

Building Batch Data Analytis solutions on Amazon Web Services

Référence : **AWS-BBDAS**

Durée : **1 jour (7 heures)**

Certification : **Aucune**

CONNAISSANCES PRÉALA

Avoir au moins un an d'expérience dans la gestion de frameworks de données open source tels qu'Apache Spark ou Apache Hadoop

PROFIL DES STAGIAIRES

Ingénieurs de plateformes de données
Architectes et opérateurs qui créent et gèrent des pipelines d'analyse de données

OBJECTIFS

Comparer les fonctionnalités et les avantages des entrepôts de données, des lacs de données et des architectures de données modernes
Concevoir et mettre en œuvre une solution d'analyse de données par lots
Identifier et appliquer les techniques appropriées, y compris la compression, pour optimiser le stockage des données
Sélectionner et déployer les options appropriées pour ingérer, transformer et stocker des données
Choisir les types d'instance et de nœud, les clusters, la mise à l'échelle automatique et la topologie de réseau appropriés pour un cas d'utilisation métier particulier
Comprendre comment le stockage et le traitement des données affectent les mécanismes d'analyse et de visualisation nécessaires pour obtenir des informations commerciales exploitables
Sécuriser les données au repos et en transit
Surveiller les charges de travail d'analyse pour identifier et résoudre les problèmes
Appliquer les meilleures pratiques de gestion des coûts

CERTIFICATION PRÉPARÉE

Aucune

MÉTHODES PÉDAGOGIQUES

6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
Remise d'une documentation pédagogique papier ou numérique pendant le stage
La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

FORMATEUR

Consultant-formateur expert Cloud

MÉTHODES D'ÉVALUATION DES ACQUIS

Auto-évaluation des acquis par le stagiaire via un questionnaire
Attestation des compétences acquises envoyée au stagiaire
Attestation de fin de stage adressée avec la facture

CONTENU DU COURS

1. Présentation de l'analyse des données et du pipeline de données

Cas d'utilisation de l'analyse de données
Utilisation du pipeline de données pour l'analyse

2. Présentation d'Amazon EMR

Utilisation d'Amazon EMR dans les solutions d'analyse
Architecture du cluster Amazon EMR
💡 *Démo interactive 1 : Lancement d'un cluster Amazon EMR*
Stratégies de gestion des coûts

3. Pipeline d'analyse de données à l'aide d'Amazon EMR : ingestion et stockage

Optimisation du stockage avec Amazon EMR
Techniques d'ingestion de données

4. Analyse de données par lots hautes performances à l'aide d'Apache Spark sur Amazon EMR

Cas d'utilisation d'Apache Spark sur Amazon EMR
Pourquoi choisir Apache Spark sur Amazon EMR
Concepts d'étincelle

💡 *Démo interactive 2 : Connectez-vous à un cluster EMR et exécutez des commandes Scala à l'aide de la commande*

Coque d'étincelle

Transformation, traitement et analytique

Utilisation de blocs-notes avec Amazon EMR

💡 *Exercice pratique 1 : Analyse de données à faible latence à l'aide d'Apache Spark sur Amazon EMR*

5. Traitement et analyse des données de lot avec Amazon EMR et Apache Hive

Utilisation d'Amazon EMR avec Hive pour traiter les données de lot
Transformation, traitement et analytique

💡 *Exercice pratique 2 : Traitement de données par lots à l'aide d'Amazon EMR avec Hive*

Présentation d'Apache HBase sur Amazon EMR

6. Traitement des données sans serveur

Traitement, transformation et analyse des données sans serveur
Utilisation d'AWS Glue avec les charges de travail Amazon EMR

💡 *Exercice pratique 3 : Orchestrer le traitement des données dans Spark à l'aide d'AWS Step Functions*

7. Sécurité et surveillance des clusters Amazon EMR

Sécurisation des clusters EMR

💡 *Démo interactive 3 : Chiffrement côté client avec EMRFS*

Surveillance et dépannage des clusters Amazon EMR

💡 *Démo : Examen de l'historique du cluster Apache Spark*

8. Conception de solutions d'analyse de données par lots

Cas d'utilisation de l'analyse de données par lots

Activité : Conception d'un flux de travail d'analyse de données par lots

9. Développement d'architectures de données modernes sur AWS

10. Architectures de données modernes

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:0171197030) ou par mail à referent.handicap@edugroupe.com pour vos éventuels besoins d'aménagements, afin de vous offrir la meilleure .