

Updated 03/25/2024

Sign up

Apache Druid training

2 days (14 hours)

Presentation

Apache Druid is a fast, modern real-time analytics database. Druid is recommended if you want a solution that performs OLAP queries and data ingestion with low latency. In fact, Apache Druid is fast on consistent queries with high concurrency. This data warehouse performs well for aggregated queries on petabyte-sized datasets. Druid can be easily deployed in any NIX environment. You can also easily integrate the tool with your existing pipelines, as both structured and semi-structured data are compatible with Apache Druid. This Apache Druid training for developers will teach you the different components of Druid, so you'll know how to ingest, prepare and transform your data. Our Apache Druid training will introduce you to the latest version of the tool, [Apache Druid 0.21](#).

Objectives

- Mastering data ingestion and preparation
- Understanding the use and operation of Apache Druid
- Using Druid for development
- Creating queries with Druid

Target audience

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

PREREQUISITES

- Knowledge of SQL
- Knowledge of databases

Apache Druid training program

Introduction to Apache Druid

- A brief history of Apache Druid and its evolution
- Overview of key features and capabilities
- Discussion of the kinds of problems Apache Druid solves

Use cases and scenarios

- Detailed examination of typical use cases such as real-time analytics, large-scale data aggregation and event-based data management
- Exploration of industry-specific scenarios, such as financial services telecommunications and e-commerce where Druid excels

Architecture and components

- Deep dive into Druid's architecture, focusing on its scalability, reliability and performance features
- Detailed description of Druid components: Historical Nodes, Broker Nodes, Nodes Coordinator, Overlord nodes, MiddleManager nodes and Indexer nodes
- Explanation of how these components interact within a Druid cluster

Installation and configuration

- Step-by-step guide to setting up a multi-node Druid cluster, including hardware and software requirements
- Best practices for configuring basic parameters to optimize performance and reliability
- Introduction to the Druid console and basic navigation tips

Data ingestion

- Overview of Druid data sources, including traditional databases, message queues and file systems
- Detailed walk-through of batch and real-time data ingestion processes, covering data formats, ingestion specifications and common pitfalls
- Practical exercises to ingest example data using both methods

Data model

- Explanation of Druid's segment-based storage architecture, including the role of segments in data storage and query performance
- Discussion of time-based data structures and how Druid optimizes for them.
time data
- Deep dive into dimensions and measurements, including how to define them and use them effectively in queries

Querying data in Druid

- Introduction to Druid's query languages, focusing on SQL for its familiarity and native queries for their flexibility.
- Techniques for creating efficient queries, including the use of aggregations and filters to refine results
- Practical exercises for querying real datasets, illustrating common use cases and optimization strategies

Indexing and tuning

- Best practices for data indexing, focusing on strategies for improving query performance and data compression
- Tuning guidelines for real-time and batch ingestion processes to maximize efficiency and accuracy

Plugin extensions and integrations

- Overview of the Druid ecosystem, highlighting popular extensions and plugins for enhanced functionality
- Step-by-step guide to integrating Druid with visualization tools such as Apache Superset and Grafana, including configuration and creation of basic dashboards
- Discussion of community and resources available for further learning and problem solving

Conclusion and Q&A

- Summary of key concepts covered in the course
- Open forum for questions, allowing participants to discuss specific challenges or topics of deeper interest

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

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.