

## **AWS : Big Data sur Amazon Web Services**

### **Formation officielle Big Data on AWS**

#### **DESCRIPTION**

Le Big Data pose aux entreprises de nombreux défis en matière de gestion de leurs données. En raison des volumétries à stocker puis à analyser, de leur variété et des délais toujours plus courts, imposés par le marché, ces défis ne peuvent être relevés par les bases de données traditionnelles. Une stratégie efficace en matière de Big Data peut aider les organisations à réduire les coûts et à améliorer leur efficacité opérationnelle, en migrant les charges de travail existantes les plus lourdes vers des technologies Big Data et, en déployant de nouvelles applications pour profiter de nouvelles opportunités.

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, vous découvrirez les solutions AWS de Big Data telles que Amazon EMR, Amazon Redshift, Amazon Kinesis et le reste de la plateforme. Vous apprendrez également à concevoir des environnements de Big Data, à utiliser Amazon DynamoDB, Amazon Redshift, Amazon Quicksight, Amazon Athena et Amazon Kinesis, mais aussi à appliquer au mieux les bonnes pratiques afin de concevoir des environnements de Big Data sécurisés et économiques.

Cette formation peut être incluse dans le parcours de préparation pour la certification Aws certified Data Analytics Speciality.

#### **OBJECTIFS PEDAGOGIQUES**

- Intégrer une solution AWS dans un écosystème de Big Data
- Choisir les options de stockage de données AWS appropriées
- Identifier et gérer les coûts et la sécurité pour une solution de Big Data
- Sécuriser une solution de Big Data
- Découvrir les options pour l'importation, le transfert et la compression de données
- Utiliser Apache Hadoop dans le cadre d'Amazon EMR

#### **Stage pratique**

Amazon Web Services

Code :

**AWS03**

Durée :

**3 jour(s) (21,00 heures)**

Exposés : **50 %**

Cas pratiques : **40 %**

Echanges d'expérience : **10 %**

#### **Inter-entreprises :**

Prochaines sessions disponibles [sur notre site web](#).

Tarif : 2 350,00 € HT / participant

#### **Intra-entreprise :**

Tarifs et dates sur demande.

- Identifier les composants d'un cluster Amazon EMR
- Lancer et configurer un cluster Amazon EMR
- Utiliser les frameworks de programmation courants 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 en temps quasi réel
- 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 Amazon QuickSight

**PUBLIC CIBLE**

- 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

**PRE-REQUIS**

- Connaissance basique des technologies de Big Data, notamment Apache Hadoop, MapReduce, HDFS et les requêtes SQL/NoSQL.
- 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. Connaissance des concepts d'entrepôt de données, de système de base de données relationnelle et de conception de base de données.

**METHODE PEDAGOGIQUE**

Formation assurée par un instructeur certifié AWS

Ce cours vous permet de tester de nouvelles compétences et d'appliquer vos connaissances à votre environnement de travail grâce à différents exercices pratiques.

**PROFIL DES INTERVENANTS**

Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

## MODALITÉS D'ÉVALUATION ET FORMALISATION À L'ISSUE DE LA FORMATION

L'évaluation des acquis se fait tout au long de la session au travers des ateliers et des mises en pratique. Une évaluation à chaud sur la satisfaction des stagiaires est réalisée systématiquement en fin de session et une attestation de formation est délivrée aux participants mentionnant les objectifs de la formation, la nature, le programme et la durée de l'action de formation ainsi que la formalisation des acquis.

## PROGRAMME PEDAGOGIQUE DETAILLE

### Jour 1

#### 1 PRÉSENTATION DES TECHNOLOGIES BIG DATA

- 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"

### Jour 2

- 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"

### Jour 3

- 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

