Stripe : Conception d'une architecture de données unifiée

Unification des systèmes OLTP, OLAP et NoSQL

Kévin Chatelain
Architecte en Intelligence Artificielle
Bloc 2 – « Concevoir et déployer des architecture de données (pour l'IA) »
Jedha 2023/2024

Gestion des données intégrées

Présentation du cadre de l’entreprise

Contexte

Stripe repose sur la gestion de données massives pour offrir des services de paiement fluides et fiables.

Cependant, l’expansion mondiale de Stripe expose l’entreprise à des défis en matière de **scalabilité**, de **performance** et de **conformité réglementaire**.

Défis

- ✔ Concevoir une architecture unifiée et performante.
- ✔ Garantir l'intégrité des transactions et des analyses.
- ✔ Exploiter les données non structurées efficacement.
- ✔ Intégrer des pipelines fiables et synchronisés.
- ✔ Assurer la conformité avec le GDPR et PCI-DSS.

Problématique

Comment optimiser la gestion et l’intégration des données ?

Section 1: Gestion des données intégrées

Présentation des architectures

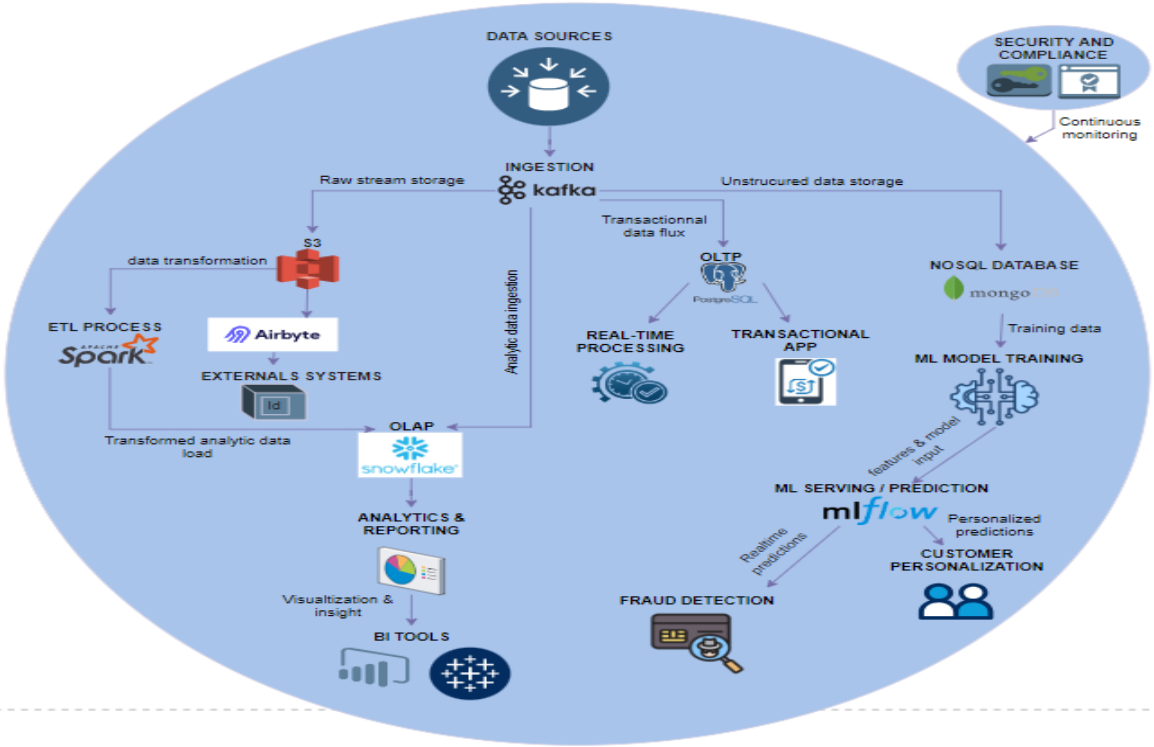
Gestion des données intégrées

Présentation des architectures

- Point 1 : Architecture de la gestion des données
.
- Point 2 : Architecture de l’OLAP
.
- Point 3 : Architecture de l’OLTP
.
- Point 4 : Architecture du NoSQL

Architecture de la gestion des données

Efficacité et conformité du flux d'information



Plan de sécurité et de conformité

Mesures pour une gestion sécurisée des données

1. **Chiffrement des données**
   Protection des données en transit et au repos via des protocoles de sécurité avancés (TLS, AES-256).

2. **Contrôle d'accès et authentification**
   Gestion des accès avec un contrôle basé sur les rôles (RBAC) et une authentification multi-facteurs (MFA).

3. **Audit et traçabilité des accès**
   Suivi des activités avec des journaux d'audit et alertes en cas d'accès suspect.

4. **Sécurité des systèmes et services**
   Protection des bases de données et des endpoints avec des configurations sécurisées et des pare-feu (WAF).

5. **Surveillance en temps réel et alertes**
   Mise en place de systèmes SIEM pour le suivi et la détection des menaces en continu.

6. **Gestion des données sensibles et conformité**
   Anonymisation et pseudonymisation des données pour se conformer aux régulations (GDPR, PCI-DSS).

7. **Plan de réponse aux incidents de sécurité**
   Préparation et formations pour une réponse rapide et efficace aux incidents de sécurité.

8. **Tests de sécurité et audits**
   Réalisation de tests de pénétration réguliers et audits de conformité.

Architecture de la gestion des données

Avantages stratégique

- **Centralisation des données** : Regroupement des données de multiples sources pour un accès et une analyse facilités.
- **Traitement en temps réel et différé** : Kafka pour l'ingestion en temps réel, S3 pour le stockage brut, offrant flexibilité et rapidité.
- **Analyses avancées** : OLAP (Snowflake) pour des insights détaillés et une exploration multi-dimensionnelle.
- **Machine Learning intégré** : MongoDB et ML Model Training pour des prédictions et une personnalisation enrichies.

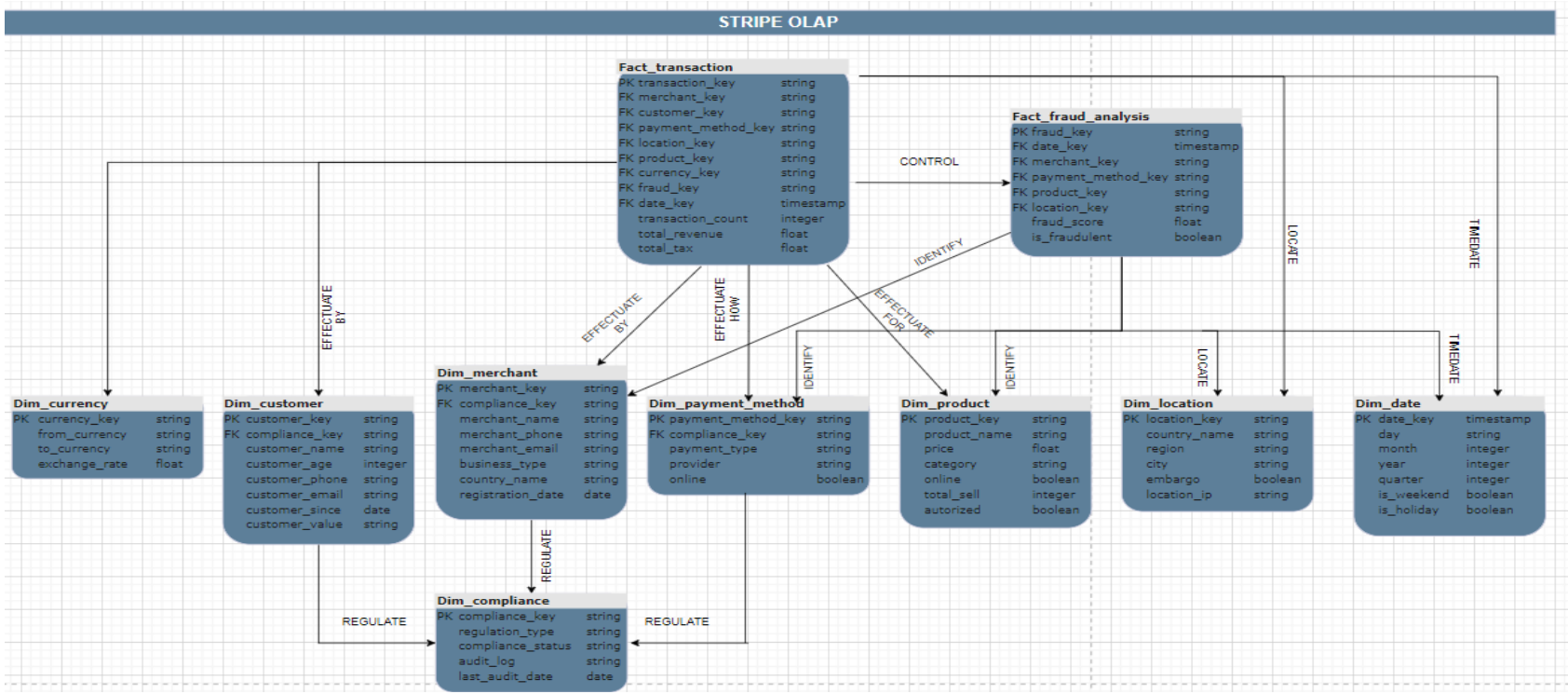
Architecture de la gestion des données

Avantages stratégique (suite)

- **Détection de fraude et personnalisation client** : Analyse en temps réel pour la fraude et des recommandations personnalisées.
- **Sécurité et conformité** : Surveillance continue pour respecter les réglementations à chaque étape.
- **Flexibilité et évolutivité** : Infrastructure scalable (Spark, Kafka, Snowflake) pour s'adapter aux besoins croissants de Stripe.

Architecture de l'OLAP

Analyses performantes et évolutives



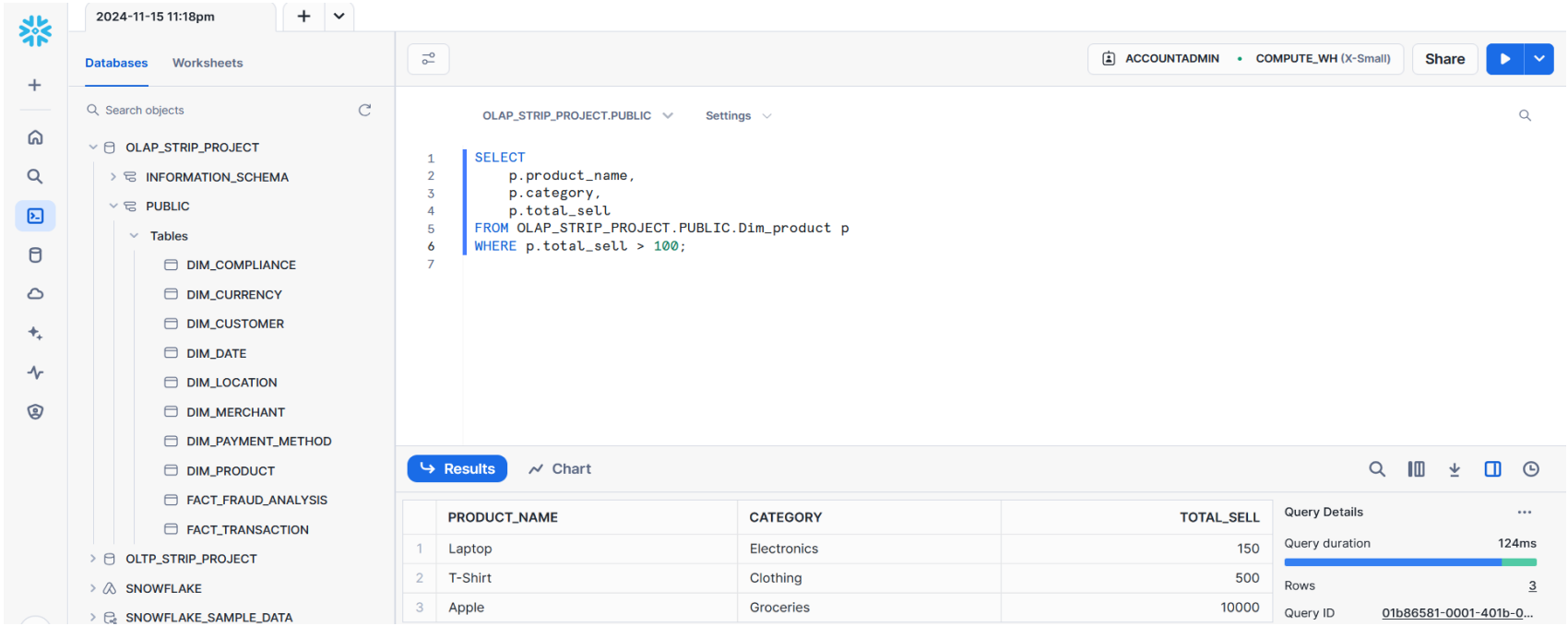
Architecture de l'OLAP

Atouts clés

- **Analyse multi-dimensionnelle efficace** : Structure en tables de faits et dimensions pour des requêtes rapides et des insights détaillés.
- **Prévention de la fraude** : Table de faits dédiée pour détecter rapidement les comportements suspects.
- **Conformité intégrée** : Suivi des régulations et audits facilités grâce à la dimension conformité.
- **Évolutivité** : Architecture flexible permettant d'ajouter de nouvelles dimensions pour s'adapter aux besoins changeants.

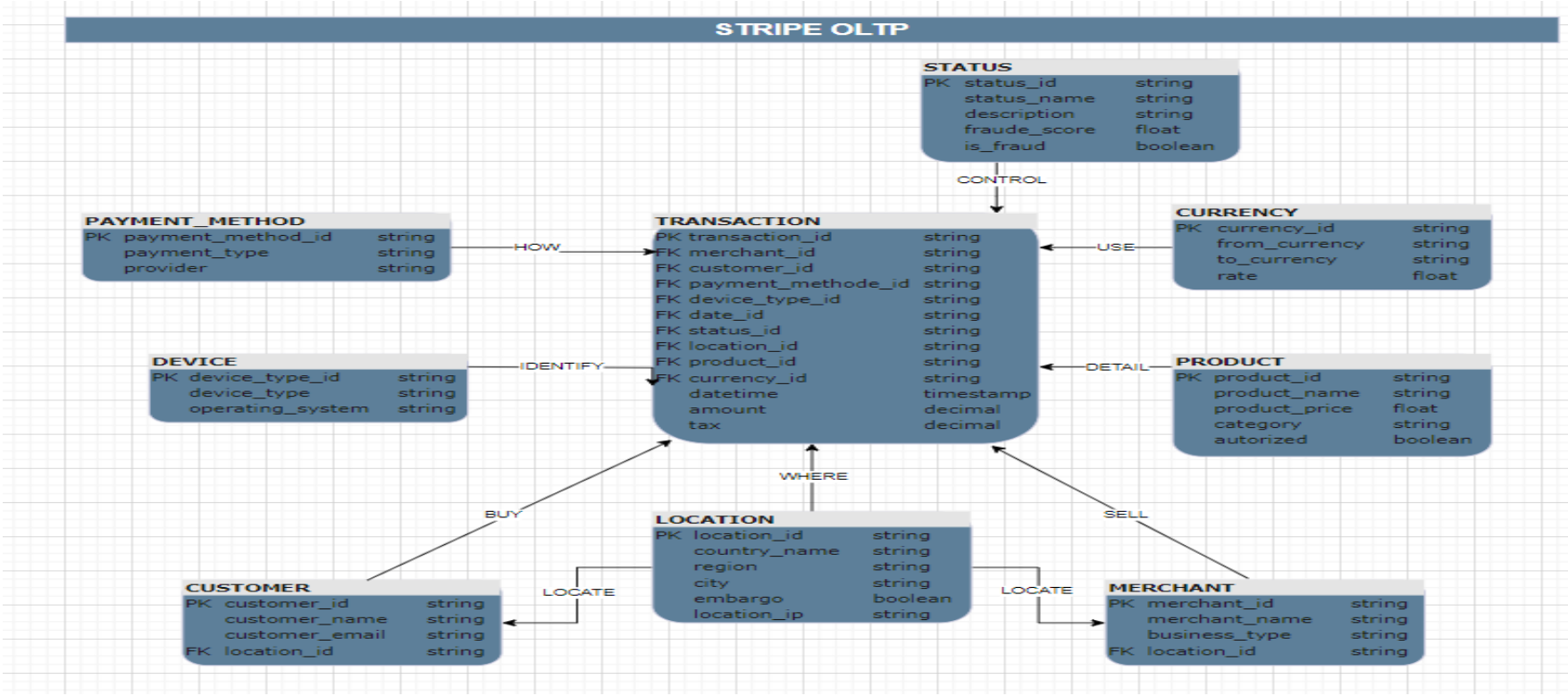
Architecture de l'OLAP

Exemple d'une requête



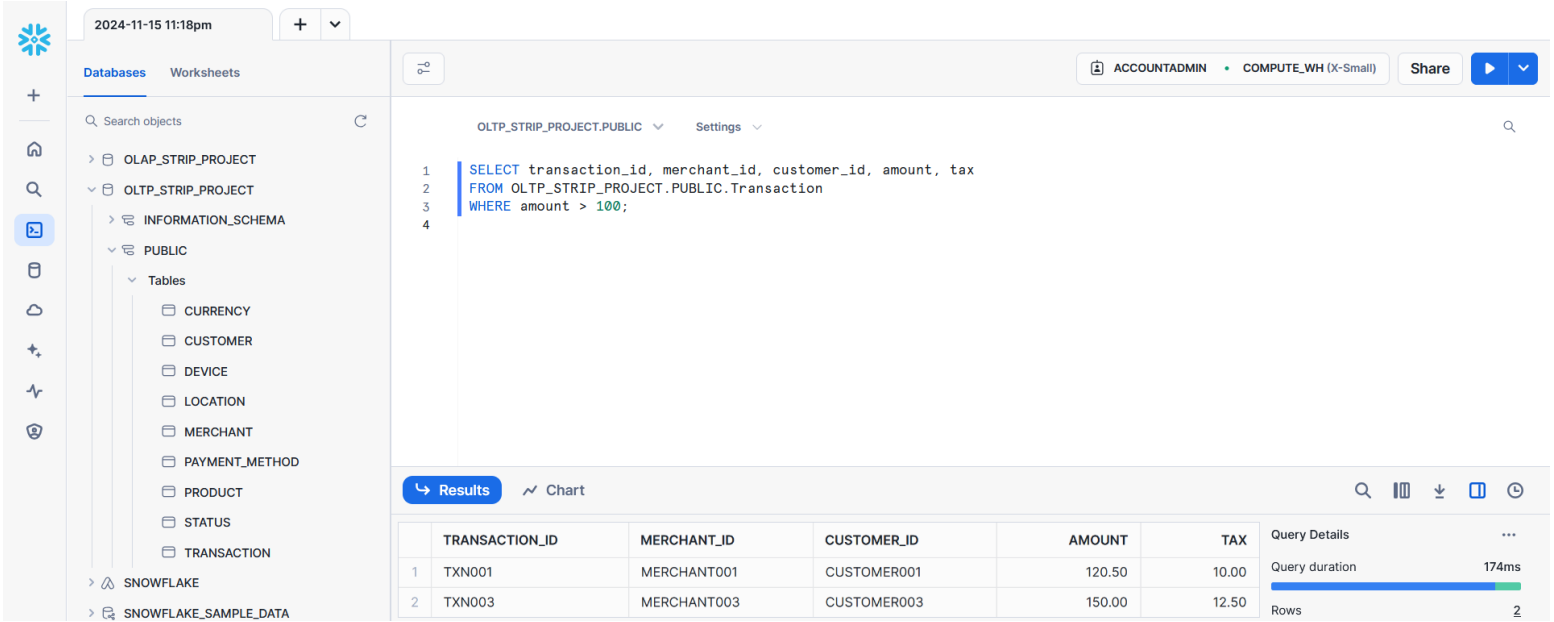
Architecture de l’OLTP

Integrité transactionelle



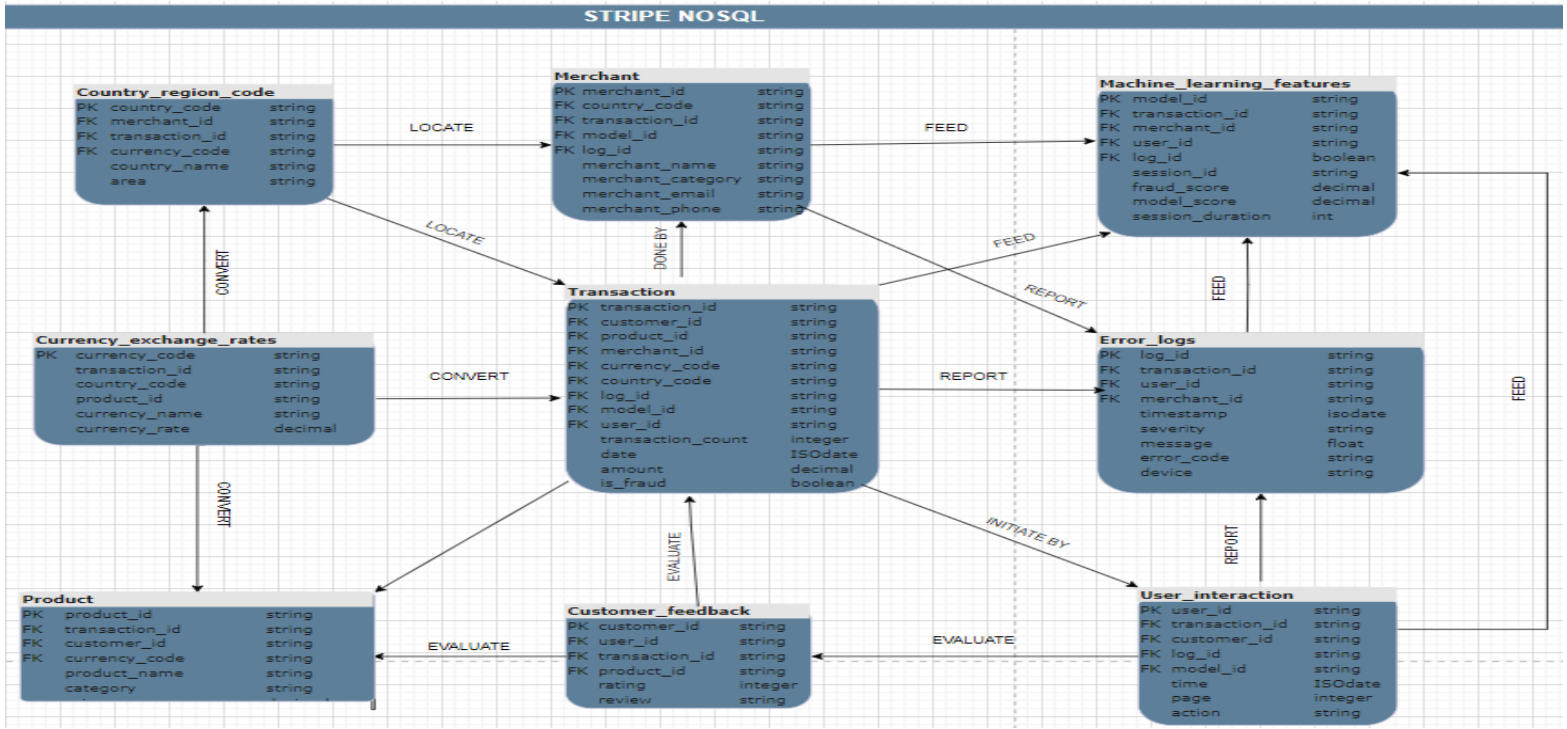
Architecture de l'OLTP

Exemple d'une requête



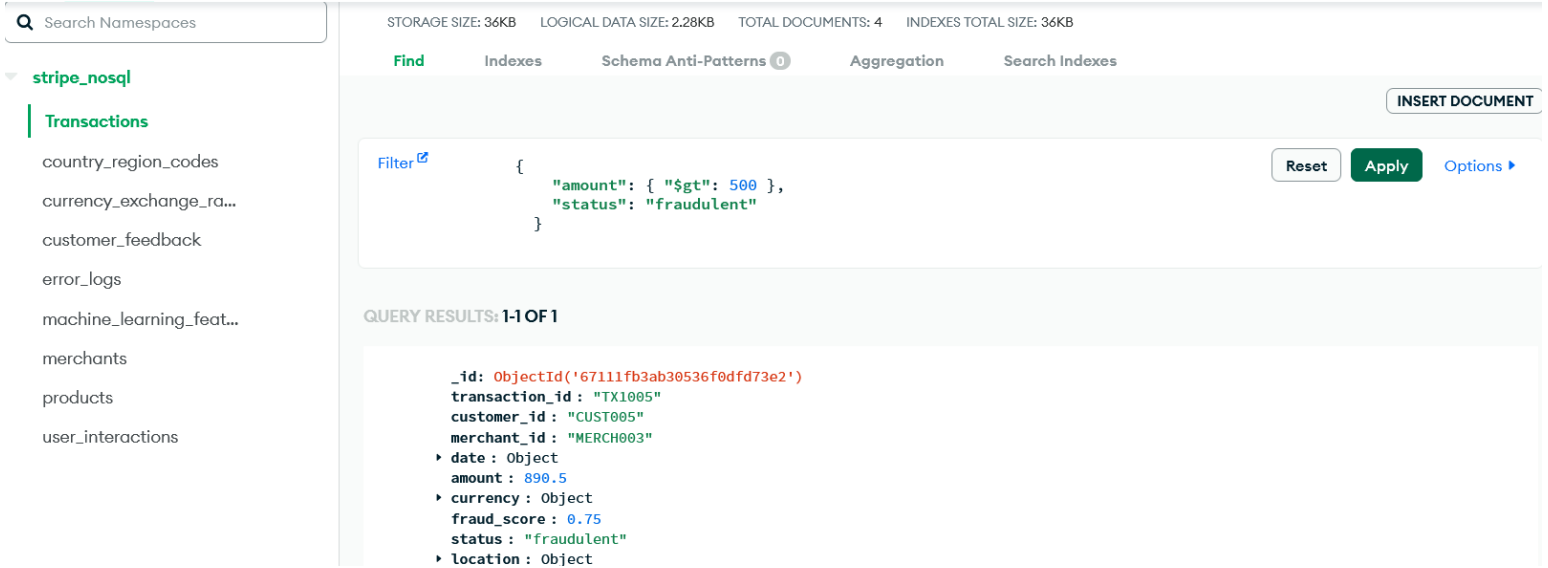
Architecture noSQL

Gestion flexible des données



Architecture de NoSQL

Exemple d'une requête



Section 2: Machine learning & securité

Présentation des stratégies adoptées

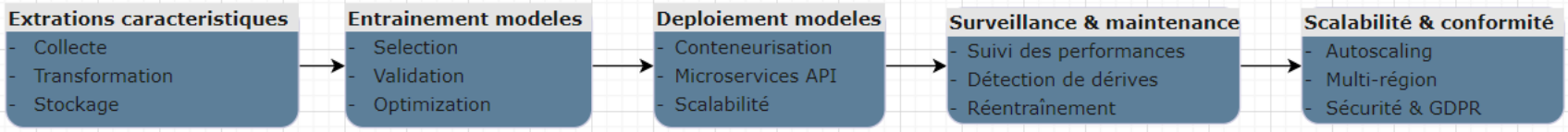
Gestion des données intégrées

Présentation des architectures

- Point 1 : **Stratégie d’intégration Machine Learning**
- Point 2 : **Plan de sécurité et de conformité**

Stratégie d’intégration Machine Learning

De l’extraction au monitoring des modeles



- **Réactivité accrue** : Feature Store centralisé pour un accès rapide aux données.
- **Déploiement optimisé** : Conteneurisation et microservices pour flexibilité et scalabilité.
- **Conformité intégrée** : Respect des normes de sécurité (GDPR, PCI-DSS) à chaque étape.
- **Précision maintenue** : Surveillance continue et réentraînement automatisé pour gérer les dérives.

Merci de votre Attention!

Dans l’attente de vos questions.