

Ambiente Hadoop

Parte 1

- Criar uma máquina virtual
- Instalar o sistema operacional Linux
- Instalar utilitários (Java, ssh, ferramentas)
- Instalar o MySQL

Atividades

Atividades

Parte 2

- Instalar o Hadoop
- Configurar o HDFS e o MapReduce
- Executar um job MapReduce no HDFS
- Instalar e configurar o Zookeeper
- Instalar e configurar o Hbase
- Instalar e configurar o Hive
- Instalar e configurar o Pig
- Instalar e configurar o Sqoop
- Instalar e configurar o Spark
- Instalar e configurar o Flume

Atividades

- Instalar e configurar a máquina virtual Cloudera
- Instalar e configurar a máquina virtual Hortonworks
- Instalar e configurar o Hadoop com Docker

Modos de Execução do Hadoop

- core-site.xml
- hdfs-site.xml
- mapred-site.xml

Modo Local
(Standalone)

Modo Pseudo-Distribuído
(Pseudo-Distributed)


Modo Totalmente Distribuído
(Fully Distributed)

Modo Local – parâmetros de arquivo não precisam ser alterados. Indicado para fase de desenvolvimento.

Modo Pseudo-Distribuído – aplicadas configurações semelhantes as necessárias para execução de um cluster, mas toda aplicação é executada numa máquina local, ou cluster de uma máquina só. Permite uma simulação.

Modo Totalmente Distribuído – utilizado para processamento distribuído da opção hadoop em um cluster de computadores, demanda edição dos arquivos de instalação para indicar máquinas que vão executar cada componente.

- core-site.xml
- hdfs-site.xml
- mapred-site.xml



Component	Property	Standalone	Pseudo-distributed	Fully distributed
Core	fs.default.name	file:/// (default)	hdfs://localhost/	hdfs://namenode/
HDFS	dfs.replication	N/A	1	3 (default)
MapReduce	mapred.job.tracker	localhost (default)	localhost:8021	jobtracker:8021

<https://www.virtualbox.org/>

download, Windows hosts,

<https://www.centos.org/>

cópia do red hat, mas totalmente gratuita

get centos now, dvd iso

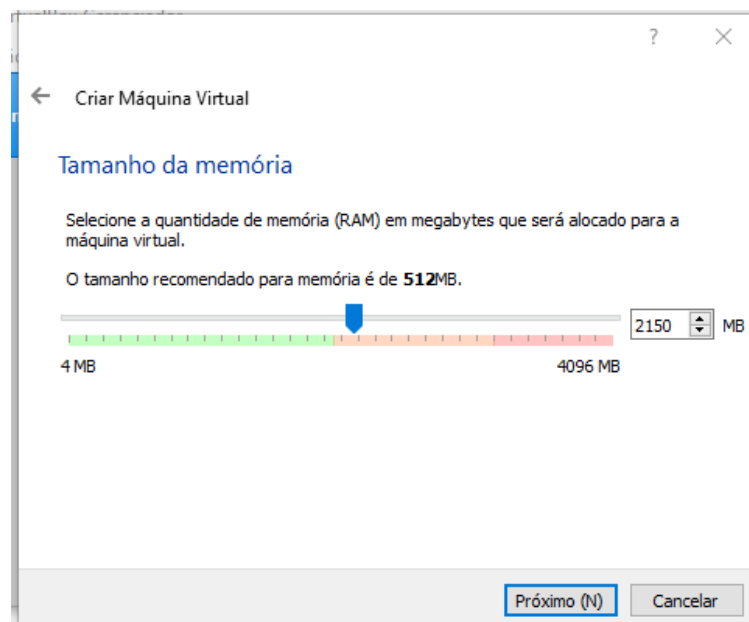
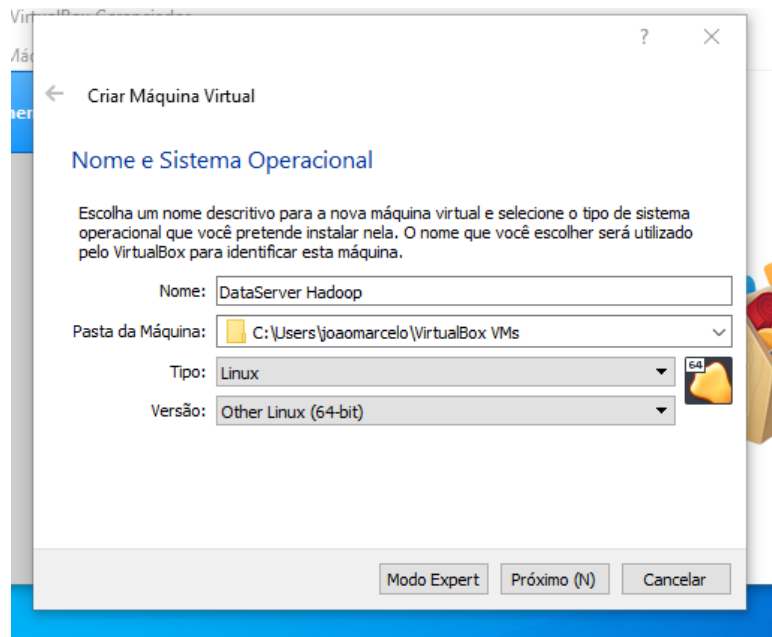


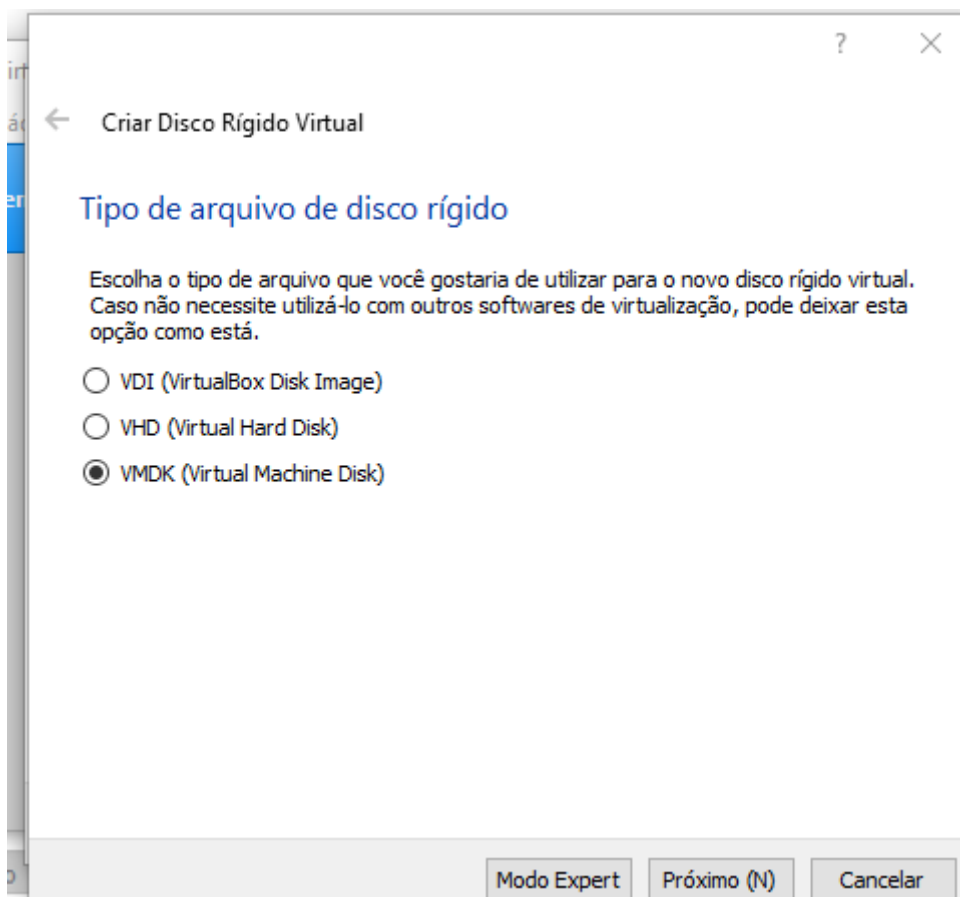
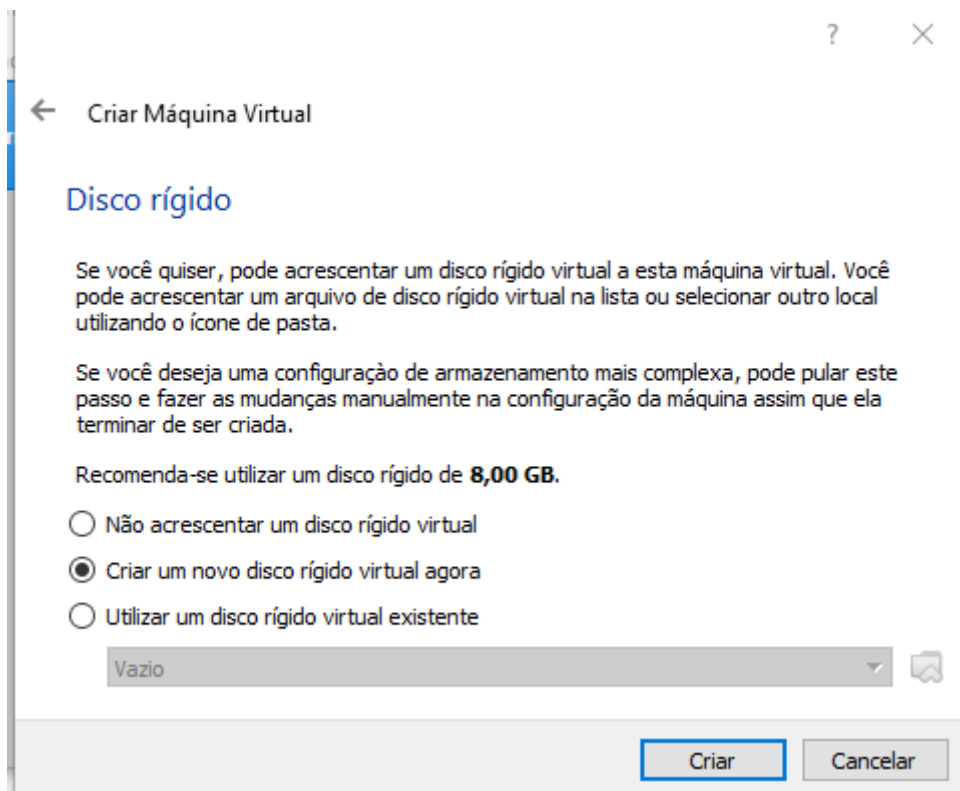
Ubuntu é bom para ambiente desktop, para ambiente servidor o centos é a melhor opção.

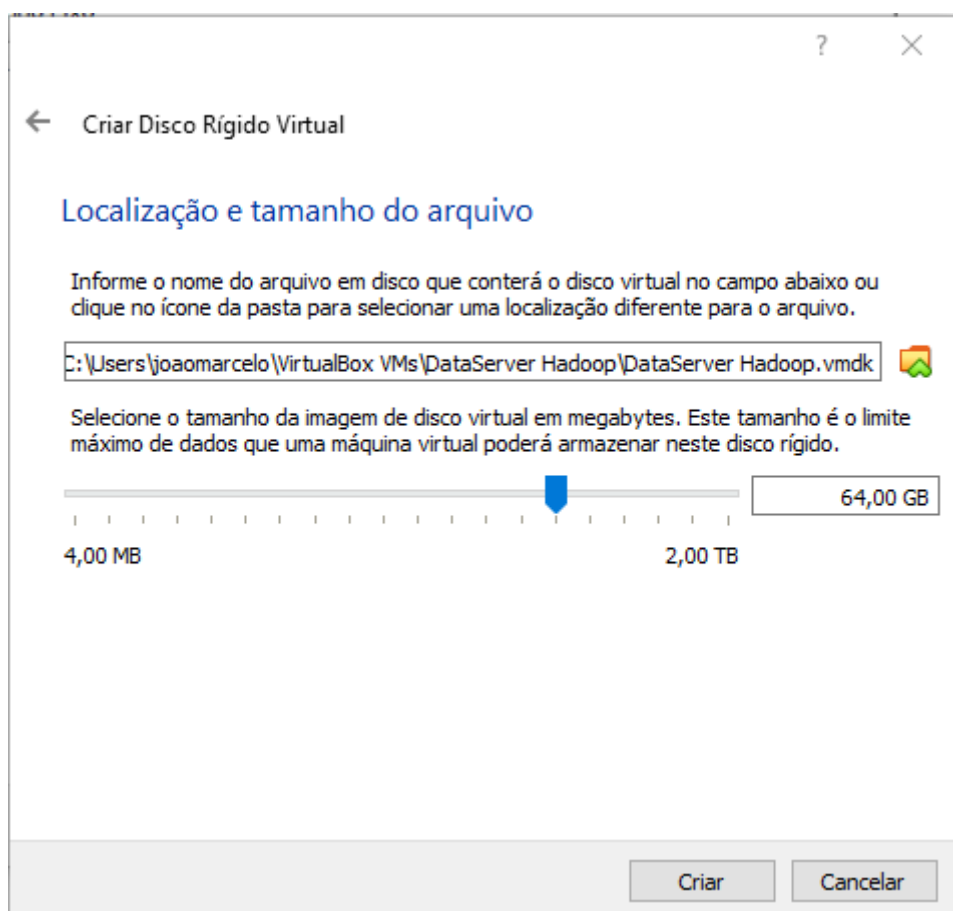
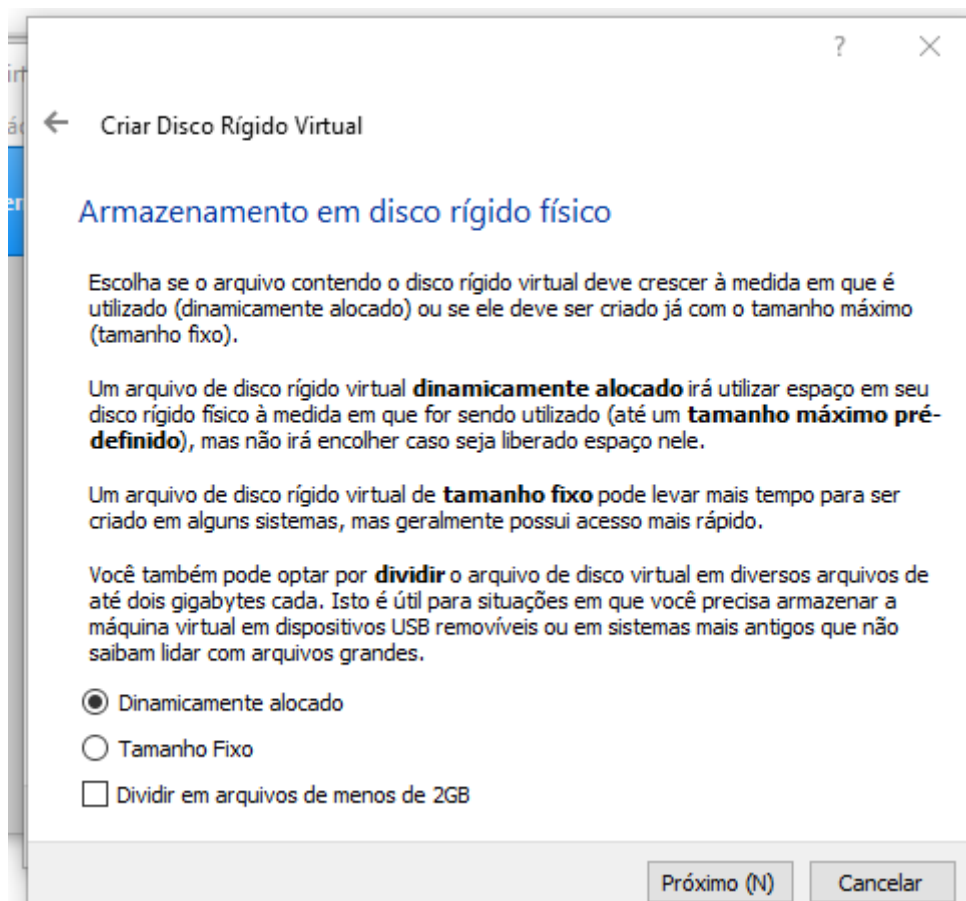
Baixando o virtual box clica em new

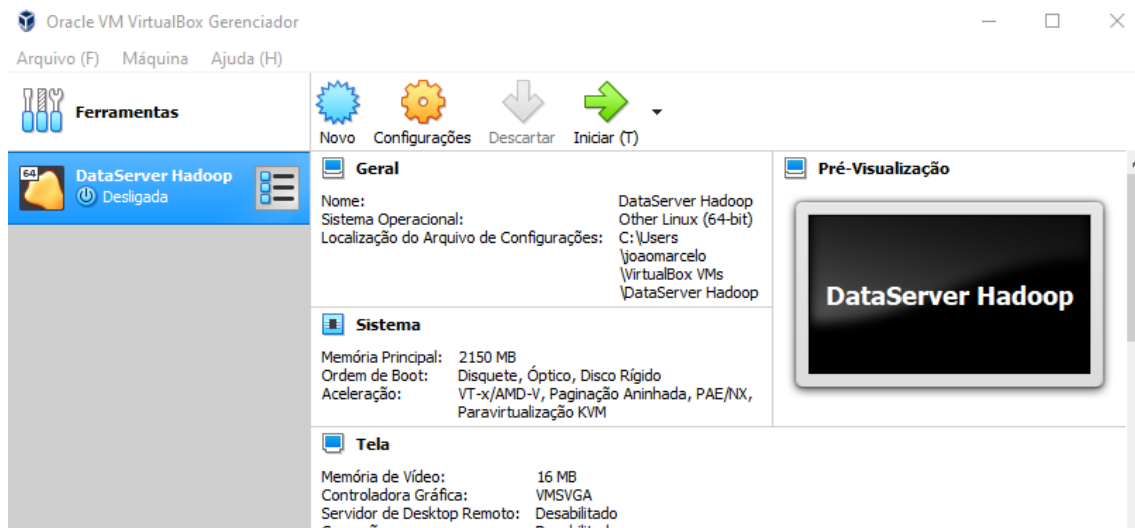
- **VirtualBox 6.0.8** (released May 13 2019)
 - ➞ Windows hosts
 - ➞ OS X hosts
 - ➞ Solaris hosts
 - Linux Hosts:

Versão utilizada

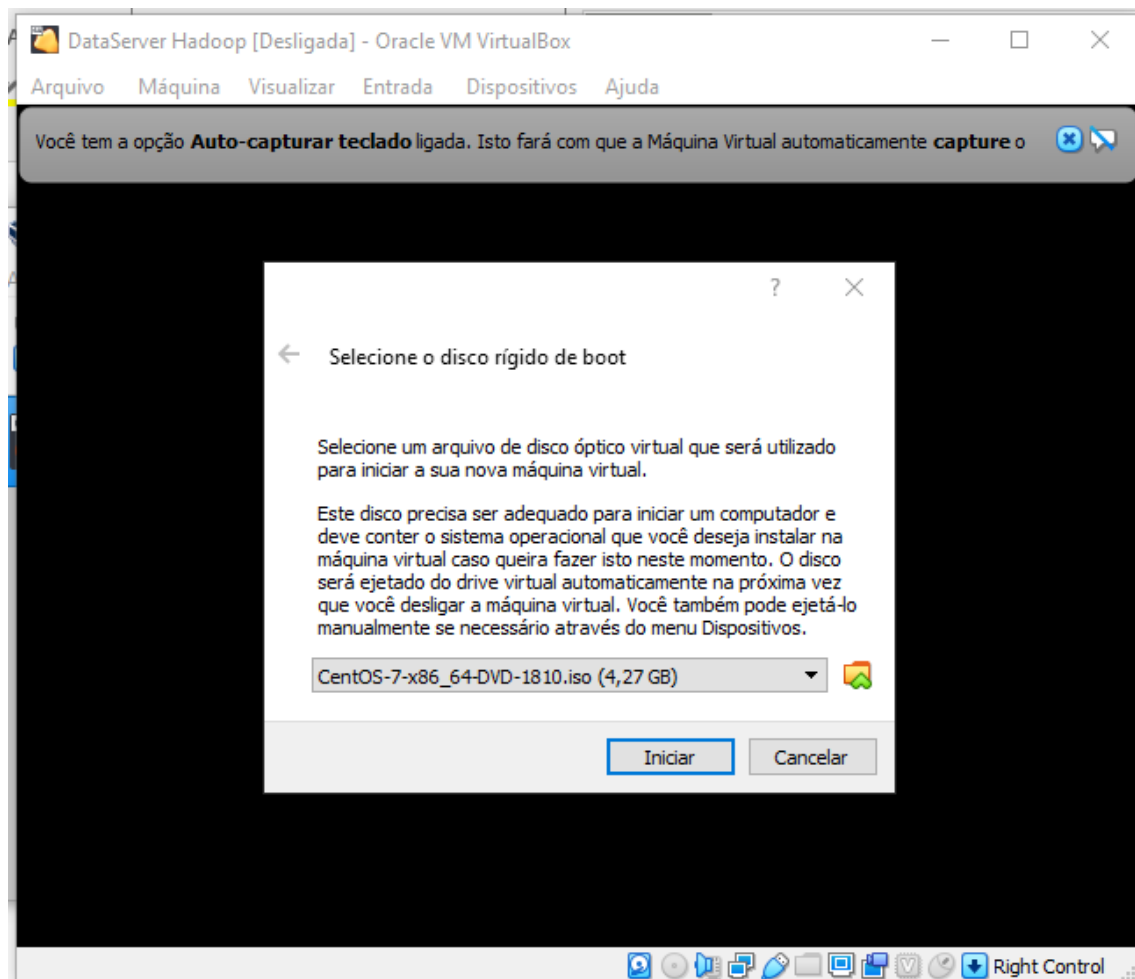








Clica no botão iniciar



SOFTWARE SELECTION

CENTOS 7 INSTALLATION

Done

us

Help!

Base Environment

☐ Minimal Install
Basic functionality.

☐ Compute Node
Installation for performing computation and processing.

☐ Infrastructure Server
Server for operating network infrastructure services.

☐ File and Print Server
File, print, and storage server for enterprises.

☐ Basic Web Server
Server for serving static and dynamic internet content.

☐ Virtualization Host
Minimal virtualization host.

☒ Server with GUI
Server for operating network infrastructure services, with a GUI.

☐ GNOME Desktop
GNOME is a highly intuitive and user friendly desktop environment.

☐ KDE Plasma Workspaces
The KDE Plasma Workspaces, a highly-configurable graphical user interface which includes a panel, desktop, system icons and desktop widgets, and many powerful KDE applications.

Add-Ons for Selected Environment

☐ Backup Server
Software to centralize your infrastructure's backups.

☐ DNS Name Server
This package group allows you to run a DNS name server (BIND) on the system.

☐ E-mail Server
Allows the system to act as a SMTP and/or IMAP e-mail server.

☐ FTP Server
Allows the system to act as an FTP server.

☐ File and Storage Server
CIFS, SMB, NFS, iSCSI, iSER, and iSNS network storage server.

☐ Hardware Monitoring Utilities
A set of tools to monitor server hardware.

☐ High Availability
Infrastructure for highly available services and/or shared storage.

☐ Identity Management Server
Centralized management of users, servers and authentication policies.

☐ Infiniband Support
Software designed for supporting clustering and grid

Arquivo

Máquina

Visualizar

Entrada

Dispositivos

Ajuda

NETWORK & HOST NAME

CENTOS 7 INSTALLATION

Done

us

Help!

Ethernet (enp0s3)
Intel Corporation 82540EM Gigabit Ethernet Controller (1000 Mb/s)

+

-

Ethernet (enp0s3)
Connected

Hardware Address 08:00:27:CC:D8:6A
Speed 1000 Mb/s
IP Address 10.0.2.15
Subnet Mask 255.255.255.0
Default Route 10.0.2.2
DNS 177.85.164.10 8.8.8.8

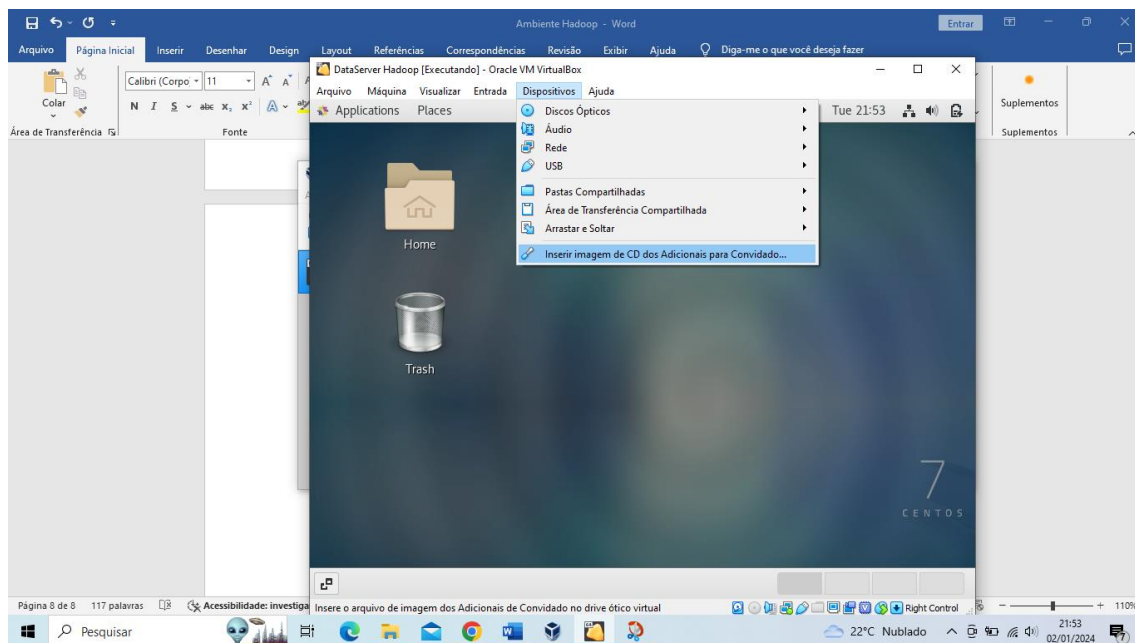
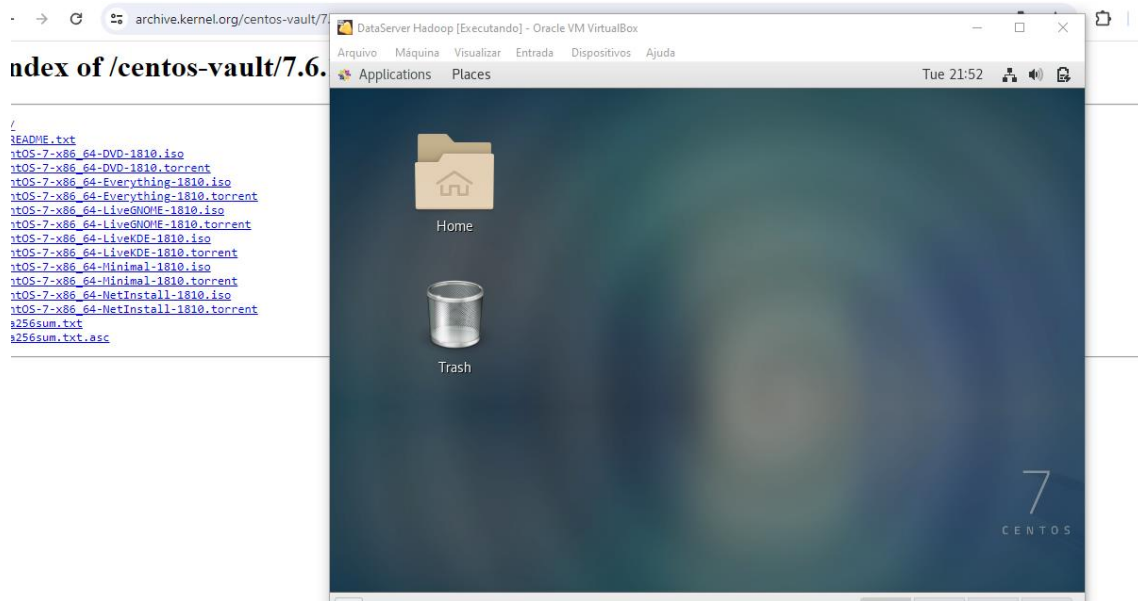
Configure...

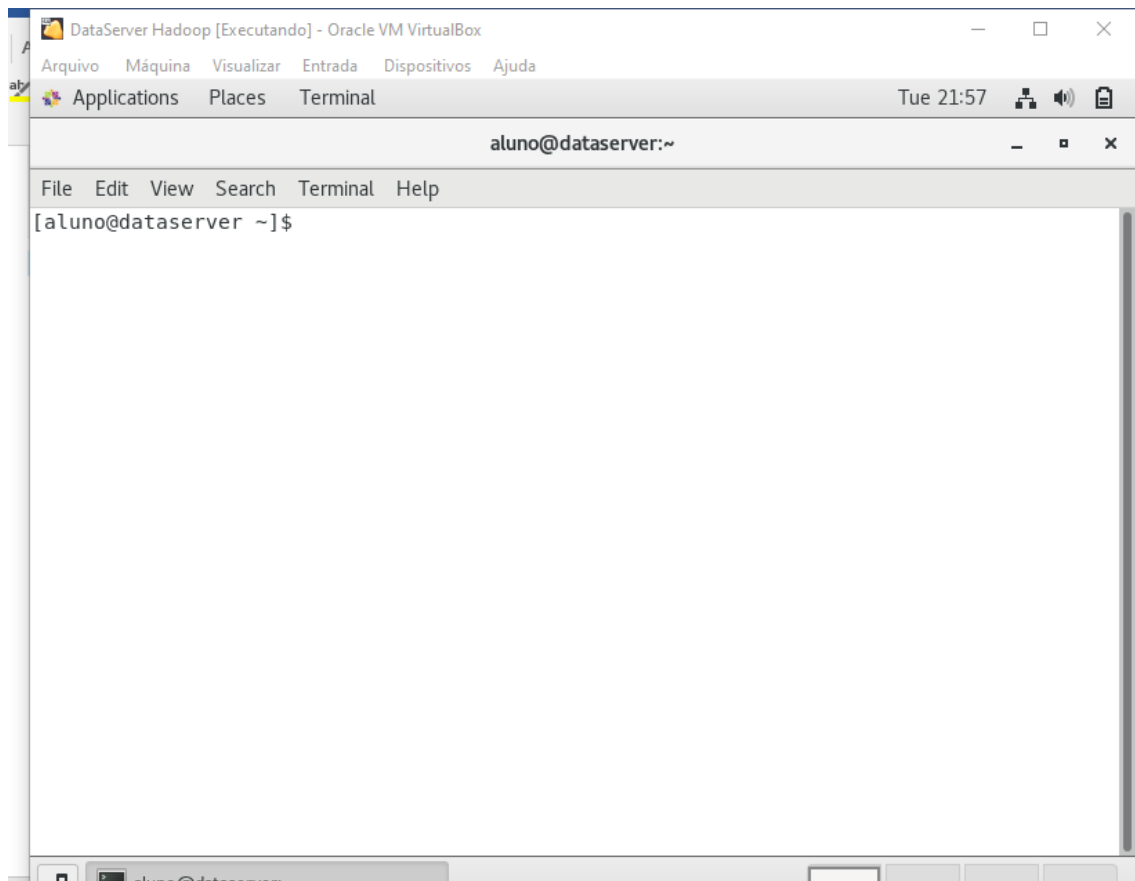
Host name: dataserver

Apply

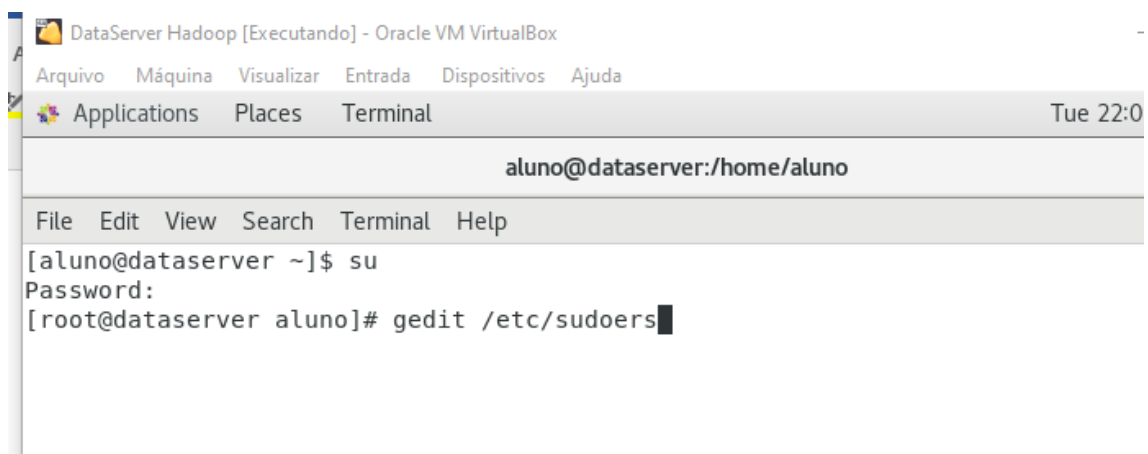
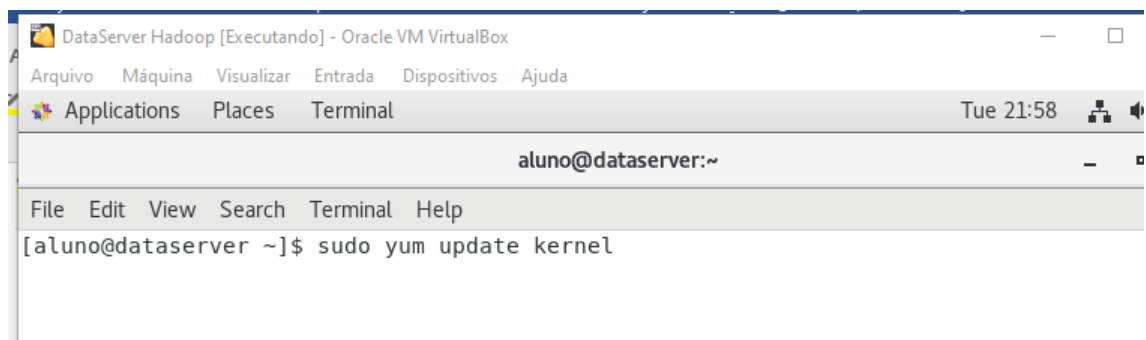
Current host name: dataserver

Right Control





Abre o terminal, em applications



Inclui aluno e salva

Arquivo Máquina Visualizar Entrada Dispositivos Ajuda

Applications Places gedit Tue 22:02

Open ***sudoers** Save

```
# Adding HOME to env_keep may enable a user to run unrestricted
# commands via sudo.
#
# Defaults    env_keep += "HOME"

Defaults    secure_path = /sbin:/bin:/usr/sbin:/usr/bin

## Next comes the main part: which users can run what software on
## which machines (the sudoers file can be shared between multiple
## systems).
## Syntax:
##
##      user    MACHINE=COMMANDS
##
## The COMMANDS section may have other options added to it.
##
## Allow root to run any commands anywhere
root    ALL=(ALL)    ALL
aluno    ALL=(ALL)    ALL

## Allows members of the 'sys' group to run networking, software,
## service management apps and more.
# %sys ALL = NETWORKING, SOFTWARE, SERVICES, STORAGE, DELEGATING, PROCESSES, LOCATE,
DRIVERS

## Allows people in group wheel to run all commands
```

Applications Places Terminal Tue 22:03

aluno@dataserver:~

File Edit View Search Terminal Help

```
connection is closed
(gedit:12013): dconf-WARNING **: 22:01:18.376: failed to commit changes to dconf: The c
onnection is closed

** (gedit:12013): WARNING **: 22:02:52.483: Set document metadata failed: Setting attri
bute metadata::gedit-spell-language not supported

** (gedit:12013): WARNING **: 22:02:52.503: Set document metadata failed: Setting attri
bute metadata::gedit-encoding not supported

** (gedit:12013): WARNING **: 22:02:56.822: Set document metadata failed: Setting attri
bute metadata::gedit-spell-language not supported

** (gedit:12013): WARNING **: 22:02:56.826: Set document metadata failed: Setting attri
bute metadata::gedit-encoding not supported

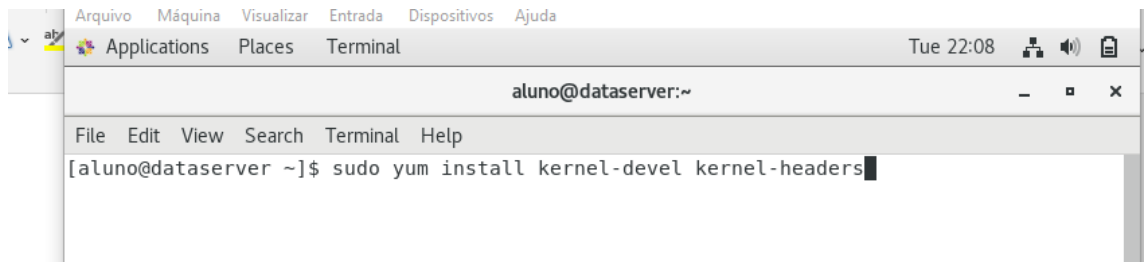
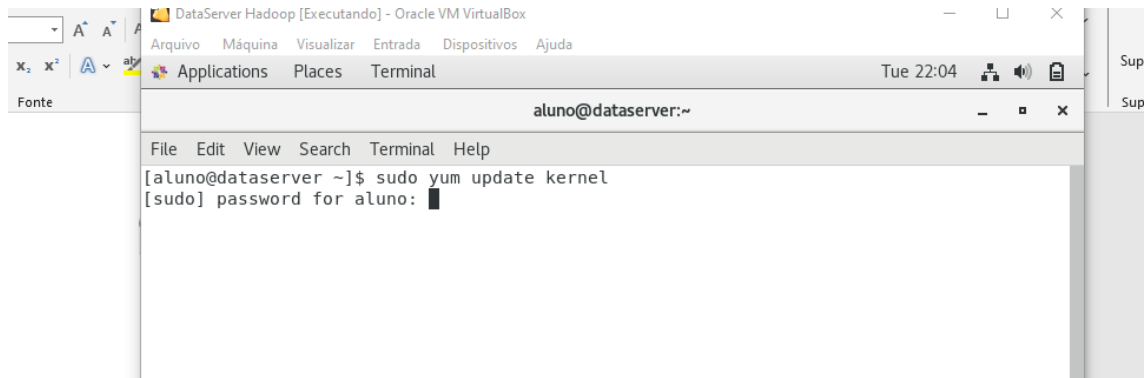
** (gedit:12013): WARNING **: 22:02:57.934: Set document metadata failed: Setting attri
bute metadata::gedit-position not supported

(gedit:12013): dconf-WARNING **: 22:02:57.965: failed to commit changes to dconf: The c
onnection is closed
[root@dataserver aluno]# exit
exit
[aluno@dataserver ~]$ clear
```

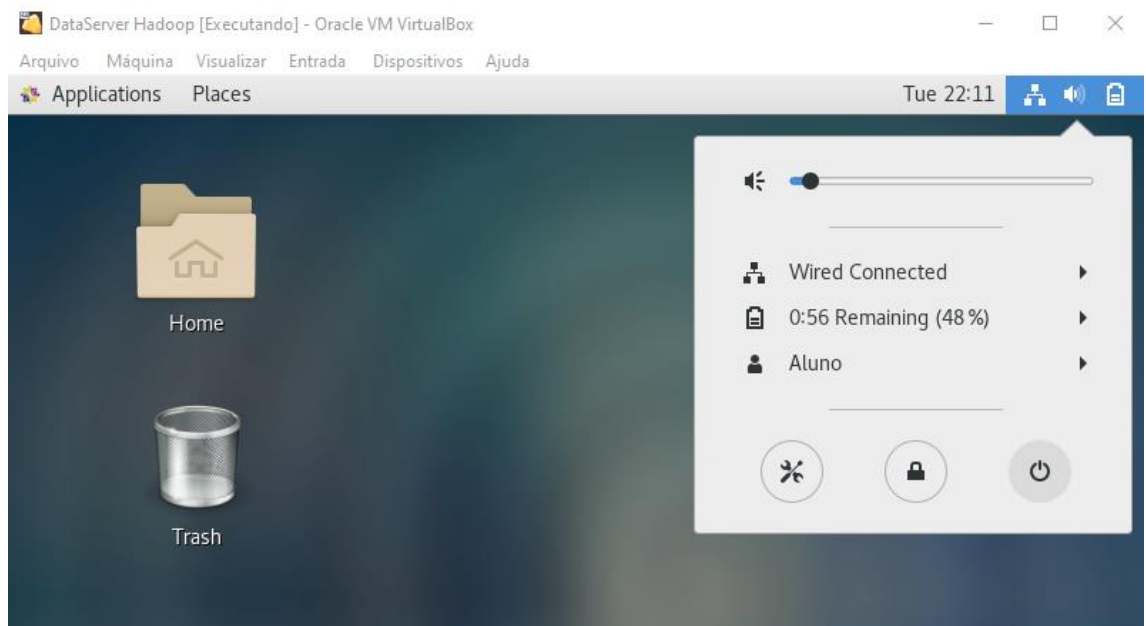
aluno@dataserver:~

investig

Right Control



Após atualizar o kernel, fazer o boot na virtual machine



```
File Edit View Search Terminal Help
Verifying archive integrity... All good.
Uncompressing VirtualBox 6.0.8 Guest Additions for Linux.....
VirtualBox Guest Additions installer
Removing installed version 6.0.8 of VirtualBox Guest Additions...
Copying additional installer modules ...
Installing additional modules ...
VirtualBox Guest Additions: Starting.
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
3.10.0-1160.105.1.el7.x86_64.

This system is currently not set up to build kernel modules.
Please install the gcc make perl packages from your distribution.
modprobe vboxguest failed
The log file /var/log/vboxadd-setup.log may contain further information.
Press Return to close this window...
```

Precisamos instalar os pacotes

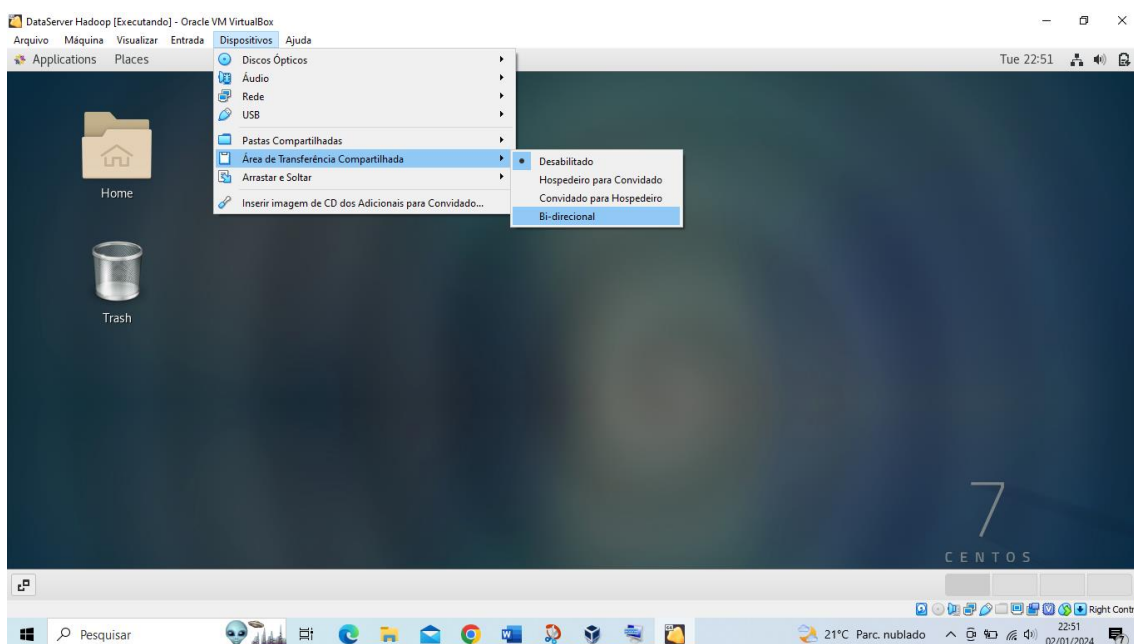
Mas antes temos de ejectar o disco

```
Applications Places Terminal Tue 22:20
aluno@dataserver:~
File Edit View Search Terminal Help
[aluno@dataserver ~]$ sudo yum install gcc make perl
```

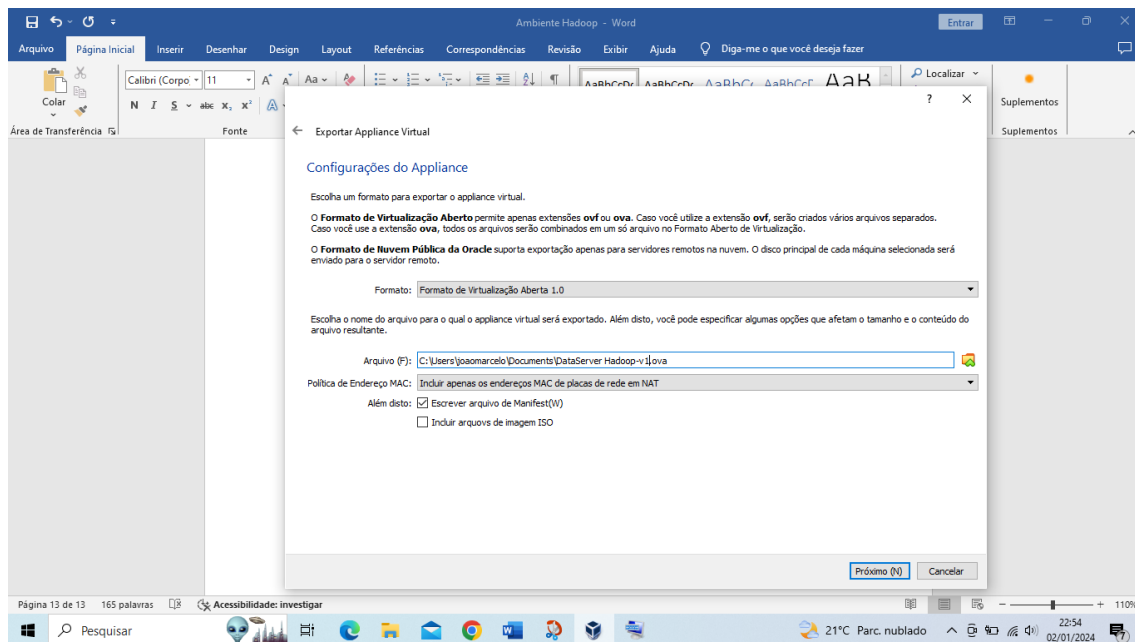
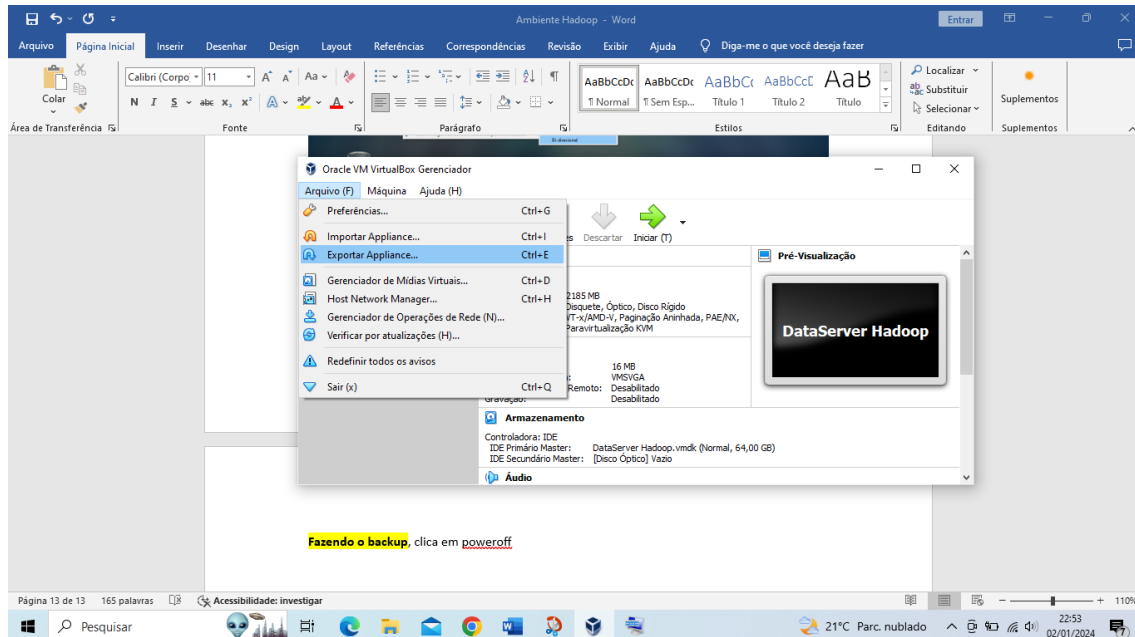
Após isso reiniciar mais uma vez

Reinicia novamente

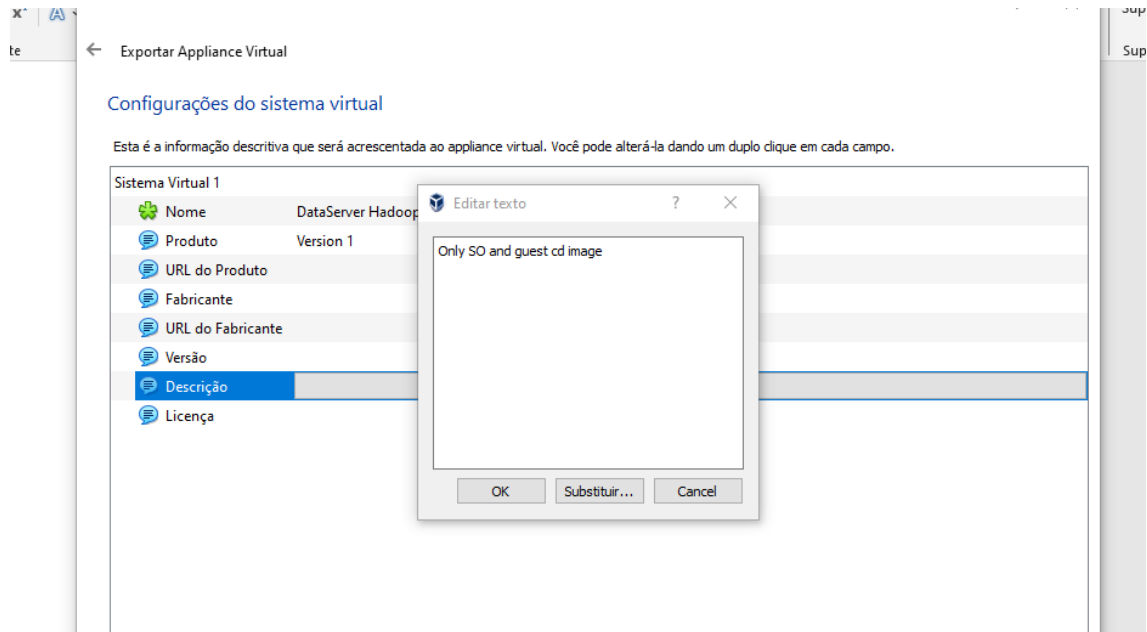
Da eject no cd



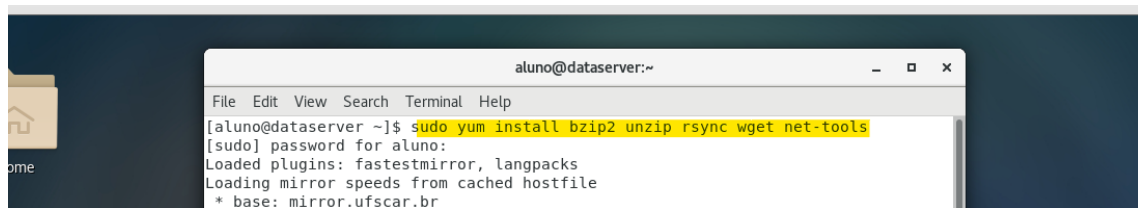
Fazendo o backup, clica em poweroff



inlui -v1



Instalando utilitários do SO



Instalando Banco de Dados SQL

sudo yum localinstall <https://dev.mysql.com/get/mysql80-community-release-el7-1.noarch.rpm>

sudo yum install mysql-community-server

entra como root su

rpm --import <https://repo.mysql.com/RPM-GPG-KEY-mysql-2022>

sudo systemctl enable mysqld

sudo systemctl start mysqld

sudo systemctl status mysqld

Instalando e configurando o SSH security shell

Protocolo de comunicação entre as máquinas do cluster

Sudo yum install openssh-server openssh-clients

Sudo systemctl enable sshd

```
aluno@dataserver:~  
File Edit View Search Terminal Help  
[aluno@dataserver ~]$ sudo systemctl enable sshd  
[aluno@dataserver ~]$ sudo systemctl start sshd  
[aluno@dataserver ~]$ sudo systemctl status sshd  
● sshd.service - OpenSSH server daemon  
   Loaded: loaded (/usr/lib/systemd/system/sshd.service; enabled; vendor preset: enabled)  
   Active: active (running) since Wed 2024-01-03 21:25:30 -03; 1min 19s ago  
     Docs: man:sshd(8)  
           man:sshd_config(5)  
  Main PID: 3330 (sshd)  
    CGroup: /system.slice/sshd.service  
            └─3330 /usr/sbin/sshd -D  
  
Jan 03 21:25:30 dataserver systemd[1]: Stopped OpenSSH server daemon.  
Jan 03 21:25:30 dataserver systemd[1]: Starting OpenSSH server daemon...  
Jan 03 21:25:30 dataserver sshd[3330]: Server listening on 0.0.0.0 port 22.  
Jan 03 21:25:30 dataserver sshd[3330]: Server listening on :: port 22.  
Jan 03 21:25:30 dataserver systemd[1]: Started OpenSSH server daemon.  
[aluno@dataserver ~]$
```

Sudo gedit /etc/ssh/sshd_config

```
*sshd_config  
/etc/ssh  
Open Save  
# $OpenBSD: sshd_config,v 1.100 2016/08/15 12:32:04 naddy Exp $  
  
# This is the sshd server system-wide configuration file. See  
# sshd_config(5) for more information.  
  
# This sshd was compiled with PATH=/usr/local/bin:/usr/bin  
  
# The strategy used for options in the default sshd_config shipped with  
# OpenSSH is to specify options with their default value where  
# possible, but leave them commented. Uncommented options override the  
# default value.  
  
# If you want to change the port on a SELinux system, you have to tell  
# SELinux about this change.  
# semanage port -a -t ssh_port_t -p tcp #PORTNUMBER  
#  
Port 22  
#AddressFamily any  
ListenAddress 0.0.0.0  
#ListenAddress ::  
  
HostKey /etc/ssh/ssh_host_rsa_key  
#HostKey /etc/ssh/ssh_host_dsa_key  
HostKey /etc/ssh/ssh_host_ecdsa_key  
HostKey /etc/ssh/ssh_host_ed25519_key  
  
# Ciphers and keying  
#RekeyLimit default none
```

Tira o suspenido

```
# Authentication:
```

```
#LoginGraceTime 2m  
PermitRootLogin no  
#StrictModes yes  
#MaxAuthTries 6
```

Tira o sustenido coloca no

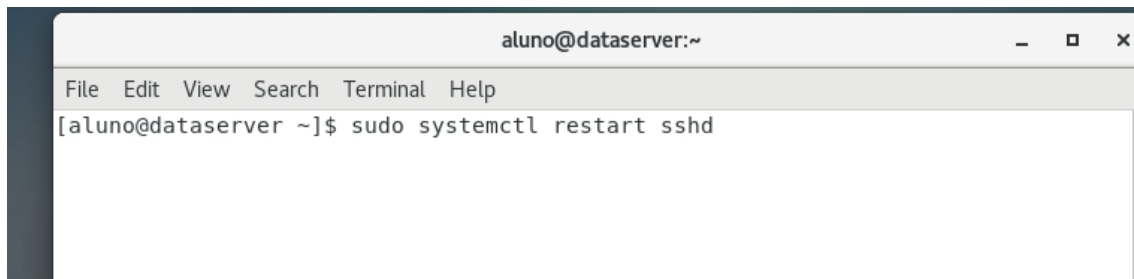
```
#LoginGraceTime 2m  
PermitRootLogin no  
#StrictModes yes  
#MaxAuthTries 6  
#MaxSessions 10  
AllowUsers aluno
```

```
#PubkeyAuthentication yes
```

```
# The default is to check both .ssh/authorized_keys and .ssh/aut  
# but this is overridden so installations will only check .ssh/a  
AuthorizedKeysFile .ssh/authorized keys
```

Inclui AllowUsers aluno

Save, fechar clear na tela

A screenshot of a terminal window titled 'aluno@dataserver:~'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The command '[aluno@dataserver ~]\$ sudo systemctl restart sshd' has been entered and executed. The terminal background is dark, and the text is light gray.

Restart

Instalando e configurando o Java JDK

java jdk é necessário para o ambiente hadoop. Apache hadoop é desenvolvido em java

java -version

remover open jdk

sudo yum list java*

clear

sudo yum -y remove java*

após remover


```
aluno@dataserver:~  
File Edit View Search Terminal Help  
[aluno@dataserver ~]$ java -version  
bash: /usr/bin/java: No such file or directory  
[aluno@dataserver ~]$
```

Baixando o oracle jdk

Java Archive Downloads - Java SE 8u211 and later - Mozilla Firefox

https://www.oracle.com/java/technologies/javase/javase8u211-later-archive-downloads.html

ORACLE Products Industries Resources Customers Partners Developers Company View Accounts Contact Sales

Windows x84 48.62 MB [server-jre-8u211-windows-x84.tar.gz](#)

Java SE Development Kit 8u212
This software is licensed under the [Oracle Technology Network License Agreement for Oracle Java SE](#)

Product / File Description	File Size	Download
Linux ARM v6/v7 Soft Float ABI	72.86 MB	jdk-8u212-linux-arm32-vfp-hflt.tar.gz
Linux ARM v6/v7 Soft Float ABI	69.77 MB	jdk-8u212-linux-arm64-vfp-hflt.tar.gz
Linux x86	174.11 MB	jdk-8u212-linux-i586.rpm

Versão 8u212 development kit

Linux x64	185.98 MB	jdk-8u212-linux-x64.tar.gz
-----------	-----------	--

Baixa este arquivo

```
aluno@dataserver:~/Downloads  
File Edit View Search Terminal Help  
[aluno@dataserver ~]$ pwd  
/home/aluno  
[aluno@dataserver ~]$ cd Downloads/  
[aluno@dataserver Downloads]$ ls  
jdk-8u212-linux-x64.tar.gz-20240104T012425Z-001.zip  
[aluno@dataserver Downloads]$
```

Descompactando o arquivo tar -xzf jdk-8u212-linux-x64.tar.gz

```
aluno@dataserver:~/Downloads
File Edit View Search Terminal Help
[aluno@dataserver ~]$ pw
bash: pw: command not found...
[aluno@dataserver ~]$ pwd
/home/aluno
[aluno@dataserver ~]$
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls
google-chrome-stable_current_x86_64.rpm jdk-8u212-linux-x64.tar.gz
[aluno@dataserver Downloads]$ tar -xzf jdk-8u212-linux-x64.tar.gz
[aluno@dataserver Downloads]$ ls
google-chrome-stable_current_x86_64.rpm jdk-8u212-linux-x64.tar.gz
jdk1.8.0_212
[aluno@dataserver Downloads]$
```

Agora precisamos fazer a configuração

Mover o diretório para outro local

```
File Edit View Search Terminal Help
[aluno@dataserver Downloads]$ clear
[aluno@dataserver Downloads]$ ls
google-chrome-stable_current_x86_64.rpm jdk-8u212-linux-x64.tar.gz
jdk1.8.0_212
[aluno@dataserver Downloads]$ cd /
[aluno@dataserver /]$ ls -la
total 24
dr-xr-xr-x. 17 root root 224 Jan 2 21:43 .
dr-xr-xr-x. 17 root root 224 Jan 2 21:43 ..
lrwxrwxrwx. 1 root root 7 Jan 2 21:25 bin -> usr/bin
dr-xr-xr-x. 5 root root 4096 Jan 2 22:30 boot
drwxr-xr-x. 20 root root 3240 Jan 3 21:18 dev
drwxr-xr-x. 136 root root 8192 Jan 3 23:00 etc
drwxr-xr-x. 3 root root 19 Jan 2 21:43 home
lrwxrwxrwx. 1 root root 7 Jan 2 21:25 lib -> usr/lib
lrwxrwxrwx. 1 root root 9 Jan 2 21:25 lib64 -> usr/lib64
drwxr-xr-x. 2 root root 6 Apr 11 2018 media
drwxr-xr-x. 2 root root 6 Apr 11 2018 mnt
drwxr-xr-x. 5 root root 62 Jan 3 23:00 opt
dr-xr-xr-x. 266 root root 0 Jan 3 21:18 proc
dr-xr-xr-x. 9 root root 4096 Jan 3 23:01 root
drwxr-xr-x. 41 root root 1300 Jan 3 23:03 run
lrwxrwxrwx. 1 root root 8 Jan 2 21:25/sbin -> usr/sbin
drwxr-xr-x. 2 root root 6 Apr 11 2018 srv
dr-xr-xr-x. 13 root root 0 Jan 3 21:17 sys
drwxrwxrwt. 34 root root 4096 Jan 3 23:16 tmp
drwxr-xr-x. 13 root root 155 Jan 2 21:25 usr
drwxr-xr-x. 20 root root 282 Jan 2 21:48 var
[aluno@dataserver /]$
```

Diretório em que vamos fazer as instalações

sudo mv jdk1.8.0_212/ /opt/jdk

```
aluno@dataserver:/opt
File Edit View Search Terminal Help
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls
google-chrome-stable_current_x86_64.rpm jdk-8u212-linux-x64.tar.gz
jdk1.8.0_212
[aluno@dataserver Downloads]$ sudo mv jdk1.8.0_212/ /opt/jdk
[sudo] password for aluno:
[aluno@dataserver Downloads]$ cd /opt/
[aluno@dataserver opt]$ ls
google jdk rh VBoxGuestAdditions-6.0.8
[aluno@dataserver opt]$
```

Configurando variável de ambiente

Gedit .bashrc



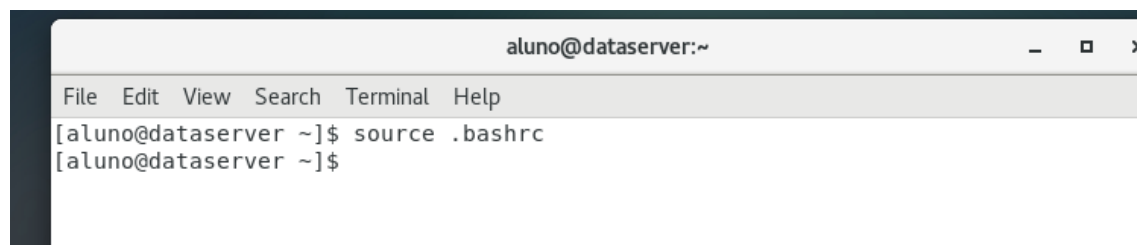
```
# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# Uncomment the following line if you don't like systemctl's auto-paging feature:
# export SYSTEMD_PAGER=

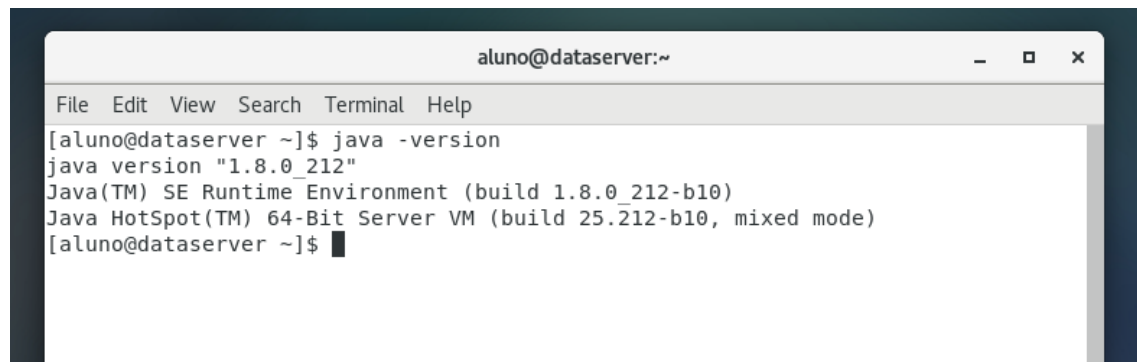
# User specific aliases and functions

# Java JDK
export JAVA_HOME=/opt/jdk
export PATH=$PATH:$JAVA_HOME/bin
```



```
aluno@dataserver:~
File Edit View Search Terminal Help
[aluno@dataserver ~]$ source .bashrc
[aluno@dataserver ~]$
```

Source .bashrc para ler as variáveis de ambiente



```
aluno@dataserver:~
File Edit View Search Terminal Help
[aluno@dataserver ~]$ java -version
java version "1.8.0_212"
Java(TM) SE Runtime Environment (build 1.8.0_212-b10)
Java HotSpot(TM) 64-Bit Server VM (build 25.212-b10, mixed mode)
[aluno@dataserver ~]$
```

Fazendo um novo backup