

AWS : Big Data sur Amazon Web Services **Formation officielle AWS "Big Data on AWS"**

Pour répondre aux besoins des entreprises, Amazon Web Services déploie un portefeuille de services de cloud computing, entièrement intégrés et couvrant l'ensemble du cycle de gestion des données : collecte, stockage, traitement / analyse, consommation / visualisation. Durant ces trois jours de formation pratique, vous découvrirez les solutions AWS de Big Data tels qu'Amazon EMR, Amazon Redshift, Amazon Kinesis et le reste de la plate-forme. Vous apprendrez également à concevoir des environnements de Big Data, à utiliser Amazon DynamoDB, Amazon Redshift, Amazon Quicksight, Amazon Athena et Amazon Kinesis et à appliquer au mieux les bonnes pratiques afin de concevoir des environnements de Big Data sécurisés et économiques.

Objectifs

- Intégrer une solution AWS dans un écosystème de Big Data
- Choisir les options de stockage de données AWS appropriées
- Comprendre et gérer les coûts et la sécurité pour une solution de Big Data
- Sécuriser une solution de Big Data
- Identifier les options pour l'importation, le transfert et la compression de données
- Utiliser Apache Hadoop dans le cadre d'Amazon EMR
- Identifier les composants d'un cluster Amazon EMR, lancer et configurer un cluster Amazon EMR
- Utiliser les frameworks compatibles avec Amazon EMR, notamment Hive, Pig et Streaming
- Utiliser Hue pour améliorer la facilité d'utilisation d'Amazon EMR
- Utiliser les analyses en mémoire avec Spark sur Amazon EMR
- Identifier les avantages de l'utilisation d'Amazon Kinesis pour le traitement de Big Data
- Utiliser Amazon Redshift pour stocker et analyser des données de manière efficace
- Exploiter Amazon Athena pour les analyses par requêtes ad hoc
- Utiliser un logiciel de visualisation pour représenter les données et les requêtes avec AWS QuickSight
- Passer dans les conditions optimales la certification internationale « Big Data on AWS »

Prérequis

- Des connaissances en langue anglaise et il est souhaitable d'avoir 6 mois d'expérience d'internet/web
- Connaissance basique des technologies de Big Data, notamment Apache Hadoop, MapReduce, HDFS et les requêtes SQL. Connaissance des concepts d'entrepôt de données, de système de base de données relationnelle et de conception de base de données.
- Avoir suivi le cours en ligne gratuit « Notions fondamentales sur le Big Data » ou de posséder un niveau d'expérience équivalent. Avoir suivi le cours « Notions techniques de base AWS » ou posséder un niveau d'expérience équivalent.
- Se munir de sa pièce d'identité pour le passage de l'examen

Public

- Architectes de solutions Big Data en charge de la conception et de l'implémentation
- Analystes de données et utilisateurs de solutions Big Data souhaitant se former sur les services et les types d'architectures soutenant les solutions de Big Data sur AWS

Durée

- 3 journées : 9h00-17h00 - 21 heures (Travail personnel non inclus)
- Intégrant 1 pause matin et après-midi et pauses déjeuners

Déroulement pédagogique

Projection du cours. La formation est en langue française, le support et l'examen sont en langue anglaise. Le cours est théorique à 50% et 50% pratique avec des discussions, partage d'expérience et étude de cas. Préparation à l'examen avec des examens QCM à blanc. Cette formation permet de préparer l'examen associé au titre de certification « Big Data on AWS ». Une attestation de formation est délivrée en fin de formation.

Déroulement de l'examen

- QCM en langue anglaise en ligne dans un centre d'examen agréé
- Obtention du certificat à partir de 720 points
- Durée : 130 minutes

Programme

Ingestion et transfert de Big Data - Streaming de Big Data et Amazon Kinesis

Atelier 1 : "Utilisation d'Amazon Kinesis pour diffuser et analyser les données de journal d'un serveur Apache"

Solutions de stockage de Big Data - Traitement et analyses de Big Data

Atelier 2 : "Utilisation d'Amazon Athena pour interroger des données de journal depuis Amazon S3"

Apache Hadoop et Amazon EMR

Atelier 3 : "Stockage et interrogation de données sur Amazon DynamoDB"

Utilisation d'Amazon EMR - Frameworks de programmation Hadoop

Atelier 4 : "Traitement des journaux du serveur avec Hive sur Amazon EMR" : Interfaces Web sur Amazon EMR

Atelier 5 : "Exécution de scripts Pig dans Hue sur Amazon EMR" : Apache Spark sur Amazon EMR

Atelier 6 : "Traitement des données de taxis new-yorkais avec Spark sur Amazon EMR"

Amazon Redshift et les technologies Big Data - Visualisation et orchestration de Big Data

Atelier 7 : "Utilisation de TIBCO Spotfire pour visualiser les données"

Gestion des coûts du Big Data - Sécurisation de vos déploiements Amazon - Modèles de conception de Big Data

Révisions et passage de l'examen final en fin de formation (voir le déroulement de l'examen)