

Formation Apach Flink

A l'issue de cette formation Flink, les participants seront en mesure d'utiliser ce framework Big data pour réaliser des traitements sur les données de leur entreprise. Ils maîtriseront les APIs DataStream, Table, et sauront utiliser Flink SQL sur des flux de données. La formation se termine par l'appropriation des bonnes pratiques pour le déploiement de Flink.

Durée

3 jours

Objectifs pédagogiques

- S'approprier les concepts fondamentaux du framework Apache Flink
- Utiliser les APIs Flink pour développer des applications Big Data
- Réaliser des traitements de données distribués en temps réel
- S'approprier les principes de base pour mettre en œuvre les bibliothèques Flink
- Connaître les possibilités d'intégration avec d'autres outils Big Data
- Maîtriser les meilleures pratiques pour déployer un cluster Flink multi-nœuds

Public

Ingénieurs Big Data, développeurs, architectes, data analysts

Prérequis

De bonnes connaissances en Java

Programme de formation

Introduction à la formation Flink

Présentation générale de la formation
Objectifs pédagogiques
Introduction à Flink : intérêt, évolution, différences avec Spark...
Ecosystème
Batch processing vs Stream processing
Flux de données parallèles
Les notions State, Time et Snapshots

Configuration d'un premier projet Flink

Maven ou Gradle
Dépendances
Anatomie d'un programme Flink

Transformer des données avec l'API DataStream

Introduction à l'API DataStream
Types de données et sérialisation, Java/Scala
Environnement d'exécution et paramètres
Data sources
Transformations stateless, stateful
Opérateurs de transformations : Map, FlatMap, Filter...
Data sinks
Debug

Streaming Analytics : event time et watermarks

Event time, ingestion time et processing time
Travailler avec les watermarks
Utiliser les Windows pour traiter des flux de données : Tumbling, Sliding, Session et Global

L'API Table et Flink SQL

Types de données

Créer un TableEnvironment

Créer et requêter des tables

Les opérateurs à connaître

Configurer et exécuter un programme Flink SQL

Utiliser les fonctions pour transformer les données

Optimisation

Connecteurs

Vue d'ensemble des possibilités d'interfaçage avec d'autres systèmes Big Data (Kafka, Cassandra, Elasticsearch, RabbitMQ...)

Librairies

Vue d'ensemble

FlinkCEP (Complex Event Processing library)

FlinkML : la librairie Apache Flink pour le Machine Learning

Gelly, l'API Flink Graph

Déploiement d'un cluster Flink

Vue d'ensemble et architecture

Les différents modes de déploiement

Configurations

Prêts pour la mise en production ? Checklist

Déployer un cluster multinoeuds

Moyens et méthodes pédagogiques

- La formation alterne entre présentations des concepts théoriques et mises en application à travers d'ateliers et exercices pratiques.
- Les participants bénéficient des retours d'expérience terrains du formateur ou de la formatrice
- Un support de cours numérique est fourni aux stagiaires

Modalités d'évaluation

- **En amont de la session de formation**, un questionnaire d'auto-positionnement est remis aux participants, afin qu'ils situent leurs connaissances et compétences déjà acquises par rapport au thème de la formation.
- **En cours de formation**, l'évaluation se fait sous forme d'ateliers, exercices et travaux pratiques de validation, de retour d'observation et/ou de partage d'expérience.
- **En fin de session**, le formateur évalue les compétences et connaissances acquises par les apprenants grâce à un questionnaire reprenant les mêmes éléments que l'auto-positionnement, permettant ainsi une analyse détaillée de leur progression.