

Flink, développer des applications pour le Big Data

Cours Pratique de 3 jours - 21h
Réf : FKB - Prix 2025 : 2 330 HT

Apache Flink est un framework big data récent. Il simplifie les traitements de flux gros débit temps réels comme les traitements batch sur d'énormes quantités de données (sur Hadoop HDFS, sur Amazon S3, sur MongoDB...). Ce cours vous permet d'installer Flink et de réaliser, en Java, des traitements variés de big data.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

- Maîtriser les concepts fondamentaux de Flink
- Développer des applications à l'aide de DataSet et DataStream API
- Traitement de données distribué avec Flink et Hadoop
- Exploiter des données avec Table API
- Avoir une première approche du machine learning

TRAVAUX PRATIQUES

Mise en pratique des notions vues en cours à l'aide du langage Java.

LE PROGRAMME

dernière mise à jour : 07/2024

1) Introduction à Apache Flink

- Historique du framework.
- Les différentes versions de Flink.
- Comparaison avec l'environnement Apache Hadoop et Apache Spark.
- Les différents modules de Flink.

Travaux pratiques : Installation et configuration de Flink. Exécution d'un premier exemple avec le comptage de mots.

2) Traitement de données à l'aide de l'API DataStream

- Environnement d'exécution et sources de données.
- Les transformations : Map, FlatMap, Filter, KeyBy, Reduce...
- Les opérations sur des flux multiples : Union, Cogroup, Connect, Join, Iterate...
- Les opérations sur Windows : Global, Tumbling, Sliding, Session...
- Le partitionnement physique personnalisé, aléatoire, rééquilibrage et redimensionnement.
- Les DataSink et les connecteurs : Kafka, X (anciennement Twitter), ElasticSearch...

Travaux pratiques : Consommation et manipulation de différents flux des données.

3) Traitement de données à l'aide de l'API Batch

- Les différents types de sources de données.
- Les transformations et agrégations.
- L'écriture des données.
- Les DataSink et les connecteurs : HDFS, S3, Avro, MongoDB.

Travaux pratiques : Manipulation de DataSets à partir de plusieurs sources de données.

PARTICIPANTS

Développeurs, architectes.

PRÉREQUIS

Bonnes connaissances du langage Java.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse psh-accueil@orsys.fr.

4) Traitement de données à l'aide de l'API Table

- Enregistrement et lecture des tables enregistrées.
- Les opérateurs : sélection, filtre, jointure, orderBy...
- Utiliser du SQL sur le flux de données.
- Traitement des événements complexes.

Travaux pratiques : Mise en place d'une analyse avec SQL sur un flux de données.

5) API Flink Graph - Gelly

- Qu'est-ce qu'un graphe ?
- Les différentes opérations.
- Créer des graphes.
- Transformations de graphes.
- Présentation de différents algorithmes.

Travaux pratiques : Manipulation de l'API travers différents exemples.

6) Déploiement de Flink

- Flink sur YARN Configurations.
- Lancement et arrêt d'un cluster.
- Soumettre un travail à Flink.
- Flink sur Google Cloud.
- Flink sur AWS.

Travaux pratiques : Configurer un cluster multinoeuds et déployer une application.

LES DATES

CLASSE À DISTANCE

2025 : 10 déc.