

# Building Streaming Data Analytics Solutions on AWS

7h

## Objectifs pédagogiques

- 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

- Ingénieurs et architectes de données

- Développeurs qui souhaitent créer et gérer des applications en temps réel et des solutions d'analyse de données en continu

## Prérequis

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

### **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