

# Relatório Lab 1 - Hadoop – Instalação em modo Single Node

Aluna: Letícia Trein Medeiros

Para a instalação do Hadoop foi necessário seguir os passos descritos no Lab 1, orientações em aula e auxílio de vários colegas. Assim, foram seguidos os passos descritos abaixo:

## Passo 1: Máquina Virtual

Foi baixada a máquina virtual disponibilizada via moodle com pré configurações (passos iniciais do Lab 1). Para rodar a máquina foi instalado o software Oracle Virtual Box. Inicialmente não houve problemas para a rodar a máquina em uma primeira vez, contudo, na vez seguinte a máquina não respondeu mais, sendo necessária re-importação da mesma com adequações de configurações para evitar que travasse novamente (redução de RAM disponível para aproximadamente 2 GB e redução de 6 para 4 núcleos para o processamento).

## Passo 2: Baixar Hadoop

Para instalação do hadoop foi necessário baixar seu arquivo no site [hadoop.apache.org](http://hadoop.apache.org) e extrair os dados do arquivo compactado na pasta downloads. (nota: este momento houve um erro, baixei o arquivo errado - o src, não o binário)

Conforme o Lab 1, seria necessária mudança dos arquivos da pasta extraída para o `usr/local/hadoop`; pasta que por sua vez não havia sido criada. Para descobrir como fazer, reassisti a aula gravada, na qual foram explicados os detalhes faltantes no Lab 1; assim, utilizando o comando `sudo mkdir`, foi criada a pasta e utilizando o `sudo mv`, foi movido os arquivos entre as pastas.

## Passo 3: Configurando o .bashrc

Segundo o Lab 1: “Você precisa editar o arquivo `.bashrc`, no home directory do usuário, adicionando as seguintes linhas:” seria necessário adequar o arquivo `.bashrc`; para isso, novamente retornei a vídeo aula e encontrei o arquivo que estaria oculto no diretório home, e com a instrução de um colega, coleí o código disponível (abaixo) no Lab 1 no final do arquivo, para abrir o arquivo e editá-lo foi necessário o comando `nano`. Foi necessária adequação do código com as mudanças das aspas duplas para simples na primeira linha e adição de aspas simples na segunda linha. Após modificação foram salvas as alterações e executado com o comando `source ~/.bashrc`

Código utilizado:

```
export JAVA_HOME='/usr/lib/jvm/java-8-openjdk-amd64'
```

```
export HADOOP_HOME='/usr/local/hadoop'
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
unset JAVA_TOOL_OPTIONS
export PDSH_RCMD_TYPE=ssh
```

#### **Passo 4: Editar arquivos do diretório etc/hadoop**

O Lab 1 indica a alteração de 5 arquivos, todos presentes no diretório `usr/local/hadoop/etc/hadoop`; contudo, neste momento, o erro cometido no segundo passo impediu que procedesse. Com a ajuda de um outro colega, identificamos que eu havia baixado o arquivo errado, de forma que os diretórios estariam diferentes nos arquivos baixados. Assim, foi baixada a versão correta do arquivo e foram alterados os seguintes arquivos: `hadoop-env.sh`; `core-site.xml`; `mapred-site.xml`; `hdfs-site.xml`; `yarn-site.xml` conforme códigos disponibilizados no Lab 1.

#### **Passo 5: Preparar o sistema de arquivos para o namenode e datanode**

Foram criados os diretórios para namenode e datanode utilizando os seguintes comandos:

```
mkdir -p /usr/local/hadoop/hadoop_data/hdfs/namenode
mkdir -p /usr/local/hadoop/hadoop_data/hdfs/datanode
sudo chown bigdata -R /usr/local/hadoop
```

Depois foi formatado o sistema de arquivos HDFS pelo comando:

```
hdfs namenode -format
```

#### **Passo 6: Iniciar o cluster e testar**

Para iniciar o HDFS e YARN foram usados inicialmente os comandos `start-dfs.sh`, [start-yarn.sh](#) e `mapred historyserver`; contudo, o código apontava que estava com a permissão negada; para solucionar, um outro colega informou que precisou reiniciar o SSH, e, após iniciação, os programas passaram a rodar. Após comprovar a execução das 6 tarefas esperadas pela função `jps`, foi realizado o teste.

Para o teste foi alterado o caminho de diretório para `$HADOOP_HOME/share/hadoop/mapreduce` e nesta pasta foi rodado o comando: `hadoop jar hadoop-mapreduce-client-jobclient-3.4.1-tests.jar TestDFSIO -write -nrFiles 5 -fileSize 10`; sendo este adaptado do Lab 1 para a versão instalada do hadoop;

Após iniciar o código, a máquina virtual trava e não consigo finalizar a visualização.