un **Lakehouse**.
  + ❌ Ils **ne sont pas visibles** dans la section “Tables”.
  + ❌ Ils **ne peuvent pas être créés dans un Warehouse**.

**🔄 Que peut-on faire avec un Shortcut ?**

| **Action** | **Possible ?** | **Commentaire** |
| --- | --- | --- |
| ⚙️ Lire les fichiers (avec Spark) | ✅ | Notebooks, PySpark, etc. |
| 📊 Interroger en SQL directement | ❌ | Pas possible tant qu’on ne crée pas une table |
| 💾 Créer une Managed Table à partir d’un Shortcut | ✅ | En important les données dans “Tables” |
| 🧩 Utiliser dans un pipeline Dataflow | ✅ | Source d’ingestion typique |

**🧠 À retenir pour les examens :**

**Un Shortcut ≠ une table.**  
Il **pointe vers des fichiers** (externes ou internes à OneLake) mais ne permet pas d'exécution SQL directe.  
Pour requêter en SQL, il faut **charger ou convertir** ces données dans une **Managed Table**.

# CASE STUDY

## Contoso

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.

### MIND MAP CONTOSO

Une image contenant texte, diagramme, capture d’écran, ligne

Le contenu généré par l’IA peut être incorrect.

### SYNHESE CONTOSO

**🧭 Synthèse – Cas Contoso (DP-700)**

Contoso modernise sa plateforme analytique avec **Microsoft Fabric** pour les analyses marketing.  
L’architecture repose sur deux **Lakehouses** (brut → modèle) selon une approche **medallion** :

* **Lakehouse1** : ingestion (bronze) + nettoyage (silver)
* **Lakehouse2** : données analytiques prêtes (gold)

**👥 Rôles :**

* **Data Engineers** : ingestion/transformation (Python/SQL), accès complet aux lakehouses, Spark & source control.
* **Data Analysts** : lecture **uniquement** des **tables Delta du gold layer** via **SQL Analytics Endpoint**.

**🔁 Ingestion :**

* Données issues de POS1 (SQL Server), MAR1 (API REST) et Amazon S3 (Parquet).
* Ingestion **en Delta format** dans **bronze**, sans traitement.
* Connexions **built-in**, **retry obligatoire**, **source control via Azure Repos**.

**🧱 Architecture :**

* Suivi strict bronze → silver → gold.
* Nettoyage lourd (MAR1) : doublons, valeurs manquantes, capitalisation.
* Suppression hebdo des fichiers orphelins Delta.
* Accès SQL (pas Spark) aux données analytiques (gold).

**🔐 Sécurité :**

* Accès des analysts limité à la **lecture SQL du gold layer**.
* Pas d’accès aux couches bronze/silver ni aux fichiers Spark.

## Litware

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 -

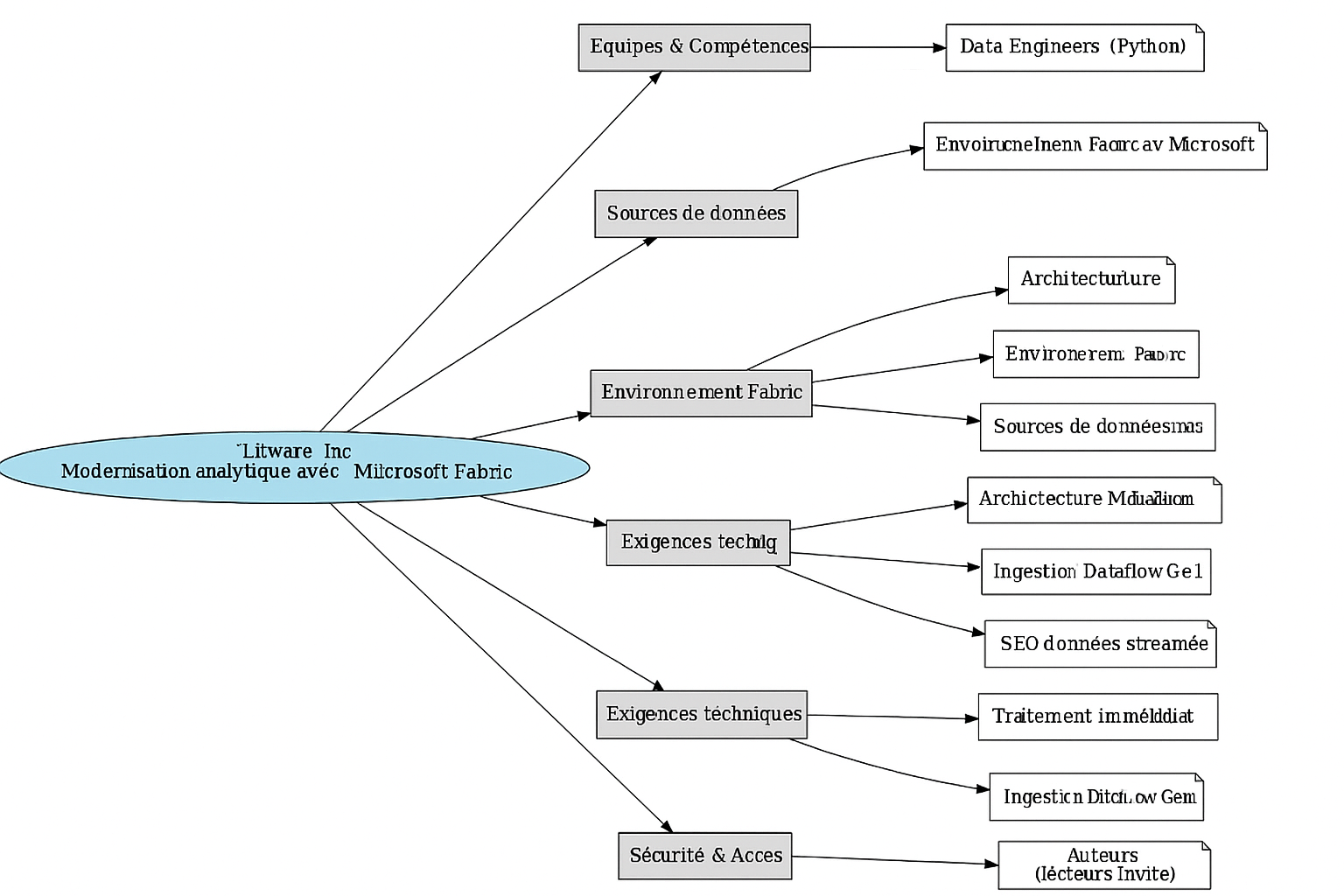
Litware identifies the following data requirements:

• 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.

### MIND MAP LITWARE



### SYNHESE LITWARE

**📚 Synthèse – Cas Litware (DP-700)**

**Litware, Inc.** est une maison d’édition gérant une librairie en ligne, des boutiques physiques et une activité de publicité en ligne pour ses auteurs.  
L’architecture analytique repose exclusivement sur **Microsoft Fabric**.

**🏗️ Environnement technique :**

* Workspace Fabric : **Workspace1** (haute concurrence activée)
* Langage de traitement : **Python**
* Architecture **medallion** :
  + Bronze & Silver : **Delta tables** créées depuis **Parquet** via notebooks
  + Gold : dans un **Warehouse** (V-Order désactivé)

**📦 Données :**

* **ERP** : reçoit logs et ventes (parquet → bronze)
* **Book covers** : images dans Azure Blob → chargées dans Files/
* **Ventes** :
  + mises à jour toutes les 6h, historisées
  + capturées via **Dataflow Gen1**
  + stockées dans AuthorSales (colonne AuthorEmail)
* **Sécurité** : auteurs = invités, authentifiés par email

**⚙️ Problèmes rencontrés :**

* **Lenteurs** ou **échecs** de requêtes ad-hoc SQL sur le warehouse
* **Pics d’activité** (nouveaux livres) = ingestion ralentie
* **Décalage** des données de vente visibles le matin

**✅ Évolutions prévues :**

* Ajout des **book reviews** (depuis **Amazon S3**) → sans copie des fichiers
* Ingestion **NRT (near-real-time)** des données **SEO** depuis une **API REST**
* Traitement immédiat des **nouvelles images** dans le Files/ du lakehouse
* Mise en place du **version control avec GitHub**, selon le **principe du moindre privilège**
* **Aucune ressource Azure externe autorisée** (Fabric only)

# Question 1

Case Study Contoso

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

What should you do?

**Propositions**

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.

### Répondu

**C. Share the lakehouse with the DataAnalysts group and grant the Read all SQL Endpoint data permission.**

**✅ Justification :**

* Le **gold layer** est dans **Lakehouse2**, qui est probablement exposé via un **SQL Endpoint**.
* Les **Data Analysts** ont besoin d’un accès **lecture uniquement**.
* L’option **C** donne un accès en lecture via SQL, ce qui correspond à leurs compétences (T-SQL) et à l’architecture cible.

**❌ B. *Share the lakehouse with the DataAnalysts group and grant the Build reports on the default semantic model permission***

* **Pourquoi c’est incorrect :**
  + Cette permission concerne **Power BI** et le **modèle sémantique**, pas directement l’accès aux données du **lakehouse**.
  + Les Data Analysts doivent accéder aux **données du gold layer**, pas seulement au modèle sémantique.
  + Cette permission ne garantit **pas l’accès aux données SQL Endpoint** du lakehouse.

**❌ D. *Share the lakehouse with the DataAnalysts group and grant the Read all Apache Spark permission***

* **Pourquoi c’est incorrect :**
  + Cette permission donne accès à **Apache Spark**, donc à des **notebooks** et à des **ressources de calcul**.
  + Cela va **au-delà de la lecture simple** et n’est **pas nécessaire** pour les Data Analysts.
  + Risque de **non-conformité** avec la politique de sécurité (lecture uniquement).

# Question 2

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?

**Propositions**

A.STORED PROCEDURE

B.CONSTRAINT

C.SCHEMA

D.FUNCTION

### Répondu

**✅ D. FUNCTION**

* Dans **Microsoft Fabric Warehouse**, la **RLS** est implémentée à l’aide de **fonctions inline table-valued (TVF)**.
* Ces fonctions retournent un ensemble de lignes filtrées selon l’utilisateur connecté.
* Ensuite, une **security policy** est créée pour appliquer cette fonction à une table cible.

**🔍 Exemple :**

**CREATE FUNCTION fn\_securitypredicate(@UserName AS sysname)**

**RETURNS TABLE**

**WITH SCHEMABINDING**

**AS**

**RETURN SELECT 1 AS fn\_securitypredicate\_result**

**WHERE @UserName = USER\_NAME(); -- ou une logique plus complexe**

**CREATE SECURITY POLICY SalesRLS**

**ADD FILTER PREDICATE dbo.fn\_securitypredicate(UserName)**

**ON dbo.SalesTable**

**WITH (STATE = ON);**

**❌ A. STORED PROCEDURE**

* Les procédures stockées sont utilisées pour encapsuler de la logique métier ou des traitements, **pas pour appliquer RLS**.

**❌ B. CONSTRAINT**

* Les contraintes (CHECK, FOREIGN KEY, etc.) sont des règles d’intégrité **au niveau des colonnes ou des lignes**, mais **ne permettent pas de filtrer dynamiquement les lignes selon l’utilisateur**.

**❌ C. SCHEMA**

* Un **schéma** est un conteneur logique pour des objets de base de données (tables, vues, etc.), **pas un mécanisme de sécurité dynamique**.

# Question 3

You have a Fabric workspace named Workspace1\_DEV that contains the following items:

10 reports

Four notebooks -

Three lakehouses -

Two data pipelines -

Two Dataflow Gen1 dataflows -

Three Dataflow Gen2 dataflows -

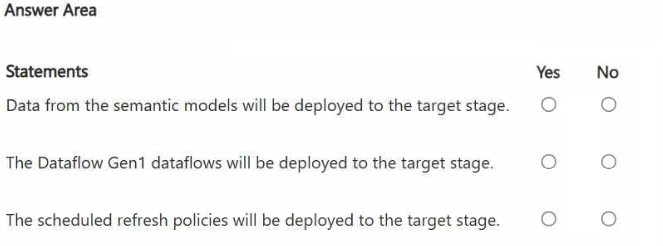
Five semantic models that each has a scheduled refresh policy

You create a deployment pipeline named Pipeline1 to move items from Workspace1\_DEV to a new workspace named Workspace1\_TEST.

You deploy all the items from Workspace1\_DEV to Workspace1\_TEST.

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

**Propositions**



1. **Data from the semantic models will be deployed to the target stage Y/N ?**
2. **The dataflow Gen1 dataflows will be deployed to the target stage Y/N ?**
3. **The schedules refresh policies will be deployed to the target stage Y/N ?**

### Répondu

1. ❌ Data from the semantic models will be deployed to the target stage: No ( Data : no)
2. ❌ The Dataflow Gen1 dataflows will be deployed to the target stage: No ( Gen1 : Obsolete non supportés)
3. ✅ The scheduled refresh policies will be deployed to the target stage: Yes (Si configurés mais oui )

# Question 4

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?

**Propositions**

A.GitHub Actions

B.a deployment pipeline

C.an Azure DevOps pipeline

### Répondu

✅ B.a deployment pipeline

Idiot !

* La question **présuppose** que tu as déjà un pipeline, mais elle te demande **ce que tu dois utiliser pour déployer un objet avec ce pipeline**.
* C’est comme si on te disait : *"Tu as déjà l’outil, que dois-tu utiliser ?"* → ben… **l’outil !**

# Question 5

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?

**Propositions**

A.a database project

B.a deployment pipeline

C.a Python script

D.a T-SQL script

### Répondu

**✅B. a deployment pipeline**

**👉 Très bonne question, et elle peut prêter à confusion si on ne connaît pas bien les outils de Microsoft Fabric.**

**✔ Pourquoi ?**

* Les **deployment pipelines** dans Microsoft Fabric incluent une **fonction de validation automatique**.
* Lorsqu’on prépare une étape de déploiement (ex. de Dev vers Test), Fabric **analyse les objets** (datasets, lakehouses, warehouses, etc.) pour détecter :
  + des **références cassées** (ex. table supprimée, vue invalide),
  + des **dépendances manquantes**,
  + des **erreurs de compatibilité**.
* Cela se fait **sans écrire de code**, donc **effort minimal**.

👉 Quand tu utilises un **deployment pipeline** pour passer un objet (comme un warehouse, un lakehouse, un semantic model, etc.) de **Dev → Test → Prod**, Fabric effectue une **validation automatique** qui peut détecter :

* Des **références cassées** (ex. une vue qui pointe vers une table supprimée).
* Des **dépendances non résolues** (ex. un rapport qui dépend d’un dataset absent).
* Des **boucles de référence** (ex. A → B → A).
* Des **objets non compatibles** avec le workspace cible

Tu peux voir des messages comme :

* "Import failure: RequestValidationFailed. 'WorkspaceId' cannot be null"
* "Deployment failed due to unresolved references in object XYZ"
* "Circular reference detected between items A and B"

Ces erreurs apparaissent **dans l’interface du pipeline**, souvent sous forme de **bannières rouges** ou de **notifications dans le panneau de déploiement**.

# Question 6

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?

**Propositions**

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

### Répondu

**✅ E. only new data added to the eventhouse**

**✔ Explication :**

* Quand tu **actives la disponibilité OneLake** pour un **eventhouse**, cela **n'affecte pas rétroactivement les données déjà présentes**.
* Seules les **nouvelles données ingérées** dans l’eventhouse **après l’activation** seront **automatiquement exposées dans OneLake**.
* Cela permet de **minimiser les coûts et les duplications**, en ne synchronisant que les données pertinentes à partir du moment de l’activation.

# Question 7

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?

**Propositions**

A.service principal

B.Microsoft Entra username and password

C.managed private endpoint

D.workspace identity

### Répondu

**✅ A. service principal**

**✔ Pourquoi ?**

* Un **service principal** est l’approche recommandée pour permettre à des outils externes (comme Azure DevOps) d’accéder de manière sécurisée à des ressources Microsoft Fabric via l’**API REST**.
* Il permet une **authentification sans interaction humaine**, ce qui est parfait pour les **pipelines CI/CD**.
* Tu peux lui attribuer des **permissions précises** sur les workspaces Fabric (ex. contributeur, lecteur, etc.).

**❌ Pourquoi les autres options ne conviennent pas :**

* **B. Microsoft Entra username and password**  
  → Non recommandé pour les automatisations. Moins sécurisé, nécessite souvent MFA, et ne respecte pas les bonnes pratiques DevOps.
* **C. managed private endpoint**  
  → Utilisé pour sécuriser les connexions réseau entre services (ex. Fabric ↔ Azure SQL), **pas pour l’authentification API**.

- Il permet à Fabric d’accéder à ces sources via le réseau privé Azure, sans passer par Internet.

- Cela améliore la sécurité, la latence, et la conformité.

- C’est particulièrement utile dans des environnements sensibles ou réglementés

* **~~D. workspace identity~~** ~~→ Ce concept n’existe pas tel quel dans Fabric. Il n’y a pas d’“identité de workspace” utilisable pour l’authentification API.~~

👉

* **Service principal** : côté Azure, pour **accéder à Fabric** depuis **Azure DevOps, GitHub, scripts, etc.**
* **Managed identity** : côté Azure, pour **autoriser une ressource Azure** (ex. Azure Function) à **accéder** **à Fabric (Model, Data, artefac..) sans gérer de secrets.**

# Question 8

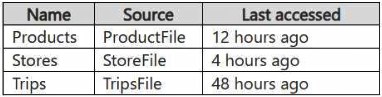
You have a Google Cloud Storage (GCS) container named storage1 that contains the files shown in the following table.

|  |  |
| --- | --- |
| Nom du fichier | Taille |
| 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.

|  |  |  |
| --- | --- | --- |
| Nom du raccourci | Fichier source | Dernier accès |
| 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?

**Propositions**

A.Stores only

B.Products only

C.Stores and Products only

D.Products, Stores, and Trips

E.Trips only

F.Products and Trips only

### Répondu

**✅ F. Products and Trips only**

👁️‍🗨️C’est la TAILLE qui compte !

* Le cache OneLake fonctionne **uniquement pour les fichiers de moins de 100 MB**.

👉 **PRECISIONS**

✅ **Un shortcut est toujours exposé via OneLake**

**1 Un shortcut implique-t-il forcément OneLake ?**

**Oui**, dans le contexte de Microsoft Fabric, un **shortcut** est **un lien logique vers un fichier ou un dossier externe**, exposé **dans OneLake** comme s’il faisait partie du lakehouse.

* Il peut pointer vers :
  + Un autre **lakehouse Fabric**
  + Un **stockage externe** (Azure Data Lake, Amazon S3, Google Cloud Storage)
* Le shortcut **n’importe pas les données**, il les **référence**.
* **OneLake** est le **mécanisme d’unification** qui permet de naviguer dans ces données comme si elles étaient locales.

✅ **Le cache dépend de la taille**

**2 Le cache dépend-il uniquement de la taille ?**

**Oui, principalement.**  
Le **cache OneLake pour les shortcuts** fonctionne selon une règle simple :

|  |  |
| --- | --- |
| Critère | Impact sur le cache |
| Taille du fichier < 100 MB | ✅ Eligible au cache |
| Taille ≥ 100 MB | ❌ Non mis en cache |
| Fréquence ou heure d’accès | ❌ N’a aucun impact sur l’éligibilité |
| Type de fichier | ✅ Tous types (parquet, csv, json...) |

Donc :

✅ **Seule la taille détermine si le fichier est mis en cache.**  
❌ **La fréquence ou l’heure de requêtage ne jouent aucun rôle.**

ℹ️⚠️ Précision Cache :

**✅ Le cache fonctionne uniquement si :**

* La **source** est **Azure Data Lake Storage Gen2 (ADLS Gen2)**
* Le **format** est compatible (Delta, Parquet, etc.)
* La **taille** du fichier est < 100Mo
* Le **cache est activé** dans le workspace

**❌ Le cache ne fonctionne pas si :**

* La source est **Amazon S3**, **Google Cloud Storage**, ou autre **non-Azure**
* Le format est non pris en charge
* Le cache n’est pas activé dans le workspace

# Question 9

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?

**Propositions**

A.an on-premises data gateway

B.a managed private endpoint

C.an integration runtime

D.a data management gateway

### Répondu

**✅ B. a managed private endpoint**

👁️‍🗨️ Car : public internet access disabled.

**👉 Managed Private Endpoint**. : Ce type d’endpoint établit une **connexion réseau privée** entre Fabric et la ressource cible **sans exposer la base de données à Internet**.

❌ A. On-premises data gateway Utilisée pour accéder à des données locales (on-premises), pas pour Azure SQL

❌ C. Integration runtime Utilisé dans Azure Data Factory ou Synapse Pipelines, pas directement dans Fabric pour Spark

❌ ~~D. Data management gateway : Ancienne technologie, remplacée par les~~ **~~integration runtimes~~**

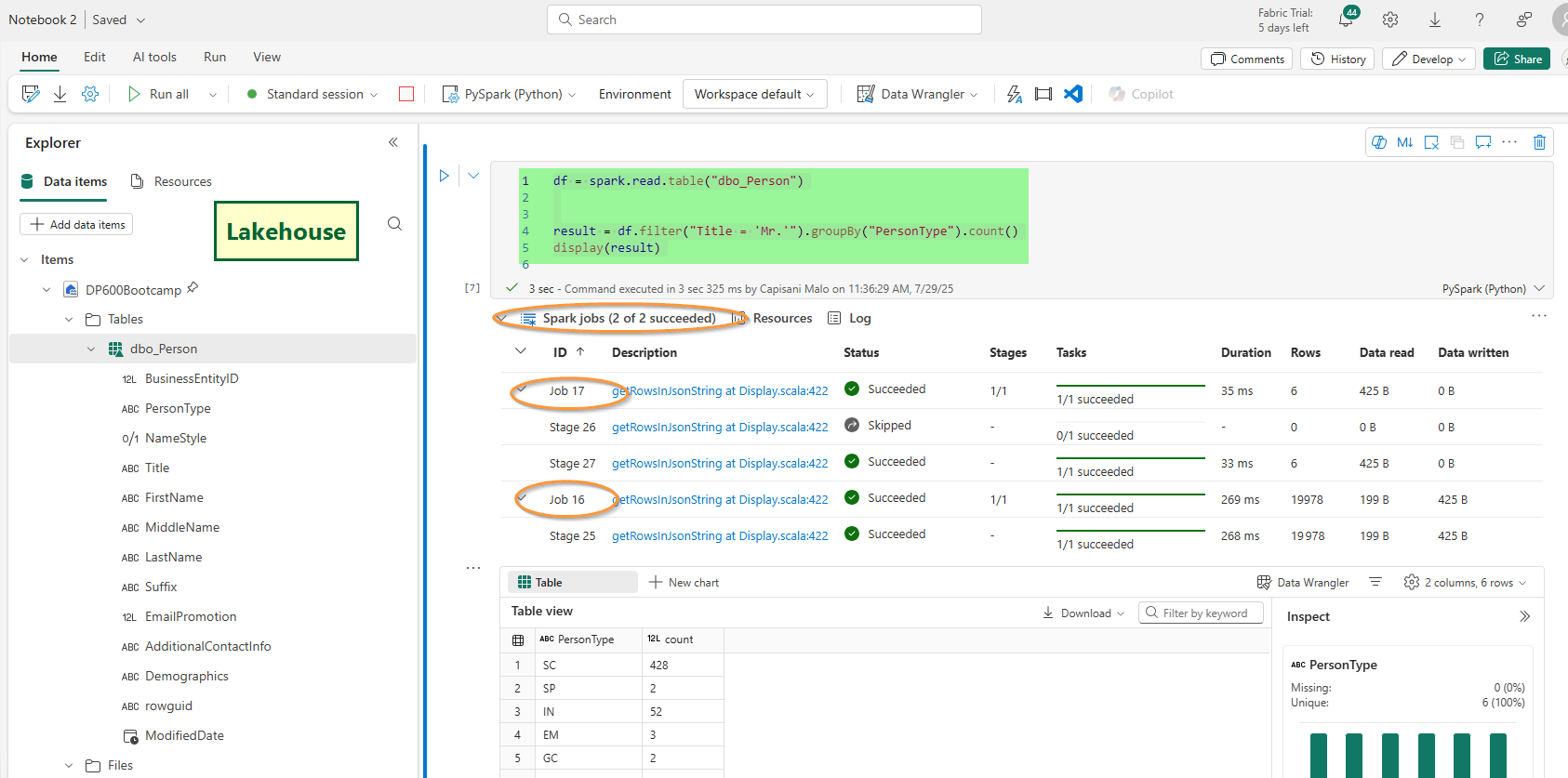
Précision : Apache Spark job

Dans le contexte de **Apache Spark**, le mot "Apache" est souvent **implicite**. Quand on parle de :

* **Spark Job**
* **Spark Session**
* **Spark Cluster**
* **Spark SQL**

… on fait bien référence à **Apache Spark**, le framework open source distribué pour le traitement de données massives.

Exemple :

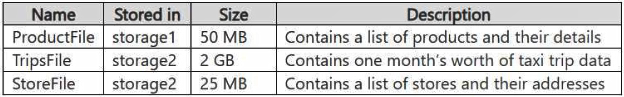


# Question 10

You have an Azure Data Lake Storage Gen2 account named storage1 and an Amazon S3 bucket named storage2.

You have the Delta Parquet files shown in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Stored in | Size | Description |
| ProductFile | Storage1 | 50 MB | Contains a list of products and their details |
| TripsFile | Storage2 | 500 MB | Contains one month’s worth of taxi trip data |
| StoredFile | Storage2 | 99 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

**Propositions**

A.Trips and Stores only

B.Products and Store only

C.Stores only

D.Products only

E.Products, Stores, and Trips

### Répondu

Tres bonne question piege

**✅ D. Products only**

Piège : **Le cache pour les raccourcis** dans Fabric **ne fonctionne que pour les sources Azure** (comme ADLS Gen2).  
Les **sources externes** comme **Amazon S3** **ne sont pas mises en cache**. Et Storage2 est sur S3 !

ℹ️⚠️ Précision Cache :

**✅ Le cache fonctionne uniquement si :**

* La **source** est **Azure Data Lake Storage Gen2 (ADLS Gen2)**
* Le **format** est compatible (Delta, Parquet, etc.)
* La **taille** du fichier est < 100Mo
* Le **cache est activé** dans le workspace

**❌ Le cache ne fonctionne pas si :**

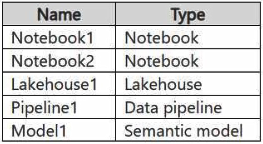
* La source est **Amazon S3**, **Google Cloud Storage**, ou autre **non-Azure**
* Le format est non pris en charge
* Le cache n’est pas activé dans le workspace

# Question 11

HOTSPOT -

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

|  |  |
| --- | --- |
| Name | Type |
| Notebook1 | Notebook |
| Notebook2 | Notebook |
| Lakehouse1 | Lakehouse |
| Pipeline1 | Data pipeline |
| Model1 | Semantic model |



For Model1, the “Keep your Direct Lake data up to date” option is disabled.

You need to configure the execution of the items to meet the following requirements:

Notebook1 must execute every weekday at 8:00 AM.

Notebook2 must execute when a file is saved to an Azure Blob Storage container.

Model1 must refresh when Notebook1 has executed successfully.

How should you orchestrate each item? To answer, select the appropriate options in the answer area.

**Propositions**

Notebook1:

* Add Notebook1 to an Apache Spark job definition.
* Add Notebook1 to Pipeline1.
* From Real-Time hub, configure the execution of Notebook1.

Notebook2:

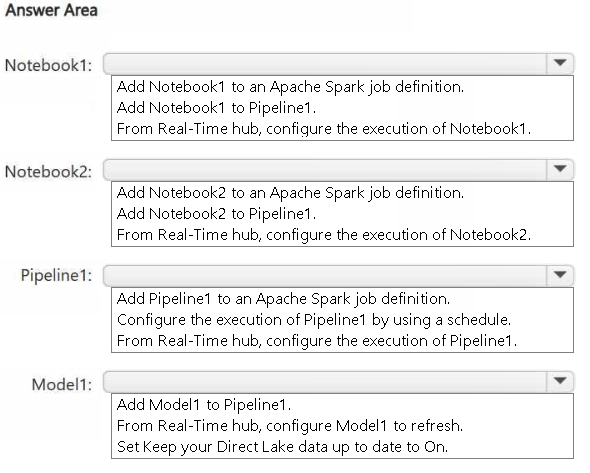
* Add Notebook2 to an Apache Spark job definition.
* Add Notebook2 to Pipeline1.
* From Real-Time hub, configure the execution of Notebook2.

Pipeline1:

* Add Pipeline1 to an Apache Spark job definition.
* Configure the execution of Pipeline1 by using a schedule.
* From Real-Time hub, configure the execution of Pipeline1.

Model1:

* Add Model1 to Pipeline1.
* From Real-Time hub, configure Model1 to refresh.
* Set "Keep your Direct Lake data up to date" to On.



### Répondu

Notebook1:

* Add Notebook1 to Pipeline1.

Notebook2:

* From Real-Time hub, configure the execution of Notebook2.

Pipeline1:

* Configure the execution of Pipeline1 by using a schedule.

Model1:

* Add Model1 to Pipeline1.

**✅ Réponses aux 4 propositions**

**🔹 Notebook1**

**Objectif** : Exécution chaque jour de semaine à 8h00  
**Bonne réponse** : ✅ **Ajouter Notebook1 à Pipeline1**  
**Justification** : La planification horaire se fait via un pipeline. Le notebook est intégré dans le pipeline, qui est ensuite planifié. A 8h00

**🔹 Notebook2**

**Objectif** : Exécution lorsqu’un fichier est ajouté dans Azure Blob Storage  
**Bonne réponse** : ✅ **Configurer l’exécution de Notebook2 via Real-Time hub**  
**Justification** : Le déclenchement basé sur un événement externe (ajout de fichier) est géré par le Real-Time hub. Notebook2 est **indépendant du pipeline** dans ce scénario.

**🔹 Pipeline1**

**Objectif** : Orchestration de Notebook1 et Model1  
**Bonne réponse** : ✅ **Configurer l’exécution de Pipeline1 par planification**  
**Justification** : Le pipeline est planifié pour s’exécuter chaque jour à 8h00, ce qui déclenche Notebook1 et ensuite Model1.

**🔹 Model1**

**Objectif** : Rafraîchissement après exécution de Notebook1  
**Bonne réponse** : ✅ **Ajouter Model1 à Pipeline1**  
**Justification** : L’option "Keep your Direct Lake data up to date" est désactivée, donc le modèle doit être rafraîchi manuellement via le pipeline, après Notebook1.

⚠️👉‼️

**🔄 Si "Keep your Direct Lake data up to date" est activée :**

**✅ Comportement :**

* **Model1** se **met automatiquement à jour** dès que les données du Lakehouse changent.
* Il **n’est plus nécessaire de le rafraîchir manuellement** via un pipeline ou une autre orchestration.

**🧠 Conséquence sur la réponse à la question :**

|  |  |
| --- | --- |
| Élément | Orchestration recommandée |
| Notebook1 | ✅ Ajouter à Pipeline1 (planifié à 8h00) |
| Notebook2 | ✅ Configurer via Real-Time hub |
| Pipeline1 | ✅ Configurer par planification |
| Model1 | ✅ Activer "Keep your Direct Lake data up to date" (aucune orchestration nécessaire) |

**❌ Donc , dans ce cas : il n’est plus nécessaire de :**

* **ajouter Model1 au pipeline**, car il se mettra à jour automatiquement.
* **configurer de déclencheur manuel** pour Model1.

# Question 12

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?

**Propositions**

A.a lakehouse

B.an eventhouse

C.a datamart

D.a warehouse

### REPONSE

# Question 13

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?

**Propositions**

A.workspace Admin

B.domain admin

C.domain contributor

D.Fabric admin

### Répondu

✅ **C. Domain Contributor**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rôle | Créer un domaine | Créer un sous-domaine | Assigner un workspace | Accès aux artefacts | Gérer les rôles du domaine |
| Fabric Admin | ✅ Oui | ✅ Oui | ✅ Oui | ❌ (sauf si aussi membre du WS) | ✅ Oui |
| Domain Admin | ✅ Oui | ✅ Oui | ✅ Oui | ❌ | ✅ Oui |
| Domain Contributor | ❌ Non | ✅ Oui | ✅ Oui | ❌ | ❌ Non |
| Workspace Admin | ❌ Non | ❌ Non | ❌ Non | ✅ Oui | ✅ (dans le WS uniquement) |

RAPPEL

|  |  |  |
| --- | --- | --- |
| Niveau | Responsabilité | Accès aux données ? |
| Domaine | Structuration, gouvernance, métadonnées, affectation ( et Création, modification, partage des artefacts ) | ❌ Non |
| Workspace | Création, modification, partage des artefacts | ✅ Oui (selon le rôle) |

🎯 Pourquoi choisir l’un plutôt que l’autre ?

✅ Domain Contributor (le plus adapté dans ton cas)

Avantages :

Suffisant pour créer des domaines et sous-domaines.

Peut affecter des workspaces aux domaines.

Respecte le principe du moindre privilège.

Limites :

Ne peut pas gérer les utilisateurs du domaine.

Ne peut pas modifier les paramètres avancés du domaine.

🔒 Domain Admin

Avantages :

Contrôle total sur un domaine spécifique.

Peut ajouter/retirer des contributeurs, modifier les paramètres.

À utiliser si : la personne doit administrer un domaine (et pas seulement le structurer).

🌐 Fabric Admin

Avantages :

Accès complet à toute la plateforme Fabric.

Inconvénients :

Trop de privilèges pour une tâche ciblée.

À éviter sauf si la personne est responsable de l’administration globale.

🧱 Workspace Admin

Avantages :

Gère le contenu d’un workspace (datasets, rapports, pipelines…).

Limites :

Aucun droit sur les domaines.

**👉 À utiliser si : la personne travaille uniquement dans un workspace.**

# Question 14

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?

**Propositions**

A.Admin

B.Member

C.Viewer

D.Contributor

### Répondu

✅ D.Contributor

**Rem** : Les roles FABRICS ont chagés par raport à POWERBI Classique

👁️‍🗨️ Phrases clef : “Update the tables in DW1”. + “You already assigned the appropriate object-level permissions to DW1”.

**pour qu’un utilisateur puisse modifier les tables d’un warehouse (DW1) :**

* Il doit :
  + ✅ Avoir **les permissions SQL appropriées dans DW1**
  + ✅ Et être au minimum **Contributor** dans le workspace (pour pouvoir accéder à DW1 dans l’interface)

RAPPEL IMPORTANT :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Artefact | Viewer | Contributor | Member | Admin | Remarques |
| Rapport | Voir | Voir, Modifier | Voir, Modifier, Exécuter | Tout accès |  |
| Dataset | Voir | Voir, Modifier | Voir, Modifier, Exécuter | Tout accès |  |
| Pipeline | Voir | Voir, Modifier | Voir, Modifier, Exécuter | Tout accès |  |
| Notebook | Voir | Voir, Modifier | Voir, Modifier, Exécuter | Tout accès |  |
| Warehouse | Voir | Voir, Modifier\* | Voir, Modifier, Exécuter\* | Tout accès | \* Requiert des permissions au niveau objet (ex. SELECT, UPDATE sur les tables) |
| Lakehouse | Voir | Voir, Modifier\* | Voir, Modifier, Exécuter\* | Tout accès | \* Requiert des permissions au niveau objet (ex. accès aux fichiers ou tables) |

# Question 15

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?

**Propositions**

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.

### Répondu

**✅ A. Share Lakehouse1 with User1 directly and select "Read all SQL endpoint data**

* ✔️ Donne accès **uniquement à la couche SQL** (tables, vues).
* ❌ Ne donne **pas accès à Spark**.
* ✔️ Ne donne **pas accès aux autres artefacts** du workspace.
* ✅ **C’est la bonne réponse** : elle respecte le **principe du moindre privilège**.

# Question 16

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.

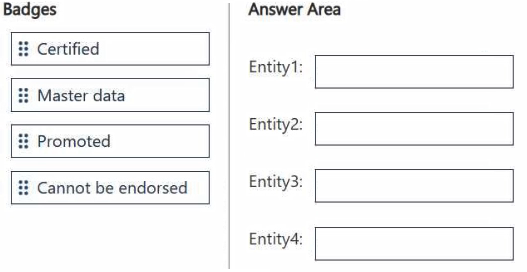
Which badge should you apply to each entity?

Each badge may be used once, more than once, or not at all.

**Propositions**

* Certified
* Master Data
* Promoted
* Cannot be endorsed

1. **Entity1 :**
2. **Entity2 :**
3. **Entity3 :**
4. **Entity4 :**



### Répondu

1. Entity1 Master Data
2. Entity2 Certified
3. Entity3 Promoted
4. Entity4 Cannot be endorsed

👁️‍🗨️ Phrases clef :

« contains data that will be used as a core organization entity”

“Available in a Power BI dashboard”

RAPPEL : **Rappel des badges disponibles**

|  |  |
| --- | --- |
| BADGE | SIGNIFICATION |
| Certified | Données officiellement validées par l’organisation, à usage critique. |
| Promoted | Données recommandées pour un usage général, mais pas encore certifiées. |
| Master Data | Données de référence centrale utilisées dans plusieurs domaines métiers. |
| Cannot be endorsed | Artefact non éligible à l’endossement (ex. : dashboards, notebooks). |

**1. Entity1 : Disponible dans un lakehouse, utilisé comme entité centrale de l’organisation**

* 🟢 **Badge approprié** : **Master Data**
* 📌 Justification : C’est une **source de vérité** utilisée dans plusieurs domaines.

**2. Entity2 : Disponible dans un semantic model, conforme aux standards de l’organisation**

* 🟢 **Badge approprié** : **Certified**
* 📌 Justification : Conforme aux normes, donc **officiellement validé**.

**3. Entity3 : Disponible dans un rapport Power BI, prêt à être partagé et réutilisé**

* 🟢 **Badge approprié** : **Promoted**
* 📌 Justification : Recommandé pour usage, mais pas nécessairement certifié.

**4. Entity4 : Disponible dans un dashboard Power BI, utilisé pour la décision exécutive**

* 🔴 **Badge approprié** : **Cannot be endorsed**
* 📌 Justification : Les **dashboards ne peuvent pas être endossés** dans Fabric

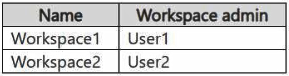
# Question 17

HOTSPOT -

You have three users named User1, User2, and User3.

You have the Fabric workspaces shown in the following table.

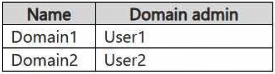
|  |  |
| --- | --- |
| Name | Workspace Admin |
| Workspace1 | User1 |
| Workspace2 | User2 |



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 |
| Domain1 | User1 |
| Domain2 | User2 |



User1 creates a new workspace named Workspace3.

You add Group1 to the default domain of Domain1.

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

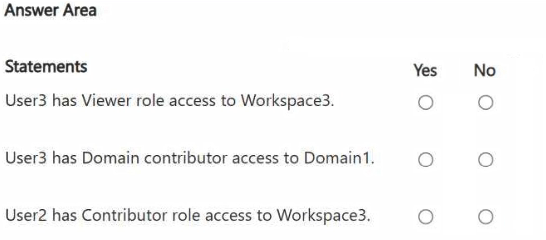
NOTE: Each correct selection is worth one point.

**Propositions**

User3 has viewer role access to Workspace3 Y/N

User3 has Domain contributor access to domain1 Y/N

User3 has contributor role access to Workspace3 Y/N



### Répondu

User3 has viewer role access to Workspace3 ❌ Non

User3 has Domain contributor access to Domain1 ✅ Oui

User3 has contributor role access to Workspace3 ❌ Non

👁️‍🗨️ **piège classique** de ce type de question : **mélanger les rôles dans les domaines et les workspaces**, qui sont **deux couches de gouvernance totalement distinctes** dans Microsoft Fabric.

Ici il faut être focus sur user3 , seul concerné par les questions

Et User3 fait partie de group1

# Question 18

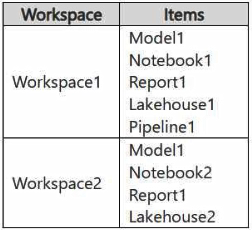
You have two Fabric workspaces named Workspace1 and Workspace2.

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

You recently modified the items in Workspaces1.

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

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



Items in Workspace1 that have the same name as items in Workspace2 are currently paired.

You need to ensure that the items in Workspace1 overwrite the corresponding items in Workspace2. The solution must minimize effort.

**Propositions**

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.

### Répondu

✅ D.Run deployPipeline1 without modifying the items in Workspace2.

# Question 19

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?

**Propositions**

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.

### Répondu

**✅ A. Folder1 is created, Pipeline1 moves to Folder1, and Lakehouse1 is deployed**

**Pourquoi ?**

* Pipeline1 est apparié → il sera **déplacé dans Folder1** dans Workspace2.
* Lakehouse1 est aussi apparié → il sera **déployé** (mis à jour).
* Folder1 est **créé automatiquement** dans Workspace2 pour refléter la structure source.

‼️ 👁️‍🗨️ La structure de dossier (comme Folder1) **n’est déployée que si elle contient un ou plusieurs éléments appariés**.

‼️ ⚠️ 👉

* ❌ **Un dossier vide n’est pas déployé**.
* Les pipelines **ne déploient pas la structure seule** si elle ne contient **aucun artefact à déployer**.

# Question 20

DRAG DROP -

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

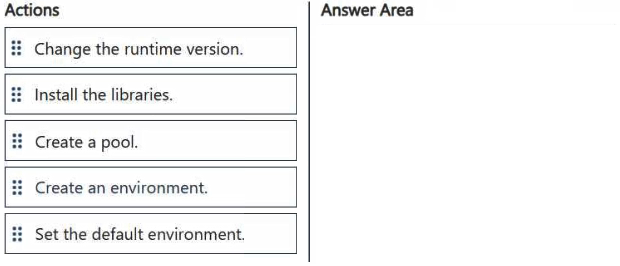
You create a Fabric workspace name Workspace1 that will be used to develop extract, transform, and load (ETL) solutions by using notebooks.

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.

**Propositions**

1. Changes The runtime version
2. Install the libraries
3. Create a Pool
4. Create an environment
5. Set the default environ



### Répondu

Phrase clef : . “The team creates Python libraries of reusable code that is used to transform data” . il faut donc un nouvel environnement pour charger ces libraries

**✅ C. Create a Pool**  
→ Crée une **infrastructure de calcul** (Spark pool) nécessaire pour exécuter des notebooks.

**✅ D. Create an environment**  
→ Crée un **environnement personnalisé** dans lequel tu peux **installer les bibliothèques** nécessaires.

**✅ E. Set the default environment**  
→ Définit cet environnement comme **par défaut** pour tous les nouveaux notebooks du workspace.

**🧠 Pourquoi pas les autres ?**

* **A. Change the runtime version** : option avancée, pas nécessaire pour rendre les bibliothèques disponibles.
* **B. Install the libraries** : cette étape est **incluse dans la création de l’environnement**.

👉 Bon, le **Processus standard pour rendre des bibliothèques disponibles par défaut dans un workspace Fabric :**

1. **Créer un pool de calcul**  
   → C’est la base : il fournit les ressources pour exécuter les notebooks.
2. **Créer un environnement personnalisé**  
   → Tu y ajoutes les bibliothèques spécifiques dont ton équipe a besoin.
3. **Associer l’environnement au pool**  
   → Cela permet au pool d’utiliser les bonnes bibliothèques.
4. **Définir ce pool comme défaut dans le workspace**  
   → Tous les nouveaux notebooks utiliseront automatiquement cet environnement.

Une image contenant texte, capture d’écran, nombre, logiciel

Le contenu généré par l’IA peut être incorrect.

# Question 21

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?

**Propositions**

A.VACUUM

B.BROADCAST

C.OPTIMIZE

D.CACHE

### REPONSE

# Question 22

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?

**Propositions**

A.Start time

B.Capacity

C.Activity name

D.Submitter

E.Item type

F.Job type

G.Location

### REPONSE

# Question 23

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?

**Propositions**

A.a Dataflow Gen1 dataflow

B.a data pipeline

C.a KQL queryset

D.a notebook

### REPONSE

# Question 24

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?

**Propositions**

A.Real-Time hub

B.OneLake data hub

C.the Admin monitoring workspace

D.Fabric Monitor

E.the Microsoft Fabric Capacity Metrics app

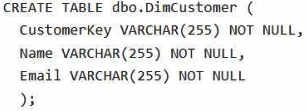
### REPONSE

# Question 25

DRAG DROP -

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

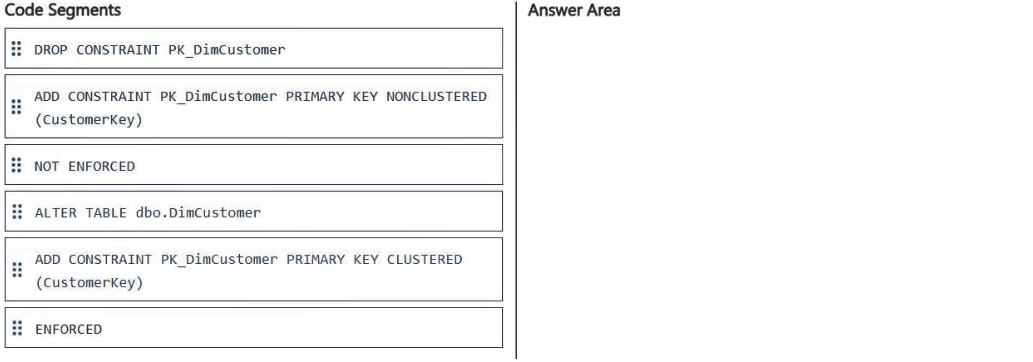
In Warehouse1, you create a table named DimCustomer by running the following statement.



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.

**Propositions**



### REPONSE

# Question 26

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?

**Propositions**

A.dynamic management views in Microsoft SQL Server Management Studio (SSMS)

B.Monitoring hub

C.dynamic management views in Azure Data Studio

D.a semantic link in a notebook

### REPONSE

# Question 27

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 -

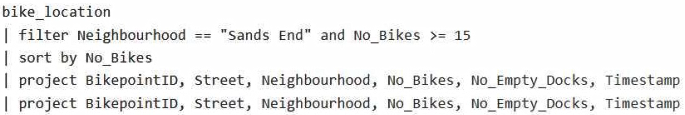
No\_Bikes -

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:



Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 28

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 -

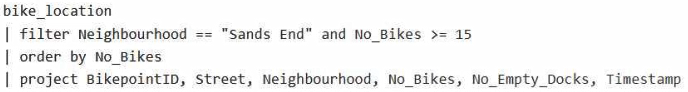
No\_Bikes -

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:



Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 29

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 -

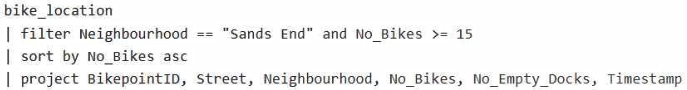
No\_Bikes -

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:



Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 30

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 -

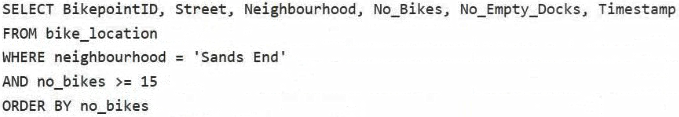
No\_Bikes -

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:



Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 31

Case Study Litware

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

How should you configure the Apache Spark settings?

**Propositions**

A.Disable high concurrency.

B.Create a custom pool.

C.Modify the number of executors.

D.Set the default environment.

### REPONSE

# Question 32

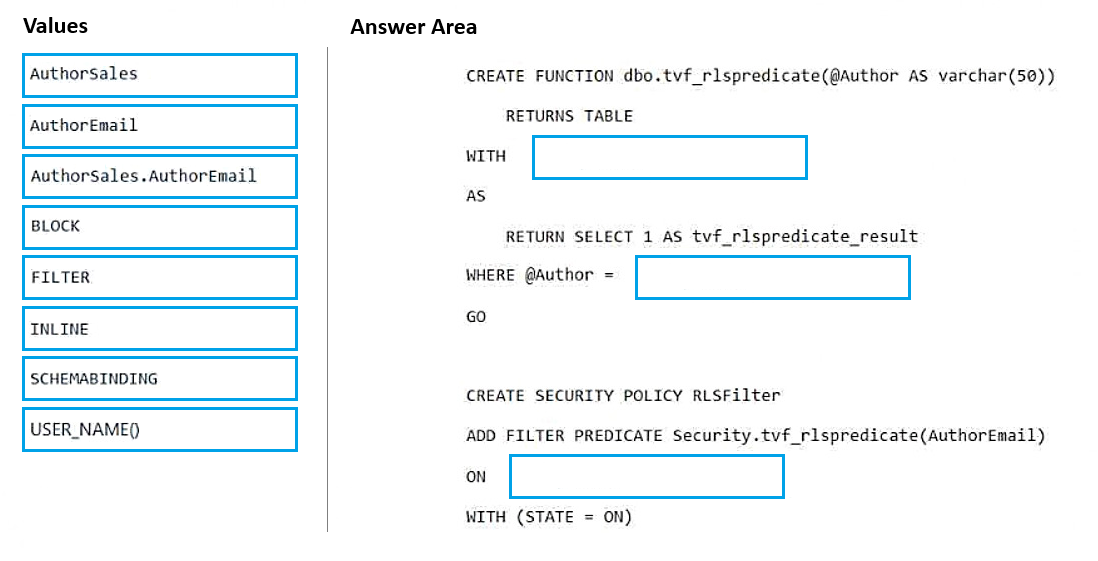
Case Study Litware

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.

**Propositions**



### REPONSE

# Question 33

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

You have a Fabric workspace named Workspace1. Workspace contains a notebook named Notebook1 that performs the following tasks:

• Loads stage data to the target tables in a lakehouse

• Triggers the refresh of a semantic model

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes.

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

Solution: You use the following code segment:

Use notebookutils.credentials.getSecret and specify the key vault URL and key vault secret.

Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 34

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?

**Propositions**

A.an Apache Spark job definition

B.a data pipeline

C.a Dataflow Gen1 dataflow

D.an eventstream

### REPONSE

# Question 35

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

You have a Fabric workspace named Workspace1. Workspace contains a notebook named Notebook1 that performs the following tasks:

• Loads stage data to the target tables in a lakehouse

• Triggers the refresh of a semantic model

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes.

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

Solution: You use the following code segment:

Use notebookutils.credentials.putSecret and specify the key vault URL and key vault secret.

Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 36

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

You have a Fabric workspace named Workspace1. Workspace contains a notebook named Notebook1 that performs the following tasks:

• Loads stage data to the target tables in a lakehouse

• Triggers the refresh of a semantic model

You plan to add functionality to Notebook1 that will use the Fabric API to monitor the semantic model refreshes.

You need to retrieve the registered application ID and secret from KeyVault1 to generate the authentication token.

Solution: You use the following code segment:

Use notebookutils.credentials.getSecret and specify the key vault URL and the name of a linked service.

Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 37

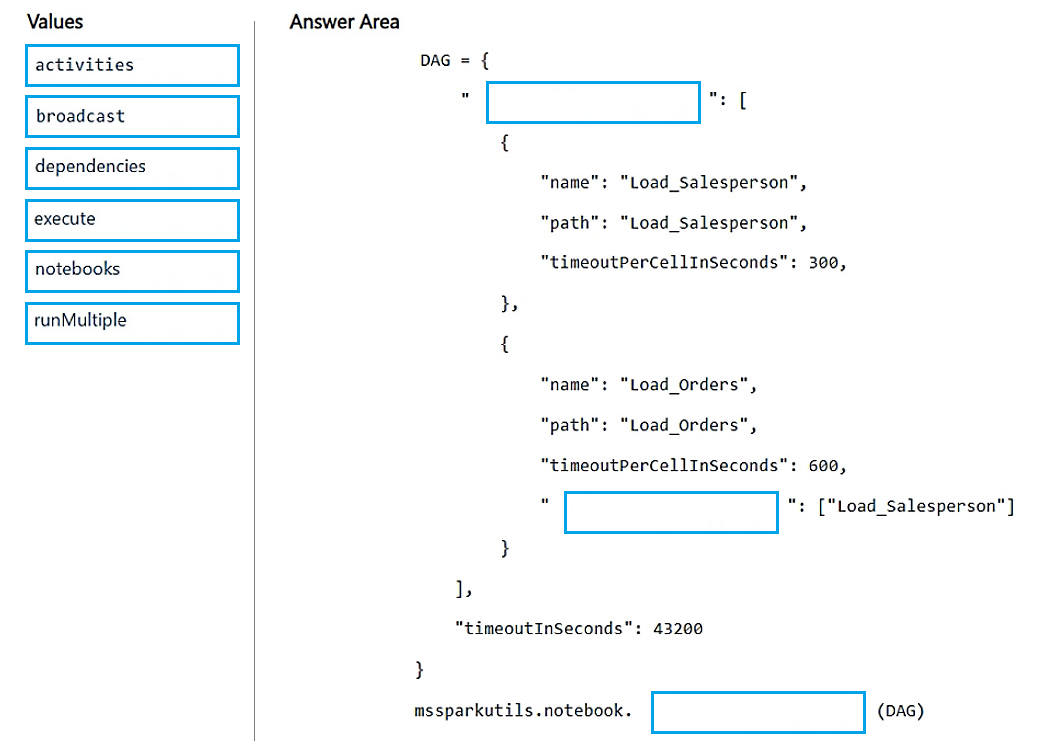
You have two Fabric notebooks named Load\_Salesperson and Load\_Orders that read data from Parquet files in a lakehouse. Load\_Salesperson writes to a Delta table named dim\_salesperson. Load\_Orders writes to a Delta table named fact\_orders and is dependent on the successful execution of Load\_Salesperson.

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.

**Propositions**



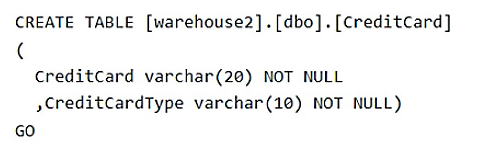
### REPONSE

# Question 38

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.

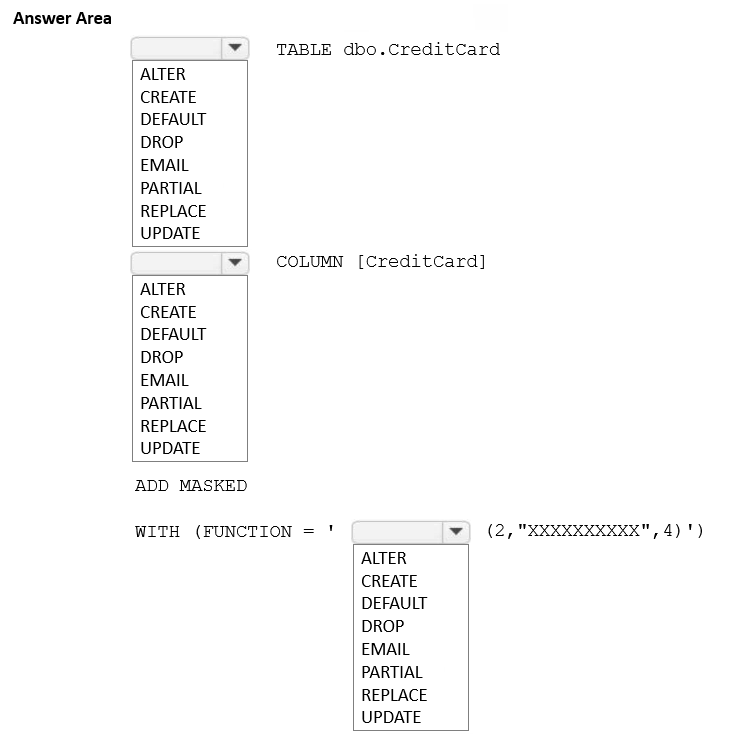


You need to ensure that the team can view only the first two characters and the last four characters of the CreditCard attribute.

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

NOTE: Each correct selection is worth one point.

**Propositions**



### REPONSE

# Question 39

You have a Fabric workspace that contains a warehouse named Warehouse1. Warehouse1 contains a table named DimCustomers. DimCustomers contains the following columns:

• CustomerName

• CustomerID

• BirthDate

• EmailAddress

You need to configure security to meet the following requirements:

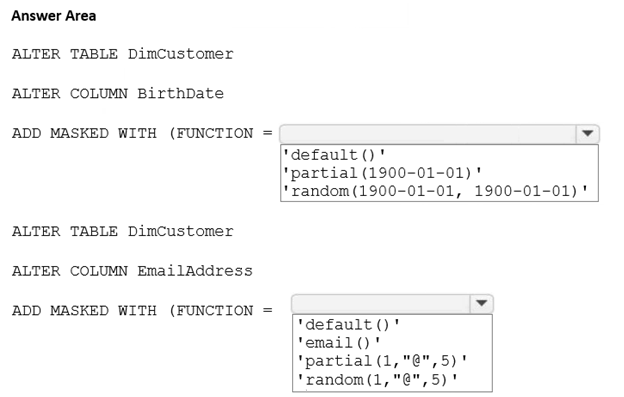
• BirthDate in DimCustomer must be masked and display 1900-01-01.

• EmailAddress in DimCustomer must be masked and display only the first leading character and the last five characters.

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

NOTE: Each correct selection is worth one point.

**Propositions**



### REPONSE

# Question 40

You have a Fabric workspace named Workspace1 that contains the following items:

• A Microsoft Power BI report named Report1

• A Power BI dashboard named Dashboard1

• A semantic model named Model1

• A lakehouse name Lakehouse1

Your company requires that specific governance processes be implemented for the items.

Which items can you endorse in Fabric?

**Propositions**

A.Lakehouse1, Model1, and Dashboard1 only

B.Lakehouse1, Model1, Report1 and Dashboard1

C.Report1 and Dashboard1 only

D.Model1, Report1, and Dashboard1 only

E.Lakehouse1, Model1, and Report1 only

### REPONSE

# Question 41

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?

**Propositions**

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)

### REPONSE

# Question 42

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

You use a notebook named Notebook1 to ingest and transform data from an external data source.

You need to execute Notebook1 as part of a data pipeline named Pipeline1. The process must meet the following requirements:

• Run daily at 07:00 AM UTC.

• Attempt to retry Notebook1 twice if the notebook fails.

• After Notebook1 executes successfully, refresh Model1.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

**Propositions**

A.Place the Semantic model refresh activity after the Notebook activity and link the activities by using the On success condition.

B.From the Schedule settings of Pipeline1, set the time zone to UTC.

C.Set the Retry setting of the Notebook activity to 2.

D.From the Schedule settings of Notebook1, set the time zone to UTC.

E.Set the Retry setting of the Semantic model refresh activity to 2.

F.Place the Semantic model refresh activity after the Notebook activity and link the activities by using an On completion condition.

### REPONSE

# Question 43

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?

**Propositions**

A.Change the MD5 hash to SHA256.

B.Increase the capacity.

C.Enable V-Order.

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

E.Create views.

### REPONSE

# Question 44

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

You plan to create a data pipeline named Pipeline1 to ingest data into Lakehouse1. You will use a parameter named param1 to pass an external value into Pipeline1. The param1 parameter has a data type of int.

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

How should you specify the parameter value?

**Propositions**

A."@pipeline().parameters.param1"

B."@{pipeline().parameters.param1}"

C."@{pipeline().parameters.[param1]}"

D."@@{pipeline().parameters.param1}"

### REPONSE

# Question 45

You have a Fabric workspace named Workspace1 that contains a lakehouse named Lakehouse1. Workspace1 contains the following items:

• A Dataflow Gen2 dataflow that copies data from an on-premises Microsoft SQL Server database to Lakehouse1

• A notebook that transforms files and loads the data to Lakehouse1

• A data pipeline that loads a CSV file to Lakehouse1

You need to develop an orchestration solution in Fabric that will load each item one after the other. The solution must be scheduled to run every 15 minutes.

Which type of item should you use?

**Propositions**

A.notebook

B.warehouse

C.Dataflow Gen2 dataflow

D.data pipeline

### REPONSE

# Question 46

You are building a Fabric notebook named MasterNotebook1 in a workspace. MasterNotebook1 contains the following code.



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

1. Notebook\_03

2. Notebook\_01

3. Notebook\_02

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

**Propositions**

A.Move the declaration of Notebook\_02 to the bottom of the Directed Acyclic Graph (DAG) definition.

B.Add dependencies to the execution of Notebook\_03.

C.Split the Directed Acyclic Graph (DAG) definition into three separate definitions.

D.Add dependencies to the execution of Notebook\_02.

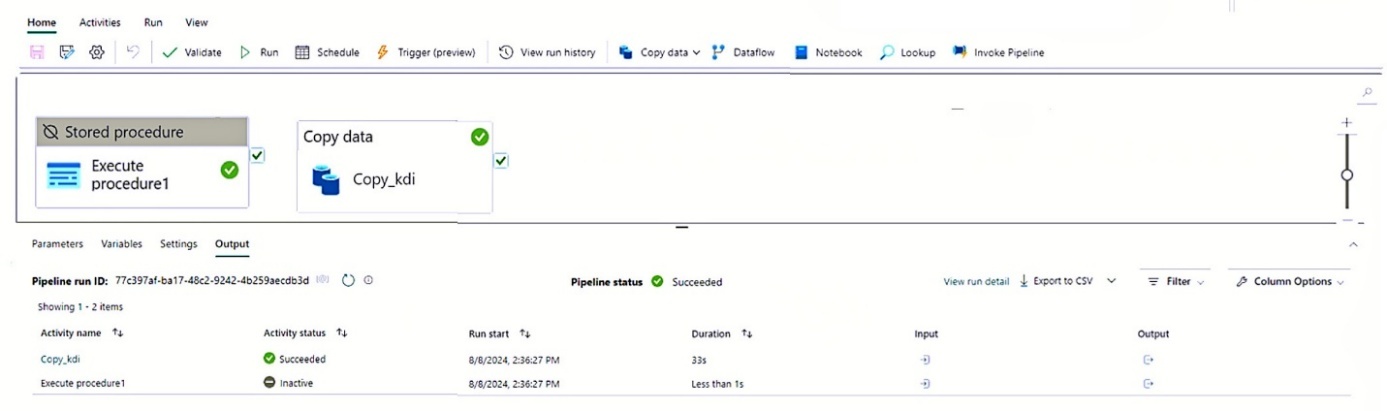
E.Change the concurrency to 3.

F.Move the declaration of Notebook\_03 to the top of the Directed Acyclic Graph (DAG) definition.

### REPONSE

# Question 47

You have a Fabric workspace that contains a data pipeline named Pipeline1 as shown in the exhibit. (Click the Exhibit tab.)



What will occur the next time Pipeline1 runs?

### REPONSE

A.Copy\_kdi will run first, and then Execute procedure1 will run.

B.Execute procedure1 will run first, and then Copy\_kdi will run.

C.Execute procedure1 will run and Copy\_kdi will be skipped.

D.Copy\_kdi will run and Execute procedure1 will be skipped.

E.Both activities will run simultaneously.

F.Both activities will be skipped.

### REPONSE

# Question 48

Case Study Contoso

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.

**Propositions**

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

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

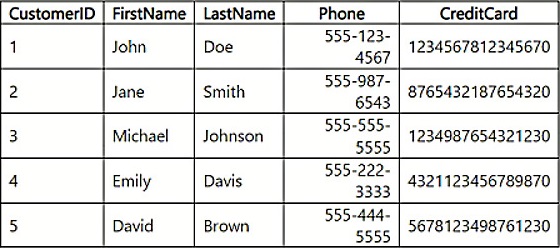
C.Configure WorkspaceA to use a Premium Per User (PPU) license.

D.Assign WorkspaceA to Cap1.

### REPONSE

# Question 49

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

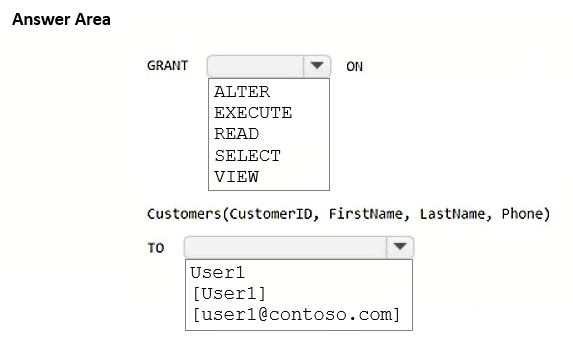


You have an internal Microsoft Entra user named User1 that has an email address of user1@contoso.com.

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.

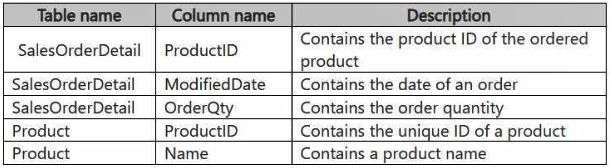
**Propositions**



### REPONSE

# Question 50

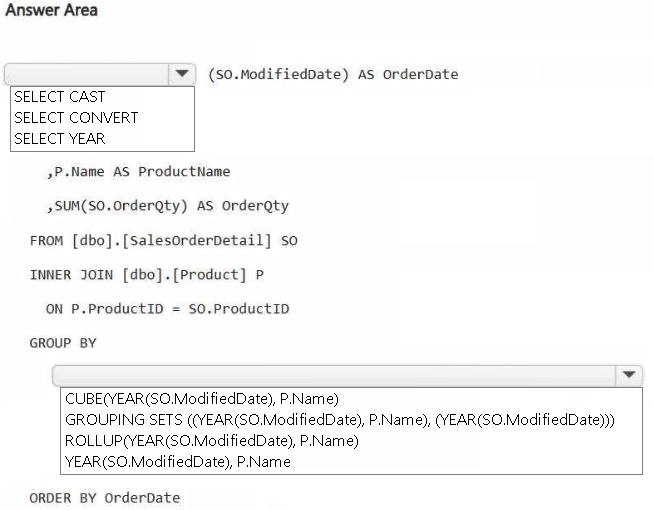
You have a Fabric workspace that contains a warehouse named DW1. DW1 contains the following tables and columns.



You need to create an output that presents the summarized values of all the order quantities by year and product. The results must include a summary of the order quantities at the year level for all the products.

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

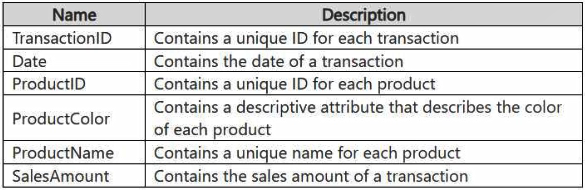
**Propositions**



### REPONSE

# Question 51

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.



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.

**Propositions**

A.Date

B.ProductName

C.ProductColor

D.TransactionID

E.SalesAmount

F.ProductID

### REPONSE

# Question 52

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?

**Propositions**

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.

### REPONSE

# Question 53

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.

**Propositions**

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.

### REPONSE

# Question 54

Case Study Litware

You need to implement the solution for the book reviews.

Which should you do?

**Propositions**

A.Create a Dataflow Gen2 dataflow.

B.Create a shortcut.

C.Enable external data sharing.

D.Create a data pipeline.

### REPONSE

# Question 55

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?

**Propositions**

A.a KQL queryset

B.an eventstream

C.a streaming dataset

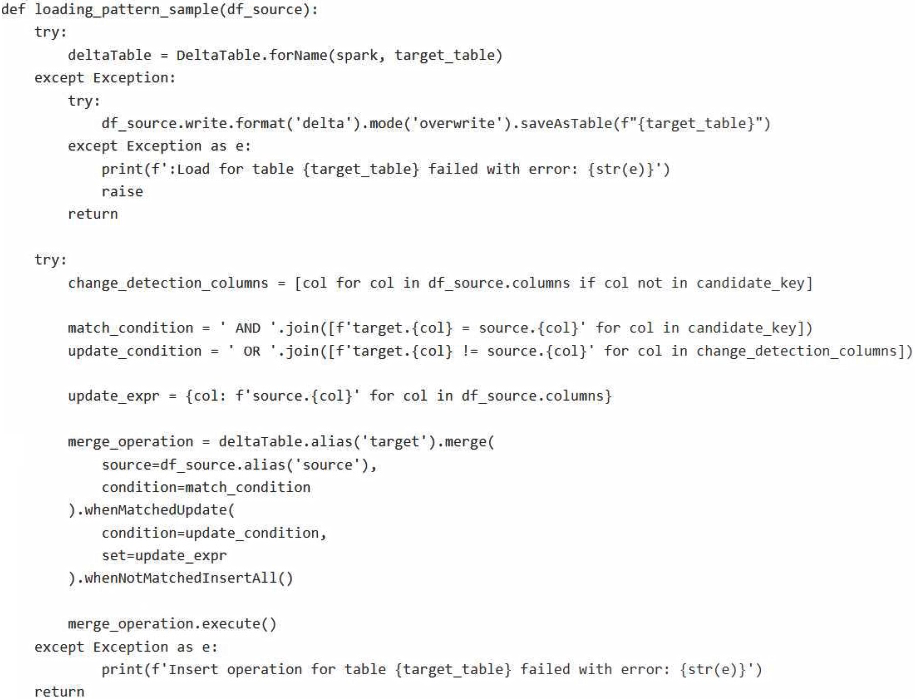
D.Apache Spark Structured Streaming

### REPONSE

# Question 56

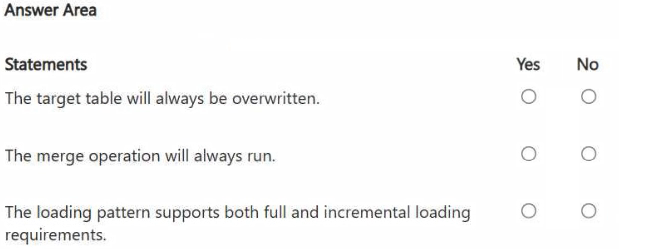
You are building a data loading pattern for Fabric notebook workloads.

You have the following code segment:



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

**Propositions**

****

### REPONSE

# Question 57

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.

**Propositions**

****

### REPONSE

# Question 58 (Idem 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?

**Propositions**

A.Eventstream

B.Dataflow Gen2

C.Streaming dataset

D.Data pipeline

### REPONSE

D. Data pipeline

* Permet d’**orchestrer l’ingestion de fichiers volumineux** (comme 500 Go par fichier).
* **Peut surveiller** un dossier source (comme un stockage ADLS Gen2, Blob, etc.)
* **Peut se déclencher automatiquement** à l’arrivée d’un nouveau fichier (via déclencheurs basés sur le stockage).
* Fournit une **gestion robuste des gros volumes**, avec **haute performance** et parallélisme.
* Le besoin ne nécessite pas de transformation — parfait pour ingestion brute.

**A. Eventstream**

❌ **Incorrect**

* Eventstream est conçu pour **des données en temps réel** et **petits événements** (télémétrie, IoT, logs…).
* **Pas adapté pour des fichiers massifs (500 Go)** ou des traitements par fichier.

**B. Dataflow Gen2**

❌ **Incorrect**

* Dataflow Gen2 est fait pour les **transformations de données**, souvent à l’aide de Power Query.
* Pas idéal pour **ingestion brute** à très haut débit sans transformation.

**C. Streaming dataset**

❌ **Incorrect**

* Conçu pour des **données en streaming** (push en temps réel), notamment dans Power BI.
* **Pas du tout adapté** pour des fichiers batch volumineux comme ici.

# Question 59 (Idem 58)

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?

**Propositions**

A.Data pipeline

B.Environment

C.KQL queryset

D.Dataflow Gen2

### REPONSE

D. Data pipeline

* Permet d’**orchestrer l’ingestion de fichiers volumineux** (comme 500 Go par fichier).
* **Peut surveiller** un dossier source (comme un stockage ADLS Gen2, Blob, etc.)
* **Peut se déclencher automatiquement** à l’arrivée d’un nouveau fichier (via déclencheurs basés sur le stockage).
* Fournit une **gestion robuste des gros volumes**, avec **haute performance** et parallélisme.
* Le besoin ne nécessite pas de transformation — parfait pour ingestion brute.

# Question 60

You have a Fabric workspace that contains an eventhouse and a KQL database named Database1. Database1 has the following:

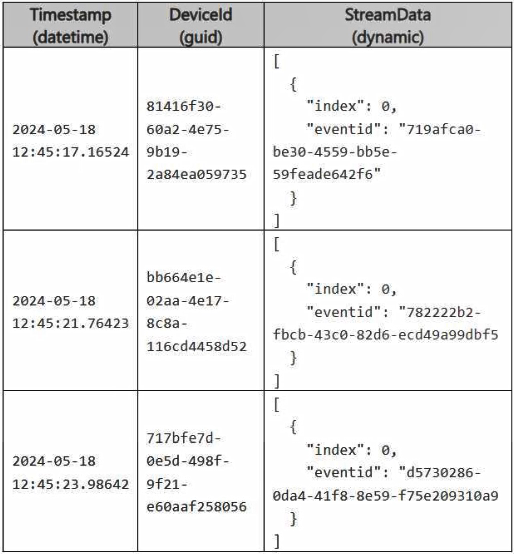
A table named Table1 -

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.



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.

**Propositions**

**A**

Une image contenant texte, capture d’écran, ligne, Police

Le contenu généré par l’IA peut être incorrect.

**B**

Une image contenant texte, ligne, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**C**

Une image contenant texte, ligne, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

**D**

Une image contenant texte, Police, ligne, nombre

Le contenu généré par l’IA peut être incorrect.

### REPONSE

Suggested Answer: B,D du site <https://www.freecram.com/discussions/Microsoft/exam-dp-700-topic-5-question-50-discussion-p1151370.html#gjson>

Reponse Chatgpt :

**✅ Réponses correctes :**

**A et D**

**🔍 Analyse des enregistrements :**

**✅ A**

* Timestamp: **2024-05-18** → Format **datetime**
* DeviceId: **GUID** → ✔️
* StreamData: contient temperature, **mais bien structuré**  
  ✅ **Valide** (champ temperature peut être ignoré si Table2 ne le définit pas)

**✅ B**

* Timestamp: **2024-05-21** → Format **datetime**
* DeviceId: **81416f30** → Format **pas un GUID complet**  
  ❌ **Non valide** → DeviceId **n'est pas au format GUID**

**(GUID = 36 caractères + 4 tirets)**

**❌ C**

* DeviceId contient : 81416f3060a24e759b192a84ea05973532dhdyt3 → trop long, contient des caractères invalides
* eventId est correct, mais ce champ est peut-être utilisé comme clé  
  ❌ **Non valide**

**✅ D**

* Timestamp: 2024-05-24 → **datetime**
* DeviceId: 81416f30-60a2-4e75-9b19-2a84ea059735 → **GUID valide**
* StreamData: correspond à un objet JSON valide  
  ✅ **Valide**

**✅ Réponses correctes :**

**A et D**

**👁️‍🗨️ Astuce : Seuls 1 et D ont un GUID valide**

# Question 61

You have a Fabric workspace.

You are debugging a statement and discover the following issues:

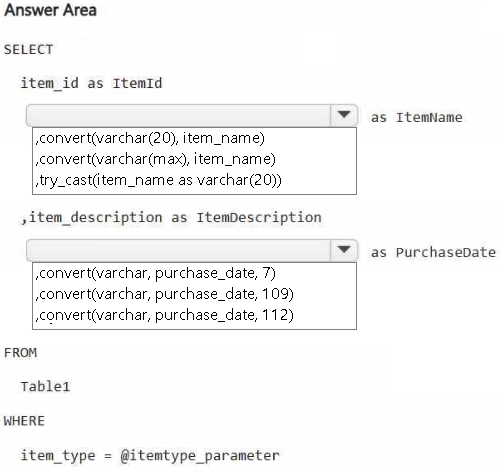
Sometimes, the statement fails to return all the expected rows.

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.



**Propositions**

### REPONSE

# Question 62

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?

**Propositions**

A.Degree of copy parallelism

B.Fault tolerance

C.Enable staging

D.Enable logging

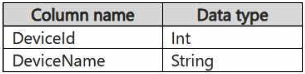
### REPONSE

# Question 63 (Idem 64, 66, 67)

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

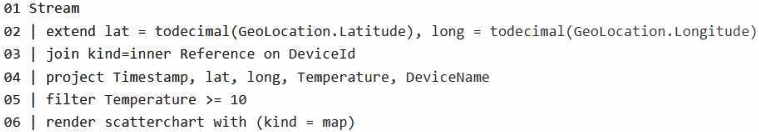


Reference contains reference data in the following format.



Both tables contain millions of rows.

You have the following KQL queryset.



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

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

Does this meet the goal?

**Propositions**

A.Yes

B.No

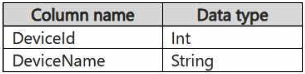
### REPONSE

# Question 64 (Idem 63, 66, 67)

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

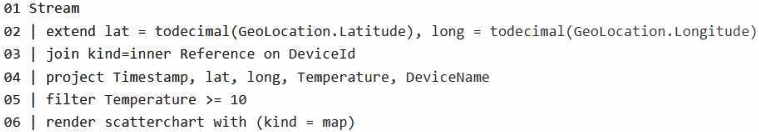


Reference contains reference data in the following format.



Both tables contain millions of rows.

You have the following KQL queryset.



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

Solution: You change project to extend.

Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

# Question 65

Case Study Litware

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

What should you do?

**Propositions**

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.

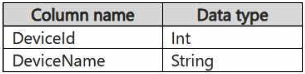
### REPONSE

# Question 66 (Idem 63, 64, 67)

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

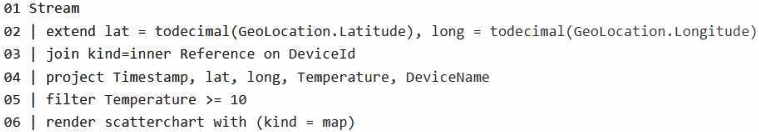


Reference contains reference data in the following format.



Both tables contain millions of rows.

You have the following KQL queryset.



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

Solution: You move the filter to line 02.

Does this meet the goal?

**Propositions**

A.Yes

B.No

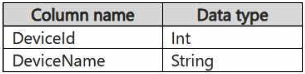
### REPONSE

# Question 67 (Idem 63, 64, 66)

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

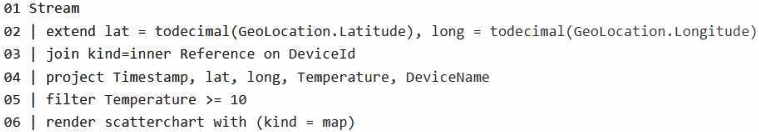


Reference contains reference data in the following format.



Both tables contain millions of rows.

You have the following KQL queryset.



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

Solution: You add the make\_list() function to the output columns.

Does this meet the goal?

**Propositions**

A.Yes

B.No

### REPONSE

❌ B.No

Le scatterchart attend des colonnes scalaires (ex: une latitude unique par ligne).

Si on utilise Make\_list (avec Summarize) , il a **une seule ligne avec des listes**, donc il ne peut pas interpréter plusieurs points.

**Requette décortiquée**

explication ligne par ligne du code (en KQL – Kusto Query Language, utilisé notamment dans Azure Data Explorer et Microsoft Fabric) :

**01 Stream**

➡️ Indique la source de données appelée Stream, probablement une table de télémétrie contenant des positions GPS et températures.

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

➡️ Convertit les coordonnées GPS (Latitude et Longitude) en nombres décimaux utilisables.

**03 | join kind=inner Reference on DeviceId**

➡️ Joint les données de Stream avec une table de référence (peut-être avec les noms des appareils), en se basant sur l’identifiant DeviceId.

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

➡️ Sélectionne les colonnes utiles : horodatage, latitude, longitude, température, et nom de l’appareil.

**05 | filter Temperature >= 10**

➡️ Ne conserve que les mesures avec une température d'au moins 10°C (ce qui est le cas des trois appareils : 19°C, 20°C, 22°C).

**06 | render scatterchart with (kind = map)**

➡️ Affiche les données sur une carte en tant que scatter chart (diagramme de dispersion), ce qui correspond exactement au rendu graphique que vous montrez.

**Précision Make\_list : Usage**

**Toujours make\_list() est utilisée avec summarize !!**

make\_list() Agrège plusieurs valeurs en liste

summarize Groupe les lignes (GROUP BY)

**Toujours make\_list() est utilisée avec summarize**

Imagine que tu as une table avec des valeurs comme ça :

**ID Valeur**

1 A

1 B

2 C

2 D

Si tu veux regrouper toutes les valeurs pour chaque ID en une liste, tu fais :

Table

| summarize liste\_valeurs = make\_list(Valeur) by ID

Cela va retourner :

**ID liste\_valeurs**

1 ["A", "B"]

2 ["C", "D"]

# Question 68

Case study Litware

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.

**Propositions**

A.a time-based schedule

B.a streaming dataflow

C.a blob storage action

D.a data pipeline

E.a notebook that uses Apache Spark Structured Streaming

F.a reflex item

### REPONSE

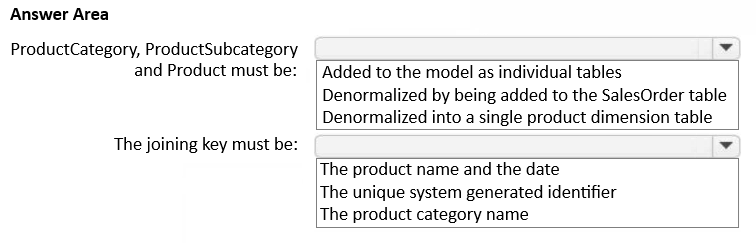
# Question 69

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.



**Propositions**

### REPONSE

# Question 70

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

Dataset1: This dataset will be added to Fabric and will have a unique primary key between the source and the destination. The unique primary key will be an integer and will start from 1 and have an increment of 1.

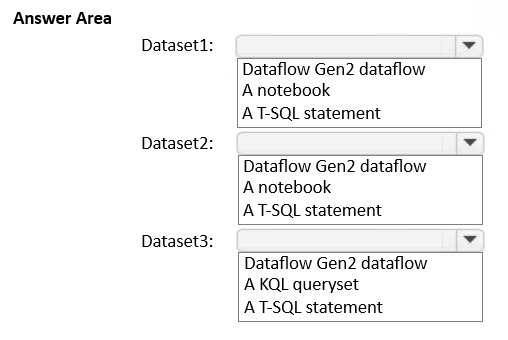
Dataset2: This dataset contains semi-structured data that uses bulk data transfer. The dataset must be handled in one process between the source and the destination. The data transformation process will include the use of custom visuals to understand and work with the dataset in development mode.

Dataset3: This dataset is in a lakehouse. The data will be bulk loaded. The data transformation process will include row-based windowing functions during the loading process.

You need to identify which type of item to use for the datasets. The solution must minimize development effort and use built-in functionality, when possible.

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

NOTE: Each correct selection is worth one point.



**Propositions**

### REPONSE

# Question 71

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?

**Propositions**

A.Type 2

B.Type 0

C.Type 1

D.Type 3

### REPONSE

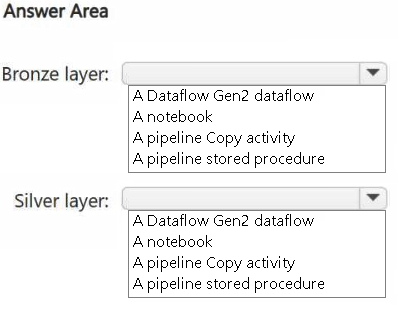
# Question 72

Case study Contoso

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.



**Propositions**

### REPONSE

# Question 73

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table named Table1 in a lakehouse. The streaming data is sourced from motorway sensors and represents the speed of cars.

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

Which windowing function should you use?

**Propositions**

A.sliding

B.hopping

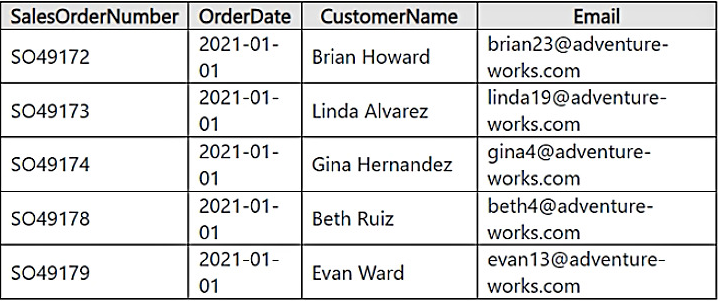
C.tumbling

D.session

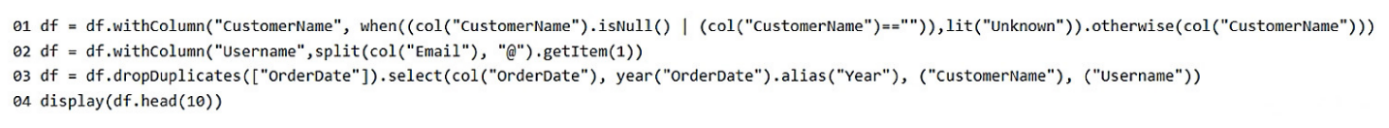
### REPONSE

# Question 74

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

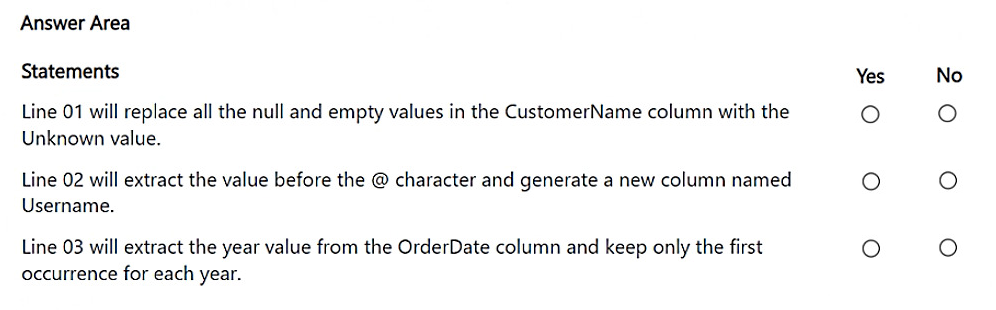


You have a notebook that contains the following code segment.



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

**Propositions**

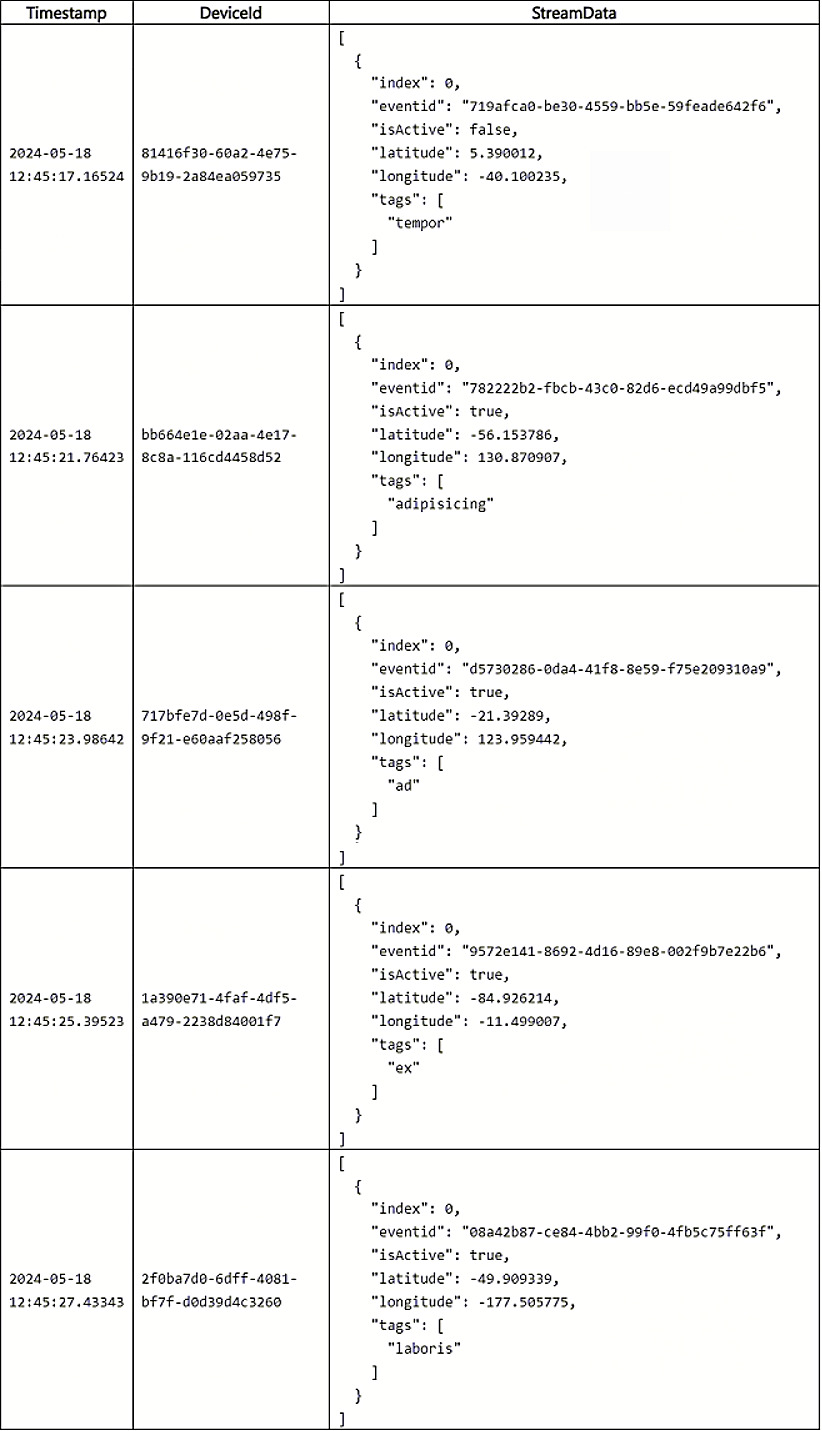


### REPONSE

# Question 75

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

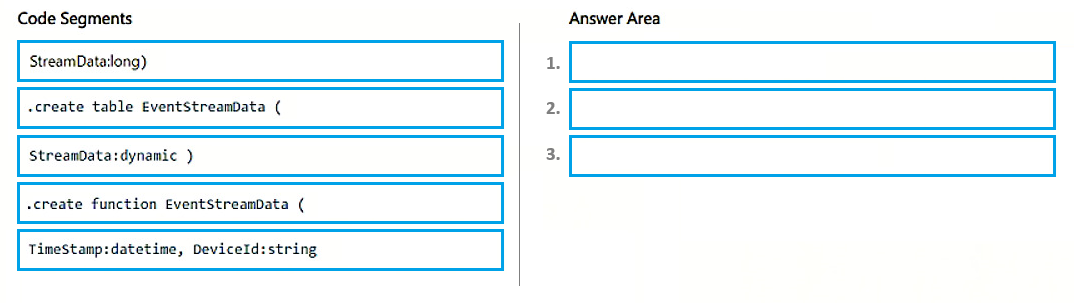
In Eventhouse1, you plan to create a table named DeviceStreamData in a KQL database. The table will contain data based on the following sample.



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

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

**Propositions**



### REPONSE

# Question 76

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?

**Propositions**

A.a data pipeline

B.an Apache Spark job definition

C.a streaming dataflow

D.a notebook

### REPONSE

# Question 77

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?

**Propositions**

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

B.a direct attributes comparison across the attributes in the DimCustomer table.

C.a direct attributes comparison for the attributes in the source table.

D.a hash function to compare the attributes in the DimCustomer table.

### REPONSE

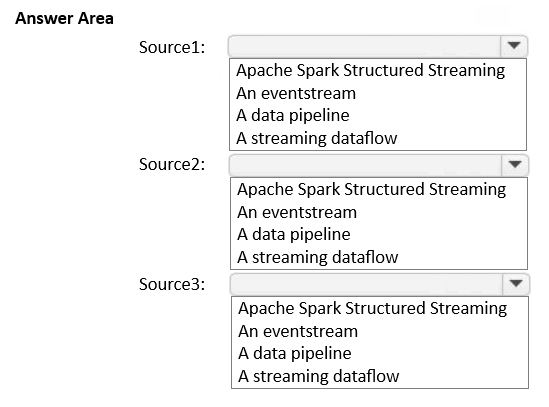
# Question 78

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



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.



**Propositions**

### REPONSE

# Question 79

Case Study Contoso

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

What should you do?

**Propositions**

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.

### REPONSE

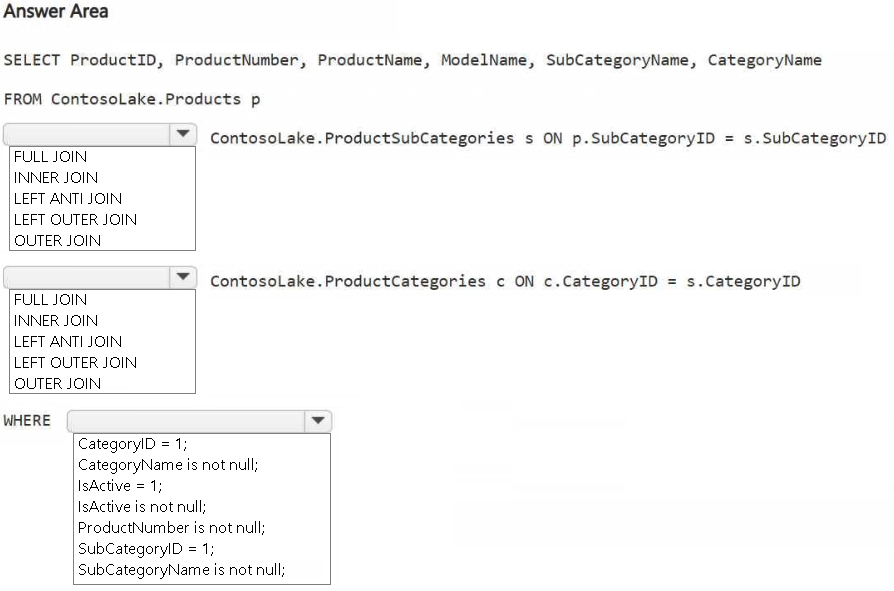
# Question 80

Case study Contoso

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.



**Propositions**

### REPONSE

# Page 17

[Microsoft - DP-700 - Page 17 | Examprepper](https://www.examprepper.co/exam/72/17)

# Question 81

**Propositions**

### REPONSE

# Question 82

**Propositions**

### REPONSE

# Question 83

**Propositions**

### REPONSE

# Question 84

**Propositions**

### REPONSE

# Question 85

**Propositions**

### REPONSE

# Question BONUS

Question posée à l’exam DP-600 ( qui est une Question DP-700)

Question 50 dans Examtopics

You have a Fabric workspace named Workspace1 and an Azure Data Lake Storage Gen2 account named storage1. Workspace1 contains a lakehouse named Lakehouse1.

You need to create a shortcut to storage1 in Lakehouse1.

Which connection and endpoint should you specify?

**Propositions**

A.connection

1.abfs

2.abfss

3.htpps

B.Endpoint

1.blob

2.dfs

3.file

### REPONSE

**✅ Connexion**

abfss — Ce protocole sécurisé est utilisé pour accéder à Azure Data Lake Storage Gen2 via SSL. C’est la version sécurisée d’abfs.

**✅ Endpoint**

dfs — L’endpoint dfs est spécifiquement conçu pour Data Lake Storage Gen2, prenant en charge les fonctionnalités de hiérarchisation et les chemins de type dossier.

* abfs est non sécurisé (sans SSL), à éviter sauf cas spécifique.
* https est utilisé pour les blobs classiques, mais ne permet pas l’accès via les chemins hiérarchiques nécessaires aux lakehouses.
* blob cible les blobs simples, file concerne Azure Files, donc pas adaptés ici.