

# INTRODUÇÃO AO HDFS

Máquina virtual (VirtualBox) Bitnami com Hadoop Instalado.

<https://bitnami.com/stack/hadoop/virtual-machine>

Hadoop é uma das mais populares plataformas para Big Data. O HDFS (Hadoop Distributed File System) é o sistema de arquivos componente do Hadoop projetado para armazenar grande volume de dados.

O HDFS possui dois tipos de nós, NameNode responsável por armazenar informações da distribuição dos arquivos e metadados, e o Datanode que é responsável os dados propriamente ditos.

## //Comandos básicos

//Listando os diretorios

```
hdfs dfs -ls /user
```

```
Found 4 items
drwxr-xr-x   - hadoop supergroup          0 2019-11-16 17:35 /home
drwxrwxrwt   - hadoop supergroup          0 2019-08-26 06:53 /tmp
drwxr-xr-x   - hadoop supergroup          0 2019-11-16 17:57 /user
```

//Criando um diretorio

```
hdfs dfs -mkdir /user/dados
```

//Incluindo um arquivo no diretório

```
hdfs dfs -put Bovespa.csv /user/dados
```

```
Found 1 items
-rw-r--r--   1 hadoop supergroup 554105111 2019-11-16 17:58
/user/dados/Bovespa.csv
```

//Copia o arquivo do hdfs para o file system local

```
hdfs dfs -get /user/dados/Bovespa.csv BovespaLocal.csv
```

```
-rw-r--r--   1 bitnami bitnami 554105111 Jun  8 16:56 BovespaLocal.csv
```

//Mostra o tamanho de um arquivo no file system

```
hdfs dfs -du /user/dados/Bovespa.csv
```

//Mostra o conteúdo final de um arquivo no file system

```
hdfs dfs -tail /user/dados/Bovespa.csv
```

//Mostra o conteúdo de um arquivo na saída padrão

```
hdfs dfs -cat /user/dados/Bovespa.csv
```

## //Outros comandos

<code>mv</code>	Move um arquivo do diretório origem para o destino
<code>rm</code>	Apaga um arquivo
<code>test -e</code>	Verifica se o arquivo existe
<code>test -z</code>	Verifica se o arquivo tem tamanho zero
<code>test -d</code>	Verifica se é um diretório
<code>touchz</code>	Cria um arquivo vazio