

Mis à jour le 27/07/2023

S'inscrire

# Formation Apache Pinot

3 jours (21 heures)

## Présentation

Déjà utilisé par de grandes entreprises du digital comme Uber, LinkedIn ou encore Slack. Apache Pinot est un datastore OLAP distribué en temps réel permettant de créer des requêtes à faibles latences. Apache Pinot intègre les données à grande échelle en étant proche du temps réel avec Apache Kafka. L'outil est un horizontalement scalable et tolérant aux fraudes. Pinot permet l'ajout de plusieurs extensions. En effet, les plugins sont très nombreux pour exploiter toutes les capacités de l'outil. Il peut aussi être utilisé avec Docker et Kubernetes. Vous pourrez également lancé des algorithmes de Machine Learning pour la détection d'anomalies. Vous pourrez créer des tables hybrides composées de tables hors lignes et connectées en temps réels. Notre formation Apache Pinot vous présentera l'outil et ses différentes features, vous saurez comment préparer et transformer vos données pour créer des requêtes OLAP à faibles latences. Cette formation Apache Pinot vous présentera la dernière version du data store, [Apache Pinot 0.7.1](#).

## Objectifs

- Comprendre l'utilité et le fonctionnement d'Apache Pinot
- Créer des requêtes avec Pinot
- Maitriser l'ingestion et la préparation de données
- Maitriser le déploiement et le monitoring

## Public visé

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

## PRÉ-REQUIS

- Connaissances en SQL
- Connaissances en base de données

## Programme de notre formation Apache Pinot

### Présentation de l'outil

- Qu'est-ce qu'Apache Pinot
- Pourquoi et quand l'utiliser?
- Installation
- Configuration

### Le fonctionnement d'Apache Pinot

- L'architecture de Pinot
- Le modèle de stockage
- Ingérer des données
- Le fonctionnement d'une requête
- Présentation des composants (cluster, controller, broker...)

### Premier pas avec Pinot

- Créer son premier cluster
- Préparation de données
- Création du schéma
- Créer un segment
- Créer sa table
- Écrire sa première requête
- Les techniques d'indexation (forward, inverted, star-tree...)

### Pinot SQL

- Identifiants vs literals
- Différences avec SQL
- Les transformations et fonctions de groupe supportées
- Requêtes CASE WHEN
- UDF et colonnes BYTES

### Programmation

- Les différents modules Pinot
- Fonctions d'agrégation personnalisées
- Manipuler des segments
- Les différents plugins

## Développement avancé

- Configuration de Pinot pour un usage avancé
- Créer un stream pour l'ingestion des données
- Transformation avancée

## Déploiement et monitoring

- Contrôle d'accès
- Les principales métriques pour le monitoring
- Créer des métriques personnalisées
- Tuning Pinot

## Module complémentaire : Docker & Kubernetes (+1 jour)

### Apache Pinot avec Docker

- Présentation de Docker
- Configuration de l'outil
- Commandes batch
- Stream processing
- Stream hybride

### Apache Pinot avec Kubernetes

- Présentation de Kubernetes
- Configuration de l'outil
- Créer son cluster
- Charger ses données avec Kafka
- Pinot Data Explorer
- Utiliser presto pour les requêtes ad hoc

## 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.

## 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.