

Building Batch Data Analytics Solutions on AWS – Formation intensive («AWSB05»)

Dans cette formation, vous apprendrez à créer des solutions d'analytique de données en lot avec Amazon EMR, Apache Spark et Apache Hadoop. Préparez-vous à la certification « AWS Certified Data Analytics » (DAS-C01).

Durée: 1 jour

Prix: 900.– excl. 8.1% TVA

Documents : Support de cours numérique officiel AWS

Contenu

Découvrez comment Amazon EMR intègre des projets open source comme Apache Hive, Hue et HBase ainsi qu'avec des services AWS comme AWS Glue et AWS Lake Formation. Cette formation aborde la collecte, l'ingestion, le catalogage, le stockage et le traitement de données dans le contexte de Spark et Hadoop. Vous apprendrez à utiliser des Notebooks EMR pour soutenir des charges de travail d'analyse et de machine learning. Vous apprendrez également comment intégrer les bonnes pratiques de sécurité, performance et gestion des coûts à Amazon EMR.

Contenu :

Module A : Aperçu de l'analytique de données et du pipeline de données

- Les cas d'utilisation de l'analytique de données
- Utiliser un pipeline de données pour l'analyse

Module 1 : Introduction à Amazon EMR

- Utiliser Amazon EMR dans des solutions d'analytique
- L'architecture en cluster d'Amazon EMR
- Démonstration interactive 1 : Lancer un cluster Amazon EMR
- Les stratégies de gestion des coûts

Module 2 : Le pipeline d'analytique de données avec Amazon EMR : Ingestion et stockage

- Optimisation du stockage avec Amazon EMR
- Techniques d'ingestion des données

Module 3 : L'analytique de données en lot de haute performance avec Apache Spark sur Amazon EMR

- Les cas d'utilisation d'Apache Spark sur Amazon EMR
- Pourquoi Apache Spark sur Amazon EMR
- Les concepts de Spark
- Démonstration interactive 2 : Se connecter à un cluster EMR et effectuer des commandes Scala avec le shell Spark
- Transformation, traitement et analyse
- Utiliser des notebooks avec Amazon EMR
- Exercice pratique 1 : L'analytique de données en lot en basse latence avec Apache Spark sur Amazon EMR

Module 4 : Traiter et analyser des données en lot avec Amazon EMR et Apache Hive

- Utiliser Amazon EMR avec Hive pour traiter des données en lot
- Transformation, traitement et analyse

- Exercice pratique 2 : Traitement de données en lot avec Amazon EMR et Hive
- Introduction à Apache HBase sur Amazon EMR

Module 5 : Traitement des données sans serveur

- Transformation, traitement et analyse des données sans serveur
- Utiliser AWS Glue avec des charges de travail Amazon EMR
- Exercice pratique 3 : Orchestrer le traitement des données dans Spark en utilisant AWS Step Functions

Module 6 : Sécuriser et surveiller les clusters Amazon EMR

- Sécuriser les clusters EMR
- Démonstration interactive 3 : Cryptage côté client avec EMRFS
- Surveiller et résoudre des problèmes de clusters Amazon EMR
- Démonstration : Examiner l'historique du cluster Apache Spark

Module 7 : Créer des solutions d'analytique de données en lot

- Les cas d'utilisation de l'analytique de données en lot
- Activité : Créer un flux d'analytique de données en lot

Module B : Développer des architectures modernes de données sur AWS

- Les architectures modernes de données

Objectifs

- Comparer les fonctionnalités et les avantages des entrepôts de données, des lacs de données et des architectures de données modernes
- Créer et mettre en œuvre une solution d'analytique des données en lot
- Identifier et appliquer des 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 l'instance, les types de nœud, le cluster, la mise à l'échelle automatique et le type de réseau pour un cas commercial en particulier
- Comprendre l'impact du stockage et du traitement des données sur les mécanismes d'analyse et de visualisation de données nécessaires pour tirer des visions commerciales précieuses
- Sécuriser les données au repos et en transit
- Surveiller les charges de travail d'analyse pour identifier et résoudre des problèmes
- Appliquer les bonnes pratiques de gestion des coûts

Méthodologie & Didactique

Cette formation comprend des présentations, des démos interactives, des exercices pratiques, des discussions et des exercices en classe.

Public cible

Cette formation s'adresse aux personnes actives dans les domaines suivants :

- Data Engineer

Et en particuliers aux :

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

- Les participantes et participants qui ont au minimum une année d'expérience en gestion de frameworks de données open source comme Apache Spark ou Apache Hadoop auront plus de facilité dans cette formation.
- Nous conseillons aux personnes qui ont besoin de mettre à jour leurs connaissances d'Apache Hadoop de suivre au préalable le cours [AWS Hadoop Fundamentals](#).
- Les participantes et participants doivent avoir au préalable suivi les formations suivantes ou s'assurer de posséder des connaissances équivalentes :
 - [AWS Technical Essentials](#) ou [Architecting on AWS](#)
 - [Building Data Lakes on AWS](#) ou [Getting started with AWS Glue](#)
- [AWS Technical Essentials – Formation intensive \(«AWSE01»\)](#)
- [Building Data Lakes on AWS – Formation intensive \(«AWSB04»\)](#)
- [Architecting on AWS – Formation intensive \(«AWSA01»\)](#)

Certification

Cette formation marque une étape essentielle vers la certification « AWS Certified Data Analytics – Specialty » pour laquelle il faut passer l'examen « AWS Certified Data Analytics – Specialty » (DAS-C01). Afin d'être pleinement préparé à l'examen de certification, nous vous conseillons de suivre également les formations suivantes :

- [Building Data Lakes on AWS](#)
- [Building Data Analytics Solutions Using Amazon Redshift](#)

L'examen, dont l'inscription se fait directement auprès d'[AWS](#), dure 180 minutes et coûte USD 300.

Informations complémentaires

Matériel

- **Support de cours** : Environ une semaine avant le début de votre formation, vous recevrez vos données d'accès (code voucher) aux supports de cours électroniques par e-mail directement de l'adresse noreply@gilmore.ca. Tous les supports de cours sont hébergés sur la plateforme evantage.gilmoreglobal.com. Veuillez suivre les instructions contenues dans l'e-mail et créer un compte avec votre adresse e-mail professionnelle (si vous n'avez pas encore de compte) pour accéder aux supports de cours.
- **Labs** : Tous les exercices des formations techniques sont hébergés sur la plateforme d'exercice officielle d'AWS digicomp.qwiklabs.com. Au début de leur formation, les participantes et participants devront créer leur propre compte sur digicomp.qwiklabs.com avec leur adresse e-mail professionnelle pour avoir accès aux labs officiels d'AWS et pouvoir effectuer les exercices pratiques.
- **Plateforme de formation** : Si vous participez à une formation virtuelle, vous recevrez l'accès à la plateforme de formation de Digicomp un jour avant le début de votre formation.
- Pour accéder aux supports de cours et exercices pendant le cours, pensez à les télécharger et à apporter votre propre tablette ou ordinateur portable.

Formations complémentaires

- [Building Streaming Data Analytics Solutions on AWS – Formation intensive \(«AWSB08»\)](#)

Avez-vous une question ou souhaitez-vous organiser un cours en entreprise ?

Nous vous conseillons volontiers au +41 22 738 80 80 ou
romandie@digicomp.ch. Retrouvez toutes les informations détaillées
concernant les dates sur [www.digicomp.ch/formations-it-
providers/amazon-web-services-aws/aws-data-engineer/cours-building-
batch-data-analytics-solutions-on-aws-formation-intensive](http://www.digicomp.ch/formations-it-providers/amazon-web-services-aws/aws-data-engineer/cours-building-batch-data-analytics-solutions-on-aws-formation-intensive)

