PLAN DE SÉCURITÉ & CONFORMITÉ

1. Objectifs du Plan

- Protéger les données sensibles (paiements, identités, logs, données client).
- Répondre aux obligations réglementaires : **PCI-DSS**, **RGPD**, **CCPA**, **PSD2**, **SOC2**.
- Garantir la sécurité du pipeline complet (Edge → App → Data Security → Storage).
- Assurer la traçabilité, l'auditabilité, la gouvernance et le contrôle d'accès.
- Minimiser les risques d'exposition, de perte, de manipulation ou d'accès non autorisé.

---

2. Sécurité par couches

2.1 Edge Layer (WAF, CDN, Anti-Bot)

Mesures clés :

Web Application Firewall (WAF)

- Protection contre : SQL Injection, XSS, CSRF, RCE, SSRF
- Signature-based + ML-based attack detection
- **Règles spécifiques PCI-DSS pour données de paiement**

CDN + TLS Termination

- TLS 1.2 minimum, TLS 1.3 préféré
- Certificats courts (rotation automatique)
- HSTS activé
- Chiffrement strict (AES-GCM, ECDHE)

Anti-Bot / Rate Limiting

- Détection comportementale
- Protection DDoS niveau L3–L7
- Throttling par IP, token, fingerprint

Objectif : bloquer 90 % des attaques avant l'infrastructure interne.

2.2 Application Layer (API Gateway, Auth, Microservices)

Mesures clés :

API Gateway

- Authentification obligatoire (OAuth2 / JWT)
- Validation stricte de schémas JSON (OpenAPI)
- Rate-limiting par clé API
- Logging centralisé + tracing distribué (OpenTelemetry)

Identity & OAuth

- MFA obligatoire pour admins
- Rotation automatique des tokens courts (5–15 min)
- RBAC / ABAC entièrement gérés par l'IAM
- Gestion granularisée des scopes d'API

Microservices

- Mutual TLS entre services
- Zero-Trust Networking
- Séparation stricte par namespace / VPC
- Secrets chargés uniquement via **Vault** hors code

2.3 Data Security Layer (IAM, KMS, Vault, DLP)

IAM — Engine RBAC / ABAC

- Attribution des permissions par rôle, jamais par utilisateur
- ABAC pour ressources sensibles (tags : pii=true, pci=true)
- Policies "deny by default"
- Just-In-Time Access pour équipes support

KMS — Gestion des clés

- Chiffrement **en transit** (TLS) et **au repos** (AES-256)

- Rotation automatique des clés tous les 90 jours
- Stockage séparé des clés maîtres (HSM ou Cloud KMS)
- Déchiffrement côté serveur uniquement

Vault — Secrets Manager

- Stockage chiffré de :
    - clés API
    - tokens OAuth
    - credentials DB
- Rotation automatique
- Accès par policies dynamiques + TTL courts
- Audit log de chaque extraction de secret

DLP — Data Loss Prevention

- Analyse en continu des données sensibles dans Lake, NoSQL, DWH
- Masquage automatique :
    - Numéro de carte (show last 4 digits)
    - Emails
    - IPs
- Blocage de l'exfiltration (alertes + quarantaines)

2.4 Storage Layer (OLTP, NoSQL, Data Lake, DWH)

OLTP (Paiements)

- Base chiffrée end-to-end
- PCI-DSS Segment avec réseau isolé
- Audit trail immuable (horodatage, utilisateur, action)
- Rétention limitée (minimisation RGPD)

NoSQL

- Gestion des logs / features
- Pseudonymisation en entrée (anonymisation irréversible en option)
- Indexation uniquement sur champs non sensibles
- Accès uniquement via microservices (pas direct)

Data Lake

- Zones isolées (Bronze / Silver / Gold)
- Données brutes chiffrées avec KMS
- Policies Lake Formation / IAM restrictives
- Masquage systématique des champs sensibles
- Versioning + immutabilité des données d'origine

DWH (Snowflake / BigQuery)

- Accès via RBAC/ABAC IAM
- Dynamic Masking (Snowflake) ou Authorized Views (BigQuery)
- Monitoring des requêtes sensibles
- Tables de conformité dédiées (logs d'accès, audit, lineage)

---

3. Gouvernance et Qualité des données

Aligné avec le pipeline et les blocs QA + Catalog.

Great Expectations

- Tests automatiques sur :
    - schémas
    - duplicats
    - nullité
    - ranges

    - cohérence inter-tables

Data Catalog (Amundsen / DataHub)

- Centralisation des métadonnées
- Classification automatique (PII, PCI, interne, public)
- Lineage end-to-end (Kafka → Lake → DWH → ML)
- Documentation obligatoire pour toute table / topic

Log immuable

- Traçabilité de toutes les transformations (dbt + Airflow)
- Conservation selon exigences RGPD / PCI-DSS

---

4. Conformité réglementaire

PCI-DSS (paiements)

- Réseau segmenté et isolé
- Chiffrement fort + rotation clés
- Journalisation immuable
- Masquage PAN
- Surveillance continue
- Contrôles automatiques dans Airflow / QA

RGPD

- Minimisation des données
- Droit d'accès / suppression automatisé
- Registre de traitement
- Pseudonymisation
- Conservation limitée
- Politique de rétention dans Data Lake et DWH

CCPA / SOC2

- Transparence sur collecte
- Accès contrôlé

- Audit périodique
- Monitoring continu des anomalies

---

5. Sécurité des pipelines (Batch + Streaming)

- Chiffrement Kafka (TLS + SASL)
- Authentification forte entre producers/consumers
- Schemas versionnés via Schema Registry
- Lecture seule sur zones Silver/Gold
- Airflow :
    - secrets via Vault
    - accès restreints
    - DAG signés / versionnés

---

6. Surveillance & Observabilité

- Prometheus : métriques sécurité (auth, latence, erreurs HTTP)
- ELK / OpenSearch : logs applicatifs + audit IAM
- Alertes SIEM (détection d'exfiltration, brute force, anomalie comportementale)
- Dashboards SOC automatisés

---

7. Plan de réponse aux incidents

- Playbooks :
    - fuite de données
    - compromission de clés / tokens
    - attaque DDoS
    - élévation de privilèges
- Rotation forcée des secrets avec Vault
- Analyse post-incident (root cause + corrections)