

QUIZ 1 — LAB 3.1

Pipeline robuste Taxi

1 Quelle est la différence fondamentale entre une erreur de transformation et une ligne invalide métier ?

2 Pourquoi est-il dangereux de laisser passer des `total_amount < 0` dans un pipeline de production ?

3 Explique la différence entre :

- un rejet technique
- un rejet métier

Donne un exemple concret pour chacun.

4 Pourquoi un pipeline doit-il être rejouable (idempotent) ?

Donne un scénario réel où cela devient critique.

5 Que se passe-t-il si le schéma du CSV change (nouvelle colonne, colonne supprimée) ?

Comment anticiper cela dans Hop ?

6 Pourquoi est-il préférable de séparer les flux "clean" et "rejected" plutôt que de simplement supprimer les lignes invalides ?

7 Quelle transformation Hop permet de filtrer les lignes selon une condition logique ?

8 Dans une architecture professionnelle, qui définit les règles de validation :

- le data engineer
 - le métier
 - les deux ?
Explique.
-

9 Pourquoi ajouter un `processing_timestamp` peut être stratégique ?

10 Si ton pipeline traite 10 millions de lignes, quelle étape risque d'être la plus coûteuse en performance ?