+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Mis à jour le 05/02/2025

S'inscrire

Formation Spark Streaming

3 jours (21 heures)

Présentation

Spark est un framework pour effectuer des calculs distribués sur un cluster d'ordinateurs. Cette formation présente la toute nouvelle version 3.5.4 apportant un lot considérable de nouveautés ainsi qu'une amélioration impressionnante des performances !

Apache Spark peut traiter rapidement une large quantité de données à grande échelle. Depuis peu, cet outil est devenu l'un des meilleurs frameworks de calcul distribué au monde. Il a l'avantage d'intégrer différents langages de programmation tel que Java, Scala, Python ou encore R.

Notre formation présente les concepts avancés de Spark Streaming, de son intégration avec Kafka, mais aussi toutes les bonnes pratiques pour réussir son déploiement en production. Les travaux pratiques sont réalisés en Scala (ou bien Python en option).

Objectifs

- Manipuler des volumes importants de données en utilisant les bonnes pratiques de Spark Streaming
- Comprendre les concepts avancés de la nouvelle API Spark Streaming v4
- Intégrer et faire cohabiter Kafka avec Spark Streaming
- Être capable d'utiliser Spark Streaming en production

Public visé

- Développeurs
- Data Engineer
- Architectes
- Administrateurs systèmes
- DevOps

Pré-requis

- Avoir idéalement suivi nos formations Spark ML ou Spark Tuning Avancé
- Connaissances de base d'un système Unix
- Connaissance de Scala, Git & Kafka

Programme de la formation Spark Streaming

Jour 1

Introduction à Spark (dans un context de streaming)

- Architecture de Spark
- Fonctionnement interne (Stage, Task, Scheduler …)
- Batch vs Stream
- Le modèle microbatch
- API DStreams avec Scala

Structured Streaming

- Introduction à l'API Structured Streaming.
- API source
- API Sink
- API fonctionnelle
- SQL streaming
- Streaming des sources Json, Csv, Paquet
- Calculer des agrégats en streaming

Jour 2

Introduction à apache Kafka

- Fonctionnement interne (Topic, partition, Offset …)
- Producer
- Consumers
- Partitioning
- Commit des offsets

Intégration Spark streaming avec Kafka

- Streaming en Source et en Sink
- Calculer des agrégats en temps réel
- Jointure Stream-static et Stream-Stream
- Watermarks
- Windowing (tumbling, sliding, reduce…)

Jour 3

Streaming avec état (Stateful Streaming)

- State store
- Les opérateurs GroupState
- Les timeouts

Spark streaming en production

- State checkpointing et fault-tolerance.
- Monitoring via Spark-UI
- Tuning

Gestion des schémas avec Avro (Optionnelle + 1 jour sur demande)

Sociétés concernées

Cette formation s’adresse à la fois aux particuliers ainsi qu’aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d’auto-évaluation nous permettant d’apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d’anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l’acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.