

Updated 07/27/2023

Sign up

Datalake training

2 days (14 hours)

Presentation

The phenomenal increase in unstructured data is driving us to master the analysis and processing of datalakes, also known as "data lakes". A data lake is a collection of large quantities of heterogeneous data. This data can now be processed operationally in real time. Mastering a data lake approach will enable you to discover rare and diverse data that can have a strong impact on your organization's performance. This 2-day datalake training course will teach you how to ingest, modify and store your data. You'll also learn how to implement a lambda architecture.

Objectives

- Understanding the Data Engineer's job
- Setting up a Datalake
- Rapidly make available large volumes of ready-to-use data

Target audience

- Data Scientists
- Devops
- Developers
- You work with data engineers

Prerequisites

Programming basics

Datalake training program

WHAT IS BIG DATA?

- Definition of Big Data - Definition of Data Engineering - Skills mapping - Main Data Engineering use cases Practical workshop: setting up the environment

THE BENEFITS OF A DATALAKE

- CAP theorem - What are ETLs and ELTs? - Definition of a datalake - Distributed file system Practical workshop: handling HDFS nodes

DATA INGESTION

- Real-time data manipulation with Kafka - Data flow management with Nifi Practical workshop: ingesting real-time data into a datalake with Nifi and Kafka

DATA TRANSFORMATION

- The difference between structured and unstructured data - The MapReduce paradigm - The Hadoop ecosystem - Data transformation with Spark Practical workshop: standardizing formats and enriching data on a data lake

DATA AVAILABILITY

- Definition of a Datawarehouse - Hyve's added value for data exposure - Practical workshop: making structured data available with Hyve - Visualizing structured data

LAMBDA ARCHITECTURE

- What is a Lambda architecture - Implementing a Lambda architecture in a Datalake Practical workshop: implementing a Lambda architecture

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.

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.