Formation AWS – Big Data & Analytics

Informations

Durée : 4 jours (28h.)

Tarif* : Nous consulter

Réf : AWBD

Niveau : Moyen

intra

Mise à jour le 18/12/25

*tarif valable jusqu'au 31/12/2026

Prochaines sessions

Contactez-nous pour connaitre nos futures sessions.

Pré-requis

- Connaissances en SQL et traitement de données distribuées (Spark/Hadoop recommandé)
- Expérience avec AWS (S3, IAM, VPC)
- Notions de streaming et ETL souhaitées

Objectifs

Objectifs pédagogiques :

- Comprendre l'écosystème Big Data et Analytics d'AWS.
- Mettre en place des clusters EMR pour le traitement distribué.
- Ingestion et traitement en streaming avec Kinesis.
- Créer des dashboards interactifs avec QuickSight.
- Appliquer les bonnes pratiques de sécurité et d'optimisation des coûts

Objectifs opérationnels :

- Concevoir, déployer et exploiter une chaîne complète Big Data sur AWS : de l'ingestion des données (batch & streaming) à leur traitement (EMR), jusqu'à la visualisation et la prise de décision via QuickSight, tout en appliquant les meilleures pratiques de sécurité, gouvernance et optimisation des coûts.

Programme

Jour 1 – Introduction au Big Data et Amazon EMR

Concepts fondamentaux du Big Data

Les 3V (Volume, Vélocité, Variété)
Architectures Big Data : Data Lake, Data Warehouse, Data Mesh
Positionnement AWS dans l'écosystème Big Data

Amazon EMR (Elastic MapReduce)

Architecture et composants : Master Node, Core Node, Task Node
Frameworks disponibles : Hadoop, Spark, Hive, Presto
Intégration avec S3 (Data Lake) et DynamoDB

Gestion des clusters

Modes de déploiement (On-Demand, Spot Instances, Auto Scaling)
Optimisation des coûts et des performances
Atelier pratique : déploiement d'un cluster EMR et exécution d'un job Spark simple (analyse de logs)

Jour 2 – Traitement de données en temps réel avec Amazon Kinesis

Amazon Kinesis Data Streams

Concepts : shards, producteurs, consommateurs
Consommation avec KCL (Kinesis Client Library)

Kinesis Data Firehose

Ingestion de données en temps réel vers S3, Redshift, OpenSearch
Transformation à la volée (Lambda)

Kinesis Data Analytics

Analyse temps réel avec SQL

Cas d’usage : analyse de logs applicatifs, clickstream, IoT

Intégration avec d’autres services AWS

Kinesis + Lambda pour traitement serverless
Kinesis + S3 + Athena pour analytics
Atelier pratique : création d’un pipeline temps réel avec Kinesis Streams + Firehose vers S3, analyse des données avec Athena

Jour 3 – Business Intelligence avec Amazon QuickSight

Introduction à QuickSight

Connexion aux sources de données : S3, Redshift, Athena, RDS
Gestion des datasets et préparation des données

Visualisation et dashboards

Création de rapports et graphiques interactifs
Paramètres, filtres et contrôle d’accès
Fonctionnalités ML intégrées (détection d’anomalies, prévisions)

Sécurité et gouvernance

Gestion des utilisateurs et groupes
Partage sécurisé des dashboards
Atelier pratique : création d’un dashboard BI complet à partir de données en temps réel ingérées par Kinesis et stockées dans S3

Jour 4 – Cas d’usage complet et gouvernance Big Data

Intégration bout en bout des services Big Data AWS

Pipeline type : ingestion (Kinesis) → traitement batch/stream (EMR + Lambda) → stockage (S3/Redshift) → visualisation (QuickSight)

Supervision et optimisation

CloudWatch Logs et métriques pour EMR et Kinesis
Alertes et optimisation des coûts des clusters EMR

Étude de cas complet

Mise en place d’une plateforme Big Data pour un site e-commerce (analyse des logs utilisateurs + tableaux de bord BI)

Préparation certification

AWS Certified Data Analytics – Specialty : thèmes couverts et exemples de questions