

Bootcamp: Arquiteto(a) de Big Data**Trabalho Prático**

Módulo 4	Processamento de dados utilizando o ecossistema Hadoop
-----------------	---

Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Configurar um sistema Hadoop em ambiente Unix;
- ✓ Aplicar ingestão de dados utilizando o Hadoop;
- ✓ Realizar análises no dataset a partir da utilização do Hadoop.

Enunciado

Acompanhe as aulas e realize a instalação de um sistema operacional Linux em uma Máquina Virtual. A seguir, faça a instalação do Hadoop, Hive e Derby. Após feita a instalação e configuração apropriadas, seguindo as notas de aulas, baixe o dataset airports.dat e responda às questões abaixo.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Baixe um programa de Máquina Virtual e instale um Sistema Operacional Linux. Recomenda-se utilizar o Ubuntu, Linux Mint, Lubuntu, ou Linux Lite.
2. Baixe os binários do Hadoop, Hive e Derby, faça as instalações e configurações. Os links encontram-se a seguir:
 - Hadoop: <https://hadoop.apache.org>
 - Hive: <https://hive.apache.org/>

- Derby: <https://db.apache.org/derby/>
3. Baixe o dataset airports.dat do seguinte link (também estará disponível no fórum da disciplina): <https://openflights.org/data.html>
 4. Insira os dados no HDFS e tente implementar as análises através de Map Reduce no Hadoop ou Hive e responda às questões.

Respostas Finais

Os alunos deverão desenvolver a prática e, depois, responder às seguintes questões objetivas: