

Modélez une infrastructure hybride dans le cloud

Étape 1 : Identification et sélection des composants cloud pour InduTechData

Introduction

Dans le cadre de la modernisation de l'infrastructure informatique d'InduTechData, il est nécessaire de sélectionner les services cloud adaptés pour gérer les données IoT, les logs applicatifs et les applications métiers, tout en assurant la compatibilité avec le système on-premise existant. Cette étape repose sur l'analyse des besoins et sur la prise en compte des flux critiques, de la sécurité et des coûts.

1. Stockage des données non structurées

Besoin :

Logs applicatifs, fichiers utilisateurs, données IoT provenant des capteurs industriels.

Service cloud proposé : Amazon S3

Justification :

- **Scalabilité illimitée :** S3 peut absorber des volumes massifs de données IoT et de logs sans nécessiter de gestion d'infrastructure.
- **Sécurité intégrée :** chiffrement au repos (SSE), contrôle d'accès via IAM, versioning et politiques de rétention.
- **Interopérabilité avec Redpanda :** S3 peut servir de data lake pour stocker les données brutes avant ou après traitement. Redpanda peut y écrire via des connecteurs Kafka.

Points de vigilance intégrés :

- Flux critiques stockés et accessibles en toute sécurité
- Données accessibles aux pipelines de traitement temps réel et batch
- Stockage optimisé pour maîtriser les coûts selon les volumes

2. Entrepôt de données (Data Warehouse)

Besoin :

Centraliser les données analytiques, exécuter des requêtes SQL complexes, alimenter les dashboards BI.

Service cloud proposé : Amazon Redshift

Justification :

- **Architecture MPP (Massively Parallel Processing)** : exécution rapide des requêtes sur de grands volumes.
- **Compatibilité SQL** : migration simple depuis SQL Server.
- **Intégration BI** : compatible avec Tableau, Power BI, QuickSight.
- **Synchronisation avec SQL Server on-premise** :
 - Utilisation de **AWS DMS** pour la réplication continue.
 - Support du **Change Data Capture (CDC)** pour synchroniser les modifications en temps réel.
 - Permet une migration progressive sans interruption des services.

Points de vigilance intégrés :

- Flux batch et temps réel identifiés
- Sécurité des transferts de données via tunnel VPN ou Direct Connect
- Impact des coûts Redshift évalué selon volumes et fréquence de requêtes

3. Traitement des données en temps réel (Streaming)

Besoin :

Ingestion et traitement des flux IoT, logs applicatifs, événements métier.

Service cloud proposé : Redpanda

Justification :

- Compatible Kafka, mais plus simple à déployer et administrer.
- Très faible consommation de ressources → idéal pour les environnements hybrides.
- Latence ultra-faible → adapté aux flux IoT et aux analyses temps réel.
- Orchestration des pipelines : intégration native avec Spark, Flink, Debezium, Python.

Alternative cloud :

- AWS MSK (Kafka managé) → plus lourd et plus coûteux.
- AWS Kinesis → propriétaire, moins flexible.

Points de vigilance intégrés :

- Flux temps réel représentés et pris en compte
- Sécurité des pipelines et des flux entre on-premise et cloud
- Choix optimisé pour réduire les coûts tout en garantissant performance et scalabilité

4. Sécurisation et gestion des accès

Besoin :

Unifier la gestion des identités entre l'infrastructure on-premise et le cloud.

Service cloud proposé : AWS Directory Service + IAM Identity Center

Justification :

- Extension de l'Active Directory on-premise vers le cloud.
- **SSO (Single Sign-On)** pour toutes les applications cloud.
- Gestion centralisée des permissions via IAM Identity Center.
- **Sécurité renforcée :**
 - MFA (authentification multi-facteurs)
 - Politiques Zero Trust
 - Intégration avec VPN ou Direct Connect pour sécuriser les flux on-premise → cloud

Points de vigilance intégrés :

- Flux d'authentification sécurisés
- Gestion centralisée pour éviter les risques liés aux comptes utilisateurs
- Maintien de la continuité et conformité avec le SI existant

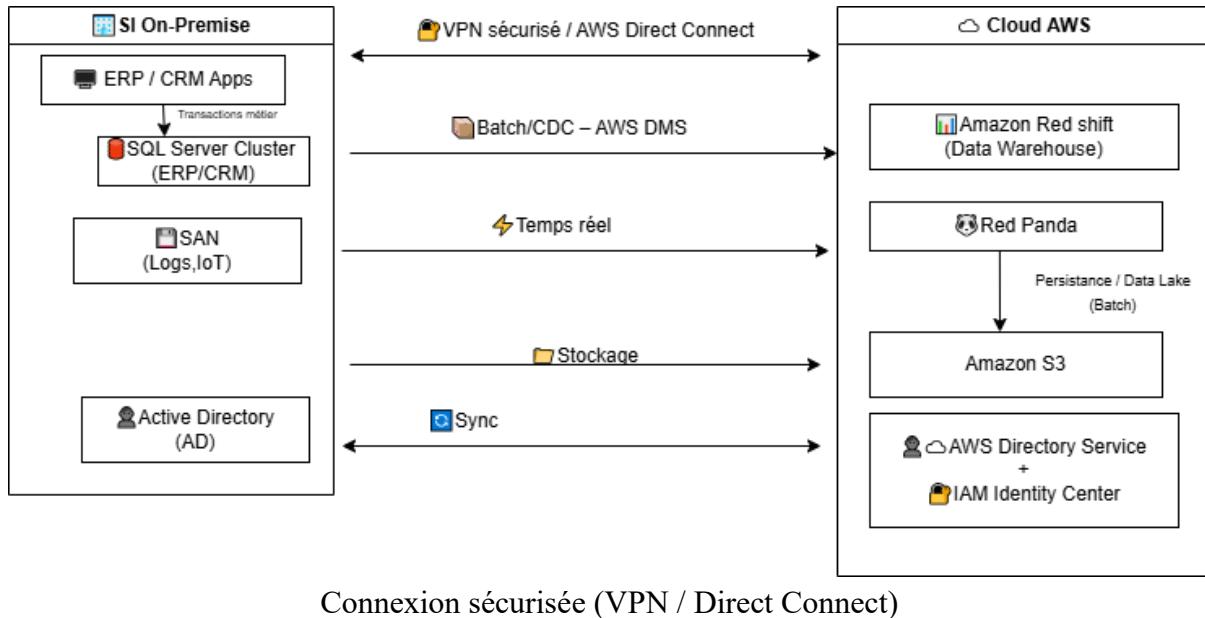
Conclusion

Cette étape permet de **sélectionner les composants cloud adaptés à l'architecture hybride d'InduTechData :**

- Amazon S3 pour le stockage des données non structurées
- Amazon Redshift pour l'entrepôt analytique
- Redpanda pour le streaming temps réel
- AWS Directory Service + IAM Identity Center pour la sécurité et la gestion des identités

Les **flux critiques, la sécurité et les coûts** ont été pris en compte pour chaque composant, conformément aux recommandations. Cette sélection constitue la base pour la modélisation graphique de l'infrastructure hybride.

Étape 2 – Représentation visuelle de l'infrastructure hybride



Connexion sécurisée (VPN / Direct Connect)

Étape 3 : Évaluation de la compatibilité de l'infrastructure hybride avec le SI existant d'InduTechData

1. Introduction

Dans le cadre de la modernisation de son système d'information, InduTechData a fait le choix d'une **architecture hybride** combinant une infrastructure **on-premise existante** et des services **cloud AWS**.

Cette architecture vise à répondre à la croissance rapide des volumes des données issus des capteurs IoT, tout en garantissant la **continuité des services métiers**, la **sécurité des données** et la **compatibilité avec le SI actuel**.

Cette évaluation analyse la **compatibilité de l'architecture proposée** avec l'environnement on-premise d'InduTechData, en abordant les aspects de **sécurité**, **interopérabilité**, **scalabilité** et **gestion des coûts**.

2. Sécurité et conformité

2.1 Protection des flux de données

La solution proposée garantit un haut niveau de sécurité pour les données sensibles, aussi bien en transit qu'au repos.

- **Redpanda** supporte nativement le chiffrement des flux via **SSL/TLS**, ce qui permet de sécuriser les échanges de données temps réel (IoT, logs) entre l'infrastructure on-premise et le cloud.
 - Les connexions entre le SI local et AWS reposent sur un **VPN sécurisé ou AWS Direct Connect**, garantissant que les flux critiques ne transitent pas en clair sur Internet.
 - Les services AWS utilisés (Amazon S3, Amazon Redshift) proposent un **chiffrement des données au repos** via AWS KMS.
- Ces mécanismes assurent la conformité avec les bonnes pratiques de sécurité et réduisent les risques de fuite ou d'altération des données.

2.2 Gestion unifiée des identités

L'architecture s'appuie sur une **extension de l'Active Directory on-premise vers le cloud** grâce à :

- **AWS Directory Service**
- **AWS IAM Identity Center**

Cette approche permet :

- Une **authentification unique (SSO)** pour les utilisateurs,
- Une gestion centralisée des identités et des permissions,
- Une continuité des politiques de sécurité existantes.

Ainsi, lorsqu'un utilisateur est ajouté ou supprimé dans l'Active Directory local, les accès cloud sont automatiquement mis à jour, garantissant une **cohérence totale des identités** entre les environnements.

3. Interopérabilité avec le SI on-premise

3.1 Intégration de Redpanda avec les systèmes existants

Redpanda est **100 % compatible avec l'API Kafka**, ce qui facilite son intégration avec les systèmes on-premise existants :

- Les applications métiers, capteurs IoT et systèmes de logs peuvent publier des événements vers Redpanda via des **producers Kafka standards**.
- Redpanda peut être connecté à des bases **SQL Server** grâce à des outils comme **Debezium** pour capter les changements de données (CDC).
- Les données peuvent être automatiquement redirigées vers Amazon S3 ou Amazon Redshift via des connecteurs Kafka.

Cette compatibilité garantit une **interopérabilité fluide**, sans refonte majeure du SI existant.

3.2 Automatisation des flux de données

L'architecture permet de gérer deux types de flux :

- **Flux temps réel** : IoT et logs → Redpanda → traitement ou archivage.
- **Flux batch / CDC** : SQL Server → AWS DMS → Amazon Redshift.

Ces flux sont **entièrement automatisés**, ce qui réduit les interventions manuelles, limite les erreurs humaines et améliore la fiabilité globale du système.

4. Scalabilité et gestion des coûts

4.1 Scalabilité de l'architecture

L'architecture proposée est conçue pour accompagner la croissance future d'InduTechData :

- **Amazon S3** offre une capacité de stockage pratiquement illimitée, idéale pour absorber l'augmentation des données IoT et des logs.
- **Amazon Redshift** permet d'augmenter ou réduire les capacités de calcul selon les besoins analytiques.
- **Redpanda** est capable de gérer un grand nombre de flux temps réel avec une faible latence et une consommation de ressources optimisée.

Cette approche garantit une **montée en charge progressive**, sans remise en cause de l'architecture globale.

4.2 Surveillance et optimisation des coûts

Afin de maîtriser les coûts cloud, plusieurs outils AWS peuvent être utilisés :

- **AWS CloudWatch** pour surveiller l'utilisation des ressources,
- **AWS Budgets** pour définir des alertes en cas de dépassement,
- **AWS Cost Explorer** pour analyser les dépenses par service.

Ces outils permettent d'anticiper les dérives budgétaires et d'optimiser les ressources consommées.

4.3 Estimation des coûts (ordre de grandeur)

Composant	Type de coût
Amazon S3	Faible au départ, dépend du volume stocké
Amazon Redshift	Coût récurrent lié aux nœuds et à la puissance de calcul

Redpanda	Coût des instances EC2, optimisé grâce à une faible consommation
AWS DMS	Coût temporaire ou continu selon la synchronisation
AWS Directory Service	Coût mensuel fixe

- L'approche hybride permet d'éviter une migration complète immédiate, ce qui **réduit les coûts initiaux** tout en permettant une évolution progressive.

5. Avantages, limitations et points d'attention

Avantages	Limitations	Points d'attention
<ul style="list-style-type: none"> -Compatibilité élevée avec le SI existant -Sécurité renforcée et gestion unifiée des identités -Scalabilité adaptée aux besoins IoT -Architecture moderne et évolutive 	<ul style="list-style-type: none"> -Dépendance partielle aux services AWS -Nécessité de compétences cloud et streaming -Coût croissant avec l'augmentation des volumes de données 	<ul style="list-style-type: none"> -Surveillance continue des coûts -Gouvernance des accès et des permissions -Supervision des flux temps réel -Plan de continuité et de reprise d'activité

6. Conclusion

L'architecture hybride proposée est **hautement compatible avec l'environnement SI on-premise d'InduTechData**. Elle permet une transition progressive vers le cloud, tout en maintenant les exigences de sécurité, de performance et de continuité métier.

Cette solution constitue une base solide pour accompagner la croissance future de l'entreprise et l'intégration de nouvelles sources de données industrielles.