

Bootcamp: Engenheiro de Dados

Desafio

Módulo 3	Solução de Dados Utilizando Ecossistema Hadoop.
-----------------	--

Objetivos

Exercitar os seguintes conceitos trabalhados no Módulo:

- ✓ Manipular o HDFS (Hadoop Distributed FileSystem, formatando o sistema de arquivos e executando operações com arquivos e diretórios.
- ✓ Acessar as ferramentas de Gerenciamento do Ecossistema Hadoop (Apache Spark).

Enunciado

Para esta atividade, o aluno deverá assistir atentamente as seguintes aulas, disponíveis no ambiente virtual de aprendizagem, **exatamente na ordem em que se apresenta abaixo**:

1. Instalando a máquina virtual;
2. Executando os comandos básicos do Ecossistema Hadoop e do HDFS;
3. Prática: trabalhando com o Apache Spark.

Atividades

Os alunos deverão desempenhar as seguintes atividades:

1. Iniciar os 5 serviços do Hadoop, por meio do comando
`/usr/local/hadoop/sbin/start-all.sh.`

2. Se você manteve os arquivos do Trabalho Prático no HDFS, você poderá utilizá-lo para realizar esta atividade. Caso contrário, será necessário importar os dados utilizando o Apache Sqoop, assim como foi feito na aula de apoio ao Trabalho Prático.
3. Executar o Apache Spark em:


```
/usr/local/spark/bin/spark-shell
```
4. Seguir os passos do vídeo “Prática: trabalhando com o Apache Spark”.

Respostas Finais

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