

Mis à jour le 01/10/2024

S'inscrire

Formation Apache Avro

2 jours (14 heures)

Présentation

Vous cherchez une solution fiable et performante pour sérialiser, stocker et échanger des données à grande échelle ? Notre formation Apache Avro vous propose une introduction complète à ce format de sérialisation open source, optimisé pour le traitement de données massives dans des environnements distribués.

[Apache Avro](#) est conçu pour permettre la sérialisation efficace des données binaires tout en facilitant leur interchangeabilité entre systèmes et langages, ce qui en fait un outil idéal pour les Data Engineers et Développeurs Big Data.

Son schéma flexible permet d'accompagner l'évolution des structures de données sans compromettre la compatibilité avec les systèmes existants.

Durant cette formation, vous apprendrez à créer et gérer des schémas Avro, à sérialiser et désérialiser des données de manière optimale, et à intégrer Avro dans vos pipelines de traitement de données distribués.

Vous découvrirez également comment utiliser Apache Avro avec d'autres outils de l'écosystème Hadoop et Kafka pour traiter des données massives en temps réel ou en mode batch.

Cette formation vous permettra de développer des compétences essentielles pour la gestion des données à grande échelle, tout en vous familiarisant avec les meilleures pratiques d'optimisation et d'évolution des schémas dans des environnements Big Data.

Comme pour toutes nos formations, elle vous sera présentée avec les [dernières ressources disponibles](#).

Objectifs

- Comprendre les principes de base et l'architecture d'Apache Avro
- Savoir structurer et valider des schémas Avro pour la sérialisation des données
- Maîtriser les processus de sérialisation et de désérialisation avec Avro
- Intégrer Apache Avro dans des systèmes distribués comme Hadoop, Kafka, et Spark
- Gérer l'évolution des schémas tout en garantissant la compatibilité avec les systèmes existants
- Optimiser les performances des pipelines de traitement de données avec Avro

Public visé

- Data Engineers
- Data Scientists
- Développeurs Big Data
- Architectes de données

Pré-requis

- Connaissance des concepts de base du traitement de données massives
- Expérience avec un langage de programmation tel que Java ou Python
- Compréhension des concepts de bases de données et de systèmes distribués
- Familiarité avec les technologies Big Data comme Hadoop, Spark, ou Kafka est un plus

PROGRAMME DE NOTRE FORMATION APACHE AVRO

Introduction à Apache Avro

- Présentation de la sérialisation des données : définition et importance
- Historique et conception d'Apache Avro dans l'écosystème Hadoop
- Comparaison avec d'autres formats de sérialisation (JSON, Protobuf, Thrift)
- Structure d'un fichier Avro : schéma et données
- Principes d'auto-description : le schéma stocké avec les données
- Avantages d'Apache Avro pour le traitement de données à grande échelle

Le modèle de schéma Avro

- Comprendre la syntaxe du schéma Avro (JSON)
- Types de données primitifs (int, long, float, etc.) et complexes (records, arrays, maps)
- Validation des schémas : assurer l'intégrité des données
- Gestion de l'évolution des schémas (ajout/suppression de champs)
- Compatibilité des schémas (backward, forward, full compatibility)
- Exemples pratiques de schémas Avro dans des applications réelles

Sérialisation et Désérialisation avec Apache Avro

- Processus de sérialisation : convertir des objets en format Avro binaire ou JSON

- Désérialisation : transformer les données Avro en objets lisibles
- Avantages du format binaire pour la performance et la réduction de la taille des fichiers
- Utilisation de bibliothèques dans différents langages (Java, Python, etc.)
- Sérialisation avec et sans schéma : différences et cas d'usage
- Optimisation des performances lors de la sérialisation des données massives

Apache Avro dans les systèmes distribués

- Utilisation d'Avro avec Apache Hadoop : stockage et traitement des fichiers Avro dans HDFS
- Intégration avec Apache Kafka pour les logs d'événements et les pipelines de streaming
- Avro avec Apache Spark : lecture et écriture de données Avro dans des applications Spark
- Cas d'utilisation dans les architectures de Big Data (streaming et batch)
- Gestion des fichiers Avro dans des environnements multi-clusters (réplication, sauvegarde)
- Optimisation des performances dans les systèmes distribués avec Avro

Outils et pratiques avancées avec Apache Avro

- Outils d'Avro : avro-tools pour manipuler et convertir des fichiers
- Compression des données Avro : formats compatibles (Snappy, Deflate, etc.)
- Meilleures pratiques pour l'évolution des schémas dans des environnements de production
- Sécurisation des données Avro (chiffrement, gestion des accès)
- Debugging et validation des schémas dans des systèmes distribués
- Surveiller et gérer les performances d'Apache Avro dans des pipelines de données

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.