Hadoop Fundamentals

Tecnologias de Informação - Data & Analytics

Live Training (também disponível em presencial)

- **Localidade:** Aveiro
- **Data:** 04 Sep 2023
- **Preço:** 1080 € (Os valores apresentados não incluem IVA. Oferta de IVA a particulares e estudantes.)
- **Horário:** Laboral das 09h30 - 16h30
- **Nível:**
- **Duração:** 18h

Sobre o curso

This Hadoop Fundamentals course teaches you the basics of Apache Hadoop and the concept of Big Data.

Destinatários

- This course targets it professionals that want to start learning Hadoop.

Objetivos

- It takes you on a journey that explains the Hadoop conceptual design
- It looks how to use the application and then manipulate data without the use of complex coding.

Pré-requisitos

- Participants of this course need to have some understanding of Java/Python programming and SQL.

Programa

- Hadoop 3: Background and Introduction
- Planning and Setting Up Hadoop Clusters
- Hadoop Distributed File System
- Developing MapReduce Applications
- Building Rich YARN Applications
- Monitoring and Administration of a Hadoop Cluster
- Demystifying Hadoop Ecosystem Components
- Other Topics in Apache Hadoop

Hadoop 3: Background and Introduction

- How it all started
- What Hadoop is and why it is important
- How Apache Hadoop works
- Hadoop 3.x releases and new features
- Choosing the right Hadoop distribution

Planning and Setting Up Hadoop Clusters

- Prerequisites for Hadoop setup
- Running Hadoop in standalone mode
- Setting up a pseudo Hadoop cluster
- Planning and sizing clusters
- Setting up Hadoop in cluster mode
- Diagnosing the Hadoop cluster

Hadoop Distributed File System

- How HDFS works
- Key features of HDFS
- Data flow patterns of HDFS
- HDFS configuration files
- Hadoop filesystem CLIs
- Working with data structures in HDFS

Developing MapReduce Applications

- How MapReduce works
- Configuring a MapReduce environment
- Understanding Hadoop APIs and packages
- Setting up a MapReduce project

- Deep diving into MapReduce APIs
- Compiling and running MapReduce jobs
- Streaming in MapReduce programming

Building Rich YARN Applications

- Understanding YARN architecture
- Key features of YARN
- Configuring the YARN environment in a cluster
- Working with YARN distributed CLI
- Deep dive with YARN application framework
- Building and monitoring a YARN application on a cluster

Monitoring and Administration of a Hadoop Cluster

- Roles and responsibilities of Hadoop administrators
- Planning your distributed cluster
- Resource management in Hadoop
- High availability of Hadoop
- Securing Hadoop clusters
- Performing routine tasks

Demystifying Hadoop Ecosystem Components

- Understanding Hadoop’s Ecosystem
- Working with Apache Kafka
- Understanding Hive
- Using HBase for NoSQL storage

Other Topics in Apache Hadoop

- Hadoop use cases in industries
- Advanced Hadoop data storage file formats
- Data analytics with Apache Spark