Updated 07/27/2023

Sign up

Apache Pinot training

3 days (21 hours)

Presentation

Already used by major digital companies such as Uber, Linkedin and Slack. Apache Pinot is a distributed, real-time OLAP datastore for creating low-latency queries. Apache Pinot integrates large-scale, near-real-time data with Apache Kafka. The tool is horizontally scalable and fraud-tolerant. Pinot allows the addition of several extensions. Indeed, there are numerous plugins to exploit all the tool's capabilities. It can also be used with Docker and Kubernetes.
You can also run Machine Learning algorithms for anomaly detection. You'll be able to create hybrid tables made up of offline tables connected in real time. Our Apache Pinot training course will introduce you to the tool and its various features, and show you how to prepare and transform your data to create low-latency OLAP queries. This Apache Pinot training course will introduce you to the latest version of the data store, Apache Pinot 0.7.1.

Objectives

- Understanding the use and operation of Apache Pinot
- Creating queries with Pinot
- Mastering data ingestion and preparation
- Master deployment and monitoring

Target audience

- Data Scientists
- Developers
- Data Analysts
- Project managers
- Consultants
- Business Analysts

PREREQUISITES

- Knowledge of SQL
- Knowledge of databases

Apache Pinot training program

Tool presentation

- What is Apache Pinot
- Why and when to use it?
- Installation
- Configuration

How Apache Pinot works

- Pinot architecture
- The storage model
- Data engineering
- How a query works
- Presentation of components (cluster, controller, broker, etc.)

First step with Pinot

- Create your first cluster
- Data preparation
- Schematic creation
- Create a segment
- Create your table
- Writing your first query
- Indexing techniques (forward, inverted, star-tree...)

Pinot SQL

- Identifiers vs literals
- Differences with SQL
- Supported transformations and group functions
- CASE WHEN queries
- UDF and BYTES columns

Programming

- The different Pinot modules
- Customized aggregation functions
- Handling segments
- The different plugins

Advanced development

- Pinot configuration for advanced use
- Create a stream for data ingestion
- Advanced processing

Deployment and monitoring

- Access control
- Key metrics for monitoring
- Create custom metrics
- Tuning Pinot

Add-on module: Docker & Kubernetes (+1 day)

Apache Pinot with Docker

- Introducing Docker
- Tool configuration
- Batch commands
- Stream processing
- Stream hybrid

Apache Pinot with Kubernetes

- Introducing Kubernetes
- Tool configuration
- Create your cluster
- Loading data with Kafka
- Pinot Data Explorer
- Use presto for ad hoc queries

Companies concerned

This course is aimed at both individuals and companies, large or small,

wishing to train its teams in a new advanced IT technology, or to acquire specific business knowledge or modern methods.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.