Programa de Aceleração de Engenharia de Dados com a DIO

How to para inicialização da Máquina Virtual	
Empresa	Everis
Criado	Edmilson Carmo - ecarmoli
Version	0.1

Esse material foi criado para auxiliar na instalaçao e configuração da VM .

Dentro da imagem criada para sua virtualBox , contém uma boa parte dos compotenentes utilizado neste programa e com isso , otimizando o aprendizado para as atividades práticas .

Na tabela abaixo contém um pouco mais de informações sobre os componentes instalados:

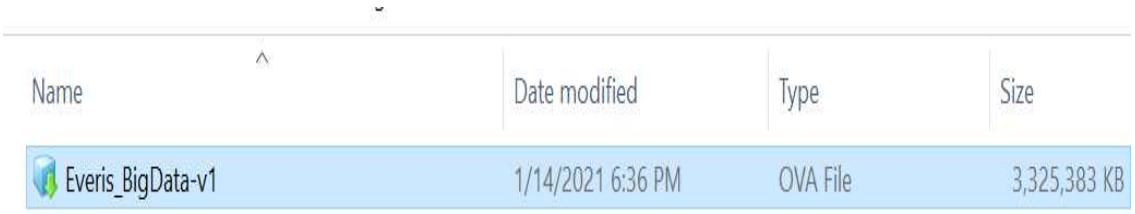
Softwares instalados	
Cloudera	CDH5.16.2
Zookeeper	3.4.5
Hadoop	2.6.0
Hive	1.1.0
Impala	2.12.0
Hbase	1.2.0
Spark	2.4.7
Cassandra	2.1.11
Python	3.7.9

Pré-requisitos

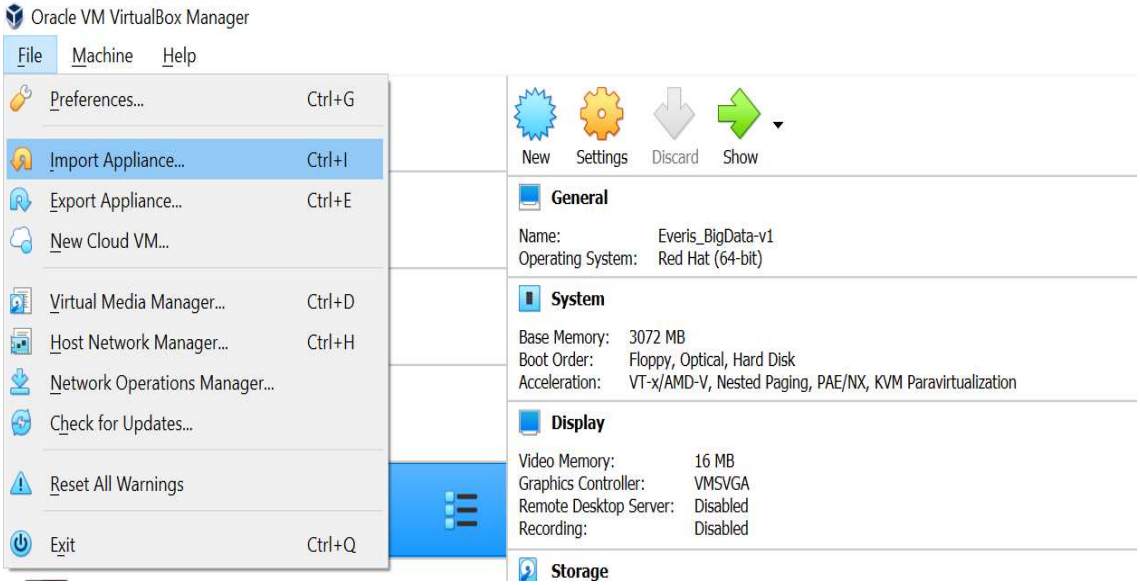
Para o melhor aproveitamento da VM , é aconselhável que a máquina onde a VM será instalada , tenha no minimo 4 GB memoria .

Requisitos
4 GB de memoria
VirtualBox Instalada

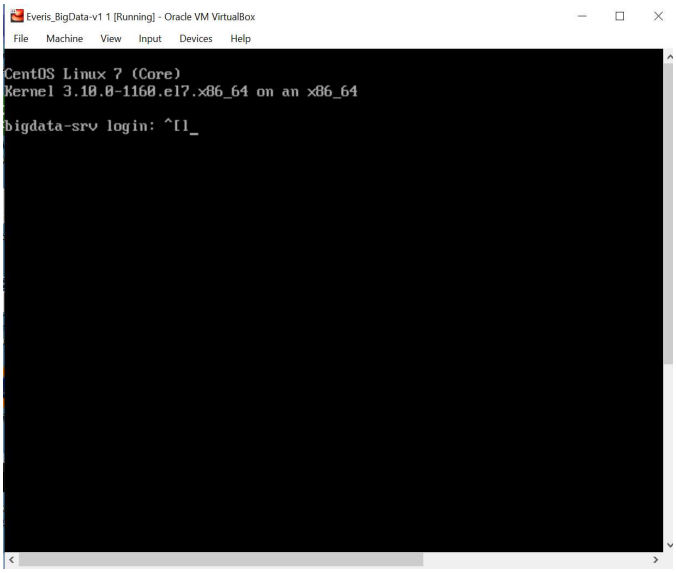
Após realizar o Download do arquivo **Everis_BigData-v1.ova**, verifique se o mesmo está conforme o exemplo abaixo:



Ao abrir a VirtualBox Manager , clique em ***File*** depois clique em ***Import Appliance***



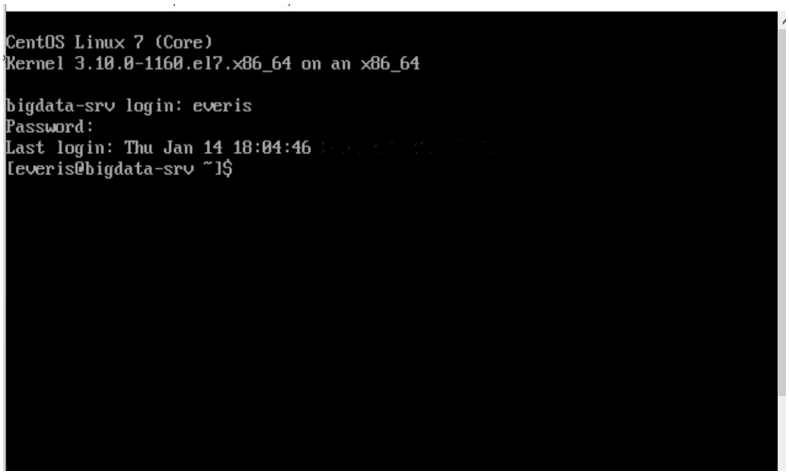
Após criação e inicialização da VM será apresentada a seguinte saida



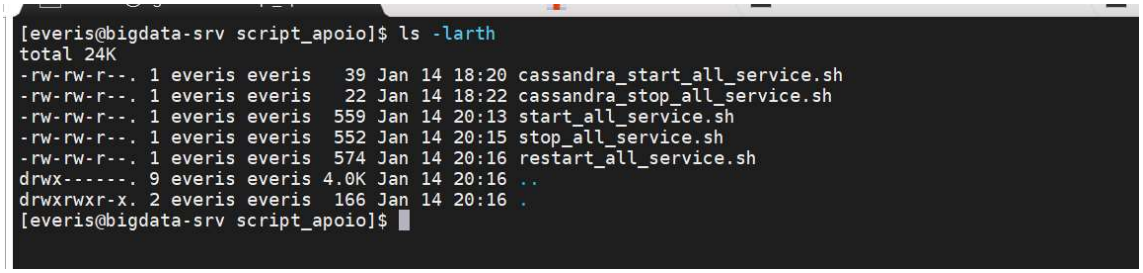
Digite usuário e senha abaixo

User: everis

Senha: everis2021



Para facilitar tais ações como: start/restart/stop dos serviços, foi incluído alguns scripts dentro da pasta **"/home/everis/script_apoio/".** Segue um exemplo:



Todos softwares já estão devidamente instalados e configurados com o mínimo de recursos necessários para o acompanhamento do treinamento.

Abaixo será apresentado um exemplo de uso para cada ferramenta.

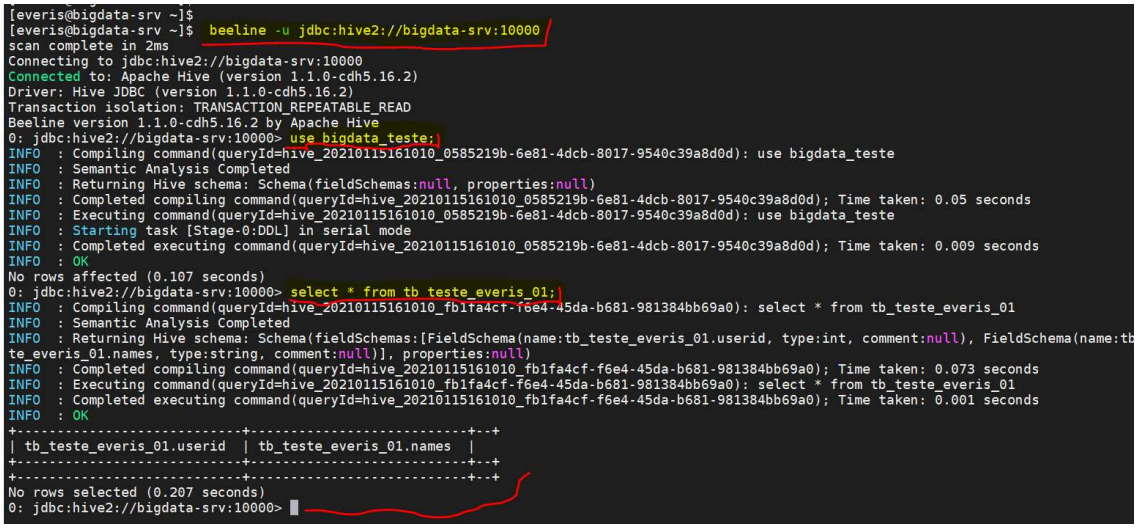
HDFS

Neste exemplo, um arquivo está sendo incluído para dentro de um diretório no HDFS.

```
[everis@bigdata-srv ~]$ hdfs dfs -ls /
Found 4 items
drwxr-xr-x   - hbase supergroup          0 2021-01-15 10:48 /hbase
drwxrwxrwt   - hdfs  supergroup          0 2021-01-15 10:59 /tmp
drwxr-xr-x   - hdfs  supergroup          0 2021-01-13 18:38 /user
drwxr-xr-x   - hdfs  supergroup          0 2021-01-13 12:58 /var
[everis@bigdata-srv ~]$
[everis@bigdata-srv ~]$ hdfs dfs -put file_teste.txt /tmp/
[everis@bigdata-srv ~]$
[everis@bigdata-srv ~]$ hdfs dfs -ls /tmp
Found 3 items
-rw-r--r--   3 everis supergroup       4784 2021-01-15 11:00 /tmp/file_teste.txt
drwx-wx-wx   - hive   supergroup          0 2021-01-13 18:46 /tmp/hive
-rw-r--r--   3 everis supergroup       1341 2021-01-15 10:54 /tmp/hive-site.xml
```

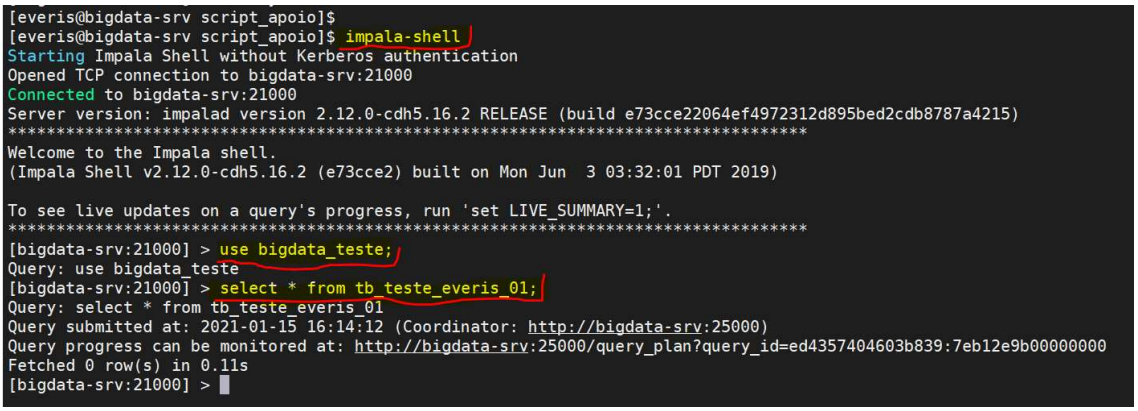
Hive

Para acessar o hive utilizando o Beeline, será necessário a utilização da string de conexão a seguir:
beeline -u jdbc:hive2://bigdata-srv:10000.



Impala

O exemplo abaixo, utiliza o **impala-shell** para acessar o Impala e então executar uma declaração em uma tabela do Hive .



Cassandra

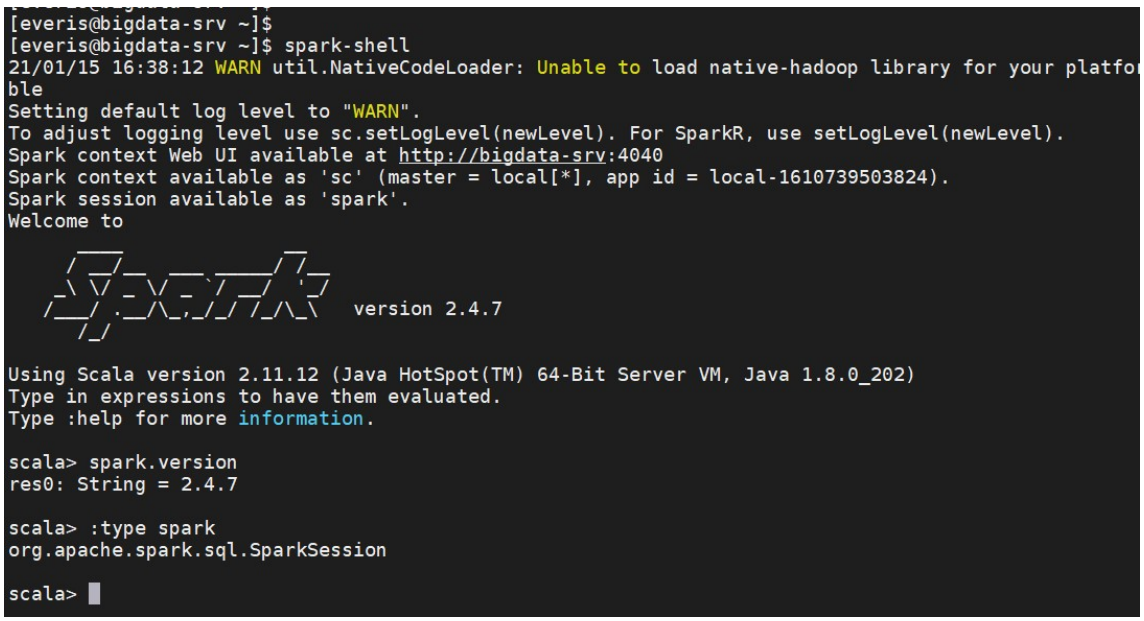
O acesso ao Cassandra será por meio do **CQL** , conforme exemplo abaixo:

Spark

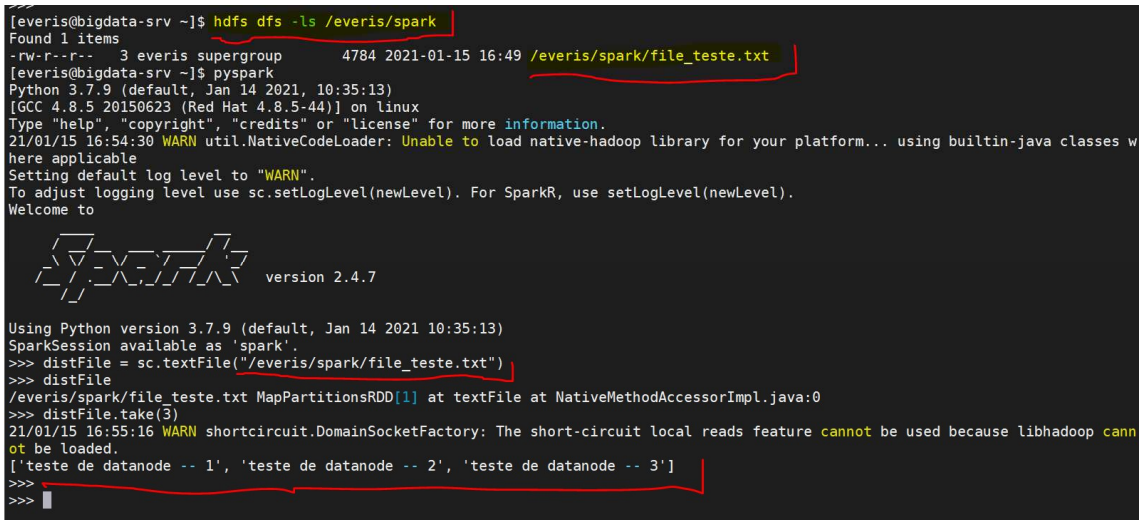
A utilização do Spark pode ser executada de algumas maneiras, em nosso exemplo será utilizado o **Spark-shell** com a liguagem Scala e o **PySpark** com a linguagem Python

Abaixo segue um exemplo para utilização do Spark:

Spark-shell(Scala)



Pyspark (Python)



Material de apoio para instalação dos componentes utilizado nesta VM:

https://docs.cloudera.com/documentation/enterprise/5-16-x/topics/cdh_intro.html

https://docs.datastax.com/en/cassandra-oss/2.1/cassandra/install/installRHEL.html