



an **NTT DATA** Company

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 instalação e configuração da VM .

Dentro da imagem criada para sua virtualBox , contém uma boa parte dos componentes 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

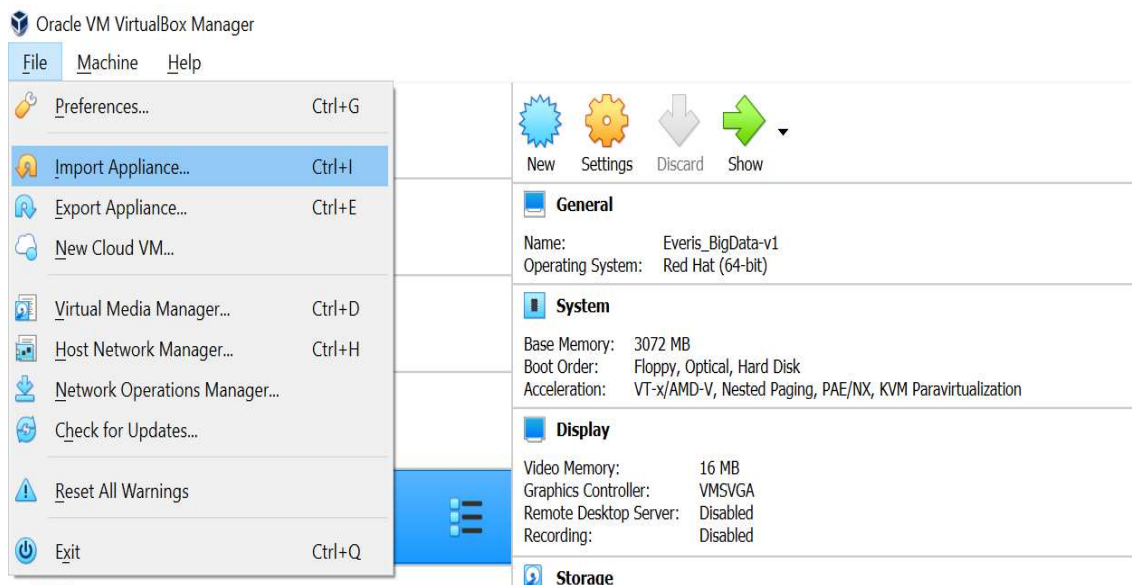
4 GB de memoria

VirtualBox Instalada

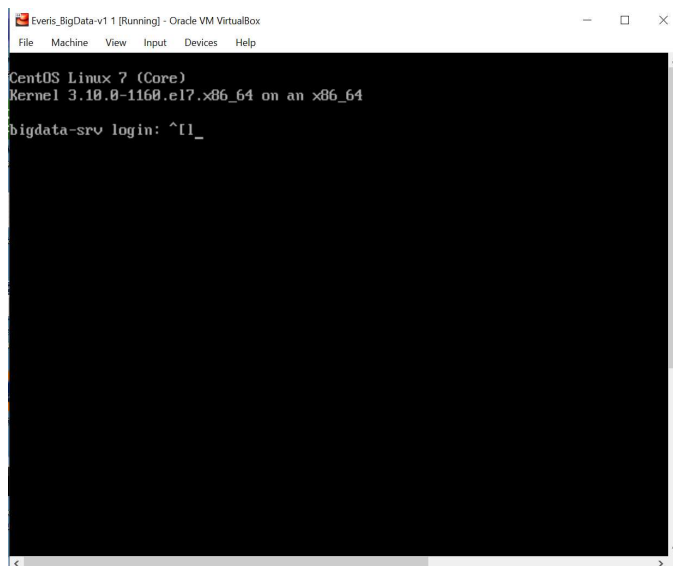
Após realizar o Download do arquivo **Everis_BigData-v1.ova**, verifique se o mesmo está conforme o exemplo abaixo:

Name	Date modified	Type	Size
 Everis_BigData-v1	1/14/2021 6:36 PM	OVA File	3,325,383 KB

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



Após criação e inicialização da VM será apresentada a seguinte saída



```
Everis_BigData-v1 1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

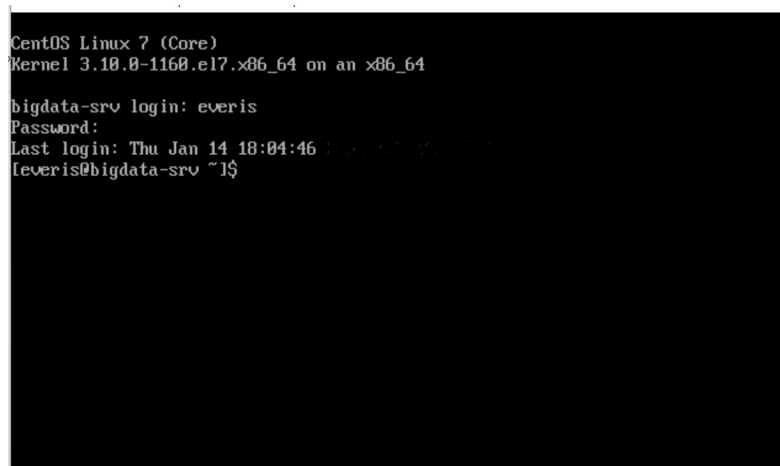
CentOS Linux 7 (Core)
Kernel 3.10.0-1160.el7.x86_64 on an x86_64
bigdata-srv login: ^[]_

```

Digite usuário e senha abaixo

User: everis

Senha: everis2021



```
Everis_BigData-v1 1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

CentOS Linux 7 (Core)
Kernel 3.10.0-1160.el7.x86_64 on an x86_64
bigdata-srv login: everis
Password:
Last login: Thu Jan 14 18:04:46 2021 from 10.0.2.15
everis@bigdata-srv ~1$

```

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

```
[everis@bigdata-srv script_apoio]$ ls -larth
total 24K
-rw-rw-r--. 1 everis everis   39 Jan 14 18:20 cassandra_start_all_service.sh
-rw-rw-r--. 1 everis everis   22 Jan 14 18:22 cassandra_stop_all_service.sh
-rw-rw-r--. 1 everis everis  559 Jan 14 20:13 start_all_service.sh
-rw-rw-r--. 1 everis everis  552 Jan 14 20:15 stop_all_service.sh
-rw-rw-r--. 1 everis everis   54 Jan 14 20:16 restart_all_service.sh
drwx-----. 9 everis everis 4.0K Jan 14 20:16 ..
drwxrwxr-x. 2 everis everis  166 Jan 14 20:16 .
```

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.

```
[everis@bigdata-srv ~]$
[everis@bigdata-srv ~]$ beeline -u jdbc:hive2://bigdata-srv:10000
scan complete in 2ms
Connecting to jdbc:hive2://bigdata-srv:10000
Connected to: Apache Hive (version 1.1.0-cdh5.16.2)
Driver: Hive JDBC (version 1.1.0-cdh5.16.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 1.1.0-cdh5.16.2 by Apache Hive
0: jdbc:hive2://bigdata-srv:10000> use bigdata_teste;
INFO : Compiling command(queryId=hive_20210115161010_0585219b-6e81-4dcb-8017-9540c39a8d0d): use bigdata_teste
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20210115161010_0585219b-6e81-4dcb-8017-9540c39a8d0d); Time taken: 0.05 seconds
INFO : Executing command(queryId=hive_20210115161010_0585219b-6e81-4dcb-8017-9540c39a8d0d): use bigdata_teste
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210115161010_0585219b-6e81-4dcb-8017-9540c39a8d0d); Time taken: 0.009 seconds
INFO : OK
No rows affected (0.107 seconds)
0: jdbc:hive2://bigdata-srv:10000> select * from tb_teste_everis_01;
INFO : Compiling command(queryId=hive_20210115161010_fb1fa4cf-f6e4-45da-b681-981384bb69a0): select * from tb_teste_everis_01
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tb_teste_everis_01.userid, type:int, comment:null), FieldSchema(name:tb_teste_everis_01.names, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20210115161010_fb1fa4cf-f6e4-45da-b681-981384bb69a0); Time taken: 0.073 seconds
INFO : Executing command(queryId=hive_20210115161010_fb1fa4cf-f6e4-45da-b681-981384bb69a0): select * from tb_teste_everis_01
INFO : Completed executing command(queryId=hive_20210115161010_fb1fa4cf-f6e4-45da-b681-981384bb69a0); Time taken: 0.001 seconds
INFO : OK
+-----+-----+
| tb_teste_everis_01.userid | tb_teste_everis_01.names |
+-----+-----+
No rows selected (0.207 seconds)
0: 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.

```
[everis@bigdata-srv script_apoio]$
[everis@bigdata-srv script_apoio]$ impala-shell
Starting Impala Shell without Kerberos authentication
Opened TCP connection to bigdata-srv:21000
Connected to bigdata-srv:21000
Server version: impalad version 2.12.0-cdh5.16.2 RELEASE (build e73cce22064ef4972312d895bed2cdb8787a4215)
*****
Welcome to the Impala shell.
(Impala Shell v2.12.0-cdh5.16.2 (e73cce2) built on Mon Jun 3 03:32:01 PDT 2019)

To see live updates on a query's progress, run 'set LIVE_SUMMARY=1;'.
*****
[bigdata-srv:21000] > use bigdata_teste;
Query: use bigdata_teste
[bigdata-srv:21000] > select * from tb_teste_everis_01;
Query: select * from tb_teste_everis_01
Query submitted at: 2021-01-15 16:14:12 (Coordinator: http://bigdata-srv:25000)
Query progress can be monitored at: http://bigdata-srv:25000/query_plan?query_id=ed4357404603b839:7eb12e9b00000000
Fetched 0 row(s) in 0.11s
[bigdata-srv:21000] >
```

Cassandra

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

```

[everis@bigdata-srv ~]$
[everis@bigdata-srv ~]$ cqlsh/
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 2.1.11 | CQL spec 3.2.1 | Native protocol v3]
Use HELP for help.
cqlsh> use tracking;
cqlsh:tracking> select * from home;

 home_id | address | alt_phone | city | contact_name | email | guest_code | main_code | phone | phone_
password | state | zip
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
H01474777 | 518 Crestview Drive | null | Beverly Hills | Jed Clampett | jclampett@bhb.com | 7778 | 5599 | 310-775-4011 | 
oil | CA | 90046

(1 rows)
cqlsh:tracking>

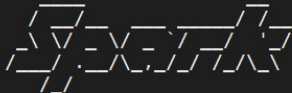
```

Spark

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

Abaixo segue um exemplo para utilização do Spark:

Spark-shell(Scala)

```
[leveris@bigdata-srv ~]$  
[leveris@bigdata-srv ~]$ spark-shell  
21/01/15 16:38:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform  
Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
Spark context Web UI available at http://bigdata-srv:4040  
Spark context available as 'sc' (master = local[*], app id = local-1610739503824).  
Spark session available as 'spark'.  
Welcome to  
 version 2.4.7  
Using Scala version 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_202)  
Type in expressions to have them evaluated.  
Type :help for more information.  
  
scala> spark.version  
res0: String = 2.4.7  
  
scala> :type spark  
org.apache.spark.sql.SparkSession  
  
scala> |
```


Pyspark (Python)

```
[everis@bigdata-srv ~]$ hdfs dfs -ls /everis/spark
Found 1 items
-rw-r--r--   3 everis supergroup      4784 2021-01-15 16:49 /everis/spark/file_teste.txt
[everis@bigdata-srv ~]$ pyspark
Python 3.7.9 (default, Jan 14 2021, 10:35:13)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-44)] on linux
Type "help", "copyright", "credits" or "license()" for more information.
21/01/15 16:54:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to

  ____ _
 / ___ \ | |
| |___ \| |_| |
| |___) | __| |
|_____|_\___|_| version 2.4.7

Using Python version 3.7.9 (default, Jan 14 2021 10:35:13)
SparkSession available as 'spark'.
>>> distFile = sc.textFile("/everis/spark/file_teste.txt")
>>> distFile
/everis/spark/file_teste.txt MapPartitionsRDD[1] at textFile at NativeMethodAccessorImpl.java:0
>>> distFile.take(3)
21/01/15 16:55:16 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used because libhadoop cannot be loaded.
['teste de datanode -- 1', 'teste de datanode -- 2', 'teste de datanode -- 3']
>>>
```

Material de apoio para instalação dos componentes utilizado nesta VM:

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



an **NTT DATA** Company

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