# Certification réussie le 23/08/2025

Score 845

|  |  |
| --- | --- |
| Une image contenant texte, capture d’écran, conception  Le contenu généré par l’IA peut être incorrect. | Une image contenant texte, Visage humain, capture d’écran, personne  Le contenu généré par l’IA peut être incorrect. |

# Table des matieres

Table des matières

[Certification réussie le 23/08/2025 1](#_Toc207028085)

[Table des matieres 1](#_Toc207028086)

[Question Video Window 7](#_Toc207028087)

[Q1. What does the ‘Take’ operator do in a KQL Query ? 7](#_Toc207028088)

[REPONSE 7](#_Toc207028089)

[Q2. What is the primary use of the Copy Data activity in a data pipeline ? 7](#_Toc207028090)

[Q3. What keyword should you use to retrieve only specific columns in a KQL query? 8](#_Toc207028091)

[USE CASE CONTOSO 10](#_Toc207028092)

[Question 1 CONTOSO 10](#_Toc207028093)

[Question 50 CONTOSO 11](#_Toc207028094)

[Question 83 CONTOSO 12](#_Toc207028095)

[Question 90 CONTOSO 12](#_Toc207028096)

[Question 91 CONTOSO 13](#_Toc207028097)

[Question 94 CONTOSO 14](#_Toc207028098)

[Question 98 CONTOSO 15](#_Toc207028099)

[Question 99 CONTOSO 17](#_Toc207028100)

[Question 109 CONTOSO 17](#_Toc207028101)

[Question 111 CONTOSO 18](#_Toc207028102)

[Question 14 DP-600 20](#_Toc207028103)

[LITWARE 22](#_Toc207028104)

[Question 31 LITWARE 22](#_Toc207028105)

[Question 32 LITWARE 22](#_Toc207028106)

[Question 61 LITWARE 23](#_Toc207028107)

[Question 72 LITWARE 24](#_Toc207028108)

[Question 75 = 97 LITWARE 24](#_Toc207028109)

[Question 75 = 97 LITWARE 25](#_Toc207028110)

[Question 76 LITWARE 26](#_Toc207028111)

[Question 103 LITWARE (ad-hoc issue) 27](#_Toc207028112)

[Question 105 LITWARE 28](#_Toc207028113)

[QUESTIONS GENERIQUES 30](#_Toc207028114)

[QUESTIONS RENCONTREES EN EXAM 30](#_Toc207028115)

[QUESTION 2 – 54 RLS 31](#_Toc207028116)

[QUESTION 3 31](#_Toc207028117)

[QUESTION 4 – 52 (Tricky la 52) 32](#_Toc207028118)

[QUESTION 5 32](#_Toc207028119)

[QUESTION 6 33](#_Toc207028120)

[Question 7 35](#_Toc207028121)

[Question 8 (/10) 35](#_Toc207028122)

[Question 9 37](#_Toc207028123)

[Question 10 (/8) 37](#_Toc207028124)

[Question 11 39](#_Toc207028125)

[Question 12 40](#_Toc207028126)

[Question 13 40](#_Toc207028127)

[Question 14 41](#_Toc207028128)

[Question 15 42](#_Toc207028129)

[Question 16 42](#_Toc207028130)

[QUESTION 17 43](#_Toc207028131)

[Question 18 46](#_Toc207028132)

[Question 19 47](#_Toc207028133)

[QUESTION 20 47](#_Toc207028134)

[Question 21 48](#_Toc207028135)

[Question 22 48](#_Toc207028136)

[Question 23 49](#_Toc207028137)

[Question 24 50](#_Toc207028138)

[Question 25 51](#_Toc207028139)

[Question 26 52](#_Toc207028140)

[Question 38 52](#_Toc207028141)

[Question 39 54](#_Toc207028142)

[Question 41 55](#_Toc207028143)

[Question 42 55](#_Toc207028144)

[Question 43 56](#_Toc207028145)

[Question 44 56](#_Toc207028146)

[Question 45 -55 57](#_Toc207028147)

[Question 55 – 45 58](#_Toc207028148)

[Question 47 59](#_Toc207028149)

[Question 46 59](#_Toc207028150)

[Question 49 EXAMPREPPER : PIPELINE 61](#_Toc207028151)

[Question 51 62](#_Toc207028152)

[Question 52 - 4 Tricky la 52 63](#_Toc207028153)

[Question 54 - 2 RLS 64](#_Toc207028154)

[Question 56 65](#_Toc207028155)

[Question 57 66](#_Toc207028156)

[Question 58 67](#_Toc207028157)

[Question 59 67](#_Toc207028158)

[Question 60 68](#_Toc207028159)

[Question 62 - 96 69](#_Toc207028160)

[QUESTION 63 - 95 70](#_Toc207028161)

[Question ??? 72](#_Toc207028162)

[Question 64 72](#_Toc207028163)

[Question 65 – 66 - 110 74](#_Toc207028164)

[Question 66 - 65 - 110 75](#_Toc207028165)

[Question 67 76](#_Toc207028166)

[Question 68 78](#_Toc207028167)

[Question 69 79](#_Toc207028168)

[Question 70 - Q MULTIPLE 80](#_Toc207028169)

[Question 71 – Q MULTIPLE 81](#_Toc207028170)

[Question 72 – Q MULTIPLE 81](#_Toc207028171)

[Question 73 – Q MULTIPLE 82](#_Toc207028172)

[Question 77 82](#_Toc207028173)

[Question 78 83](#_Toc207028174)

[Question 79 – Similaire … à 78 84](#_Toc207028175)

[Question 80 85](#_Toc207028176)

[Question 81 87](#_Toc207028177)

[Question 82 88](#_Toc207028178)

[Question 84 89](#_Toc207028179)

[Question 85 89](#_Toc207028180)

[Question 86 90](#_Toc207028181)

[Question 87 92](#_Toc207028182)

[Question 88 92](#_Toc207028183)

[Question 89 93](#_Toc207028184)

[Question 92 94](#_Toc207028185)

[Question 93 95](#_Toc207028186)

[Question 95 – 63 97](#_Toc207028187)

[Question 96 - 62 99](#_Toc207028188)

[Question 100 100](#_Toc207028189)

[Question 101 102](#_Toc207028190)

[Question 102 103](#_Toc207028191)

[Question 104 103](#_Toc207028192)

[Question 106 105](#_Toc207028193)

[Question 107 106](#_Toc207028194)

[Question 108 107](#_Toc207028195)

[Question 110 – 65 - 66 107](#_Toc207028196)

[Question 112 108](#_Toc207028197)

[Question 113 109](#_Toc207028198)

[Question 114 111](#_Toc207028199)

[Question 115 112](#_Toc207028200)

[Question 116 113](#_Toc207028201)

[Question 117 114](#_Toc207028202)

[Question 117 114](#_Toc207028203)

[Question 118 115](#_Toc207028204)

[SYNTHESE GENERIQUES 116](#_Toc207028205)

[SYNTHESE CONTOSO 117](#_Toc207028206)

[SYNTHESE LITWARE 120](#_Toc207028207)

[SCD TYPE 122](#_Toc207028208)

[EXO Eventstream 123](#_Toc207028209)

[EXO Apache Spark Structured Streaming ASSS 130](#_Toc207028210)

[ANNEXES 135](#_Toc207028211)

[Annexe Question 84 135](#_Toc207028212)

# Question Video Window

## Q1. What does the ‘Take’ operator do in a KQL Query ?

1. Filter data based on a condition
2. Retrieves a specified number of rows from a table
3. Aggregates data
4. Joins two tables together

## REPONSE

✅**B. Retrieves a specified number of rows from a table**

**Explanation:**  
In KQL, take is used to return only a specific number of rows from the query result set.  
Example KQL :

TableName

| take 10

## Q2. What is the primary use of the Copy Data activity in a data pipeline ?

1. To apply transformation to data during ingestion
2. To ingest data from external source into a lakehouse file or table
3. To merge data from multiple sources
4. To delete existing data before copying new data

**✅ B. To ingest data from an external source into a lakehouse file or table**

**Explication**

* Dans Microsoft Fabric (ou Azure Data Factory), l’activité **Copy Data** permet de **copier des données depuis une source externe** (SQL, API, S3, ADLS, etc.) vers une destination (souvent un Lakehouse, un Warehouse ou un autre stockage).
* Elle peut inclure **quelques transformations légères** (mapping de colonnes, conversions de type), mais ce n’est pas son but principal.
* L’objectif principal reste **l’ingestion** de données dans l’environnement cible

## Q3. What keyword should you use to retrieve only specific columns in a KQL query?

1. Filter
2. Select
3. Project
4. Retrieve

✅C.Project

Question ????

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

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

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

How should you specify the parameter value?

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

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

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

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

✅ A. "@pipeline().parameters.param1"

**Explication**

Ce qu’on te demande :

Tu dois écrire une expression dans le pipeline qui retourne la valeur de param1 comme un entier.

➡️ Autrement dit :  
Tu dois écrire une formule dans le pipeline qui utilise ce paramètre et qui donne la valeur entière qu’on lui a passée.

**param1 est bien un paramètre d’entrée.  
La phrase "returns param1 as an int value" signifie que l’expression doit produire une valeur entière à partir de ce paramètre, pas que le pipeline retourne param1.**

**🧩 Les propositions :**

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

**Phrases-clés à retenir :**

**🔹 1. Expression pure → Pas d’accolades**

**"Expression pure, pas de couture : pas d’accolades."**  
📌 Utilisée dans les champs comme Set Variable, If Condition, Derived Column.

➡️ Syntaxe : @pipeline().parameters.param1

**🔹 2. Interpolation dans chaîne → Accolades**

**"Texte dynamique ? Accolades automatiques !"**  
📌 Utilisée pour construire des chaînes : noms de fichiers, chemins, messages.

➡️ Syntaxe : "file\_@{pipeline().parameters.param1}.csv"

**🔹 3. Type du paramètre guide le résultat**

**"Le type défini dicte le type retourné."**  
📌 Si param1 est un int, l’expression retourne un int, même avec @{...} — mais **le contexte d’évaluation** peut forcer une conversion.

**🔹 4. Champ d’expression = pas de guillemets, pas d’accolades**

**"Expression = directe, sans décor."**  
📌 Pas de guillemets, pas de @{...} — juste l’expression brute.

**🔹 5. Champ de texte = guillemets + accolades**

**"Texte = guillemets + accolades pour insérer."**  
📌 Tu construis une chaîne → tu interpolles avec @{...}.

# USE CASE CONTOSO

## Question 1 CONTOSO

[DP-700 Exam - Free Actual Q&As, Page 1 | ExamTopics](https://www.examtopics.com/exams/microsoft/dp-700/view/)

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

What should you do?

A. Add the DataAnalyst group to the Viewer role for WorkspaceA.

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

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

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

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

Cela permet aux analysts de lire les tables exposées via le **SQL Endpoint** du lakehouse. Si on configure ce SQL Endpoint pour n’exposer que les tables de la **gold layer**, ils auront l’accès en lecture voulu, et **aucun accès** direct aux données bronze/silver.

## Question 50 CONTOSO

[Microsoft - DP-700 - Page 10 | Examprepper](https://www.examprepper.co/exam/73/10)

You need to ensure that WorkspaceA can be configured for source control.

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

NOTE: Each correct selection is worth one point.

A.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.

**✅ Réponse correcte :**

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

**D** Assign WorkspaceA to Cap1.

| **Option** | **Analyse** | **Correct ?** |
| --- | --- | --- |
| **A.** From Tenant setting, set *Users can synchronize workspace items with their Git repositories* to Enabled. | C’est **exactement** le paramètre à activer au niveau du tenant pour permettre la synchronisation avec Git. | ✅ |
| **B.** From Tenant setting, set *Users can sync workspace items with GitHub repositories* to Enabled. | Ce paramètre **n’existe pas** sous cette formulation. GitHub est un type de dépôt Git, mais le paramètre est générique. | ❌ |
| **C.** Configure WorkspaceA to use a Premium Per User (PPU) license. | Le contrôle de version via Git **ne dépend pas** du type de licence PPU. | ❌ |
| **D.** Assign WorkspaceA to Cap1. | Pour activer la synchronisation Git, le workspace doit être **assigné à une capacité Fabric** (comme Cap1). | ✅ |

## Question 83 CONTOSO

[Microsoft - DP-700 - Page 17 | Examprepper](https://www.examprepper.co/exam/73/17)

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.

**Bronze layer** 🔽 (choix possible)

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

**Silver layer** 🔽 (choix possible)

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

**✅ Réponse recommandée**

| **Layer** | **Méthode recommandée** |
| --- | --- |
| Bronze layer | **A pipeline Copy activity** |
| Silver layer | **A notebook** |

**mots-clés vers la bonne réponse**

* **Bronze layer → données brutes → Pipeline Copy activity (simple, Fabric-native, peu de dev).**
* **Silver layer → transformations complexes → Notebook (puissant, versionnable, compatible Python/SQL).**

**Justification rapide**

* **Bronze layer** : Le pipeline avec **Copy activity** est idéal pour ingérer les données brutes de POS1 (SQL Server sur VM Azure) vers le lakehouse en format Delta, sans transformation préalable. Il respecte les contraintes de Fabric, les préférences de l’équipe, et les exigences de source control.
* **Silver layer** : Les notebooks permettent d’appliquer des transformations complexes en Python ou SQL (préférés par les data engineers), comme le nettoyage, la déduplication, et la normalisation. Ils sont aussi source-controlables et orchestrables.

## Question 90 CONTOSO

[Microsoft - DP-700 - Page 18 | Examprepper](https://www.examprepper.co/exam/73/18)

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

What should you recommend?

A.Add a ForEach activity to the data pipeline.

B.Configure retries for the Copy data activity.

C.Call a notebook from the data pipeline.

D.Configure Fault tolerance for the Copy data activity.

✅ **B.Configure retries for the Copy data activity.**

Rappel

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.

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

**✅ B. Configure retries for the Copy data activity**

**🎯 Pourquoi ?**

* C’est une **fonction intégrée** dans Fabric.
* Elle permet de **réessayer automatiquement** en cas d’erreur temporaire.
* Elle **minimise l’effort de développement**.
* Elle cible **exactement** le problème décrit.

## Question 91 CONTOSO

[Microsoft - DP-700 - Page 19 | Examprepper](https://www.examprepper.co/exam/73/19)

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

What should you include in the recommendation?

A.a data pipeline that includes a Copy data activity

B.a data pipeline that includes a Delete data activity

C.a notebook that runs the VACUUM command

D.a notebook that runs the OPTIMIZE command

Mot Clef :

| **Mot-clé / Phrase** | **Ce que ça implique** |
| --- | --- |
| **"old files that are no longer referenced by a Delta table log"** | Il s’agit de **fichiers orphelins** dans le Delta Lake. |
| **"must be removed"** | Il faut une **suppression physique** des fichiers. |
| **"meet the technical requirements"** | La solution doit être **compatible Fabric**, **source-controllable**, et **automatisable**. |

**✅ C. a notebook that runs the VACUUM command**

**🎯 Pourquoi ?**

* VACUUM est **spécialement conçu** pour supprimer les fichiers non référencés dans Delta Lake.
* Il respecte les exigences :
  + **Automatisable** (via notebook planifié).
  + **Source-controllable** (stocké dans Azure Repos).
  + **Compatible Fabric**.
  + **Minimise le développement** (commande simple).

## Question 94 CONTOSO

[Microsoft - DP-700 - Page 19 | Examprepper](https://www.examprepper.co/exam/73/19)

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

What should you do?

A.Create a workspace identity and enable high concurrency for the notebooks.

B.Create a shortcut and ensure that caching is disabled for the workspace.

C.Create a workspace identity and use the identity in a data pipeline.

D.Create a shortcut and ensure that caching is enabled for the workspace.

**Mots-clés à repérer**

| **Mot-clé / Phrase** | **Ce que ça implique** |
| --- | --- |
| **"Amazon S3 bucket"** | Source externe, inter-cloud. |
| **"Minimize egress costs"** | Ne pas transférer inutilement les données. |
| **"Prevent saving a copy"** | Ne pas ingérer ni mettre en cache → accès direct. |

✅ **B. Create a shortcut and ensure that caching is disabled for the workspace**

**🎯 Pourquoi ?**

* Le **shortcut** permet un accès direct aux fichiers S3.
* **Désactiver le cache** évite toute copie locale.
* Cela respecte :
  + ✅ **Minimisation des coûts d’egress**
  + ✅ **Pas de duplication des données**
  + ✅ **Accès contrôlé et optimisé**

## Question 98 CONTOSO

[Microsoft - DP-700 - Page 20 | Examprepper](https://www.examprepper.co/exam/73/20)

You need to create the product dimension.

How should you complete the Apache Spark SQL code?

SELECT

ProductID,

ProductNumber,

ProductName,

ModelName,

SubCategoryName,

CategoryName

FROM

ContosoLake.Products p

[Choice 1]

ContosoLake.ProductSubCategories s ON p.SubCategoryID = s.SubCategoryID

[Choice 2]

ContosoLake.ProductCategories c ON c.CategoryID = s.CategoryID

WHERE

[Choice 3]

Choice 1:

1. FULL JOIN
2. INNER JOIN
3. LEFT ANTI JOIN
4. LEFT OUTER JOIN
5. OUTER JOIN

Choice 2:

1. FULL JOIN
2. INNER JOIN
3. LEFT ANTI JOIN
4. LEFT OUTER JOIN
5. OUTER JOIN

Choice 3:

1. CategoryID = 1;
2. CategoryName is not null;
3. IsActive = 1 ;
4. IsActive is not null;
5. ProductNumber is not null;
6. SubCategoryID = 1;
7. SubCategoryName is not null;

Mots clef :

| **Élément** | **Implication** |
| --- | --- |
| **"only active products"** | Il faut filtrer sur IsActive = 1. |
| **"Some product categories and subcategories NOT assigned to any product... must be omitted"** | Il faut des **INNER JOIN** pour exclure les éléments non liés. |

**✅ Choice 1 : B. INNER JOIN**  
**Référence :**

*"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."*  
💡 → Un INNER JOIN entre **Products** et **ProductSubCategories** permet de **supprimer** les sous-catégories sans produits (évite les NULL).

**✅ Choice 2 : B. INNER JOIN**  
**Référence :**

*"Some product categories and subcategories are NOT assigned to any product. They are NOT analytically relevant and must be omitted..."*  
💡 → Un INNER JOIN entre **ProductSubCategories** et **ProductCategories** élimine les catégories qui ne sont pas liées à des produits (par propagation).

**✅ Choice 3 : C. IsActive = 1**  
**Référence :**

*"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."*  
💡 → La clause WHERE IsActive = 1 filtre uniquement les produits actifs.

## Question 99 CONTOSO

[Microsoft - DP-700 - Page 20 | Examprepper](https://www.examprepper.co/exam/73/20)

You need to populate the MAR1 data in the bronze layer.

Which two types of activities should you include in the pipeline? Each correct answer presents part of the solution.

A.ForEach

B.Copy data

C.WebHook

D.Stored procedure

**A. ForEach** ✅

*Chaque entité MAR1 ayant un endpoint distinct, il faut un mécanisme pour itérer sur la liste d’URL/endpoints.*  
💡 ForEach permet de boucler sur une collection d’éléments (ici, les 7 endpoints).

**B. Copy data** ✅

*"Development effort must be minimized and a built-in connection must be used to import the source data."*  
💡 L’activité Copy data est la méthode native dans Fabric Data Pipeline pour **copier des données** depuis une source REST API vers un Lakehouse (Delta format).

## Question 109 CONTOSO

[Microsoft - DP-700 - Page 22 | Examprepper](https://www.examprepper.co/exam/73/22)

You need to ensure that the data engineers are notified if any step in populating the lakehouses fails. The solution must meet the technical requirements and minimize development effort.

What should you use?

[Choice 1] : To identify the failure

1. Une activité Fail
2. Une activité If condition
3. Une condition On failure dependency
4. Une condition On completion dependency

[Choice 2] : To send the notification

1. Une activité Teams
2. Une activité Invoke pipeline
3. Une activité Office365Outlook

**Contexte du use case**

“Business users perform ad-hoc queries against the warehouse.” “Reports sometimes run for two hours and fail to load as expected.” “The SQL query failed while running.

✅ **[Choice 1]** → **3. Une condition On failure dependency**

|  |  |
| --- | --- |
| **1. Une activité Fail** | ❌ Sert à **provoquer** un échec, pas à le détecter |
| **2. Une activité If condition** | ❌ Nécessite une logique personnalisée → effort de développement |
| 🟩 **3. Une condition On failure dependency** | ✅ Permet de **déclencher une action uniquement si une activité précédente échoue** |
| **4. Une condition On completion dependency** | ❌ Se déclenche même si l’activité réussit → pas ciblé sur l’échec |

✅ **[Choice 2]** → **3. Une activité Office365Outlook**

**MOT CLEF : “If any step in populating the lakehouses fails, an email must be sent to the data engineers.”**

|  |  |
| --- | --- |
| **1. Une activité Teams** | ❌ Pas mentionné dans les exigences (email requis) |
| **2. Une activité Invoke pipeline** | ❌ Sert à appeler un autre pipeline, pas à envoyer une notification |
| 🟩 **3. Une activité Office365Outlook** | ✅ Permet d’envoyer un **email** via Outlook, intégré à Fabric, avec **peu de configuration** |

## Question 111 CONTOSO

[Microsoft - DP-700 - Page 23 | Examprepper](https://www.examprepper.co/exam/73/23)

You need to schedule the population of the medallion layers to meet the technical requirements.

What should you do?

A.Schedule a data pipeline that calls other data pipelines.

B.Schedule a notebook.

C.Schedule an Apache Spark job.

D.Schedule multiple data pipelines.

✅ A. Schedule a data pipeline that calls other data pipelines

**Extraits clés du use case**

Voici les exigences directement liées à la planification :

* “The new lakehouses must follow a medallion architecture by using the following three layers: bronze, silver, and gold.”
* “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.”
* “Development effort must be minimized.”

➡️ On doit orchestrer les étapes **dans un ordre précis**, avec **gestion des erreurs**, et **planification centralisée**.

| **Option** | **Pertinence** | **Justification** |
| --- | --- | --- |
| 🟩 **A. Schedule a data pipeline that calls other data pipelines** | ✅ **Correct** | Permet d’orchestrer les étapes bronze → silver → gold dans un **pipeline maître**, avec **gestion des dépendances**, **retries**, et **notifications**. |
| **B. Schedule a notebook** | ❌ Trop limité pour orchestrer plusieurs couches et gérer les erreurs entre étapes |  |
| **C. Schedule an Apache Spark job** | ❌ Trop bas niveau, ne permet pas d’orchestration complète ni de gestion des dépendances |  |
| **D. Schedule multiple data pipelines** | ❌ Risque de perte de contrôle sur l’ordre, la synchronisation, et la gestion des erreurs |  |

²

**Objectif**

Contoso veut **automatiser** le traitement des données en suivant une architecture **medallion** :

* **Bronze** : données brutes
* **Silver** : données nettoyées
* **Gold** : données prêtes pour l’analyse

Et surtout, **chaque couche doit être remplie avant de passer à la suivante**, avec **gestion des erreurs** et **notification par email** si quelque chose échoue.

**🧠 Pourquoi le choix A est le bon ?**

**✅ A. Schedule a data pipeline that calls other data pipelines**

Imagine un **pipeline principal** (appelons-le Pipeline\_Medallion) qui orchestre les trois étapes :

1. Il **appelle le pipeline Bronze** pour ingérer les données brutes.
2. Une fois le Bronze terminé avec succès, il **appelle le pipeline Silver** pour nettoyer les données.
3. Ensuite, il **appelle le pipeline Gold** pour créer les dimensions et les modèles analytiques.

Et si **l’un des pipelines échoue**, le pipeline principal peut :

* **Arrêter l’exécution**
* **Envoyer un email** aux data engineers
* **Faire un retry automatique** si besoin

Ce comportement est **nativement supporté** dans Fabric avec les **pipelines de données**, qui permettent :

* L’**orchestration** de plusieurs étapes
* La **gestion des dépendances**
* La **planification** (exécution régulière)
* L’**intégration avec les alertes et les logs**

## Question 14 DP-600

Case Study CONTOSO

You need to migrate the Research division data for Productline2. The solution must meet the data preparation requirements.

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

NOTE: Each correct selection is worth one point.

**Answer Area**

|  |
| --- |
| df=spark.read.format('csv').options(header=”true”,inferSchema=”true”).load([“fsi@storage1.dfs.core.windows.net/files/productline2”])  df.write.mode("overwrite").format(  [ Choice 1 ]  ).save(  [ Choice 2 ]  ) |

**Propositions**

Choice 1 :

1. csv
2. delta
3. parquet

Choice 2 :

1. productline2
2. Tables/productline2
3. Tables/research/productline2

* Choice 1 : 2 : delta → pour créer une table managée compatible avec Lakehouse Explorer.
* Choice 2 : 3 : Tables/research/productline2→ pour respecter la structure logique et permettre le filtrage dans OneLake Data Hub par département.
* Respecte la hiérarchie imposée : Tables = managed tables, research = département, productline2 = dataset.

# LITWARE

## Question 31 LITWARE

[Microsoft - DP-700 - Page 7 | Examprepper](https://www.examprepper.co/exam/73/7)

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

How should you configure the Apache Spark settings?

A.Disable high concurrency.

B.Create a custom pool.

C.Modify the number of executors.

D.Set the default environment.

**B. Create a custom pool**  
✅ Correct — un **custom Spark pool** permet d’allouer des ressources dédiées et d’isoler les workloads. On pourrait en avoir un pour bronze et un autre pour silver.

## Question 32 LITWARE

[Microsoft - DP-700 - Page 7 | Examprepper](https://www.examprepper.co/exam/73/7)

You need to ensure that the authors can see only their respective sales data. How should you complete the statement?

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

CREATE FUNCTION dbo.tvf\_rlspredicate(@Author AS varchar(50))

RETURNS TABLE

WITH [ Choice 1]

AS

RETURN SELECT 1 AS tvf\_rlspredicate\_result

WHERE @Author = [Choice 2]

GO

CREATE SECURITY POLICY RLSFilter

ADD FILTER PREDICATE Security.tvf\_rlspredicate(AuthorEmail)

ON [ Choice 3]

WITH (STATE = ON)

PROPOSITIONS :

1. AuthorSales
2. AuthorEmail
3. AuthorSales.AuthorEmail
4. BLOCK
5. FILTER
6. INLINE
7. SCHEMABINDING
8. USER\_NAME()

✅ **Choice 1** → **G. SCHEMABINDING**

* Obligatoire pour une fonction table utilisée comme prédicat de sécurité RLS.

✅ **Choice 2** → **H. USER\_NAME()**

* Permet de comparer l’auteur courant à la colonne d’e-mail stockée dans la table.

✅ **Choice 3** → **A. AuthorSales**

* C’est la table sur laquelle la politique RLS s’applique.

## Question 61 LITWARE

[Microsoft - DP-700 - Page 13 | Examprepper](https://www.examprepper.co/exam/73/13)

You need to implement the solution for the book reviews.

Which should you do?

A.Create a Dataflow Gen2 dataflow.

B.Create a shortcut.

C.Enable external data sharing.

D.Create a data pipeline.

la phrase clé associée est :

**"Make the book reviews available in the lakehouse without making a copy of the data."**

cible vraiment la **mise à disposition** de la donnée, pas le **traitement automatique**.

✅ **B. Create a shortcut**

## Question 72 LITWARE

[Microsoft - DP-700 - Page 15 | Examprepper](https://www.examprepper.co/exam/73/15)

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

What should you do?

A.Spilt the dataflow into two dataflows.

B.Configure scheduled refresh for the dataflow.

C.Configure incremental refresh for the dataflow. Set Store rows from the past to 1 Month.

D.Configure incremental refresh for the dataflow. Set Refresh rows from the past to 1 Year.

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

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

**Remarque sur set store rows :**

**Ca fait toujours le job chargement DELTA mais ne conserve dans le modele que 1 mois**

## Question 75 = 97 LITWARE

[Microsoft - DP-700 - Page 15 | Examprepper](https://www.examprepper.co/exam/73/15)

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

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

NOTE: Each correct selection is worth one point.

A.a time-based schedule

B.a streaming dataflow

C.a blob storage action

D.a data pipeline

E.a notebook that uses Apache Spark Structured Streaming

F.a reflex item

**Rappel du besoin** :

*When a new book cover image arrives in the Files folder, process the image as soon as possible.*

➡️ Donc, il faut **déclencher un traitement à l’arrivée d’un fichier** (pas sur horaire) → c’est un traitement *event-driven*.

**✅ Réponse**

**D. a data pipeline**  
**F. a reflex item**

**Analyse des options**

* **A. a time-based schedule** → ❌ Non, on ne veut pas attendre une exécution programmée.
* **B. a streaming dataflow** → ❌ Non, streaming dataflows sont faits pour flux tabulaires (JSON, CSV, etc.), pas pour images binaires.
* **C. a blob storage action** → ❌ Pas vraiment. On n’est pas dans Azure Logic Apps ici, et Fabric ne parle pas de “blob storage action” pour ce besoin.
* **D. a data pipeline** → ✅ Oui, pour orchestrer le traitement et exécuter les étapes dès l’événement.
* **E. a notebook that uses Apache Spark Structured Streaming** → ❌ Pas nécessaire, on ne parle pas d’un flux continu, juste d’un déclenchement ponctuel à chaque fichier.
* **F. a reflex item** → ✅ Oui, Reflex dans Fabric permet de déclencher automatiquement un pipeline/notebook dès l’arrivée d’un fichier dans le Lakehouse/OneLake.

## Question 75 = 97 LITWARE

[Microsoft - DP-700 - Page 20 | Examprepper](https://www.examprepper.co/exam/73/20)

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

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

NOTE: Each correct selection is worth one point.

A.an activator item

B.a data pipeline

C.a blob storage action

D.a time-based schedule

E.a streaming dataflow

F.a notebook that uses Apache Spark Structured Streaming

**Réponse correcte**

✅ **A. an activator item**

✅ **B. a data pipeline**

**🧩 Justification finale**

* **Activator item** : Déclenche le pipeline dès qu’un nouveau fichier image est détecté dans le dossier Files.
* **Data pipeline** : Orchestration du traitement des images (extraction de métadonnées, resize, etc.).

## Question 76 LITWARE

[Microsoft - DP-700 - Page 16 | Examprepper](https://www.examprepper.co/exam/73/16)

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

A.a streaming dataflow

B.a streaming dataset

C.a notebook that uses Apache Spark Structured Streaming

D.an eventstream

**“Litware plans to manage Search Engine Optimization (SEO) for the authors. The SEO data will be streamed from a REST API.”** **“Process the SEO data in near-real-time (NRT).”**

**Ce que cela implique :**

* Les données SEO sont **streamées depuis une API REST** → donc **flux continu**, pas batch.
* Le traitement doit être en **near-real-time** → donc **latence minimale**, ingestion rapide.
* Le workflow doit être **entièrement dans Fabric** → pas d’Azure Event Hub, pas de Logic Apps.

**✅D. an eventstream**

**🧩 Justification finale**

* **Eventstream** est le composant Fabric dédié à l’ingestion de données en streaming.
* Il permet de connecter des **sources comme des API REST**, via des connecteurs personnalisés.
* Il est **optimisé pour le traitement near-real-time**, avec faible latence.
* Il respecte la contrainte de **n’utiliser que des services Fabric**.

## Question 103 LITWARE (ad-hoc issue)

<https://www.examprepper.co/exam/73/21>

You need to troubleshoot the ad-hoc query issue.

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

SELECT last\_run\_start\_time, last\_run\_command

FROM [Choice 1]

WHERE last\_run\_total\_elapsed\_time\_ms > 7200000

AND [Choice 2]

[Choice 1] :

1. queryinsights.exec\_requests\_history
2. queryinsights.exec\_sessions\_history
3. queryinsights.frequently\_run\_queries
4. queryinsights.long\_running\_queries

[Choice 2] :

1. max\_run\_total\_elapsed\_time\_ms > 7200000
2. median\_total\_elapsed\_time\_ms > 7200000
3. number\_of\_canceled\_runs > 1
4. number\_of\_failed\_runs > 1
5. number\_of\_runs > 1

Focus dans le use-case :

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.

✅**Choice 1** → queryinsights.long\_running\_queries

***Seule long\_running\_queries*** *contient toutes les métriques sur le nombre d’exécutions, d’annulations et d’échecs.*

✅**Choice 2** → number\_of\_failed\_runs > 1

*Deja max\_run.. & median\_total.. ne sont pas des colonnes possibles existantes : Eject*

*Mais comme 7200000ms = 2h .. on peut être piegé*

Précision

Queryinsights =

**✅ queryinsights.exec\_requests\_history**

* **Contenu** : chaque requête SQL terminée (pas en cours).

**A RETENIR**

* **DM = en cours,   vues Dynamic Management Views (DMVs).**
* **QueryInsights = historique  (30 jours)**
* **Travaillent toutes les 2 sur le WS uniquement**

## Question 105 LITWARE

<https://www.examprepper.co/exam/73/21>

What should you do to optimize the query experience for the business users?

A.Enable V-Order.

B.Create and update statistics.

C.Run the VACUUM command.

D.Introduce primary keys.

FOCUS:

**Contexte du use case**

“Business users perform ad-hoc queries against the warehouse.” “Reports sometimes run for two hours and fail to load as expected.” “The SQL query failed while running.”

➡️ Il y a un **problème de performance** dans le **warehouse**, et les requêtes sont **lentes ou échouent**.

**📌 Note importante**

Le use case précise que **V-Order is disabled** dans le gold layer (warehouse). Donc **l’activer** est une action directe pour améliorer les performances.

🟩 **A. Enable V-Order**

V-Order est une **optimisation de stockage** dans le warehouse Fabric. Elle améliore la **vitesse de lecture**, la **compression**, et les **performances de requêtes**.

# QUESTIONS GENERIQUES

# QUESTIONS RENCONTREES EN EXAM

## QUESTION 2 – 54 RLS

<https://www.examprepper.co/exam/73/1>

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

A. STORED PROCEDURE

B. CONSTRAINT

C. SCHEMA

D. FUNCTION

**D. FUNCTION** ✅ → C’est bien ce qu’il faut pour définir la logique de filtrage de RLS.

❌ **A. STORED PROCEDURE** → Pas utilisé pour filtrer automatiquement les lignes selon l’utilisateur connecté.

❌ **B. CONSTRAINT** → Gère l'intégrité des données, pas la sécurité.

❌ **C. SCHEMA** → Sert à regrouper des objets, pas à filtrer dynamiquement les lignes selon les utilisateurs.

## QUESTION 3

<https://www.examprepper.co/exam/73/1>

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

Data from the semantic models will be deployed to the target stage. YES/NO?

The Dataflow Gen1 dataflows will be deployed to the target stage. YES/NO?

The scheduled refresh policies will be deployed to the target stage. YES/NO?

❌NO : Les datas ne sont pas déployées

❌NO : Dataflow Gen1 est obsolete

✅YES : Meme si il faut les réactiver après deploiement

## QUESTION 4 – 52 (Tricky la 52)

<https://www.examprepper.co/exam/73/1>

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?

**👉1ere liste de Propositions (Question 4)**

A.GitHub Actions

B.a deployment pipeline

C.an Azure DevOps pipeline

✅ B.a deployment pipeline

**👉2de liste de Propositions (Question 52)**

A.an Azure DevOps pipeline

B.an eventstream

C.GitHub Actions

✅A. an Azure DevOps pipeline ✅ (plus cohérent avec écosystem FABRIC)

## QUESTION 5

<https://www.examprepper.co/exam/73/1>

You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse1.

You plan to deploy Warehouse1 to a new workspace named Workspace2.

As part of the deployment process, you need to verify whether Warehouse1 contains invalid references. The solution must minimize development effort.

What should you use?

A.a database project

B.a deployment pipeline

C.a Python script

D.a T-SQL script

**🟩 B. a deployment pipeline**

Dans Microsoft Fabric, les **deployment pipelines** permettent non seulement de déplacer des artefacts (comme un warehouse) entre les environnements **Dev → Test → Prod**, mais aussi de :

* **Analyser automatiquement les dépendances et références** (datasets, lakehouses, pipelines, etc.)
* **Détecter les références invalides** ou manquantes avant le déploiement
* **Minimiser l’effort de développement**, car tout est intégré dans l’interface Fabric

## QUESTION 6

<https://www.examprepper.co/exam/73/2>

You have a Fabric workspace that contains a Real-Time Intelligence solution and an eventhouse.

Users report that from OneLake file explorer, they cannot see the data from the eventhouse.

You enable OneLake availability for the eventhouse.

What will be copied to OneLake?

A.only data added to new databases that are added to the eventhouse

B.only the existing data in the eventhouse

C.no data

D.both new data and existing data in the eventhouse

E.only new data added to the eventhouse

la question suppose que c'est la **première activation** de la disponibilité OneLake.

Mais aussi que l’on coche le defaut ‘existing table’

**✅** D.both new data and existing data in the eventhouse

**Ref :**

|  |
| --- |
| dans la doc , <https://learn.microsoft.com/fr-fr/fabric/real-time-intelligence/event-house-onelake-availability>  il y a cette phrase " ImportantDésactiver disponibilité oneLake supprime vos données de OneLake. Lorsque vous **réactivez la disponibilité**, seules les nouvelles données sont mises à disposition dans OneLake sans retour des données supprimées." |
| il est également écrit :  **Activer la disponibilité de OneLake** Vous pouvez activer la disponibilité de OneLake au niveau d'une base de données KQL ou d'une table. Pour activer la Disponibilité de OneLake, accédez à la section OneLake du volet d’informations de votre base de données ou table KQL. Définissez la Disponibilité sur Activée. Capture d’écran de la section OneLake du volet Détails de la base de données montrant l’option Disponibilité mise en surbrillance. La base de données s’actualise automatiquement. Capture d’écran des détails de la section OneLake dans Real-Time Intelligence une fois la disponibilité définie sur Activé. L’option permettant d’exposer des données à OneLake est activée. Vous avez activé la disponibilité de OneLake dans votre base de données KQL. Vous pouvez désormais accéder à toutes les nouvelles données ajoutées à votre base de données au chemin OneLake donné au format Delta Lake. Vous pouvez également choisir de créer un raccourci OneLake à partir d’un Lakehouse ou d’un Data Warehouse, ou d’interroger les données directement via le mode Direct Lake de Power BI.  Ambiguité avec cette phrase : "Vous pouvez désormais accéder à toutes les nouvelles données ajoutées à votre base de données au chemin OneLake donné au format Delta Lake" ..  **toutes nouvelles données ! on considere seulement nouvelles ..**  **.. mais sauf si existing table (defaut) a été cochée** |

Depuis la question, je suppose que l’option Also apply to existing tables a été ajoutée… ca change tout… ☹

Donc aucune certitudes même après avoir fait le test dans Fabric .. et avec acces powerbi Desktop .. ces règles ne semblent plus s’appliquer

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

Le contenu généré par l’IA peut être incorrect.

## Question 7

<https://www.examprepper.co/exam/73/2>

You have a Fabric workspace named Workspace1.

You plan to integrate Workspace1 with Azure DevOps.

You will use a Fabric deployment pipeline named deployPipeline1 to deploy items from Workspace1 to higher environment workspaces as part of a medallion architecture. You will run deployPipeline1 by using an API call from an Azure DevOps pipeline.

You need to configure API authentication between Azure DevOps and Fabric.

Which type of authentication should you use?

A.service principal

B.Microsoft Entra username and password

C.managed private endpoint

D.workspace identity

**✅ A. service principal**

Pour appeler l’API Fabric (ex. déclencher deployPipeline1) depuis un pipeline Azure DevOps :

* Tu dois **enregistrer une application dans Microsoft Entra ID** (Azure AD).
* Lui attribuer les **permissions API Fabric** (via rôle ou scope).
* Utiliser le **client ID / secret** pour obtenir un **token OAuth2**.
* Appeler l’API Fabric avec ce token.

## Question 8 (/10)

<https://www.examprepper.co/exam/73/2>

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

Copilot

Lorsque le **cache des raccourcis est activé** dans un workspace Fabric :

* **Les données sont mises en cache automatiquement** lorsqu’un raccourci est **accédé**.
* **Le cache est conservé pendant 24 heures** après le dernier accès.
* **Au-delà de 24 heures sans accès**, les données ne sont plus servies depuis le cache, mais directement depuis la source (ici GCS).

Mistral

**Règles de Cache pour les Raccourcis :**

* **Cache Activé :** Le cache pour les raccourcis est activé dans le workspace.
* **Condition de Mise en Cache :** Les données d'un raccourci sont mises en cache si le raccourci a été accédé récemment. Par défaut, les données des raccourcis accédées dans les dernières 24 heures sont mises en cache.

**Analyse des Raccourcis :**

* **Products (ProductFile) :** Dernier accès il y a 12 heures. Cela signifie que les données sont encore dans la fenêtre de 24 heures et seront donc récupérées depuis le cache.
* **Stores (StoreFile) :** Dernier accès il y a 4 heures. Cela signifie que les données sont encore dans la fenêtre de 24 heures et seront donc récupérées depuis le cache.
* **Trips (TripsFile) :** Dernier accès il y a 48 heures. Cela signifie que les données ne sont plus dans la fenêtre de 24 heures et ne seront donc pas récupérées depuis le cache.

ChatGpt

**cache des raccourcis Fabric**

* Quand on crée un **shortcut** dans un lakehouse vers un stockage externe (GCS, ADLS, S3…), Fabric peut mettre en place un **cache** pour améliorer les performances.
* **Règle clé** (documentation officielle Microsoft) :
  + Le cache est utilisé uniquement si **la dernière lecture a eu lieu il y a moins de 24 heures**.
  + Si le raccourci n’a pas été accédé depuis plus de 24 heures, les données sont relues depuis la source externe.

**✅ C. Stores and Products only**

## Question 9

<https://www.examprepper.co/exam/73/2>

You have a Fabric workspace named Workspace1 that contains an Apache Spark job definition named Job1.

You have an Azure SQL database named Source1 that has public internet access disabled.

You need to ensure that Job1 can access the data in Source1.

What should you create?

A.an on-premises data gateway

B.a managed private endpoint

C.an integration runtime

D.a data management gateway

**✅ B. Managed private endpoint**  
→ Fonctionnalité de Fabric permettant à un **Spark job** ou à un **pipeline** d’accéder à une ressource Azure privée (ex. SQL DB, Storage) **sans passer par Internet**, via un endpoint managé. ✅

## Question 10 (/8)

<https://www.examprepper.co/exam/73/2>

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

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

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Stored in | Size | Description |
| ProductFile | Storage1 | 50 MB | Contains a list of products and their details |
| TripsFile | Storage2 | 2 GB | Contains one month’s worth of taxi trip data |
| StoredFile | Storage2 | 25 MB | Contains a list of stores and their addresses |

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

* A shortcut to ProductFile aliased as Products
* A shortcut to StoreFile aliased as Stores
* A shortcut to TripsFile aliased as Trips

The data from which shortcuts will be retrieved from the cache

**Propositions**

A.Trips and Stores only

B.Products and Store only

C.Stores only

D.Products only

E.Products, Stores, and Trips

1. Il doit provenir d’une **source externe compatible** :
   * ✅ Amazon S3
   * ✅ S3-compatible
   * ✅ Google Cloud Storage (GCS)
   * ✅ On-premises via Gateway

Non supporté pour

* + ❌ ADLS Gen2
  + ❌ OneLake interne

1. Il doit faire **≤ 1 Go**
2. Il doit avoir été accédé récemment (selon la politique de rétention, par défaut 24h)

✅ C.Stores only = AS3 + < 1Go

## Question 11

<https://www.examprepper.co/exam/73/3>

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

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

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

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

Notebook1 must execute every weekday at 8:00 AM.

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

Model1 must refresh when Notebook1 has executed successfully.

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

**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.

✅ **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**

## Question 12

<https://www.examprepper.co/exam/73/3>

You have a Fabric workspace.

You have semi-structured data.

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

What should you use to store the data?

A.a lakehouse

B.an eventhouse

C.a datamart

D.a warehouse

✅ A.a lakehouse

⚠️ Ne pas penser que KQK = eventhouse. Sous prétexte que eventhouse n’accepte **que** KQL

* Le **Lakehouse** est le **seul artefact dans Fabric** qui permet une **interopérabilité complète entre Spark, T-SQL et KQL**, tout en supportant les **formats semi-structurés**.

## Question 13

<https://www.examprepper.co/exam/73/3>

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

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

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

Create a new domain for the sales department.

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

Assign Workspace1 to the east region subdomain.

Assign Workspace2 to the west region subdomain.

The solution must follow the principle of least privilege.

Which role should you assign to the user?

A.workspace Admin

B.domain admin

C.domain contributor

D.Fabric admin

✅ B.domain admin

Pas de piege ! Création Domain mais less privilege : Domain admin

## Question 14

<https://www.examprepper.co/exam/73/3>

You have a Fabric workspace named Workspace1 that contains a warehouse named DW1 and a data pipeline named Pipeline1.

You plan to add a user named User3 to Workspace1.

You need to ensure that User3 can perform the following actions:

View all the items in Workspace1.

Update the tables in DW1.

The solution must follow the principle of least privilege.

You already assigned the appropriate object-level permissions to DW1.

Which workspace role should you assign to User3?

A.Admin

B.Member

C.Viewer

D.Contributor

✅ D.Contributor

Pas de piège pour le rôle sur les artefacts du WS

Et précision faite et important pour le DWH  : (sinon ca marche pas .. )

CONTRIBUTOR permet de **modifier les objets** (comme les tables dans DW1), **à condition que les permissions spécifiques au niveau de l'objet (DW1) aient déjà été configurées** (ce qui est le cas ici).

## Question 15

<https://www.examprepper.co/exam/73/3>

You have a Fabric capacity that contains a workspace named Workspace1. Workspace1 contains a lakehouse named Lakehouse1, a data pipeline, a notebook, and several Microsoft Power BI reports.

A user named User1 wants to use SQL to analyze the data in Lakehouse1.

You need to configure access for User1. The solution must meet the following requirements:

Provide User1 with read access to the table data in Lakehouse1.

Prevent User1 from using Apache Spark to query the underlying files in Lakehouse1.

Prevent User1 from accessing other items in Workspace1.

What should you do?

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

B.Assign User1 the Viewer role for Workspace1. Share Lakehouse1 with User1 and select Read all SQL endpoint data.

C.Share Lakehouse1 with User1 directly and select Build reports on the default semantic model.

D.Assign User1 the Member role for Workspace1. Share Lakehouse1 with User1 and select Read all SQL endpoint data.

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

⚠️

Au début je pense : C. Share Lakehouse1 and select Build reports on the default semantic model

Mais ❌ Cela donne accès au modèle sémantique, pas aux tables SQL du lakehouse. Ne permet pas d’écrire des requêtes SQL sur les tables.

## Question 16

[Microsoft - DP-700 - Page 4 | Examprepper](https://www.examprepper.co/exam/73/4)

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?

Entity 1: ?  
Entity 2: ?  
Entity 3: ?  
Entity 4: ?

Propositions :  
Certified

Master data

Promoted

Cannot be endorsed

Reponses :

Entity 1: MASTER  
Entity 2: CERTIFIED .. meme sans approbation mentionnée, ca matche avec les standard quand meme !  
Entity 3: PROMOTED  
Entity 4: CERTIFIED

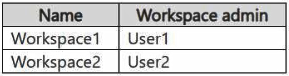
## QUESTION 17

<https://www.examprepper.co/exam/72/4>

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

You have the Fabric workspaces shown in the following table.

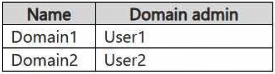
|  |  |
| --- | --- |
| Name | Workspace Admin |
| Workspace1 | User1 |
| Workspace2 | User2 |



You have a security group named Group1 that contains User1 and User3.

The Fabric admin creates the domains shown in the following table.

|  |  |
| --- | --- |
| Name | Domaine Admin |
| Domain1 | User1 |
| Domain2 | User2 |



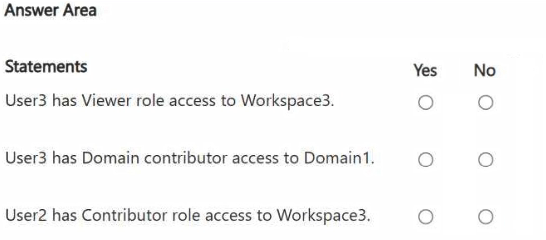
User1 creates a new workspace named Workspace3.

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.

| **Statement** | **Yes** | **No** |
| --- | --- | --- |
| User3 has Viewer role access to Workspace3. | ⭘ | ⭘ |
| User3 has Domain contributor access to Domain1. | ⭘ | ⭘ |
| User2 has Contributor role access to Workspace3. | ⭘ | ⭘ |



**✅ Background Summary:**

**Users:**

* **User1**
* **User2**
* **User3**

**Workspaces:**

* **Workspace1**: Admin = User1
* **Workspace2**: Admin = User2
* **Workspace3**: Created by **User1**

**Security Group:**

* **Group1** = {User1, User3}

**Domains:**

* **Domain1**: Admin = User1
* **Domain2**: Admin = User2

**Action Taken:**

* **Group1 is added to the default domain of Domain1**

**Statement 1:**

**❌ NO pour User3 has Viewer role access to Workspace3.**

🟡 **user3 devient contributor sur Domain1 par le biais de son appartenance à group1 mais n’pas de rôle défini sur le Workspace**

* **User3 n’a aucun rôle explicite dans Workspace3, ni directement ni via son appartenance à Group1.**
* **Son accès est délégué par le domaine (Domain1), ce qui lui permet d’interagir avec les objets Fabric du domaine, y compris ceux présents dans Workspace3**
* **Sans apparaître dans la liste des membres du workspace.**

**Statement 2:**

**✅ YES pour User3 has Domain contributor access to Domain1.**

🟡 **Analysis:**

* **Group1**, which includes User3, was **added to the *default domain* of Domain1**.
* In Fabric, **adding a user or group to the default domain grants the "Domain contributor" role**.
* That means **User3 now has domain contributor permissions for Domain1.**

**Statement 3:**

**❌ NO pour User2 has Contributor access to Workspace3.**

🟡 **Analysis:**

* **Aucun rapport entre User2 et Workspace3 , ni pas une attribution directe ou indirecte**

**✅ Final Answers:**

| **Statement** | **Yes** | **No** |
| --- | --- | --- |
| User3 has Viewer role access to Workspace3. |  | ✅ |
| User3 has Domain contributor access to Domain1. | ✅ |  |
| User2 has Contributor role access to Workspace3. |  | ✅ |

| **Statement** | **Réponse** | **Justification** |
| --- | --- | --- |
| **User3 has Viewer role access to Workspace3.** | ❌ **Non** | User3 n’a pas reçu explicitement le rôle Viewer. En tant que membre de Group1, il devient **Domain Contributor**, ce qui donne un accès étendu au objets du domain1 **mais aucun rôle (pas même viewer ) direct ou indirecte sur le Workspace**. |
| **User3 has Domain contributor access to Domain1.** | ✅ **Oui** | Group1 est ajouté au domaine par défaut de Domain1. Selon la documentation officielle, ses membres deviennent automatiquement **Domain Contributors** pour les workspaces assignés à ce domaine. |
| **User2 has Contributor role access to Workspace3.** | ❌ **Non** | User2 n’a aucun lien avec Workspace3 ni avec Domain1. Aucun rôle ne lui a été attribué. |

## Question 18

<https://www.examprepper.co/exam/73/4>

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.

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

Pas de piège

## Question 19

<https://www.examprepper.co/exam/73/4>

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.

## QUESTION 20

[Microsoft - DP-700 - Page 4 | Examprepper](https://www.examprepper.co/exam/73/4)

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.  
1.Change the runtime version

2.Install the libraries  
3.Create a pool  
4.Create an environment

5.Set the default environment

1. **Create an environment**  
   → Crée un environnement personnalisé dans lequel tu pourras installer les bibliothèques nécessaires.
2. **Install the libraries**  
   → Dans cet environnement, installe les bibliothèques Python que ton équipe utilise (via pip ou conda).
3. **Set the default environment**  
   → Configure Workspace1 pour utiliser cet environnement comme **environnement par défaut** pour tous les nouveaux notebooks.

## Question 21

<https://www.examprepper.co/exam/73/5>

You have a Fabric workspace that contains a lakehouse and a notebook named Notebook1. Notebook1 reads data into a DataFrame from a table named Table1 and applies transformation logic. The data from the DataFrame is then written to a new Delta table named Table2 by using a merge operation.

You need to consolidate the underlying Parquet files in Table1.

Which command should you run?

A.VACUUM

B.BROADCAST

C.OPTIMIZE

D.CACHE

✅ C.OPTIMIZE

Optimize ! entre autres les petits fichiers

## Question 22

<https://www.examprepper.co/exam/73/5>

You have five Fabric workspaces.

You are monitoring the execution of items by using Monitoring hub.

You need to identify in which workspace a specific item runs.

Which column should you view in Monitoring hub?

A.Start time

B.Capacity

C.Activity name

D.Submitter

E.Item type

F.Job type

G.Location

✅ G.Location

Pas de piège

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

Le contenu généré par l’IA peut être incorrect.

## Question 23

[Microsoft - DP-700 - Page 5 | Examprepper](https://www.examprepper.co/exam/73/5)

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

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

You need to copy data from Database1 to Warehouse1.

Which item should you use?

A.a Dataflow Gen1 dataflow

B.a data pipeline

C.a KQL queryset

D.a notebook

✅ B.a data pipeline

Pas de piège

## Question 24

[Microsoft - DP-700 - Page 5 | Examprepper](https://www.examprepper.co/exam/73/5)

You have a Fabric workspace that contains a warehouse named DW1. DW1 is loaded by using a notebook named Notebook1.

You need to identify which version of Delta was used when Notebook1 was executed.

What should you use?

A.Real-Time hub

B.OneLake data hub

C.the Admin monitoring workspace

D.Fabric Monitor

E.the Microsoft Fabric Capacity Metrics app

✅ D.Fabric Monitor

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

Le contenu généré par l’IA peut être incorrect.

## Question 25

[Microsoft - DP-700 - Page 5 | Examprepper](https://www.examprepper.co/exam/73/5)

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

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

CREATE TABLE dbo.DimCustomer (

CustomerKey VARCHAR(255) NOT NULL,

Name VARCHAR(255) NOT NULL,

Email VARCHAR(255) NOT NULL

);

You need to set the Customerkey column as a primary key of the DimCustomer table.

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

1. DROP CONSTRAINT PK\_DimCustomer
2. ADD CONSTRAINT PK\_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)
3. NOT ENFORCED
4. ALTER TABLE dbo.DimCustomer
5. ADD CONSTRAINT PK\_DimCustomer PRIMARY KEY CLUSTERED (CustomerKey)
6. ENFORCED

* **ALTER TABLE dbo.DimCustomer** : Indique que vous modifiez la table existante.
* **ADD CONSTRAINT PK\_DimCustomer PRIMARY KEY NONCLUSTERED (CustomerKey)** : ***CLUSTERED*** *non supporté en Fabric, donc pas de clustered.*
* **NOT ENFORCED** :

***Obligatoire Dans Fabric Warehouse, toutes les contraintes de clé primaire doivent être déclarées avec NOT ENFORCED.***

***ENFORCED non supporté sur les clefs primaires dans Fabric***

## Question 26

[Microsoft - DP-700 - Page 6 | Examprepper](https://www.examprepper.co/exam/73/6)

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

You need to dynamically execute and monitor the refresh progress of Model1.

What should you use?

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

B.Monitoring hub

C.dynamic management views in Azure Data Studio

D.a semantic link in a notebook

✅ B.Monitoring hub

**"dynamically execute and monitor"** signifie que tu veux :

* **Déclencher un rafraîchissement à la demande** (pas seulement via une planification automatique)
* **Suivre en temps réel** l’état d’avancement du processus

Mais cela **n’implique pas nécessairement une automatisation ou un déclenchement programmatique**

## 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 :

Copier

CREATE TABLE [warehouse2].[dbo].[CreditCard]

(

CreditCard varchar(20) NOT NULL,

CreditCardType varchar(10) NOT NULL

)

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.

The statement to complete is :

Copier

[Choice 1] TABLE dbo.CreditCard

[Choice 2] COLUMN [CreditCard]

ADD MASKED

WITH (FUNCTION = '[Choice 3] (2,'XXXXXXXX',4))

**Choice 1 :**

* ADD
* ALTER
* CREATE
* DEFAULT
* DROP
* EMAIL
* PARTIAL
* REPLACE
* UPDATE

**Choice 2 :**

* ADD
* ALTER
* CREATE
* DEFAULT
* DROP
* EMAIL
* PARTIAL
* REPLACE
* UPDATE
* MASKED

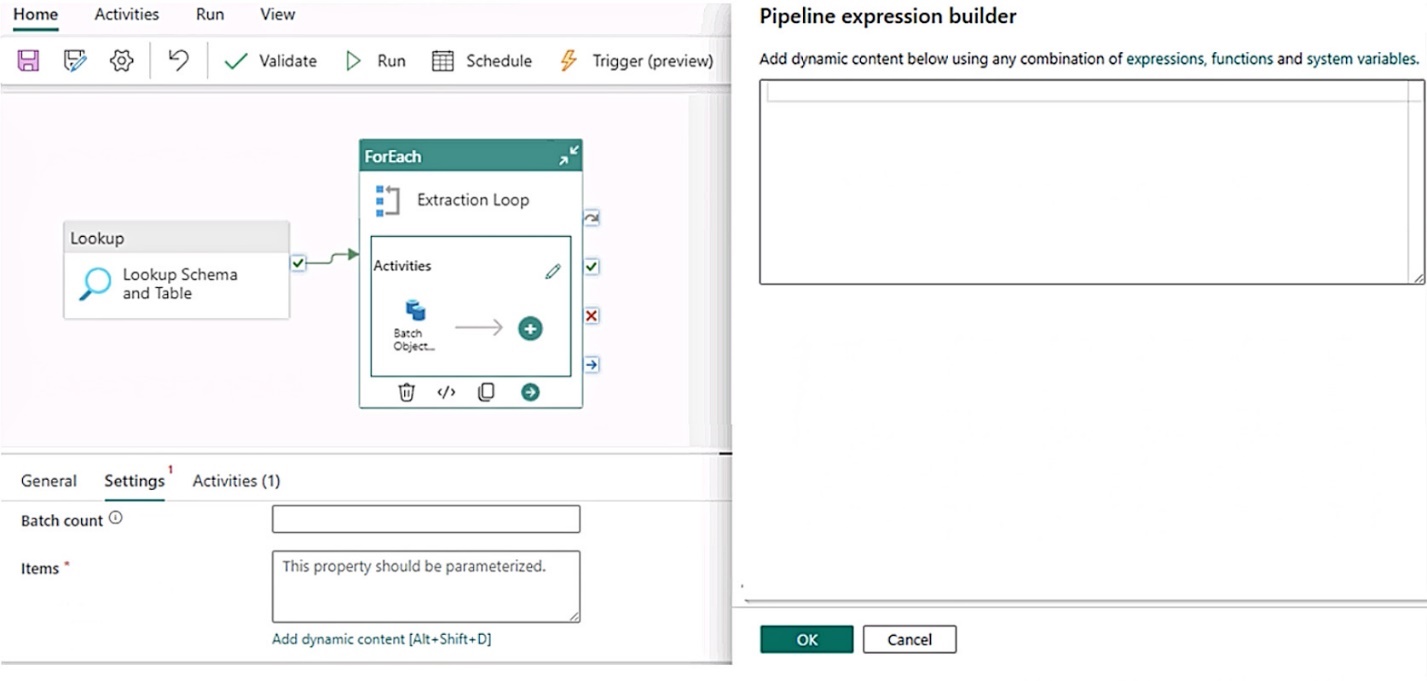
**Choice 3 :**

* ALTER
* CREATE
* DEFAULT
* DROP
* EMAIL
* PARTIAL
* REPLACE
* UPDATE

## Question 39

[Microsoft - DP-700 - Page 8 | Examprepper](https://www.examprepper.co/exam/73/8)

You are building a data orchestration pattern by using a Fabric data pipeline named Dynamic Data Copy as shown in the exhibit. (Click the Exhibit tab.)

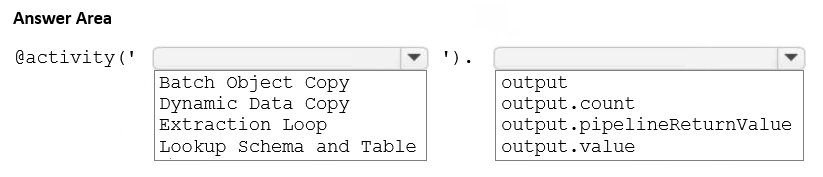


Dynamic Data Copy does NOT use parametrization.

You need to configure the ForEach activity to receive the list of tables to be copied.

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

NOTE: Each correct selection is worth one point.



**@activity('Lookup Schema and Table').output.value**

## Question 41

<https://www.examprepper.co/exam/73/9>

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

• A Microsoft Power BI report named Report1

• A Power BI dashboard named Dashboard1

• A semantic model named Model1

• A lakehouse name Lakehouse1

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

Which items can you endorse in Fabric?

A.Lakehouse1, Model1, and Dashboard1 only

B.Lakehouse1, Model1, Report1 and Dashboard1

C.Report1 and Dashboard1 only

D.Model1, Report1, and Dashboard1 only

E.Lakehouse1, Model1, and Report1 only

✅ E.Lakehouse1, Model1, and Report1 only

<https://learn.microsoft.com/en-us/fabric/governance/endorsement-overview#types-of-items-that-can-be-endorsed>

: Types of items that can be endorsed **All Fabric items and Power BI items** except Power BI dashboards can be promoted or certified.

## Question 42

<https://www.examprepper.co/exam/73/9>

You have a Fabric workspace named Workspace1.

Your company acquires GitHub licenses.

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

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

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

B.Actions (Read and write) only

C.Contents (Read and write) only

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

✅ C.Contents (Read and write) only

permission minimale requise

## Question 43

<https://www.examprepper.co/exam/73/9>

You have a Fabric workspace named Workspace1.

You plan to configure Git integration for Workspace1 by using an Azure DevOps Git repository.

An Azure DevOps admin creates the required artifacts to support the integration of Workspace1.

Which details do you require to perform the integration?

A.the organization, project, Git repository, and branch

B.the personal access token (PAT) for Git authentication and the Git repository URL

C.the project, Git repository, branch, and Git folder

D.the Git repository URL and the Git folder

✅ A.the organization, project, Git repository, and branch

 you need to specify:

1. **Azure DevOps organization** – e.g., https://dev.azure.com/your-org
2. **Project name** – the Azure DevOps project that contains the repo
3. **Git repository name** – the repo where Fabric artifacts will be stored
4. **Branch name** – the branch to sync with (typically main or master)

These are required to link the workspace to the correct location in Azure DevOps.

## Question 44

<https://www.examprepper.co/exam/73/9>

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.

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.

A,B,C

✅ 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.

## Question 45 -55

<https://www.examprepper.co/exam/73/9>

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

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

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

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

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

Provide the best query performance.

Minimize operational costs.

Which should you do?

A.Change the MD5 hash to SHA256.

B.Increase the capacity.

C.Enable V-Order.

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

E.Create views.

✅ C.Enable V-Order.

C’est fait pour optimize !

## Question 55 – 45

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

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

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

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

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

• Provide the best query performance.

• Minimize operational costs.

Which should you do?

A.Create views.

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

C.Change the MD5 hash to SHA256.

D.Increase the capacity.

E.Disable V-Order on the warehouse.

✅ A.Create views.

Si **V-Order est déjà activé**, alors **l’option A (Create views)** est probablement **la meilleure première action**

## Question 47

<https://www.examprepper.co/exam/73/10>

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

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

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

• A data pipeline that loads a CSV file to Lakehouse1

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

Which type of item should you use?

A.notebook

B.warehouse

C.Dataflow Gen2 dataflow

D.data pipeline

✅ D.data pipeline

Ya pas de pieges

## Question 46

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

DAG = {

"activities": [

{

"name": "execute\_notebook\_1",

"path": "notebook\_01",

"timeoutPerCellInSeconds": 600,

"args": {

"input\_value": "999"

},

"retry": 1,

"retryIntervalInSeconds": 30

},

{

"name": "execute\_notebook\_2",

"path": "notebook\_02",

"timeoutPerCellInSeconds": 400,

"args": {

"input\_value": "888"

},

"retry": 1,

"retryIntervalInSeconds": 30

},

{

"name": "execute\_notebook\_3",

"path": "notebook\_03",

"timeoutPerCellInSeconds": 600,

"args": {

"input\_value": "777"

},

"retry": 1,

"retryIntervalInSeconds": 30

},

{

"name": "execute\_notebook\_3",

"path": "notebook\_03",

"timeoutPerCellInSeconds": 600,

"args": {

"input\_value": "777"

},

"retry": 1,

"retryIntervalInSeconds": 30

}

],

"timeoutInSeconds": 43200,

"concurrency": 0

}

mssparkutils.notebook.runMultiple(DAG, {"displayDAGviaGraphviz": True})

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

1. Notebook\_03

2. Notebook\_01

3. Notebook\_02

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

NOTE: Each correct selection is worth one point.

**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.

**✅ Réponse “attendue par l’examen”**

* **D** → ajouter la dépendance pour Notebook\_02
* **F** → déclarer Notebook\_03 en “racine” (bien qu’en pratique ce soit inutile)

**Piège** : l’examen veut **deux actions**, et l’une d’elles est un faux indicateur visuel (“move top”) pour faire croire que l’ordre JSON importe.

 **D** seule → assure que Notebook\_02 attend Notebook\_01

 **F** → **aucun effet réel**, juste un “piège QCM”

 Pour un **ordre strict réel** : il faut ajouter explicitement une dépendance :

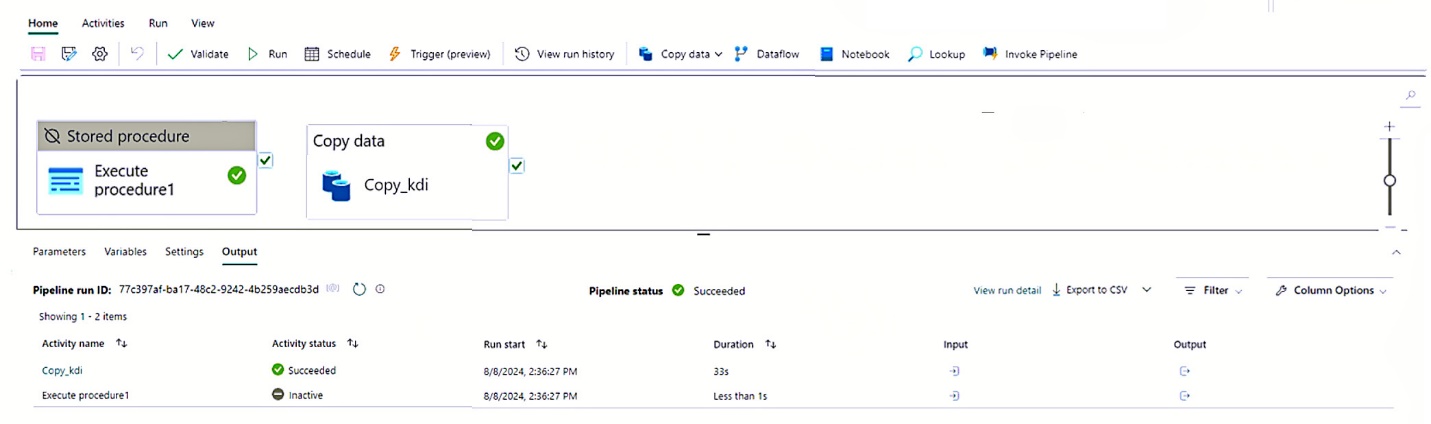
* Notebook\_01 dépend de Notebook\_03
* Notebook\_02 dépend de Notebook\_01

En réalité Notebook\_3 va s’executer en parallèle …

## Question 49 EXAMPREPPER : PIPELINE

[Microsoft - DP-700 - Page 10 | Examprepper](https://www.examprepper.co/exam/73/10)

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?

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.

**✅ Correct Answer: D. Copy\_kdi will run and Execute procedure1 will be skipped.**

**🧠 Justification :**

* **Copy\_kdi** :
  + Statut actuel : **Succeeded** → elle a été exécutée avec succès.
  + Rien n’indique qu’elle est désactivée → elle **sera exécutée à nouveau** lors du prochain run.
* **Execute procedure1** :
  + Statut : **Inactive** dans le panneau de sortie.
  + Pas de flèche de dépendance.
  + Présence d’une **coche verte** → pas d’erreur, mais **pas de run**.
  + Cela indique qu’elle est **désactivée** ou **conditionnée**, donc **elle sera ignorée** lors du prochain run **tant que son état ne change pas**.

**❌ Pourquoi les autres réponses sont incorrectes :**

| **Option** | **Pourquoi c’est faux** |
| --- | --- |
| A | Execute procedure1 est inactive → elle ne sera pas exécutée. |
| B | L’ordre est inversé, et Execute procedure1 ne s’exécutera pas. |
| C | Copy\_kdi n’est pas désactivée → elle sera exécutée. |
| E | Pas de parallélisme ici, et une activité est inactive. |
| F | Copy\_kdi est active → elle sera exécutée. |

## Question 51

<https://www.examprepper.co/exam/73/11>

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

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

You have an internal Microsoft Entra user named User1 that has an email address of 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?

**GRANT [Choice 1] ON**

**Customers(CustomerID, FirstName, LastName, Phone)**

**TO [Choice2]**

**Propositions**

**Choice 1 :**

1. **ALTER**
2. **EXECUTE**
3. **READ**
4. **SELECT**
5. **VIEW**

**Choice 2 :**

1. **User1**
2. **[User1]**
3. **[User1@contoso.com]**

**GRANT SELECT ON**

**Customers(CustomerID, FirstName, LastName, Phone)**

**TO [User1@contoso.com]**

👉 **Parmi les propositins , avec GRANT , seuls ALTER, EXECUTE et SELECT sont possibles**

⚠️ **READ = inexistant en GRANT**

👁**️‍**🗨**️ Phrase clef memo : Entra Id = @mail et** ⚠️ → Dans Microsoft Fabric (SQL endpoint), on utilise l’adresse mail complète de l’utilisateur Entra ID, entre crochets. [User1@contoso.com]

## Question 52 - 4 Tricky la 52

<https://www.examprepper.co/exam/73/11>

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?

**👉2de liste de Propositions (Question 52)**

A.an Azure DevOps pipeline

B.an eventstream => ce sera celui là

C.GitHub Actions

✅A. an Azure DevOps pipeline ✅ (plus cohérent avec écosystem FABRIC)

**👉1ere liste de Propositions (Question 4)**

A.GitHub Actions

B.a deployment pipeline

C.an Azure DevOps pipeline

✅ B.a deployment pipeline

## Question 54 - 2 RLS

<https://www.examprepper.co/exam/73/11>

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

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

You need to ensure that the sales representatives can see only their respective data.

Which warehouse object do you require to implement RLS?

A.SECURITY POLICY

B.TABLE

C.TRIGGER

D.STORED PROCEDURE

**✅ A. SECURITY POLICY**

**Explication :**

* **SECURITY POLICY** :
  + Dans un **Fabric Warehouse**, la sécurité au niveau des lignes (RLS) est implémentée en créant une **politique de sécurité** (SECURITY POLICY).
  + Une politique de sécurité utilise une **fonction en ligne (inline table-valued function)** pour filtrer les lignes de la table en fonction de l'utilisateur connecté.

Donc function derrière..

## Question 56

<https://www.examprepper.co/exam/73/12>

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

|  |  |
| --- | --- |
| **Table name** | **Column name** |
| SalesOrderDetail | ProductID |
| SalesOrderDetail | ModifiedDate |
| SalesOrderDetail | OrderQty |
| Product | ProductID |
|  |  |
| Product | Name |

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

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

SELECT

[ Choice 1 ? ] (SO.ModifiedDate AS DATE) AS OrderDate,

P.Name AS ProductName,

SUM(SO.OrderQty) AS OrderQty

FROM [dbo].[SalesOrderDetail] SO

INNER JOIN [dbo].[Product] P

ON P.ProductID = SO.ProductID

GROUP BY

[ Choice 2 ? ]

ORDER BY OrderDate

**Propositions**

**Propositions**

Choice 1:

1. CAST
2. CONVERT
3. YEAR

Choice 2 :

1. CUBE(YEAR(SO.ModifiedDate), P.Name)
2. GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate)))
3. ROLLUP(YEAR(SO.ModifiedDate), P.Name)
4. YEAR(SO.ModifiedDate), P.Name

Question Pas claire dans l'objectif. Mais je comprends par "yearly sum of order quantities for all the products in each row" comme un cumul annuel de TOUS les produits présent dans chaque lignes de l'année

✅ Choice 1 : YEAR ( On veut grouper par Année)  
✅ Choice 2 : 2.GROUPING SETS ((YEAR(SO.ModifiedDate), P.Name), (YEAR(SO.ModifiedDate))) : YEAR, PRODUCT, SUM(QTE ANNEE )

**💡 Astuce :**  
👁️‍🗨️ phrase clef : yearly sum of order quantities for all the products in each row . Reflexion la plus longue ...réponse est la fonction la plus longue ...

## Question 57

<https://www.examprepper.co/exam/73/12>

You have a Fabric workspace named Workspace1 that contains a warehouse named Warehouse1.

You plan to deploy Warehouse 1 to a new workspace named Workspace2.

As part of the deployment process, you need to verify whether Warehouse1 contains invalid references. The solution must minimize development effort and provide detailed information about the invalid references.

What should you use?

A.a dbt project

B.a deployment pipeline

C.a Python script

D.a database project

✅ B.a deployment pipeline

Y a pas de piege

## Question 58

<https://www.examprepper.co/exam/73/12>

You have a Fabric workspace that contains a lakehouse named Lakehouse1. Data is ingested into Lakehouse1 as one flat table. The table contains the following columns.

|  |  |
| --- | --- |
| **Name** | **Description** |
| TransactionID | Contains a unique ID for each transaction |
| Date | Contains the date of a transaction |
| ProductID | Contains a unique ID for each product |
| ProductColor | Contains a descriptive attribute that describes the color of each product |
| ProductName | Contains a unique name for each product |
| SalesAmount | Contains the sales amount of a transaction |

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

You need to prepare the data.

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

**Propositions**

A.Date

B.ProductName

C.ProductColor

D.TransactionID

E.SalesAmount

F.ProductID

## Question 59

<https://www.examprepper.co/exam/73/12>

You have a Fabric workspace named Workspace1 that contains a notebook named Notebook1.

In Workspace1, you create a new notebook named Notebook2.

You need to ensure that you can attach Notebook2 to the same Apache Spark session as Notebook1.

What should you do?

A.Enable high concurrency for notebooks.

B.Enable dynamic allocation for the Spark pool.

C.Change the runtime version.

D.Increase the number of executors.

**✅ A. Enable high concurrency for notebooks**

Dans Fabric, pour que plusieurs notebooks partagent **la même session Spark**, il faut que le **Spark pool** soit configuré en **mode haute concurrence**. Ce mode permet :

* Le **partage de session Spark** entre notebooks.
* Une **exécution parallèle** plus efficace.
* Une **réduction du temps de démarrage** des notebooks.

Remarque : **B. Enable dynamic allocation** : utile pour ajuster les ressources dynamiquement, mais **ne permet pas le partage de session**.

## Question 60

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.

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.

✅ **D. Assigner au data engineer le rôle Contributor sur Workspace1**  
👉 Étape indispensable pour qu’il puisse écrire dans *Customer* (les droits d’écriture se gèrent au niveau du workspace).

✅ **F. Créer un nouveau workspace (Workspace2) avec un nouveau lakehouse (Lakehouse2)**  
👉 On prépare un espace cloisonné qui isolera la table sensible (*Employee*).

✅ **E. Migrer la table *Employee* vers Lakehouse2**  
👉 On retire la donnée PII de Lakehouse1, ce qui garantit que le Contributor n’aura jamais accès à ces informations sensibles, même indirectement.

## Question 62 - 96

<https://www.examprepper.co/exam/73/13>

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

BikepointID -

Street -

Neighbourhood -

Latitude -

Longitude -

No\_Bikes -

No\_Empty\_Docks -

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

What should you use?

A.a KQL queryset

B.an eventstream

C.a streaming dataset

D.Apache Spark Structured Streaming

✅ B.an eventstream

Pas de piege, on peut faire du filtrage dans l’eventstream

* Connecteur natif avec **Azure Event Hubs**.
* Permet de filtrer en temps réel (via transformations intégrées low-code).
* Peut écrire directement dans un **Lakehouse**.
* Bénéficie de la **rétention Event Hub** (configurable jusqu’à 7 jours) → donc les 2 jours requis sont garantis côté Event Hub.
* Minimise l’effort de développement.

Exemple filtrage

# Version Transformation avec filtrage  (exclure < 10 , = conserver >= 10)

df\_filtered = df\_stream.filter(df\_stream["value"] >= 10)

df\_transformed = df\_filtered.withColumn("value\_squared", df\_filtered["value"] \*\* 2)

## QUESTION 63 - 95

<https://www.examprepper.co/exam/73/13>

You are building a data loading pattern for Fabric notebook workloads. You have the following code segment:

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

Le contenu généré par l’IA peut être incorrect.

|  |
| --- |
| **def loading\_pattern\_sample(df\_source):**  **try:**  **deltaTable = DeltaTable.forName(spark, target\_table)**  **except Exception:**  **try:merge**  **df\_source.write.format('delta').mode('overwrite').saveAsTable(f"{target\_table}")**  **except Exception as e:**  **print(f'Load for table {target\_table} failed with error: {str(e)}')**  **raise**  **return**  **try:**  **change\_detection\_columns = [col for col in df\_source.columns if col not in candidate\_key]**  **match\_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate\_key])**  **update\_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change\_detection\_columns])**  **update\_expr = {col: f'source.{col}' for col in df\_source.columns}**  **merge\_operation = deltaTable.alias('target').merge(**  **source=df\_source.alias('source'),**  **condition=match\_condition**  **).whenMatchedUpdate(**  **condition=update\_condition,**  **set=update\_expr**  **).whenNotMatchedInsertAll()**  **merge\_operation.execute()**  **except Exception as e:**  **print(f'Insert operation for table {target\_table} failed with error: {str(e)}')**  **return** |

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

1. The target table will always be overwritten YES/NO
2. The merge operation will always run YES/NO
3. The loading pattern supports both full and incremental loading requirements YES/NO

*( 95 : The loading pattern supports both full and incremental loading requirements)*

❌ 1 : NO . Overwritten est activé lors de la création de table (target\_table), La création de table n’intervient que si son «assignation» échoue dans la ligne « **deltaTable = DeltaTable.forName(spark, target\_table) »**

❌ 2 : NO . Seulement si l’assignation de la table destination (target\_table) réussit (signifiant qu’elle existe), sinon, on entre dans le except de création de la table destination (target\_table) qui termine avec un return.

✅ 3. YES :

- la table existe (son assignation réussit) : On merge

- la table n’existe pas (son assignation échoue) : On overwrite (full )

*En fait, pour toujours charger ( même si table n’est pas accessible et donc créée) il faudrait poursuivre gérer le chargement de la nouvelle table pour poursuivre .. et donc supprimer le Return*

|  |
| --- |
| def loading\_pattern\_sample(df\_source):  try:  deltaTable = DeltaTable.forName(spark, target\_table)  except Exception:  try:  # Création d'une nouvelle table Delta avec les données de df\_source  df\_source.write.format('delta').mode('overwrite').saveAsTable(f"{target\_table}")  **# Rechargement de la table nouvellement créée**  **deltaTable = DeltaTable.forName(spark, target\_table)**  except Exception as e:  # En cas d'échec de la création de la table, on imprime l'erreur et on la relance  print(f'Load for table {target\_table} failed with error: {str(e)}')  raise  *~~return~~*  try:  # Identification des colonnes à surveiller pour les changements  change\_detection\_columns = [col for col in df\_source.columns if col not in candidate\_key]  # Création des conditions pour la fusion  match\_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate\_key])  update\_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change\_detection\_columns])  update\_expr = {col: f'source.{col}' for col in df\_source.columns}  # Exécution de l'opération de fusion pour mettre à jour ou insérer des données  merge\_operation = deltaTable.alias('target').merge(  source=df\_source.alias('source'),  condition=match\_condition  ).whenMatchedUpdate(  condition=update\_condition,  set=update\_expr  ).whenNotMatchedInsertAll()  merge\_operation.execute()  except Exception as e:  # En cas d'échec de l'opération de fusion, on imprime l'erreur  print(f'Insert operation for table {target\_table} failed with error: {str(e)}') |

## Question ???

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

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

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

Le contenu généré par l’IA peut être incorrect.

## Question 64

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

SELECT

OrderLineID order\_line\_id,

OrderDate order\_date,

c.customer\_key,

c.customer\_id,

Quantity order\_quantity,

UnitPrice unit\_price,

TaxRate tax\_rate

FROM

Lakehouse1.orderlines o

INNER JOIN

Lakehouse2.dim\_customer c

ON o.customerid = c.customer\_id

AND

[Choice 1]

AND

[Choice 2]

Choice 1 :

1. c.is\_current = 1
2. o.OrderDate > c.valid\_to\_datetime
3. o.OrderDate >= c.valid\_from\_datetime

Choice 2 :

1. c.is\_current = 1
2. o.OrderDate < c.valid\_to\_datetime
3. o.OrderDate <= c.valid\_from\_datetime

!! c.is\_current = 1 ?

Cette condition ne permettrait de récupérer que la version actuelle du client.

Analyse historique

**Choice 1**: 3.OrderDate >= valid\_from\_datetime **FROM**

**Choice 2:** 2. OrderDate < valid\_to\_datetime **TO**

* Permet de retrouver la version du client valide à la date de la commande

## Question 65 – 66 - 110

<https://www.examprepper.co/exam/73/13>

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

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

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

Trigger the process when a new file is added.

Provide the highest throughput.

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

A.Eventstream

B.Dataflow Gen2

C.Streaming dataset

D.Data pipeline

✅ D.Data pipeline

500go !

| **Rang** | **Option** | **Pertinence** | **Justification** |
| --- | --- | --- | --- |
| 1️⃣ | **Data pipeline** | ✅ **Le plus adapté** | Conçu pour l’ingestion de **fichiers massifs**, avec **débit élevé** via l’activité Copy. Peut être déclenché par planification ou via API/Logic App. |
| 2️⃣ | **Eventstream** | ⚠️ Partiellement adapté | Réagit à des événements (Event Hub, IoT Hub), mais **pas conçu pour des fichiers de 500 Go**. Idéal pour des **flux d’événements légers**, pas pour du batch massif. |
| 3️⃣ | **Dataflow Gen2** | ❌ Peu adapté | Ne supporte **pas le déclenchement automatique** basé sur l’arrivée de fichiers. Moins performant pour des fichiers volumineux. |
| *~~4️⃣~~* | ***~~KQL Queryset~~*** | *~~❌ Inadapté~~* | *~~Sert uniquement à~~* ***~~interroger des données temps réel~~*** *~~dans un KQL DB,~~* ***~~pas à les ingérer~~****~~.~~* |
| *~~5️⃣~~* | ***~~Environment~~*** | *~~❌ Hors sujet~~* | *~~Ce n’est~~* ***~~pas un outil d’ingestion~~****~~, mais un conteneur de configuration pour Spark/Notebooks.~~* |

## Question 66 - 65 - 110

<https://www.examprepper.co/exam/73/13>

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

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

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

Trigger the process when a new file is added.

Provide the highest throughput.

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

A.Eventstream

B.Environment

C.KQL queryset

D.Data pipeline

✅ D.Data pipeline

| **Rang** | **Option** | **Pertinence** | **Justification** |
| --- | --- | --- | --- |
| 1️⃣ | **Data pipeline** | ✅ **Le plus adapté** | Conçu pour l’ingestion de **fichiers massifs**, avec **débit élevé** via l’activité Copy. Peut être déclenché par planification ou via API/Logic App. |
| 2️⃣ | **Eventstream** | ⚠️ Partiellement adapté | Réagit à des événements (Event Hub, IoT Hub), mais **pas conçu pour des fichiers de 500 Go**. Idéal pour des **flux d’événements légers**, pas pour du batch massif. |
| *~~3️⃣~~* | ***~~Dataflow Gen2~~*** | *~~❌ Peu adapté~~* | *~~Ne supporte~~* ***~~pas le déclenchement automatique~~*** *~~basé sur l’arrivée de fichiers. Moins performant pour des fichiers volumineux.~~* |
| 4️⃣ | **KQL Queryset** | ❌ Inadapté | Sert uniquement à **interroger des données temps réel** dans un KQL DB, **pas à les ingérer**. |
| 5️⃣ | **Environment** | ❌ Hors sujet | Ce n’est **pas un outil d’ingestion**, mais un conteneur de configuration pour Spark/Notebooks. |

## Question 67

<https://www.examprepper.co/exam/73/14>

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

A table named Table1 -

A table named Table2 -

An update policy named Policy1 -

Policy1 sends data from Table1 to Table2.

The following is a sample of the data in Table2.

|  |  |  |
| --- | --- | --- |
| Timestamp (datetime) | DeviceId (guid) | StreamData (dynamic) |
| 2024-05-18 12:45:17.16524 | 81416f30-60a2-4e75-9b19-2a84ea059735 | [  {  "index": 0,  "eventid": "719afca0-be30-4559-bb5e-59fade642f6"  }  ] |
| 2024-05-18 12:45:21.76423 | bb664e1e-02aa-4e17-8c8a-116cd4458fd2 | [  {  "index": 0,  "eventid": "782222b2-fbcb-43c0-82d6-ecd49a99dbf5"  }  ] |
| 2024-05-18 12:45:23.98642 | 717bfe7d-0e5d-498f-9f21-e60aaf258056 | [  {  "index": 0,  "eventid": "d5730286-0da4-41f8-8e59-f75e209310a9"  }  ] |

Recently, the following actions were performed on Table1:

An additional element named temperature was added to the StreamData column.

The data type of the Timestamp column was changed to date.

The data type of the 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.

|  |  |  |  |
| --- | --- | --- | --- |
| A. | Timestamp (datetime) | DeviceId (guid) | StreamData (dynamic) |
|  | 2024-05-18 | 81416f30-60a2-4e75-9b19-2a84ea059735 | [  {  "index": 40,  "eventid": "729afca2-be30-4559-bb5e-59fade642f3",  "temperature": 32  }  ] |
| B. | Timestamp (datetime) | DeviceId (guid) | StreamData (dynamic) |
|  | 2024-05-21 | 81416f30 | [  {  "index": 0,  "eventid": "719afca0-be30-4559-bb5e-5werade642f6",  "temperature": 27  }  ] |
| C. | Timestamp (datetime) | DeviceId (guid) | StreamData (dynamic) |
|  | 2024-05-23 | 81416f3060a24e759b192a84ea059735dhdyte3 | [  {  "index": 0,  "eventid": "719afca0-be30-4559-bb5e-59fade642f6"  }  ] |
| D. | Timestamp (datetime) | DeviceId (guid) | StreamData (dynamic) |
|  | 2024-05-24 | 81416f30-60a2-4e75-9b19-2a84ea059735 | [  {  "index": 0,  "eventid": "719afca0-be30-4559-bb5e-59fade642f6"  }  ] |

✅ A : car dynamic correct

✅ D : car dynamic : correct meme si pas de nouvelle col

Dynamic (schema flexible)

**Et surtout : indice: GUID incorrect sur B & C !!!!**

**A**

* Timestamp = 2024-05-18 → c’est une date (sans heures).
* Dans Table2 on attend datetime.
* Conversion implicite date → datetime est **valide** (exemple : 2024-05-18 00:00:00).
* DeviceId = guid valide (81416f30-60a2-4e75-9b19-2a84ea059735).
* StreamData contient "temperature" mais c’est du dynamic → aucun problème.  
  ✅ **A passe.**

**D**

* Timestamp = date → convertible vers datetime. ✔
* DeviceId = 81416f30-60a2-4e75-9b19-2a84ea059735 → GUID valide. ✔
* StreamData = dynamic correct.  
  ✅ **D passe.**

## Question 68

<https://www.examprepper.co/exam/73/14>

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.

Select item\_id as Itemid

[ Choice 1 ? ] as ItemName

, item.description as ItemDecription

[ Choice 2 ? ] as PurchaseDate

FROM

Table1

WHERE

Item\_type = @Itemtype\_parameter

**Propositions**

Choice 1 :

1. .convert(varchar(20), item\_name)
2. .convert(varchar(max), item\_name)
3. .try.cast(item\_name as varchar(20))

Choice 2 :

1. .convert(varchar, purchase\_date, 7)
2. .convert(varchar, purchase\_date, 109)
3. .convert(varchar, purchase\_date, 112)

* **Choice 1 : 3. try\_cast(item\_name as varchar(20))**
* **Choice 2 : 1. .convert(varchar, purchase\_date, 7)**

Les IA sont completement au fraise avec cette question 7 existe bel et bien

<https://learn.microsoft.com/fr-fr/sql/t-sql/functions/cast-and-convert-transact-sql?view=sql-server-ver17>

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

Le contenu généré par l’IA peut être incorrect.

## Question 69

<https://www.examprepper.co/exam/73/14>

You are developing a data pipeline named Pipeline1.

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

What should you configure?

A.Degree of copy parallelism

B.Fault tolerance

C.Enable staging

D.Enable logging

✅ C.Enable staging

This is crucial when the **source and sink(destination) are different types of systems**, especially **Snowflake → Fabric warehouse**.

## Question 70 - Q MULTIPLE

<https://www.examprepper.co/exam/73/14>

Q MULTIPLE : You have a KQL database that contains two tables named Stream and Reference.

Stream contains streaming data in the following format.

|  |  |
| --- | --- |
| **Column name** | **Data type** |
| Timestamp | Datetime |
| GeoLocation | Dynamic |
| Temperature | Decimal |
| Deviceld | Int |

Reference contains reference data in the following format.

|  |  |
| --- | --- |
| **Column name** | **Data type** |
| Deviceld | Int |
| DeviceName | String |

Both tables contain millions of rows.

You have the following KQL queryset.

Stream

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

| join kind=inner Reference on DeviceId

| project Timestamp, lat, long, Temperature, DeviceName

| filter Temperature >= 10

| render scatterchart with (kind = map)

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

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

Does this meet the goal?

A.Yes

B.No

❌ NO

En plus ca change tout .. Changer le type de jointure de \*\*inner\*\* à \*\*outer\*\* **n’améliore pas les performances** — au contraire, cela peut **les dégrader :**

Une **outer join** (par exemple leftouter) conserve **toutes les lignes de la table de gauche**, même si aucune correspondance n’est trouvée dans la table de droite.

## Question 71 – Q MULTIPLE

<https://www.examprepper.co/exam/73/15>

Solution: You change project to extend.

Does this meet the goal?

A.Yes

B.No

❌ NO

Ca change tout aussi ici …

* **extend** ajoute ou modifie des colonnes **sans supprimer les autres**.
* **project** sélectionne **uniquement les colonnes spécifiées**, ce qui **réduit le volume de données** en sortie.

## Question 72 – Q MULTIPLE

<https://www.examprepper.co/exam/73/15>

Solution: You move the filter to line 02.

Does this meet the goal?

A.Yes

B.No

✅YES

Effectivement ,  **KQL est séquentiel**

Déplacer le **filtre Temperature >= 10 avant la jointure** dans la requête KQL **améliore les performances**, car cela **réduit le volume de données** à traiter **avant** la jointure.

## Question 73 – Q MULTIPLE

<https://www.examprepper.co/exam/73/15>

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

Does this meet the goal?

A.Yes

B.No

❌ NO

Au contraire !!

Ca fait un tableau !

C’est une fonction **d’agrégation** qui regroupe des valeurs dans une liste dynamique.

## Question 77

<https://www.examprepper.co/exam/73/16>

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.

**ProductCategory, ProductSubcategory and Product must be:**

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

**The joining key must be:**

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

**ProductCategory, ProductSubcategory and Product must be**

✅ Denormalized into a single product dimension table

**The joining key must be**

✅ The unique system generated identifier

## Question 78

[Microsoft - DP-700 - Page 16 | Examprepper](https://www.examprepper.co/exam/73/16)

Your company has three newly created data engineering teams named Team1, Team2, and Team3 that plan to use Fabric. The teams have the following personas:

• Team1 consists of members who currently use Microsoft Power BI. The team wants to transform data by using by a low-code approach.

• Team2 consists of members that have a background in Python programming. The team wants to use PySpark code to transform data.

• Team3 consists of members who currently use Azure Data Factory. The team wants to move data between source and sink environments by using the least amount of effort.

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

What should you recommend for each team?

1. Data pipeline
2. Notebook (Spark)
3. Dataflow Gen2

| **Équipe** | **Persona** | **Outil recommandé** |
| --- | --- | --- |
| Team1 | transform ! Power BI, low-code | **Dataflow Gen2** |
| Team2 | Python ! Python, PySpark | **Notebook (Spark)** |
| Team3 | move data between source  + low effort !!  Azure Data Factory | **Data pipeline** |

## Question 79 – Similaire … à 78

<https://www.examprepper.co/exam/73/16>

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?**

Dataset1 :

1.Dataflow Gen2 dataflow

2.A notebook

3.A T-SQL statement

Dataset2 :

1.Dataflow Gen2 dataflow

2.A notebook

3.A T-SQL statement

Dataset3 :

1.Dataflow Gen2 dataflow

2.A KQL queryset

3.A T-SQL statement

**✅ Correction des réponses**

* **Dataset1 : T-SQL statement**
* **Dataset2 : Dataflow Gen2 dataflow**
* **Dataset3 : T-SQL statement**

Piege Dataset2 :

* **“Development mode”** → on est dans une phase de **préparation / shaping des données**, pas encore dans l’analyse finale (Power BI reports).
* **“Understand and work with the dataset”** → ça fait penser à du **profiling interactif** (voir la distribution des colonnes, identifier les valeurs nulles, repérer les anomalies).
* **“Custom visuals”** dans ce contexte → ce sont les **visualisations intégrées de Power Query / Dataflow Gen2** (exemple : petits graphiques sous les colonnes, histogrammes de distribution, indicateurs de qualité des données).
* Tout cela est du **low-code/no-code**, exactement comme tu dis : **mode dev visuel pour transformer les données**.

## Question 80

[Microsoft - DP-700 - Page 16 | Examprepper](https://www.examprepper.co/exam/73/16)

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

• Key

• Status

• LastModified

The data source contains a table named Status\_Source that has the same columns as Status\_Target. Status\_Source is used to populate Status\_Target.

In a notebook name Notebook1, you load Status\_Source to a DataFrame named sourceDF and Status\_Target to a DataFrame named targetDF.

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

• For all the matching records that have the same value of key, update the value of LastModified in Status\_Target to the value of LastModified in Status\_Source.

• Insert all the records that exist in Status\_Source that do NOT exist in Status\_Target.

• Set the value of Status in Status\_Target to inactive for all the records that were last modified more than seven days ago and that do NOT exist in Status\_Source.

How should you complete the statement?

…

(targetDF

.merge(sourceDF, "sourceDF.Key" = "targetDF.Key")

[Choice 1]

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

)

(values = {

"targetDF.Key": "sourceDF.Key",

"targetDF.LastModified": "sourceDF.LastModified",

"targetDF.Status": "sourceDF.Status"

}

[Choice 2]

)

[Choice 3]

condition="targetDF.LastModified" > (current\_date() - INTERVAL '7' DAY)",

set = {"targetDF.Status": ""inactive""}

)

.execute()

…

Liste des propositions pour chaque choix :

Choice 1 :

1. .whenMatchedInsert(
2. .whenMatchedUpdate(
3. .whenNotMatchedBySourceInsert(
4. .whenNotMatchedBySourceUpdate(
5. .whenNotMatchedInsert(
6. .whenNotMatchedUpdate(

Choice 2 :

1. .whenMatchedInsert(
2. .whenMatchedUpdate(
3. .whenNotMatchedBySourceInsert(
4. .whenNotMatchedBySourceUpdate(
5. .whenNotMatchedInsert(
6. .whenNotMatchedUpdate(

Choice 3 :

1. .whenMatchedInsert(
2. .whenMatchedUpdate(
3. .whenNotMatchedBySourceInsert(
4. .whenNotMatchedBySourceUpdate(
5. .whenNotMatchedInsert(
6. .whenNotMatchedUpdate(

✅ Choice 1 .whenMatchedUpdate(

✅ Choice 2 .whenNotMatchedInsert(

✅ Choice 3 .whenNotMatchedBySourceUpdate(

## Question 81

[Microsoft - DP-700 - Page 17 | Examprepper](https://www.examprepper.co/exam/73/17)

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

In a warehouse, you create a control table named Control.Object as shown.

|  |  |  |
| --- | --- | --- |
|  | Schema\_name | table\_name |
| 1 | Warehouse | ColdRoomTemperatures |
| 2 | Warehouse | Colors |
| 3 | Warehouse | PackageTypes |
| 4 | Warehouse | StockGroups |
| 5 | Warehouse | StockItems |
| 6 | dbo | BuildVersion |
| 7 | dbo | ErrorLog |
| 8 | Application | SystemParameters |
| 9 | Purchasing | PurchaseOrderLines |
| 10 | Purchasing | PurchaseOrders |

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

Which three actions should you perform in sequence?

Propositions

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

✅ 1. Lookup : Récupérer la liste des tables à copier.

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

✅ 2. ForEach : Itérer sur chaque table de la liste.

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

✅ 3. Copy data : Copier les données de chaque table source vers le lakehouse.

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

## Question 82

<https://www.examprepper.co/exam/73/17>

You are implementing a medallion architecture in a Fabric lakehouse.

You plan to create a dimension table that will contain the following columns:

• ID

• CustomerCode

• CustomerName

• CustomerAddress

• CustomerLocation

• ValidFrom

• ValidTo

You need to ensure that the table supports the analysis of historical sales data by customer location at the time of each sale.

Which type of slowly changing dimension (SCD) should you use?

A.Type 2

B.Type 0

C.Type 1

D.Type 3

✅ A.Type 2

0 : Rien

1 : Ecrase

2 : Historise

3 : 1 seule Version

## Question 84

<https://www.examprepper.co/exam/73/17>

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

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

Which windowing function should you use?

A.sliding

B.hopping

C.tumbling

D.session

✅ C.tumbling

 grouped by non-overlapping and contiguous time intervals of one minute

**la fenêtre tumbling est la plus simple et la plus couramment utilisée**

(voir en annexe . ANNEXE Question 84 )

## Question 85

[Microsoft - DP-700 - Page 17 | Examprepper](https://www.examprepper.co/exam/73/17)

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

| **SalesOrderNumber** | **OrderDate** | **CustomerName** | **Email** |
| --- | --- | --- | --- |
| SO49172 | 2021-01-01 | Brian Howard | [brian23@adventure-works.com](mailto:brian23@adventure-works.com) |
| SO49173 | 2021-01-01 | Linda Alvarez | [linda19@adventure-works.com](mailto:linda19@adventure-works.com) |
| SO49174 | 2021-01-01 | Gina Hernandez | [gina4@adventure-works.com](mailto:gina4@adventure-works.com) |
| SO49178 | 2021-01-01 | Beth Ruiz | [beth4@adventure-works.com](mailto:beth4@adventure-works.com) |
| SO49179 | 2021-01-01 | Evan Ward | [evan13@adventure-works.com](mailto:evan13@adventure-works.com) |

You have a notebook that contains the following code segment.

01 df = df.withColumn("CustomerName", when((col("CustomerName").isNull()) | (col("CustomerName") == ""), lit("Unknown")).otherwise(col("CustomerName")))

02 df = df.withColumn("Username", split(col("Email"), "@").getItem(0))

03 df = df.dropDuplicates(["OrderDate"]).select(col("OrderDate"), year("OrderDate").alias("Year"), ("CustomerName"), ("Username"))

04 display(df.head(10))

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

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Statement** | **Yes** | **No** | | --- | --- | --- | | Line 01 will replace all the null and empty values in the CustomerName column with the Unknown value. | ⭘ | ⭘ | | Line 02 will extract the value before the @ character and generate a new column named Username. | ⭘ | ⭘ | | Line 03 will extract the year value from the OrderDate column and keep only the first occurrence for each year. | ⭘ | ⭘ | |

01: ✅YES ok la commande avec CustomerName").isNull()) est assez clair

02: ✅YES c’est bien un split sur @

03: ❌NO le dropduplicate est sur Orderdate ( Orderdate = la date entier pas seulement l’année !!!! )

## Question 86

[Microsoft - DP-700 - Page 18 | Examprepper](https://www.examprepper.co/exam/73/18)

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

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

|  |  |  |
| --- | --- | --- |
| **Timestamp** | **Deviceld** | **StreamData** |
| 2024-05-18 12:45:17.16524 | 81416f30-60a2-4e75-9b19-2a84ea059735 | { |
|  |  | "index": 0, |
|  |  | "eventid": "719afca0-be30-4559-bb5e-59fade642f6", |
|  |  | "isActive": false, |
|  |  | "latitude": 5.390012, |
|  |  | "longitude": -40.100235, |
|  |  | "tags": ["tempor"] |
|  |  | } |
| 2024-05-18 12:45:21.76423 | bb664e1e-02aa-4e17-8c8a-116cd4458d52 | { |
|  |  | "index": 0, |
|  |  | "eventid": "782222b2-fbcb-43c0-82d6-ecd49a99dbf5", |
|  |  | "isActive": true, |
|  |  | "latitude": -56.153786, |
|  |  | "longitude": 130.870907, |
|  |  | "tags": ["adipisicing"] |
|  |  | } |
| 2024-05-18 12:45:23.98642 | 717bfe7d-0e5d-498f-9f21-e60aaf258056 | { |
|  |  | "index": 0, |
|  |  | "eventid": "d5730286-0da4-41f8-8e59-f75e209310a9", |
|  |  | "isActive": true, |
|  |  | "latitude": -21.39289, |
|  |  | "longitude": 123.959442, |
|  |  | "tags": ["ad"] |
|  |  | } |
| 2024-05-18 12:45:25.39523 | 1a390e71-4af4-4df5-a479-2238d84001f7 | { |
|  |  | "index": 0, |
|  |  | "eventid": "9572e141-8692-4d16-89e8-002f9b7e22b6", |
|  |  | "isActive": true, |
|  |  | "latitude": -84.902614, |
|  |  | "longitude": -11.499007, |
|  |  | "tags": ["ex"] |
|  |  | } |
| 2024-05-18 12:45:27.43343 | 2f0ba7d0-6dff-4081-bf7f-d0d39d4c3260 | { |
|  |  | "index": 0, |
|  |  | "eventid": "08a42b87-ce84-4bb2-99f0-4fb5c75ff63f", |
|  |  | "isActive": true, |
|  |  | "latitude": -49.909339, |
|  |  | "longitude": -177.505775, |
|  |  | "tags": ["laboris"] |
|  |  | } |

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

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

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

**2 . .create table EventStreamData (** *(piege , nom de table different mais syntaxe ok)*

**5 . TimeStamp:datetime, DeviceId:string**  (2 1ere colonnes)

**3 . StreamData:dynamic )**  ( derniere colonne)

## Question 87

[Microsoft - DP-700 - Page 18 | Examprepper](https://www.examprepper.co/exam/73/18)

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

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

You need to copy data from Database1 to Warehouse1.

Which item should you use?

A.a data pipeline

B.an Apache Spark job definition

C.a streaming dataflow

D.a notebook

✅ A.a data pipeline

Qui correspond à la définition :

**"The primary use of the Copy Data activity in a data pipeline is to ingest data from an external source into a lakehouse file or table."**

## Question 88

[Microsoft - DP-700 - Page 18 | Examprepper](https://www.examprepper.co/exam/73/18)

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

DimCustomer contains 100 columns and 20 million rows. The columns are of various data types, including int, varchar, date, and varbinary.

You need to identify incoming changes to the table and update the records when there is a change.

The solution must minimize resource consumption.

What should you use to identify changes to attributes?

A. a 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.

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

la fonction de Hash transforme toutes les valeurs d'un enregistrement en une seule chaîne de caractères (ou une séquence de bits) de longueur fixe. Courte et plus facile ensuite avec un autre enregistrement d'un autre table => plus rapide , moins couteux que comparer colone par colonne

En fait il faudrait A puis D mais

la question semble se concentrer sur **l'identification des changements entrants** (incoming changes). Autrement dit, elle demande comment détecter les changements dans les données sources avant de les comparer à la table de destination.

## Question 89

[Microsoft - DP-700 - Page 18 | Examprepper](https://www.examprepper.co/exam/73/18)

You have an Azure SQL database named DB1.

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

You discover that events are NOT being propagated to EventStreamDB1.

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

What should you do?

A.Create a read-only replica of DB1.

B.Create an Azure Stream Analytics job.

C.Enable Extended Events for DB1.

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

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

Pour activer CDC sur **Azure SQL Database**, vous devez d'abord activer CDC au niveau de la base de données, puis pour chaque table spécifique. Cela permet de capturer les modifications apportées aux tables et de les rendre disponibles pour des outils comme **Fabric Eventstream**.

## Question 92

[Microsoft - DP-700 - Page 19 | Examprepper](https://www.examprepper.co/exam/73/19)

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

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

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

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

How should you complete the query? To answer, drag the appropriate values the correct targets. Each value may be used once, more than once, or not at all.

Readings

| filter City == "Kansas"

[choice 1] by Timestamp

[choice 2] PrevMeterReading = prev(MeterReading),

PrevTimestamp = prev(Timestamp),

[choice 3] City, Area, MeterReading, Timestamp, PrevMeterReading, PrevTimestamp

Propositions

1. evaluate
2. extend
3. lookup
4. project
5. sort
6. summarize
7. take

**✅ sort**

**✅ extend**

**✅ project**

* **[choice 1] → sort by Timestamp**  
  ✅ On trie les données par Timestamp pour garantir que prev() s’applique correctement à la ligne précédente dans l’ordre temporel.
* **[choice 2] → extend PrevMeterReading = prev(MeterReading), PrevTimestamp = prev(Timestamp)**  
  ✅ extend permet d’ajouter des colonnes calculées. Ici, on utilise prev() pour récupérer la valeur précédente de MeterReading et Timestamp.
* **[choice 3] → project City, Area, MeterReading, Timestamp, PrevMeterReading, PrevTimestamp**  
  ✅ project sélectionne les colonnes à afficher dans le résultat final.

## Question 93

[Microsoft - DP-700 - Page 19 | Examprepper](https://www.examprepper.co/exam/73/19)

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

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

The solution must minimize development effort.

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

Voici la transcription des choix pour chaque source en listes de choix texte :

**Source1 :**

1. Apache Spark Structured Streaming
2. An eventstream
3. A data pipeline
4. A streaming dataflow

**Source2 :**

1. Apache Spark Structured Streaming
2. An eventstream
3. A data pipeline
4. A streaming dataflow

**Source3 :**

1. Apache Spark Structured Streaming
2. An eventstream
3. A data pipeline
4. A streaming dataflow

✅ **Source1 → Eventstream**

✅ **Source2 → Eventstream**

✅ **Source3 → Apache Spark Structured Streaming**

**🚫 Règle éliminatoire**

* **Streaming dataflow** : limité à **1 MB par message** → **éliminé d’office** pour les 3 sources (10 MB, 25 MB, 5 MB).
* **Data pipeline** = batch, pas du streaming → **éliminé d’office** pour les 3 sources

**Source1 (10 MB, semi-structured)**

**✅ Eventstream**

* **Spark Structured Streaming** = lourd en dev.
* **Eventstream** accepte jusqu’à 30 MB → ✅ réponse optimale.

**Source2 (25 MB, structured, 19 colonnes)**

**✅ Eventstream**

* **Spark Structured Streaming** = lourd en dev.
* **Eventstream** gère 25 MB structuré → ✅.

**Source3 (5 MB, unstructured, images)**

**✅ Spark Structured Streaming**

* Seul **Spark Structured Streaming** sait traiter de l’unstructured en continu.

De plus .. **Eventstream** ne gère pas les **blobs/images**.

👉 **Streaming Dataflow** : direct à rejeter si > 1 MB ou données unstructured.

👉 **Eventstream** : à rejeter si > 30 MB ou données unstructured.

👉 **Pipeline** : toujours rejeter si on demande du *streaming natif* (c’est batch/near real-time).

## Question 95 – 63

<https://www.examprepper.co/exam/73/19>

You are building a data loading pattern for Fabric notebook workloads. You have the following code segment:

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

Le contenu généré par l’IA peut être incorrect.

|  |
| --- |
| **def loading\_pattern\_sample(df\_source):**  **try:**  **deltaTable = DeltaTable.forName(spark, target\_table)**  **except Exception:**  **try:merge**  **df\_source.write.format('delta').mode('overwrite').saveAsTable(f"{target\_table}")**  **except Exception as e:**  **print(f'Load for table {target\_table} failed with error: {str(e)}')**  **raise**  **return**  **try:**  **change\_detection\_columns = [col for col in df\_source.columns if col not in candidate\_key]**  **match\_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate\_key])**  **update\_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change\_detection\_columns])**  **update\_expr = {col: f'source.{col}' for col in df\_source.columns}**  **merge\_operation = deltaTable.alias('target').merge(**  **source=df\_source.alias('source'),**  **condition=match\_condition**  **).whenMatchedUpdate(**  **condition=update\_condition,**  **set=update\_expr**  **).whenNotMatchedInsertAll()**  **merge\_operation.execute()**  **except Exception as e:**  **print(f'Insert operation for table {target\_table} failed with error: {str(e)}')**  **return** |

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

1. The target table will always be overwritten YES/NO
2. The merge operation will always run YES/NO
3. The loading pattern supports both full and incremental loading requirements YES/NO

*( 63 : The code support incremental loading and the initial table creation)*

❌ 1 : NO . Overwritten est activé lors de la création de table (target\_table), La création de table n’intervient que si son «assignation» échoue dans la ligne « **deltaTable = DeltaTable.forName(spark, target\_table) »**

❌ 2 : NO . Seulement si l’assignation de la table destination (target\_table) réussit (signifiant qu’elle existe), sinon, on entre dans le except de création de la table destination (target\_table) qui termine avec un return.

✅ 3. YES :

- la table existe (son assignation réussit) : On merge

- la table n’existe pas (son assignation échoue) : On overwrite (full )

*En fait, pour toujours charger ( même si table n’est pas accessible et donc créée) il faudrait poursuivre gérer le chargement de la nouvelle table pour poursuivre .. et donc supprimer le Return*

|  |
| --- |
| def loading\_pattern\_sample(df\_source):  try:  deltaTable = DeltaTable.forName(spark, target\_table)  except Exception:  try:  # Création d'une nouvelle table Delta avec les données de df\_source  df\_source.write.format('delta').mode('overwrite').saveAsTable(f"{target\_table}")  **# Rechargement de la table nouvellement créée**  **deltaTable = DeltaTable.forName(spark, target\_table)**  except Exception as e:  # En cas d'échec de la création de la table, on imprime l'erreur et on la relance  print(f'Load for table {target\_table} failed with error: {str(e)}')  raise  *~~return~~*  try:  # Identification des colonnes à surveiller pour les changements  change\_detection\_columns = [col for col in df\_source.columns if col not in candidate\_key]  # Création des conditions pour la fusion  match\_condition = ' AND '.join([f'target.{col} = source.{col}' for col in candidate\_key])  update\_condition = ' OR '.join([f'target.{col} != source.{col}' for col in change\_detection\_columns])  update\_expr = {col: f'source.{col}' for col in df\_source.columns}  # Exécution de l'opération de fusion pour mettre à jour ou insérer des données  merge\_operation = deltaTable.alias('target').merge(  source=df\_source.alias('source'),  condition=match\_condition  ).whenMatchedUpdate(  condition=update\_condition,  set=update\_expr  ).whenNotMatchedInsertAll()  merge\_operation.execute()  except Exception as e:  # En cas d'échec de l'opération de fusion, on imprime l'erreur  print(f'Insert operation for table {target\_table} failed with error: {str(e)}') |

## Question 96 - 62

[Microsoft - DP-700 - Page 20 | Examprepper](https://www.examprepper.co/exam/73/20)

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

* BikepointID -
* Street -
* Neighbourhood -
* Latitude -
* Longitude -
* No\_Bikes -
* No\_Empty\_Docks -

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

What should you use?

A.an eventstream

B.Apache Spark Structured Streaming

C.a streaming dataset

D.a KQL queryset

✅ B.an eventstream

Pas de piege, on peut faire du filtrage dans l’eventstream

* Connecteur natif avec **Azure Event Hubs**.
* Permet de filtrer en temps réel (via transformations intégrées low-code).
* Peut écrire directement dans un **Lakehouse**.
* Bénéficie de la **rétention Event Hub** (configurable jusqu’à 7 jours) → donc les 2 jours requis sont garantis côté Event Hub.
* Minimise l’effort de développement.

Exemple filtrage

# Version Transformation avec filtrage  (exclure < 10 , = conserver >= 10)

df\_filtered = df\_stream.filter(df\_stream["value"] >= 10)

df\_transformed = df\_filtered.withColumn("value\_squared", df\_filtered["value"] \*\* 2)

## Question 100

<https://www.examprepper.co/exam/73/20>

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

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

You need to denormalize the tables and include the ContractType and StartDate columns in the Employee table. The solution must meet the following requirements:

Ensure that the StartDate column is of the date data type.

Ensure that all the rows from the Employee table are preserved and include any matching rows from the Contract table.

Ensure that the result set displays the total number of employees per contract type for all the contract types that have more than two employees.

How should you complete the statement?

WITH result AS (

SELECT e.EmployeeID

, e.EmployeeName,

, e.EmployeePosition,

, c.ContractType,

, [Choice1] (date, c.StartDate) as StartDate

FROM Employee AS e

[Choice2] Contract AS c on c.EmployeeID = e.EmployeeID

)

SELECT

COUNT(DISTINCT EmployeeID) AS TotalEmployees,

ContractType

FROM result

GROUP BY ContractType

[Choice3] COUNT(DISTINCT EmployeeID) > 2

**Choice1 :**

* CAST
* CONVERT
* REPLACE
* SUBSTRING

**Choice2 :**

* CROSS JOIN
* INNER JOIN
* LEFT OUTER JOIN
* RIGHT OUTER JOIN

**Choice3 :**

* CONTAINS
* HAVING
* UNMIT
* WHERE
* **✅ [Choice1] : CAST** **Explication :** La colonne StartDate doit être convertie en type date. CAST est la fonction la plus simple et directe pour convertir un type Datetime2 en date.
* **✅ [Choice2] : LEFT OUTER JOIN** **Explication :** Pour préserver toutes les lignes de la table Employee et inclure les lignes correspondantes de la table Contract, un LEFT OUTER JOIN est nécessaire.
* **✅ [Choice3] : HAVING** **Explication :** Pour filtrer les résultats après une agrégation (COUNT(DISTINCT EmployeeID)), la clause HAVING est utilisée.

## Question 101

<https://www.examprepper.co/exam/73/21>

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

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

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

What should you use for the data processor and filtering?

**Data processor :**

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

**Filtering :**

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

**Data processor :**

* ✅ An eventstream with an external data source

**Filtering :**

* ✅ An eventstream processor

Le custom endpoint d’Eventstream est utilisé quand on veut envoyer la donnée vers un système externe (exemple : webhook, Kafka, service REST).  
Ici, la destination est un Lakehouse, qui est déjà une destination native d’Eventstream.

KQL est sur un eventhouse , il ne peut agir sur la source ( Azure Event Hub ) ni sur la destination (Lakehouse)

## Question 102

<https://www.examprepper.co/exam/73/21>

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

You need to modify Eventstream1 to include the standard deviation of the temperature.

Which transform operator should you include in the Eventstream1 logic?

A.Expand

B.Group by

C.Union

D.Aggregate

✅ D.Aggregate : Standard deviation ( **Stdev()** – **standard deviation (écart-type)** ) est une fonction stat d’ agrégation comme Max, Min, AVG, …)

## Question 104

<https://www.examprepper.co/exam/73/21>

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

You discover that an EventStream1 transformation fails.

You need to find the following error information:

The error details, including the occurrence time

The total number of errors -

What should you use?

**To find the error details :**

* Data insights
* Data preview
* Details
* Runtime logs

**To find the total number of errors :**

* Data insights
* Data preview
* Details
* Runtime logs

**To find the error details :**

✅ **Runtime logs** → ce sont les **logs d’exécution** où tu vois les erreurs avec le message précis et l’heure d’occurrence.

**To find the total number of errors :**

✅ **Data insights** → tableau de bord avec **métriques globales** sur le flux (taux de messages traités, nombre d’événements en erreur, latence, etc.).

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

Le contenu généré par l’IA peut être incorrect.

## Question 106

<https://www.examprepper.co/exam/73/22>

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

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

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

What should you use?

A.the Microsoft Fabric Capacity Metrics app

B.views from the queryinsights schema

C.Query activity

D.the sys.dm\_exec\_requests dynamic management view

Pas evident .. tout depend vraiment du degré d’analyse souhaité

Mais on suppose que **queryinsights.query** est disponible, DONC

**❌ Query activity (interface UI) :**

* **Ne permet pas de filtrer sur les 60 dernières min (24h , 7d, 30d)**

**✅ queryinsights.query :**

* **Permet de filtrer sur les 60 dernières min moyennant code**

|  |  |  |
| --- | --- | --- |
| Option | Utilité | Limite |
| B. queryinsights | Requête SQL, filtrage dynamique, automatisation | Nécessite SQL |
| C. Query activity | Interface visuelle, facile à lire | Pas de requêtage SQL, pas d’export direct |

Cas B : permet de préciser les 60 dernières minutes

SELECT

    request\_time,

    user\_principal\_name,

    statement\_text

FROM

    queryinsights.query

WHERE

    request\_time >= DATEADD(minute, -60, CURRENT\_TIMESTAMP)

ORDER BY

    request\_time DESC;

Cas A: interface et filtrage seulement 24h, 7d, 30d…

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

Le contenu généré par l’IA peut être incorrect.

## Question 107

<https://www.examprepper.co/exam/73/22>

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

You need to monitor the refresh history of Model1 and visualize the refresh history in a chart.

What should you use?

A.the refresh history from the settings of Model1

B.a notebook

C.a Dataflow Gen2 dataflow

D.a data pipeline

Chelou .. y a pas de ‘Vraie’ réponse .. A , parait normal, mais y a pas de graphique … ce que peut faire un notebook mais qui amène de la complexité : créer un code, avec l’authentification ,l’utilisation d’un API, la création du graphe….

Entre A & B

Apparemment, plutôt ✅ A qui est le choix «NATIF » de monitoring .. ensuite, cela implique un export pour une exploitation en graphe ..

## Question 108

<https://www.examprepper.co/exam/73/22>

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

You need to minimize the load time of DW1.

What should you do?

A.Enable V-Order.

B.Create statistics.

C.Drop statistics.

D.Disable V-Order.

## Question 110 – 65 - 66

<https://www.examprepper.co/exam/73/22>

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

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

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

• Trigger the process when a new file is added.

• Provide the highest throughput.

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

A.KQL queryset

B.Streaming dataset

C.Notebook

D.Dataflow Gen2

## Question 112

[Microsoft - DP-700 - Page 23 | Examprepper](https://www.examprepper.co/exam/73/23)

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

| **VendorID** | **tpep\_pickup\_datetime** | **tpep\_dropoff\_datetime** | **passenger\_count** | **trip\_distance** | **PULocationID** | **DOLocationID** | **payment\_type** | **total\_amount** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | 2022-06-06T11:08:32Z | 2022-06-06T11:22:17Z | 1 | 0.17 | 231 | 50 | 2 | 7.12 |
| 2 | 2022-06-06T11:12:05Z | 2022-06-06T11:20:43Z | 1 | 1.02 | 161 | 163 | 1 | 10.56 |
| 1 | 2022-06-06T11:15:00Z | 2022-06-06T11:25:32Z | 1 | 1.07 | 142 | 230 | 2 | 17.12 |
| 2 | 2022-06-06T11:29:54Z | 2022-06-06T11:49:34Z | 2 | 2.07 | 162 | 236 | 2 | 12.01 |
| 1 | 2022-06-06T11:50:50Z | 2022-06-06T12:07:24Z | 2 | 2.65 | 140 | 142 | 1 | 7.89 |

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

One of the queries must partition RunningTotalAmount by VendorID.

The other query must create a column named FirstPickupDateTime that shows the first value of each hour from tpep\_pickup\_datetime partitioned by payment\_type.

How should you complete each query?.

**Statement1:**

TaxiData

| sort by VendorID asc

| extend RunningTotalAmount = [Choice 1](total\_amount, VendorID == prev(VendorID))

**Choice 1**

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

**Statement2:**

TaxiData

| sort by tpep\_pickup\_datetime asc, payment\_type asc

| extend FirstPickupDateTime = [Choice 2] (tpep\_pickup\_datetime, 1h, 0m, payment\_type == prev(payment\_type))

**Choice 2**

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

Objectif :

*Partitionner RunningTotalAmount par VendorID*  
Cela signifie : calcul cumulatif (running total) par groupe (VendorID).  
En KQL, la fonction appropriée pour un cumul progressif est **row\_cumsum()**.

✅ **Choice 1 = row\_cumsum**

Objectif :

*Créer une colonne FirstPickupDateTime qui montre la première valeur de chaque heure (1h) partitionnée par payment\_type*

En KQL, pour récupérer la **première valeur** d’un champ dans un intervalle (bucket), on utilise row\_rank\_min() appliqué à la date, ou on filtre après un classeme

✅ **Choice 2 =Row\_rank\_min**

## Question 113

<https://www.examprepper.co/exam/73/23>

You are processing streaming data from an external data provider.

You have the following code segment.

datatable (Location:string, Company:string, UnitsSold:long)

[

"New York", "Contoso", 300,

"New York", "Litware", 1000,

"New York", "Relecloud", 300,

"New York", "Fabrikam", 200,

"Seattle", "Contoso", 300,

"Seattle", "Litware", 100,

"Seattle", "Fabrikam", 100,

"San Francisco", "Relecloud", 500,

"San Francisco", "Litware", 500,

"Washington DC", "Litware", 300,

"Washington DC", "Contoso", 400

]

| sort by Location desc, UnitsSold desc

| extend Rank=row\_rank\_dense(UnitsSold, prev(Location) != Location)

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

|  |  |
| --- | --- |
| **Statements** | **YES/NO ?** |
| Litware from New York will be displayed at the top of the result set. | YES/NO |
| Fabrikam in Seattle will have value = 2 in the Rank column. | YES/NO |
| Litware in San Francisco will have the same value in the Rank column as Litware in New York. | YES/NO |

| **Affirmation** | **Réponse** | **Justification** |
| --- | --- | --- |
| **Litware from New York will be displayed at the top of the result set.** | **❌ NO** | **Elle n’est pas en haut car le tri est par Location DESC, donc Washington DC apparaît avant New York.** |
| **Fabrikam in Seattle will have value = 2 in the Rank column.** | **✅ YES** | **Elle partage le même UnitsSold (100) que Litware Seattle, donc rang dense = 2.** |
| **Litware in San Francisco will have the same value in the Rank column as Litware in New York.** | **✅ YES** | **Les deux ont le rang 1 dans leur groupe respectif (UnitsSold le plus élevé dans chaque Location).** |

**CAR .. Voici le résultat :**

* **Washington DC: valeurs présentes : 400, 300**
  1. **Contoso, 400: Rang 1**
  2. **Litware, 300: Rang 2**
* **Seattle: valeurs présentes : 300,100**
  1. **Contoso, 300: Rang 1**
  2. **Litware, 100: Rang 2**
  3. **Fabrikam, 100: Rang 2**
* **San Francisco: valeurs présentes : 500**
  1. **Relecloud, 500: Rang 1**
  2. **Litware, 500: Rang 1**
* **New York: valeurs présentes : 1000,300,200**
  1. **Litware, 1000: Rang 1**
  2. **Contoso, 300: Rang 2**
  3. **Relecloud, 300: Rang 2**
  4. **Fabrikam, 200: Rang 3**

## Question 114

[**Microsoft - DP-700 - Page 23 | Examprepper**](https://www.examprepper.co/exam/73/23)

You have a Fabric workspace that contains a lakehouse named Lakehouse1. Lakehouse1 contains a Delta table named Table1.

You analyze Table1 and discover that Table1 contains 2,000 Parquet files of 1 MB each.

You need to minimize how long it takes to query Table1.

What should you do?

A.Disable V-Order and run the OPTIMIZE command.

B.Disable V-Order and run the VACUUM command.

C.Run the OPTIMIZE and VACUUM commands.

**🟩 C. Run the OPTIMIZE and VACUUM commands**

**OPTIMIZE**

* Combine les petits fichiers en fichiers plus gros.
* Réduit le nombre de fichiers à lire, ce qui accélère les requêtes.
* Peut aussi appliquer V-Order si activé (mais ce n’est pas obligatoire ici).

**VACUUM**

* Supprime les fichiers obsolètes (non référencés dans le Delta log).
* Réduit le volume total de stockage et améliore la gestion du cache.

## Question 115

[Microsoft - DP-700 - Page 23 | Examprepper](https://www.examprepper.co/exam/73/23)

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

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

You need to monitor Warehouse1 to identify the names of users that are actively running queries.

Which view should you use?

A.sys.dm\_exec\_connections

B.sys.dm\_exec\_requests

C.queryinsights.long\_running\_queries

D.queryinsights.frequently\_run\_queries

E.sys.dm\_exec\_sessions

**E. sys.dm\_exec\_sessions** ✅

* Shows all active sessions, including **login\_name** (user) and session status.
* Can be filtered by status = 'running' to find who is actively running queries **right now**.
* Does not require joins just to get the user name.

The view sys.dm\_exec\_sessions:

* Returns **one row per authenticated session**
* Includes columns like:
  + login\_name → the **user name**
  + status → whether the session is **running**, **sleeping**, etc.
  + program\_name, host\_name, client\_interface\_name → useful for identifying the **source of the query**

**A RETENIR**

* **DM = en cours,   vues Dynamic Management Views (DMVs).**
* **QueryInsights = historique  (30 jours)**
* **Travaillent toutes les 2 sur le WS uniquement**

**Correct answer:** **E. sys.dm\_exec\_sessions**  
Because it directly lists **currently active sessions with user names**, which is exactly what you need for monitoring who is running queries in Warehouse1 in real time.

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

Le contenu généré par l’IA peut être incorrect.

## Question 116

<https://www.examprepper.co/exam/73/24>

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table in a lakehouse.

You need to remove files that are older than seven days and are no longer in use.

Which command should you run?

A.VACUUM

B.COMPUTE

C.OPTIMIZE

D.CLONE

Pas de piège :

✅ A.VACUUM

## Question 117

<https://www.examprepper.co/exam/73/24>

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

You discover that Pipeline1 keeps failing.

You need to identify which SQL query was executed when the pipeline failed.

What should you do?

A.From Monitoring hub, select the latest failed run of Pipeline1, and then view the output JSON.

B.From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.

C.From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemReadFailed.

D.From Real-time hub, select Fabric events, and then review the details of Microsoft. Fabric.ItemUpdateFailed.

## Question 117

[Microsoft - DP-700 - Page 24 | Examprepper](https://www.examprepper.co/exam/73/24)

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

You discover that Pipeline1 keeps failing.

You need to identify which SQL query was executed when the pipeline failed.

What should you do?

A.From Monitoring hub, select the latest failed run of Pipeline1, and then view the output JSON.

B.From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.

C.From Real-time hub, select Fabric events, and then review the details of Microsoft.Fabric.ItemReadFailed.

D.From Real-time hub, select Fabric events, and then review the details of Microsoft. Fabric.ItemUpdateFailed.

✅ B.From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.

Reproduit . le code sql est dans le input

## Question 118

[Microsoft - DP-700 - Page 24 | Examprepper](https://www.examprepper.co/exam/73/24)

You have a Fabric notebook named Notebook1 that has been executing successfully for the last week.

During the last run, Notebook1executed nine jobs.

You need to view the jobs in a timeline chart.

What should you use?

A.Real-Time hub

B.Monitoring hub

C.the job history from the application run

D.Spark History Server

E.the run series from the details of the application run

✅ E.the run series from the details of the application run

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

Le contenu généré par l’IA peut être incorrect.

# SYNTHESE GENERIQUES

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **REF** | **Thème** | **Question (résumé)** | **Réponse correcte** | **Justification courte** | **MEMO ou Piège à éviter** |
| 5 | 49 | PPL RUN | screenshot | D.Copy\_kdi will run and Execute procedure1 will be skipped. |  | Inactive = definitif |
| 9 |  | KQL Take’ operator |  | **B. Retrieves a specified number of rows from a table** |  |  |
|  |  | primary use of the Copy Data activity in a data pipeline |  | **B. To ingest data from an external source into a lakehouse file or table** |  |  |
|  |  | keyword should you use to retrieve only specific columns in a KQL |  | **C.Project** |  |  |
|  |  | PPL param |  | **A."@pipeline().parameters.param1"** |  | **La phrase "returns param1 as an int value" signifie que l’expression doit produire une valeur entière à partir de ce paramètre, pas que le pipeline retourne param1.** |
|  | 77 | implement a star schema for DW1 (product) |  | ✅ **ProductCategory, ProductSubcategory, Product** → **3. Denormalized into a single product dimension table**  ✅**Joining key** → **2. The unique system generated identifier** |  |  |
|  | 79 | to process the following three datasets by using Fabric |  | Dataset1 T-SQL statement  Dataset2 Notebook  Dataset3 T-SQL statement |  |  |
|  | 80 | implement an incremental loading pattern by using Notebook1 |  | ✅ Choice 1 .whenMatchedUpdate(  ✅ Choice 2 .whenNotMatchedInsert(  ✅ Choice 3 .whenNotMatchedBySourceUpdate( |  |  |
|  | 117 | SQL query was executed when the pipeline failed. |  | ✅ B.From Monitoring hub, select the latest failed run of Pipeline1, and then view the input JSON.  Reproduit . |  | le code sql est dans le input |

# SYNTHESE CONTOSO

| **#** | **REF** | **Thème** | **Question (résumé)** | **Réponse correcte** | **Justification courte** | **MEMO ou Piège à éviter** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | Accès gold layer (détail) | Data Analysts access gold LH | **C. Share lakehouse + Read all SQL Endpoint data** | Lecture via SQL Endpoint, pas accès fichiers | “Viewer” ou “Spark” donnerait accès bronze/silver  “Build reports” ne donne pas accès SQL Endpoint |
| 2 | 50 | WS A config Source control |  | **A : From TENANT Git Repo**  **D : Assigne WS A to Cap1** |  |  |
| 3 | 83 | POS1 – population |  | **Bronze : 3.A PPL copy activity**  **Silver : 2.A notebook** |  |  |
| 4 | 90 | MAR1 – issue | Relancer ingestion si erreur transitoire REST API MAR1 | **B.Configure retries for the Copy data activity.** |  | Éviter code perso pour retry |
| 5 | 91 | Handle Old Files |  | **C.a notebook that runs the VACUUM command** |  |  |
| 6 | 94 | usage of the data in the Amazon S3 bucket |  | **B. Create a shortcut and ensure that caching is disabled for the workspace** |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| 7 | 99 | MAR1 – Population |  | **A. ForEach**  **B. Copy data** |  |  |
| 8 | 98 | Product Dim |  | **B. INNER JOIN**  **B. INNER JOIN**  **C. IsActive = 1** |  |  |
| 9 | 109 | Notif any step failed |  | **3. Une condition On failure dependency**  **3. Une activité Office365Outlook** |  |  |
| 10 | 111 | Médalion Population |  | **A.Schedule a data pipeline that calls other data pipelines.** |  | Ne pas choisir plusieurs pipelines indépendants (risque désynchronisation) |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

# SYNTHESE LITWARE

| **#** | **REF** | **Thème** | **Question (résumé)** | **Réponse correcte** | **Justification courte** | **MEMO ou Piège à éviter** |
| --- | --- | --- | --- | --- | --- | --- |
|  | 31 | the bronze and silver layers run in isolation |  | **B. Create a custom pool** |  |  |
|  | 32 | authors can see only their respective sales data |  | ✅ **Choice 1** → **G. SCHEMABINDING**  ✅ **Choice 2** → **H. USER\_NAME()**  ✅ **Choice 3** → **A. AuthorSales** |  | ✅ **Choice 1** → **G. SCHEMABINDING**   * Obligatoire pour une fonction table utilisée comme prédicat de sécurité RLS.   ✅ **Choice 2** → **H. USER\_NAME()**   * Permet de comparer l’auteur courant à la colonne d’e-mail stockée dans la table.   ✅ **Choice 3** → **A. AuthorSales**   * C’est la table sur laquelle la politique RLS s’applique. |
|  | 61 | You need to implement the solution for the book reviews. |  | ✅ **B. Create a shortcut** |  |  |
|  | 72 | resolve the sales Team data issue. |  | ✅ **E. Configure incremental refresh for the dataflow. Set Refresh rows from the past to 1 Month.** |  | Ca fait toujours le job chargement DELTA mais ne conserve dans le modele que 1 mois |
|  | 75 | create a workflow for the new book cover images. |  | **D. a data pipeline** **F. a reflex item** |  |  |
|  | 76 | ingest the SEO data |  | **D.An eventstream** |  |  |
|  | 97 | create a workflow for the new book cover images. |  | **A. an Activator Item** **F. a reflex item** |  |  |
|  | 103 | ad-hoc query issue |  | **[Choice 1] → queryinsights.exec\_requests\_history**  **[Choice 2] → number\_of\_failed\_runs > 1** |  |  |

# SCD TYPE

**🧩 SCD Type 0 – Fixed Dimension**

* **Aucune mise à jour** des attributs.
* Les valeurs restent **figées** dans le temps.
* Utilisé pour des données qui ne changent jamais (ex : date de naissance).

**🔁 SCD Type 1 – Overwrite**

* Les **valeurs sont écrasées** lors d’une mise à jour.
* **Pas d’historique** conservé.
* Simple à implémenter, mais ne permet pas d’analyse historique.

**🧠 SCD Type 2 – Add New Row**

* Une **nouvelle ligne est ajoutée** à chaque changement d’attribut.
* Chaque version est **datée** avec des colonnes comme valid\_from et valid\_to.
* Permet une **analyse historique complète**.
* Nécessite une gestion des clés techniques (surrogate key).

**🧮 SCD Type 3 – Add New Column**

* Une **nouvelle colonne** est ajoutée pour stocker l’ancienne valeur.
* Permet de conserver **une version antérieure**, mais **pas plusieurs**.
* Utile pour des comparaisons entre l’état actuel et précédent.

# EXO Eventstream

INGESTION LH

|  |
| --- |
| # Welcome to your new notebook  # Type here in the cell editor to add code!  from pyspark.sql.types import StructType, StructField, StringType, IntegerType  # Schéma des données JSON  schema = StructType([      StructField("id", IntegerType(), True),      StructField("name", StringType(), True),      StructField("value", IntegerType(), True)  ])  # Chemin d'entrée dans OneLake  input\_path = "abfss://975b49e1-0a26-4dad-9061-2b5091bc3ffe@onelake.dfs.fabric.microsoft.com/1ce5fb78-ac21-471b-9f71-aeda68ea1313/Files/stream\_input/"  # Lecture en streaming  df\_stream = spark.readStream \      .schema(schema) \      .option("maxFilesPerTrigger", 1) \      .json(input\_path)  # Transformation  df\_transformed = df\_stream.withColumn("value\_squared", df\_stream["value"] \*\* 2)  # Chemin complet ABFSS pour l’écriture  output\_path\_abfss = "abfss://975b49e1-0a26-4dad-9061-2b5091bc3ffe@onelake.dfs.fabric.microsoft.com/1ce5fb78-ac21-471b-9f71-aeda68ea1313/Tables/stream\_output"  # Écriture en streaming  query = df\_transformed.writeStream \      .format("delta") \      .outputMode("append") \      .option("checkpointLocation", output\_path\_abfss + "/\_checkpoint") \      .start(output\_path\_abfss)  # Attendre quelques secondes que le stream démarre  import time  time.sleep(5)  # Enregistrement de la table dans Fabric (chemin relatif)  spark.sql("""      CREATE TABLE IF NOT EXISTS logical\_stream\_output      USING DELTA      LOCATION 'Tables/stream\_output'  """)  # Lecture et affichage  df = spark.read.format("delta").load("Tables/stream\_output")  display(df) |

|  |
| --- |
|  |

|  |
| --- |
| spark.catalog.listTables() |

|  |
| --- |
|  |

Une image contenant texte, logiciel, nombre, Police

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, logiciel, Icône d’ordinateur, Page web

Le contenu généré par l’IA peut être incorrect.

KQL queryset

['TBL\_EVH']

| where passenger\_count == 4

| order by passenger\_count, trip\_distance desc

| take 5

| project tpep\_pickup\_datetime, tpep\_dropoff\_datetime, passenger\_count, trip\_distance, total\_amount, fare\_amount

Une image contenant texte, nombre, logiciel, ligne

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, capture d’écran, logiciel, Logiciel multimédia

Le contenu généré par l’IA peut être incorrect.

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

Le contenu généré par l’IA peut être incorrect.

ONELAKE

Une image contenant texte, capture d’écran, logiciel, Icône d’ordinateur

Le contenu généré par l’IA peut être incorrect.

Une image contenant texte, capture d’écran, logiciel, Icône d’ordinateur

Le contenu généré par l’IA peut être incorrect.

# EXO Apache Spark Structured Streaming ASSS

Contexte actuel :   
A) j'ai un Lakehouse (LH), "DLKH\_LABO" ( j'utilise l'id technique : 975b49e1-0a26-4dad-9061-2b5091bc3ffe.

B) dans un Workspace LABO-WS.   
C) La section Files de mon LH contient un dossier "stream\_input" et ce dossier un Fichier source "stream\_input\_sample.json

EXO : principe du ASSS, Dès qu’un nouveau fichier source entre, je peux faire en sorte de mettre à jour une table DELTA dans la section Delta de mon Lakehouse.

Divers fichier en entrée avec un stucture telle que :

{"id": 1, "name": "Sensor0A", "value": 11}

{"id": 2, "name": "Sensor0B", "value": 13}

{"id": 3, "name": "Sensor0C", "value": 14}

{"id": 4, "name": "Sensor0F", "value": 20}

Le code fera un produit au carré de la valeur

Code :

# Welcome to your new notebook

# Type here in the cell editor to add code!

from pyspark.sql.types import StructType, StructField, StringType, IntegerType

# Schéma des données JSON

schema = StructType([

    StructField("id", IntegerType(), True),

    StructField("name", StringType(), True),

    StructField("value", IntegerType(), True)

])

# Chemin d'entrée dans OneLake

input\_path = "abfss://975b49e1-0a26-4dad-9061-2b5091bc3ffe@onelake.dfs.fabric.microsoft.com/1ce5fb78-ac21-471b-9f71-aeda68ea1313/Files/stream\_input/"

# Lecture en streaming

df\_stream = spark.readStream \

    .schema(schema) \

    .option("maxFilesPerTrigger", 1) \

    .json(input\_path)

# Transformation

df\_transformed = df\_stream.withColumn("value\_squared", df\_stream["value"] \*\* 2)

# Version Transformation avec filtrage  (exclure < 10 )

#df\_filtered = df\_stream.filter(df\_stream["value"] >= 10)

#df\_transformed = df\_filtered.withColumn("value\_squared", df\_filtered["value"] \*\* 2)

# Chemin complet ABFSS pour l’écriture

output\_path\_abfss = "abfss://975b49e1-0a26-4dad-9061-2b5091bc3ffe@onelake.dfs.fabric.microsoft.com/1ce5fb78-ac21-471b-9f71-aeda68ea1313/Tables/stream\_output"

# Écriture en streaming

query = df\_transformed.writeStream \

    .format("delta") \

    .outputMode("append") \

    .option("checkpointLocation", output\_path\_abfss + "/\_checkpoint") \

    .start(output\_path\_abfss)

# Attendre quelques secondes que le stream démarre

import time

time.sleep(5)

# Enregistrement de la table dans Fabric (chemin relatif)

spark.sql("""

    CREATE TABLE IF NOT EXISTS logical\_stream\_output

    USING DELTA

    LOCATION 'Tables/stream\_output'

""")

# Lecture et affichage

df = spark.read.format("delta").load("Tables/stream\_output")

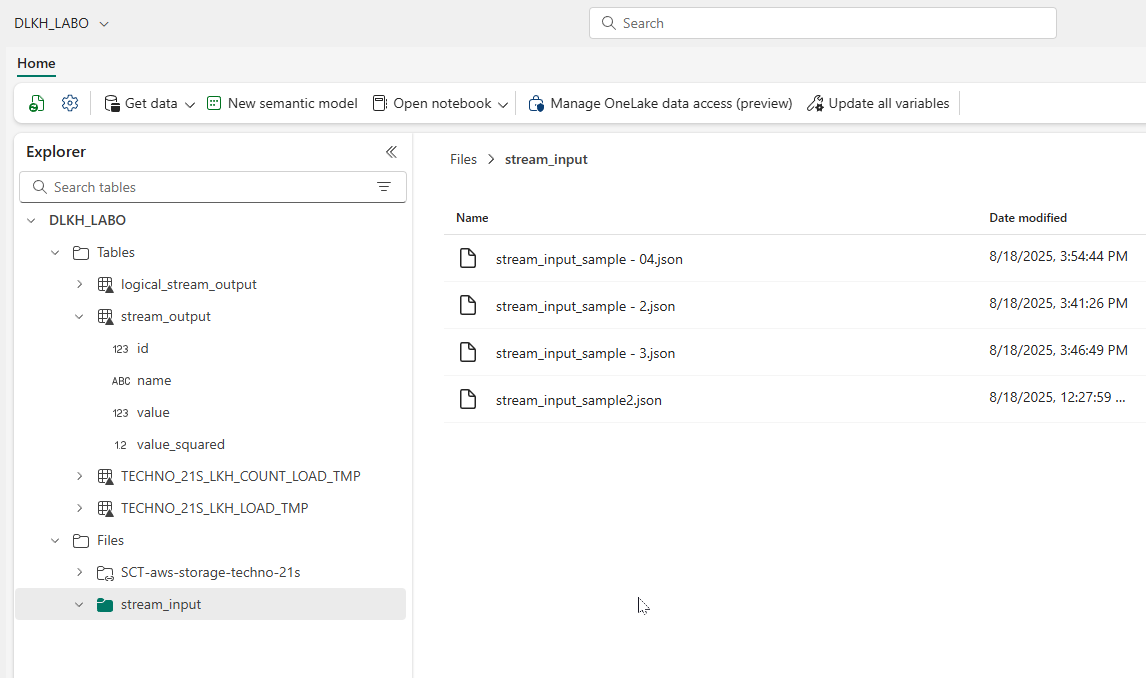
display(df)

spark.catalog.listTables()

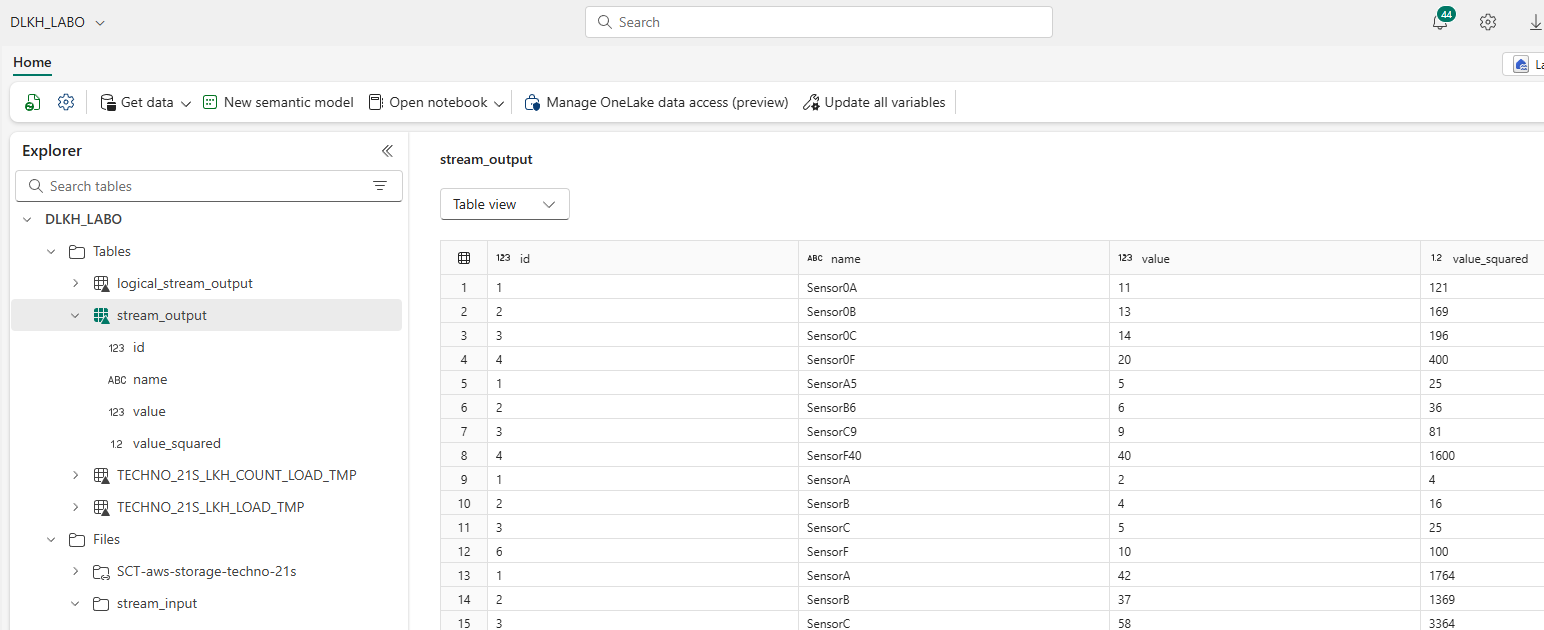
en fait, c'est spark.readStream qui est la fonction master du streaming ASSS , detectant un nouveau fichier

puis writestream et start écrivant dans le dossier output de Table

depot des fichier dans le input :



Ecriture dans le output ( Table Delta )



Résumé Exo

**📘 Exercice : Traitement en streaming de fichiers JSON vers une table Delta dans Fabric**

**🎯 Objectif**

Lire en continu des fichiers JSON déposés dans un dossier Files/stream\_input d’un Lakehouse Fabric, transformer les données, et les écrire dans une table Delta Tables/stream\_output, avec création d’une table logique SQL.

**🧱 Contexte technique**

* **Workspace ID** : 975b49e1-0a26-4dad-9061-2b5091bc3ffe
* **Lakehouse ID** : 1ce5fb78-ac21-471b-9f71-aeda68ea1313
* **Lakehouse** : DLKH\_LABO
* **Source** : Fichiers JSON dans Files/stream\_input
* **Destination** : Table Delta dans Tables/stream\_output
* **Table logique SQL** : logical\_stream\_output

**🧪 Contenu du fichier JSON (exemple)**

**🧠 Code Spark complet**

**📋 Création de la table logique (à exécuter dans un notebook Spark attaché au Lakehouse)**

**👀 Lecture des données**

**🔁 Pour que le flux fonctionne en continu :**

* Le stream doit rester actif (query.isActive == True).
* Chaque nouveau fichier JSON ajouté dans stream\_input sera automatiquement traité.
* Si le notebook est fermé ou redémarré, il faut relancer la cellule writeStream.

**🛠️ Solutions pour garder le stream actif durablement**

**🔁 Option 1 : Garder le notebook ouvert**

* Laisse le notebook Spark ouvert dans Fabric.
* Ne redémarre pas la session.
* Tu peux ajouter une cellule de monitoring :

Code Python :

A placer dans une cellule dédiée du notebook spark, juste après celle qu idemarre le streaming

while query.isActive:

    print("Stream is running...")

    time.sleep(10)

**🧪 Option 2 : Automatiser avec un pipeline Fabric**

* Crée un **pipeline Fabric** qui exécute ton notebook à intervalles réguliers (ex : toutes les 5 minutes).
* Cela relance le stream si nécessaire.
* Tu peux ajouter une logique pour vérifier si le stream est déjà actif ou non.

**🧱 Option 3 : Déploiement dans un environnement Spark dédié**

* Si tu veux une solution industrielle, tu peux déployer ton code dans un cluster Spark (hors Fabric) avec un job Spark permanent.

Exemple avec notebook monitoring :

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

Le contenu généré par l’IA peut être incorrect.

# ANNEXES

## Annexe Question 84

[Microsoft - DP-700 - Page 17 | Examprepper](https://www.examprepper.co/exam/73/17)

You have a Fabric workspace that contains an eventstream named EventStream1. EventStream1 outputs events to a table named Table1 in a lakehouse. The streaming data is sourced from motorway sensors and represents the speed of cars. You need to add a transformation to EventStream1 to average the car speeds. The speeds must be grouped by non-overlapping and contiguous time intervals of one minute. Each event must belong to exactly one window. Which windowing function should you use?

A. sliding

B. hopping

C. tumbling

D. session

**✅ Tumbling :**

* Tu veux **des fenêtres de 1 minute fixes et continues**
* Tu veux **zéro chevauchement**
* Tu veux **chaque événement assigné à une seule fenêtre**

VIDEO SUR WINDOW FUNCTION

[Préparation à l'examen DP-700 : Fonctions de fenêtrage Eventstream | Ingénieur de données Microso...](https://www.youtube.com/watch?v=xwLF0YLMbqY)

|  |
| --- |
| Une image contenant texte, capture d’écran, circuit, Appareils électroniques  Le contenu généré par l’IA peut être incorrect. |

|  |
| --- |
| Une image contenant texte, capture d’écran, circuit, Appareils électroniques  Le contenu généré par l’IA peut être incorrect. |

|  |
| --- |
| Une image contenant texte, capture d’écran, circuit, Logiciel multimédia  Le contenu généré par l’IA peut être incorrect. |

|  |
| --- |
| Une image contenant texte, capture d’écran, circuit  Le contenu généré par l’IA peut être incorrect. |

|  |
| --- |
| Une image contenant texte, capture d’écran, circuit, Appareils électroniques  Le contenu généré par l’IA peut être incorrect. |

Différences :

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

Le contenu généré par l’IA peut être incorrect.