

Créer des solutions d'analyse de données par lots sur Amazon EMR

Date et durée
Code formation : AWS26FR Durée : 1 jour Nombre d'heures : 20 heures
Description
Amazon EMR (Elastic Map Reduce) simplifie l'exécution de frameworks Big Data comme Spark et Hadoop pour l'analyse de données par lots. Il permet de traiter efficacement de grands volumes de données (ETL, analyse de logs, etc.). EMR offre flexibilité, optimisation des coûts et intégration avec d'autres services AWS. Cette formation en data analytics vous fournit les compétences nécessaires pour maîtriser ces concepts et créer des solutions optimales. Vous découvrirez la conception, le déploiement et l'optimisation de solutions d'analyse de données EMR sur AWS, en vous appuyant sur les bonnes pratiques du secteur. Ainsi, vous serez capable de mettre en œuvre des pipelines de données efficaces, de choisir les services AWS appropriés, de garantir la sécurité des données et d'optimiser les coûts. À l'issue de ce programme d'une journée, vous maîtriserez également les compétences clés pour concevoir et implémenter des solutions d'analyse de données par lots et pour développer des architectures de données modernes sur AWS. 
<i>En tant que partenaire de formation premium (ATP) agréé par Amazon Web Services, Oo2 vous propose des formations qualifiantes et certifiantes qui respectent les normes de qualité rigoureuses de l'organisme.</i>
Objectifs
À l'issue de la formation AWS Amazon EMR, vous atteindrez les objectifs suivants : • comparer les caractéristiques 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 les données ; • choisir les types d'instances et de nœuds, les clusters, l'auto-scaling 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 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.

Points forts
• Formateurs experts AWS certifiés : bénéficiez de l'expertise de formateurs reconnus et certifiés par AWS dans le domaine de l'analyse de données. • Mise en pratique interactive : maîtrisez les techniques et les outils d'analyse de données AWS à travers des démos et des exercices. Vous serez ainsi préparé aux défis concrets du développement de solutions d'analyse de données sur AWS. • Acquisition des compétences clés : le contenu de la formation est rigoureusement aligné sur les objectifs de la conception et de la mise en œuvre de solutions d'analyse de données par lots et le développement d'architectures de données modernes sur AWS.
Modalités d'évaluation
Travaux Pratiques Etude de cas
Pré-requis
<i>Suivre cette formation AWS nécessite les prérequis suivants :</i> • Au moins 1 an d'expérience en gestion de frameworks de données open source, tel que sur Apache Spark ou Apache Hadoop.
Public
<i>Cette formation s'adresse aux publics suivants :</i> • les ingénieurs de plateforme de données qui conçoivent et déploient des solutions d'analyse de données. • les architectes et les opérateurs de données qui sont responsables de la construction et de la gestion des pipelines d'analyse de données.
Programme
Module 1 : comprendre l'analyse des données et du pipeline de données • Les cas d'utilisation de l'analyse des données. • L'utilisation du pipeline de données pour l'analyse. Module 2 : découvrir Amazon EMR • L'utilisation d'Amazon EMR dans les solutions analytiques. • L'architecture du cluster Amazon EMR. • Les stratégies de gestion des coûts. <i>Labs :</i> • Lancement d'un cluster Amazon EMR. Module 3 : explorer le pipeline d'analyse des données • L'ingestion et le stockage • L'optimisation du stockage avec Amazon EMR.

- Les techniques d'ingestion des données.

Module 4 : effectuer une analyse de données par lots haute performance avec Apache Spark

- Les cas d'utilisation d'Apache Spark sur Amazon EMR.
- Pourquoi Apache Spark sur Amazon EMR.
- Les concepts Spark.
- La transformation, le traitement et l'analyse.
- L'utilisation de notebooks avec Amazon EMR.

Labs :

- Connexion à un cluster EMR et exécution de commandes Scala à l'aide du shell Spark.
- Analyse de données à faible latence avec Apache Spark sur Amazon EMR.

Module 5 : traiter et analyser des données par lots avec Amazon EMR et Apache Hive

- L'utilisation d'Amazon EMR avec Hive pour traiter des données par lots.
- La transformation, le traitement et l'analyse.
- Introduction à Apache HBase sur Amazon EMR.

Labs :

- Traitement de données par lots en utilisant Amazon EMR avec Hive.

Module 6 : traiter des données sans serveur

- Le traitement, la transformation et l'analyse des données sans serveur.
- L'utilisation d'AWS Glue avec les charges de travail Amazon EMR.

Labs :

- Orchestrer le traitement des données dans Spark à l'aide d'AWS Step Functions.

Module 7 : assurer la sécurité et la surveillance des clusters

- Introduction à la sécurisation des clusters EMR.
- La surveillance et le dépannage des clusters Amazon EMR.

Labs :

- Cryptage côté client avec EMRFS.
- Examen de l'historique du cluster Apache Spark.

Module 8 : concevoir des solutions d'analyse de données par lots

- Les cas d'utilisation de l'analyse de données par lots.

Travaux pratiques :

- Conception d'un flux de travail d'analyse de données par lots.

Module 9 : développer des architectures de données modernes sur AWS

- Introduction aux architectures de données modernes.