

Développement d'une Data Platform (Lakehouse) dans le cloud Azure

Gérald Herrera

1. Objectifs du projet

- Déployer une architecture Lakehouse complète dans le cloud azure.
- Mettre en place un pipeline automatisé et gouverné.
- Explorer des outils d'intelligence artificielle dans Databricks.

3. Technologies utilisées

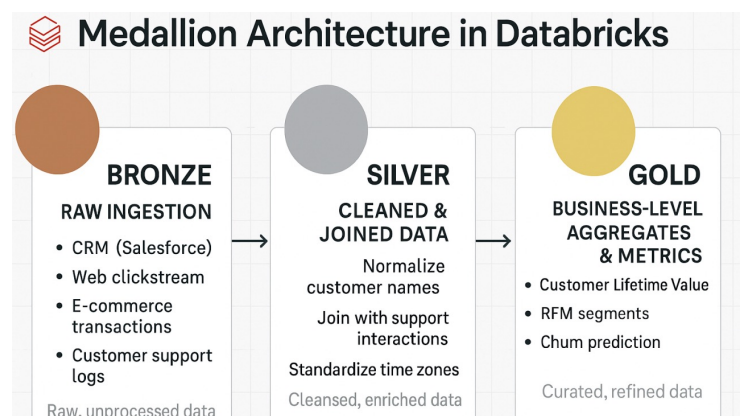
- Azure Databricks
- Azure SQL Database
- Microsoft PowerBI
- ML Flow
- Terraform



2. Contexte

Le data lakehouse est une architecture, ouverte, de gestion des données. Il associe la souplesse, le faible coût et l'échelle des data lakes aux possibilités de gestion et aux transactions ACID des data warehouses.

4. Architecture cible



5. Explorations avancées

- Entraînement d'un modèle de classification supervisée.
- Suivi des expériences avec MLflow.
- Test d'un modèle de langage (LLM).



6. Résultats

- Mise en œuvre d'un pipeline automatisé, traçable et évolutif.
- Production de tableaux de bord Power BI.
- Automatisation du déploiement complet via Terraform.