

Título: Relatório o instalação hadoop máquina virtual

Aluno: Rafael De Pauli Baptista

Disciplina: Introdução a Big Data

Relatório detalhando a instalação do ambiente single-node do hadoop em uma máquina virtual sendo rodada na solução VirtualBox da Oracle.

Recursos utilizados:

- [VirtualBox](#);
- [Imagem](#) de uma máquina virtual contendo a versão Linux LUbuntu;
- [Hadoop](#). Para esse processo, foi utilizada a versão 3.3.1;

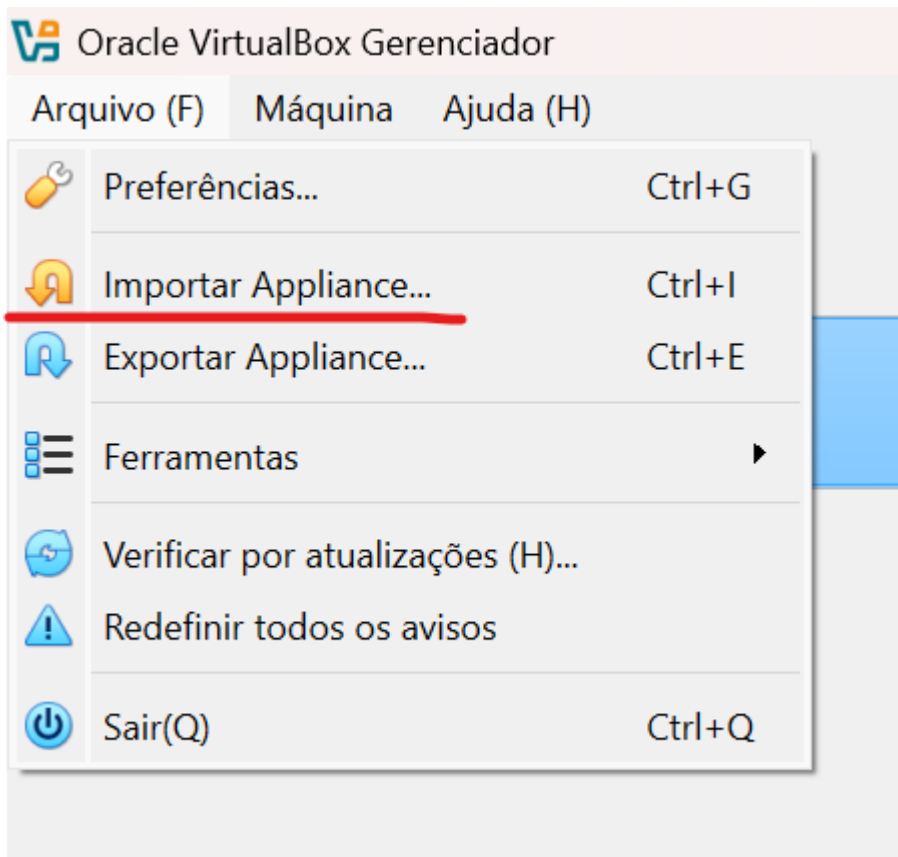
Procedimento:

1. Instalação do VirtualBox

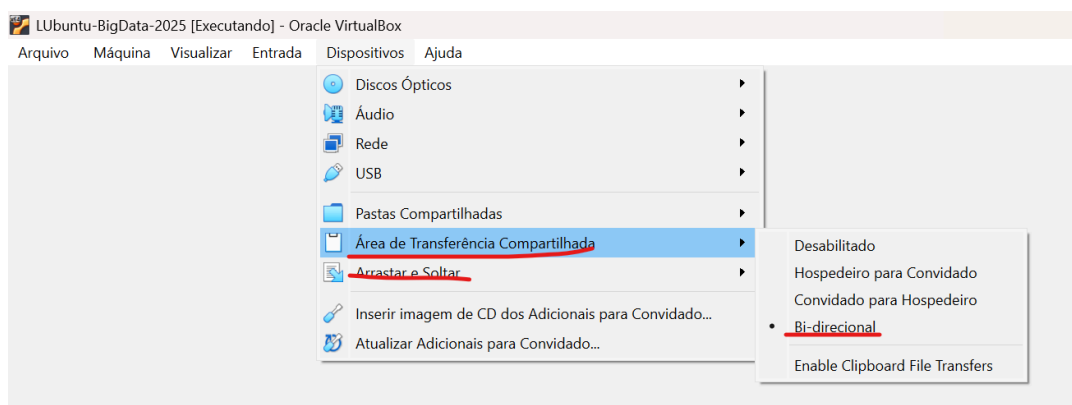
Primeiramente foi realizado download e a instalação do software VirtualBox para podermos rodar a máquina virtual disponibilizada pelo professor.

2. Importação e configuração da imagem da máquina virtual

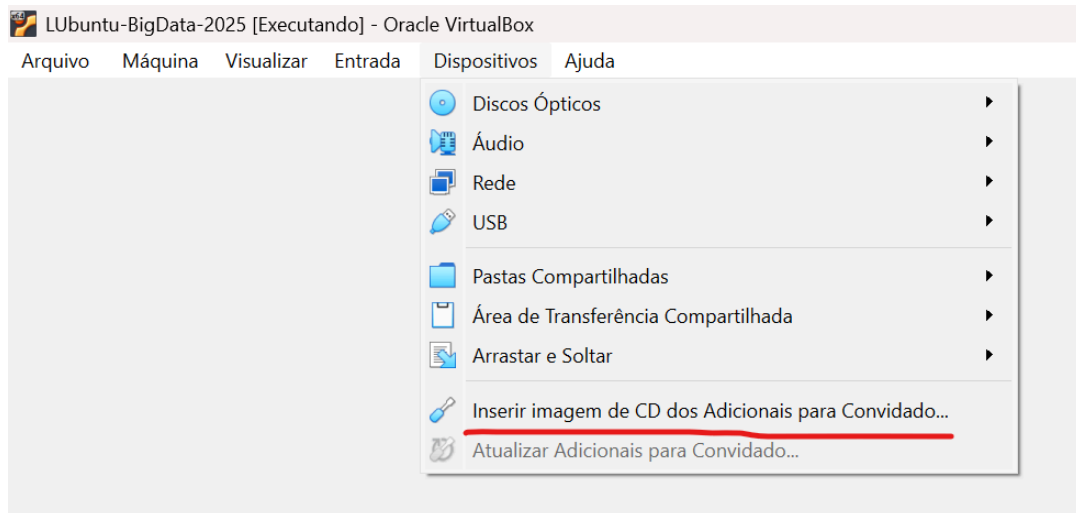
- Importar a imagem da máquina virtual para o VirtualBox;



- Iniciar máquina virtual;
- Para facilitar a configuração do ambiente, foi habilitado as funcionalidades **Área de Transferência Compartilhada** e **Arrastar e Soltar**;
 - Para isso, primeiramente foi habilitado as devidas opções no software VirtualBox;



- Após isso foi necessário instalar alguns pacotes adicionais na máquina virtual;
 - Inserir imagem de CD dos Adicionais para Convidados



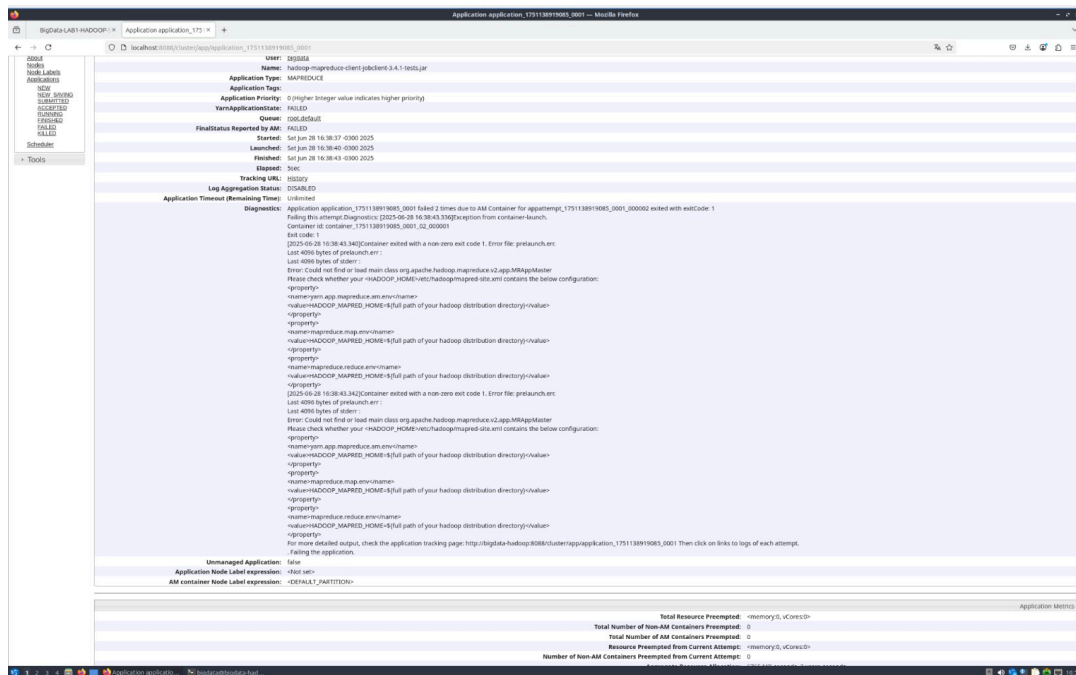
- Executar comando `$ sudo`

`/media/<USER>/VBox_GAs_7.1.10/VBoxLinuxAdditions.run`

```
Arquivo  Ações  Editar  Exibir  Ajuda
bigdata@bigdata-hadoop: /media/bigdata/VBox_GAs_7.1.10 x
bigdata@bigdata-hadoop: /media/bigdata/VBox_GAs_7.1.10$ pwd
/media/bigdata/VBox_GAs_7.1.10
bigdata@bigdata-hadoop: /media/bigdata/VBox_GAs_7.1.10$ sudo ./VBoxLinuxAdditions.run
[sudo] senha para bigdata:
Verifying archive integrity... 100% MD5 checksums are OK. All good.
Uncompressing VirtualBox 7.1.10 Guest Additions for Linux 100%
VirtualBox Guest Additions installer
Removing installed version 7.1.10 of VirtualBox Guest Additions...
update-initramfs: Generating /boot/initrd.img-6.11.0-28-generic
VirtualBox Guest Additions: Starting.
VirtualBox Guest Additions: Setting up modules
VirtualBox Guest Additions: Building the VirtualBox Guest Additions kernel
modules. This may take a while.
VirtualBox Guest Additions: To build modules for other installed kernels, run
VirtualBox Guest Additions: /sbin/rcvboxadd quicksetup <version>
VirtualBox Guest Additions: or
VirtualBox Guest Additions: /sbin/rcvboxadd quicksetup all
VirtualBox Guest Additions: Building the modules for kernel 6.11.0-28-generic.
update-initramfs: Generating /boot/initrd.img-6.11.0-28-generic
bigdata@bigdata-hadoop: /media/bigdata/VBox_GAs_7.1.10$
```

3. Instalação do Hadoop

Devido a diferentes versões do manual de instalação do hadoop (um disponibilizado no Moodle e outro pelo Whatsapp), foi encontrado o seguinte problema no momento de rodar o teste `$ hadoop jar hadoop-mapreduce-client-jobclient-3.4.1-tests.jar TestDFSIO -write -nrFiles 5 -fileSize 10`



O erro era devido a configuração do arquivo **yarn-site.xml**.

A solução do problema foi na edição correta do arquivo **yarn-site.xml**.

Abaixo destaco o trecho do arquivo que corrigiu o problema:

yarn-site.xml (dentro da tag **configuration**):

```
<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>
<property>
  <name>yarn.nodemanager.env-whitelist</name>
  <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLAS
SPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_HOME,PATH,LANG,TZ,HADOOP
MAPRED_HOME</value>
</property>
```

4. Evidências teste TestDFSIO

Segue abaixo imagens evidenciando o sucesso da instalação do hadoop:

```
bigdata@bigdata-hadoop: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-client-jobclient-3.4.1-tests.jar TestDFSIO -write -nrFiles 3 -fileSize 100
2025-06-30 20:15:50,534 INFO fs.TestDFSIO: TestDFSIO.1.8
2025-06-30 20:15:50,535 INFO fs.TestDFSIO: nrFiles = 3
2025-06-30 20:15:50,536 INFO fs.TestDFSIO: nrBytes (MB) = 100.0
2025-06-30 20:15:50,536 INFO fs.TestDFSIO: bufferSize = 1000000
2025-06-30 20:15:50,536 INFO fs.TestDFSIO: baseDir = /benchmarks/TestDFSIO
2025-06-30 20:15:50,726 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2025-06-30 20:15:51,640 INFO fs.TestDFSIO: creating control file: 104857600 bytes, 3 files
2025-06-30 20:15:52,827 INFO fs.TestDFSIO: created control files for: 3 files
2025-06-30 20:15:52,957 INFO client.DefaultHARMFaloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-06-30 20:15:53,428 INFO client.DefaultHARMFaloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-06-30 20:15:54,049 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bigdata/.staging/job_1751324424014_0001
2025-06-30 20:15:54,799 INFO mapreduce.JobSubmitter: number of splits:3
2025-06-30 20:15:55,086 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1751324424014_0001
2025-06-30 20:15:55,086 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-06-30 20:15:55,352 INFO conf.Configuration: resource-types.xml not found
2025-06-30 20:15:55,352 INFO resource.ResourceUtil: Unable to find 'resource-types.xml'.
2025-06-30 20:15:56,081 INFO impl.YarnClientImpl: Submitted application application_1751324424014_0001
2025-06-30 20:15:56,200 INFO mapreduce.Job: The url to track the job: http://bigdata-hadoop:8088/proxy/application_1751324424014_0001/
2025-06-30 20:15:56,205 INFO mapreduce.Job: Running job: job_1751324424014_0001
2025-06-30 20:16:06,446 INFO mapreduce.Job: Job job_1751324424014_0001 running in uber mode : false
2025-06-30 20:16:06,448 INFO mapreduce.Job: map 0% reduce 0%
2025-06-30 20:16:15,659 INFO mapreduce.Job: map 100% reduce 0%
2025-06-30 20:16:21,739 INFO mapreduce.Job: map 100% reduce 100%
2025-06-30 20:16:22,764 INFO mapreduce.Job: Job job_1751324424014_0001 completed successfully
2025-06-30 20:16:22,856 INFO mapreduce.Job: Counters: 54
```

5. Evidências teste WordCount

Para realizar o teste do WordCount, foi utilizado o [tutorial](#).

Segue abaixo imagens evidenciando o sucesso da instalação do hadoop:

```
bigdata@bigdata-hadoop: /usr/local/hadoop/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-3.4.1.jar wordcount /count_words/input /count_words/output
2025-06-30 20:26:41,923 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2025-06-30 20:26:41,648 INFO client.DefaultHARMFaloverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2025-06-30 20:26:42,089 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/bigdata/.staging/job_1751324424014_0002
2025-06-30 20:26:42,432 INFO input.FileInputFormat: Total input files to process : 1
2025-06-30 20:26:42,941 INFO mapreduce.JobSubmitter: number of splits:1
2025-06-30 20:26:43,127 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1751324424014_0002
2025-06-30 20:26:43,127 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-06-30 20:26:43,387 INFO conf.Configuration: resource-types.xml not found
2025-06-30 20:26:43,387 INFO resource.ResourceUtil: Unable to find 'resource-types.xml'.
2025-06-30 20:26:43,535 INFO impl.YarnClientImpl: Submitted application application_1751324424014_0002
2025-06-30 20:26:43,590 INFO mapreduce.Job: The url to track the job: http://bigdata-hadoop:8088/proxy/application_1751324424014_0002/
2025-06-30 20:26:43,591 INFO mapreduce.Job: Running job: job_1751324424014_0002
2025-06-30 20:26:50,886 INFO mapreduce.Job: Job job_1751324424014_0002 running in uber mode : false
2025-06-30 20:26:50,887 INFO mapreduce.Job: map 0% reduce 0%
2025-06-30 20:26:56,021 INFO mapreduce.Job: map 100% reduce 0%
2025-06-30 20:27:06,265 INFO mapreduce.Job: map 100% reduce 100%
2025-06-30 20:27:06,354 INFO mapreduce.Job: Job job_1751324424014_0002 completed successfully
2025-06-30 20:27:06,354 INFO mapreduce.Job: Counters: 54
```