



Introduction to Big Data Administration

Tecnologias de Informação - Data & Analytics

- **Nível:** Intermédio
 - **Duração:** 28h
-

Sobre o curso

Este curso responde à crescente necessidade das organizações em trabalharem grandes quantidades de dados, onde por vezes é necessário cruzar várias ferramentas e métodos para se conseguirem alcançar os resultados esperados.

Durante esta formação, iremos trabalhar com algumas das ferramentas mais relevantes e com maior potencial nesta área, tal como:

- *Hadoop*: plataforma de software Java de computação distribuída voltada para clusters e processamento de grandes volumes de dados, com atenção a tolerância a falhas. Foi inspirado no MapReduce e no GoogleFS (GFS);
 - *Hive, Pig*: Data Warehouse desenvolvido em cima do Apache Hadoop para consulta e análise de dados. Oferece uma interface semelhante ao SQL para consulta de dados em diferentes bancos de dados e sistemas de arquivos integrados do Hadoop;
 - *Impala*: Mecanismo de consulta SQL de código aberto massivamente paralelo (MPP) para dados armazenados num cluster com Apache Hadoop;
 - *Flume, Sqoop*: Software distribuído, e disponível para obter, agregar e mover grandes quantidades de dados de forma eficiente.
-

Destinatários

Todos aqueles que pretendam trabalhar com Big Data, nomeadamente:

- Instalar/Configurar/Administrar uma infraestrutura de Big Data On Premises;
 - Criar uma infraestrutura híbrida On Premises / Cloud;
 - Escalar a infraestrutura física para a cloud.
-

Pré-requisitos

- Recomenda-se conhecimentos de bases de dados relacionais, java e redes
 - Serão necessários conhecimentos de Inglês;
 - Não tem requisitos a nível de habilitações académicas.
-

Programa

- Introduction to Big Data;
- Hadoop (Data Management);
- Hive, Pig (Data Access);
- Impala (Data Analysis);
- Flume, Sqoop (Data Integration)

Hadoop (Data Management)

- Setting up a single node cluster
- Installing software
- Start Cluster
- Standalone Operation
- Execution
- YARN

Hive, Pig (Data Access)

- Installation and Configuration
- DDL Operations
 - Creating Hive Tables
 - Browsing through Tables
 - Altering and Dropping Tables
 - Metadata Store
- DML Operations
- SQL Operations
 - Example Queries
 - SELECTS and FILTERS
 - GROUP BY
 - JOIN
 - MULTITABLE INSERT
 - STREAMING

Impala (Data Analysis)

- Installation and Configuration
- Administration
- SQL Reference
 - Comments
 - Data Types
 - Transactions
 - SQL Statements
 - DDL
 - DML

Flume, Sqoop (Data Integration)

- Installation and Configuration
- Administration