

Building Modern Data Analytics Solutions on AWS

4jours / 28h

Objectifs pédagogiques

Building Batch Data Analytics Solutions on AWS

- 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 analytiques pour identifier et résoudre les problèmes
- Appliquer les meilleures pratiques de gestion des coûts

Building Data Analytics Solutions Using Amazon Redshift

- 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 d'entrepôt de données
- 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 analytiques pour identifier et résoudre les problèmes
- Appliquer les meilleures pratiques de gestion des coûts

Building Data Lakes on AWS

- Appliquer les méthodologies de lac de données dans la planification et la conception d'un lac de données
- Articuler les composants et les services requis pour créer un lac de données AWS
- Sécurisez un lac de données avec les autorisations appropriées
- Ingérer, stocker et transformer des données dans un lac de données
- Interroger, analyser et visualiser des données dans un lac de données

Building Streaming Data Analytics Solutions on AWS

- Comprendre les fonctionnalités et les avantages d'une architecture de données moderne.
- Découvrir comment les services de streaming AWS s'intègrent dans une architecture de données moderne.
- Concevoir et mettre en œuvre une solution d'analyse de données en continu
- Identifier et appliquer les techniques appropriées, telles que la compression, le partitionnement et le partitionnement, pour optimiser le stockage des données
- Sélectionner et déployez les options appropriées pour ingérer, transformer et stocker des données en temps réel et en temps quasi réel
- Choisir les flux, les clusters, les sujets, l'approche de mise à l'échelle 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
- Données de streaming sécurisées au repos et en transit

- Surveiller les charges de travail analytiques pour identifier et résoudre les problèmes
- Appliquer les meilleures pratiques de gestion des coûts

Public cible

- Architectes et opérateurs qui construisent et gèrent des pipelines d'analyse de données
- Ingénieurs et architectes de données
- Ingénieurs plateformes de données
- Ingénieurs d'entrepôt de données
- Développeurs qui souhaitent créer et gérer des applications en temps réel et des solutions d'analyse de données en continu
- Architectes de solutions

Prérequis

Building Batch Data Analytics Solutions on AWS

Les participants ayant au moins un an d'expérience dans la gestion de frameworks de données open source tels qu'Apache Spark ou Apache Hadoop bénéficieront de ce cours. Nous suggérons le cours [AWS Hadoop Fundamentals](#) pour ceux qui ont besoin d'un rappel sur Apache Hadoop.

Nous recommandons aux participants de ce cours d'avoir suivi les cours suivants :

- [AWS Technical Essentials](#) ou [Architecting sur AWS](#)
- Building Data Lakes on AWS ou [Getting Started with AWS Glue](#)

Building Data Analytics Solutions Using Amazon Redshift

Les participants ayant au moins un an d'expérience dans la gestion d'entrepôts de données bénéficieront de ce cours.

Nous recommandons aux participants de ce cours d'avoir suivi les formations :

- [AWS Cloud Practitioner Essentials](#) ou [AWS Technical Essentials](#)
- Building Data Lakes on AWS

Building Data Lakes on AWS

Nous recommandons aux participants de ce cours d'avoir :

- Suivi la formation [AWS Cloud Practitioner Essentials](#) ou [AWS Technical Essentials](#)
- Un an d'expérience dans la création de pipelines d'analyse de données ou avoir suivi le cours numérique [Data Analytics Fundamentals](#)

Building Streaming Data Analytics Solutions on AWS

Nous recommandons aux participants de ce cours d'avoir :

- Au moins un an d'expérience en analyse de données ou une expérience directe dans la création d'applications en temps réel ou de solutions d'analyse en continu. Nous suggérons le livre blanc [Streaming Data Solutions on AWS](#) pour ceux qui ont besoin d'un rappel sur les concepts de streaming.
- Suivi la formation [Architecting on AWS](#) ou [Data Analytics Fundamentals](#)
- Suivi la formation Building Data Lakes on AWS

Programme

Building Batch Data Analytics Solutions on AWS

Module A : 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

Module 1 : Présentation d'Amazon EMR

- Utilisation d'Amazon EMR dans les solutions d'analyse
- Architecture de cluster Amazon EMR
- Stratégies de gestion des coûts

Module 2 : 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

Module 3 : 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 Apache Spark sur Amazon EMR
- Concepts de Spark

- Transformation, traitement et analytique
- Utilisation de blocs-notes avec Amazon EMR
- Mise en pratique 1 : Analyse de données à faible latence à l'aide d'Apache Spark sur Amazon EMR

Module 4 : Traitement et analyse des données de lot avec Amazon EMR et Apache Hive

- Utilisation d'Amazon EMR avec Hive pour traiter les données par lots
- Transformation, traitement et analytique
- Introduction à Apache HBase sur Amazon EMR
- Mise en pratique 2 : traitement de données par lots à l'aide d'Amazon EMR avec Hive

Building Data Analytics Solutions Using Amazon Redshift

Module A : 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

Module 1 : Utilisation d'Amazon Redshift dans le pipeline d'analyse de données

- Pourquoi Amazon Redshift pour l'entreposage de données ?
- Présentation d'Amazon Redshift

Module 2 : Présentation d'Amazon Redshift

- Architecture Amazon Redshift
- Démo interactive 1 : Visite de la console Amazon Redshift
- Fonctionnalités Amazon Redshift
- Mise en pratique : Charger et interroger des données dans un cluster Amazon Redshift

Module 3 : Ingestion et stockage

- Ingestion
- Distribution et stockage des données
- Interroger des données dans Amazon Redshift
- Mise en pratique : Analyse de données à l'aide d'Amazon Redshift Spectrum

Module 4 : Traitement et optimisation des données

- Transformation des données

- Interrogation avancée
- La gestion des ressources
- Automatisation et optimisation
- Mise en pratique : Transformation et interrogation des données dans Amazon Redshift

Module 5 : Sécurité et surveillance des clusters Amazon Redshift

- Sécurisation du cluster Amazon Redshift
- Surveillance et dépannage des clusters Amazon Redshift

Module 6 : Conception de solutions d'analyse d'entrepôt de données

- Revue de cas d'utilisation de l'entrepôt de données
- Activité : Conception d'un workflow d'analyse d'entrepôt de données

Module B : Développement d'architectures de données modernes sur AWS

- Architectures de données modernes

Building Data Lakes on AWS

Module 1 : Introduction aux lacs de données

- Décrire la valeur des lacs de données
- Comparer les lacs de données et les entrepôts de données
- Décrire les composants d'un lac de données
- Reconnaître les architectures communes construites sur des lacs de données

Module 2 : Ingestion, catalogage et préparation des données

- Décrire la relation entre le stockage du lac de données et l'ingestion de données
- Décrire les robots AWS Glue et comment ils sont utilisés pour créer un catalogue de données
- Identifier le formatage, le partitionnement et la compression des données pour un stockage et une requête efficaces
- Mise en pratique : Configurer un lac de données simple

Module 3 : Traitement et analyse des données

- Reconnaître comment le traitement des données s'applique à un lac de données
- Utiliser AWS Glue pour traiter les données dans un lac de données

- Décrire comment utiliser Amazon Athena pour analyser les données dans un lac de données

Module 4 : Création d'un lac de données avec AWS Lake Formation

- Décrire les fonctionnalités et les avantages d'AWS Lake Formation
- Utiliser AWS Lake Formation pour créer un lac de données
- Comprendre le modèle de sécurité AWS Lake Formation
- Mise en pratique : Créer un lac de données à l'aide d'AWS Lake Formation

Module 5 : Configurations supplémentaires de Lake Formation

- Automatisez AWS Lake Formation à l'aide de plans et de flux de travail
- Appliquer des contrôles de sécurité et d'accès à AWS Lake Formation
- Faites correspondre les enregistrements avec AWS Lake Formation FindMatches
- Visualisez les données avec Amazon QuickSight
- Mise en pratique : Automatiser la création d'un lac de données à l'aide des plans AWS Lake Formation
- Mise en pratique : Visualisation des données à l'aide d'Amazon QuickSight

Module 6 : Architecture et révision des cours

- Vérification des connaissances après le cours
- Revue d'architecture
- Revue de cours

Building Streaming Data Analytics Solutions on AWS

Module A : 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

Module 1 : Utilisation des services de streaming dans le pipeline d'analyse de données

- L'importance de l'analyse des données en continu
- Le pipeline d'analyse de données en continu
- Concepts de streaming

Module 2 : Introduction aux services de diffusion en continu AWS

- Services de données en continu dans AWS

- Amazon Kinesis dans les solutions d'analyse
- Utilisation d'Amazon Kinesis Data Analytics
- Présentation d'Amazon MSK
- Présentation de Spark Streaming
- Mise en pratique : Configuration d'un pipeline de diffusion en continu avec Amazon Kinesis

Module 3 : Utilisation d'Amazon Kinesis pour l'analyse de données en temps réel

- Explorer Amazon Kinesis à l'aide d'une charge de travail clickstream
- Création de données Kinesis et de flux de diffusion
- Construire des producteurs de flux
- Construire des consommateurs de flux
- Création et déploiement d'applications Flink dans Kinesis Data Analytics
- Mise en pratique : analyses en continu avec Amazon Kinesis Data Analytics et Apache Flink

Module 4 : Sécurisation, surveillance et optimisation d'Amazon Kinesis

- Optimisez Amazon Kinesis pour obtenir des informations commerciales exploitables
- Bonnes pratiques de sécurité et de surveillance

Module 5 : Utilisation d'Amazon MSK dans les solutions d'analyse de données en continu

- Cas d'utilisation pour Amazon MSK
- Création de clusters MSK
- Ingestion de données dans Amazon MSK
- Transformation et traitement dans Amazon MSK
- Mise en pratique : Introduction au contrôle d'accès avec Amazon MSK

Module 6 : Sécurisation, surveillance et optimisation d'Amazon MSK

- Optimisation d'Amazon MSK
- Démonstration : Mise à l'échelle du stockage Amazon MSK
- Sécurité et surveillance
- Mise en pratique : pipeline de diffusion en continu Amazon MSK et déploiement d'applications

Module 7 : Conception de solutions d'analyse de données en continu

- Examen des cas d'utilisation
- Exercice de classe : Conception d'un flux de travail d'analyse de données en continu

Module B : Développement d'architectures de données modernes sur AWS

- Architectures de données modernes