Updated on 12/27/2023

Sign up

# Apache Doris training: large-scale data analysis
3 days (21 hours)

## Presentation

Our Apache Doris training course will give you the skills and knowledge you need to exploit the full capabilities of Apache Doris for real-time data analysis. You'll learn how to use this modern, scalable and efficient data warehouse, which stands out for its speed and simplicity in processing large-scale data.

The course will familiarize you with the fundamental concepts of Apache Doris, including its Massively Parallel Processing (MPP) architecture, data storage models and SQL optimization techniques.

Our program will also cover basic configuration and operations through to more advanced topics such as performance optimization, data integration and Python connectivity.

Like all our training courses, it will run on the latest version of Apache Doris.

## Objectives

- **Configuring Apache Doris**
- Integrate and manage data flows
- **Integrating Python with Doris**

## Target audience

- **Data engineers**
- Data analyst

## Prerequisites

- Knowledge of databases and SQL
- Basic knowledge of Python

# Our Apache Doris training program

## Introduction to Apache Doris

- **What is Apache Doris?**
- Tool features
- Applications and case studies

## Environment configuration

- Installation
- Basic parameters
- FE and BE
- Exercise: setting up a simple cluster

## Basic operations

- Dashboard Doris
- Data types
- Data import
- SQL queries
- Data storage

## Advanced SQL operations

- Advanced SQL queries
- Query optimization
- Joins and subqueries
- Window function and application
- Aggregation and filtering

## Data integration

- **Batch data loading**
- Real-time data ingestion methods
- Integration with Kafka
- Integration with Flink

## Performance optimization

- Cluster scaling
- Materialized views
- Performance optimization
- Cluster monitoring
- Indexation

## Integration with Python

- Python configuration
- Connection with Doris
- Data handling
- Data visualization with python libraries

## Security

- Privileges and access control
- configure SSL
- Metadata
- Backup strategies

## Advanced functions

- JSON
- Lakehouse architecture
- Dynamic score
- Real-time analysis
- Optimization for the cloud

## Continuing education

- **Using apache community resources**
- Follow-up on updates
- Contribute to the project

# Companies concerned

This training 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 on entry to training complies with Qualiopi quality criteria. As soon as enrolment is confirmed, the learner receives a self-assessment questionnaire enabling us to

assess their estimated level of proficiency in different types of technology, and their 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 internal security difficulties within the company (intra-company or virtual classroom) that 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.