

Mis à jour le 25/03/2024

S'inscrire

# Formation Apache Druid

2 jours (14 heures)

## Présentation

Apache Druid est une base de données analytics en temps réel rapide et moderne. Druid est recommandé si vous désirez une solution effectuant des requêtes OLAP et des ingestions de données avec peu de latences. En effet, Apache Druid est rapide sur des requêtes consistantes à haute concurrence. Ce data warehouse performe pour des requêtes agrégées sur des jeux de données de la taille d'un petabyte. Druid peut être déployé facilement dans n'importe quel environnement NIX. Vous pouvez aussi facilement intégrer l'outil avec vos pipelines existants, les données structurées et semi-structurées sont compatibles avec Apache Druid. Cette formation Apache Druid pour les développeurs vous enseignera les différents composants de Druid, vous saurez comment ingérer, préparer et transformer vos données. Notre formation Apache Druid vous présentera la dernière version de l'outil, à savoir [Apache Druid 0.21](#).

## Objectifs

- Maîtriser l'ingestion et la préparation de données
- Comprendre l'utilité et le fonctionnement d'Apache Druid
- Savoir utiliser Druid pour le développement
- Créer des requêtes avec Druid

## Public visé

- Data Scientists
- Data Engineers
- Développeurs
- Data Analysts
- Chefs de projet
- Consultants
- Business Analysts

## PRÉ-REQUIS

- Connaissances en SQL
- Connaissances en base de données

## Programme de notre formation Apache Druid

### Introduction à Apache Druid

- Brève histoire d'Apache Druid et son évolution
- Aperçu de ses fonctionnalités et capacités principales
- Discussion sur le type de problèmes qu'Apache Druid résout

### Cas d'utilisation et scénarios

- Examen détaillé des cas d'utilisation typiques tels que l'analytique en temps réel, l'agrégation de données à grande échelle et la gestion de données basée sur les événements
- Exploration de scénarios spécifiques à l'industrie comme les services financiers, les télécommunications et le commerce électronique où Druid excelle

### Architecture et composants

- Plongée profonde dans l'architecture de Druid, en se concentrant sur ses caractéristiques de scalabilité, fiabilité et performance
- Description détaillée des composants de Druid : Nœuds Historiques, Nœuds Broker, Nœuds Coordinateur, Nœuds Overlord, Nœuds MiddleManager et Nœuds Indexer
- Explication de l'interaction de ces composants au sein d'un cluster Druid

### Installation et configuration

- Guide étape par étape pour mettre en place un cluster Druid multi-nœuds, incluant les prérequis matériels et logiciels
- Meilleures pratiques pour configurer les paramètres de base afin d'optimiser la performance et la fiabilité
- Introduction à la console de Druid et conseils de navigation de base

### Ingestion de données

- Vue d'ensemble des sources de données de Druid, incluant les bases de données traditionnelles, les files d'attente de messages et les systèmes de fichiers
- Parcours détaillé des processus d'ingestion de données en batch et en temps réel, couvrant les formats de données, les spécifications d'ingestion et les pièges communs
- Exercices pratiques pour ingérer des données d'exemple en utilisant les deux méthodes

### Modèle de données

- Explication de l'architecture de stockage basée sur les segments de Druid, incluant le rôle des segments dans le stockage de données et la performance des requêtes
- Discussion sur les structures de données basées sur le temps et comment Druid optimise pour les données temporelles
- Plongée profonde dans les dimensions et mesures, y compris comment les définir et les utiliser efficacement dans les requêtes

## Interrogation de données dans Druid

- Introduction aux langages de requête de Druid, en se concentrant sur SQL pour sa familiarité et les requêtes natives pour leur flexibilité.
- Techniques pour créer des requêtes efficaces, incluant l'utilisation d'agrégations et de filtres pour affiner les résultats
- Exercices pratiques pour interroger des ensembles de données réels, illustrant les cas d'utilisation communs et les stratégies d'optimisation

## Indexation et tuning

- Meilleures pratiques pour l'indexation des données, se concentrant sur des stratégies pour améliorer la performance des requêtes et la compression des données
- Directives de tuning pour les processus d'ingestion en temps réel et en batch pour maximiser l'efficacité et l'exactitude

## Extensions et intégrations de Plugin

- Vue d'ensemble de l'écosystème de Druid, mettant en lumière les extensions et plugins populaires pour une fonctionnalité améliorée
- Guide étape par étape pour l'intégration de Druid avec des outils de visualisation tels qu'Apache Superset et Grafana, incluant la configuration et la création de tableaux de bord basiques
- Discussion sur la communauté et les ressources disponibles pour un apprentissage supplémentaire et la résolution de problèmes

## Conclusion et Q&A

- Récapitulatif des concepts clés abordés dans le cours
- Forum ouvert pour les questions, permettant aux participants de discuter de défis spécifiques ou de sujets d'intérêt plus profond

## 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.