

# Rapport de Modernisation de l'Infrastructure de Données – Commerce Électronique

## 1. Conception de l'Architecture Azure

Pour répondre aux besoins de l'entreprise, nous proposons une architecture Azure scalable et sécurisée intégrant les services suivants :

### Ingestion de données

- **Azure Data Factory (ADF)** : pour orchestrer les flux d'ingestion depuis les bases SQL Server/PostgreSQL.
- **Azure Event Hubs** : pour collecter les journaux de navigation et clics utilisateurs en temps réel.
- **Azure Logic Apps** : pour automatiser la récupération de données issues des réseaux sociaux via APIs.

### Stockage des données

- **Azure Data Lake Storage Gen2** : stockage centralisé pour données structurées et non structurées.
- **Azure Blob Storage** : pour les objets binaires (images, vidéos, logs).

### Traitement & analyse

- **Azure Stream Analytics** : traitement en temps réel des données d'Event Hubs.
- **Azure Synapse Analytics** : analyse des données en différé (Big Data + Data Warehousing).
- **Azure Databricks** : pour le machine learning (entraînement de modèles de recommandation et détection de fraude).

### Restitution des résultats

- **Power BI** : tableau de bord interactif pour visualiser les ventes, alertes de fraude et recommandations.
- **Azure Machine Learning Service** : déploiement des modèles en production via endpoint API.

### Sécurité et conformité

- **Azure Key Vault** : pour la gestion des secrets et clés de chiffrement.
- **Azure Active Directory (AAD)** : gestion des accès utilisateurs et authentification.
- **Microsoft Defender for Cloud** : surveillance et sécurité proactive.
- **Azure Policy & Blueprints** : conformité RGPD, ISO 27001.

## 2. Justification des choix technologiques

- **Azure Data Factory** est idéal pour l'intégration de données hétérogènes, avec plus de 90 connecteurs disponibles.
- **Event Hubs** assure un débit élevé pour la télémétrie et les logs utilisateurs.
- **Synapse Analytics** combine entrepôt de données SQL et Big Data.
- **Azure Databricks** est optimisé pour les cas complexes de ML et s'intègre nativement à Synapse et Data Lake.
- **Power BI** offre une visualisation claire des KPIs pour les équipes métiers.
- **Key Vault et AAD** assurent une protection robuste des données sensibles.

## 3. Description du flux de données

### Ingestion :

- Données transactionnelles sont extraites par ADF.
- Journaux de navigation envoyés en continu via Event Hubs.
- Données sociales intégrées via Logic Apps.

### Stockage :

- Tous les flux convergent vers Data Lake Storage, en format brut.

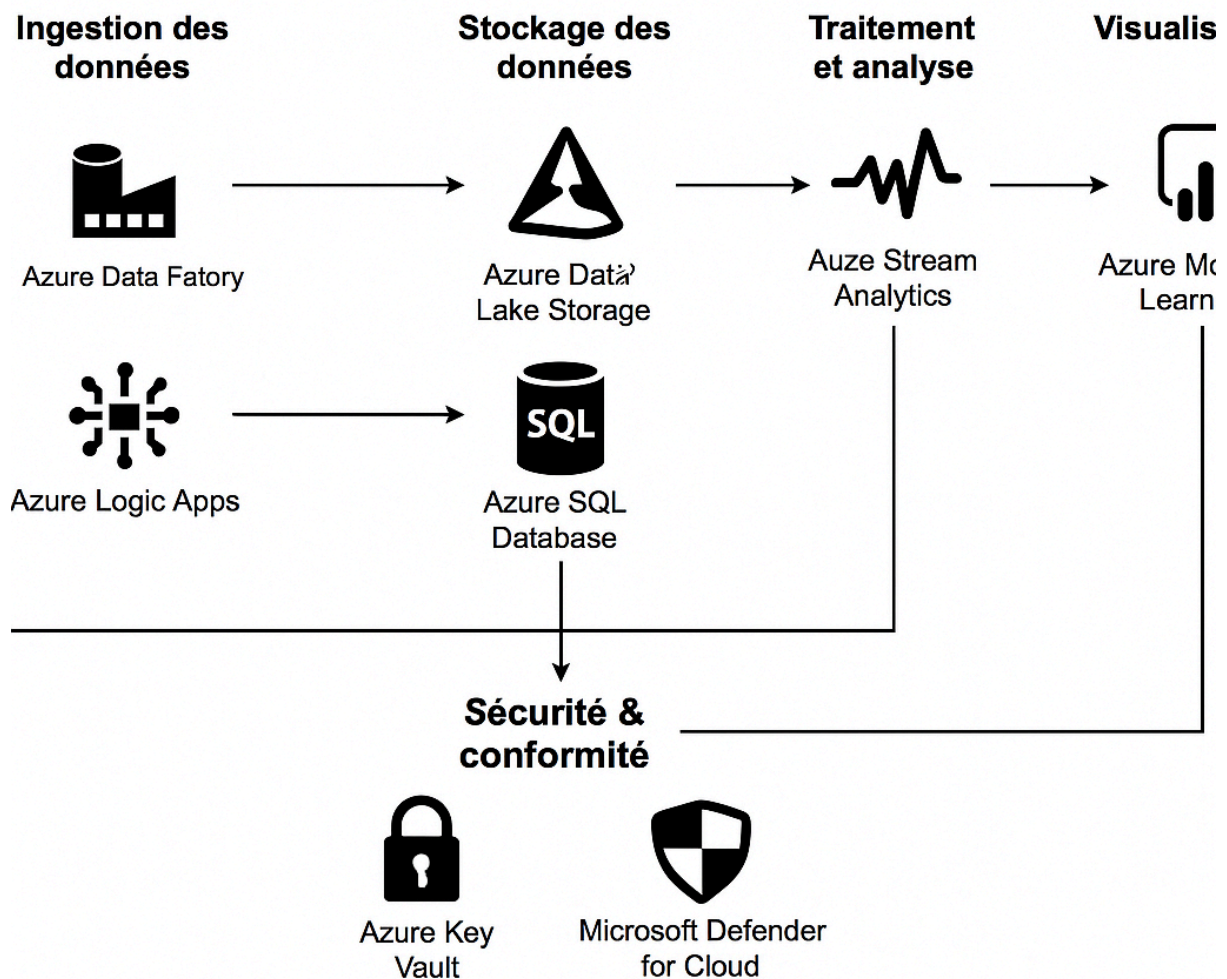
### Traitement :

- Stream Analytics analyse les événements en direct.
- Synapse traite les données historiques.
- Databricks entraîne et exécute des modèles de ML.

### Restitution :

- Résultats visualisés dans Power BI.
- Modèles déployés via Azure ML pour recommandations personnalisées.

## 4. Schéma d'architecture



## 5. Mesures de sécurité et de conformité adaptées à un environnement e-commerce

Dans le cadre d'un environnement de commerce électronique, la sécurité et la conformité des données représentent des enjeux majeurs, en particulier vis-à-vis des réglementations comme le RGPD. L'architecture proposée intègre des services et bonnes pratiques garantissant un haut niveau de protection des données clients et commerciales.

L'architecture garantit un haut niveau de sécurité et de conformité grâce à plusieurs services Azure. **Azure Key Vault** protège les secrets et clés de chiffrement, tandis que **Azure Active Directory** gère les identités avec authentification multi-facteur et contrôle

des accès. Le réseau est sécurisé par des **VNet** et **NSG**, et la surveillance continue est assurée par **Microsoft Defender for Cloud**.

Pour la conformité (notamment RGPD), **Azure Policy** impose des règles (chiffrement, restriction géographique, audit), et **Azure Blueprints** facilite le déploiement d'environnements conformes. Enfin, **Azure Monitor** et **Log Analytics** assurent la traçabilité des actions et des accès.

En résumé, l'ensemble permet de protéger les données, prévenir les intrusions, assurer la conformité réglementaire, et renforcer la confiance des utilisateurs.

### Bénéfices de l'architecture :

- **Scalabilité** : chaque service est managé, extensible selon la charge.
- **Réactivité** : décision en temps réel grâce à l'analyse streamée.
- **Personnalisation** : recommandations intelligentes augmentent le panier moyen.
- **Sécurité** : architecture conforme RGPD, avec contrôle d'accès fin et journalisation.

### Risques et Atténuation :

Risque	Atténuation
Latence sur flux temps réel	Monitoring + autoscaling Event Hubs
Échec ingestion API	Retry Logic + Alerte Logic App
Fuite de données sensibles	Key Vault + AAD + Ségrégation des rôles
Mauvaise qualité des données	Intégration de tests de validation dans ADF et Databricks