

# AWS – Big Data on Amazon Web Services

## Description

Dans ce cours vous découvrirez les solutions Big Data basées sur le cloud comme Amazon EMR, Amazon Redshift, Amazon Kinesis et le reste de la plate-forme Big Data AWS. Apprenez à utiliser Amazon EMR pour traiter des données à l'aide du vaste écosystème d'outils Hadoop comme Hive et Hue, créez des environnements Big Data, travaillez avec Amazon DynamoDB, Amazon Redshift, Amazon QuickSight, Amazon Athena et Amazon Kinesis, et concevez des environnements Big Data pour la sécurité et rentabilité.

### Prix de l'inscription en Présentiel (CHF)

2500

### Prix de l'inscription en Virtuel (CHF)

2500

### Contenu du cours

#### Module 1: Présentation du Big Data

- Qu'est-ce que le Big Data
- Le pipeline Big Data
- Principes architecturaux Big Data

#### Module 2: Ingestion et transfert de Big Data

- Vue d'ensemble: ingestion de données
- Transfert de données

#### Module 3: Streaming Big Data et Amazon Kinesis

- Traitement de flux de Big Data
- Amazon Kinesis
- Amazon Kinesis Data Firehose
- Flux vidéo Amazon Kinesis
- Analyse de données Amazon Kinesis

#### Module 4: Solutions de stockage Big Data

- Options de stockage de données AWS
- Concepts de solutions de stockage
- Facteurs dans le choix d'un magasin de données

#### Module 5: Traitement et analyse du Big Data

- Traitement et analyse de Big Data
- Amazon Athéna

#### Module 6: Apache Hadoop et Amazon EMR

- Introduction à Amazon EMR et Apache Hadoop
- Bonnes pratiques pour l'ingestion de données
- Amazon EMR

- Architecture Amazon EMR

### **Module 7: Utilisation d'Amazon EMR**

- Développer et exécuter votre application
- Lancement de votre cluster
- Gestion de la sortie de vos travaux terminés

### **Module 8: Cadres de programmation Hadoop**

- Cadres Hadoop
- Autres frameworks à utiliser sur Amazon EMR

### **Module 9: Interfaces Web sur Amazon EMR**

- Hue sur Amazon EMR
- Suivi de votre cluster

### **Module 10: Apache Spark sur Amazon EMR**

- Apache Spark
- Utilisation de Spark

### **Module 11: Utilisation d'AWS Glue pour automatiser les charges de travail ETL**

- Qu'est-ce qu'AWS Glue?
- AWS Glue: orchestration des tâches

### **Module 12: Amazon Redshift et le Big Data**

- Entrepôts de données vs bases de données traditionnelles
- Amazon Redshift
- Architecture Amazon Redshift

### **Module 13: Sécuriser vos déploiements Amazon**

- Sécurisation de vos déploiements Amazon
- Présentation de la sécurité Amazon EMR
- Présentation d'AWS Identity and Access Management (IAM)
- Sécurisation des données
- Présentation de la sécurité d'Amazon Kinesis
- Présentation de la sécurité Amazon DynamoDB
- Présentation de la sécurité Amazon Redshift

### **Module 14: Gérer les coûts du Big Data**

- Considérations relatives au coût total pour Amazon EMR
- Modèles de tarification Amazon EC2
- Modèles de tarification Amazon Kinesis
- Considérations de coût pour Amazon DynamoDB
- Considérations de coût et modèles de tarification pour Amazon Redshift
- Optimiser les coûts avec AWS

### **Module 15: Visualiser et orchestrer le Big Data**

- Visualisation du big data
- Amazon QuickSight
- Orchestrer un workflow big data

## Module 16: Modèles de conception Big Data

- Architectures communes

## Module 17: Synthèse du cours

- Et après?

## Lab / Exercices

- Laboratoires officiels AWS

## Documentation

- Support de cours numérique inclus

## Examen

- Ce cours prépare à la certification AWS Certified Data Analytics - Specialty. Si vous souhaitez passer cet examen, merci de contacter notre secrétariat qui vous communiquera son prix et s'occupera de toutes les démarches administratives nécessaires pour vous

## Profils des participants

- Personnes responsables de la conception et de la mise en oeuvre de solutions Big Data, à savoir Solutions
- Architectes et administrateurs SysOps
- Scientifiques et analystes de données intéressés à en savoir plus sur les solutions Big Data sur AWS

## Connaissances Préalables

- Connaissance de base des technologies Big Data, y compris Apache Hadoop, HDFS et les requêtes SQL / NoSQL
- Formation numérique gratuite sur les principes de base de l'analyse des données ou expérience équivalente
- Connaissance pratique des principaux services AWS et de la mise en oeuvre du cloud public
- Compréhension de base de l'entreposage de données, des systèmes de bases de données relationnelles et de la conception de bases de données
- Avoir suivi ou maîtriser les notions incluses dans le cours suivant : [AWS Technical Essentials](#)

## Objectifs

- Utiliser Apache Hadoop avec Amazon EMR
- Lancer et configurer un cluster Amazon EMR
- Utilisez des cadres de programmation courants pour Amazon EMR, notamment Hive, Pig et Streaming
- Utilisez Hue pour améliorer la facilité d'utilisation d'Amazon EMR
- Utiliser des analyses en mémoire avec Spark sur Amazon EMR
- Comprendre comment des services comme AWS Glue, Amazon Kinesis, Amazon Redshift, Amazon Athena et Amazon QuickSight peuvent être utilisés avec des charges de travail Big Data

## Niveau

---

Intermédiaire

**Durée (Nombre de Jours)**

3

**Reference**

AWS-11