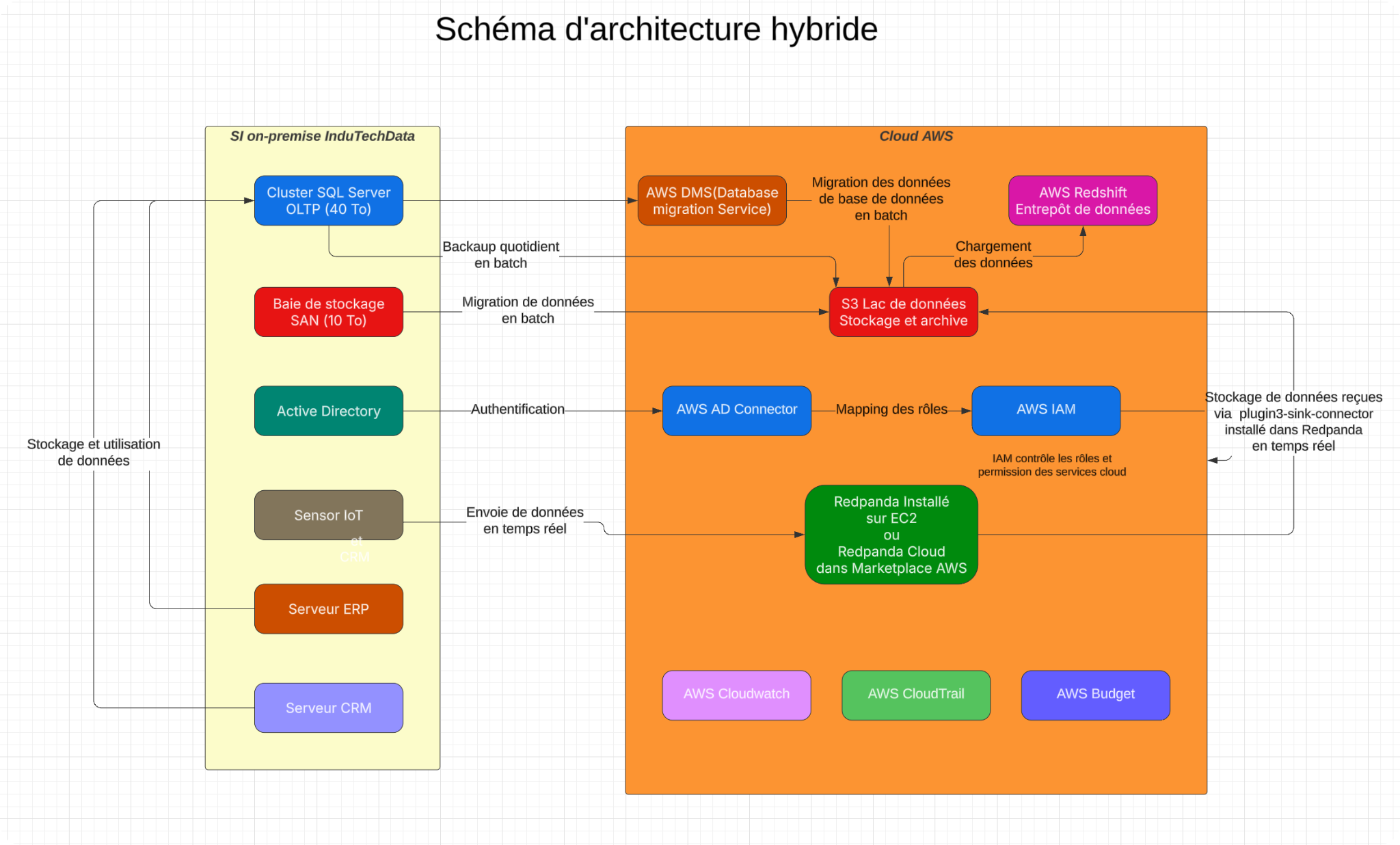
Choix de composants AWS cloud



Amazon S3: Stockage de données non structurées

Description : Service hautement scalable capable de gérer des pétaoctets de données, Excellent rapport coût/performance pour le stockage d'objets, compatible avec Redpanda via des connecteurs natifs avec le plugin (s3-sink-connector) à installer via la commande `rpk plugin install s3-sink-connector` , il dispose plusieures classes de stockage avec S3 qui ont des facturations différentes, on peut configurer le lifecycle management pour changer des classes de stockage qui coûte moins élevé à un moment donné(par exemple 30 jours à un cool tier et 90 jours à un archivage tier).

Sécurité:

* En transit:
  + TLS 1.2/1.3 forcé pour toutes les communications
  + Vérification des certificats SSL
  + Validation MD5 des contenus transmis
* En repos:
  + Chiffrement côté serveur avec KMS
  + Clés gérées par AWS (SSE-KMS)
  + Bucket policy forçant le HTTPS
  + Vérification d'intégrité automatique
* Contrôle d'accès :
  + IAM avec conditions strictes
  + Accès restreint par IP/VPC possible
  + Audit trail via CloudTrail
  + Monitoring des accès

Coût pour 50 To par mois:

* S3 Standard : accès fréquent (1228.8 USD)
* S3 Standard-IA : accès peu fréquent (670.72 USD)
* S3 Glacier : archivage (208.35 USD)
* S3 Glacier Deep Archive : archivage long terme (92.95 USD)

Cas d'usage dans l’architecture hybride :

* + Data Lake pour stockage des données(données IoT, logs, etc …)
* Backups de données du server SQL
* Sauvegarde des données IoT reçu dans Redpanda qui permet ensuite d’exploiter par Redshift

Amazon Redshift: Entrepôt de données

Description: C’est une base de données columnaire (optimisée pour l'analytique) , il a une architecture MPP (Massive Parallel Processing), il est compatible avec des requêtes SQL, facilitant la migration depuis SQL Server en utilisant AWS DMS (Database Migration Service), il est intégré native avec S3 pour le chargement de données. Un nœud de Redshift peut prendre en maximum 60 To de données et un cluster peut avoir jusqu’à 128 nœuds, donc un maximum de 7.68 Petabytes.

Coût: Il a des coûts différents selon si c’est Redshift (selon les types de nœuds aussi) ou Redshift Serverless.

Limitation:

* Limite de 5 To par opération COPY
* Consomme des ressources Redshift pendant le chargement
* Plus cher (compute nodes actifs pendant le chargement)
* Moins flexible pour la transformation des données

Coût par mois:

* Redshift Serverless 128 RPU 1585.02 USD
* Redshift 2 instances ra3.plus 1,846.90 USD

Cluster de Redpanda dans EC2: Traitement temps réel

Description: c’est un cluster de Redpanda qu’on va installer sur les machines virtuelles AWS EC2 pour reçevoir les flux de données en temps réel.

Sécurité:

* Configuration de VPC dédié
* Groupe de sécurité qui permet de contrôler les accès et sorties dans le réseau
* Haute disponibilité
  + 1 noeud leader et 2 noeuds worker
  + Auto Scaling Group pour la récupération automatique
  + Route 53 pour le DNS failover
  + Multi-AZ deployment

Stockage:

* EBS pour la persistance de données
* Synchronisation de données à AWS S3 avec s3-sink-connector qu’on installe dans Redpanda

Coût :

* 3 \* r6i.2xlarge 679.23 USD par mois

AWS DMS(Databases Migration Service): Migration de données

Description: il permet la migration de bases de données vers AWS, de réplication continue des données, il support de migrations homogènes et hétérogènes

Coût par mois pour 40 To de données avec 4 DCUs: 606.28USD

AWS Directory Service (AD Connector): Gestion des identités

Description: C’est une extension transparente de l'Active Directory on-premise vers le cloud avec Single Sign-On (SSO) pour tous les services, il conserve des politiques de sécurité existantes et il n’y a pas de duplication des identités

AWS Cloud Watch: Gestion des identités

Description: c’ est un service de surveillance qui collecte et suit les métriques, les logs et les alarmes pour l'infrastructure AWS. Il offre des tableaux de bord personnalisables pour visualiser les performances en temps réel, permet de créer des alertes basées sur des seuils et peut déclencher des actions automatiques. CloudWatch est principalement utilisé pour le monitoring des performances, la collecte de métriques automatiques des services AWS et la création de métriques personnalisées. Il permet également la centralisation et l'analyse des logs applicatifs.

Cas d'usage dans l’architecture hybride :

* Surveillance des coûts et création d’alerte si le budget atteint à un certain niveau
* Surveillance des services cloud
* Analyse des logs des services

Coût pour 100 Go par mois:

* 60.32 USD

AWS CloudTrail: audit et sécurité de l'infrastructure AWS

Description: Il enregistre automatiquement tous les appels d'API effectués dans votre compte AWS, fournissant ainsi un historique complet des actions réalisées sur votre infrastructure. Ces enregistrements incluent l'identité de l'appelant API, le moment de l'appel, l'adresse IP source, les paramètres de la requête et les éléments de réponse. Cette piste d'audit détaillée est essentielle pour la conformité réglementaire, l'analyse de sécurité et le dépannage opérationnel.

AWS Budgets: définir et suivre les dépenses AWS

Description: il permet de suivre des dépenses mensuelles/annuelles, des instances réservées. Et aussi de définir des budgets par service/projet et de configurer des alertes de budget.

Coût:

* 2 budgets gratuits par mois
* 0,02 USD par budget supplémentaire/jour
* Alertes incluses