



Kafka Streams & Confluent KSQL

Durée 3 jour(s) (KAFKA-KSQL)

Développer des applications de stream-processing au dessus de Kafka

Formation officielle



Description

Au cours de ce cours pratique, vous apprendrez à utiliser Confluent KSQL et la librairie Apache Kafka Streams dans le but de développer des applications de streaming. Vous apprendrez le rôle du streaming dans la construction le pipeline de distribution de données moderne, discuterez des concepts architecturaux et des différents composants de la plateforme Confluent.

Objectifs

- Vous apprendrez à développer des applications pour transformer, enrichir, filtrer et agréger des flux de données en temps réel avec Confluent KSQL et la librairie Apache Kafka Streams.
- Vous apprendrez comment tester, surveiller, sécuriser et mettre à l'échelle ces applications de streaming.
- Vous apprendrez comment Kafka Streams et KSQL s'intègrent à la plateforme Confluent et les solutions Kafka Connect, Schema Registry, REST Proxy et Confluent Control Center (C3).

Public

Cette formation est conçu pour les développeurs d'applications, de chaînes ETL (extraire, transformer et charger), les ingénieurs DevOps et Data qui ont besoin d'interagir avec des clusters Kafka comme source de données temps réel pour transformer, enrichir et joindre des flux pour découvrir des anomalies, analyser des comportements ou surveiller des systèmes complexes.

Prérequis

- Les participants doivent être familiarisés avec le développement en Java et avoir des notions avec SQL.
- Être familiarisé avec les concepts associés à Apache Kafka, la production et la consommation de messages.
- Il est recommandé de suivre au préalable la formation Confluent Kafka Developer.

Répartition

50% Théorie, 50% Pratique

Programme

Fondamentaux

- Les logs
- La réplication des logs
- Les concepts de Topics, Partitions et de Segments
- Kafka Streams
- Dualité entre les notions de Stream et Table• Stream-Table Dualism
- Les jobs de Stream Processing

Cas d'utilisation de KSQL

- Pourquoi KSQL
- Exemples de cas d'utilisations
- KSQL et licensing

Vue d'ensemble de KSQL et écosystème

- KSQL & Kafka = Easy
- Utilisation interactive de KSQL
- Architecture de KSQL
- KSQL CLI
- Les modes serveur de KSQL
- KSQL et Confluent Control Center (C3)

Utiliser KSQL

- Kafka Streams et Tables
- Message Kafka et Formats de données
- Manipulation et agrégation de données
- Fonctions définies par l'utilisateur (UDF)
- Enrichissement des données et jointures
- Agrégations fenêtrées (Windowed opération)
- Métrique et observabilité
- Tests et Monitoring
- Astuces, pièges et limitations

Déploiement et utilisation de KSQL

- Les bonnes pratiques et patterns
- Sécuriser KSQL
- Élasticité et évolutivité
- Tolérance aux pannes
- Healthcheck

Architecture de Kafka Streams

- Motivation et évolution
- Les caractéristiques
- Liberté de choix

Kafka Streams Application Anatomie

- Anatomie d'une application Streams
- Configuration d'une application Streams
- Topologie d'une application Streams

Kafka Streams DSL

- Opérations Stateless et Stateful
- Kafka Streams DSL
- Les opération de type fenêtre glissante
- L'API processeur

Test des applications Kafka Streams

- Les catégories de test
- Tests unitaires
- Tests d'intégration avec Test Driver et Kafka embarqué

Surveillance des applications Kafka Streams

- Utilisation de la surveillance basée sur JMX
- Utilisation de Confluent Control Center pour la surveillance

Sécurisation des applications Kafka Streams

- Pourquoi la sécurité est nécessaire
- Aperçu de la sécurité
- Fonctions de sécurité côté client
- Les ACLs requis
- Chiffrement des données en transit

Dimensionnement et mise à l'échelle des applications Kafka Streams

- Scalabilité et élasticité
- Définir le nombre d'instances par application
- Gestion de la mémoire
- Dimensionnement et positionnement des tâches
- Stateless vs Stateful
- Troubleshooting