



***LeCloudFacile.com***

## **Fiche de révision Analytique sur AWS**

[Amazon Redshift](#)

[Amazon EMR](#)

[Amazon Athena](#)

[Amazon Quicksight](#)

[AWS Glue](#)

[Suivez-nous](#)

---

## Amazon Redshift

- **Introduction**
  - Amazon Redshift est une base de données basée sur PostgreSQL conçue pour le traitement analytique en ligne (OLAP) et l'entreposage de données, contrairement à RDS qui est utilisé pour le traitement transactionnel en ligne (OLTP). Redshift est particulièrement adapté pour l'analyse des données et l'entreposage à grande échelle.
- **Fonctionnement de Redshift**
  - **OLAP vs OLTP**
    - **OLTP (Online Transaction Processing)** : Utilisé pour les transactions courantes, comme les bases de données RDS.
    - **OLAP (Online Analytical Processing)** : Utilisé pour l'analyse et l'entreposage des données, ce qui est le rôle de Redshift.
- **Caractéristiques Principales**
  - Stockage Columnar : Les données sont stockées en colonnes plutôt qu'en lignes, optimisant ainsi les performances pour les requêtes analytiques.
  - Massively Parallel Processing (MPP) : Redshift utilise un moteur de requête massivement parallèle pour exécuter rapidement des calculs complexes.
  - Scalabilité : Capable de gérer des pétaoctets de données.
  - Interface SQL : Permet d'interroger les données avec des commandes SQL familières.
- **Cas d'Utilisation**
  - Entreposage de Données : Idéal pour stocker et analyser de grandes quantités de données.
  - Business Intelligence (BI) : Intégré avec des outils BI comme QuickSight ou Tableau pour créer des tableaux de bord et des visualisations à partir des données.
  - Rapports et Analyses : Utilisé pour générer des rapports et effectuer des analyses approfondies des ensembles de données.
- **Redshift Serverless**
  - Aperçu

- Redshift Serverless est une fonctionnalité qui permet d'exécuter Redshift sans gérer l'infrastructure sous-jacente. Cette option est particulièrement utile pour les charges de travail analytiques où la mise à l'échelle automatique est nécessaire.
- **Fonctionnement de Redshift Serverless**
  - **Activation** : Activez Redshift Serverless sur votre compte AWS.
  - **Connexion** : Utilisez l'éditeur de requêtes Redshift ou tout autre outil compatible pour commencer à écrire vos requêtes.
  - **Provisionnement Automatique** : Redshift Serverless provisionne et met à l'échelle automatiquement la capacité en fonction de la charge de travail et des requêtes.
  - **Facturation** : Payez uniquement pour le calcul et le stockage utilisés pendant l'analyse, réduisant ainsi les coûts.
- **Avantages de Redshift**
  - Performances : Jusqu'à 10 fois plus rapide que les autres entrepôts de données pour l'analyse des données.
  - Efficacité Coût : Paiement à l'utilisation pour le calcul et le stockage, particulièrement avec Redshift Serverless.
  - Simplicité : Gestion automatisée de l'infrastructure avec Redshift Serverless, simplifiant les opérations et réduisant les besoins en maintenance.

## Amazon EMR

- **Introduction**
  - Amazon EMR (Elastic MapReduce) n'est pas une base de données traditionnelle mais un service de traitement de données massives qui permet de créer des clusters Hadoop sur AWS. Ce service est conçu pour analyser et traiter de grandes quantités de données de manière distribuée.
- **Fonctionnement de Amazon EMR**
  - **Qu'est-ce qu'un Cluster Hadoop ?**
  - **Hadoop** : Technologie open source qui permet à plusieurs serveurs de travailler ensemble pour analyser de grandes quantités de données.
  - **Cluster Hadoop** : Ensemble de serveurs fonctionnant en collaboration pour traiter des données massives.
- **Rôle de Amazon EMR**

- Provisionnement Automatisé : EMR s'occupe de provisionner et de configurer les instances EC2 nécessaires pour former un cluster Hadoop.
- Mise à l'Échelle Automatique : EMR ajuste automatiquement la taille du cluster en fonction des besoins de la charge de travail.
- Utilisation des Instances Spot : Intégré aux instances Spot pour réduire les coûts d'exécution des clusters.
- **Écosystème Big Data**
  - Projets Compatibles : EMR supporte des projets populaires de l'écosystème Hadoop tels que Apache Spark, HBase, Presto, et Flink.
  - Traitement de Données Distribués : Permet d'exécuter des tâches de traitement de données distribuées sur des centaines d'instances EC2.
- **Cas d'Utilisation**
  - Traitement des Données : Idéal pour les tâches de traitement de données volumineuses.
  - Apprentissage Automatique : Utilisé pour exécuter des algorithmes de machine learning à grande échelle.
  - Indexation Web : Permet d'indexer de grandes quantités de données web.
  - Big Data en Général : Toute application nécessitant le traitement et l'analyse de données massives.
- **Avantages d' Amazon EMR**
  - Gestion Simplifiée : EMR simplifie la gestion des clusters Hadoop, réduisant la complexité du provisionnement et de la configuration.
  - Scalabilité : Capacité à ajuster automatiquement la taille des clusters en fonction des besoins de traitement.
  - Coût-Efficacité : Utilisation des instances Spot pour réduire les coûts et payer uniquement pour les ressources utilisées.

## Amazon Athena

- **Introduction**
  - Amazon Athena est un service de requête interactif et sans serveur permettant d'analyser des données directement dans Amazon S3 en utilisant SQL standard. Il est conçu pour rendre l'analyse des données rapide et facile sans nécessiter de chargement préalable des données.
- **Architecture et Technologie**

- **Sans Serveur** : Athena est un service sans serveur, ce qui signifie que vous n'avez pas besoin de gérer d'infrastructure sous-jacente.
- **Langage SQL** : Utilise SQL standard pour interroger les fichiers stockés dans Amazon S3.
- **Moteur Presto** : Athena est construit sur le moteur de requête Presto, conçu pour effectuer des requêtes rapides et distribuées.
- **Processus de Fonctionnement**
  - **Chargement des Données** : Les utilisateurs chargent leurs données dans Amazon S3.
  - **Interrogation des Données** : Les requêtes SQL sont exécutées directement sur les fichiers stockés dans S3.
  - **Formats de Données supportés** : Athena prend en charge divers formats de fichiers, notamment CSV, JSON, Avro, Parquet et ORC.
  - **Analyse et Reporting** : Les résultats des requêtes peuvent être utilisés pour créer des rapports et des visualisations, notamment via Amazon QuickSight.
- **Tarifification**
  - **Coût** : Le coût est basé sur la quantité de données analysées, à environ 5 \$ par téraoctet de données scannées.
  - **Économies** : L'utilisation de formats de données compressés ou partitionnés peut réduire les coûts en diminuant la quantité de données scannées.
- **Cas d'Utilisation**
  - **Analyses de Business Intelligence** : Idéal pour les rapports BI et les visualisations de données.
  - **Analyse des Journaux** : Parfait pour analyser les journaux de flux, les journaux d'API, les journaux d'ELB, etc.
  - **Analyse Ad Hoc** : Utilisé pour des requêtes analytiques ponctuelles sur des données stockées dans S3.
  - **ETL et Préparation des Données** : Utilisé pour transformer et préparer des données avant de les charger dans des entrepôts de données ou des data lakes.
- **Avantages de Amazon Athena**
  - **Facilité d'Utilisation** : Aucune infrastructure à gérer, rendant le service très facile à utiliser.
  - **Rapidité et Efficacité** : Permet d'exécuter des requêtes rapidement sur de grandes quantités de données.

- **Flexibilité des Formats de Données** : Prend en charge plusieurs formats de fichiers, offrant une grande flexibilité dans le traitement des données.
- **Intégration avec S3** : Fonctionne directement avec les données stockées dans Amazon S3, éliminant le besoin de les déplacer.

## Amazon Quicksight

- **Introduction**
  - Amazon QuickSight est un service de business intelligence (BI) sans serveur, basé sur le cloud, qui permet de créer des tableaux de bord interactifs et de visualiser des données. QuickSight utilise l'apprentissage automatique pour fournir des insights avancés.
- **Création de Tableaux de Bord**
  - **Visualisation des Données** : Permet de créer divers types de graphiques et de tableaux pour représenter visuellement des données.
  - **Interactivité** : Les tableaux de bord sont interactifs, permettant aux utilisateurs de filtrer et d'explorer les données en profondeur.
  - **Accès Web et Mobile** : Les tableaux de bord peuvent être consultés via des navigateurs web et des applications mobiles.
- **Performance et Évolutivité**
  - **Rapide et Évolutif** : QuickSight est conçu pour être rapide et peut s'adapter automatiquement à la charge, ce qui est utile pour les entreprises de toutes tailles.
  - **Sans Serveur** : Aucune infrastructure à gérer, ce qui simplifie l'utilisation et la gestion du service.
  - **Tarifcation par Session** : Le coût est basé sur l'utilisation par session, ce qui le rend économique et flexible.
- **Intégrations**
  - QuickSight s'intègre avec de nombreux services et sources de données AWS, notamment :
  - **Bases de Données** : RDS, Aurora, Redshift.
  - **Services de Stockage** : Amazon S3.
  - **Services d'Analyse** : Amazon Athena.

- **Applications Externes** : QuickSight peut également se connecter à des sources de données tierces, telles que des bases de données on-premise ou d'autres services cloud.
- **Cas d'Utilisation**
  - **Analyse Commerciale** : Fournir des insights commerciaux basés sur les données.
  - **Création de Visualisations** : Créer des graphiques et des tableaux pour représenter visuellement les données.
  - **Analyse Ad Hoc** : Réaliser des analyses ponctuelles pour répondre à des questions spécifiques.
  - **Reporting** : Générer des rapports et des tableaux de bord pour différentes parties prenantes.
- **Avantages de Amazon QuickSight**
  - **Simplicité** : Interface utilisateur intuitive, facilitant la création de visualisations sans avoir besoin de compétences techniques avancées.
  - **Apprentissage Automatique** : Intégré pour fournir des insights avancés et des prédictions.
  - **Collaboration** : Permet de partager facilement les tableaux de bord et les insights avec les membres de l'équipe ou les parties prenantes externes.
  - **Accessibilité** : Accessible depuis n'importe quel navigateur web ou application mobile, offrant une flexibilité maximale.

## AWS Glue

- **AWS Glue**
  - AWS Glue est un service entièrement géré par AWS qui facilite l'extraction, la transformation et le chargement (ETL) de données pour le traitement analytique. Il est conçu pour simplifier le processus de préparation des données en automatisant de nombreuses tâches fastidieuses.
- **ETL Géré**
  - **Extraction** : Permet d'extraire des données à partir de diverses sources telles que Amazon S3, Amazon RDS, Amazon Redshift, etc.
  - **Transformation** : Facilite la transformation des données en appliquant des scripts Python ou en utilisant des jobs Glue basés sur Spark.

- 
- **Chargement** : Charge les données transformées dans des entrepôts de données, des lacs de données ou d'autres destinations AWS.
  - **Sans Serveur**
    - **Architecture sans Serveur** : AWS Glue élimine le besoin de provisionner et de gérer des serveurs, ce qui permet de se concentrer uniquement sur la logique de transformation des données.
    - **Évolutivité Automatique** : Gère automatiquement la mise à l'échelle en fonction des volumes de données et des besoins de calcul.
  - **Intégration et Automatisation**
    - **Intégration avec d'autres Services AWS** : Se connecte facilement à d'autres services AWS comme Amazon S3, Amazon RDS, Amazon Redshift, etc., pour l'intégration de données.
    - **Automatisation des Tâches** : Planifie et orchestre automatiquement les jobs ETL, ce qui réduit la gestion manuelle et les erreurs potentielles.
  - **Utilisation et Avantages**
    - **Simplicité d'Utilisation** : Interface conviviale pour configurer, gérer et surveiller les jobs ETL sans nécessiter une expertise approfondie en infrastructure.
    - **Flexibilité** : Supporte une variété de sources de données et offre des options de transformation flexibles pour répondre aux besoins spécifiques de chaque application.
    - **Optimisation des Performances** : Utilise le moteur Apache Spark pour des transformations rapides et évolutives sur de gros volumes de données.

---

## Suivez-nous

- **Site web** : <https://lecloudfacile.com>
- **Youtube** : <https://www.youtube.com/@lecloudfacile>
- **Linkedin** : <https://www.linkedin.com/company/lecloudfacile/>
- **Udemy** :  
<https://www.udemy.com/course/nouveau-aws-cloud-practitioner-clf-c02/?referralCode=8CE99E6C2100F1998BDF>
- **Communauté WhatsApp** :  
<https://chat.whatsapp.com/HleIILVOJ9xCJKX8VhbLSr>