

## 1 Modo *Single Node*

Nesta seção é apresentado a configuração do Apache Hadoop no modo *Single Node*. Com a finalização desta seção será possível montar um ambiente capaz de executar rotinas *MapReduce* e gerenciar arquivos de forma distribuída com o HDFS.

Para instalação do Apache Hadoop em modo *Single Node* é necessário atender os seguintes pré-requisitos:

- Tenha o sistema operacional Ubuntu 14.04 ou superior instalado;
- Tenha o Java instalado;
- Tenha o servidor *ssh* devidamente instalado e configurado; e
- Configure o *ssh* para o acesso sem senha com todos os endereços de IP das máquinas desejadas incluindo o *localhost*.

1. Instale o Java. Caso já tenha instalado, desconsidere esta etapa:

### Listing 1: Instalação do Java 8 Oracle

```
~$ sudo add-apt-repository -y ppa:webupd8team/java \
&& sudo apt-get update \
&& sudo apt-get install oracle-java8-installer -y
```

Ou caso prefira o OpenJDK:

### Listing 2: Instalação do OpenJDK

```
~$ sudo apt-get install openjdk-8-jdk -y
```

2. Verifique o *path* (caminho) de instalação do Java:

### Listing 3: Verificar caminho de instalação do Java

```
~$ update-alternative --config java
```

3. Adicione ao final do arquivo */etc/environments* a variável de ambiente *JAVA\_HOME* ao *PATH* do sistema:

### Listing 4: Adicionar variável de ambiente *JAVA\_HOME*

```
~$ sudo echo 'JAVA_HOME="/usr/lib/jvm/java-8-jdk'" >> /etc/environments
```

4. Atualize as variáveis de ambiente do S.O:

### Listing 5: Atualize as variáveis de ambiente

```
~$ source /etc/environment
```

5. Verifique se o *path* está correto:

Listing 6: Verificar o *path* do JAVA\_HOME

```
~$ echo $JAVA_HOME
```

6. Instale o servidor *ssh* com o seguinte comando. Caso já tenha instalado, desconsidere esta etapa:

Listing 7: Instalação do *ssh*

```
~$ sudo apt-get install ssh -y
```

7. Verifique se a máquina tem acesso a *ssh* sem senha ao *localhost*:

Listing 8: Verificar *ssh* sem senha no *localhost*

```
~$ ssh localhost
```

8. Caso seja solicitado uma senha, faça o seguinte procedimento:

Crie uma chave pública de acesso ao *ssh*:

Listing 9: Criar uma chave pública para acesso ao *ssh*

```
~$ ssh-keygen -t rsa -P ''
```

Agora com a chave criada, adicione ela ao diretório de chaves de acesso autorizadas:

```
~$ cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
```

9. Faça o *download* do Apache Hadoop, neste caso a versão 2.7.3:

Acesse o site oficial do Hadoop para o *download*:

<http://hadoop.apache.org/releases.html>

10. Descompacte o Apache Hadoop no diretório */opt*:

Listing 10: Comando para descompactar arquivo tar.gz

```
~$ sudo tar -zxvf ~/Downloads/hadoop-2.7.3.tar.gz -C /opt
```

11. Renomeia o diretório do Apache Hadoop com o comando:

Listing 11: Comando para renomear diretório

```
~$ sudo mv /opt/hadoop-2.7.3 /opt/hadoop
```

12. Configure as variáveis de ambiente do Apache Hadoop:

Edite o arquivo `/etc/profile` ou `/.bashrc`. Caso prefira que todos usuários do sistema tenham as variáveis de ambiente:

Listing 12: Edite o arquivo `/etc/profile`

```
~$ nano /etc/profile
```

Ou caso prefira que apenas o seu usuário tenham as variáveis de ambiente:

Listing 13: Edite o arquivo `/.bashrc`

```
~$ nano ~/.bashrc
```

13. Adicione ao final do arquivo as seguintes variáveis de ambiente:

Listing 14: Adicione as seguintes variáveis de ambiente

```
export JAVA_HOME=$JAVA_HOME
export HADOOP_INSTALL=/opt/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
```

14. Atualize as variáveis de ambiente:

Listing 15: Atualizar as variáveis de ambiente do S.O

```
~$ source /etc/profile
~$ source ~/.bashrc
```

15. Edite o arquivo de variáveis de ambiente do Hadoop:

Listing 16: Editar arquivo `hadoop-env.sh`

```
~$ sudo nano $HADOOP_INSTALL/etc/hadoop/hadoop-env.sh
```

16. Edite a variável de ambiente `JAVA_HOME`:

Listing 17: Adicionar *path* (caminho) do `JAVA_HOME`

```
~$ export JAVA_HOME=<YOUR_JAVA_HOME_PATH>
```

17. Crie o diretório que será o ponto de montagem para o sistemas de arquivos HDFS:

Listing 18: Criação dos diretórios para o HDFS

```
~$ mkdir -p ~/hadoop_store/hdfs/namenode
~$ mkdir -p ~/hadoop_store/hdfs/datanode
```

18. Configure o arquivo hdfs-site.xml com o caminho das pastas criadas:

Listing 19: Editar o arquivo hdfs-site.xml

```
~$ sudo nano $HADOOP_INSTALL/etc/hadoop/hdfs-site.xml
```

19. Adicione as seguintes propriedades:

Listing 20: Adicione as seguintes propriedades

```
<property>
<name>dfs.replication</name>
<value>1</value>
<description></description>
</property>

<property>
<name>dfs.namenode.name.dir</name>
<value>file:#####NAMENODE_FOLDER_PATH#####</value>
</property>

<property>
<name>dfs.datanode.data.dir</name>
<value>file:#####DATANODE_FOLDER_PATH#####</value>
</property>
```

20. Crie um diretório *tmp* na pasta de instalação do Hadoop:

Listing 21: Criar diretório *tmp* do HDFS

```
~$ mkdir -p ~/hadoop_store/hdfs/tmp
```

21. Configure o arquivo core-site.xml:

Listing 22: Editar arquivo core-site.xml

```
~$ sudo nano $HADOOP_INSTALL/etc/hadoop/core-site.xml
```

22. Adicione as seguintes propriedades:

**Listing 23: Adicione as seguintes propriedades**

```
<property>
<name>hadoop.tmp.dir</name>
<value>#####TMP_FOLDER_PATH#####</value>
<description></description>
</property>

<property>
<name>fs.default.name</name>
<value>hdfs://localhost:54310</value>
<description></description>
</property>
```

23. Configure o arquivo mapred-site.xml:

**Listing 24: Crie um arquivo chamado mapred-site.xml**

```
~$ sudo cp $HADOOP_INSTALL/etc/hadoop/mapred-site.xml.template
$HADOOP_INSTALL/etc/hadoop/mapred-site.xml
```

24. Adicione a propriedade:

**Listing 25: Adicione a seguinte propriedade**

```
<property>
<name>mapred.job.tracker</name>
<value>localhost:54311</value>
<description></description>
</property>
```

25. Configure o arquivo o arquivo yarn-site.xml:

**Listing 26: Edite o arquivo yarn-site.xml**

```
~$ sudo nano yarn-site.xml
```

26. Adicione as seguintes propriedades:

**Listing 27: Adicione as seguintes propriedades**

```
<property>
<name>yarn.resourcemanager.hostname</name>
<value>localhost</value>
<description></description>
</property>
```

```
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>

<property>
<name>
yarn.nodemanager.aux-services.mapreduce_shuffle.class
</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
```

27. Formate o sistema de arquivos HDFS:

**Listing 28: Formate o sistema de arquivos HDFS**

```
~$ hadoop namenode -format
```

28. Inicie os serviços do Hadoop:

**Listing 29: Inicie o Apache Hadoop**

```
~$ start-dfs.sh && start-yarn.sh
```

29. Verifique os processos do Hadoop:

**Listing 30: Verifique os processos do Apache Hadoop**

```
~$ jps
```

30. No console tem que aparecer uma saída semelhante a essa:

**Listing 31: Verifique a saída do console**

```
9757 Jps
8950 DataNode
9145 SecondaryNameNode
9438 NodeManager
8799 NameNode
9311 ResourceManager
```