

Data Science Academy

www.datascienceacademy.com.br

Engenharia de Dados com Hadoop e Spark

Instalação e Configuração do Ecosistema Hadoop

Sumário

1.	Versão	4
2.	Configuração do Ambiente	5
2.1.	Criação da Máquina Virtual no VirtualBox.....	6
2.2.	Instalação do Sistema Operacional.....	15
2.3.	Instalação da Interface Gráfica Básica	30
2.4.	Instalação de Utilitários do Sistema Operacional.....	46
2.5.	Instalação do MySQL.....	57
3.	Instalação do servidor ssh	67
4.	Instalação do Java 8	81
4.1.	Instalação do JRE	81
4.2.	Instalação do JDK	84
5.	Instalação e Configuração do Hadoop	96
5.1.	Desabilitando ipv6	96
5.2.	Configuração do ssh	98
5.3.	Download e Instalação do Hadoop	108
5.3.1.	Editando o arquivo hosts.....	108
5.3.2.	Download do Hadoop	110
5.4.	Configuração do Hadoop.....	116
5.4.1.	Editar arquivos de configuração do Hadoop	116
5.4.2.	Formatando o Namenode	133
5.4.3.	Iniciando o Hadoop	135
5.4.4.	Iniciando o Yarn.....	138
5.5.	Processando Big Data.....	142
6.	Instalação e Configuração do Zookeeper.....	160
6.1.	Download e Instalação do Zookeeper	160
6.2.	Configurando do Zookeeper.....	164
7.	Instalação e Configuração do HBase.....	176
7.1.	Download e Instalação do HBase	176
7.2.	Configurando o HBase.....	180
8.	Instalação e Configuração do Hive	191
8.1.	Download e Instalação do Hive	191
8.2.	Configurando o Hive	196
9.	Instalação e Configuração do Pig.....	207
9.1.	Download e Instalação do Pig	207
9.2.	Configurando do Pig.....	212
10.	Instalação e Configuração do Spark	219
10.1.	Download e Instalação do Spark	219
11.	Instalação e Configuração do Sqoop.....	229
11.1.	Download do Sqoop.....	229
11.2.	Configuração do Sqoop	234

12.	Instalação e Configuração do Apache Flume.....	242
13.	Instalação e Configuração do Ambari (Opcional)	252

1. Versão

Este documento foi criado pela equipe Data Science Academy e pode ser distribuído livremente, desde que se faça menção à fonte.

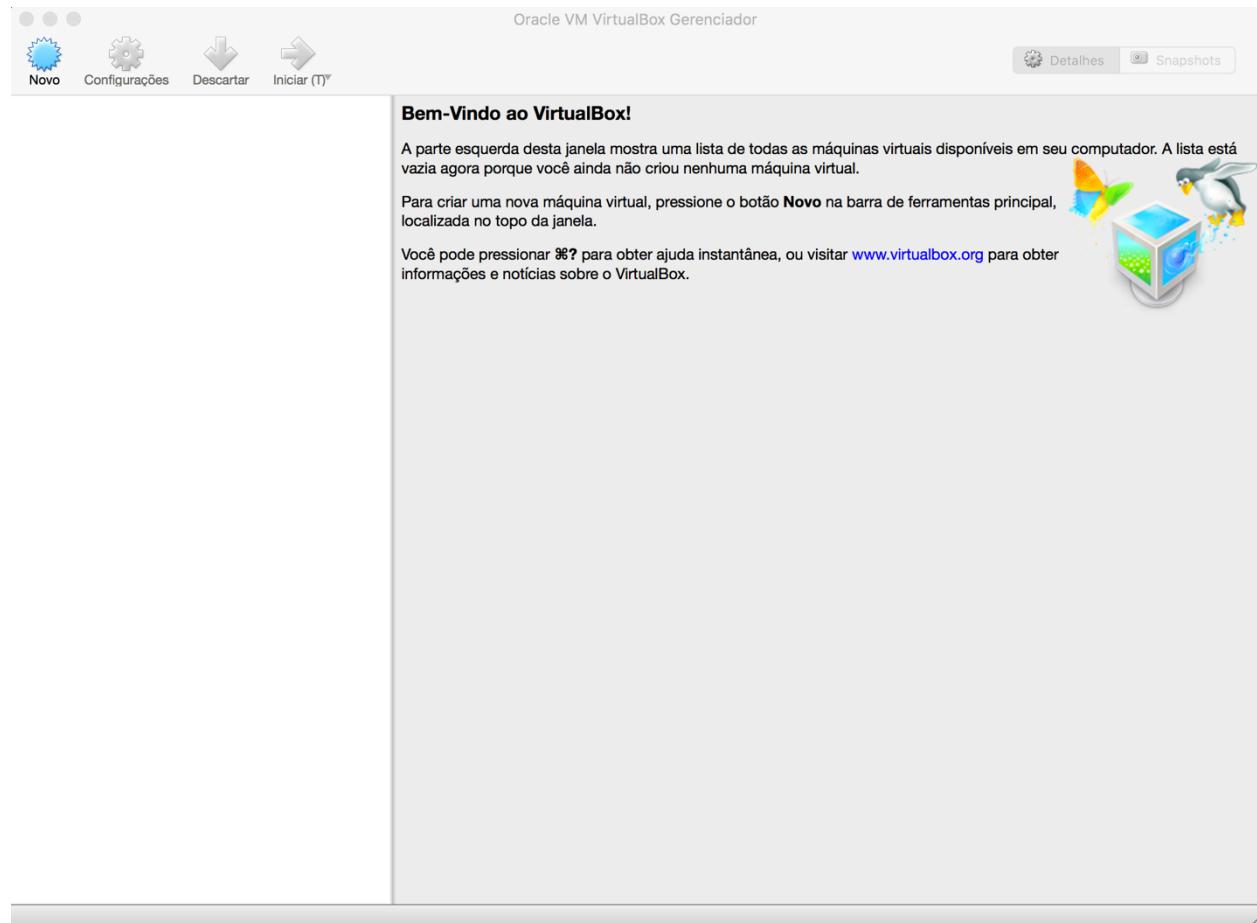
Versão	Ação	Data
1.0	Criação do documento	25/05/2016
1.1	Procedimento de instalação de mais aplicativos	29/06/2016
1.2	Procedimento de instalação do MySQL	15/07/2016
1.3	Atualização de versão de software	02/09/2016
1.4	Atualização de versão de software	25/09/2016

2. Configuração do Ambiente

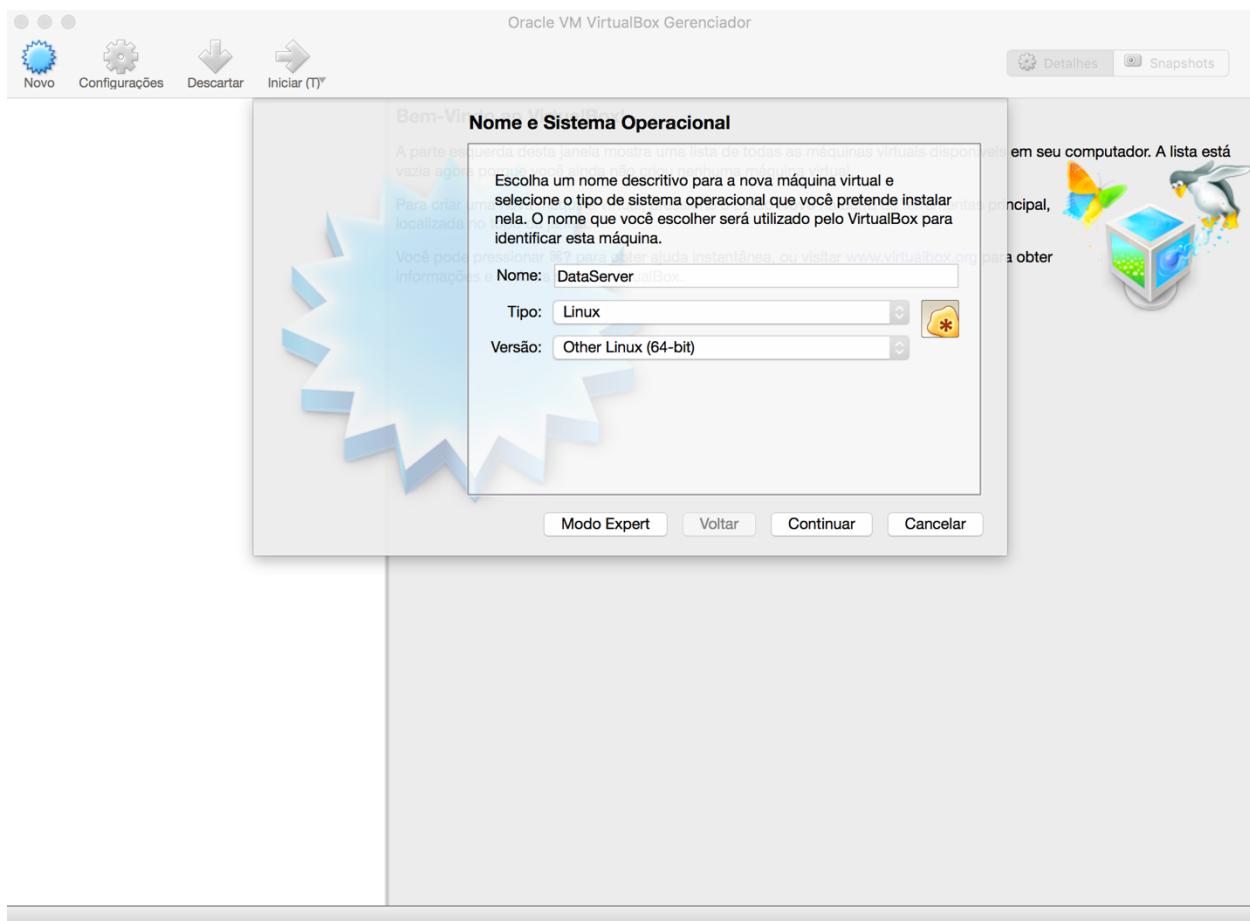
Item	Versão
Virtual Box	5.16
Sistema Operacional	CentOS 7.2 (64 bits) ou CentOS 6.8 (32 bits)
Interface Gráfica	Gnome
Firefox Web Browser	45.4
Java	1.8.0_102
Apache Hadoop	2.7.3
Apache Zookeeper	3.4.9
Apache Hbase	1.2.3
Apache Hive	2.1.0
Apache Pig	0.16.0
Apache Spark	2.0.0
Apache Sqoop	1.4.6-hadoop-2.0.4-alpha
Apache Flume	1.6.0
Apache Ambari	2.4.1

2.1. Criação da Máquina Virtual no VirtualBox

O Oracle VM Virtual Box é gratuito e pode ser baixado em <https://www.virtualbox.org>. Existem versões disponíveis para Windows, MAC, Linux e Solaris. Aqui utilizaremos a versão 5.0 e o tutorial será o mesmo independente do sistema operacional do seu computador. Caso você não esteja usando máquina virtual, pode pular direto para o item 2.2.

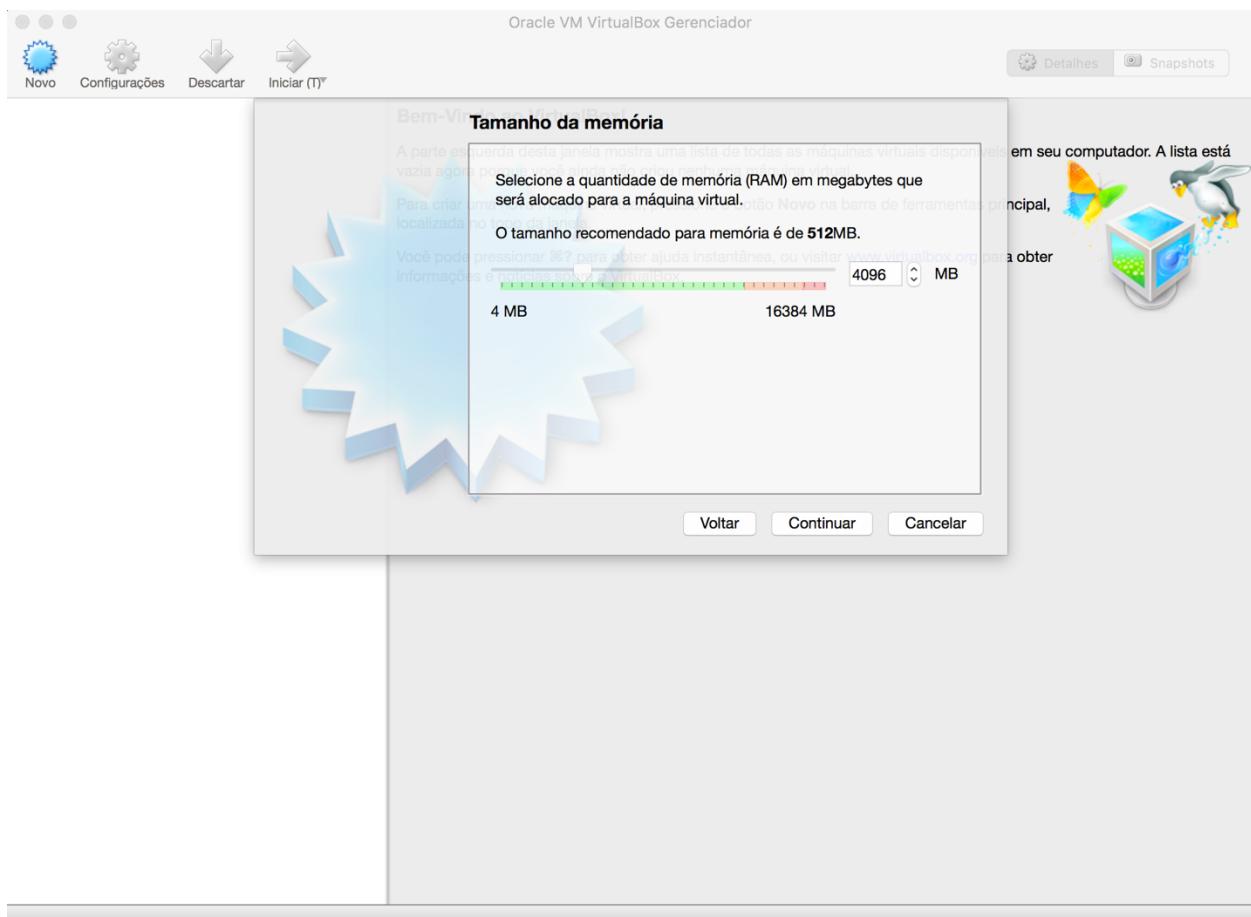


Abrindo o Gerenciador do Oracle Virtual Box

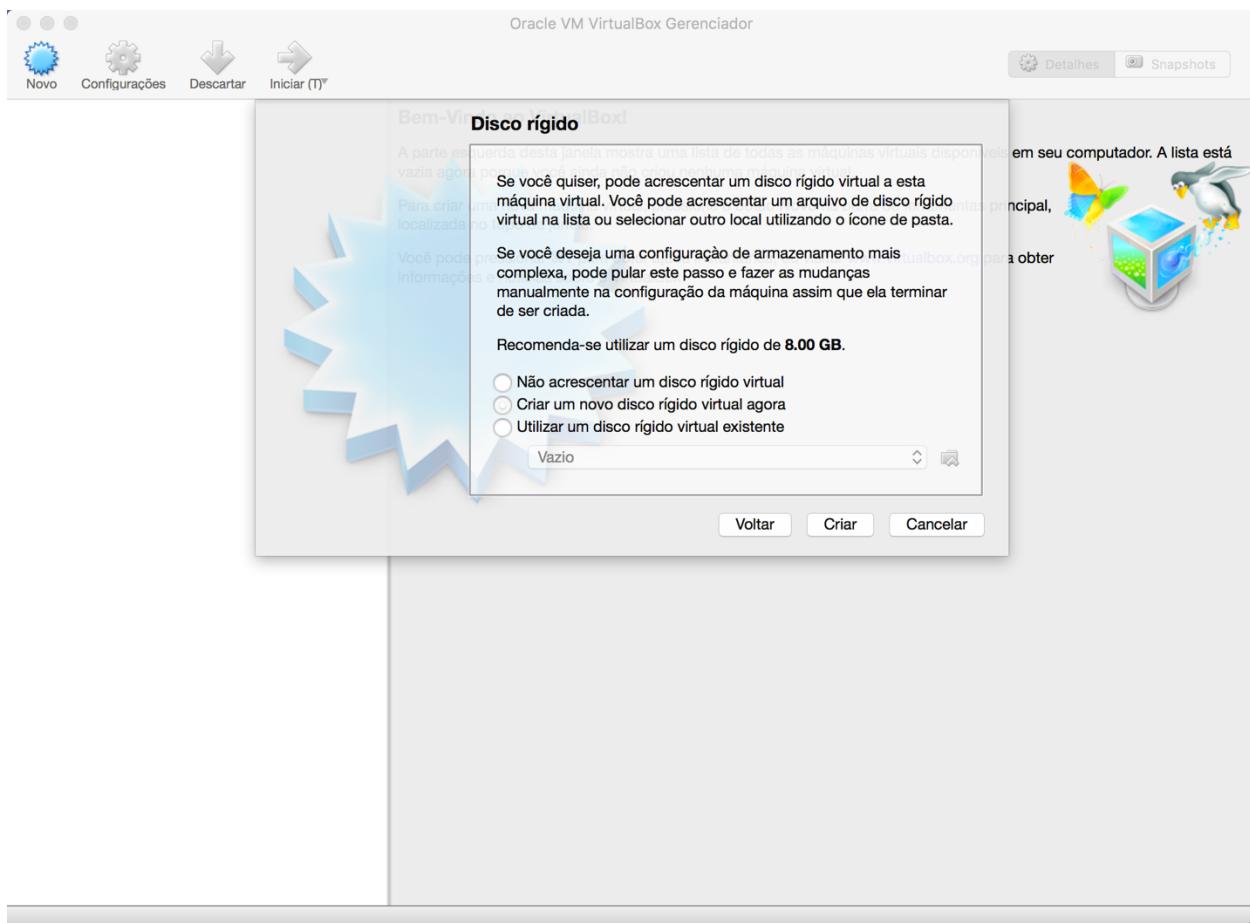


Definindo o nome da máquina virtual e a versão do sistema operacional

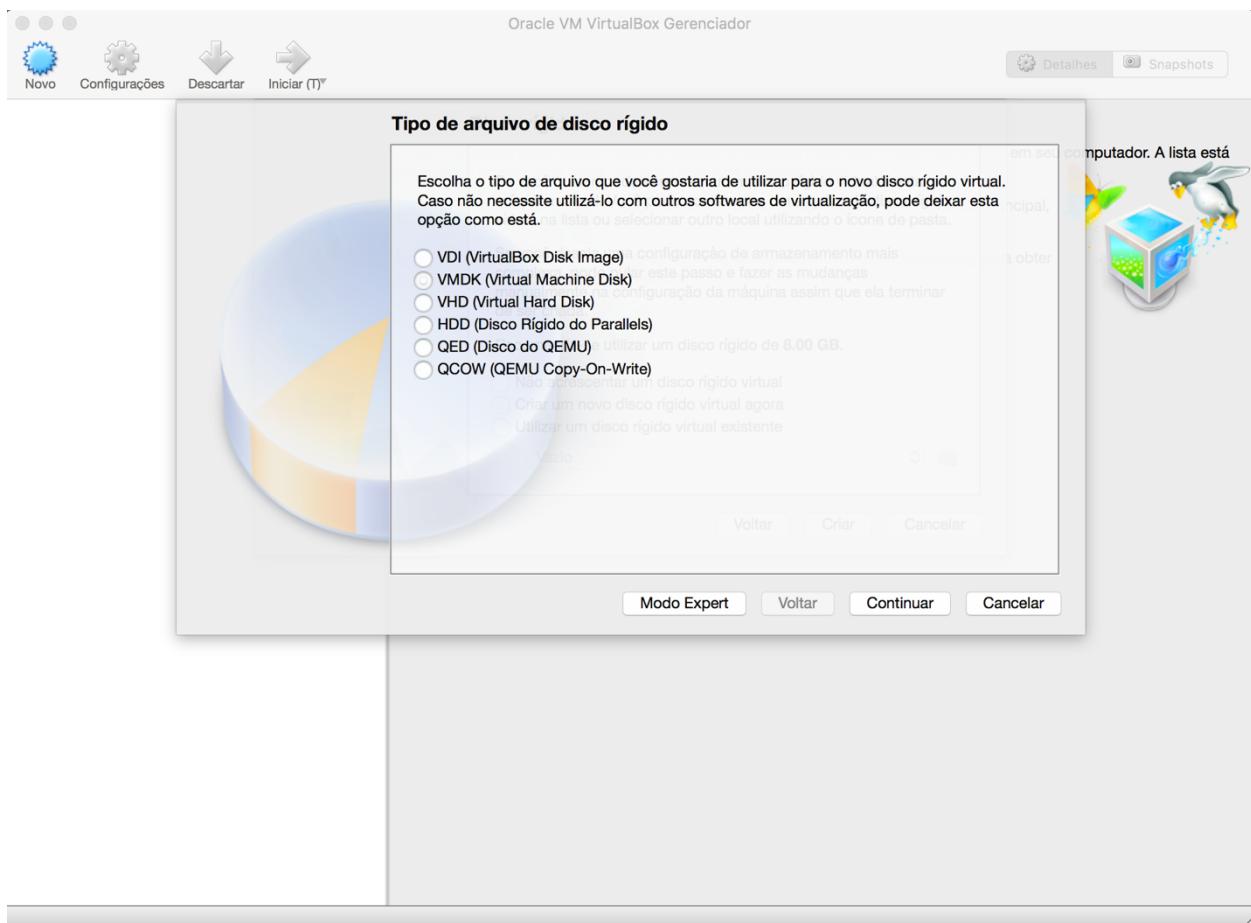
Instalação e Configuração do Ecosistema Hadoop



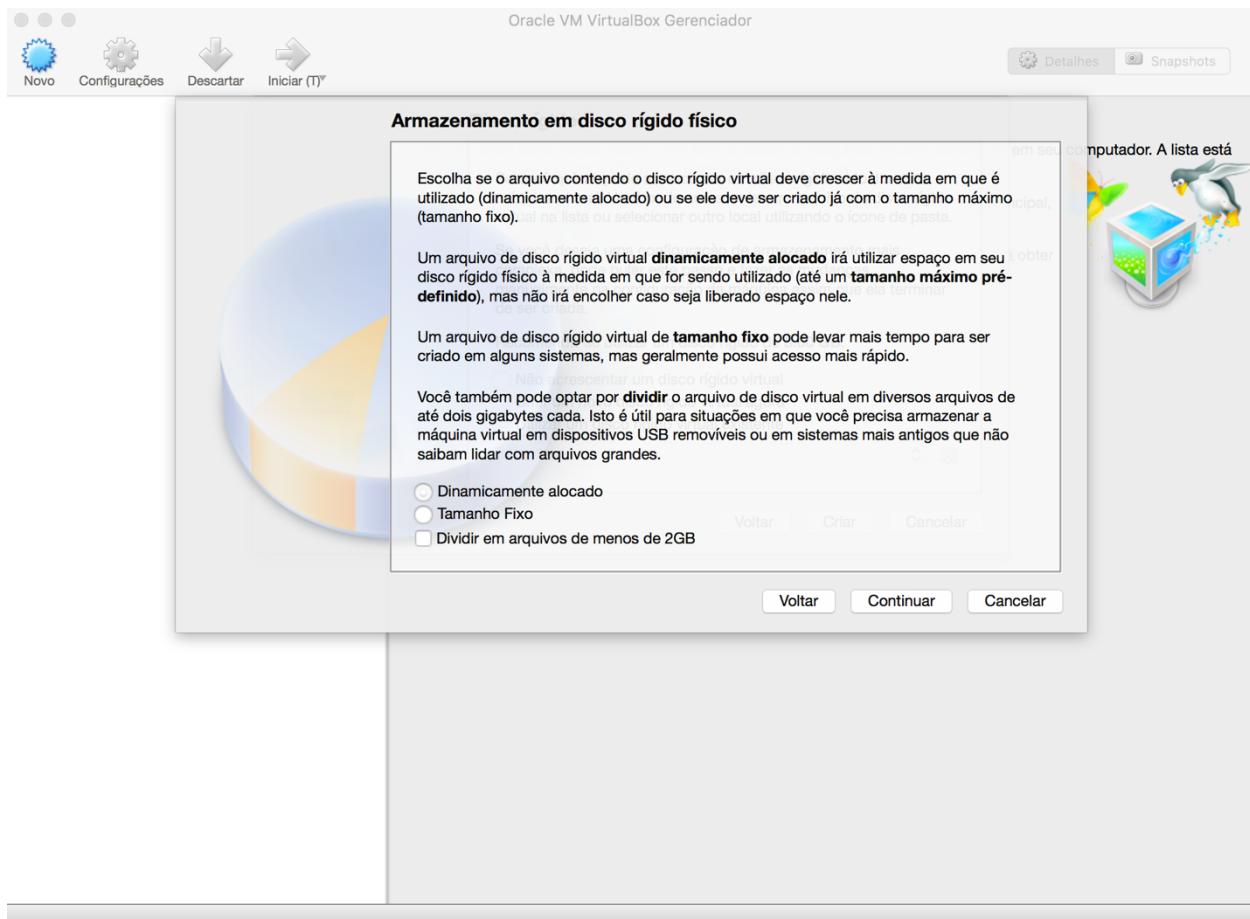
4 GB de memória são recomendados, mas não obrigatórios



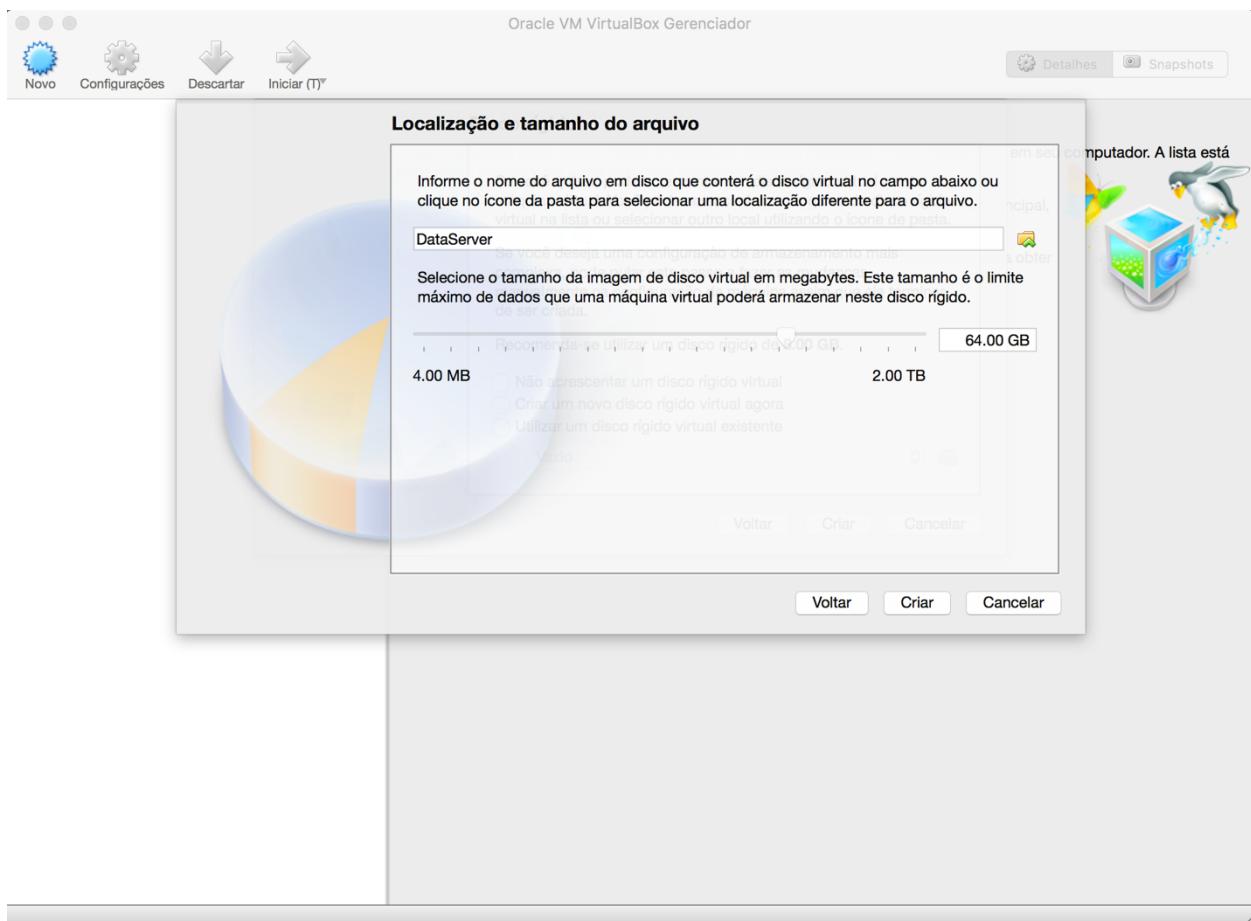
Criar um novo disco rígido virtual



Selecionando a opção VMDK

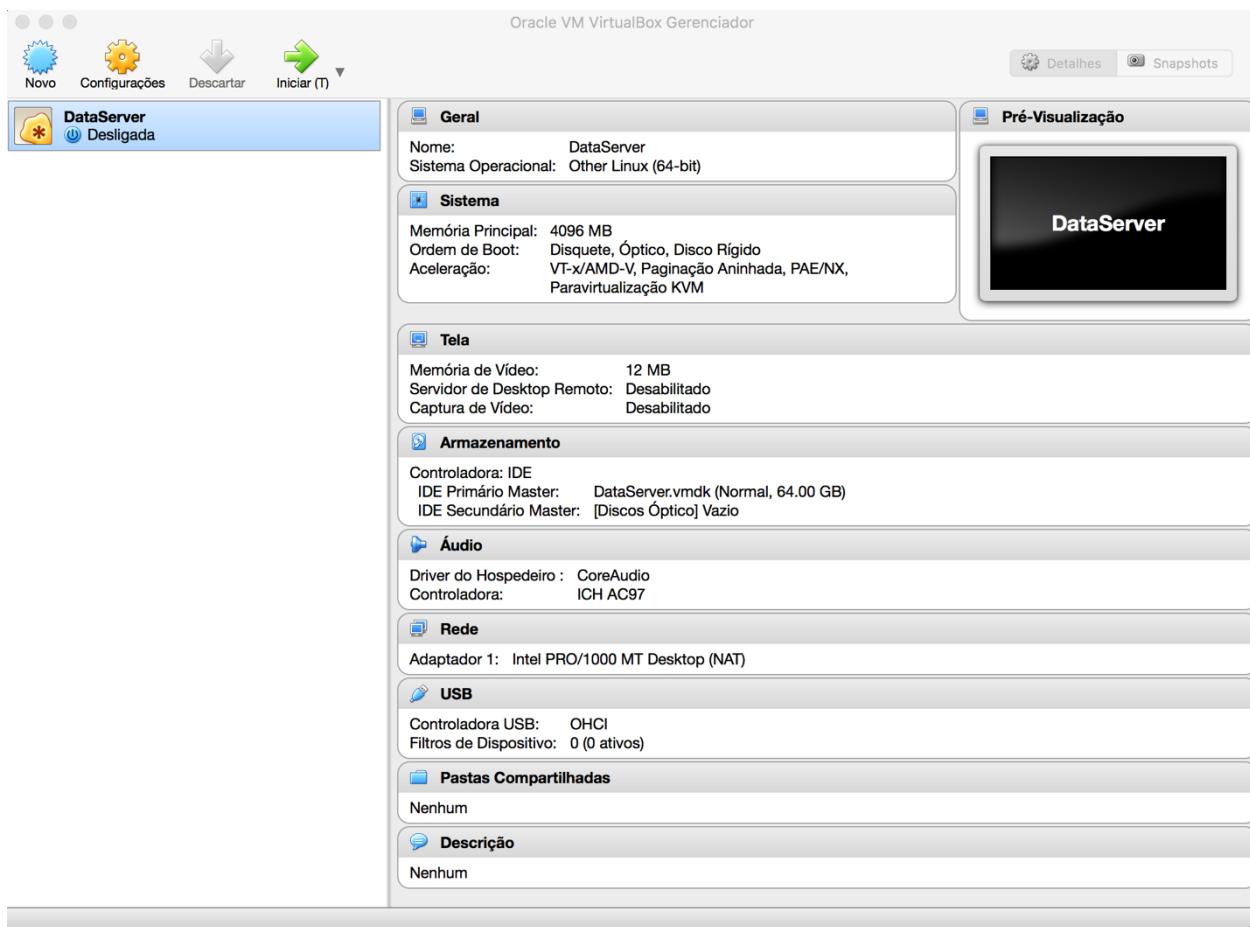


O disco deve ser alocado dinamicamente

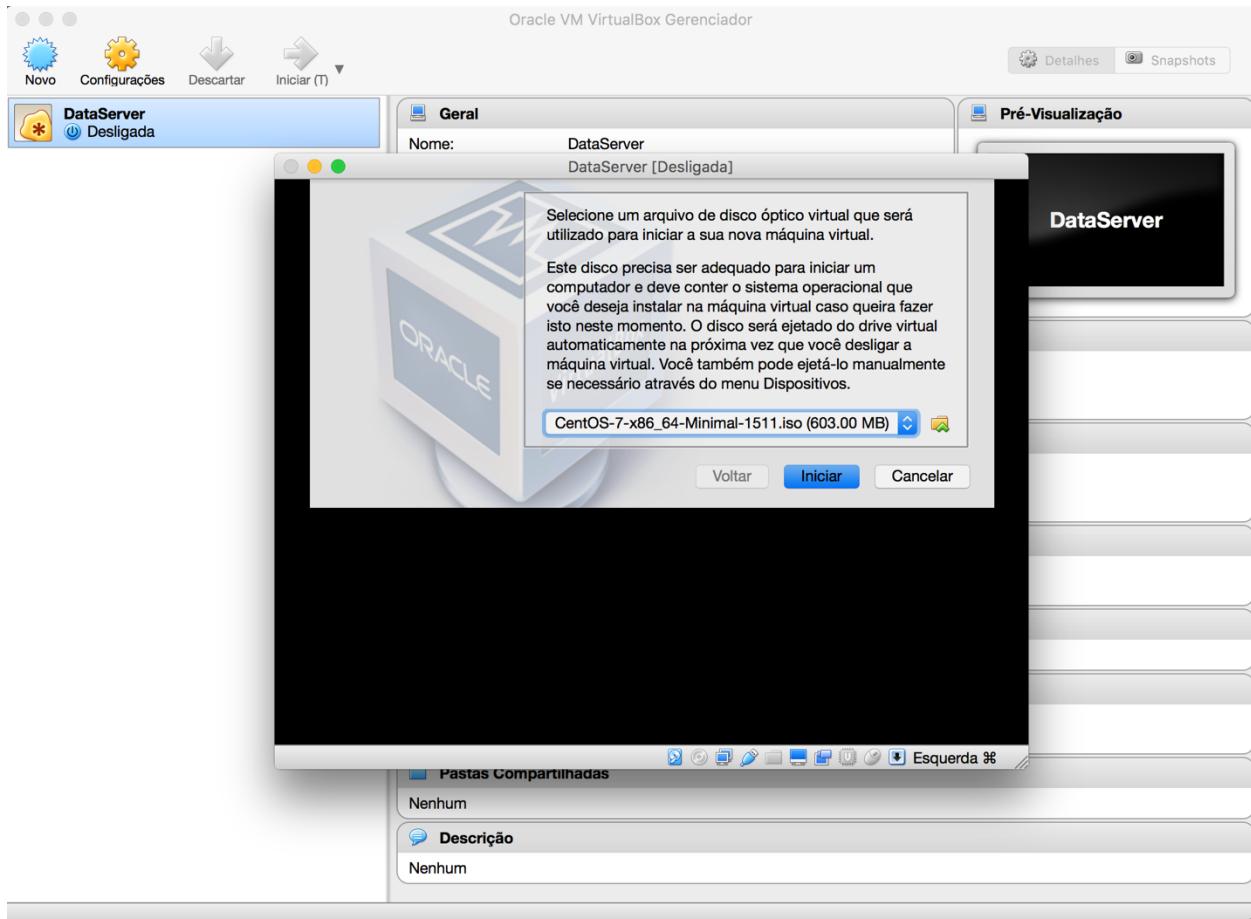


Seleciona 64 GB para o disco virtual

Instalação e Configuração do Ecosistema Hadoop



Máquina virtual criada



Selecione a mídia de instalação do sistema operacional

Utilizaremos o Cent OS versão 7. Recomendamos a mídia Minimal ISO, pois construiremos nosso sistema a partir da instalação básica do sistema operacional

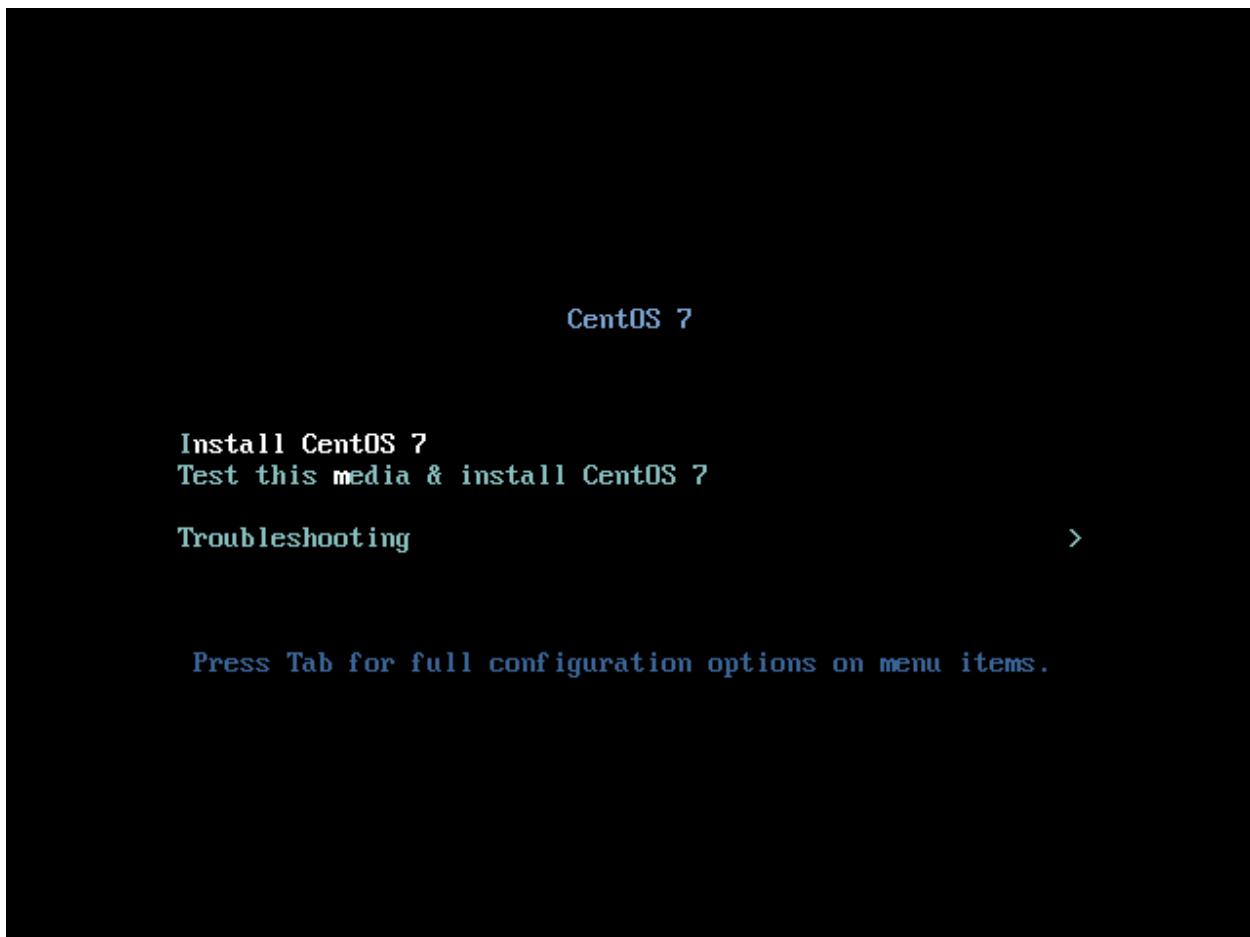
CentOS 64 bits (versão 7.2):

http://isoredirect.centos.org/centos/7/isos/x86_64/CentOS-7-x86_64-Minimal-1511.iso

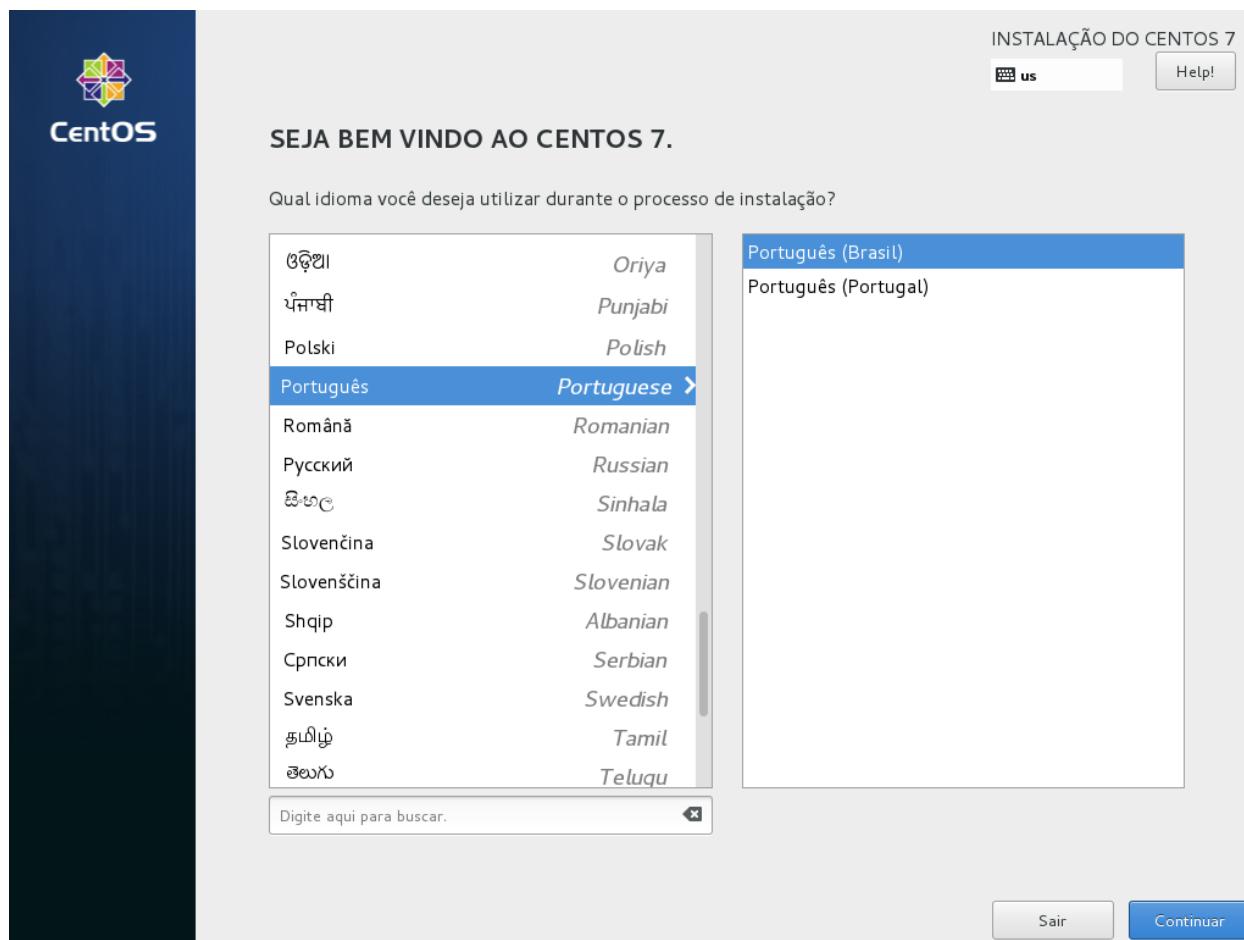
CentOS 32 bits (versão 6.8):

<http://centos.mirror.netelligent.ca/centos/>

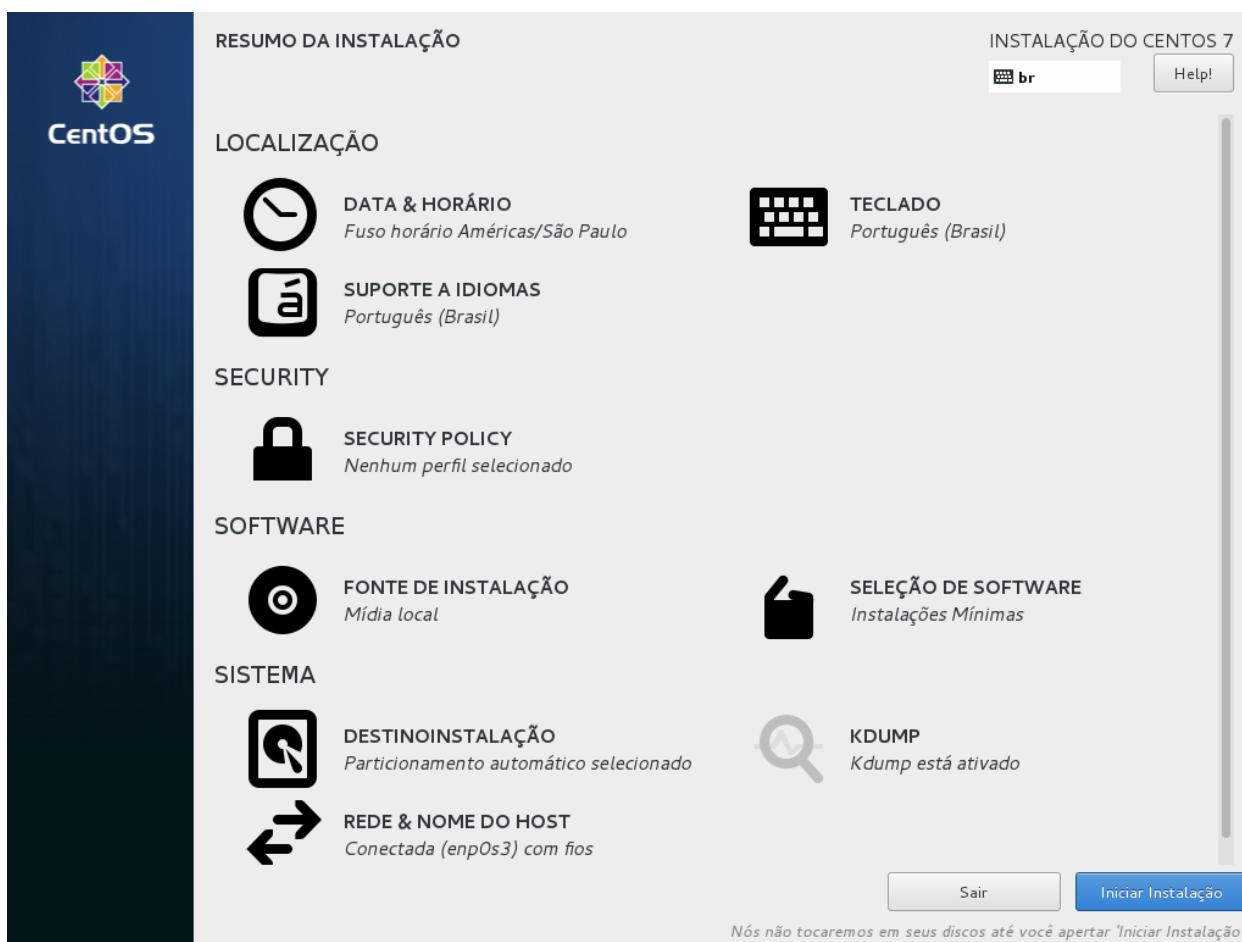
2.2. Instalação do Sistema Operacional



Instalação do Sistema Operacional CentOS 7



Seleção do idioma



Opções de configuração



Timezone

LAYOUT DO TECLADO

Finalizado

INSTALAÇÃO DO CENTOS 7

br Help!

Quais layouts de teclado você deseja usar no seu sistema? Você deve mover um layout qualquer para a parte superior da lista para selecioná-lo como padrão.

Inglês (EUA)

Português (Brasil)

Testar a configuração de layout abaixo:

Alteração do layout não configurado.

Opções

+ - ^ v ☰

Layout do teclado

SUporte a idiomas

Finalizado

INSTALAÇÃO DO CENTOS 7
br Help!

Selecionar um suporte de idioma adicional para ser instalado:

Português	Portuguese
Română	Romanian
Русский	Russian
සිංහල	Sinhala
Slovenčina	Slovak
Slovenščina	Slovenian
Shqip	Albanian
Српски	Serbian
Svenska	Swedish
தமிழ்	Tamil
తెలుగు	Telugu
Тоҷикӣ	Tajik
ไทย	Thai
Türkçe	Turkish
Українська	Ukrainian
اردو	Urdu
Tiếng Việt	Vietnamese

Português (Brasil)
<input checked="" type="checkbox"/> Português (Portugal)

|

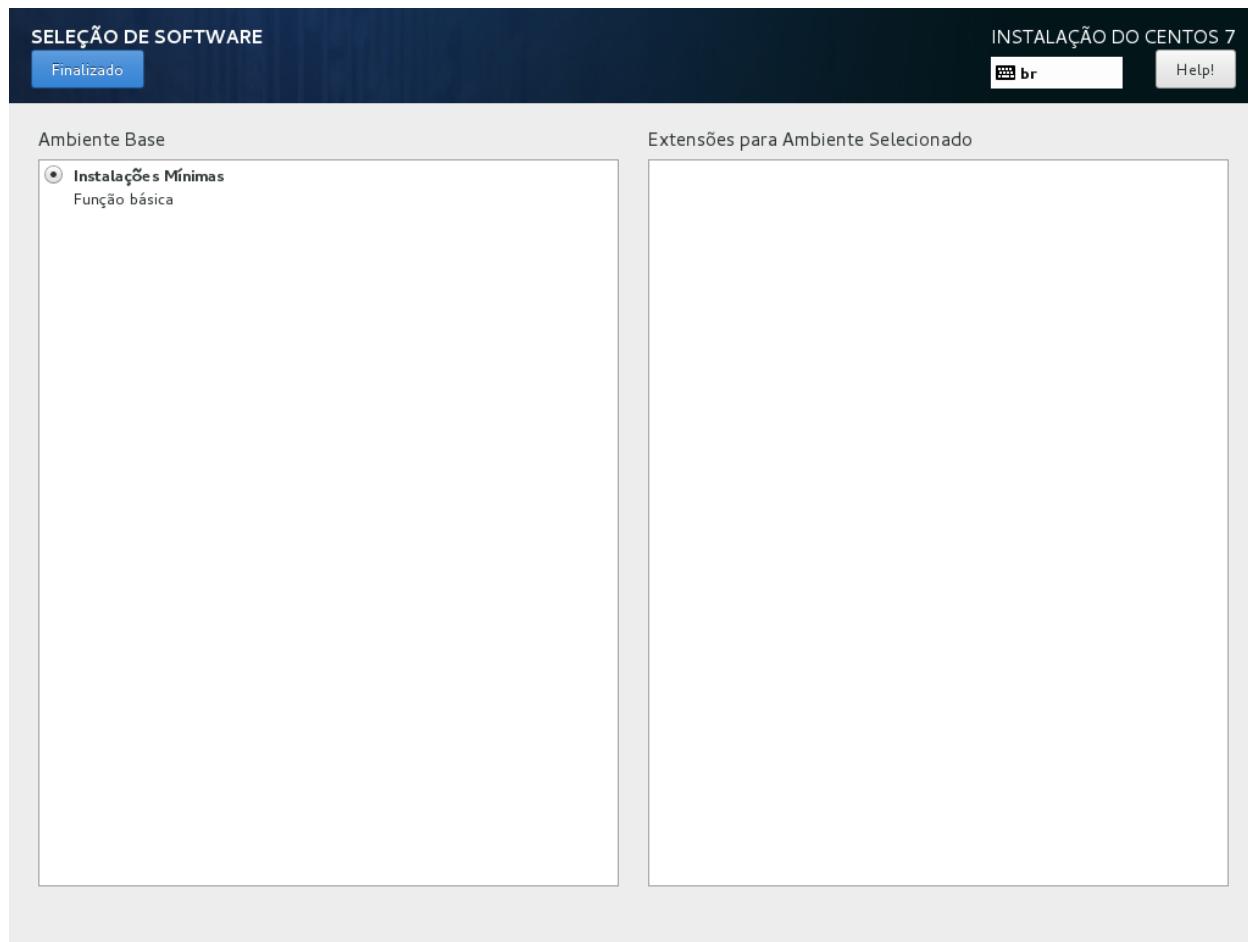
Idioma do sistema operacional

The screenshot shows a software interface for managing security policies on a CentOS 7 system. At the top, there's a dark header bar with the text "SECURITY POLICY" and "Finalizado" on the left, and "INSTALAÇÃO DO CENTOS 7", "br", and "Help!" on the right. Below the header, there's a toolbar with buttons for "Change content" and "Apply security policy: ON". A dropdown menu labeled "Choose profile below:" is open, showing several options:

- Default**: Described as the implicit XCCDF profile, containing no rules.
- Standard System Security Profile**: Described as ensuring a standard security base for CentOS Linux 7.
- Draft PCI-DSS v3 Control Baseline for CentOS Linux 7**: Described as a *draft* profile for PCI-DSS v3.
- CentOS Profile for Cloud Providers (CPCP)**: Described as a *draft* SCAP profile for CentOS Cloud Providers.
- Common Profile for General-Purpose Systems**: This option is highlighted with a blue background and a green checkmark icon to its right. It is described as containing items common to general-purpose desktop and server installations.
- Pre-release Draft STIG for CentOS Linux 7 Server**: Described as being developed under the DoD consensus model to become a STIG in coordination with DISA FSO.

Below the dropdown menu, there's a button labeled "Select profile". Further down, under the heading "Changes that were done or need to be done:", there's a note with a lightbulb icon followed by the text "Não há regras para a fase de pré-instalação".

Política de segurança padrão



Instalação Mínima do Sistema Operacional – A interface gráfica será instalada manualmente

DESTINAÇÃO DA INSTALAÇÃO

Finalizado

INSTALAÇÃO DO CENTOS 7

br Help!

Seleção de Dispositivo

Selecione o(s) dispositivo(s) nos quais você gostaria de instalar. Eles permanecerão intocados até que você clique no botão do menu principal "Iniciar Instalação".

Discos Padrões Locais

64 GiB

ATA VBOX HARDDISK
sda / 992,5 KiB livre

Discos não selecionados aqui não serão tocados.

Discos especializados & de Rede

 Adicionar um disco...

Discos não selecionados aqui não serão tocados.

Outras opções de armazenamento

Particionamento

Configurar automaticamente o particionamento. Eu irei configurar o particionamento.

Eu gostaria de disponibilizar espaço adicional.

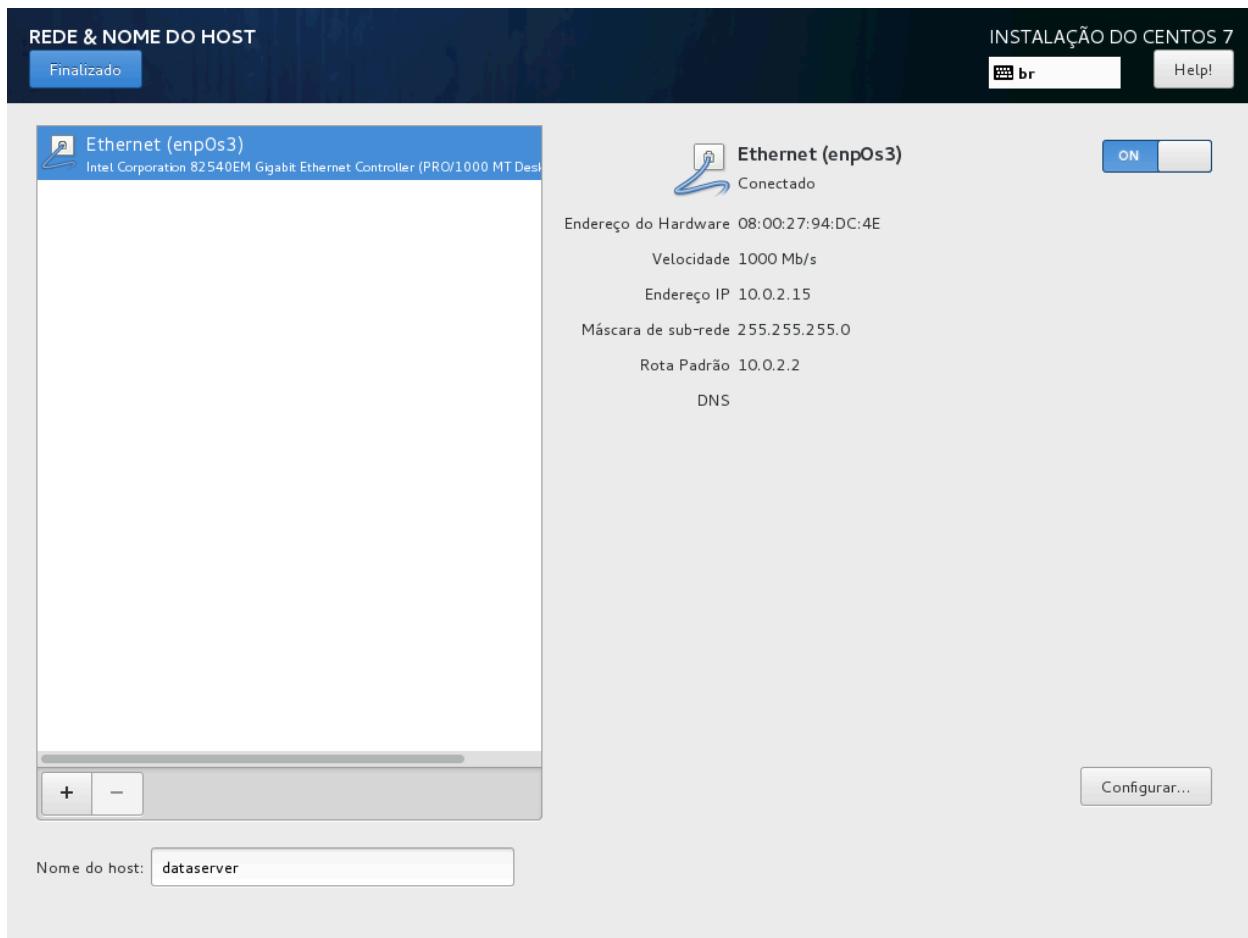
Criptografando

Criptografar meus dados. Você definirá uma senha a seguir.

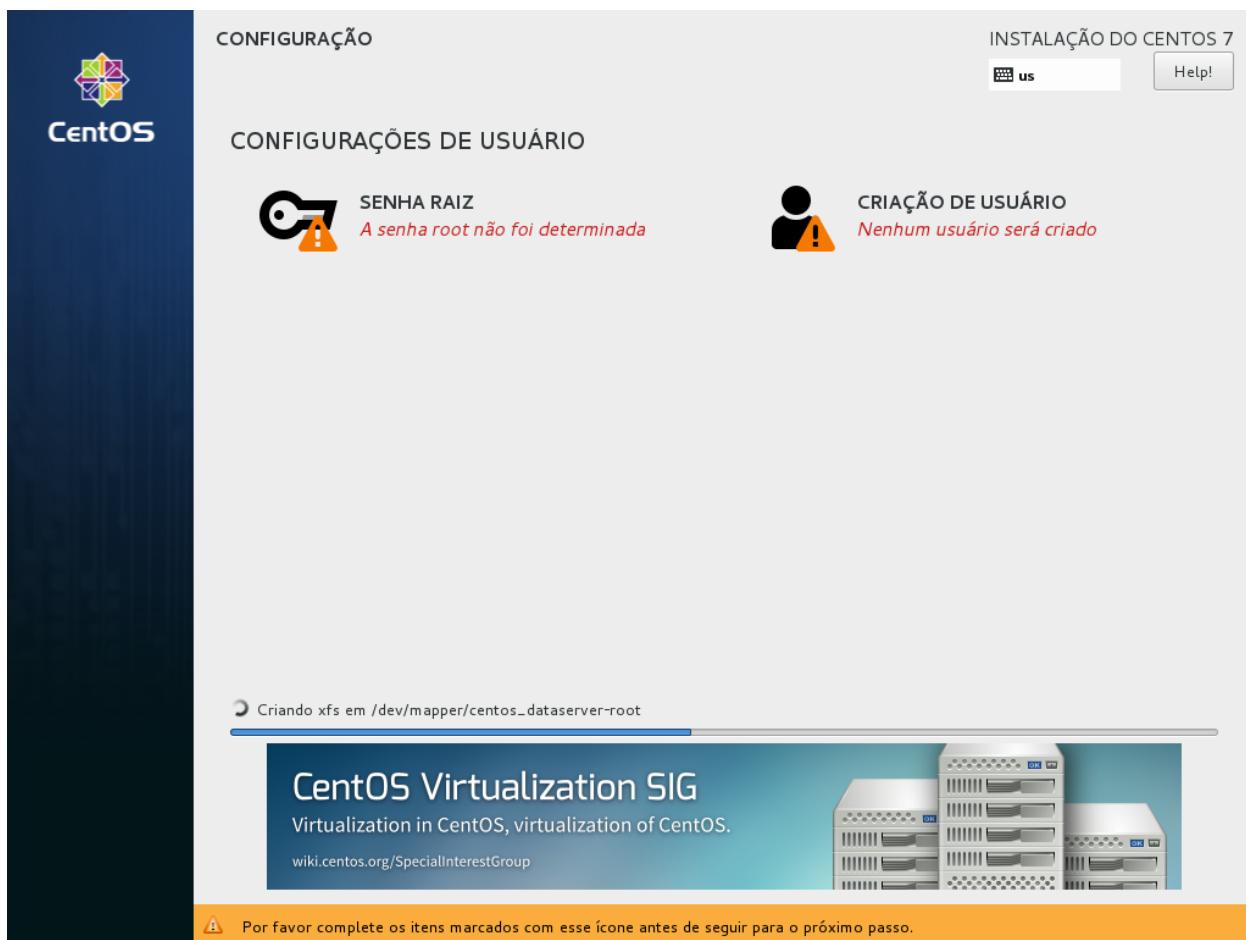
[Resumo de disco cheio e carregador de inicialização...](#)

1 discos selecionados; 64 GiB capacidade; 992,5 KiB livre

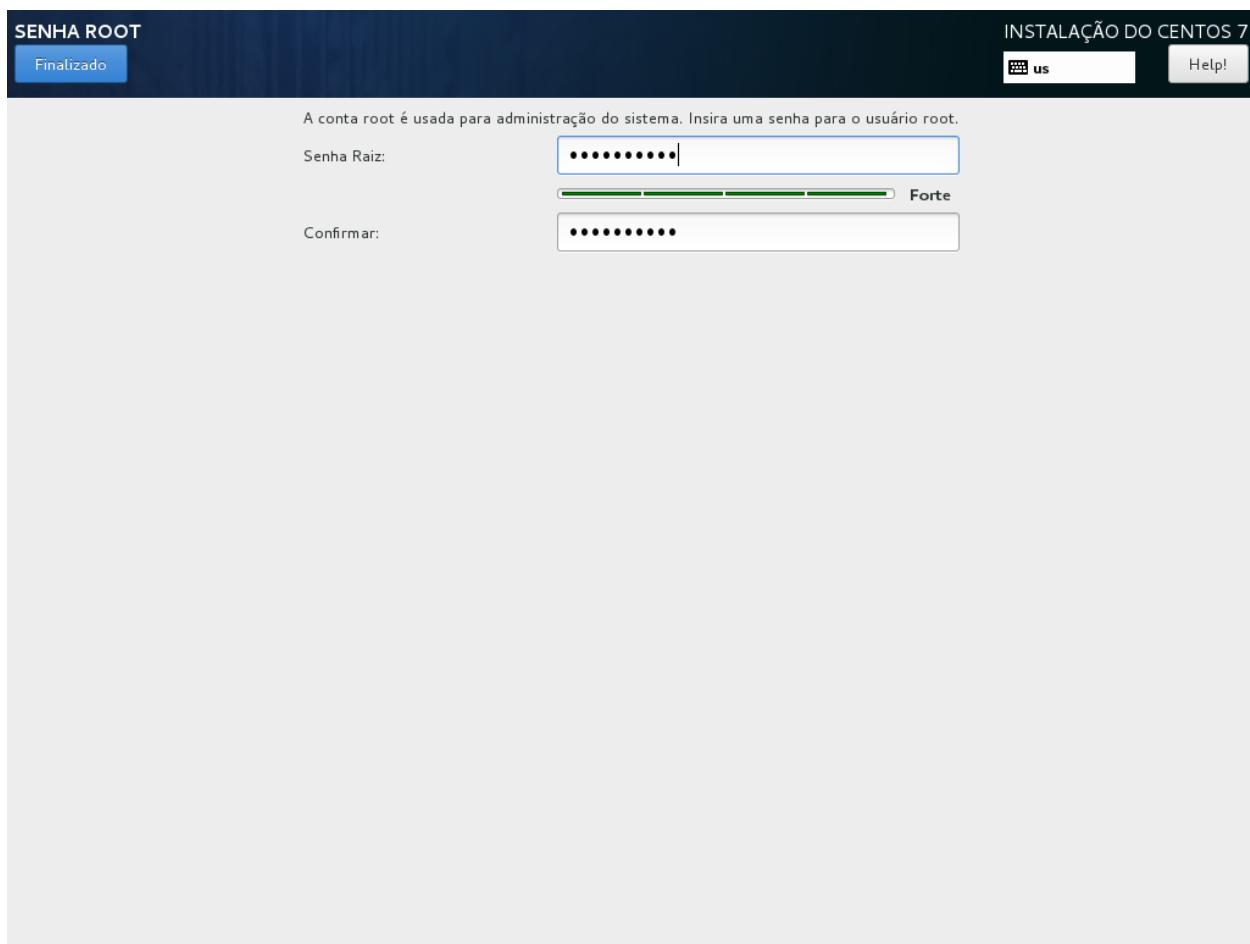
Disco



Configuração de Rede e nome do servidor – **dataserver**
Certifique-se de habilitar a opção de ativar a Ethernet (botão on)



Definir senha do root – usuário administrador



Definir senha do root – usuário administrador
Senha: dsacademy

Instalação e Configuração do Ecosistema Hadoop

CRIAR USUÁRIO

Finalizado

INSTALAÇÃO DO CENTOS 7

us Help!

Nome Completo: Aluno

Nome do usuário: aluno

Tip: Mantenha o seu nome de usuário com menos de 32 caracteres e não use espaços.

Tornar esse usuário administrador

É necessário uma senha para utilizar essa conta

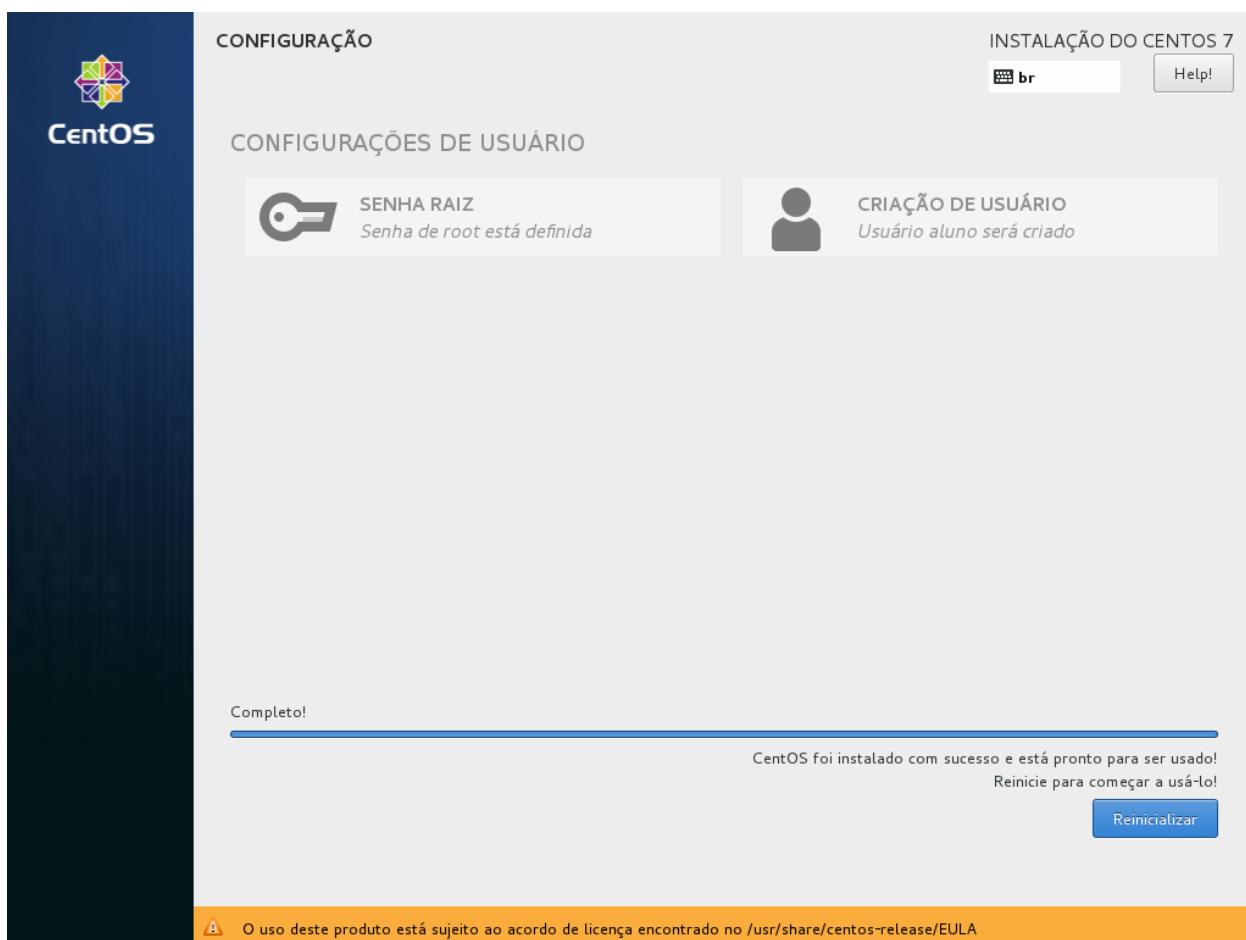
Senha: ••••••••

Confirmar senha: •••••••|

Avaliação da senha: Forte

[Avançado...](#)

Criação de um usuário – Aluno
(username: aluno, senha: dsacademy)



Conclusão da instalação

```
CentOS Linux 7 (Core)
Kernel 3.10.0-327.el7.x86_64 on an x86_64

dataserver login: _
```

Tela inicial de login

2.3. Instalação da Interface Gráfica Básica

```
[root@dataserver ~]# yum groupinstall "X Window System"
```

yum groupinstall “X Window System”

```
plymouth-plugin-label.x86_64 0:0.8.9-0.24.20140113.el7.centos
plymouth-plugin-two-step.x86_64 0:0.8.9-0.24.20140113.el7.centos
plymouth-theme-charge.x86_64 0:0.8.9-0.24.20140113.el7.centos
xcb-util.x86_64 0:0.4.0-2.el7
xkeyboard-config.noarch 0:2.14-1.el7
xorg-x11-drv-ati.x86_64 0:7.5.0-3.el7
xorg-x11-drv-dummy.x86_64 0:0.3.6-21.el7
xorg-x11-drv-evdev.x86_64 0:2.9.2-2.el7
xorg-x11-drv-fbdev.x86_64 0:0.4.3-20.el7
xorg-x11-drv-intel.x86_64 0:2.99.917-8.20150615.el7
xorg-x11-drv-nouveau.x86_64 1:1.0.11-2.el7
xorg-x11-drv-qxl.x86_64 0:0.1.1-18.el7
xorg-x11-drv-synaptics.x86_64 0:1.8.2-1.el7
xorg-x11-drv-v4l.x86_64 0:0.2.0-42.el7
xorg-x11-drv-vesa.x86_64 0:2.3.2-20.el7
xorg-x11-drv-vmmouse.x86_64 0:13.0.0-11.el7
xorg-x11-drv-vmware.x86_64 0:13.0.2-7.20150211git8f0cf7c.el7
xorg-x11-drv-void.x86_64 0:1.4.1-1.el7
xorg-x11-drv-wacom.x86_64 0:0.29.0-1.el7
xorg-x11-server-common.x86_64 0:1.17.2-10.el7
xorg-x11-server-utils.x86_64 0:7.7-14.el7
xorg-x11-xkb-utils.x86_64 0:7.7-12.el7
```

Concluído!

```
[root@dataserver ~]# _
```

Concluído com sucesso

```
[root@dataserver ~]# yum install gnome-classic-session_
```

yum install gnome-classic-session

```
samba-libs.x86_64 0:4.2.3-12.el7_2
sbc.x86_64 0:1.0-5.el7
sound-theme-freedesktop.noarch 0:0.8-3.el7
speex.x86_64 0:1.2-0.19.rc1.el7
startup-notification.x86_64 0:0.12-8.el7
taglib.x86_64 0:1.8-7.20130218git.el7
telepathy-filesystem.noarch 0:0.0.2-6.el7
telepathy-glib.x86_64 0:0.24.0-1.el7
telepathy-logger.x86_64 0:0.8.0-5.el7
totem-pl-parser.x86_64 0:3.10.5-1.el7
tracker.x86_64 0:1.2.6-3.el7
udisks2.x86_64 0:2.1.2-6.el7
upower.x86_64 0:0.99.2-1.el7
usbmuxd.x86_64 0:1.0.8-11.el7
vino.x86_64 0:3.14.2-1.el7
webkitgtk3.x86_64 0:2.4.9-5.el7
webrtc-audio-processing.x86_64 0:0.1-5.el7
zenity.x86_64 0:3.8.0-5.el7
```

Dependência(s) atualizada(s):
avahi-autoipd.x86_64 0:0.6.31-15.el7_2.1
avahi-libs.x86_64 0:0.6.31-15.el7_2.1

Concluído!

```
[root@dataserver ~]# _
```

Concluído com sucesso

```
[root@dataserver ~]# yum install control-center gnome-terminal nautilus-open-terminal liberation-mono-fonts_
```

yum install control-center gnome-terminal nautilus-open-terminal liberation-mono-fonts

```
Instalando : vte-profile-0.38.3-2.el7.x86_64 1/6
Instalando : vte291-0.38.3-2.el7.x86_64 2/6
Instalando : 1:liberation-fonts-common-1.07.2-15.el7.noarch 3/6
Instalando : 1:liberation-mono-fonts-1.07.2-15.el7.noarch 4/6
Instalando : gnome-terminal-3.14.3-3.el7.x86_64 5/6
Instalando : nautilus-open-terminal-0.20-3.el7.x86_64 6/6
Verifying  : 1:liberation-fonts-common-1.07.2-15.el7.noarch 1/6
Verifying  : nautilus-open-terminal-0.20-3.el7.x86_64 2/6
Verifying  : vte291-0.38.3-2.el7.x86_64 3/6
Verifying  : 1:liberation-mono-fonts-1.07.2-15.el7.noarch 4/6
Verifying  : vte-profile-0.38.3-2.el7.x86_64 5/6
Verifying  : gnome-terminal-3.14.3-3.el7.x86_64 6/6

Instalados:
  gnome-terminal.x86_64 0:3.14.3-3.el7
  liberation-mono-fonts.noarch 1:1.07.2-15.el7
  nautilus-open-terminal.x86_64 0:0.20-3.el7

Dependência(s) instalada(s):
  liberation-fonts-common.noarch 1:1.07.2-15.el7
  vte-profile.x86_64 0:0.38.3-2.el7
  vte291.x86_64 0:0.38.3-2.el7

Concluído!
[root@dataserver ~]# _
```

Concluído com sucesso

```
[root@dataserver ~]# yum install kernel-devel
```

yum install kernel-devel

```
perl-File-Temp.noarch 0:0.23.01-3.el7
perl-Filter.x86_64 0:1.49-3.el7
perl-Getopt-Long.noarch 0:2.40-2.el7
perl-HTTP-Tiny.noarch 0:0.033-3.el7
perl-PathTools.x86_64 0:3.40-5.el7
perl-Pod-Escapes.noarch 1:1.04-286.el7
perl-Pod-Perldoc.noarch 0:3.20-4.el7
perl-Pod-Simple.noarch 1:3.28-4.el7
perl-Pod-Usage.noarch 0:1.63-3.el7
perl-Scalar-List-Utils.x86_64 0:1.27-248.el7
perl-Socket.x86_64 0:2.010-3.el7
perl-Storable.x86_64 0:2.45-3.el7
perl-Text-ParseWords.noarch 0:3.29-4.el7
perl-Time-HiRes.x86_64 4:1.9725-3.el7
perl-Time-Local.noarch 0:1.2300-2.el7
perl-constant.noarch 0:1.27-2.el7
perl-libs.x86_64 4:5.16.3-286.el7
perl-macros.x86_64 4:5.16.3-286.el7
perl-parent.noarch 1:0.225-244.el7
perl-podlators.noarch 0:2.5.1-3.el7
perl-threads.x86_64 0:1.87-4.el7
perl-threads-shared.x86_64 0:1.43-6.el7
```

Concluído!

[root@dataserver ~]# _

Concluído com sucesso

```
[root@dataserver ~]# yum update_
```

Atualização do sistema operacional

```
lvm2-libs.x86_64 7:2.02.130-5.el7_2.1
nss.x86_64 0:3.19.1-19.el7_2
nss-sysinit.x86_64 0:3.19.1-19.el7_2
nss-tools.x86_64 0:3.19.1-19.el7_2
nss-util.x86_64 0:3.19.1-9.el7_2
numactl-libs.x86_64 0:2.0.9-6.el7_2
openssh.x86_64 0:6.6.1p1-23.el7_2
openssh-clients.x86_64 0:6.6.1p1-23.el7_2
openssh-server.x86_64 0:6.6.1p1-23.el7_2
openssl.x86_64 1:1.0.1e-51.el7_2.4
openssl-libs.x86_64 1:1.0.1e-51.el7_2.4
polkit.x86_64 0:0.112-6.el7_2
procps-ng.x86_64 0:3.3.10-5.el7_2
python-perf.x86_64 0:3.10.0-327.10.1.el7
python-pyudev.noarch 0:0.15-7.el7_2.1
selinux-policy.noarch 0:3.13.1-60.el7_2.3
selinux-policy-targeted.noarch 0:3.13.1-60.el7_2.3
systemd.x86_64 0:219-19.el7_2.4
systemd-libs.x86_64 0:219-19.el7_2.4
systemd-sysv.x86_64 0:219-19.el7_2.4
tuned.noarch 0:2.5.1-4.el7_2.2
tzdata.noarch 0:2016b-1.el7
```

Concluído!

```
[root@dataserver ~]# _
```

Concluído

```
[root@dataserver ~]# unlink /etc/systemd/system/default.target_
```

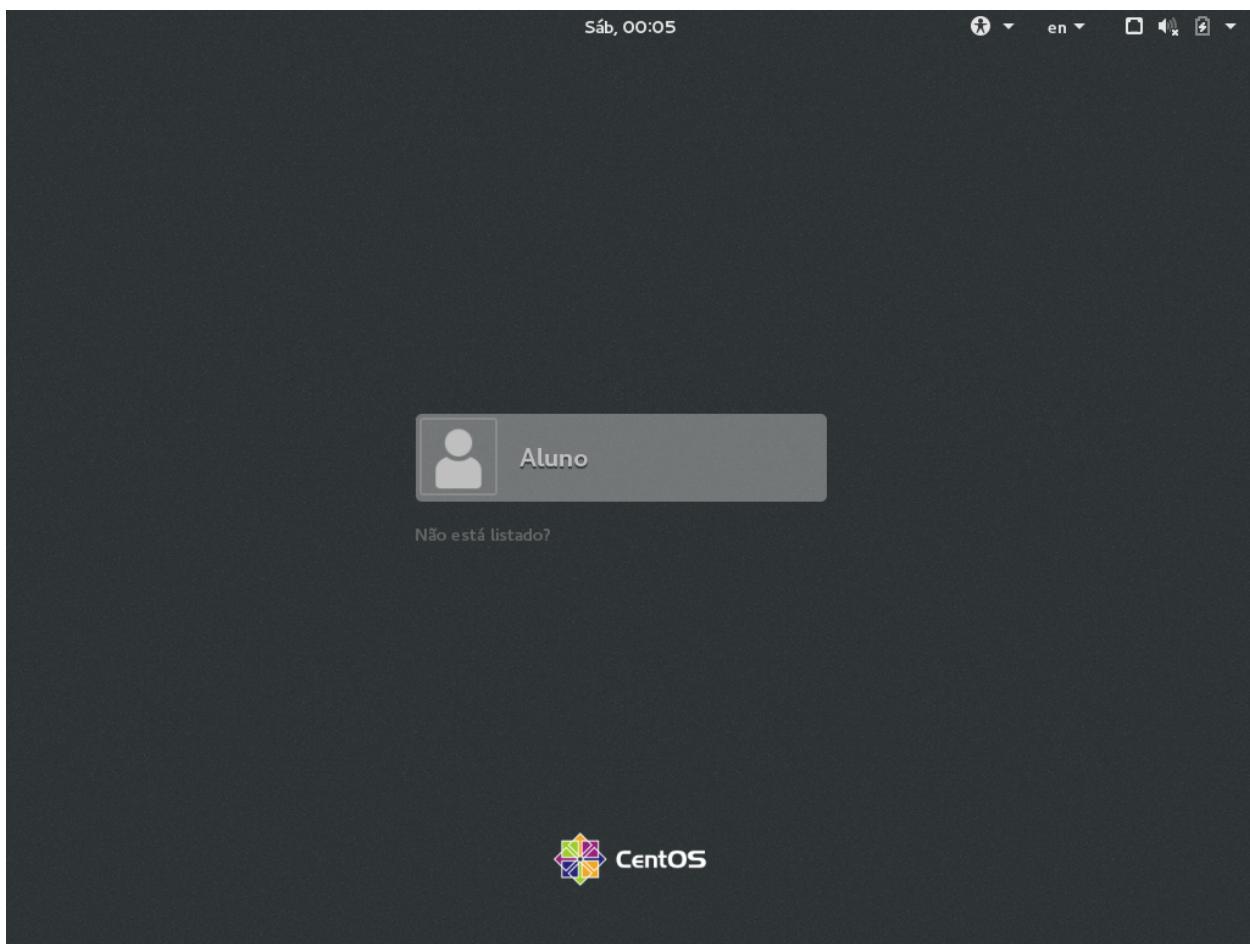
unlink /etc/systemd/system/default.target

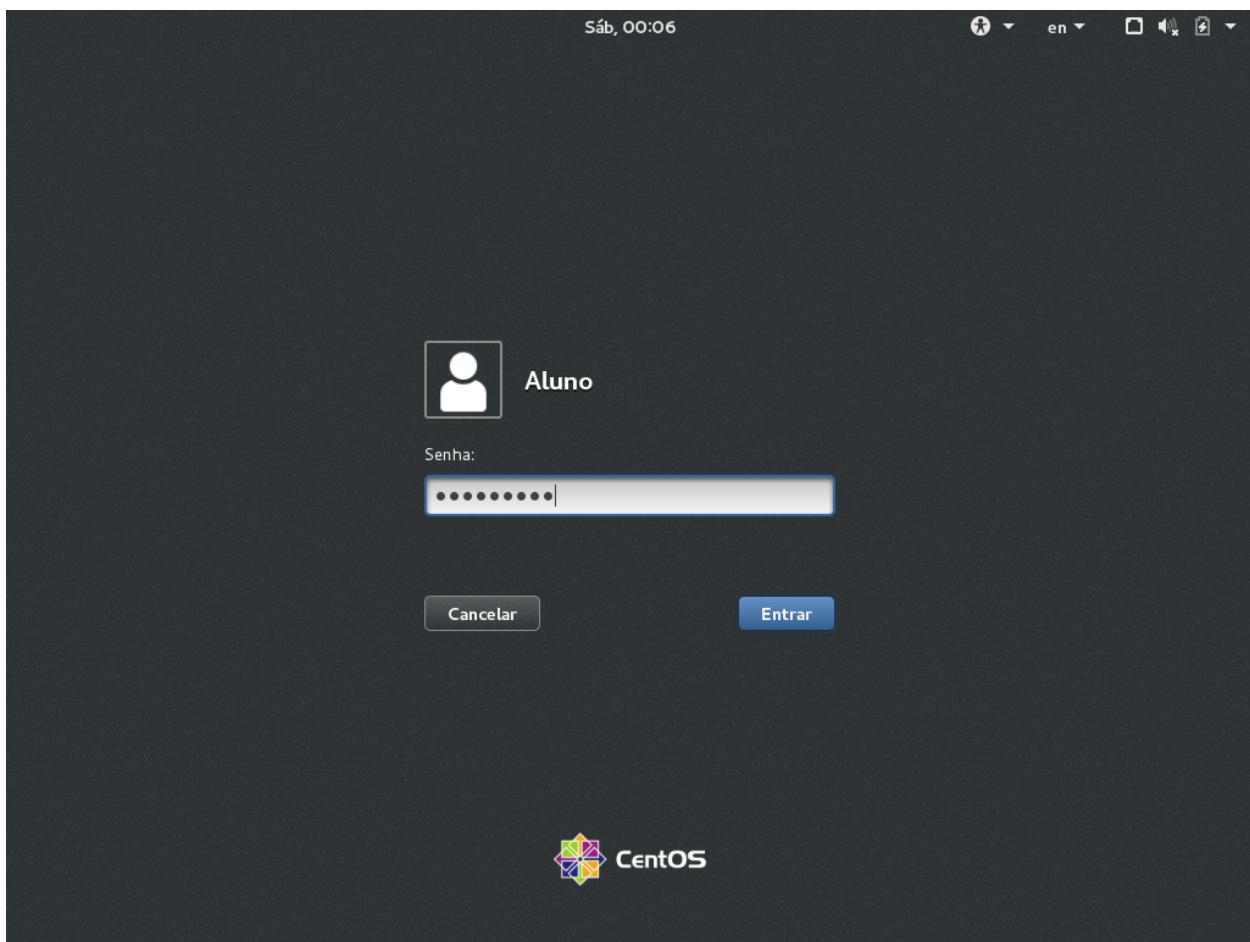
```
[root@dataserver ~]# ln -sf /lib/systemd/system/graphical.target /etc/systemd/system/default.target
```

ln -sf /lib/systemd/system/graphical.target /etc/systemd/system/default.target

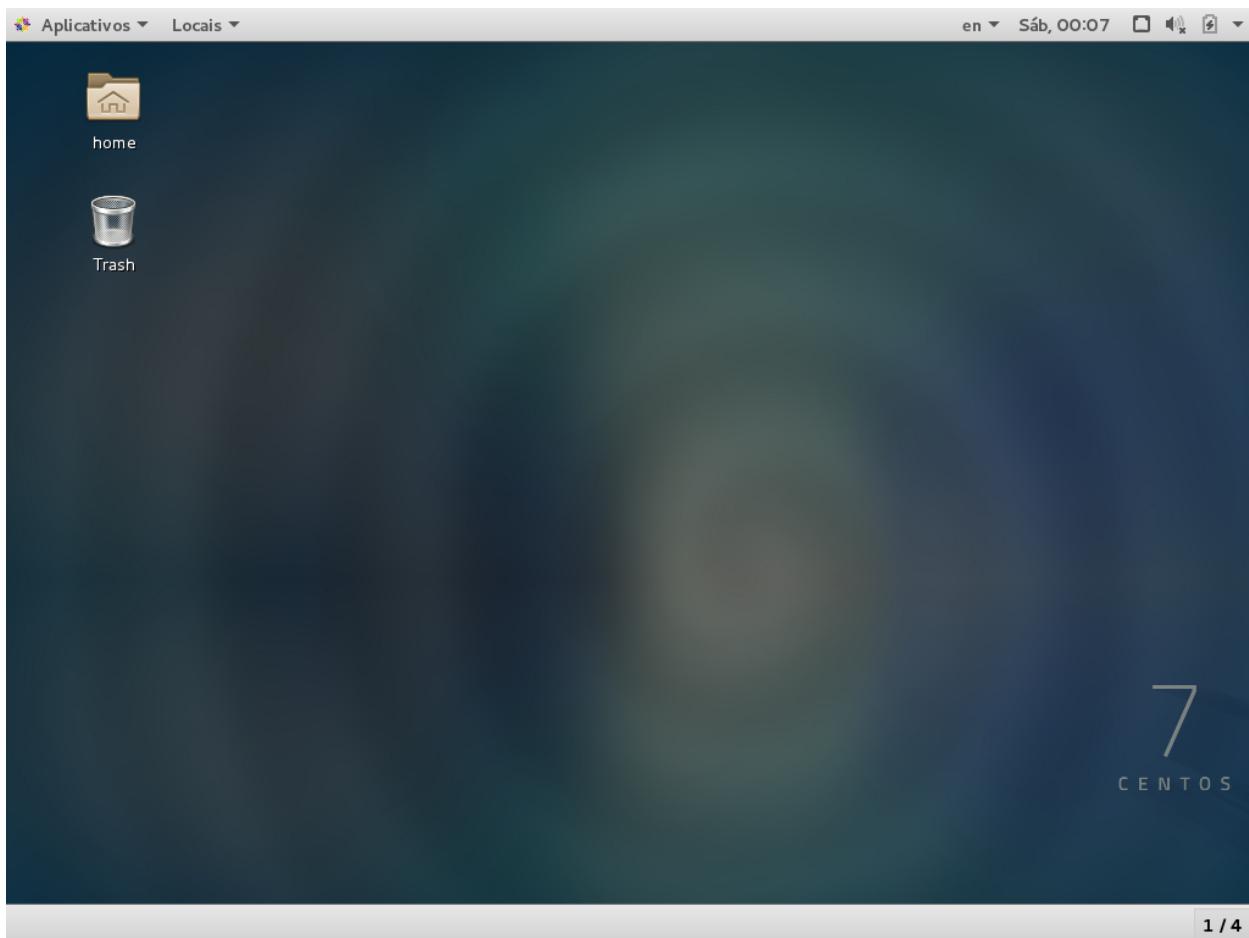
```
[root@dataserver ~]# unlink /etc/systemd/system/default.target  
[root@dataserver ~]# ln -sf /lib/systemd/system/graphical.target /etc/systemd/sy  
stem/default.target  
[root@dataserver ~]# reboot_
```

reboot





Senha



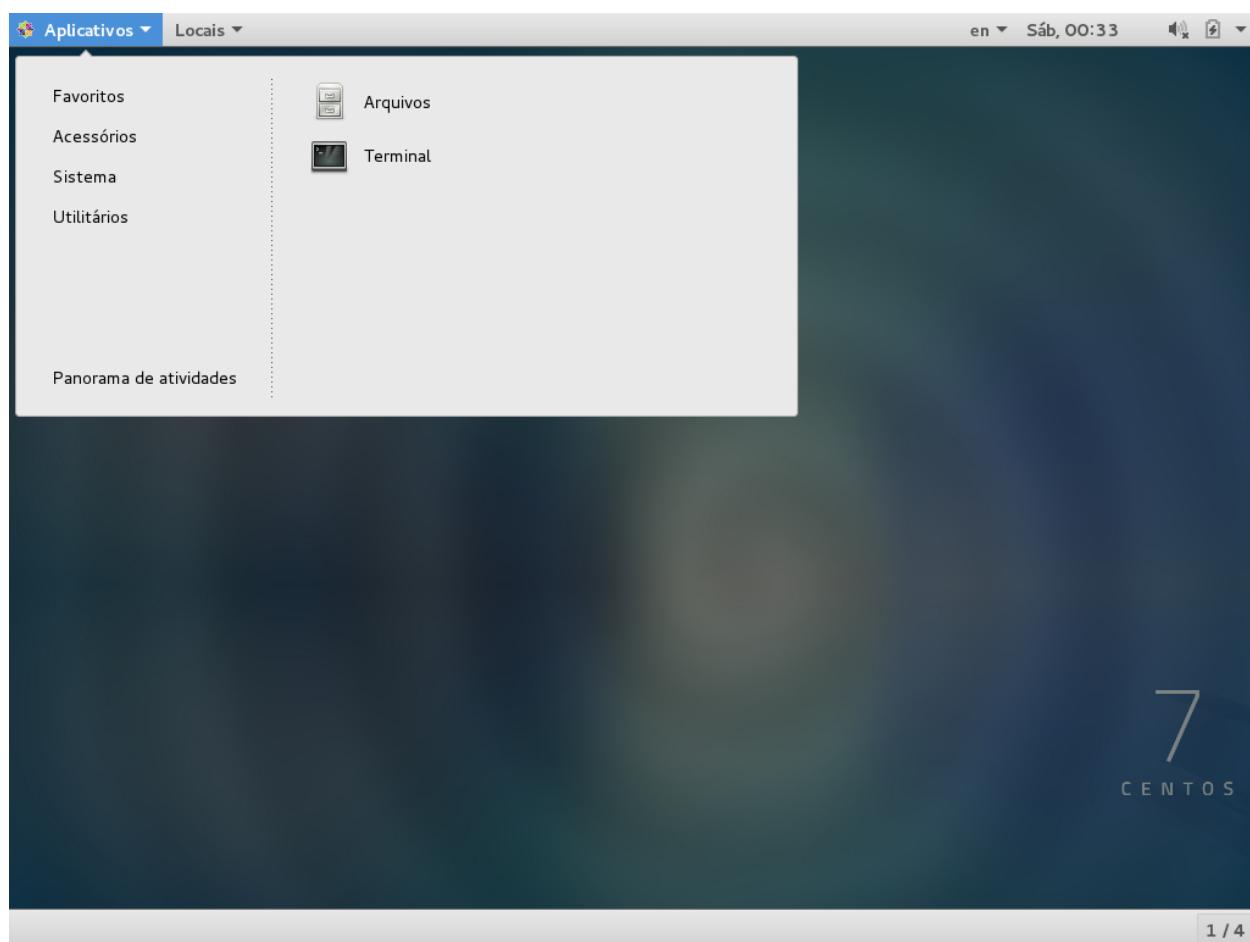
Instalação concluída com sucesso

Primeiro checkpoint:

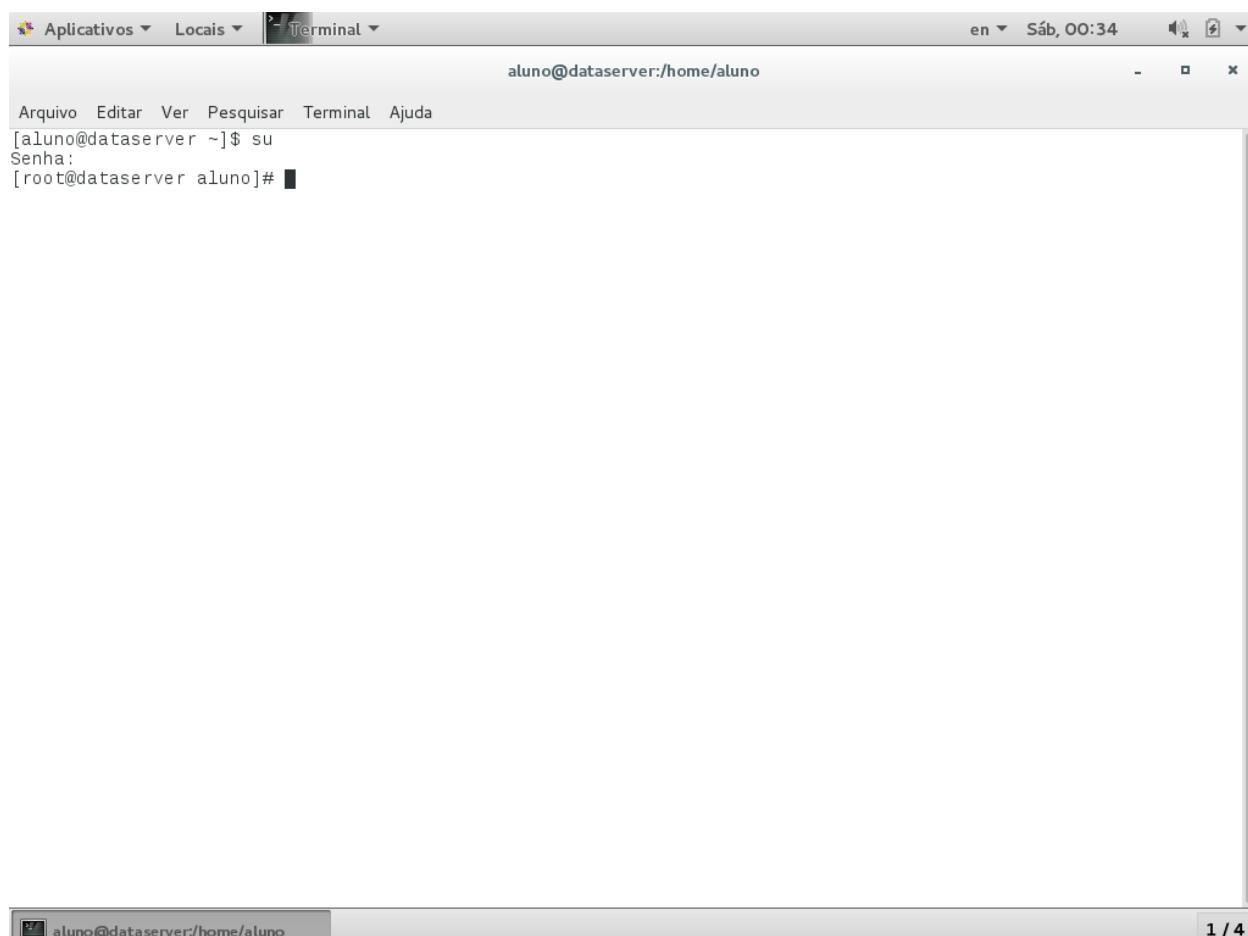
Clique no meu File – Export Appliance.
Será gerada uma cópia de segurança da sua máquina virtual.

→ VM: DataServer-v1.0.ova (Apenas SO)

2.4. Instalação de Utilitários do Sistema Operacional



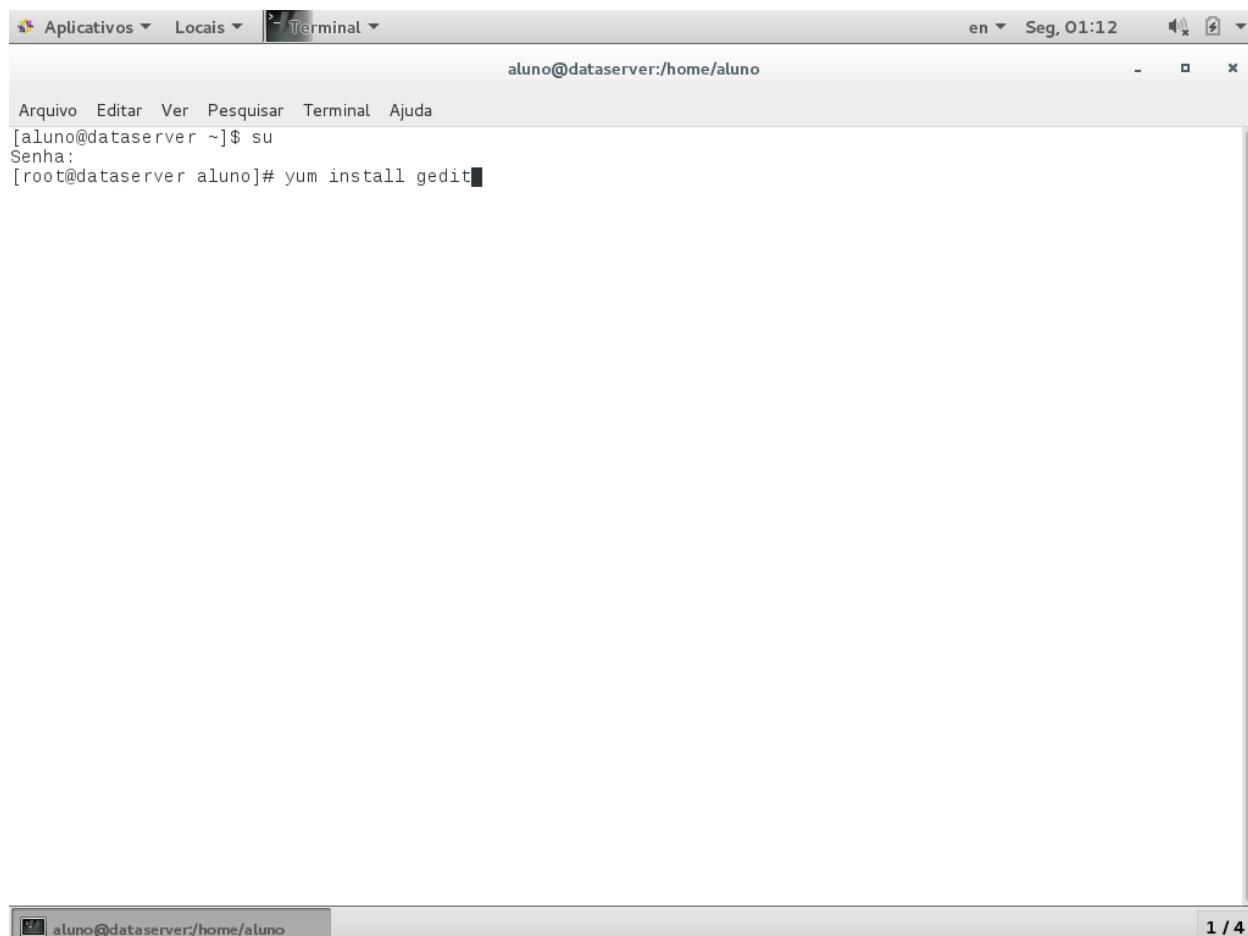
Abrindo o terminal



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:/home/aluno". The user has typed the command "[aluno@dataserver ~]\$ su" followed by their password. The terminal shows the user switching to root mode, indicated by the prompt "[root@dataserver aluno]#".

Efetuar login como root, usando o comando su

1 / 4

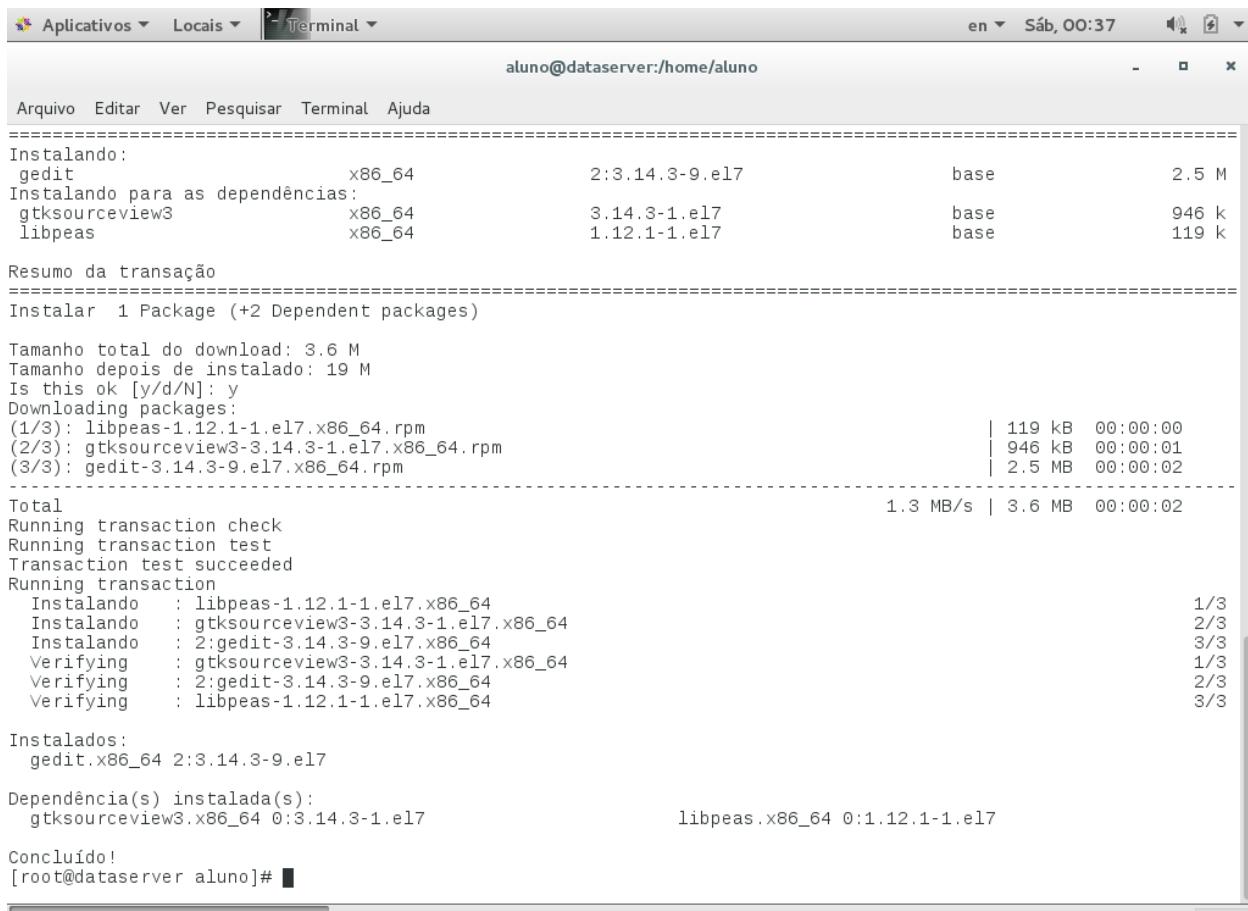


A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~\$'. The user types the command '[root@dataserver ~]\$ su' followed by their password. After becoming root, they type '[root@dataserver ~]# yum install gedit'. The terminal window has a standard window title bar with icons for application, location, and terminal, and a status bar at the bottom.

Instalar o editor de texto gedit, com o comando **yum install gedit**

1 / 4

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux terminal window titled "Terminal". The window shows the command-line interface for installing the "gedit" package. The output of the command "sudo yum install gedit" is displayed, including dependency resolution, package download progress, transaction tests, and final installed packages.

```
aluno@dataserver:/home/aluno
=====
Instalando:
  gedit           x86_64        2:3.14.3-9.el7      base       2.5 M
Instalando para as dependências:
  gtksourceview3  x86_64        3.14.3-1.el7      base       946 k
  libpeas         x86_64        1.12.1-1.el7      base       119 k

Resumo da transação
=====
Instalar 1 Package (+2 Dependent packages)

Tamanho total do download: 3.6 M
Tamanho depois de instalado: 19 M
Is this ok [y/d/N]: y
Downloading packages:
(1/3): libpeas-1.12.1-1.el7.x86_64.rpm | 119 kB 00:00:00
(2/3): gtksourceview3-3.14.3-1.el7.x86_64.rpm | 946 kB 00:00:01
(3/3): gedit-3.14.3-9.el7.x86_64.rpm | 2.5 MB 00:00:02
Total                                         1.3 MB/s | 3.6 MB 00:00:02

Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Instalando : libpeas-1.12.1-1.el7.x86_64                               1/3
  Instalando : gtksourceview3-3.14.3-1.el7.x86_64                         2/3
  Instalando : 2:gedit-3.14.3-9.el7.x86_64                           3/3
  Verificando: gtksourceview3-3.14.3-1.el7.x86_64                         1/3
  Verificando: 2:gedit-3.14.3-9.el7.x86_64                           2/3
  Verificando: libpeas-1.12.1-1.el7.x86_64                           3/3

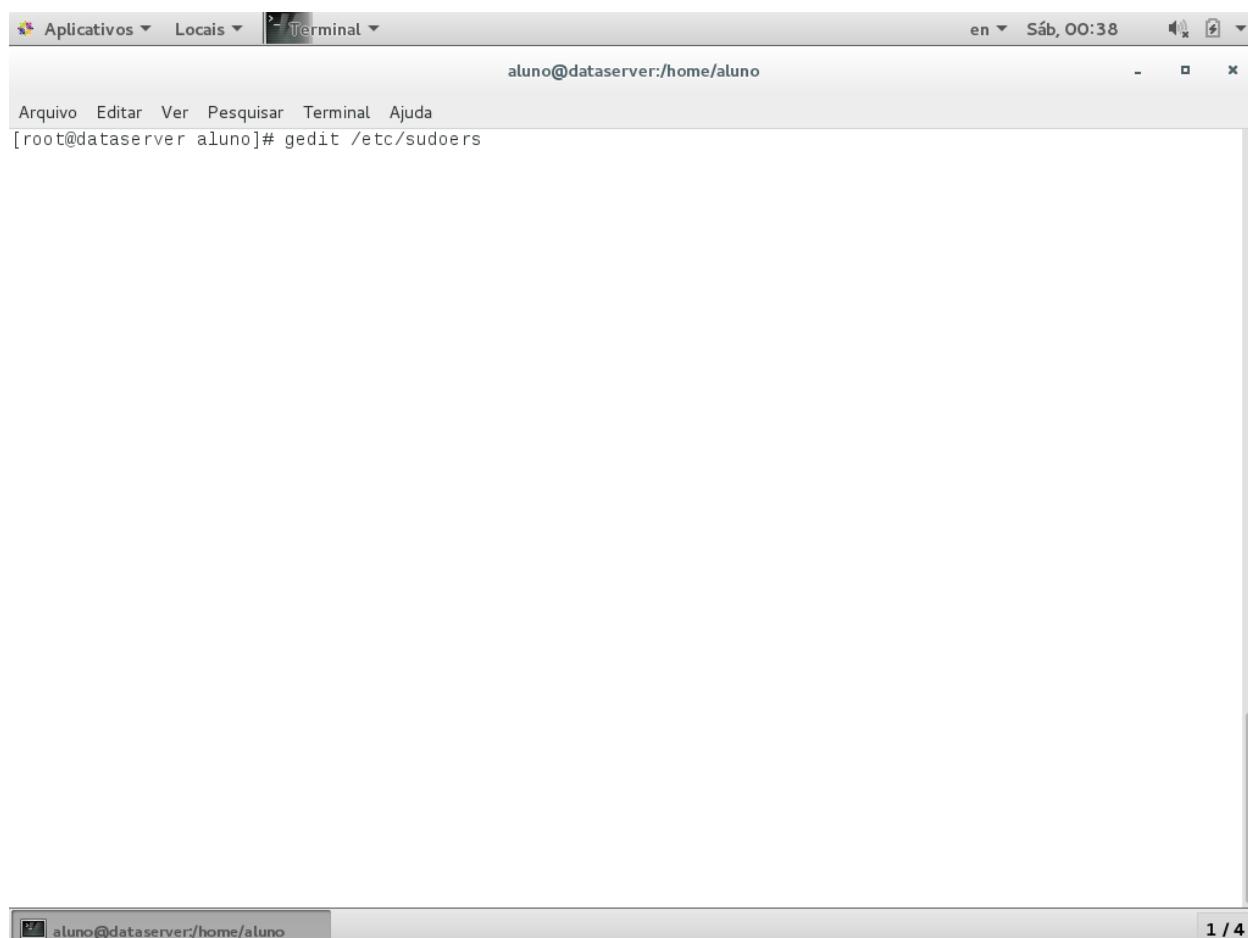
Instalados:
  gedit.x86_64 2:3.14.3-9.el7

Dependência(s) instalada(s):
  gtksourceview3.x86_64 0:3.14.3-1.el7                                libpeas.x86_64 0:1.12.1-1.el7

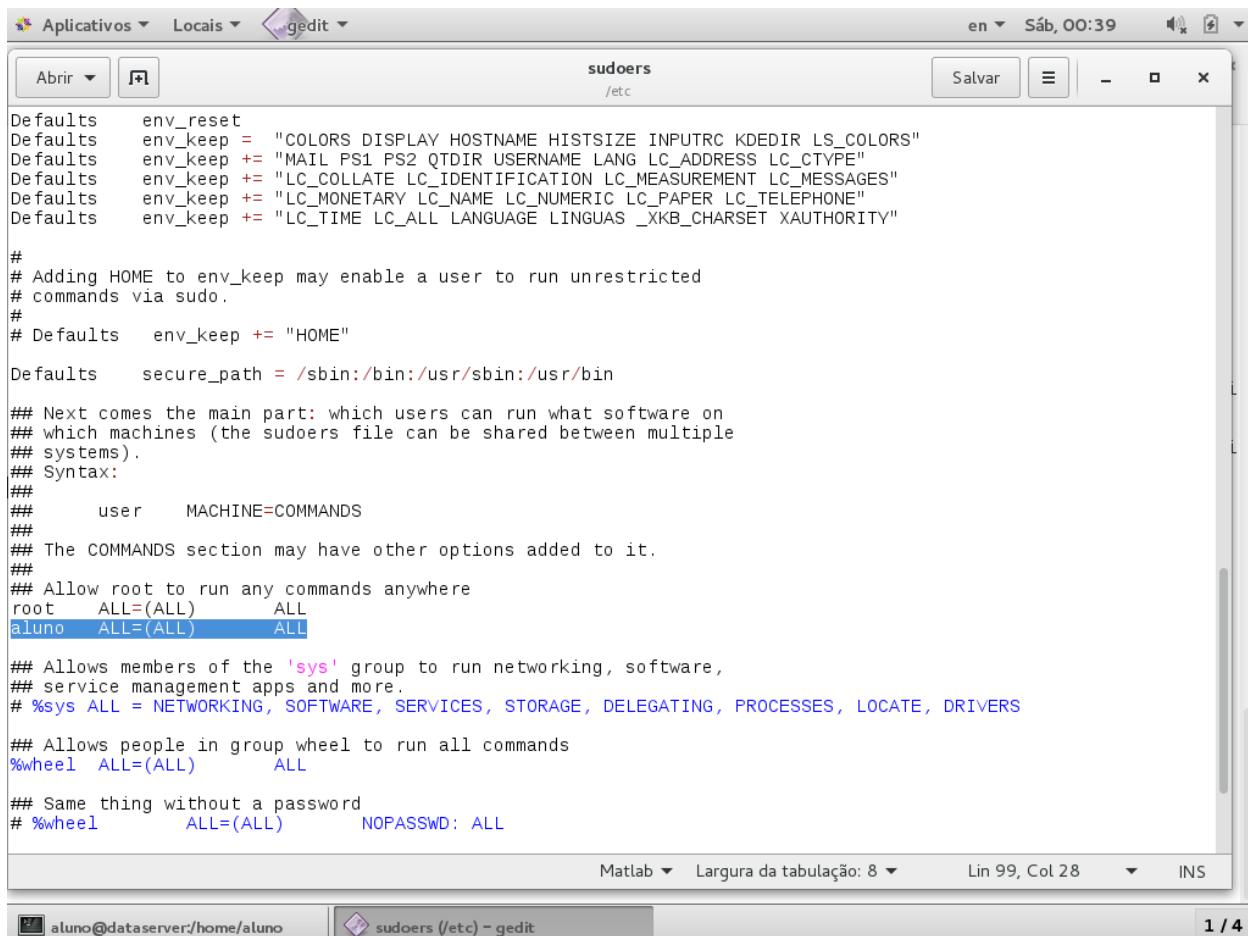
Concluído!
[root@dataserver aluno]#
```

gedit instalado

1 / 4



Editar o arquivo /etc/sudoers usando o gedit



```

Defaults    env_reset
Defaults    env_keep  = "COLORS DISPLAY HOSTNAME HISTSIZE INPUTRC KDEDIR LS_COLORS"
Defaults    env_keep += "MAIL PS1 PS2 QTDIR USERNAME LANG LC_ADDRESS LC_CTYPE"
Defaults    env_keep += "LC_COLLATE LC_IDENTIFICATION LC_MEASUREMENT LC_MESSAGES"
Defaults    env_keep += "LC_MONETARY LC_NAME LC_NUMERIC LC_PAPER LC_TELEPHONE"
Defaults    env_keep += "LC_TIME LC_ALL LANGUAGE LINGUAS _XKB_CHARSET XAUTHORITY"

#
# 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
%wheel   ALL=(ALL)      ALL

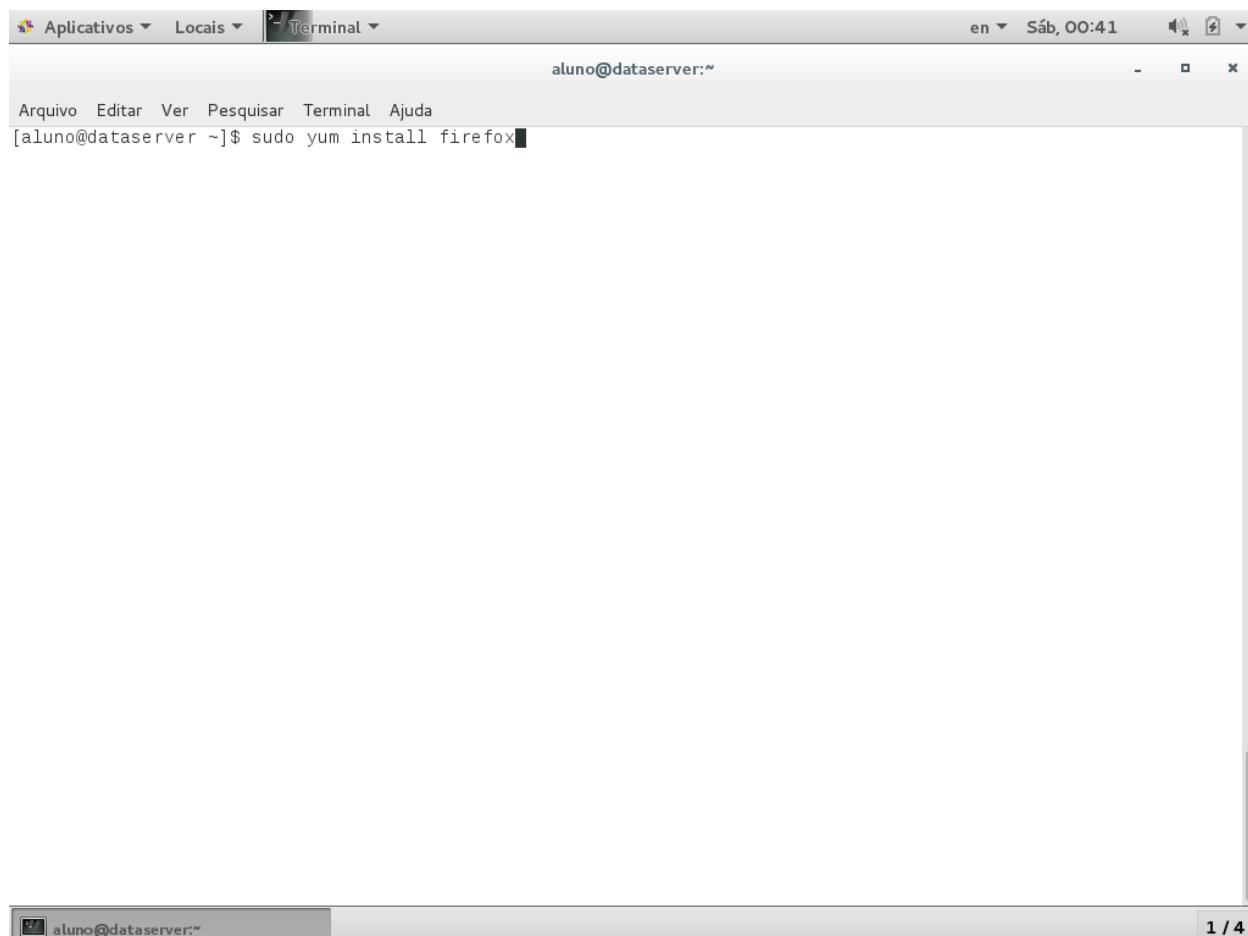
## Same thing without a password
# %wheel     ALL=(ALL)      NOPASSWD: ALL

```

Matlab ▾ Largura da tabulação: 8 ▾ Lin 99, Col 28 ▾ INS

aluno@dataserver:/home/aluno | sudoers (/etc) - gedit 1 / 4

Incluir no arquivo, a linha marcada acima e salvar o arquivo. Isso permitirá o usuário aluno executar comandos de administrador (root)



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The status bar at the top shows 'en' and 'Sáb, 00:41'. The terminal menu bar includes 'Arquivo', 'Editar', 'Ver', 'Pesquisar', 'Terminal', and 'Ajuda'. The command line shows the user 'aluno@dataserver ~\$' followed by the command 'sudo yum install firefox'. The bottom status bar shows the user 'aluno@dataserver ~\$' and the page number '1 / 4'.

Conectado como usuário aluno, instalar o Firefox com o comando: **sudo yum install firefox**

```

Aplicativos Locais Terminal
aluno@dataserver:~ - x

Arquivo Editar Ver Pesquisar Terminal Ajuda
liberation-sans-fonts      noarch      1:1.07.2-15.el7      base      279 k
libvpx                      x86_64      1.3.0-5.el7_0       base      498 k

Resumo da transação
=====
Instalar 1 Package (+3 Dependent packages)

Tamanho total do download: 72 M
Tamanho depois de instalado: 133 M
Is this ok [y/d/N]: y
Downloading packages:
(1/4): centos-indexhtml-7-9.el7.centos.noarch.rpm | 92 kB 00:00:00
(2/4): liberation-sans-fonts-1.07.2-15.el7.noarch.rpm | 279 kB 00:00:00
(3/4): libvpx-1.3.0-5.el7_0.x86_64.rpm           | 498 kB 00:00:01
(4/4): firefox-38.6.0-1.el7.centos.x86_64.rpm     | 72 MB 00:00:25
-----
Total                                         2.9 MB/s | 72 MB 00:00:25

Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Instalando : 1:liberation-sans-fonts-1.07.2-15.el7.noarch          1/4
  Instalando : centos-indexhtml-7-9.el7.centos.noarch                2/4
  Instalando : libvpx-1.3.0-5.el7_0.x86_64                         3/4
  Instalando : firefox-38.6.0-1.el7.centos.x86_64                  4/4
  Verifying   : libvpx-1.3.0-5.el7_0.x86_64                         1/4
  Verifying   : centos-indexhtml-7-9.el7.centos.noarch                2/4
  Verifying   : firefox-38.6.0-1.el7.centos.x86_64                  3/4
  Verifying   : 1:liberation-sans-fonts-1.07.2-15.el7.noarch          4/4

Instalados:
  firefox.x86_64 0:38.6.0-1.el7.centos

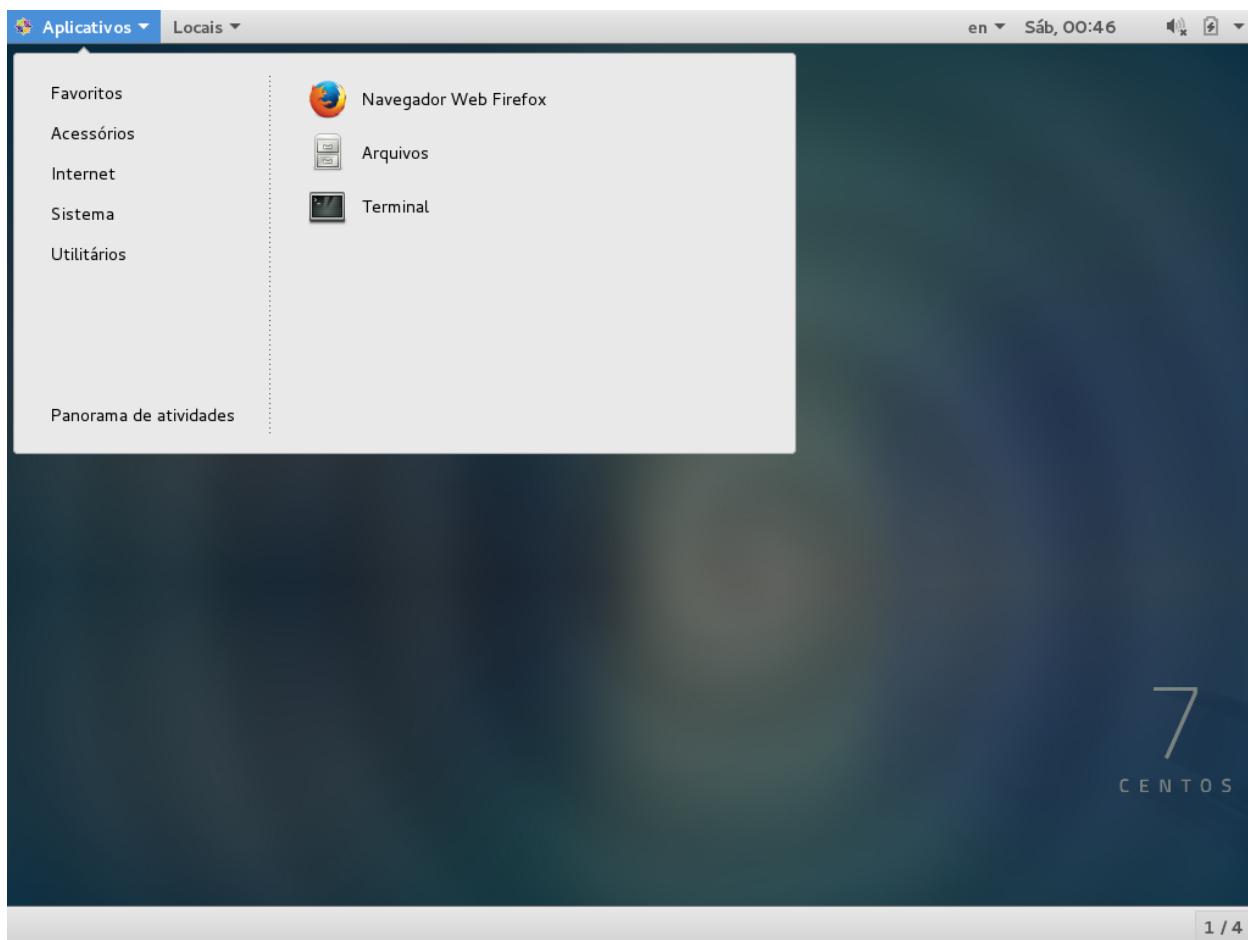
Dependência(s) instalada(s):
  centos-indexhtml.noarch 0:7-9.el7.centos           liberation-sans-fonts.noarch 1:1.07.2-15.el7
  libvpx.x86_64 0:1.3.0-5.el7_0

Concluído!
[aluno@dataserver ~]$ 

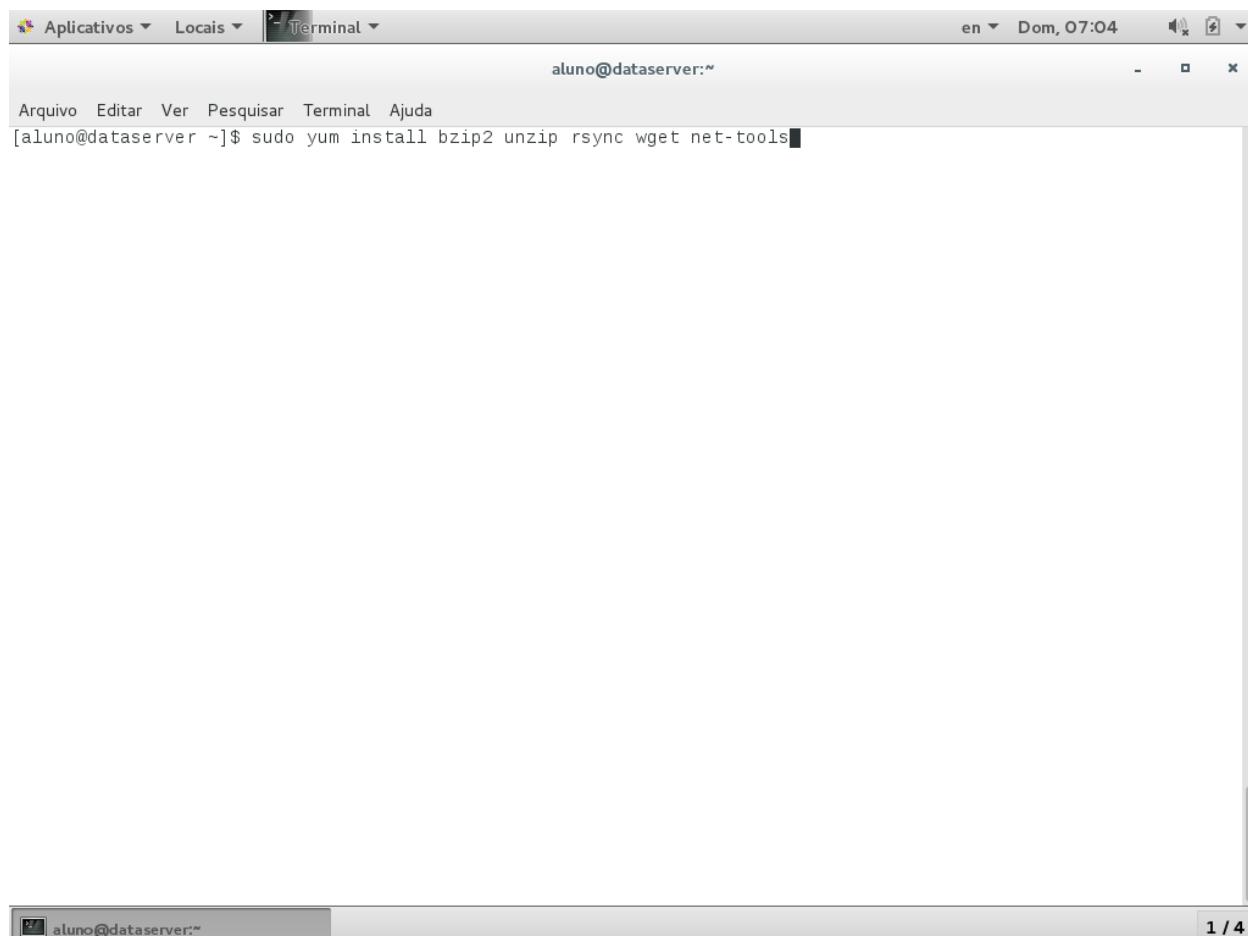
```

Instalação concluída

1 / 4



Firefox instalado



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en Dom, 07:04". The terminal prompt is "aluno@dataserver:~". Below the prompt, the user has typed the command: "[aluno@dataserver ~]\$ sudo yum install bzip2 unzip rsync wget net-tools". The terminal window is part of a larger desktop interface with other application icons visible in the top bar.

Instalar outros aplicativos: bzip2, unzip, rsync, wget e net-tools

1 / 4

```

Aplicativos Locais Terminal
aluno@dataserver:~ - x

Arquivo Editar Ver Pesquisar Terminal Ajuda
unzip           x86_64      6.0-15.el7          base      166 k
wget            x86_64      1.14-10.el7_0.1    base      545 k

Resumo da transação
=====
Instalar 5 Packages

Tamanho total do download: 1.4 M
Tamanho depois de instalado: 4.0 M
Is this ok [y/d/N]: y
Downloading packages:
(1/5): bzip2-1.0.6-13.el7.x86_64.rpm | 52 kB 00:00:00
(2/5): wget-1.14-10.el7_0.1.x86_64.rpm | 545 kB 00:00:00
(3/5): unzip-6.0-15.el7.x86_64.rpm   | 166 kB 00:00:00
(4/5): net-tools-2.0-0.17.20131004git.el7.x86_64.rpm | 304 kB 00:00:01
(5/5): rsync-3.0.9-17.el7.x86_64.rpm | 360 kB 00:00:02
-----
Total                                         666 kB/s | 1.4 MB 00:00:02

Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Instalando : bzip2-1.0.6-13.el7.x86_64          1/5
  Instalando : net-tools-2.0-0.17.20131004git.el7.x86_64 2/5
  Instalando : wget-1.14-10.el7_0.1.x86_64        3/5
  Instalando : rsync-3.0.9-17.el7.x86_64         4/5
  Instalando : unzip-6.0-15.el7.x86_64          5/5
  Verificando: bzip2-1.0.6-13.el7.x86_64          1/5
  Verificando: net-tools-2.0-0.17.20131004git.el7.x86_64 2/5
  Verificando: wget-1.14-10.el7_0.1.x86_64        3/5
  Verificando: rsync-3.0.9-17.el7.x86_64         4/5
  Verificando: unzip-6.0-15.el7.x86_64          5/5

Instalados:
  bzip2.x86_64 0:1.0.6-13.el7      net-tools.x86_64 0:2.0-0.17.20131004git.el7      rsync.x86_64 0:3.0.9-17.el7
  unzip.x86_64 0:6.0-15.el7       wget.x86_64 0:1.14-10.el7_0.1

Concluído!
[aluno@dataserver ~]$ 

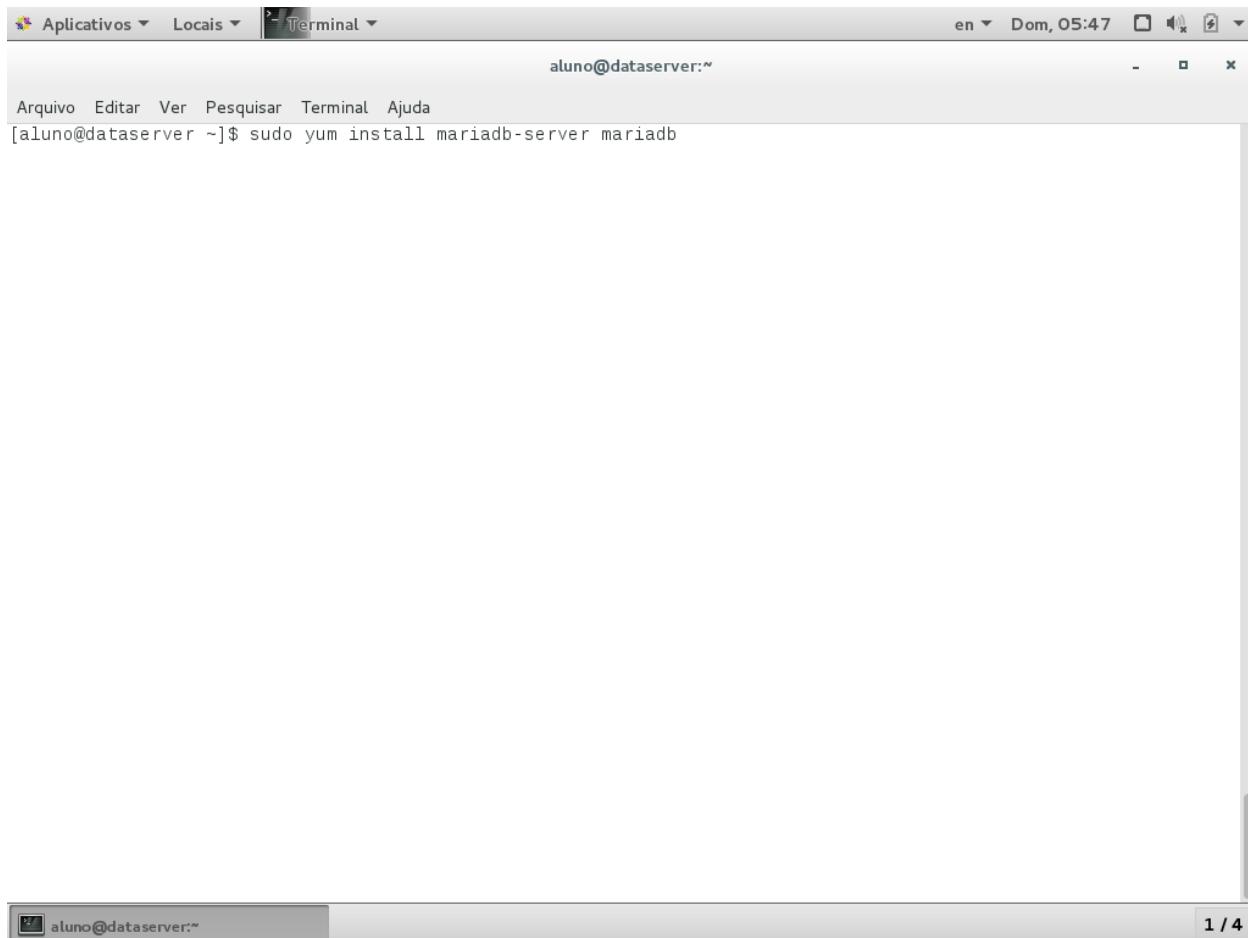
```

Aplicativos instalados

1 / 4

2.5. Instalação do MySQL

Obs: O CentOS 7 substituiu o MySQL (banco de dados relacional) pelo MariaDB. Mas ainda é possível usar comandos MySQL com o MariaDB, que na prática são muito parecidos. Usaremos o MariaDB para fazer os testes de importação de dados de bancos de dados relacionais para o Hadoop.



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~'. The user has typed the command 'sudo yum install mariadb-server mariadb' and is waiting for the output. The desktop interface includes a menu bar with 'Aplicativos', 'Locais', and 'Terminal'. The system tray shows 'en Dom, 05:47' and various icons. The bottom status bar shows 'aluno@dataserver ~' and '1 / 4'.

Instalação do MariaDB

```

Aplicativos Locais Terminal
aluno@dataserver:~ - x

Arquivo Editar Ver Pesquisar Terminal Ajuda
(10/10): mariadb-server-5.5.44-2.el7.centos.x86_64.rpm | 11 MB 00:00:11
-----|-----|-----|-----|-----|-----|-----|-----|
Total 1.9 MB/s | 21 MB 00:00:11

Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
Instalando : perl-Data-Dumper-2.145-3.el7.x86_64 1/10
Instalando : perl-Compress-Raw-Zlib-2.061-4.el7.x86_64 2/10
Instalando : 1:mariadb-5.5.44-2.el7.centos.x86_64 3/10
Instalando : perl-Net-Daemon-0.48-5.el7.noarch 4/10
Instalando : perl-Compress-Raw-Bzip2-2.061-3.el7.x86_64 5/10
Instalando : perl-IO-Compress-2.061-2.el7.noarch 6/10
Instalando : perl-P1RPC-0.2020-14.el7.noarch 7/10
Instalando : perl-DBI-1.627-4.el7.x86_64 8/10
Instalando : perl-DBD-MySQL-4.023-5.el7.x86_64 9/10
Instalando : 1:mariadb-server-5.5.44-2.el7.centos.x86_64 10/10
Verifying : perl-Compress-Raw-Bzip2-2.061-3.el7.x86_64 1/10
Verifying : perl-Net-Daemon-0.48-5.el7.noarch 2/10
Verifying : perl-Data-Dumper-2.145-3.el7.x86_64 3/10
Verifying : 1:mariadb-5.5.44-2.el7.centos.x86_64 4/10
Verifying : 1:mariadb-server-5.5.44-2.el7.centos.x86_64 5/10
Verifying : perl-P1RPC-0.2020-14.el7.noarch 6/10
Verifying : 1:perl-Compress-Raw-Zlib-2.061-4.el7.x86_64 7/10
Verifying : perl-DBI-1.627-4.el7.x86_64 8/10
Verifying : perl-IO-Compress-2.061-2.el7.noarch 9/10
Verifying : perl-DBD-MySQL-4.023-5.el7.x86_64 10/10

Instalados:
mariadb.x86_64 1:5.5.44-2.el7.centos           mariadb-server.x86_64 1:5.5.44-2.el7.centos

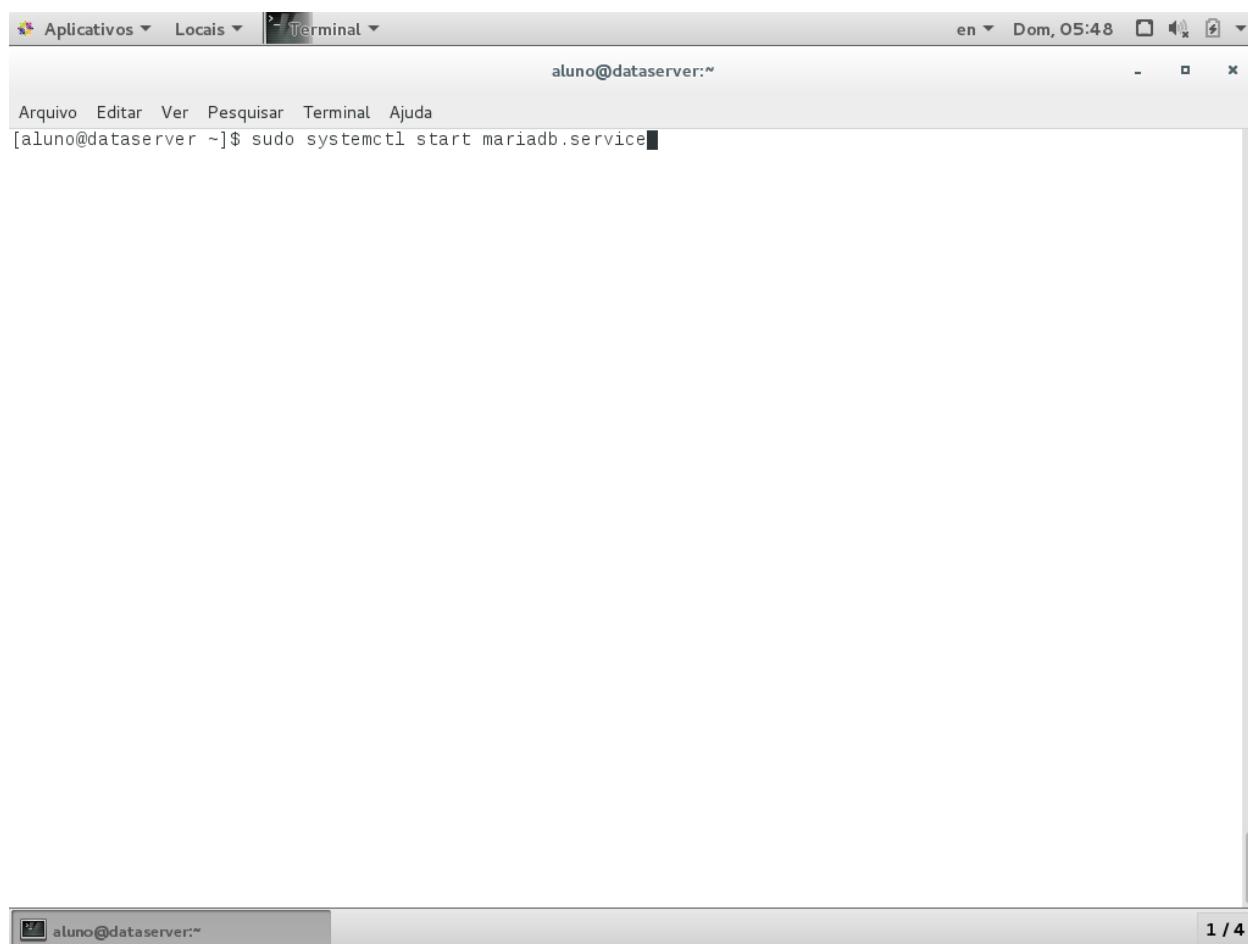
Dependência(s) instalada(s):
perl-Compress-Raw-Bzip2.x86_64 0:2.061-3.el7   perl-Compress-Raw-Zlib.x86_64 1:2.061-4.el7
perl-DBD-MySQL.x86_64 0:4.023-5.el7            perl-DBI.x86_64 0:1.627-4.el7
perl-Data-Dumper.x86_64 0:2.145-3.el7          perl-IO-Compress.noarch 0:2.061-2.el7
perl-Net-Daemon.noarch 0:0.48-5.el7             perl-P1RPC.noarch 0:0.2020-14.el7

Concluído!
[aluno@dataserver ~]$ ■

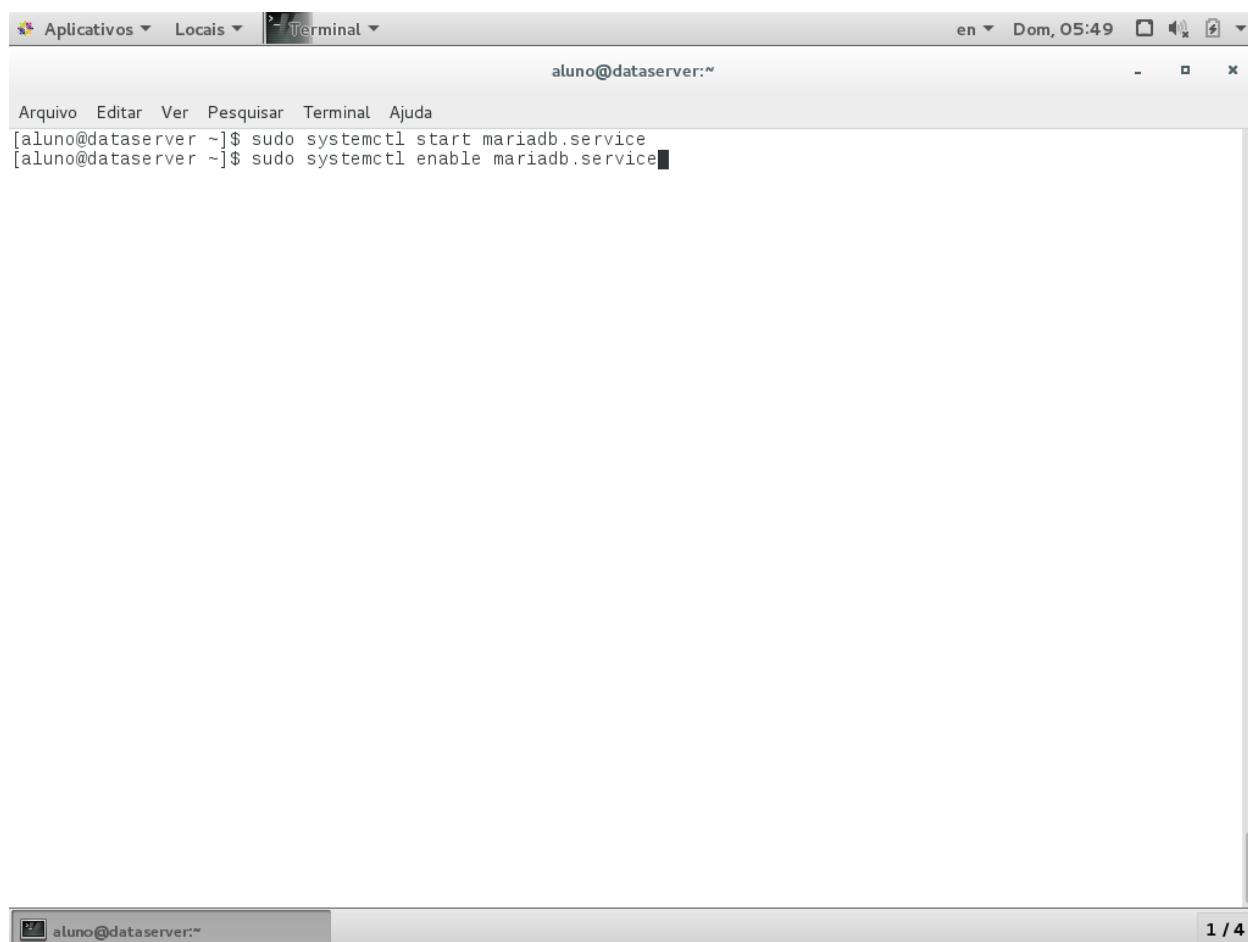
```

Instalação concluída

1 / 4



Iniciando o serviço

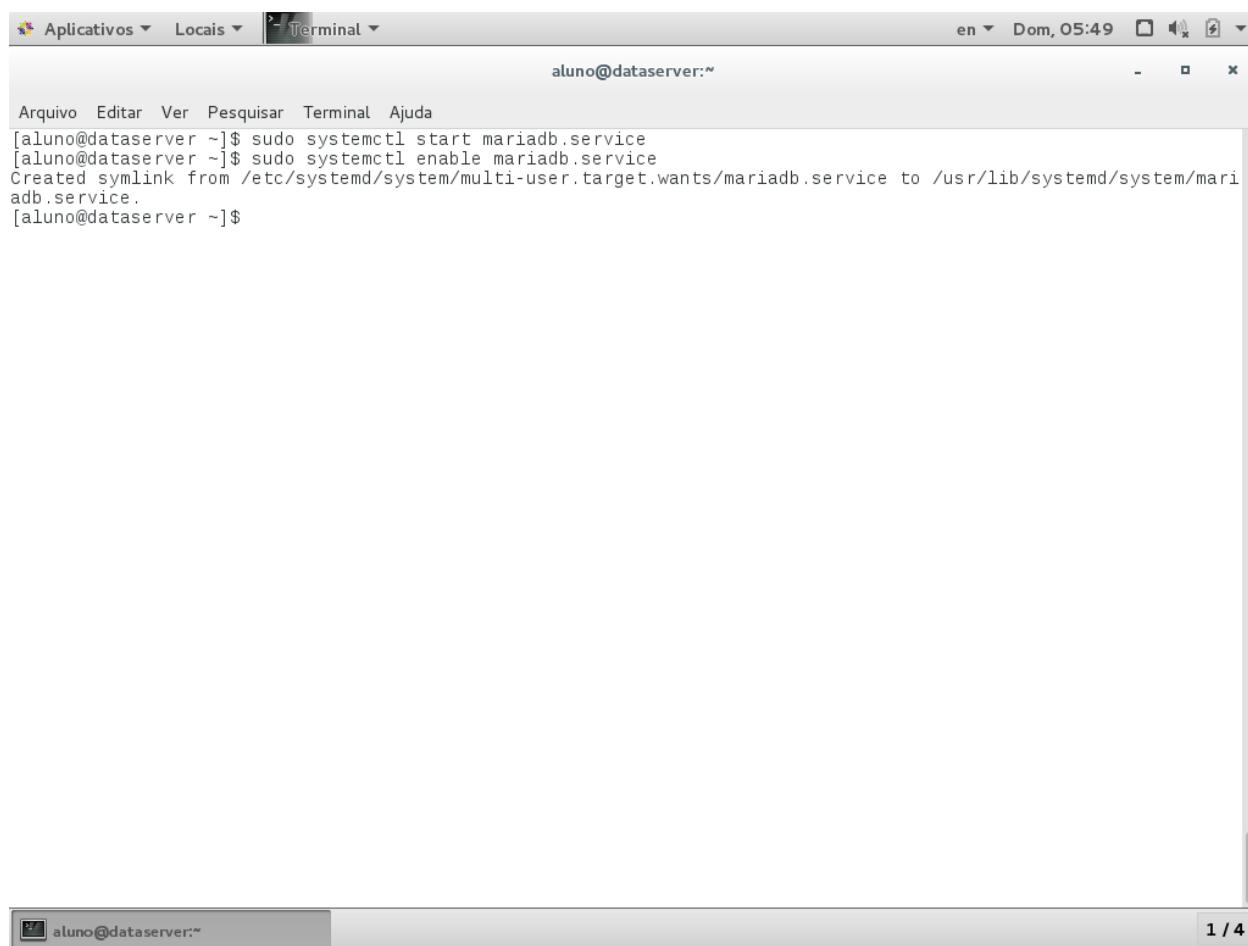


A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~". The user has run the following commands:

```
[aluno@dataserver ~]$ sudo systemctl start mariadb.service
[aluno@dataserver ~]$ sudo systemctl enable mariadb.service
```

The terminal window is part of a larger interface with a menu bar at the top and a status bar at the bottom indicating "1 / 4".

Habilitando a inicialização no boot

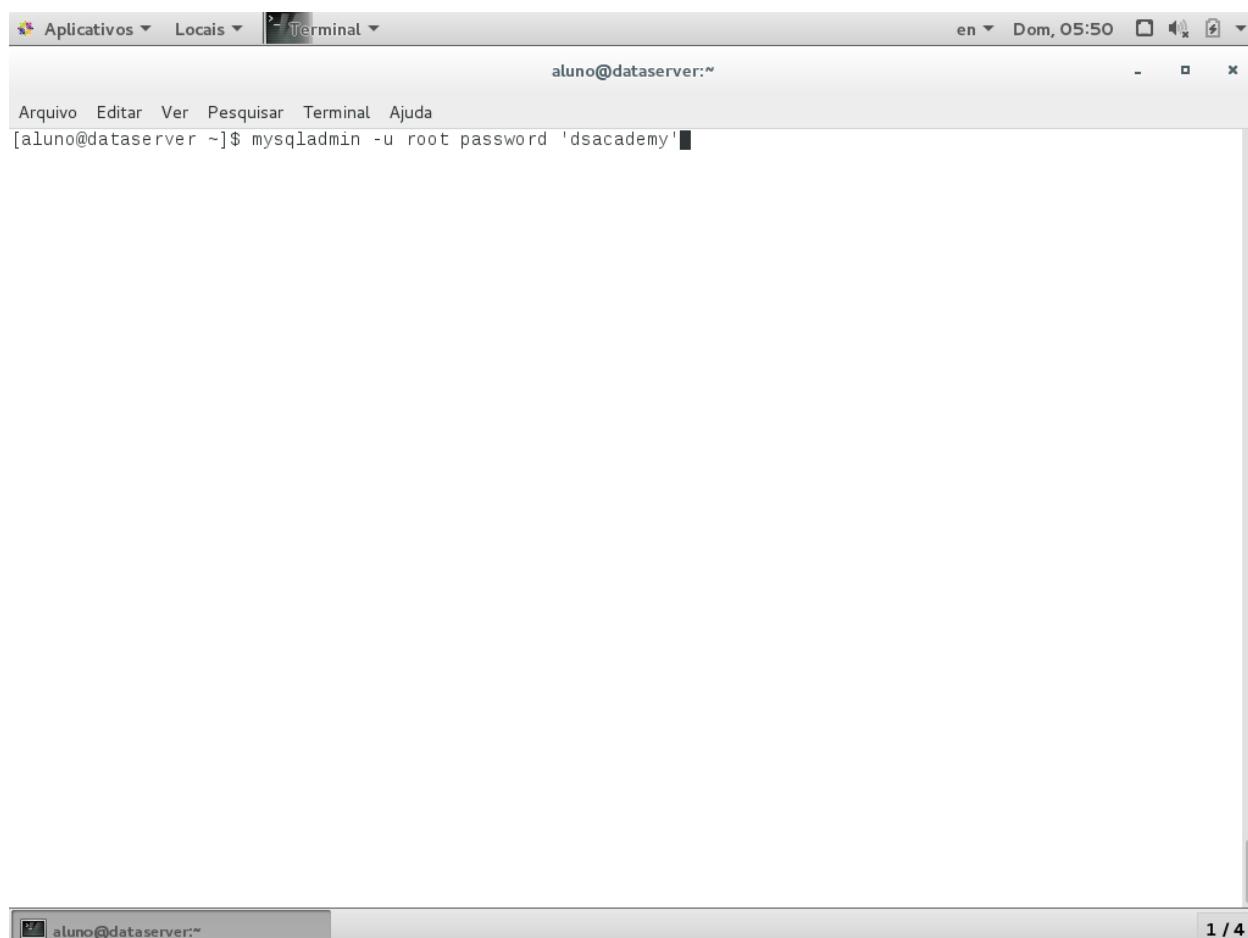


A screenshot of a Linux terminal window titled "Terminal". The window shows a command-line session where the user is enabling the MariaDB service. The session starts with the user's name "aluno" followed by the command "sudo systemctl start mariadb.service", then "sudo systemctl enable mariadb.service", and finally a message indicating that a symlink was created from /etc/systemd/system/multi-user.target.wants/mariadb.service to /usr/lib/systemd/system/mariadb.service. The terminal window has a standard top bar with icons for applications, locations, and system status, and a bottom status bar.

```
[aluno@dataserver ~]$ sudo systemctl start mariadb.service
[aluno@dataserver ~]$ sudo systemctl enable mariadb.service
Created symlink from /etc/systemd/system/multi-user.target.wants/mariadb.service to /usr/lib/systemd/system/mariadb.service.
[aluno@dataserver ~]$
```

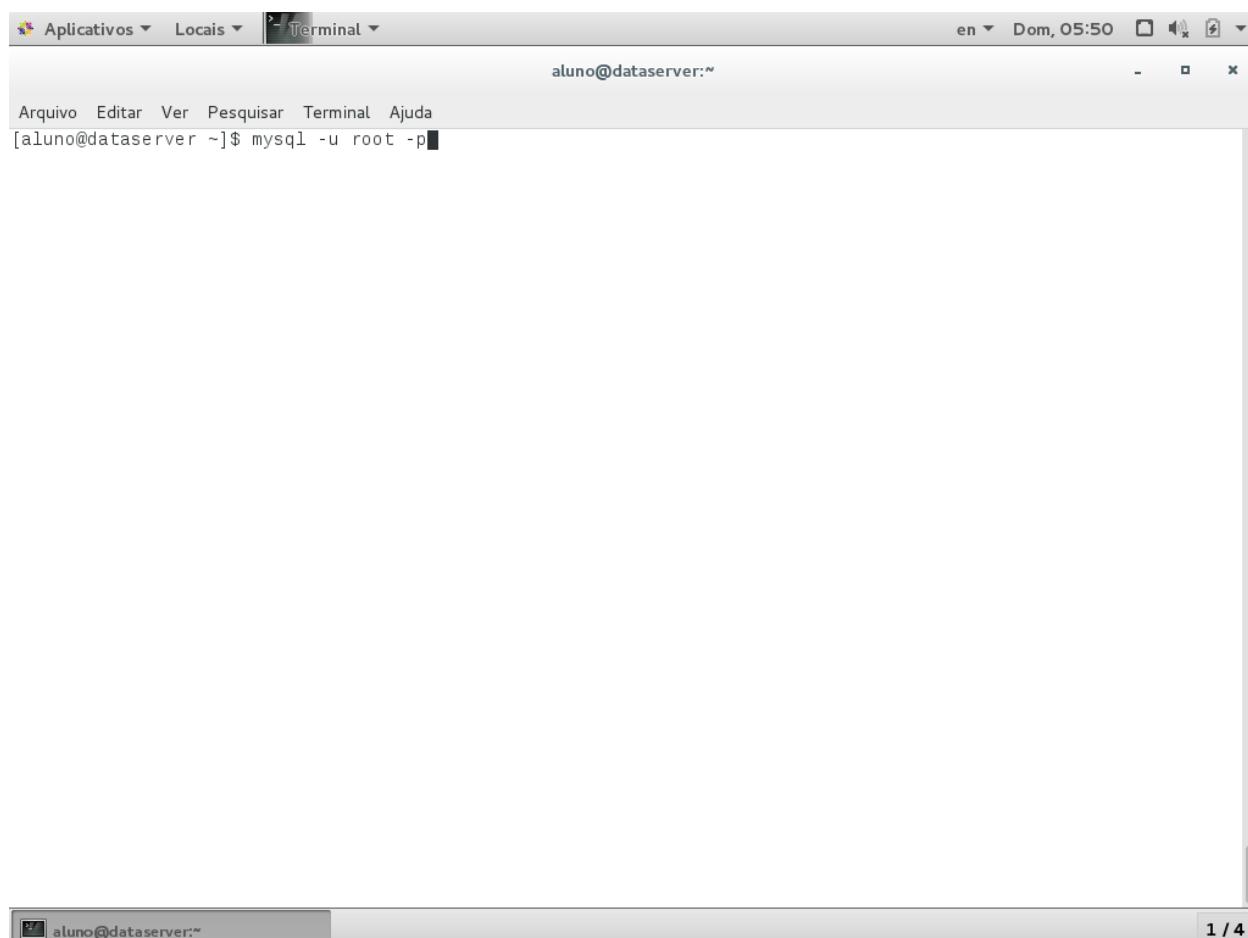
Habilitado

1 / 4

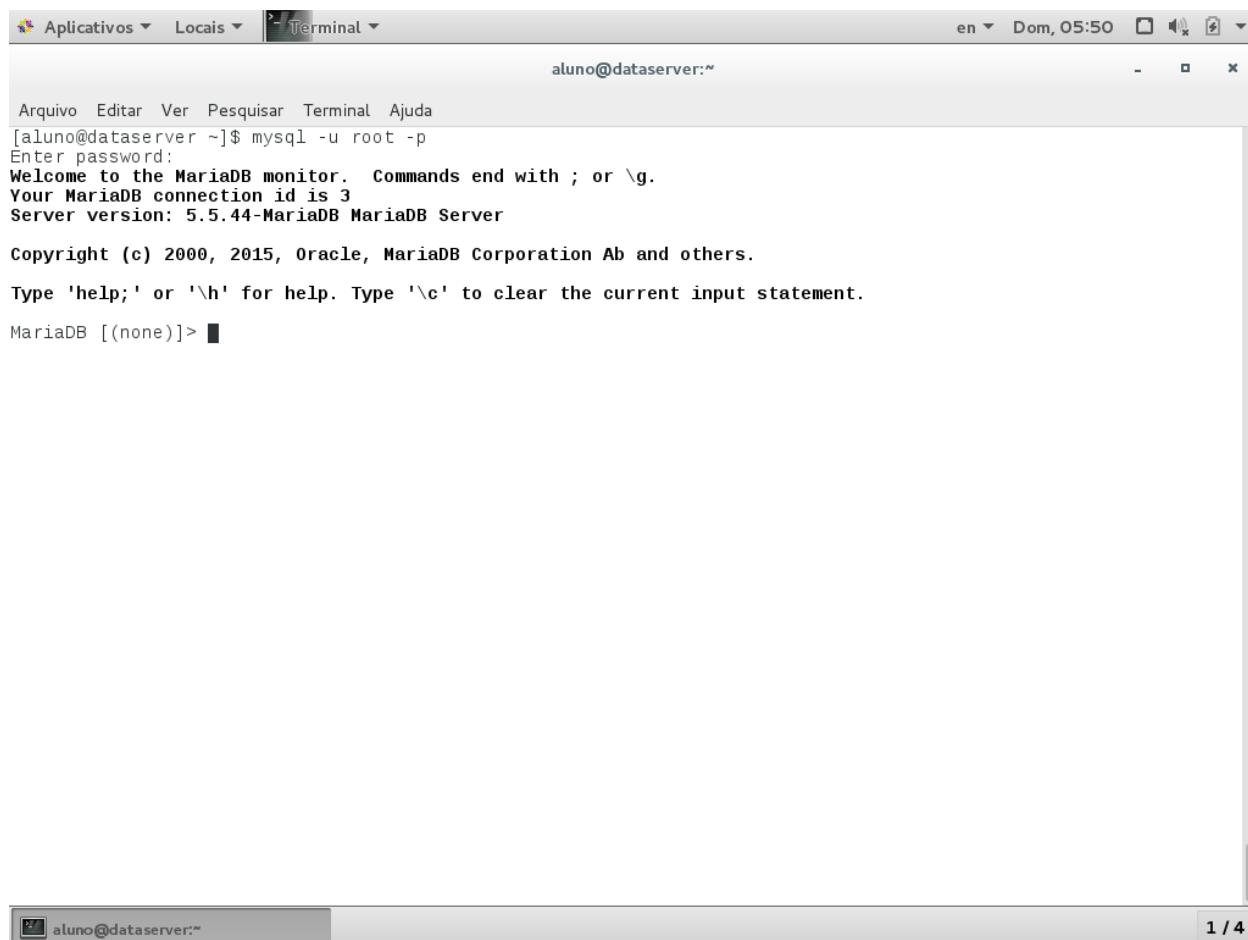


A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~". The user has typed the command "[aluno@dataserver ~]\$ mysqladmin -u root password 'dsacademy'" into the terminal. The terminal interface includes a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows "aluno@dataserver:~" and "1 / 4".

Definindo a senha do administrador (senha: dsacademy)



Iniciando o console



A screenshot of a terminal window titled "Terminal". The window shows a MySQL session connected as "aluno" to a MariaDB monitor. The session starts with the command "mysql -u root -p". It prompts for a password and displays the welcome message: "Welcome to the MariaDB monitor. Commands end with ; or \g.". It also shows the connection ID (3), server version (5.5.44-MariaDB), and copyright information. The prompt "MariaDB [(none)]>" is visible at the bottom.

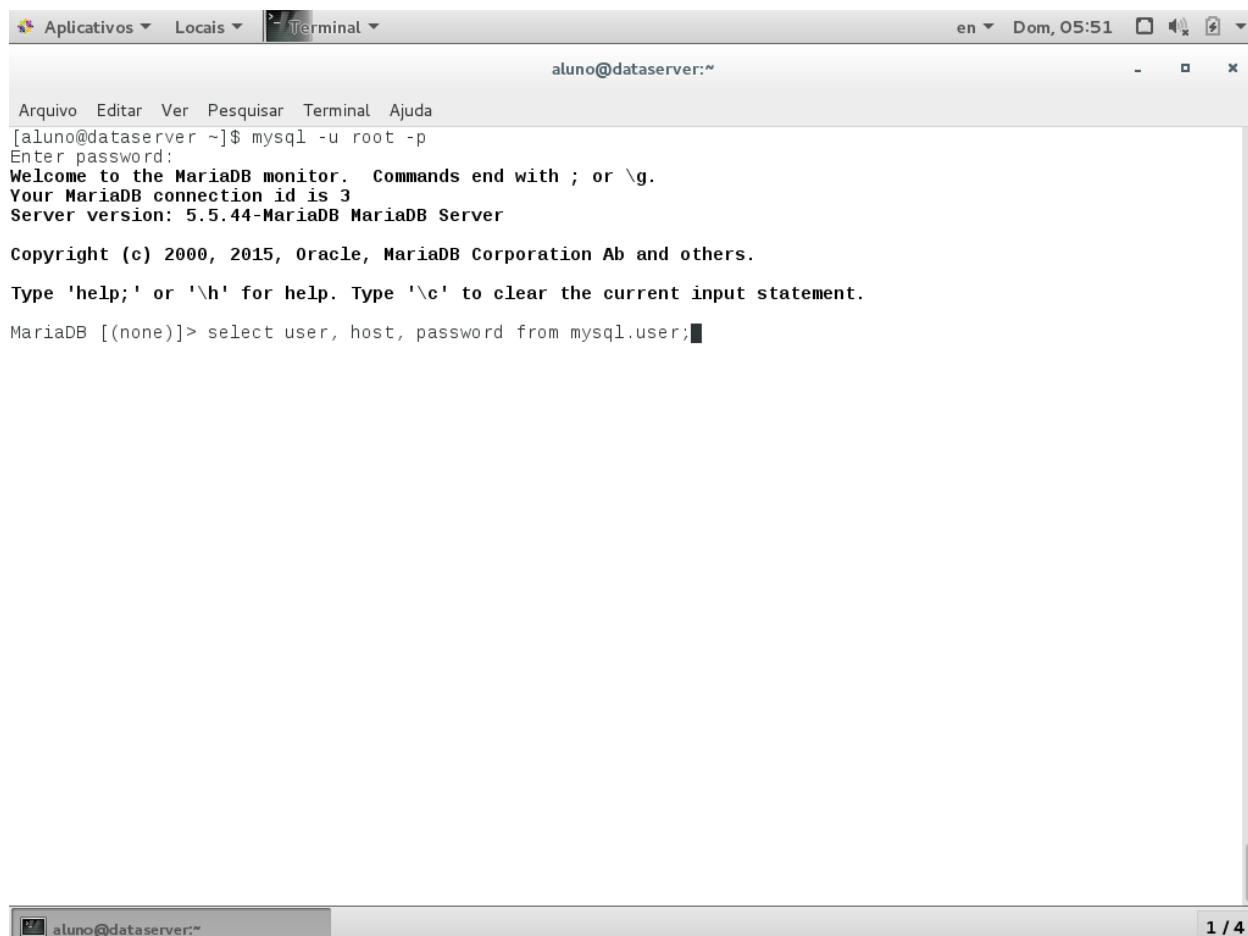
```
Arquivo Editar Ver Pesquisar Terminal Ajuda
[aluno@dataserver ~]$ mysql -u root -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MariaDB connection id is 3
Server version: 5.5.44-MariaDB MariaDB Server

Copyright (c) 2000, 2015, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

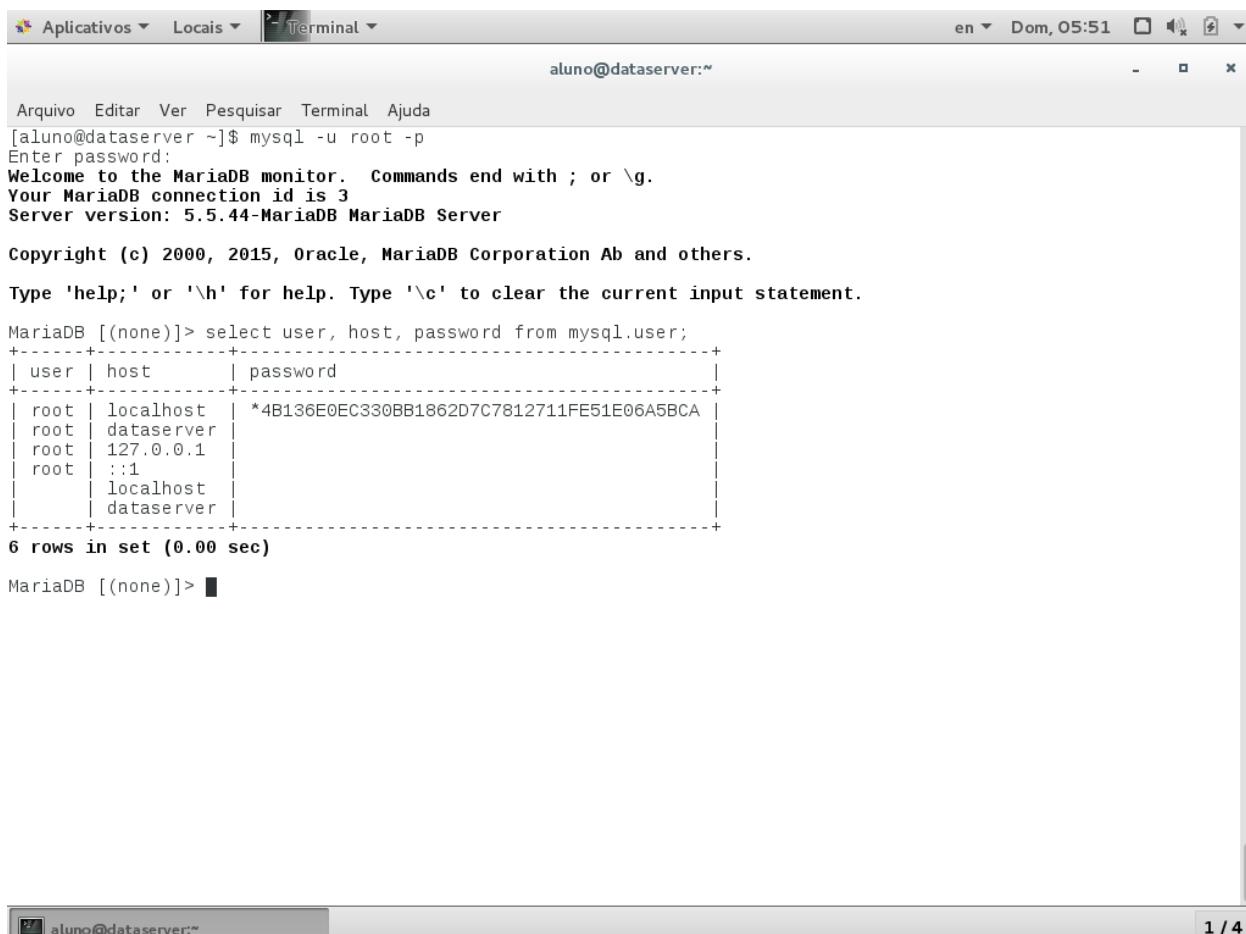
MariaDB [(none)]>
```

Sucesso!!!



A screenshot of a terminal window titled "Terminal". The window shows a MySQL session on a MariaDB monitor. The user is connected as "aluno" at "dataserver". The session starts with the MySQL prompt "[aluno@dataserver ~]\$". The user enters "mysql -u root -p" and is prompted for a password. The MySQL monitor displays its welcome message, connection information ("Your MariaDB connection id is 3"), and server version ("Server version: 5.5.44-MariaDB MariaDB Server"). The user then types "select user, host, password from mysql.user;" and presses Enter. The terminal window has a standard OS X-style interface with a menu bar and status bar.

Executando uma query



A screenshot of a Linux terminal window titled "Terminal". The window shows a MySQL session on a MariaDB monitor. The user is connected as "root" from "localhost" with connection ID 3. The server version is 5.5.44-MariaDB. The user runs a query to select users from the "mysql.user" table, resulting in 6 rows. The output shows the user, host, and password for each entry.

```
Arquivo Editar Ver Pesquisar Terminal Ajuda
[aluno@dataserver ~]$ mysql -u root -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MariaDB connection id is 3
Server version: 5.5.44-MariaDB MariaDB Server

Copyright (c) 2000, 2015, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

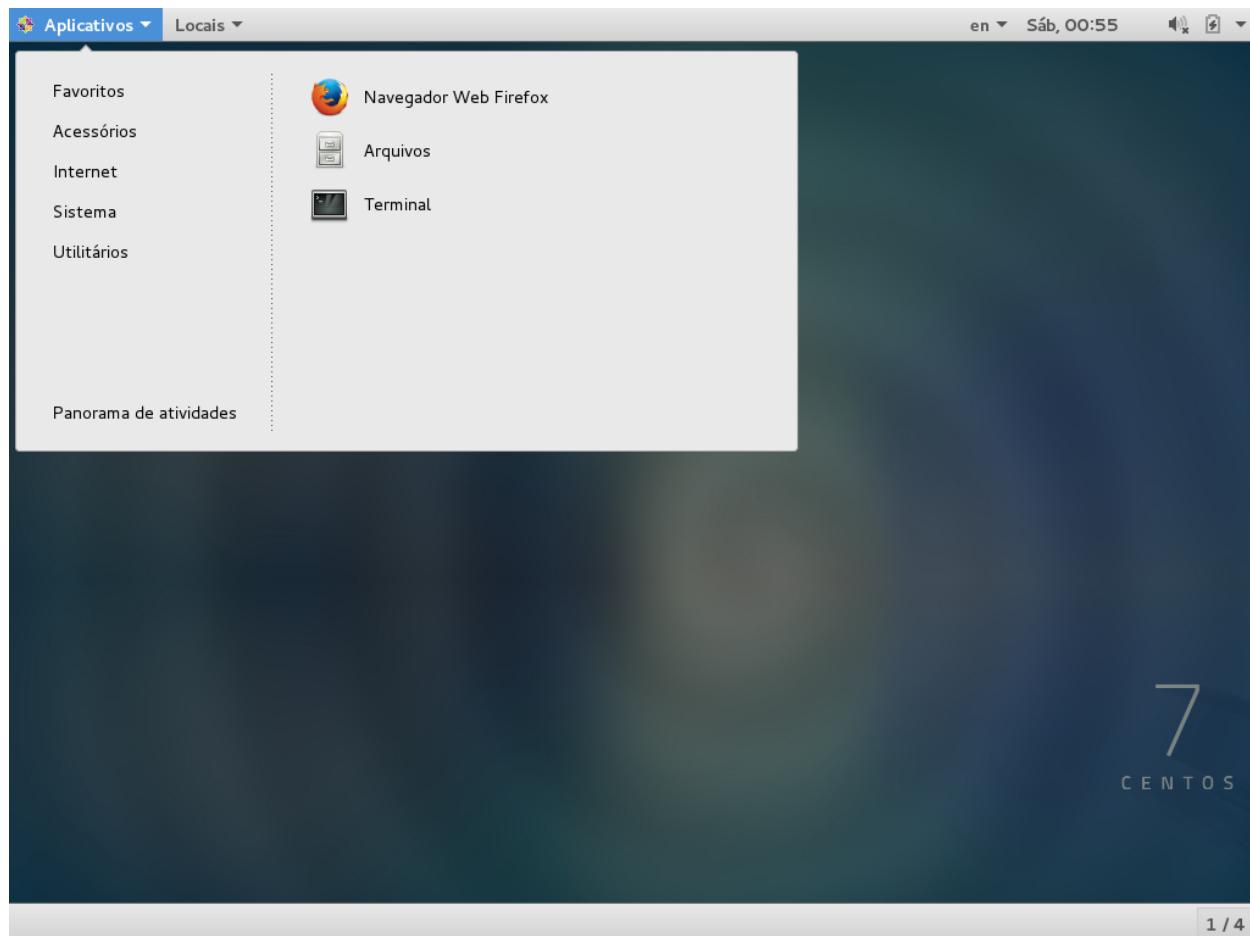
MariaDB [(none)]> select user, host, password from mysql.user;
+-----+-----+
| user | host   | password          |
+-----+-----+
| root | localhost | *4B136E0EC330BB1862D7C7812711FE51E06A5BCA |
| root | dataserver |                               |
| root | 127.0.0.1  |                               |
| root | ::1       |                               |
|      | localhost |                               |
|      | dataserver |                               |
+-----+-----+
6 rows in set (0.00 sec)

MariaDB [(none)]>
```

Query executada com sucesso

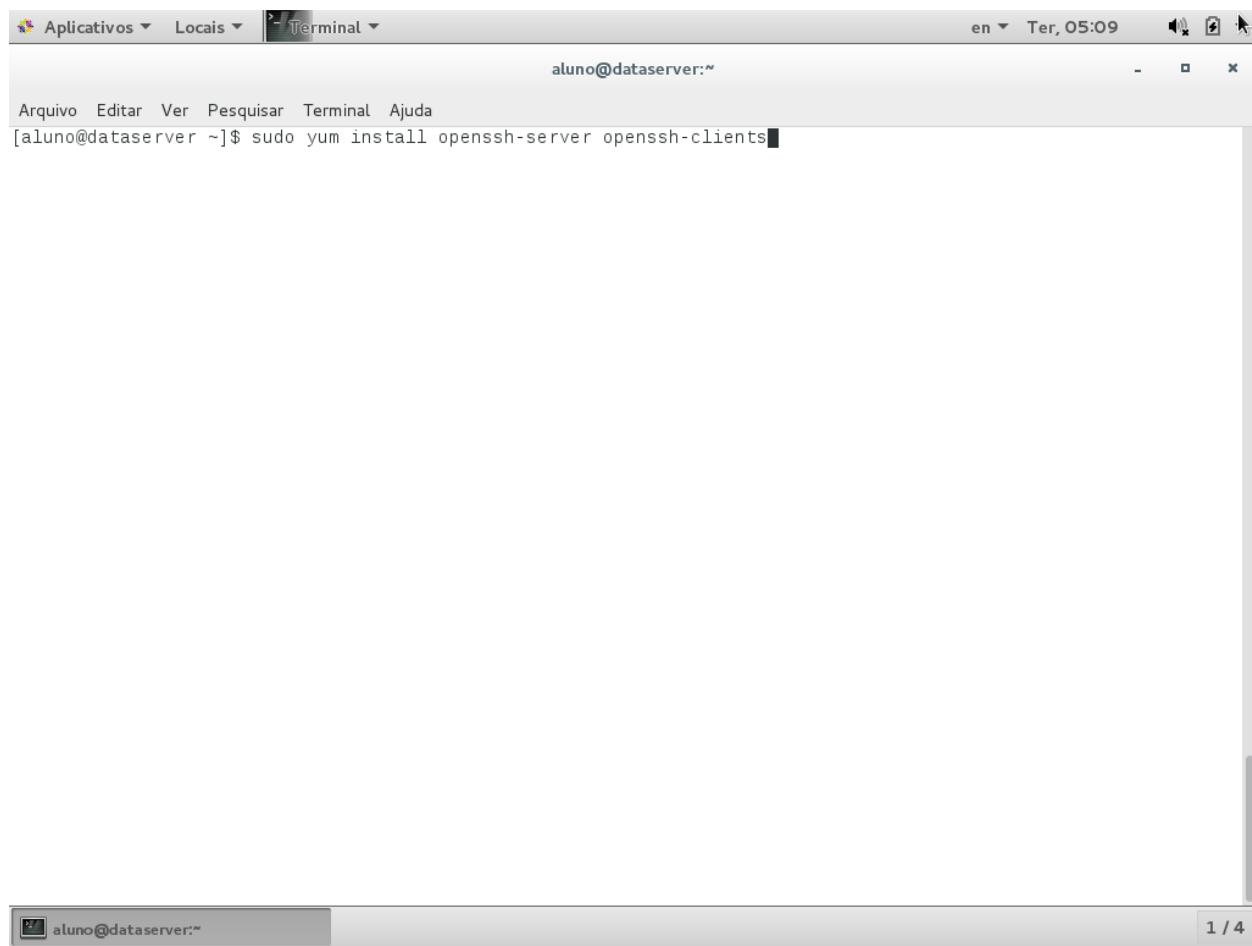
1 / 4

3. Instalação do servidor ssh



Abrindo o terminal

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The status bar at the top right shows 'en Ter, 05:09'. The terminal prompt is 'aluno@dataserver:~'. Below the prompt, the command 'sudo yum install openssh-server openssh-clients' is visible. The bottom left corner of the terminal window shows the same prompt 'aluno@dataserver:~'. The bottom right corner of the window has a status bar with '1 / 4'.

```
aluno@dataserver:~$ sudo yum install openssh-server openssh-clients
```

sudo yum install openssh-server openssh-clients

```

Aplicativos Locais Terminal
aluno@dataserver:~ - x

Arquivo Editar Ver Pesquisar Terminal Ajuda

Resumo da transação
=====
Upgrade 2 Packages (+1 Dependent package)

Tamanho total do download: 1.5 M
Is this ok [y/d/N]: y
Downloading packages:
Delta RPMs disabled because /usr/bin/applydeltarpm not installed.
(1/3): openssh-6.6.1p1-23.el7_2.x86_64.rpm | 435 kB 00:00:00
(2/3): openssh-server-6.6.1p1-23.el7_2.x86_64.rpm | 436 kB 00:00:01
(3/3): openssh-clients-6.6.1p1-23.el7_2.x86_64.rpm | 639 kB 00:00:02
Total                                         598 kB/s | 1.5 MB 00:00:02

Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Atualizando : openssh-6.6.1p1-23.el7_2.x86_64          1/6
  Atualizando : openssh-server-6.6.1p1-23.el7_2.x86_64    2/6
  Atualizando : openssh-clients-6.6.1p1-23.el7_2.x86_64   3/6
  Limpeza      : openssh-clients-6.6.1p1-22.el7.x86_64    4/6
  Limpeza      : openssh-server-6.6.1p1-22.el7.x86_64     5/6
  Limpeza      : openssh-6.6.1p1-22.el7.x86_64           6/6
  Verificando  : openssh-server-6.6.1p1-23.el7_2.x86_64   1/6
  Verificando  : openssh-clients-6.6.1p1-23.el7_2.x86_64  2/6
  Verificando  : openssh-6.6.1p1-23.el7_2.x86_64         3/6
  Verificando  : openssh-clients-6.6.1p1-22.el7.x86_64    4/6
  Verificando  : openssh-6.6.1p1-22.el7.x86_64           5/6
  Verificando  : openssh-server-6.6.1p1-22.el7.x86_64    6/6

Atualizados:
  openssh-clients.x86_64 0:6.6.1p1-23.el7_2           openssh-server.x86_64 0:6.6.1p1-23.el7_2

Dependência(s) atualizada(s):
  openssh.x86_64 0:6.6.1p1-23.el7_2

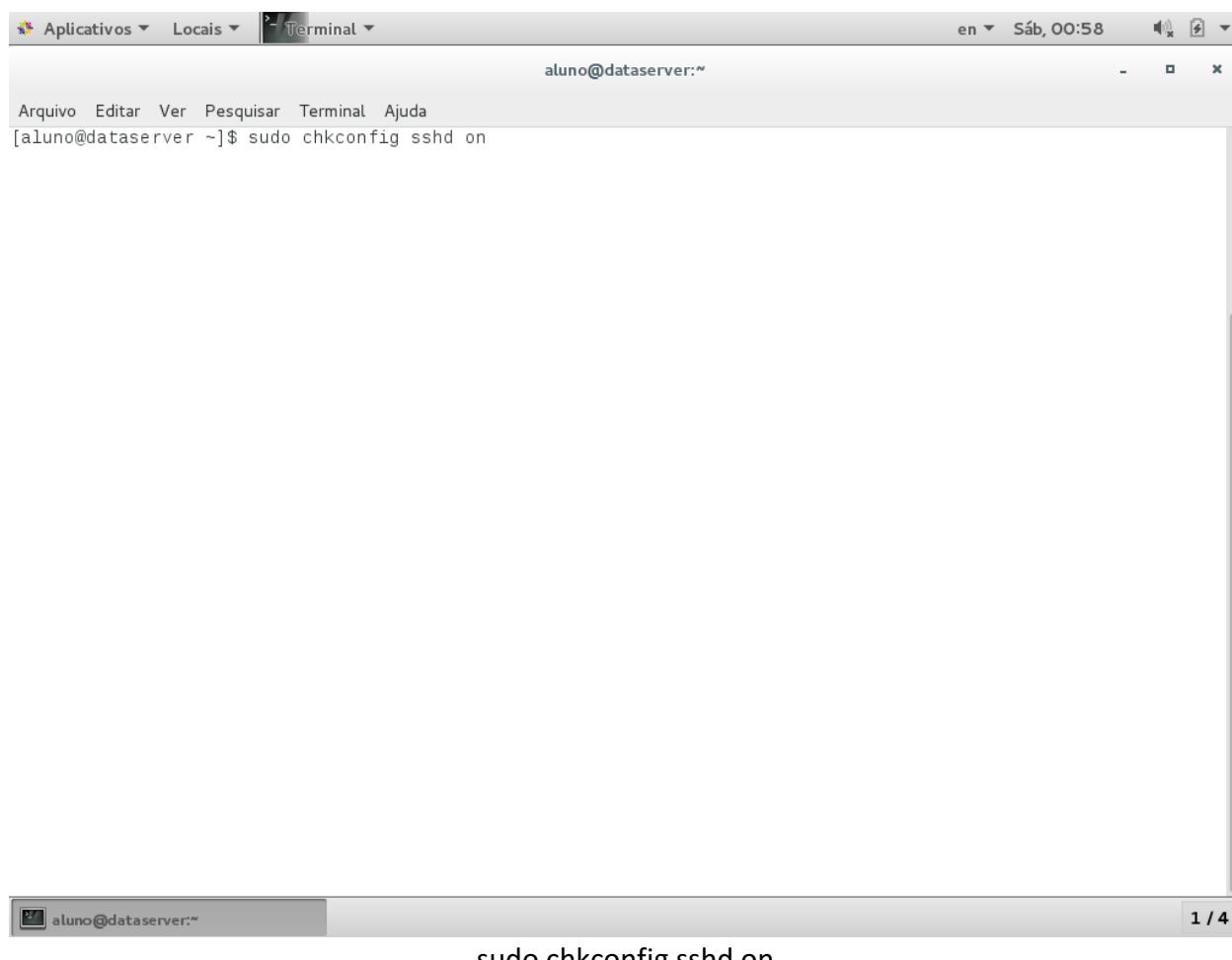
Concluído!
[aluno@dataserver ~]$ ■

```

Concluído

1 / 4

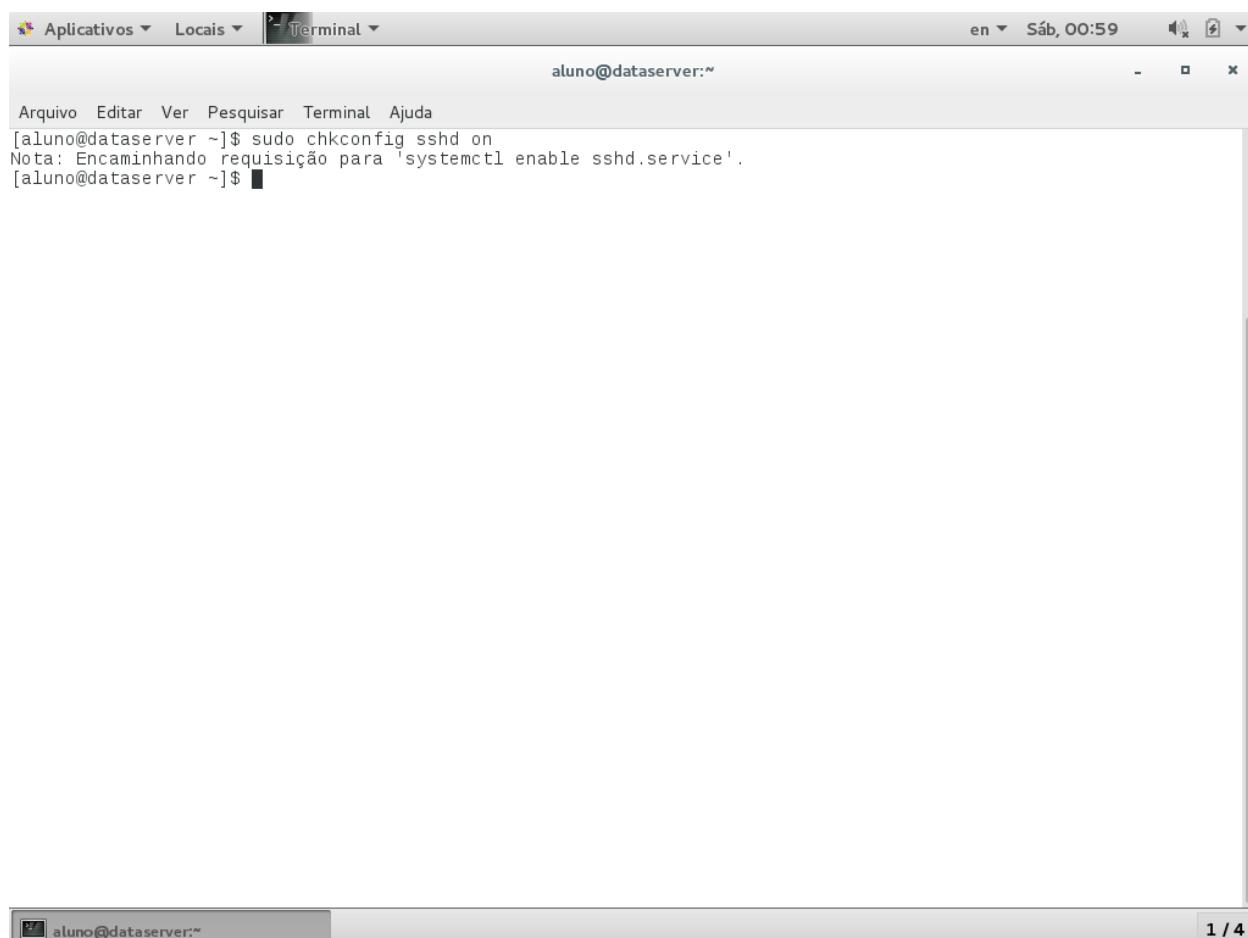
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en" and "Sáb, 00:58". The terminal prompt is "aluno@dataserver:~". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command "sudo chkconfig sshd on" is visible in the terminal window. The bottom status bar shows "aluno@dataserver:~" and "1 / 4".

```
aluno@dataserver:~$ sudo chkconfig sshd on
```

Instalação e Configuração do Ecosistema Hadoop

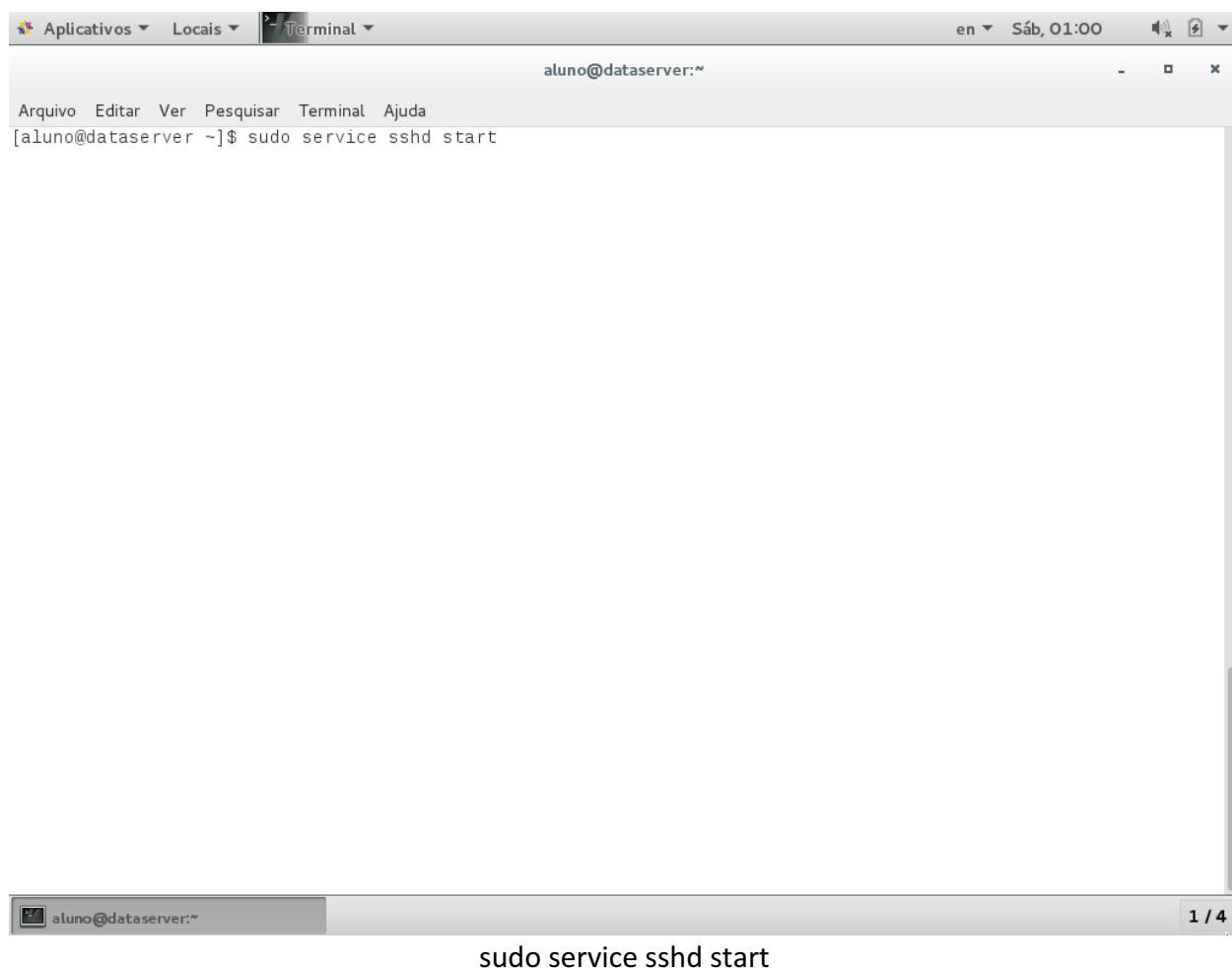


A screenshot of a Linux terminal window titled "Terminal". The window shows the command `sudo chkconfig sshd on` being run, which triggers a note about enabling the service via `systemctl enable sshd.service`. The terminal window has a standard top bar with icons for applications, locations, and system status (language: en, date: Sáb, 00:59). The bottom status bar shows the user "aluno@dataserver:~" and a page number indicator "1 / 4".

```
[aluno@dataserver ~]$ sudo chkconfig sshd on
Nota: Encaminhando requisição para 'systemctl enable sshd.service'.
[aluno@dataserver ~]$
```

Ok

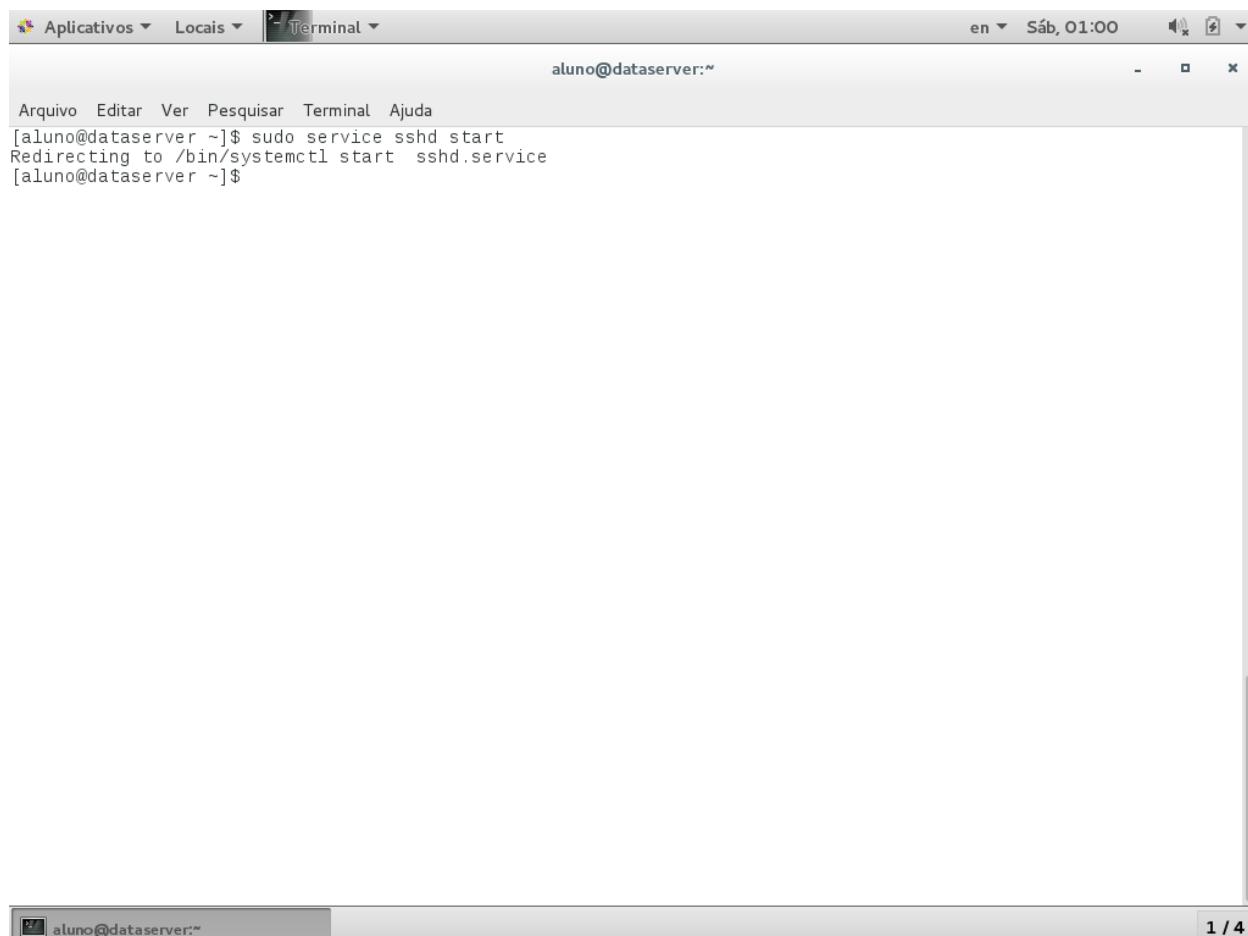
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en" and "Sáb, 01:00". The terminal prompt is "aluno@dataserver:~". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command "sudo service sshd start" is typed into the terminal. The bottom status bar shows the user icon, the terminal prompt again, and "1 / 4".

```
aluno@dataserver:~$ sudo service sshd start
```

Instalação e Configuração do Ecosistema Hadoop

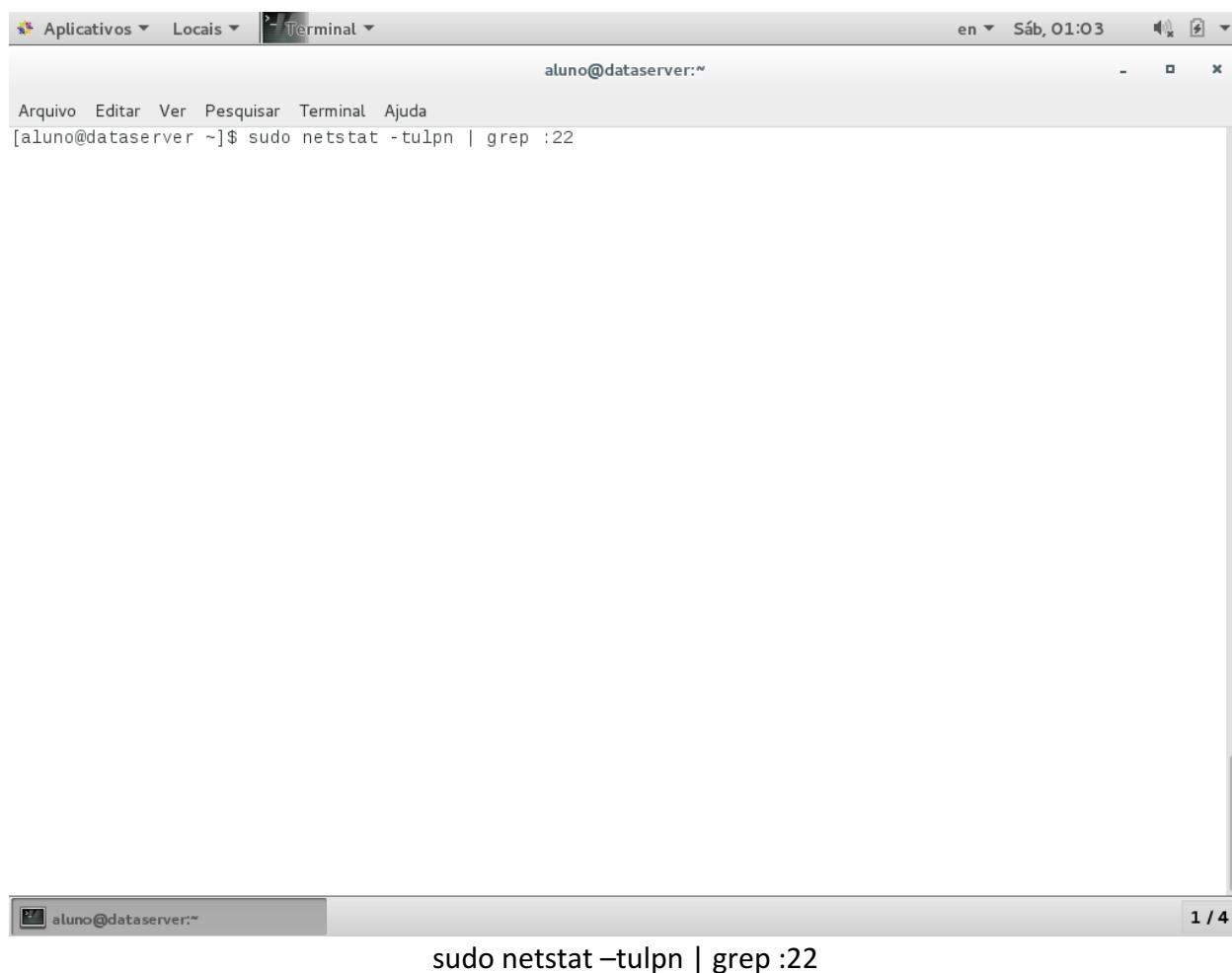


A screenshot of a Linux terminal window titled "Terminal". The window has a menu bar with "Aplicativos", "Locais", and "Terminal". The status bar shows "en" and "Sáb, 01:00". The terminal prompt is "aluno@dataserver:~". The user runs the command "[aluno@dataserver ~]\$ sudo service sshd start" and receives the output "Redirecting to /bin/systemctl start sshd.service". The terminal window has scroll bars on the right side.

```
[aluno@dataserver ~]$ sudo service sshd start
Redirecting to /bin/systemctl start sshd.service
[aluno@dataserver ~]$
```

Ok

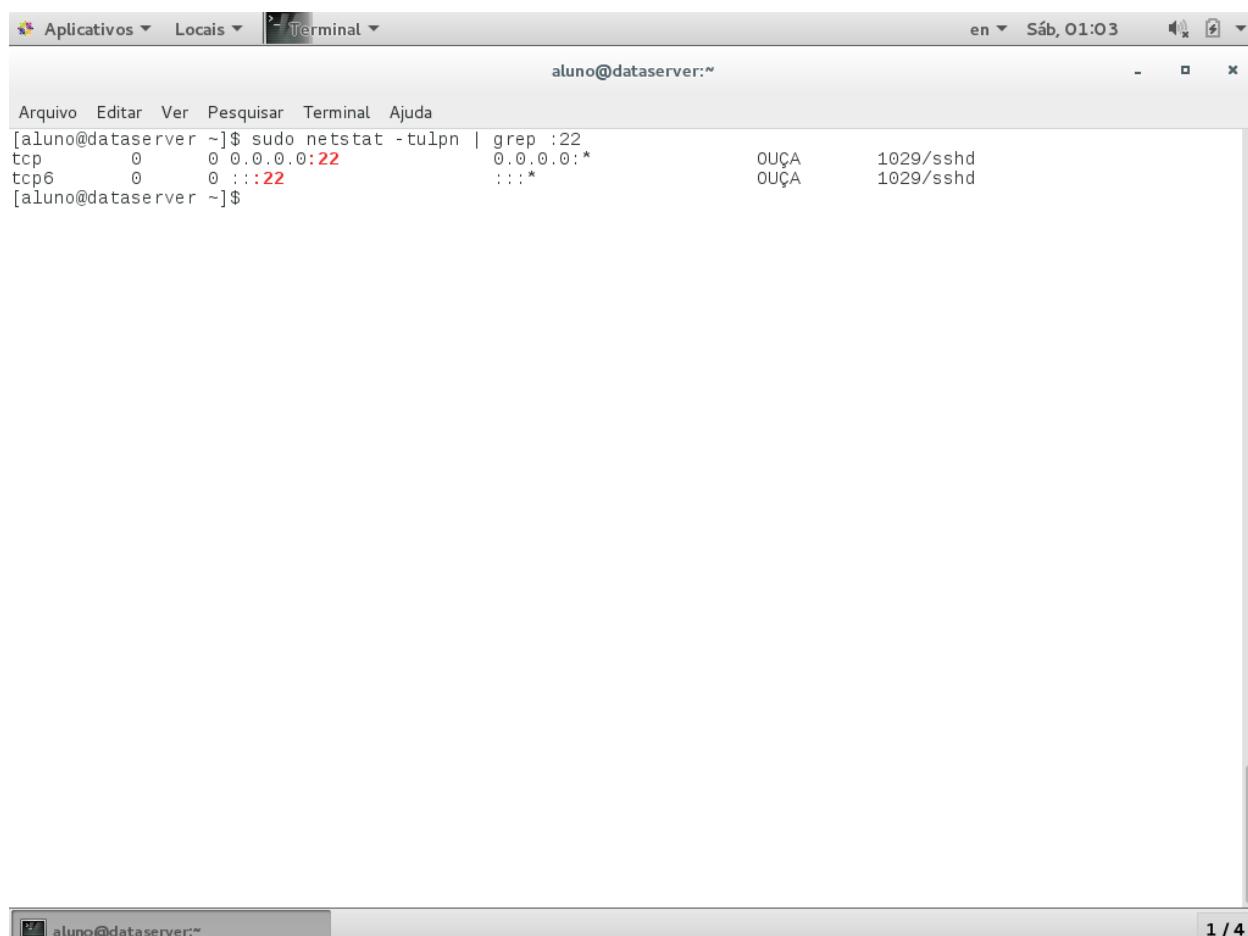
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~". The user has run the command "sudo netstat -tulpn | grep :22". The terminal window is part of a larger interface with a menu bar at the top and a status bar at the bottom indicating "1 / 4".

```
aluno@dataserver:~$ sudo netstat -tulpn | grep :22
```

Instalação e Configuração do Ecosistema Hadoop

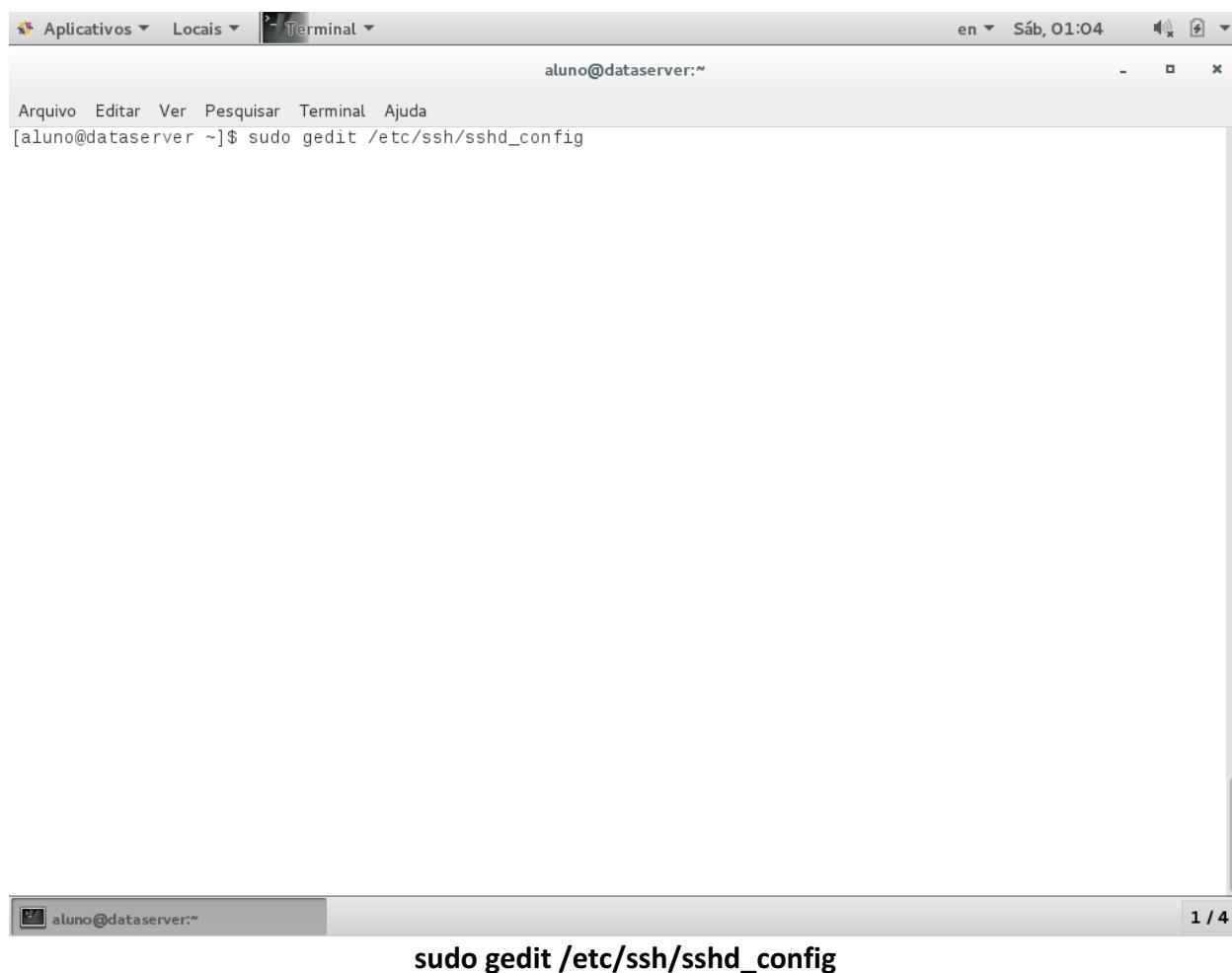


A screenshot of a Linux terminal window titled "Terminal". The window shows the command "sudo netstat -tulpn | grep :22" being run by the user "aluno" at the host "dataserver". The output of the command is displayed, showing two entries for port 22: one for TCP (tcp) and one for TCP6 (tcp6). Both entries show the process "sshd" running with PID 1029. The terminal window has a standard Linux interface with menu bars and system status indicators.

```
[aluno@dataserver ~]$ sudo netstat -tulpn | grep :22
tcp        0      0 0.0.0.0:22              0.0.0.0:*               OUÇA
tcp6       0      0 :::22                 :::*                  OUÇA
[aluno@dataserver ~]$
```

Ok

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "aluno@dataserver:~" and the date/time "Sáb, 01:04". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The terminal window contains the command "[aluno@dataserver ~]\$ sudo gedit /etc/ssh/sshd_config". The bottom status bar shows the user icon "aluno@dataserver:~" and the page number "1 / 4".

sudo gedit /etc/ssh/sshd_config

```

*sshd_config
/etc/ssh

#      $OpenBSD: sshd_config,v 1.93 2014/01/10 05:59:19 djm 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 ::

# The default requires explicit activation of protocol 1
#Protocol 2

# HostKey for protocol version 1
#HostKey /etc/ssh/ssh_host_key
# HostKeys for protocol version 2
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

# Lifetime and size of ephemeral version 1 server key
#KeyRegenerationInterval 1h
#ServerKeyBits 1024

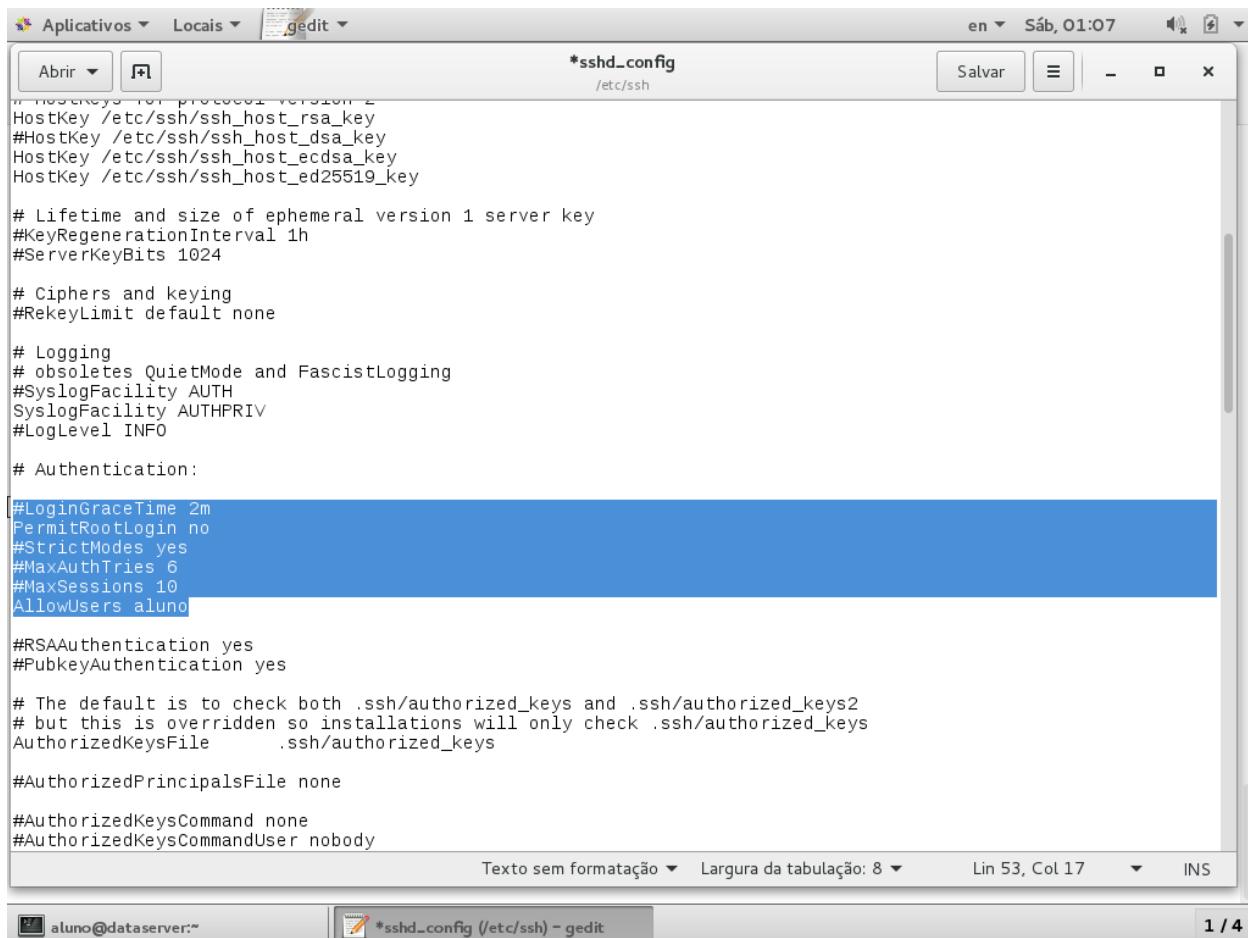
# Ciphers and keying
#RekeyLimit default none

```

Texto sem formatação ▾ Largura da tabulação: 8 ▾ Lin 19, Col 22 ▾ INS

1 / 4

Primeira parte da configuração ssh. Remover o símbolo (#) de comentário das 3 linhas marcadas acima



```
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

# Lifetime and size of ephemeral version 1 server key
#KeyRegenerationInterval 1h
#ServerKeyBits 1024

# Ciphers and keying
#RekeyLimit default none

# Logging
# obsoletes QuietMode and FascistLogging
#SyslogFacility AUTH
SyslogFacility AUTHPRIV
#LogLevel INFO

# Authentication:

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

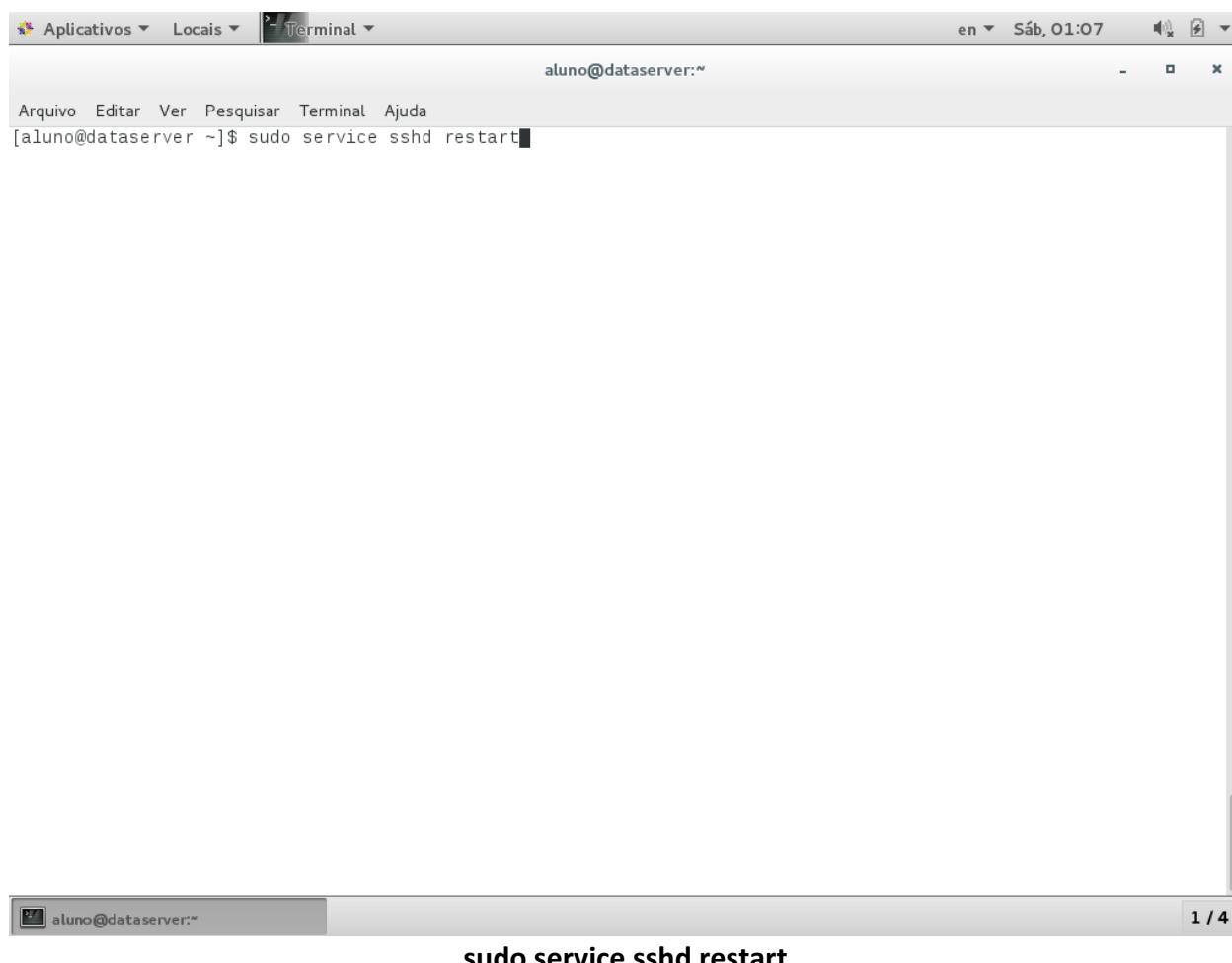
#RSAAuthentication yes
#PubkeyAuthentication yes

# The default is to check both .ssh/authorized_keys and .ssh/authorized_keys2
# but this is overridden so installations will only check .ssh/authorized_keys
AuthorizedKeysFile      .ssh/authorized_keys

#AuthorizedPrincipalsFile none

#AuthorizedKeysCommand none
#AuthorizedKeysCommandUser nobody
```

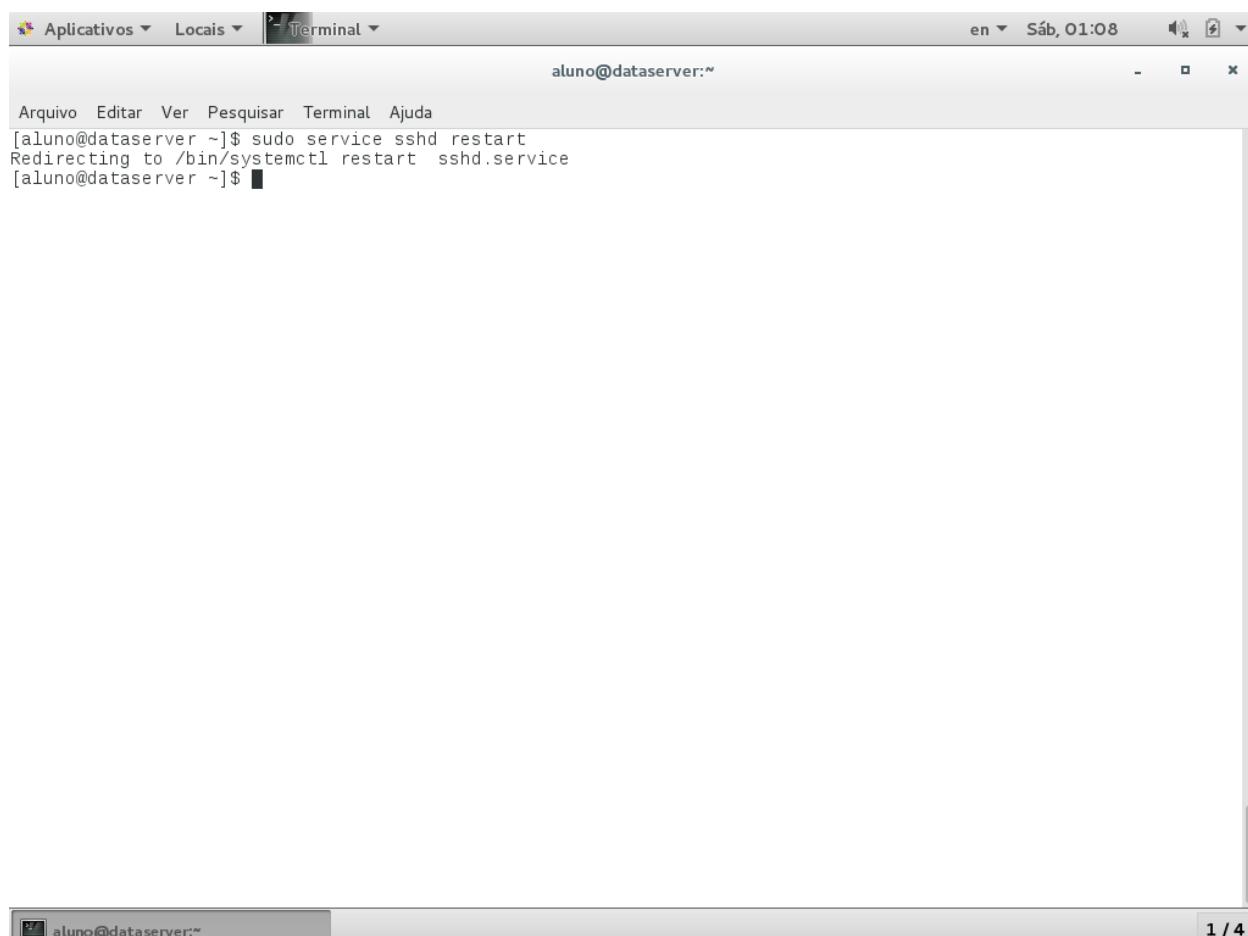
Segunda parte da configuração do ssh



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en" and "Sáb, 01:07". The terminal prompt is "aluno@dataserver:~". Below the prompt, the command "sudo service sshd restart" is visible. The bottom status bar shows the user icon, the terminal title, and "1 / 4".

```
aluno@dataserver:~$ sudo service sshd restart
```

Instalação e Configuração do Ecosistema Hadoop



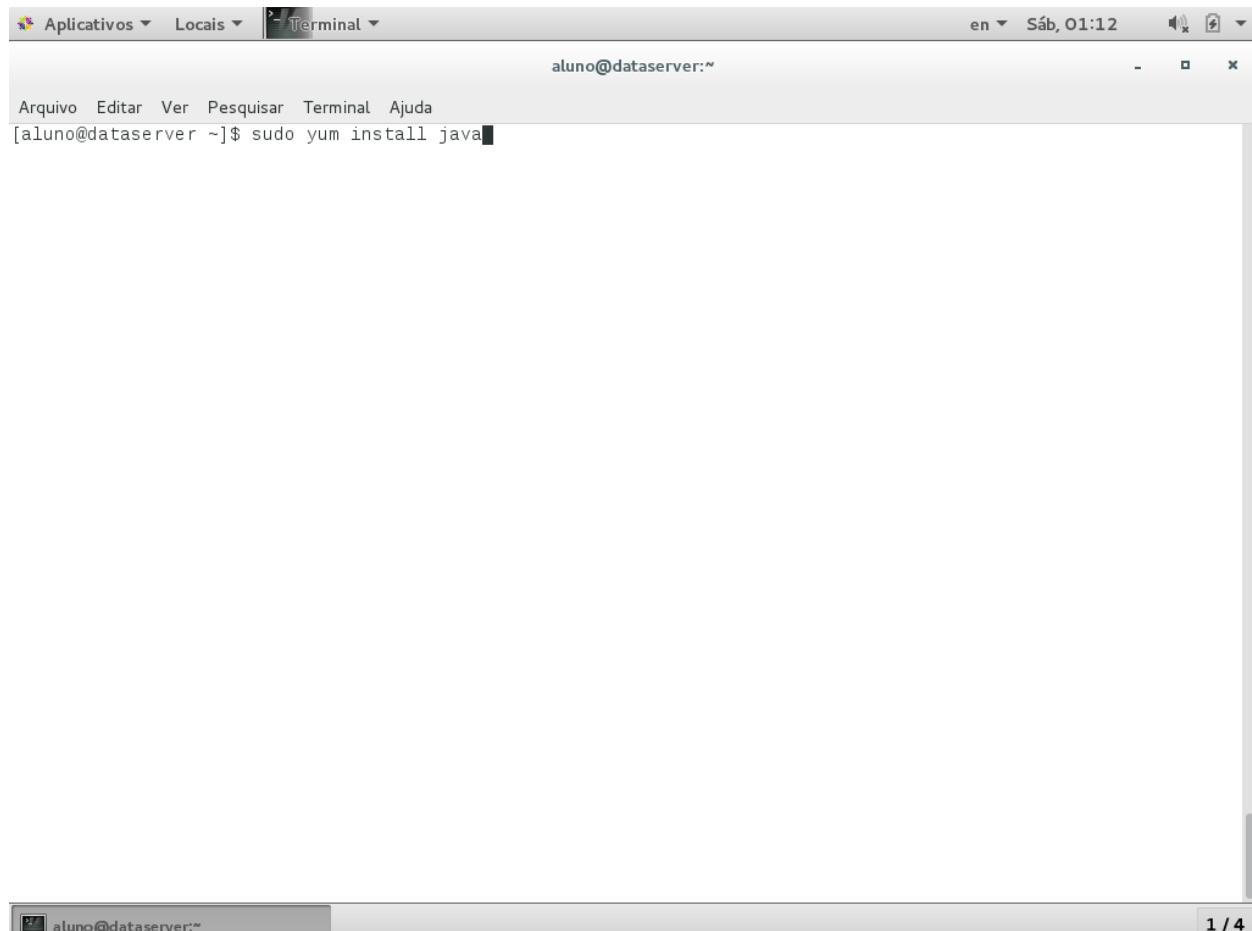
A screenshot of a Linux terminal window titled "Terminal". The window has a menu bar with "Aplicativos", "Locais", and "Terminal". The status bar shows "en" and "Sáb, 01:08". The terminal prompt is "aluno@dataserver:~". The user runs the command "sudo service sshd restart", which outputs "Redirecting to /bin/systemctl restart sshd.service". The terminal has scroll bars on the right side.

```
[aluno@dataserver ~]$ sudo service sshd restart
Redirecting to /bin/systemctl restart sshd.service
[aluno@dataserver ~]$
```

Ok

4. Instalação do Java 8

4.1. Instalação do JRE



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~". Below the prompt, the command "sudo yum install java" is visible in the text area. The desktop interface includes a menu bar with "Aplicativos", "Locais", and "Terminal". The system tray shows "en", "Sáb, 01:12", and other icons. A status bar at the bottom indicates "1 / 4".

Instalação do JRE – Comando **sudo yum install java**



```

Aplicativos Locais Terminal
aluno@dataserver:~ - x

Arquivo Editar Ver Pesquisar Terminal Ajuda

Total 334 kB/s | 33 MB 00:01:41
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Instalando : 1:xorg-x11-font-utils-7.5-20.el7.x86_64 1/10
  Instalando : ttmkdir-3.0.9-42.el7.x86_64 2/10
  Instalando : xorg-x11-fonts-Type1-7.5-9.el7.noarch 3/10
  Instalando : lksctp-tools-1.0.13-3.el7.x86_64 4/10
  Instalando : python-lxml-3.2.1-4.el7.x86_64 5/10
  Instalando : python-javapackages-3.4.1-11.el7.noarch 6/10
  Instalando : javapackages-tools-3.4.1-11.el7.noarch 7/10
  Instalando : tzdata-java-2016c-1.el7.noarch 8/10
  Instalando : 1:java-1.8.0-openjdk-headless-1.8.0.77-0.b03.el7_2.x86_64 9/10
  Instalando : 1:java-1.8.0-openjdk-1.8.0.77-0.b03.el7_2.x86_64 10/10
  Verificando : tzdata-java-2016c-1.el7.noarch 1/10
  Verificando : 1:java-1.8.0-openjdk-headless-1.8.0.77-0.b03.el7_2.x86_64 2/10
  Verificando : xorg-x11-fonts-Type1-7.5-9.el7.noarch 3/10
  Verificando : python-javapackages-3.4.1-11.el7.noarch 4/10
  Verificando : python-lxml-3.2.1-4.el7.x86_64 5/10
  Verificando : lksctp-tools-1.0.13-3.el7.x86_64 6/10
  Verificando : javapackages-tools-3.4.1-11.el7.noarch 7/10
  Verificando : ttmkdir-3.0.9-42.el7.x86_64 8/10
  Verificando : 1:java-1.8.0-openjdk-1.8.0.77-0.b03.el7_2.x86_64 9/10
  Verificando : 1:xorg-x11-font-utils-7.5-20.el7.x86_64 10/10

Instalados:
  java-1.8.0-openjdk.x86_64 1:1.8.0.77-0.b03.el7_2

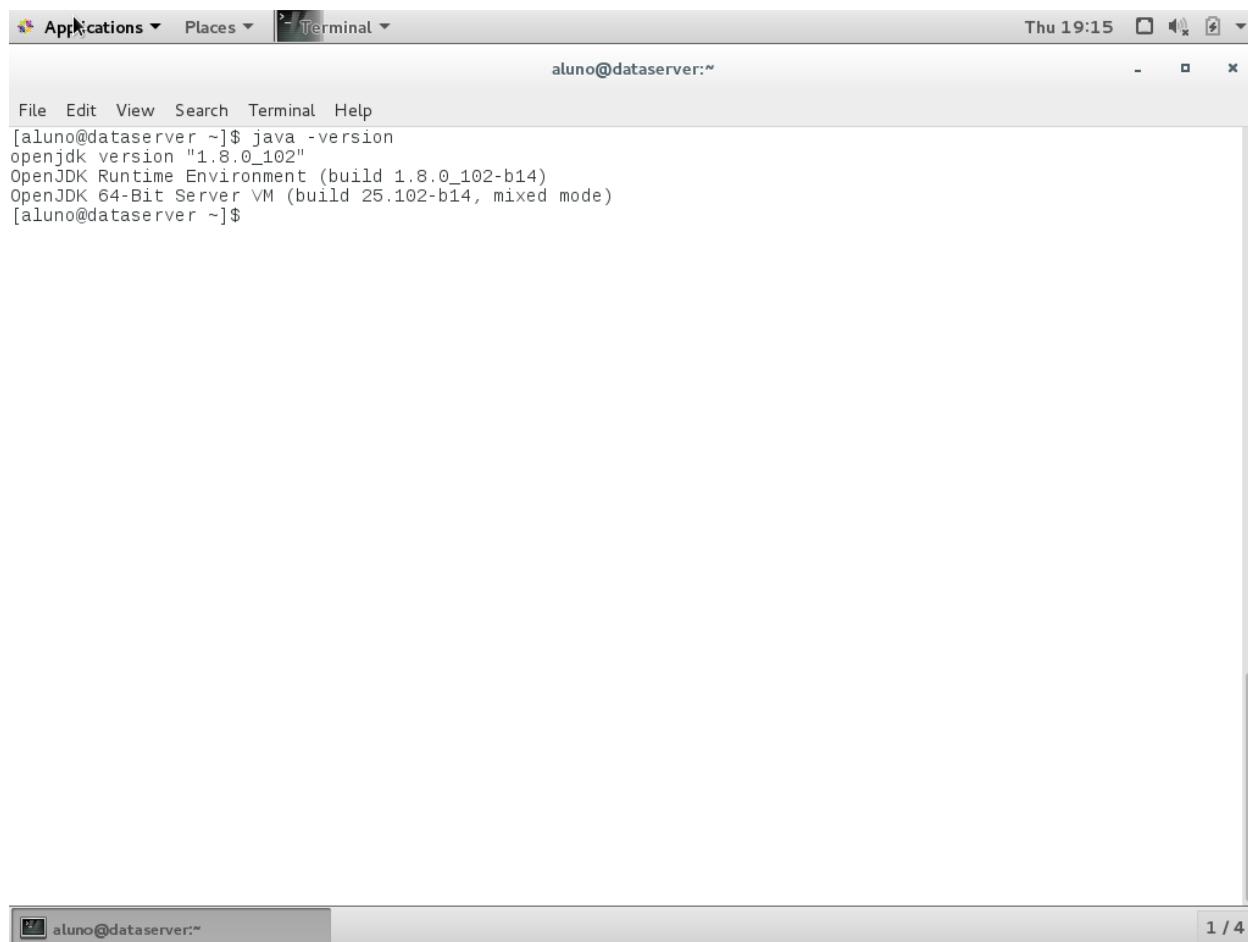
Dependência(s) instalada(s):
  java-1.8.0-openjdk-headless.x86_64 1:1.8.0.77-0.b03.el7_2      javapackages-tools.noarch 0:3.4.1-11.el7
  lksctp-tools.x86_64 0:1.0.13-3.el7                            python-javapackages.noarch 0:3.4.1-11.el7
  python-lxml.x86_64 0:3.2.1-4.el7                            ttmkdir.x86_64 0:3.0.9-42.el7
  tzdata-java.noarch 0:2016c-1.el7                             xorg-x11-font-utils.x86_64 1:7.5-20.el7
  xorg-x11-fonts-Type1.noarch 0:7.5-9.el7

Concluído!
[aluno@dataserver ~]$ ■

```

Concluído

1 / 4



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~". The user has run the command "java -version" and the output is:

```
[aluno@dataserver ~]$ java -version
openjdk version "1.8.0_102"
OpenJDK Runtime Environment (build 1.8.0_102-b14)
OpenJDK 64-Bit Server VM (build 25.102-b14, mixed mode)
```

java -version → Java instalado

4.2. Instalação do JDK

Java SE Development Kit 8u102

You must accept the [Oracle Binary Code License Agreement for Java SE](#) to download this software.

Thank you for accepting the Oracle Binary Code License Agreement for Java SE; you may now download this software.

Product / File Description	File Size	Download
Linux x86	160.35 MB	jdk-8u102-linux-i586.rpm
Linux x86	175.03 MB	jdk-8u102-linux-i586.tar.gz
Linux x64	158.35 MB	jdk-8u102-linux-x64.rpm
Linux x64	173.03 MB	jdk-8u102-linux-x64.tar.gz
Mac OS X	227.35 MB	jdk-8u102-macosx-x64.dmg
Solaris SPARC 64-bit	139.59 MB	jdk-8u102-solaris-sparcv9.tar.Z
Solaris SPARC 64-bit	98.98 MB	jdk-8u102-solaris-sparcv9.tar.gz
Solaris x64	140.02 MB	jdk-8u102-solaris-x64.tar.Z
Solaris x64	96.24 MB	jdk-8u102-solaris-x64.tar.gz
Windows x86	189.2 MB	jdk-8u102-windows-i586.exe
Windows x64	194.68 MB	jdk-8u102-windows-x64.exe

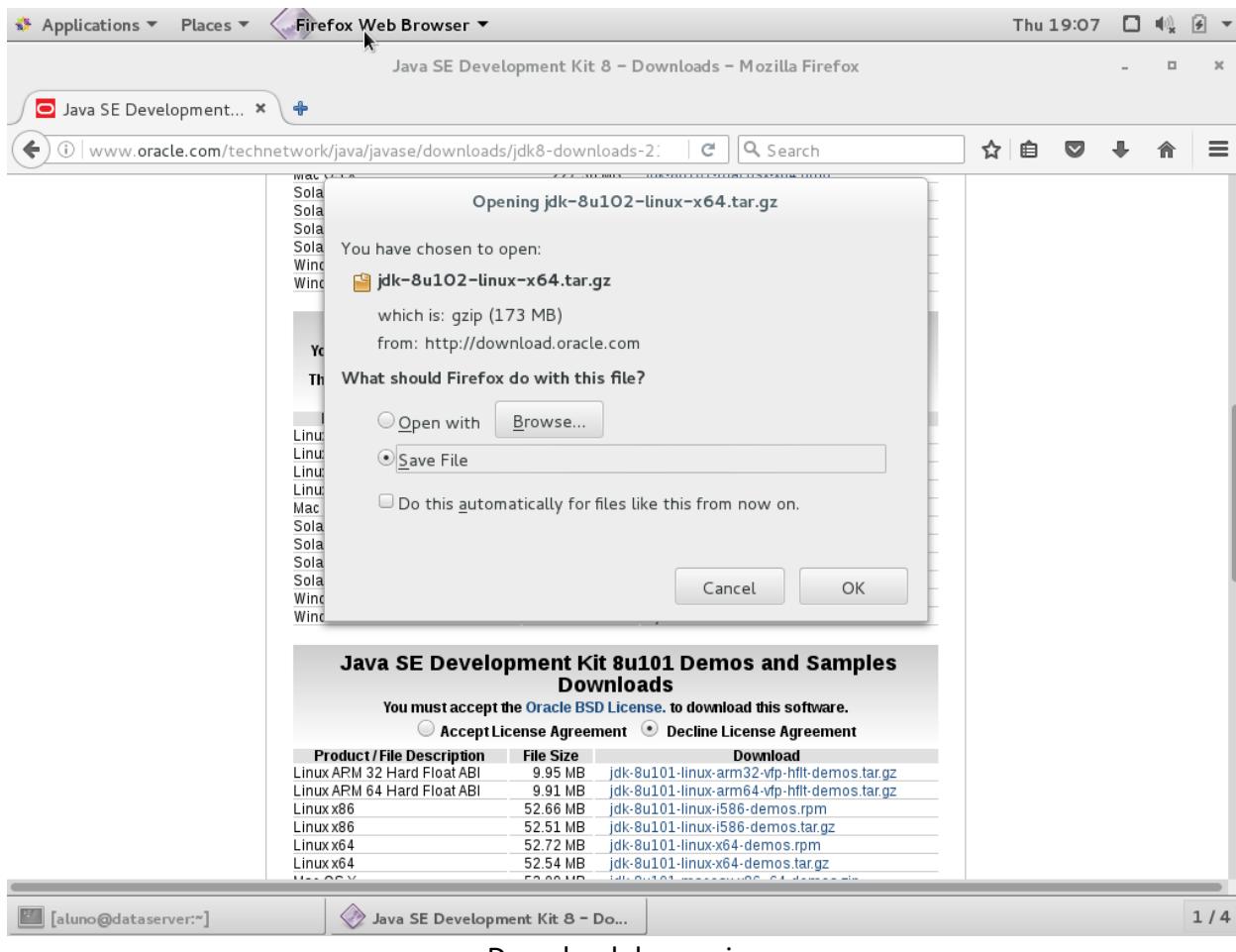
Java SE Development Kit 8u101 Demos and Samples

You must accept the [Oracle BSD License](#) to download this software.

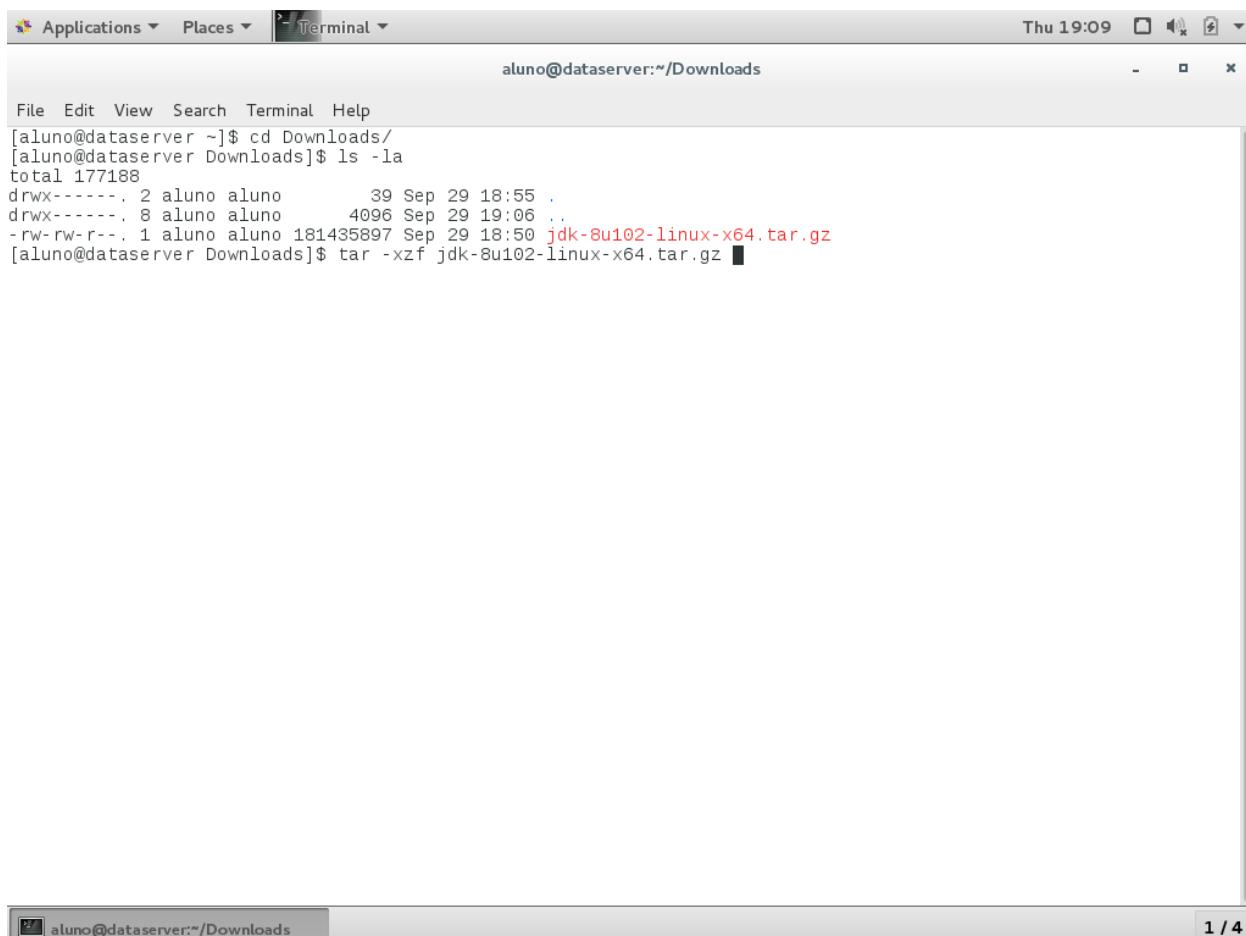
Accept License Agreement Decline License Agreement

Product / File Description	File Size	Download
Linux APM 32 Hard Float ABI	9.95 MB	jdk-8u101-linux-arm32-vfp-hf-f-demos.tar.gz
Linux APM 64 Hard Float ABI	9.91 MB	jdk-8u101-linux-arm64-vfp-hf-f-demos.tar.gz
Linux x86	52.66 MB	jdk-8u101-linux-i586-demos.rpm
Linux x86	52.51 MB	jdk-8u101-linux-i586-demos.tar.gz
Linux x64	52.72 MB	jdk-8u101-linux-x64-demos.rpm
Linux x64	52.54 MB	jdk-8u101-linux-x64-demos.tar.gz
Mac OS X	53.09 MB	jdk-8u101-macosx-x86_64-demos.zip
Solaris SPARC 64-bit	13.55 MB	jdk-8u101-solaris-sparcv9-demos.tar.Z
Solaris SPARC 64-bit	9.34 MB	jdk-8u101-solaris-sparcv9-demos.tar.gz
Solaris x64	13.51 MB	jdk-8u101-solaris-x64-demos.tar.Z
Solaris x64	9.29 MB	jdk-8u101-solaris-x64-demos.tar.gz
Windows x86	53.81 MB	jdk-8u101-windows-i586-demos.zip
Windows x64	53.83 MB	jdk-8u101-windows-x64-demos.zip

No site da Oracle, fazer o download do JDK



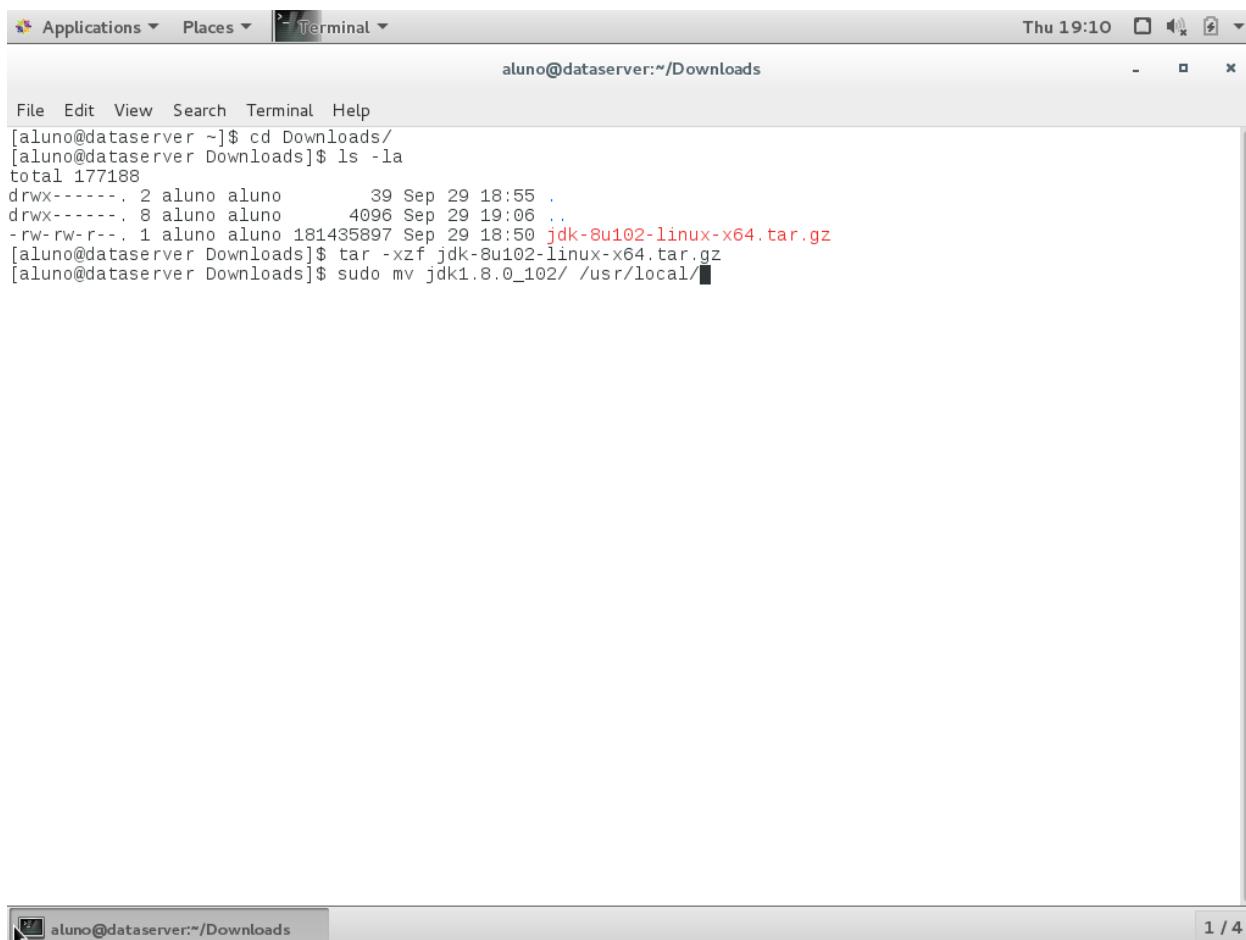
Download do arquivo



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~/Downloads'. The user has run the command 'ls -la' to list files in the Downloads directory. The output shows a file named 'jdk-8u102-linux-x64.tar.gz' which is highlighted in red. The terminal then executes the command 'tar -xzf jdk-8u102-linux-x64.tar.gz'.

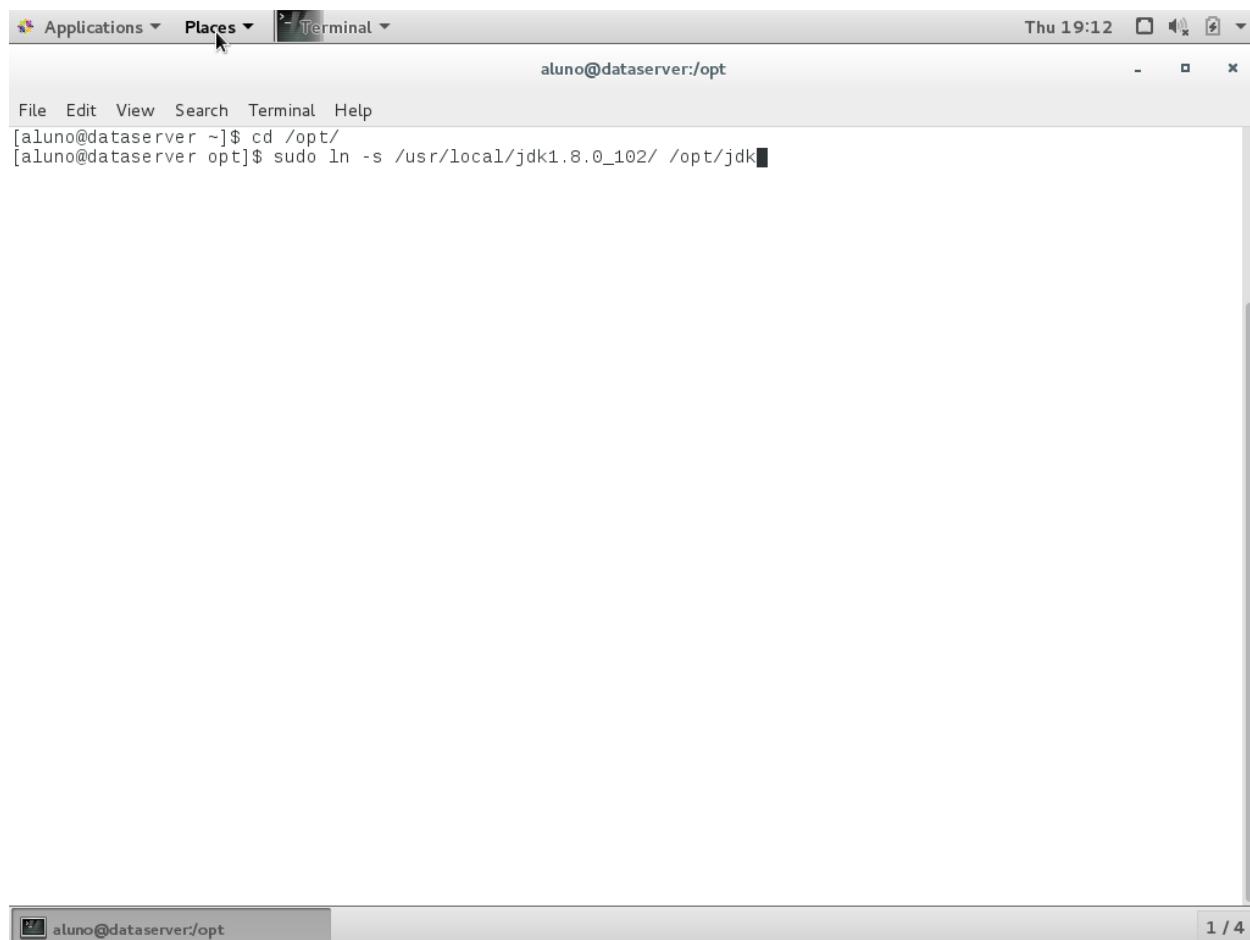
```
[aluno@dataserver ~]$ cd Downloads/  
[aluno@dataserver Downloads]$ ls -la  
total 177188  
drwx----- . 2 aluno aluno 39 Sep 29 18:55 .  
drwx----- . 8 aluno aluno 4096 Sep 29 19:06 ..  
-rw-rw-r-- . 1 aluno aluno 181435897 Sep 29 18:50 jdk-8u102-linux-x64.tar.gz  
[aluno@dataserver Downloads]$ tar -xzf jdk-8u102-linux-x64.tar.gz
```

Executar o comando tar para descompactar o arquivo: **tar -xzf jdk-8u102-linux-x64.tar.gz**



```
File Edit View Search Terminal Help
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 177188
drwx----- 2 aluno aluno 39 Sep 29 18:55 .
drwx----- 8 aluno aluno 4096 Sep 29 19:06 ..
-rw-rw-r-- 1 aluno aluno 181435897 Sep 29 18:50 jdk-8u102-linux-x64.tar.gz
[aluno@dataserver Downloads]$ tar -xzf jdk-8u102-linux-x64.tar.gz
[aluno@dataserver Downloads]$ sudo mv jdk1.8.0_102/ /usr/local/
```

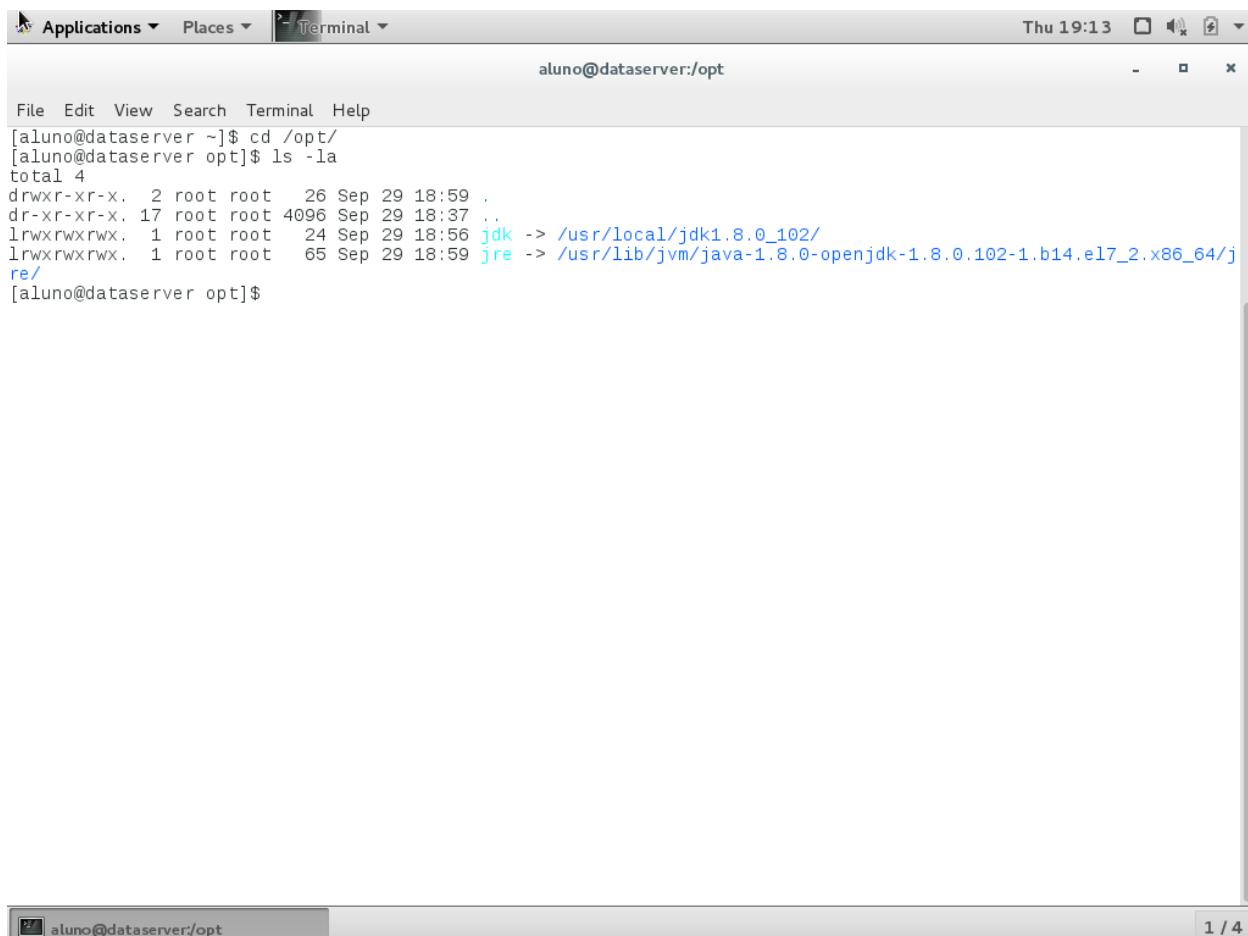
Mover o diretório do JDK



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~/opt'. The user has run the command 'sudo ln -s /usr/local/jdk1.8.0_102/ /opt/jdk' in the terminal.

```
[aluno@dataserver ~]$ cd /opt/  
[aluno@dataserver opt]$ sudo ln -s /usr/local/jdk1.8.0_102/ /opt/jdk
```

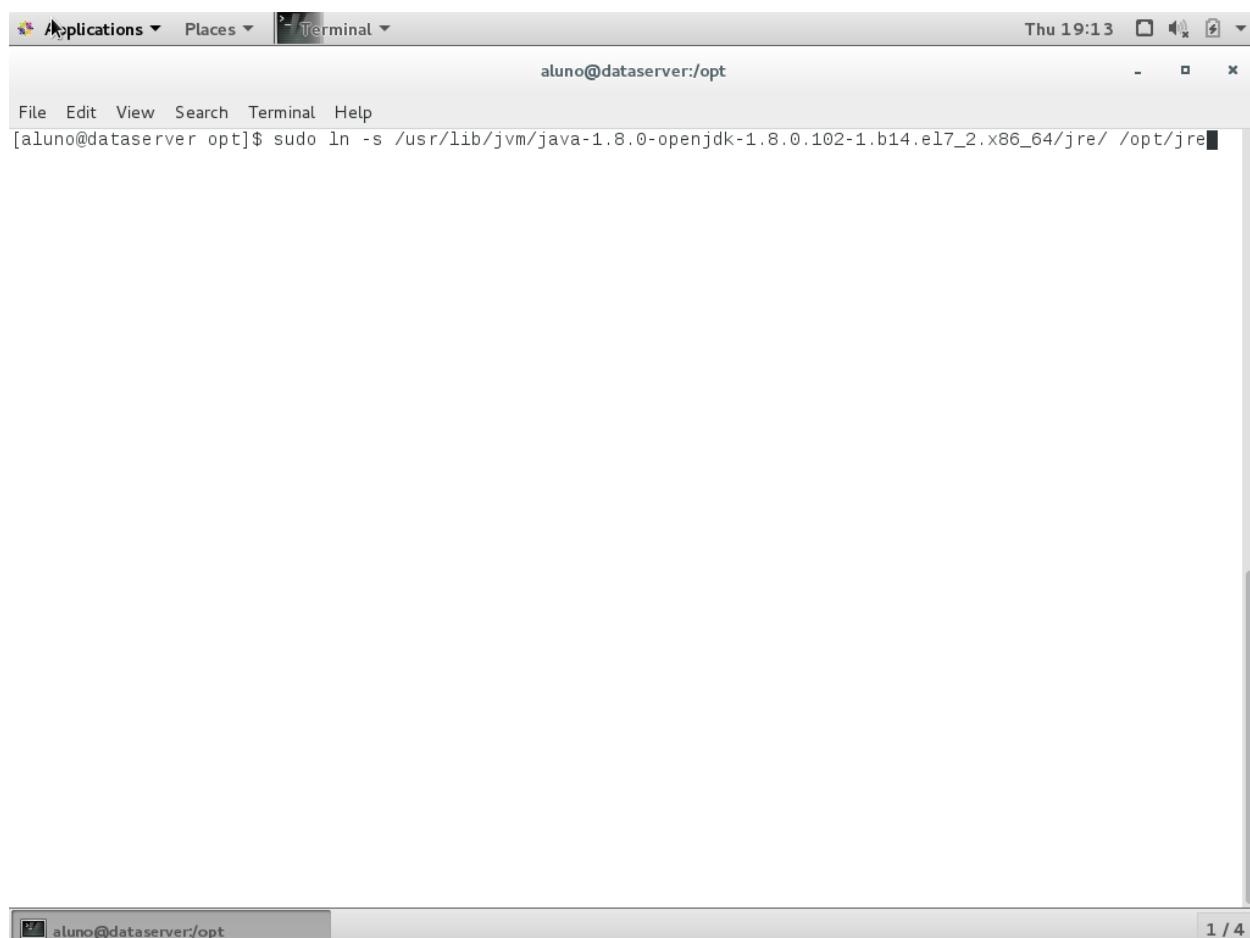
Criaremos links simbólicos para facilitar as configurações posteriores



A screenshot of a Linux terminal window titled "Terminal". The window shows a command-line session with the user "aluno" at "dataserver" in the "/opt" directory. The user runs "ls -la" to list files, which shows a symbolic link named "jdk" pointing to "/usr/local/jdk1.8.0_102/" and another symbolic link named "jre" pointing to "/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/". The terminal interface includes a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The status bar at the bottom indicates "aluno@dataserver: /opt" and "1 / 4".

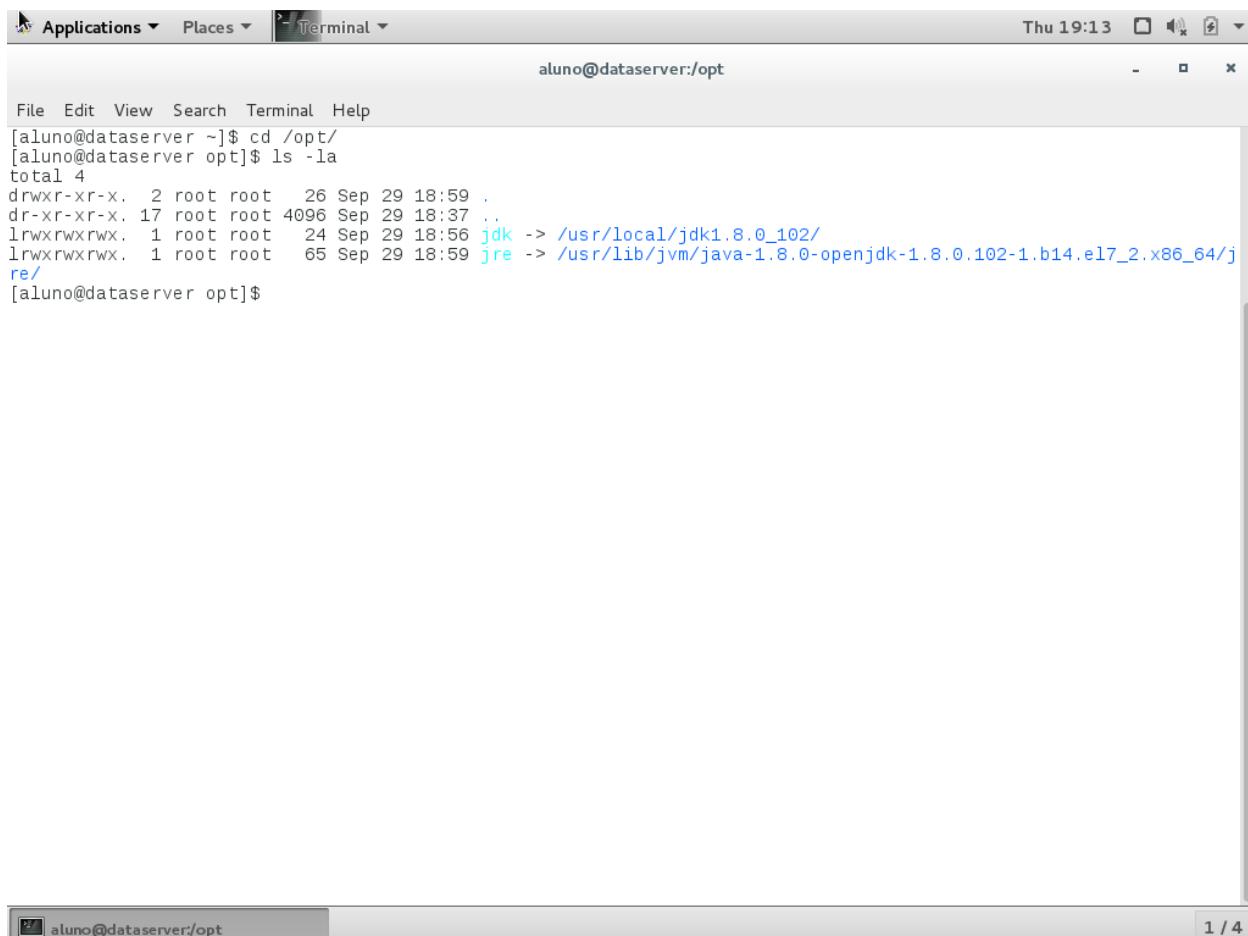
```
[aluno@dataserver ~]$ cd /opt/
[aluno@dataserver opt]$ ls -la
total 4
drwxr-xr-x. 2 root root 26 Sep 29 18:59 .
dr-xr-xr-x. 17 root root 4096 Sep 29 18:37 ..
lrwxrwxrwx. 1 root root 24 Sep 29 18:56 jdk -> /usr/local/jdk1.8.0_102/
lrwxrwxrwx. 1 root root 65 Sep 29 18:59 jre -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/
[aluno@dataserver opt]$
```

Link criado. O JDK agora aponta para /opt/jdk



A screenshot of a terminal window titled "Terminal". The window shows the command: `sudo ln -s /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.el7_2.x86_64/jre/ /opt/jre`. The terminal is running on a Linux system, indicated by the prompt `aluno@dataserver:~$`. The window has a standard title bar with "Applications", "Places", and "Terminal" buttons, and a status bar at the bottom showing "1 / 4".

Vamos criar um link também para o JRE

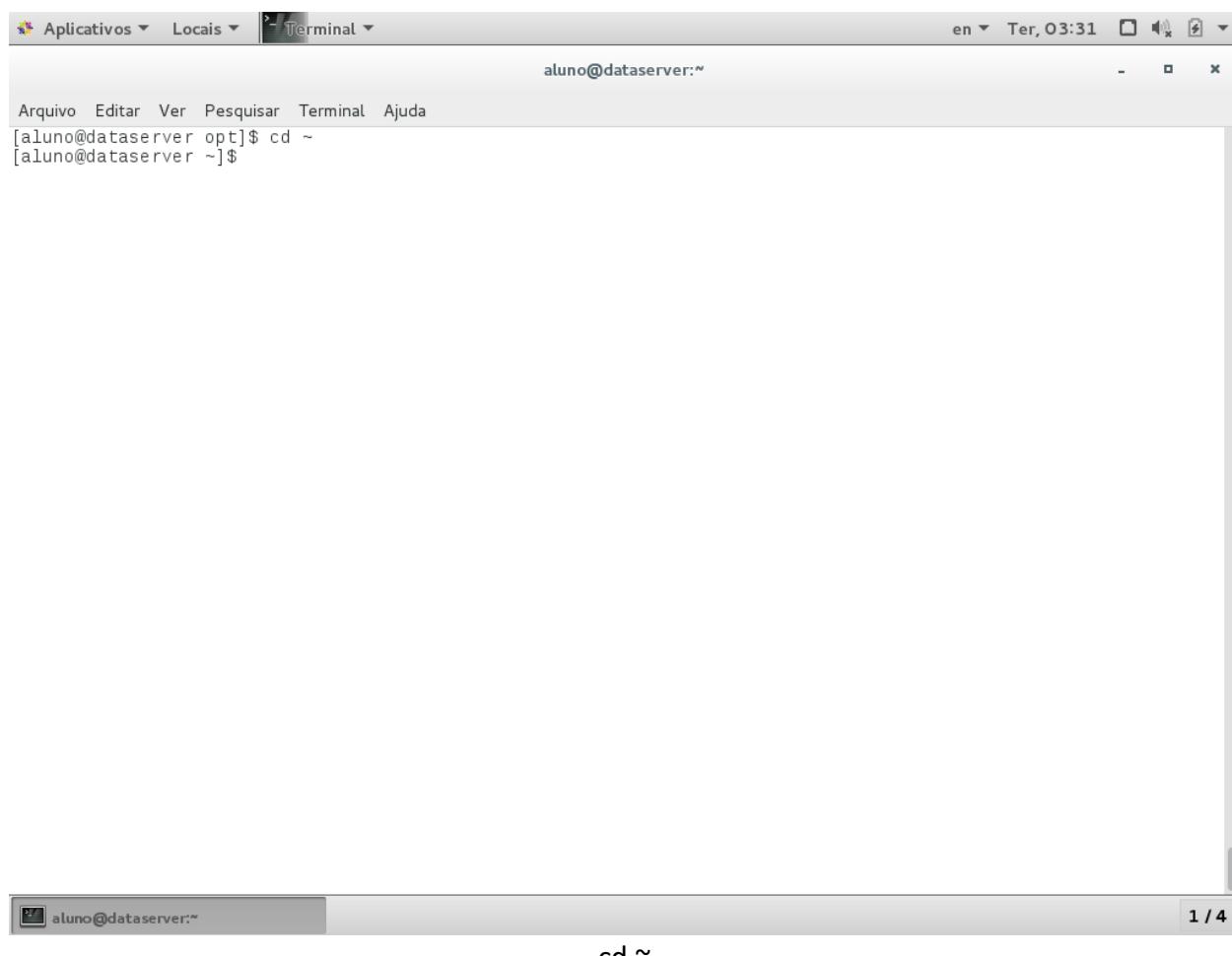


A screenshot of a Linux terminal window titled "Terminal". The window shows the user's session: "aluno@dataserver:~/opt\$". The terminal displays the output of the command "ls -la" in the "/opt" directory. The output shows four files: a dot file (.), a dot-dot file (..), a symbolic link "jdk" pointing to "/usr/local/jdk1.8.0_102/", and a symbolic link "jre" pointing to "/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/". The terminal interface includes a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The status bar at the bottom indicates "1 / 4".

```
[aluno@dataserver ~]$ cd /opt/
[aluno@dataserver opt]$ ls -la
total 4
drwxr-xr-x. 2 root root 26 Sep 29 18:59 .
dr-xr-xr-x. 17 root root 4096 Sep 29 18:37 ..
lrwxrwxrwx. 1 root root 24 Sep 29 18:56 jdk -> /usr/local/jdk1.8.0_102/
lrwxrwxrwx. 1 root root 65 Sep 29 18:59 jre -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/
[aluno@dataserver opt]$
```

O JRE agora aponta para /opt/jre

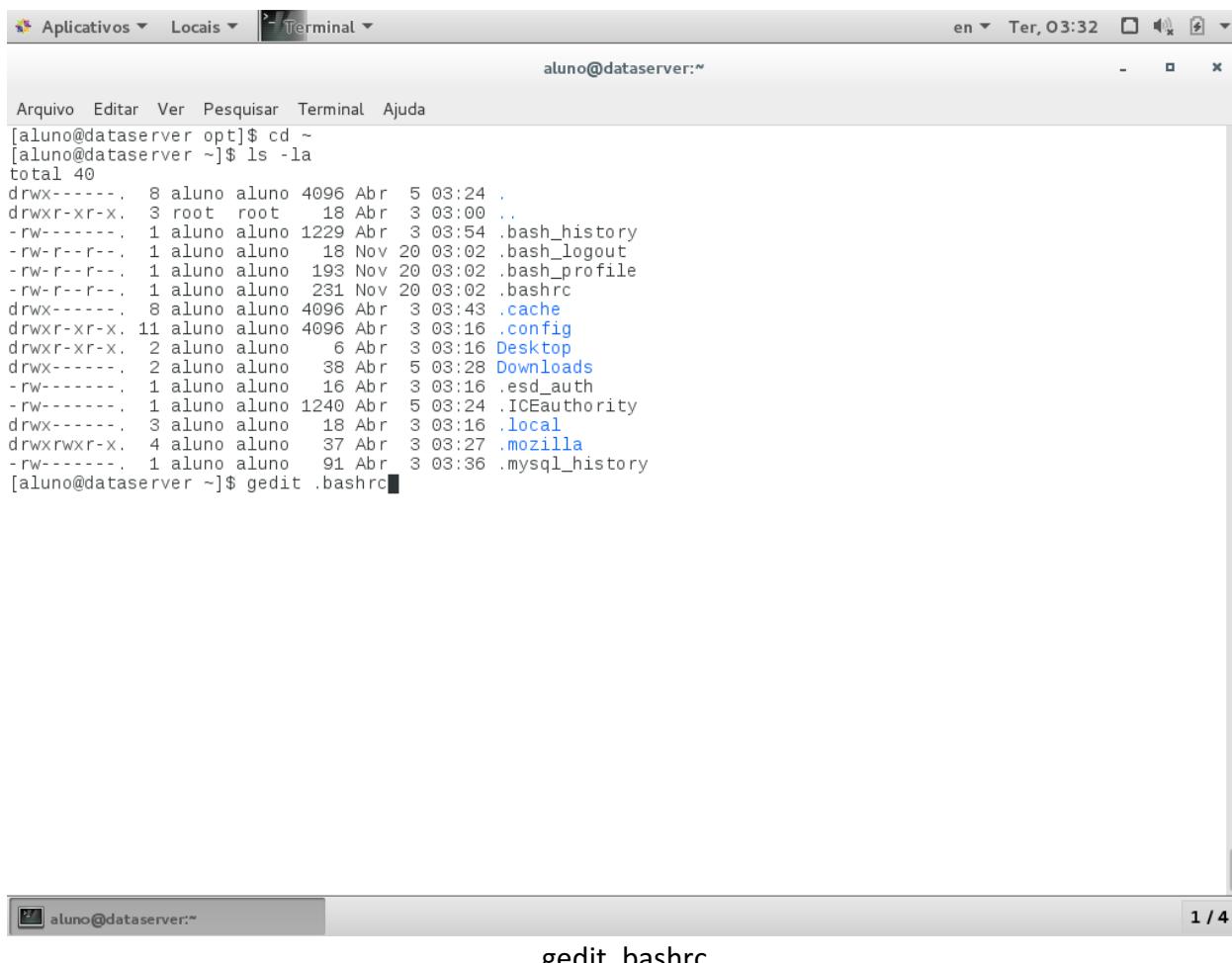
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en Ter, 03:31". The terminal prompt is "aluno@dataserver:~". The menu bar includes "Aplicativos", "Locais", and "Terminal". The main area of the terminal shows the command "[aluno@dataserver opt]\$ cd ~" being typed. The bottom status bar shows "aluno@dataserver:~" and "1 / 4".

```
aluno@dataserver:~\n[aluno@dataserver opt]$ cd ~\n[aluno@dataserver ~]$
```

cd ~

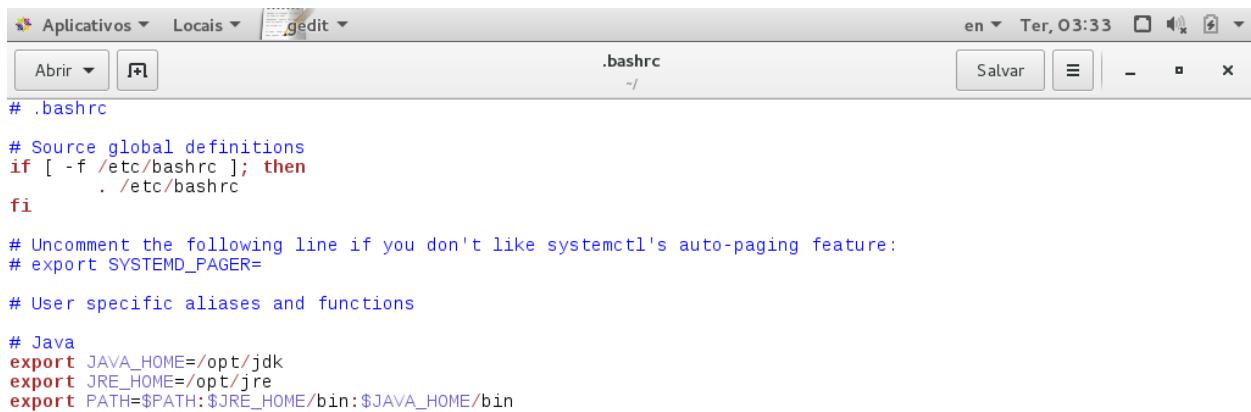


A screenshot of a Linux desktop environment showing a terminal window. The terminal title is "Terminal" and the user is "aluno@dataserver:~". The window contains the following text:

```
[aluno@dataserver opt]$ cd ~  
[aluno@dataserver ~]$ ls -la  
total 40  
drwx----- 8 aluno aluno 4096 Abr  5 03:24 .  
drwxr-xr-x  3 root  root   18 Abr  3 03:00 ..  
-rw-----  1 aluno aluno 1229 Abr  3 03:54 .bash_history  
-rw-r--r--  1 aluno aluno  18 Nov 20 03:02 .bash_logout  
-rw-r--r--  1 aluno aluno 193 Nov 20 03:02 .bash_profile  
-rw-r--r--  1 aluno aluno 231 Nov 20 03:02 .bashrc  
drwx----- 8 aluno aluno 4096 Abr  3 03:43 .cache  
drwxr-xr-x 11 aluno aluno 4096 Abr  3 03:16 .config  
drwxr-xr-x  2 aluno aluno   6 Abr  3 03:16 Desktop  
drwx-----  2 aluno aluno  38 Abr  5 03:28 Downloads  
-rw-----  1 aluno aluno  16 Abr  3 03:16 .esd_auth  
-rw-----  1 aluno aluno 1240 Abr  5 03:24 .ICEauthority  
drwx-----  3 aluno aluno  18 Abr  3 03:16 .local  
drwxrwxr-x  4 aluno aluno  37 Abr  3 03:27 .mozilla  
-rw-----  1 aluno aluno  91 Abr  3 03:36 .mysql_history  
[aluno@dataserver ~]$ gedit .bashrc
```

The terminal window has a toolbar at the top with "Aplicativos", "Locais", and "Terminal". The status bar shows "en Ter, 03:32". The bottom of the window shows the command "gedit .bashrc".

Instalação e Configuração do Ecosistema Hadoop



```
# .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=

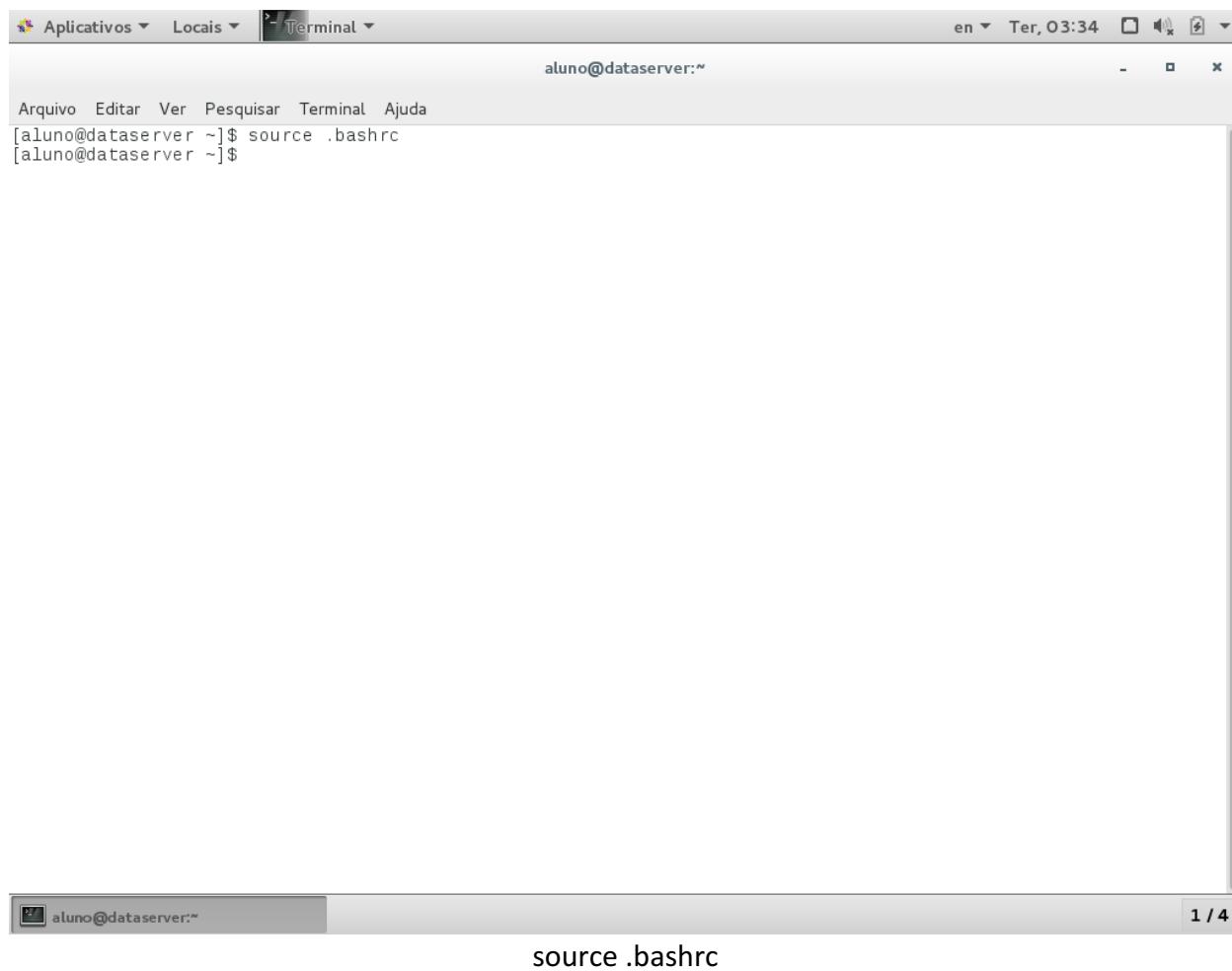
# User specific aliases and functions

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



```
sh ▾ Largura da tabulação: 8 ▾ Lin 16, Col 47 ▾ INS
aluno@dataserver:~ | .bashrc (~/) - gedit
1 / 4
```

Editar as variáveis de ambiente conforme acima e salvar o arquivo



A screenshot of a Linux terminal window titled "Terminal". The window shows the user's home directory (~) and the command "source .bashrc" being run. The terminal interface includes a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom indicates "aluno@dataserver:~". The window has standard OS X-style controls (minimize, maximize, close) in the top right corner. A vertical scroll bar is visible on the right side of the terminal window.

```
aluno@dataserver:~$ source .bashrc
```

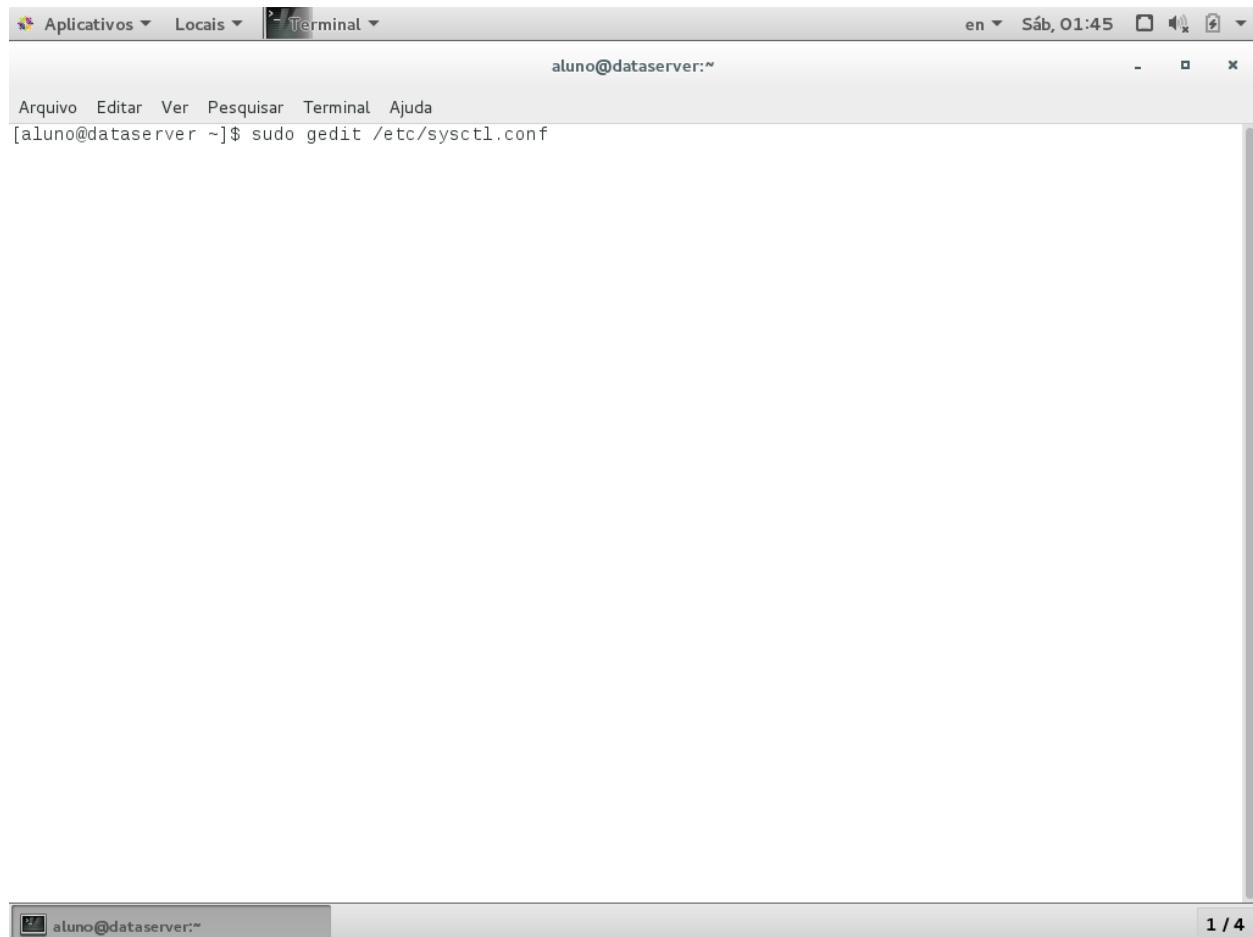
Segundo checkpoint:

Clique no meu File – Export Appliance.
Será gerada uma cópia de segurança da sua máquina virtual.

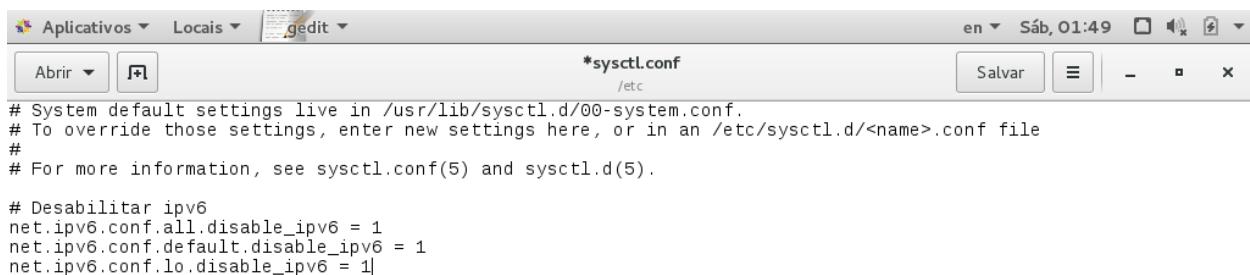
→ VM: DataServer-v2.0.ova (SO e Utilitários)

5. Instalação e Configuração do Hadoop

5.1. Desabilitando ipv6

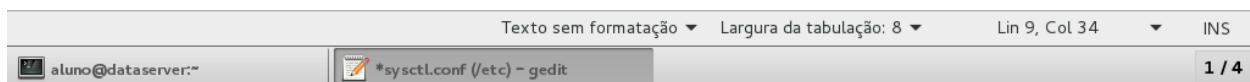


Editar o arquivo /etc/sysctl.conf para desabilitar o ipv6



```
# System default settings live in /usr/lib/sysctl.d/00-system.conf.
# To override those settings, enter new settings here, or in an /etc/sysctl.d/<name>.conf file
#
# For more information, see sysctl.conf(5) and sysctl.d(5).

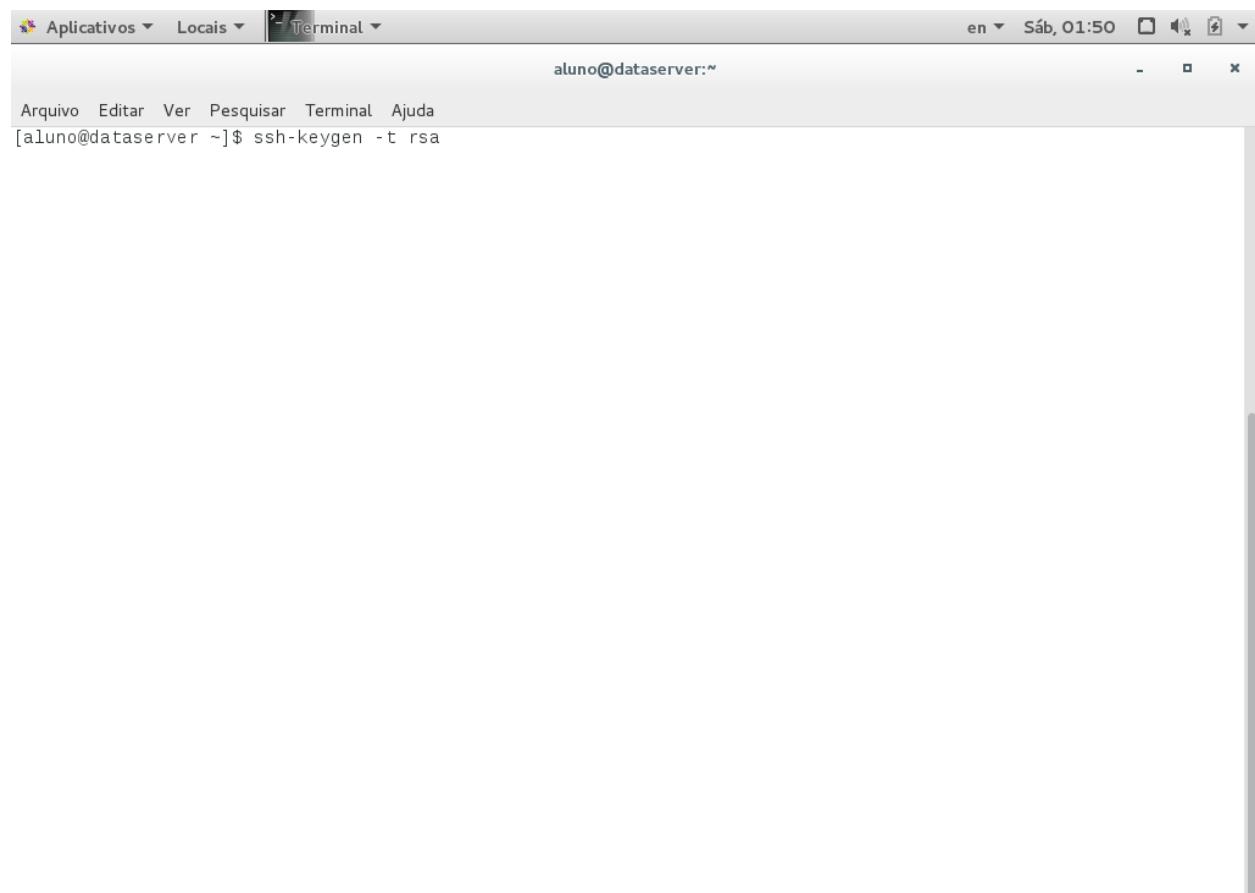
# Desabilitar ipv6
net.ipv6.conf.all.disable_ipv6 = 1
net.ipv6.conf.default.disable_ipv6 = 1
net.ipv6.conf.lo.disable_ipv6 = 1
```



```
aluno@dataserver:~ | *sysctl.conf (/etc) - gedit
```

Incluir as linhas acima e salvar o arquivo

5.2. Configuração do ssh



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en" and "Sáb, 01:50". The terminal prompt is "aluno@dataserver:~". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command "ssh-keygen -t rsa" is typed into the terminal. The bottom status bar shows "aluno@dataserver:~" and "1 / 4".

```
aluno@dataserver:~$ ssh-keygen -t rsa
```

ssh-keygen -t rsa



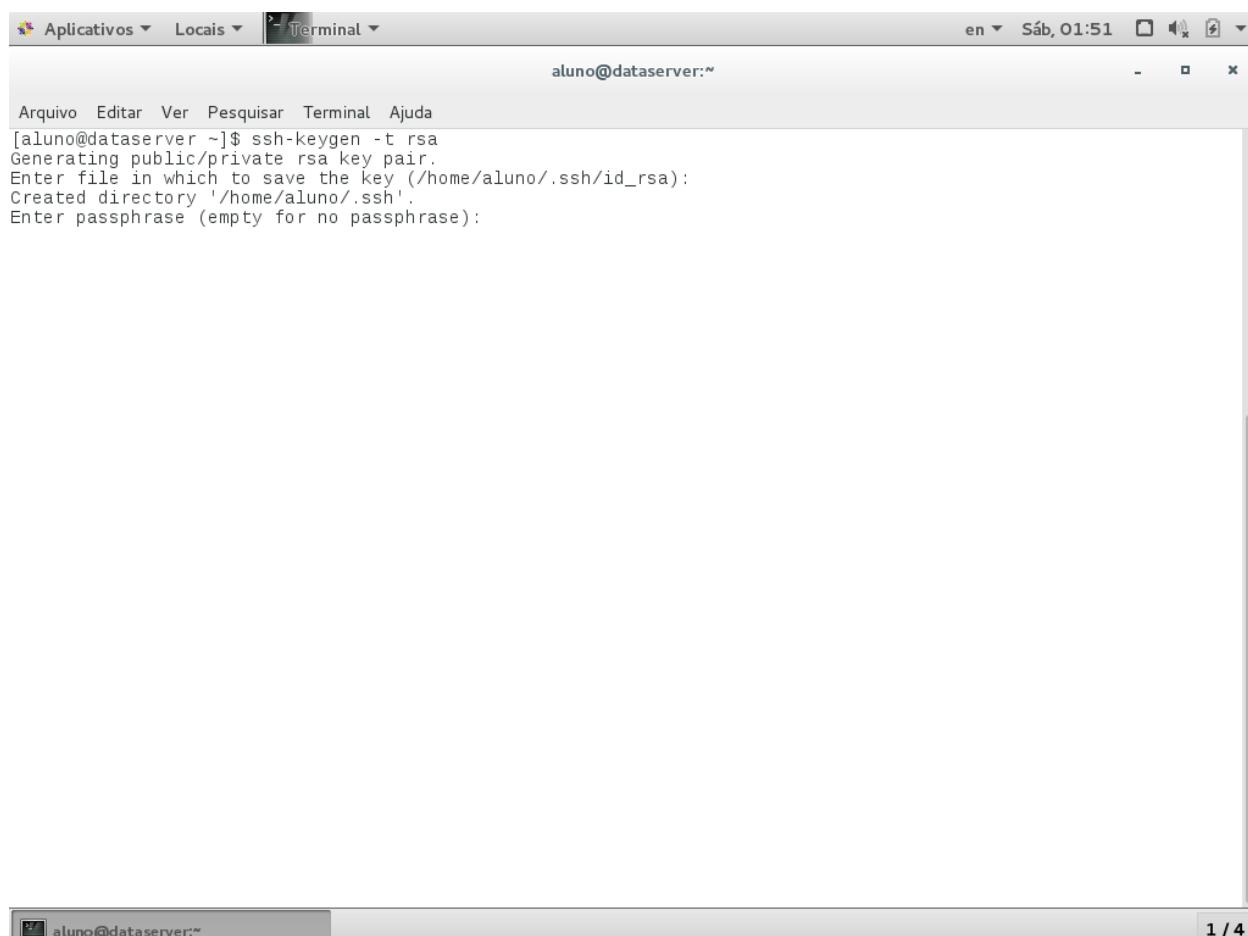
A screenshot of a Linux terminal window titled "Terminal". The window shows the command `ssh-keygen -t rsa` being run by the user "aluno" on a host named "dataserver". The terminal displays the output of the command, which includes the generation of a public/private key pair and a prompt to enter a file name to save the key. The window has a standard title bar with icons for application, location, and terminal, and a status bar at the bottom indicating the user is "aluno@dataserver:~".

```
[aluno@dataserver ~]$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/aluno/.ssh/id_rsa):
```

Pressionar Enter

1 / 4

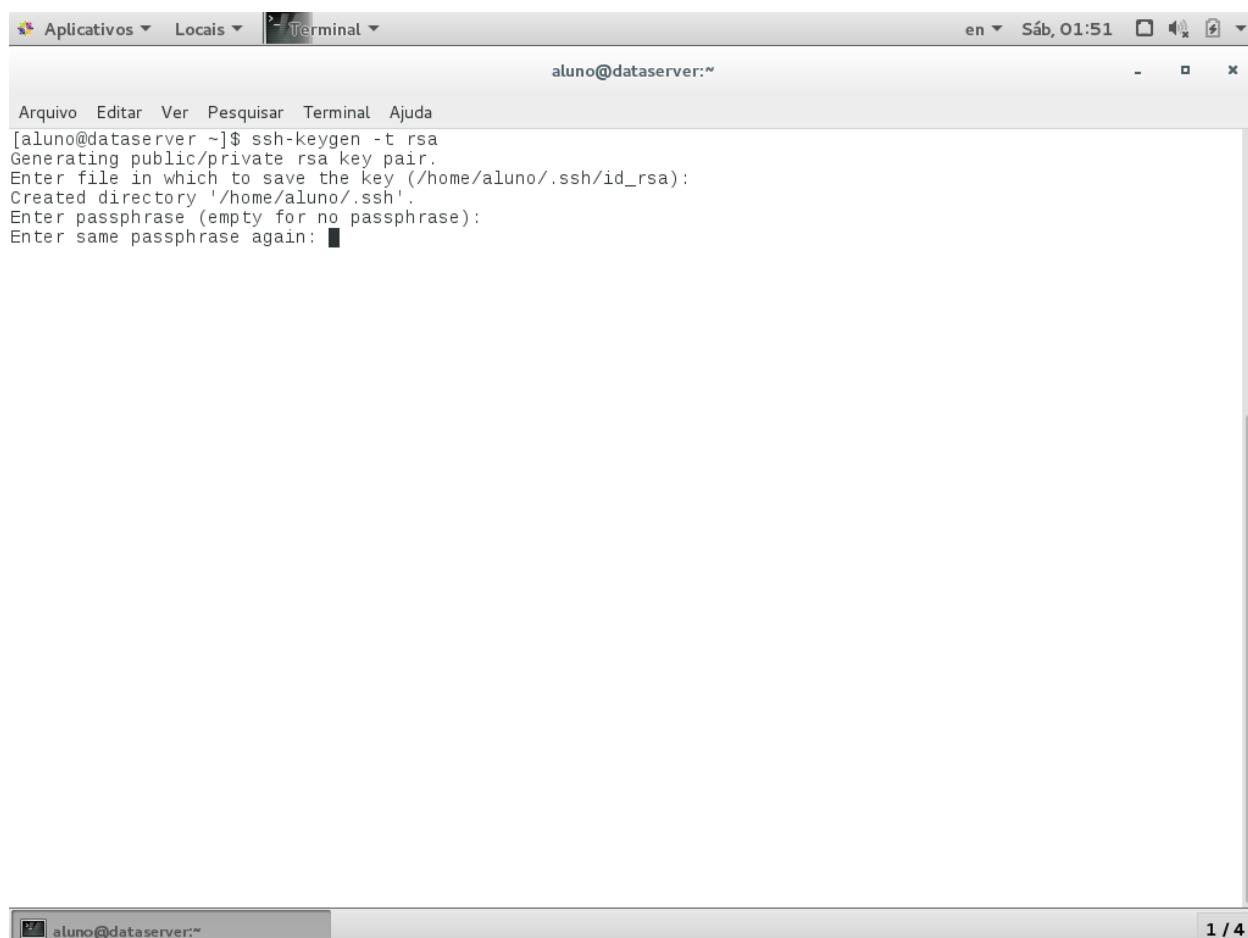
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux terminal window titled "Terminal". The window shows the command `ssh-keygen -t rsa` being run by the user `aluno`. The terminal output indicates the generation of an RSA key pair, prompting the user to enter a file name for the key and a passphrase. The terminal interface includes a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows the user's name and the current date and time.

```
[aluno@dataserver ~]$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/aluno/.ssh/id_rsa):
Created directory '/home/aluno/.ssh'.
Enter passphrase (empty for no passphrase):
```

Pressionar Enter



A screenshot of a Linux terminal window titled "Terminal". The window shows the command `ssh-keygen -t rsa` being run by the user "aluno" at the host "dataserver". The terminal displays the process of generating an RSA key pair, prompting the user to enter a file name for the key and a passphrase. The terminal interface includes a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom indicates the user is at the prompt `aluno@dataserver:~`. The window has standard Linux window controls (minimize, maximize, close) and a vertical scroll bar.

```
[aluno@dataserver ~]$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/aluno/.ssh/id_rsa):
Created directory '/home/aluno/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again: ■
```

Pressionar Enter

Chave de segurança gerada

The screenshot shows a terminal window titled "Terminal" with the command-line interface. The user is generating an RSA key pair using the ssh-keygen command. The terminal output includes the key fingerprint and the public key content. The user then appends the public key to the authorized_keys file.

```
[aluno@dataserver ~]$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/aluno/.ssh/id_rsa):
Created directory '/home/aluno/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/aluno/.ssh/id_rsa.
Your public key has been saved in /home/aluno/.ssh/id_rsa.pub.
The key fingerprint is:
b6:f0:42:fd:d2:b6:7e:10:af:b3:6a:5d:a9:b9:12:98 aluno@dataserver
The key's randomart image is:
+--[ RSA 2048]----+
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
+-----+
[aluno@dataserver ~]$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

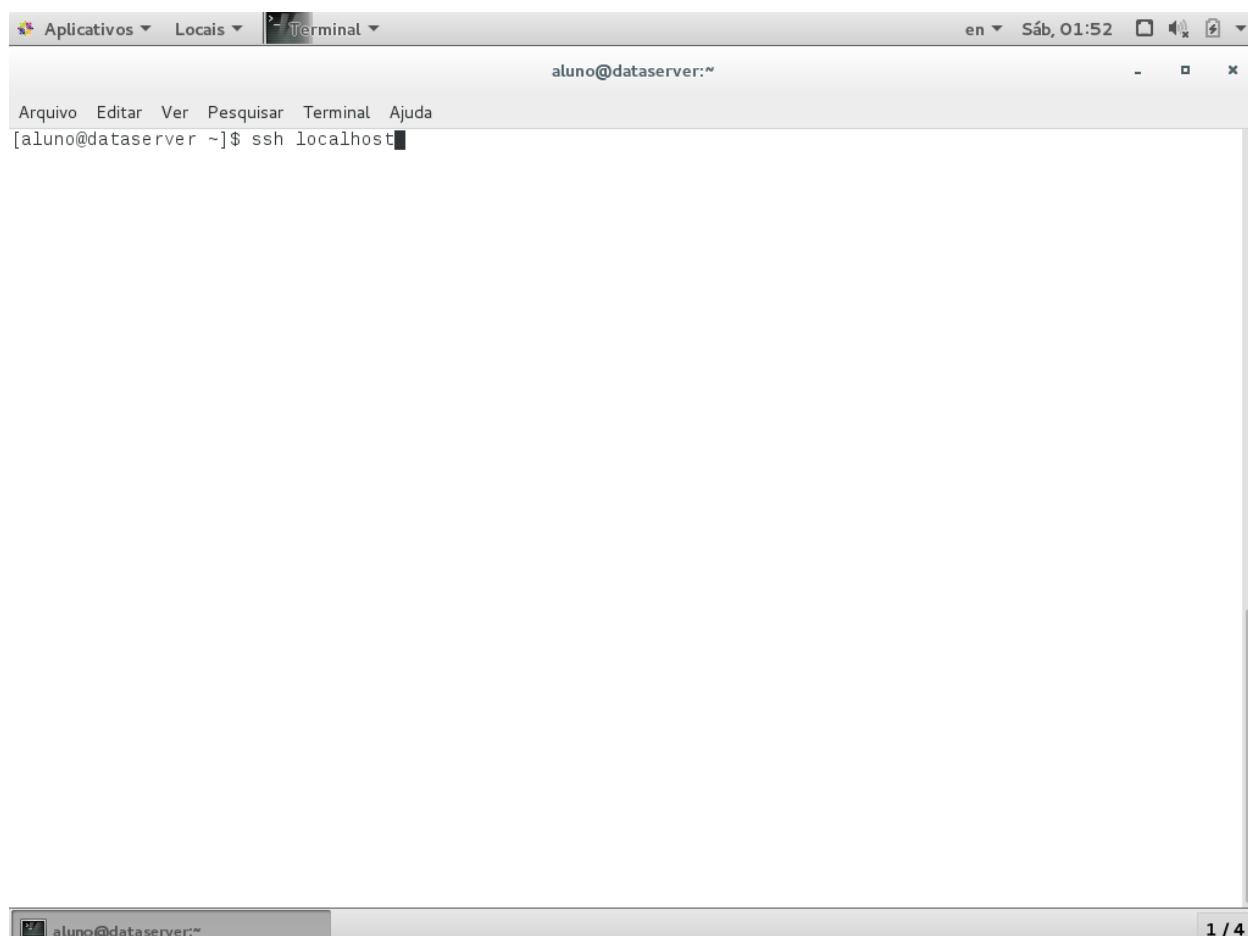
A screenshot of a Linux terminal window titled "Terminal". The window shows the command-line interface with the user "aluno" logged in at "dataserver". The terminal displays the following sequence of commands and their output:

```
[aluno@dataserver ~]$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/aluno/.ssh/id_rsa):
Created directory '/home/aluno/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/aluno/.ssh/id_rsa.
Your public key has been saved in /home/aluno/.ssh/id_rsa.pub.
The key fingerprint is:
b6:f0:42:fd:d2:b6:7e:10:af:b3:6a:5d:a9:b9:12:98 aluno@dataserver
The key's randomart image is:
+--[ RSA 2048]----+
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
|          .          |
+-----+
[aluno@dataserver ~]$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
[aluno@dataserver ~]$ chmod 0600 ~/.ssh/authorized_keys
[aluno@dataserver ~]$
```

The terminal window also includes a status bar at the bottom with the text "aluno@dataserver:~" and a page number indicator "1 / 4".

chmod 0600 ~/ssh/authorized_keys

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en" and "Sáb, 01:52". The terminal prompt is "aluno@dataserver:~". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command "ssh localhost" is typed into the terminal. The bottom status bar shows "aluno@dataserver:~" and "1 / 4".

```
aluno@dataserver:~$ ssh localhost
```

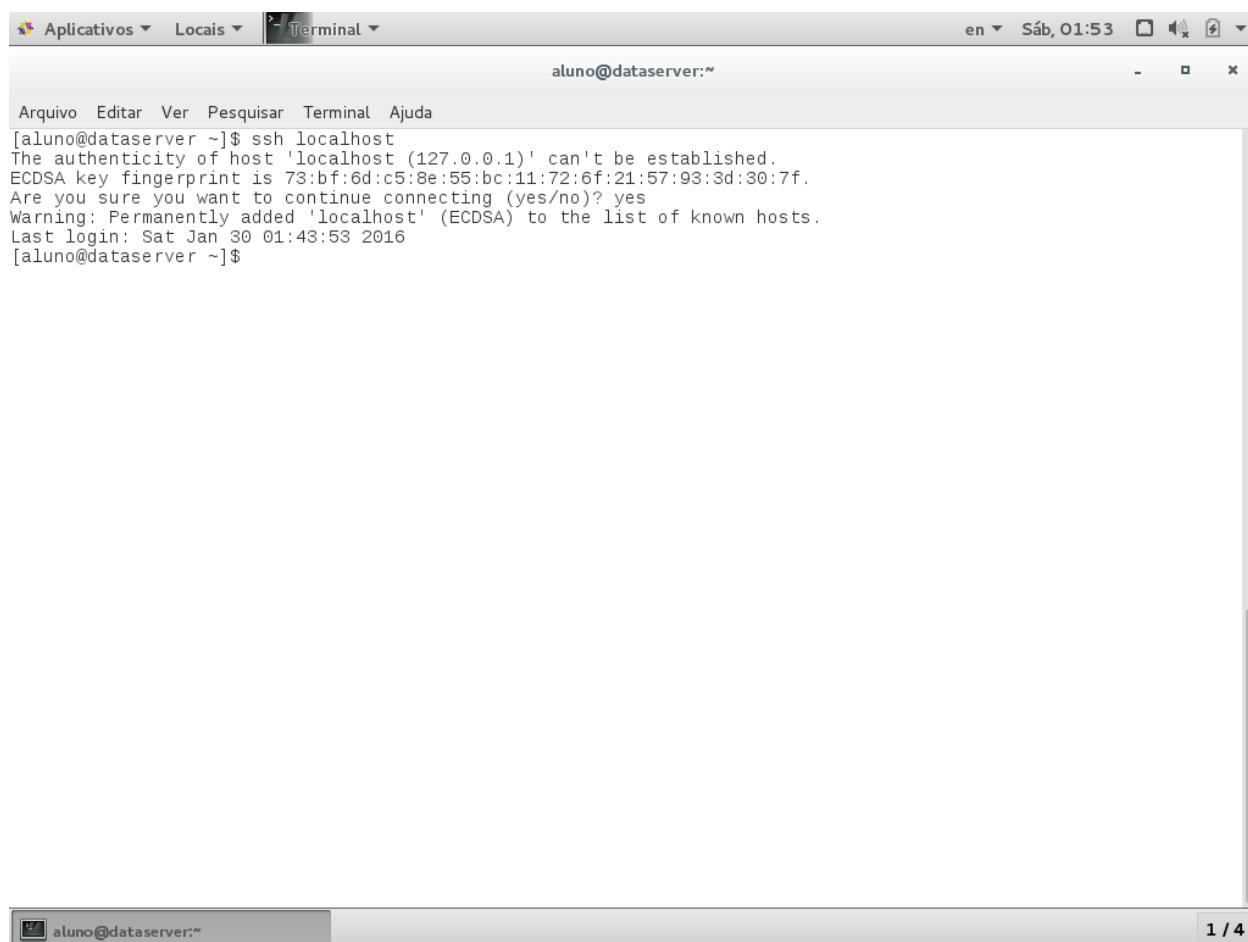


A screenshot of a terminal window titled "Terminal". The window shows a command-line interface with the user "aluno" connected to "localhost". The terminal displays a warning message about the host key fingerprint and asks if the user wants to continue connecting. The user has typed "yes" at the prompt.

```
[aluno@dataserver ~]$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is 73:bf:6d:c5:8e:55:bc:11:72:6f:21:57:93:3d:30:7f.
Are you sure you want to continue connecting (yes/no)? yes
```

1 / 4

Yes



A screenshot of a terminal window titled "Terminal". The window shows the command "ssh localhost" being run, followed by the output of the SSH handshake. The output includes the host's fingerprint, a confirmation prompt asking if the user wants to continue connecting, and a message indicating that the host has been added to the list of known hosts. The terminal window has a standard OS X-style interface with a menu bar at the top.

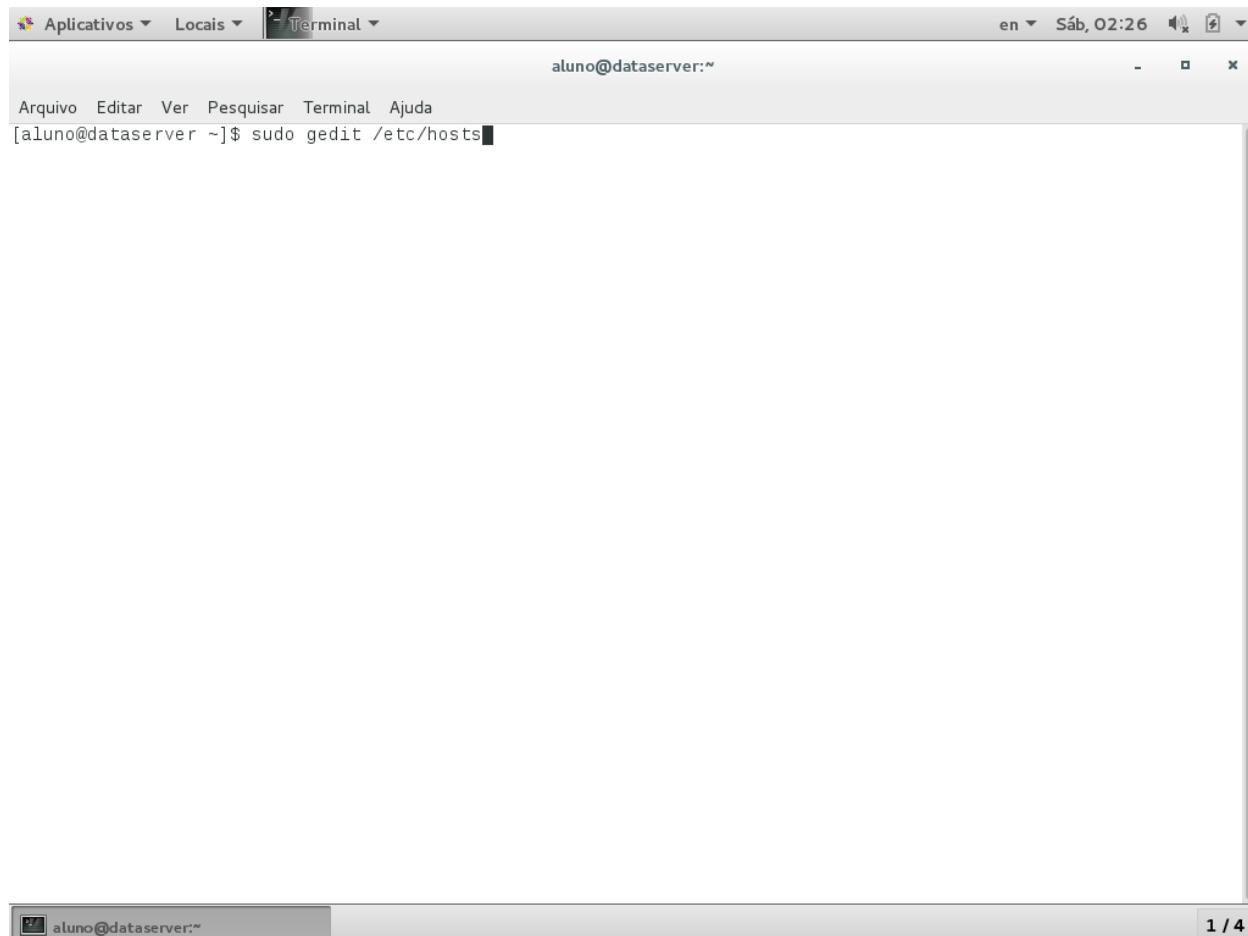
```
[aluno@dataserver ~]$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is 73:bf:6d:c5:8e:55:bc:11:72:6f:21:57:93:3d:30:7f.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Last login: Sat Jan 30 01:43:53 2016
[aluno@dataserver ~]$
```

Conexão ssh sem senha. Parabéns, seu servidor está pronto para receber o Hadoop!!

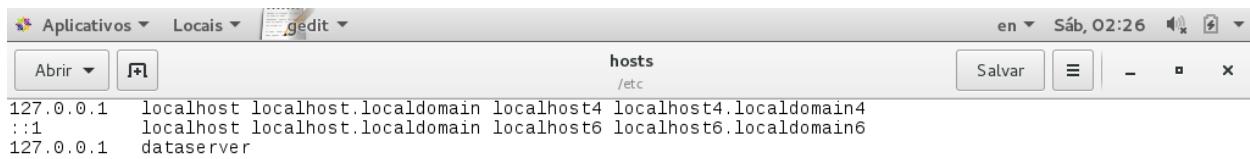
1 / 4

5.3. Download e Instalação do Hadoop

5.3.1. Editando o arquivo hosts



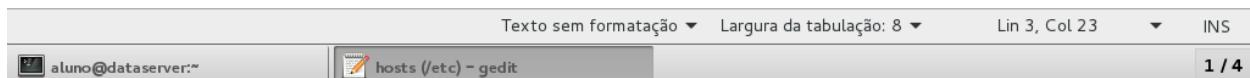
Editar o arquivo hosts



A screenshot of the gedit text editor window. The title bar shows "Aplicativos", "Locais", "gedit", "en", "Sáb, 02:26", and various window control buttons. The main area displays the "/etc/hosts" file with the following content:

```
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
127.0.0.1 dataserver
```

Buttons at the bottom right include "Salvar" and window control icons.



A screenshot of a terminal window titled "hosts (/etc) - gedit". The window shows the same content as the previous screenshot. The terminal prompt is "aluno@dataserver:~" and the status bar indicates "Lin 3, Col 23" and "1 / 4".

Incluir a última linha conforme acima

5.3.2. Download do Hadoop

Thu 20:25

Apache Hadoop Releases - Mozilla Firefox

Apache Hadoop Rel... [+ New Tab](#)

hadoop.apache.org/releases.html

Search

Apache > Hadoop >

APACHE 

Top Wiki

About

- Welcome
- [Releases](#)
- Download
- Release Notes
- Release Versioning
- Mailing Lists
- Issue Tracking
- Who We Are?
- Who Uses Hadoop?
- Buy Stuff
- Sponsorship
- Thanks
- Privacy Policy
- Bylaws
- Committer criteria
- License

Documentation

Related Projects

Search with Apache Solr [Search](#)

Last Published: 09/07/2016 12:32:06 [PDF](#)

Apache Hadoop Releases

Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-256.

Version	Release Date	Tarball	GPG	SHA-256
3.0.0-alpha1	03 September, 2016	source	signature	checksum file
		binary	signature	checksum file
2.7.3	25 August, 2016	source	signature	227785DC 6E3E6EF8..
		binary	signature	D489DF38 08244B90..
2.6.4	11 February, 2016	source	signature	F7550961 18316335..
		binary	signature	C58F08D2 E0B13035..
2.5.2	19 Nov, 2014	source	signature	139EF872 09C5637E..
		binary	signature	0BD84850 A3825208..

To verify Hadoop releases using GPG:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the signature file `hadoop-X.Y.Z-src.tar.gz.asc` from [Apache](#).
3. Download the [Hadoop KEYS](#) file.
4. `gpg --import KEYS`
5. `gpg --verify hadoop-X.Y.Z-src.tar.gz.asc`

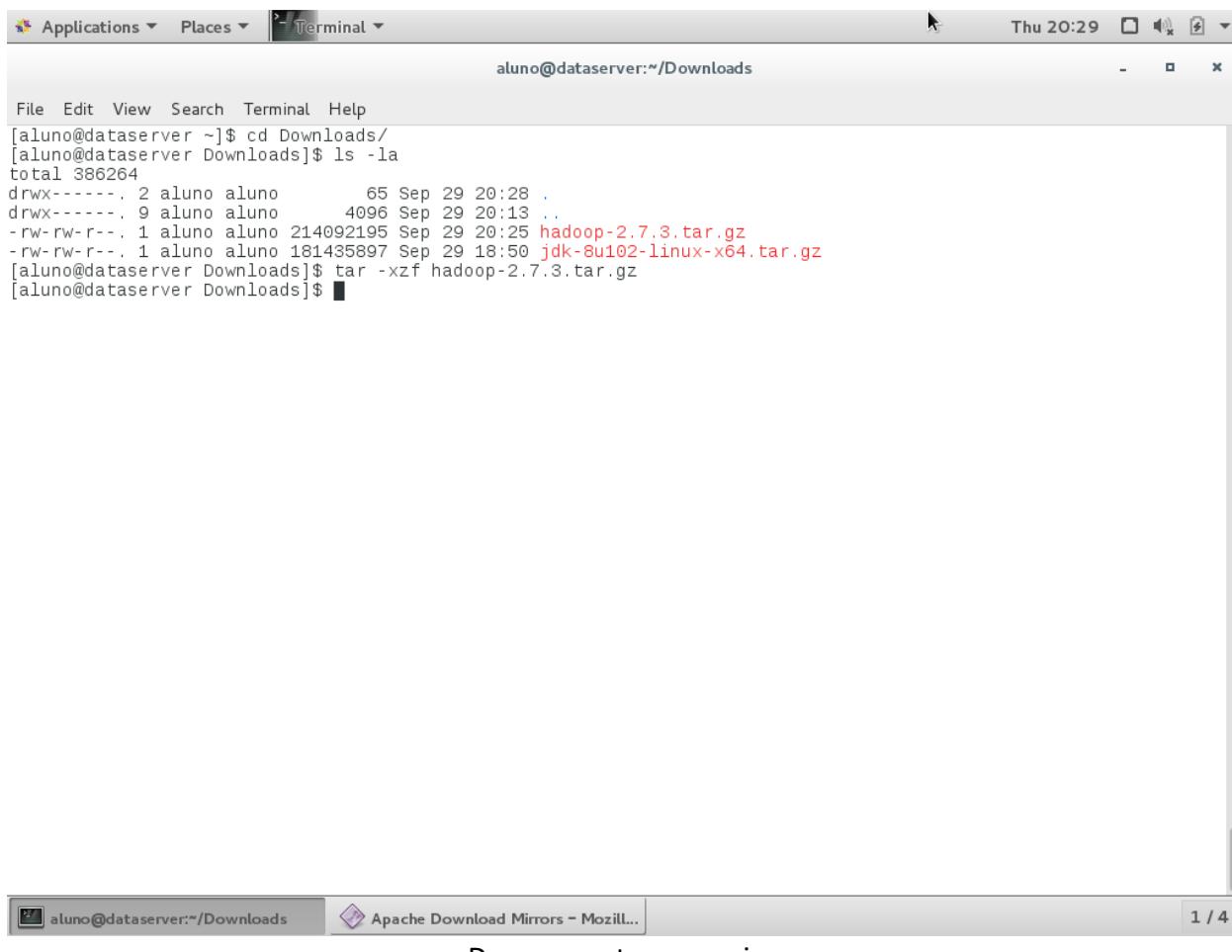
Acessar a página de downloads do Hadoop e selecionar a opção binary. O arquivo será baixado no diretório /home/aluno/Downloads



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~'. Below the prompt, the command 'wget http://apache.mirror.gtcomm.net/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz' is visible. The window has a standard title bar with icons for Applications, Places, and Terminal, and a system tray at the top right showing the date and time.

Outra opção para fazer o download pela linha de comando

Instalação e Configuração do Ecosistema Hadoop



The screenshot shows a terminal window titled 'Terminal' with the command line interface. The user is in the '/Downloads' directory and has run the following commands:

```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 386264
drwx----- 2 aluno aluno 65 Sep 29 20:28 .
drwx----- 9 aluno aluno 4096 Sep 29 20:13 ..
-rw-rw-r-- 1 aluno aluno 214092195 Sep 29 20:25 hadoop-2.7.3.tar.gz
-rw-rw-r-- 1 aluno aluno 181435897 Sep 29 18:50 jdk-8u102-linux-x64.tar.gz
[aluno@dataserver Downloads]$ tar -xzf hadoop-2.7.3.tar.gz
[aluno@dataserver Downloads]$
```

The terminal window has a title bar with 'Applications', 'Places', and 'Terminal'. The status bar at the bottom shows the date and time as 'Thu 20:29'. Below the terminal window, there is a browser tab for 'Apache Download Mirrors - Mozilla...' and a page number indicator '1 / 4'.

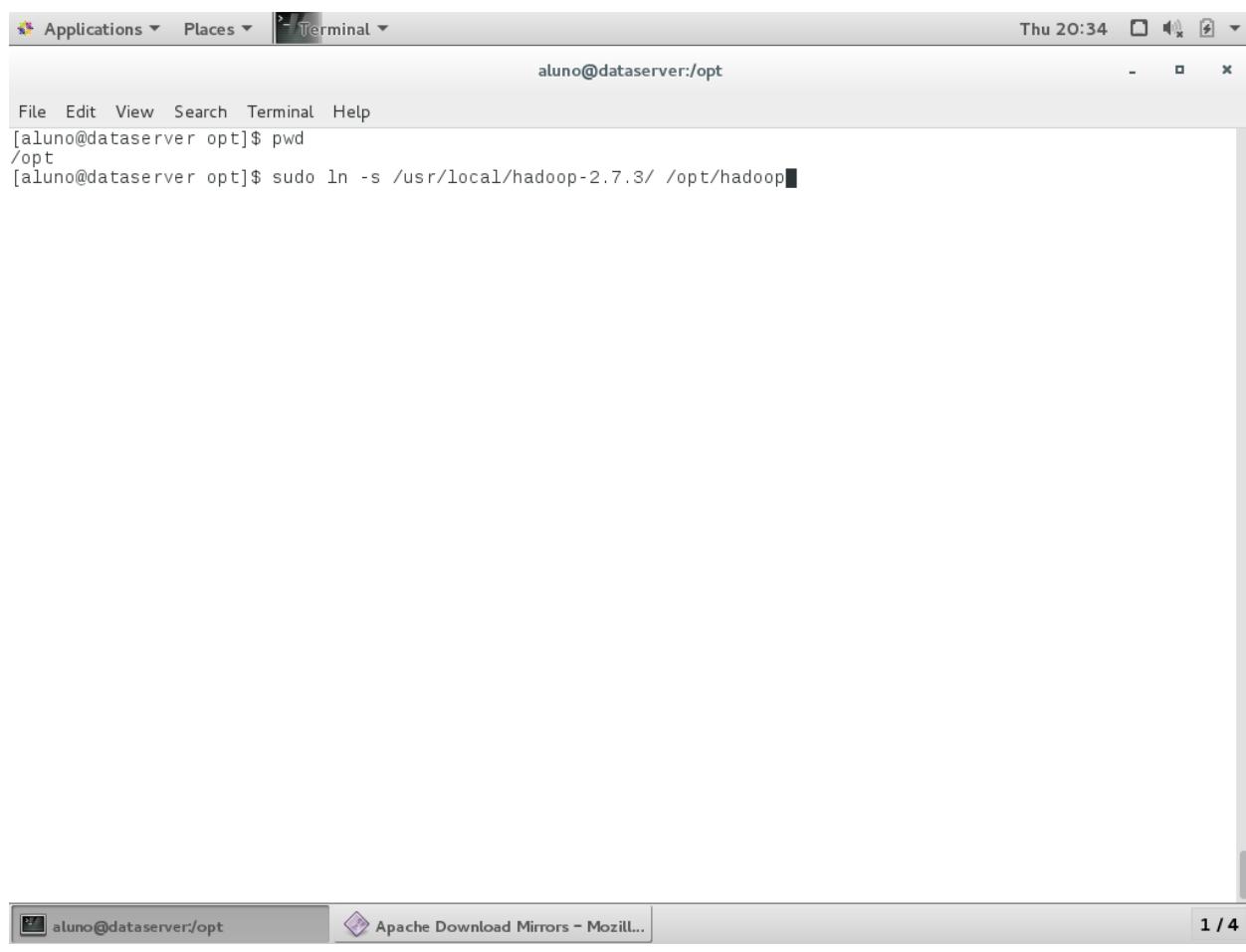
Descompactar o arquivo

Instalação e Configuração do Ecosistema Hadoop

The screenshot shows a terminal window titled 'Terminal' with the command-line interface. The user is navigating to the 'Downloads' directory and listing its contents. They then extract the 'hadoop-2.7.3.tar.gz' file and move it to the '/usr/local/' directory.

```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 386264
drwx----- 2 aluno aluno 65 Sep 29 20:28 .
drwx----- 9 aluno aluno 4096 Sep 29 20:13 ..
-rw-rw-r-- 1 aluno aluno 214092195 Sep 29 20:25 hadoop-2.7.3.tar.gz
-rw-rw-r-- 1 aluno aluno 181435897 Sep 29 18:50 jdk-8u102-linux-x64.tar.gz
[aluno@dataserver Downloads]$ tar -xzf hadoop-2.7.3.tar.gz
[aluno@dataserver Downloads]$ sudo mv hadoop-2.7.3 /usr/local/
```

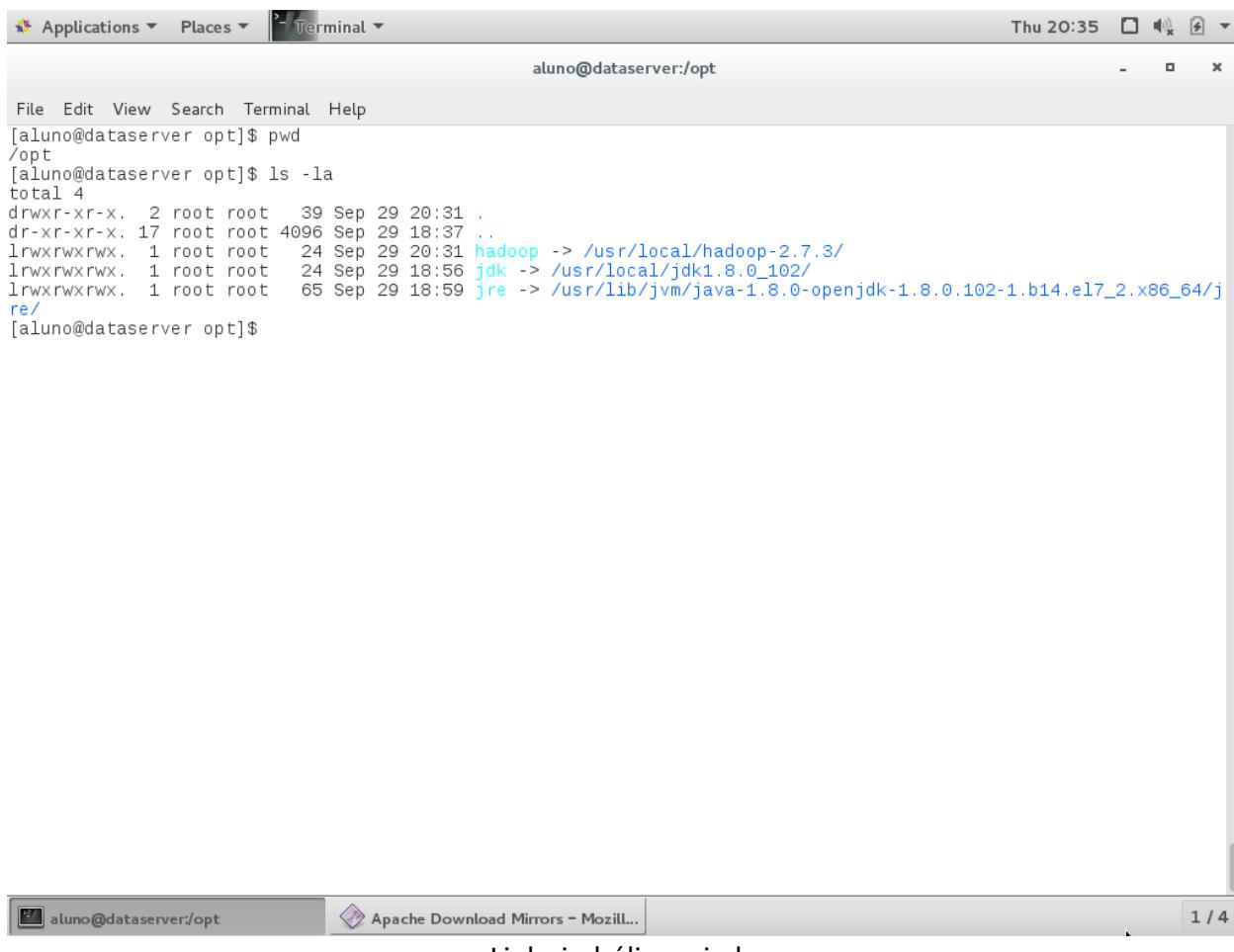
Mover o diretório para /usr/local/hadoop-2.7.3



The screenshot shows a terminal window titled 'Terminal' with the command line interface. The terminal window is part of a desktop environment with a menu bar at the top. The terminal content shows the user navigating to the '/opt' directory and creating a symbolic link named 'hadoop' pointing to '/usr/local/hadoop-2.7.3/'. The terminal window has a title bar with 'Applications', 'Places', and 'Terminal' buttons, and a status bar at the bottom.

```
aluno@dataserver:~$ pwd  
/opt  
[aluno@dataserver opt]$ sudo ln -s /usr/local/hadoop-2.7.3/ /opt/hadoop
```

Criar um link simbólico em /opt/hadoop



The screenshot shows a terminal window titled "Terminal" with the command-line interface. The window title bar includes "Applications", "Places", "Terminal", and the date/time "Thu 20:35". The terminal window itself has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The command history shows:

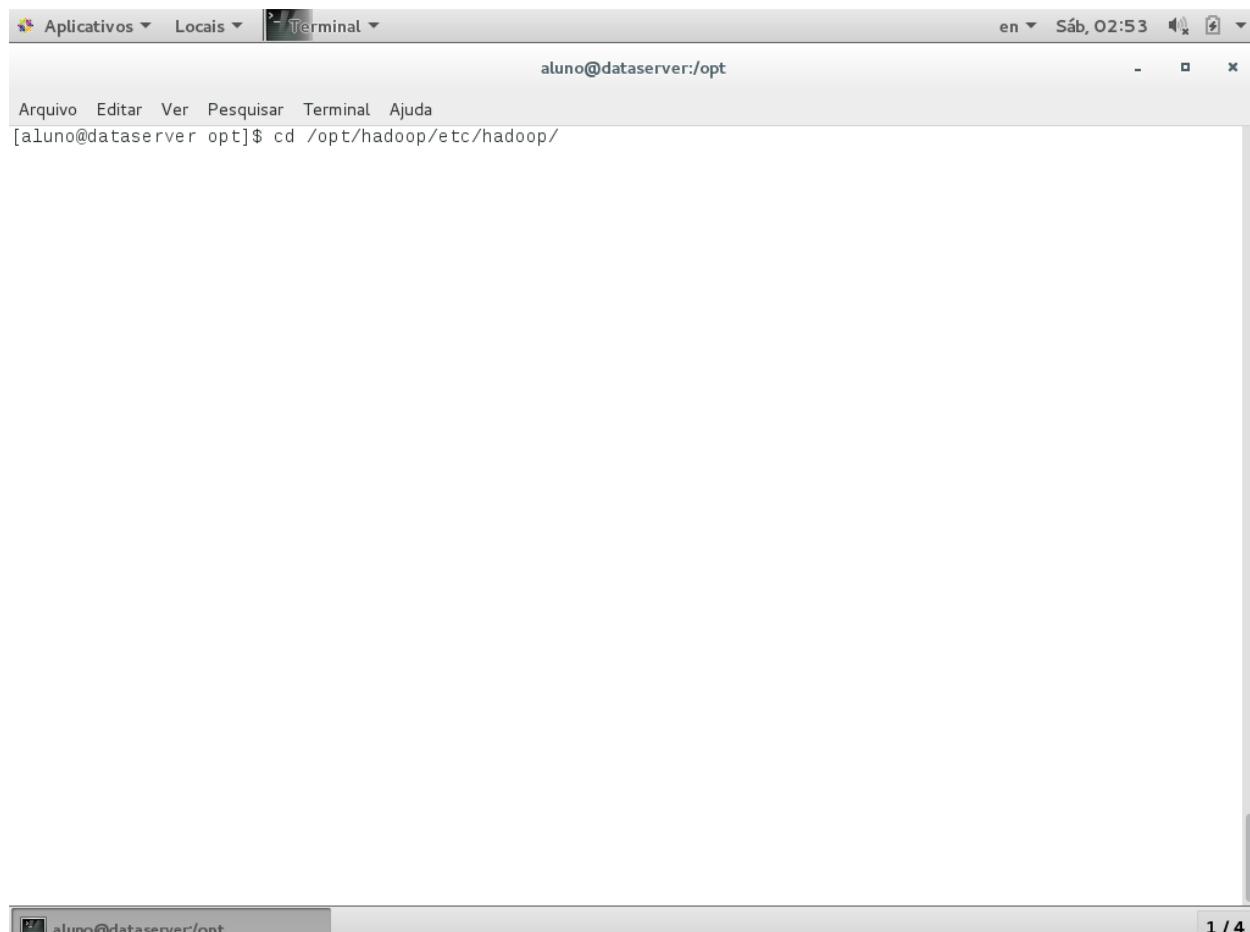
```
[aluno@dataserver opt]$ pwd  
/opt  
[aluno@dataserver opt]$ ls -la  
total 4  
drwxr-xr-x. 2 root root 39 Sep 29 20:31 .  
dr-xr-xr-x. 17 root root 4096 Sep 29 18:37 ..  
lrwxrwxrwx. 1 root root 24 Sep 29 20:31 hadoop -> /usr/local/hadoop-2.7.3/  
lrwxrwxrwx. 1 root root 24 Sep 29 18:56 jdk -> /usr/local/jdk1.8.0_102/  
lrwxrwxrwx. 1 root root 65 Sep 29 18:59 jre -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/  
[aluno@dataserver opt]$
```

The status bar at the bottom of the terminal window shows the user "aluno@dataserver:opt" and the URL "Apache Download Mirrors – Mozilla...". The page number "1 / 4" is also visible.

Link simbólico criado

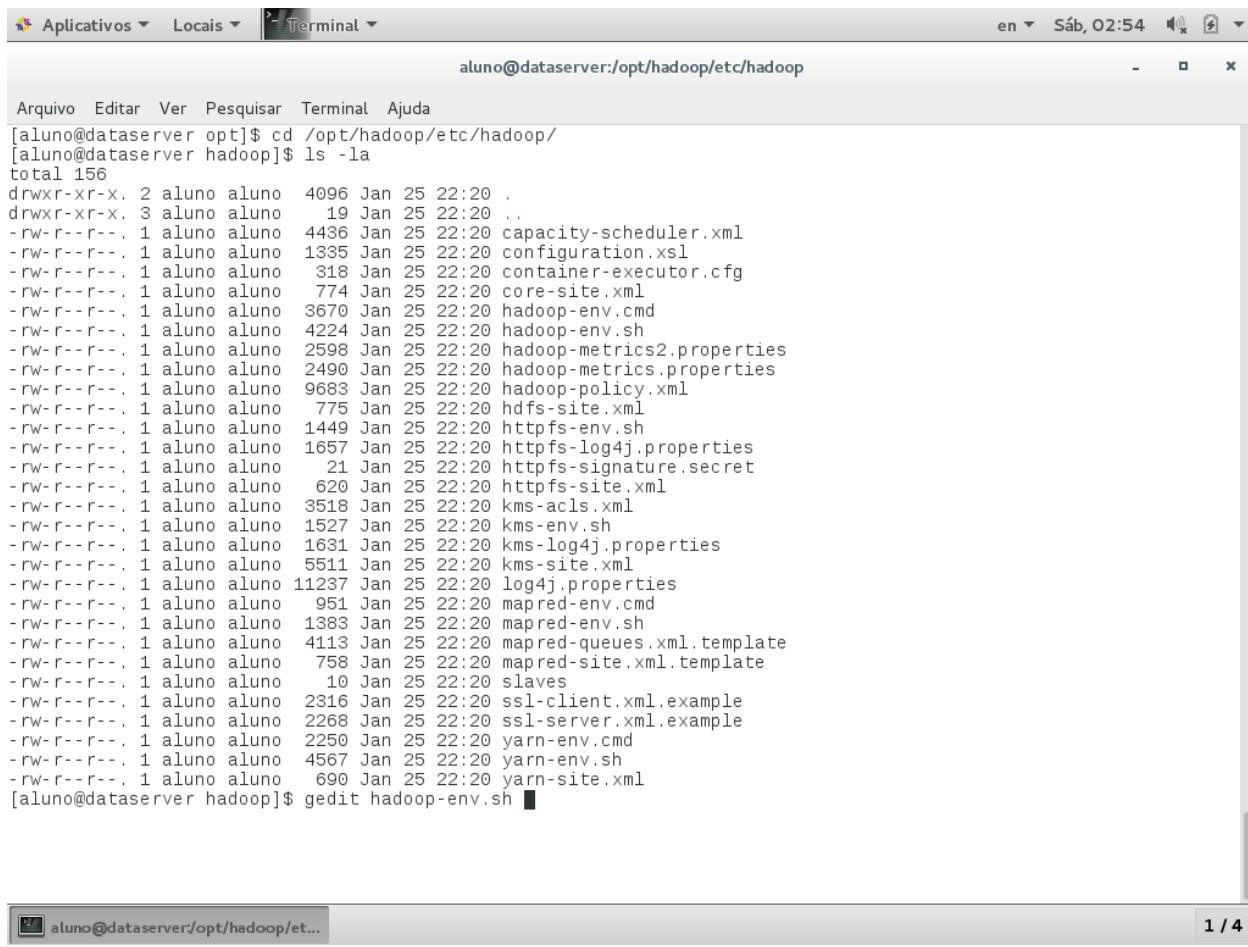
5.4. Configuração do Hadoop

5.4.1. Editar arquivos de configuração do Hadoop



```
aluno@dataserver:~$ cd /opt/hadoop/etc/hadoop/
```

Os arquivos de configuração do Hadoop estão em
[Diretório de instalação do Hadoop]/etc/hadoop
Nesse caso: /opt/hadoop/etc/hadoop



```
[aluno@dataserver opt]$ cd /opt/hadoop/etc/hadoop/
[aluno@dataserver hadoop]$ ls -la
total 156
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 .
drwxr-xr-x. 3 aluno aluno 19 Jan 25 22:20 ..
-rw-r--r--. 1 aluno aluno 4436 Jan 25 22:20 capacity-scheduler.xml
-rw-r--r--. 1 aluno aluno 1335 Jan 25 22:20 configuration.xsl
-rw-r--r--. 1 aluno aluno 318 Jan 25 22:20 container-executor.cfg
-rw-r--r--. 1 aluno aluno 774 Jan 25 22:20 core-site.xml
-rw-r--r--. 1 aluno aluno 3670 Jan 25 22:20 hadoop-env.cmd
-rw-r--r--. 1 aluno aluno 4224 Jan 25 22:20 hadoop-env.sh
-rw-r--r--. 1 aluno aluno 2598 Jan 25 22:20 hadoop-metrics2.properties
-rw-r--r--. 1 aluno aluno 2490 Jan 25 22:20 hadoop-metrics.properties
-rw-r--r--. 1 aluno aluno 9683 Jan 25 22:20 hadoop-policy.xml
-rw-r--r--. 1 aluno aluno 775 Jan 25 22:20 hdfs-site.xml
-rw-r--r--. 1 aluno aluno 1449 Jan 25 22:20 httpfs-env.sh
-rw-r--r--. 1 aluno aluno 1657 Jan 25 22:20 httpfs-log4j.properties
-rw-r--r--. 1 aluno aluno 21 Jan 25 22:20 httpfs-signature.secret
-rw-r--r--. 1 aluno aluno 620 Jan 25 22:20 httpfs-site.xml
-rw-r--r--. 1 aluno aluno 3518 Jan 25 22:20 kms-acls.xml
-rw-r--r--. 1 aluno aluno 1527 Jan 25 22:20 kms-env.sh
-rw-r--r--. 1 aluno aluno 1631 Jan 25 22:20 kms-log4j.properties
-rw-r--r--. 1 aluno aluno 5511 Jan 25 22:20 kms-site.xml
-rw-r--r--. 1 aluno aluno 11237 Jan 25 22:20 log4j.properties
-rw-r--r--. 1 aluno aluno 951 Jan 25 22:20 mapred-env.cmd
-rw-r--r--. 1 aluno aluno 1383 Jan 25 22:20 mapred-env.sh
-rw-r--r--. 1 aluno aluno 4113 Jan 25 22:20 mapred-queues.xml.template
-rw-r--r--. 1 aluno aluno 758 Jan 25 22:20 mapred-site.xml.template
-rw-r--r--. 1 aluno aluno 10 Jan 25 22:20 slaves
-rw-r--r--. 1 aluno aluno 2316 Jan 25 22:20 ssl-client.xml.example
-rw-r--r--. 1 aluno aluno 2268 Jan 25 22:20 ssl-server.xml.example
-rw-r--r--. 1 aluno aluno 2250 Jan 25 22:20 yarn-env.cmd
-rw-r--r--. 1 aluno aluno 4567 Jan 25 22:20 yarn-env.sh
-rw-r--r--. 1 aluno aluno 690 Jan 25 22:20 yarn-site.xml
[aluno@dataserver hadoop]$ gedit hadoop-env.sh
```

Editar o arquivo hadoop-env.sh

1 / 4

```

# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License. You may obtain a copy of the License at
#
#     http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# Set Hadoop-specific environment variables here.

# The only required environment variable is JAVA_HOME. All others are
# optional. When running a distributed configuration it is best to
# set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.

# The java implementation to use.
export JAVA_HOME=/opt/jdk

# Pasta onde o Hadoop estiver instalador
export HADOOP_PREFIX=/opt/hadoop

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol. Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.
#export JSVC_HOME=${JSVC_HOME}

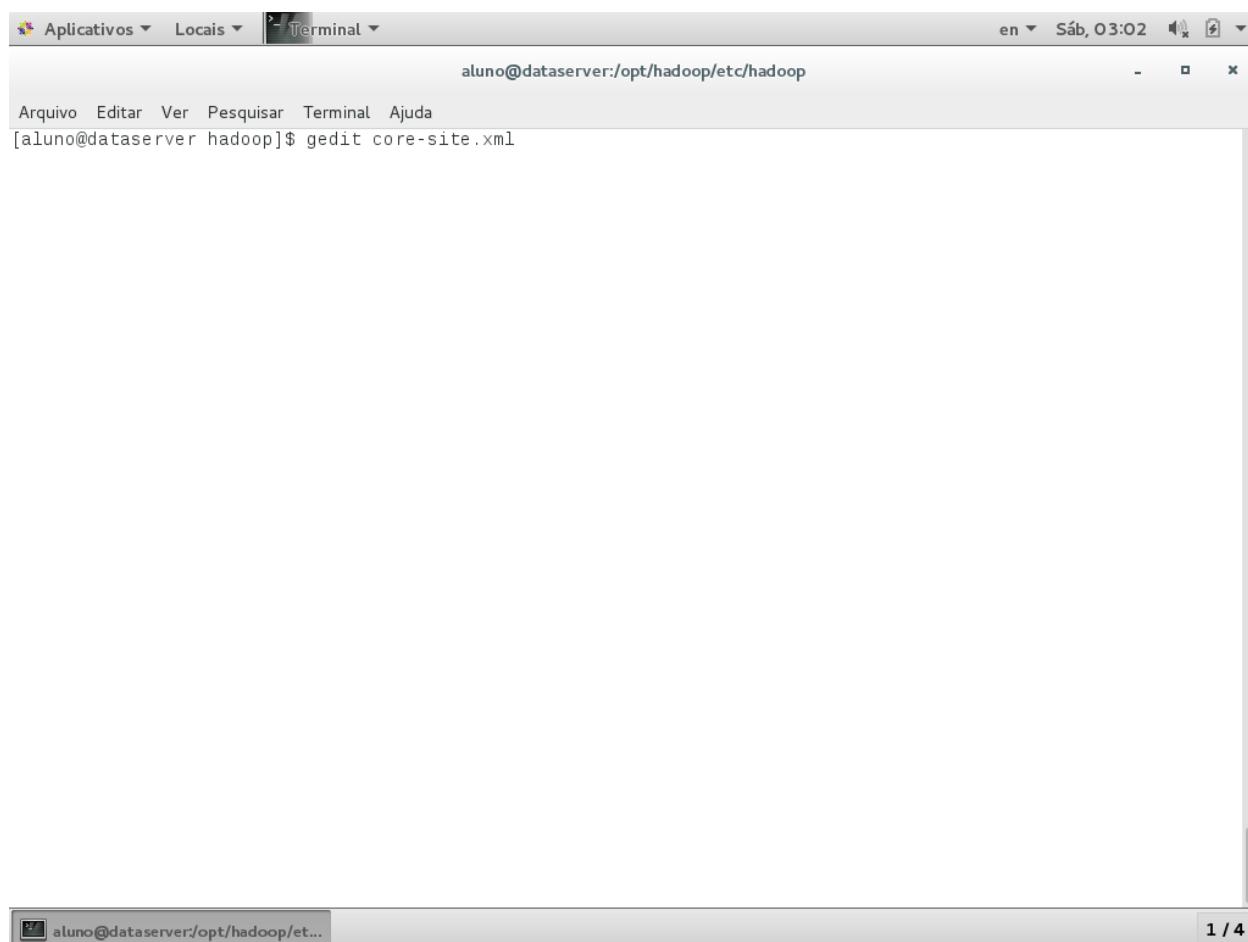
export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}
```

sh ▾ Tab Width: 8 ▾ Ln 28, Col 33 ▾ INS

aluno@dataserver:~ | hadoop-env.sh (/opt/hadoop/etc/...) 1 / 4

[Editar/acrescentar as linhas acima](#)

Instalação e Configuração do Ecosistema Hadoop



Editar o arquivo core-site.xml

Instalação e Configuração do Ecosistema Hadoop

The screenshot shows a terminal window titled "gedit" with the file "core-site.xml" open. The file path is "/opt/hadoop/etc/hadoop". The content of the file is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

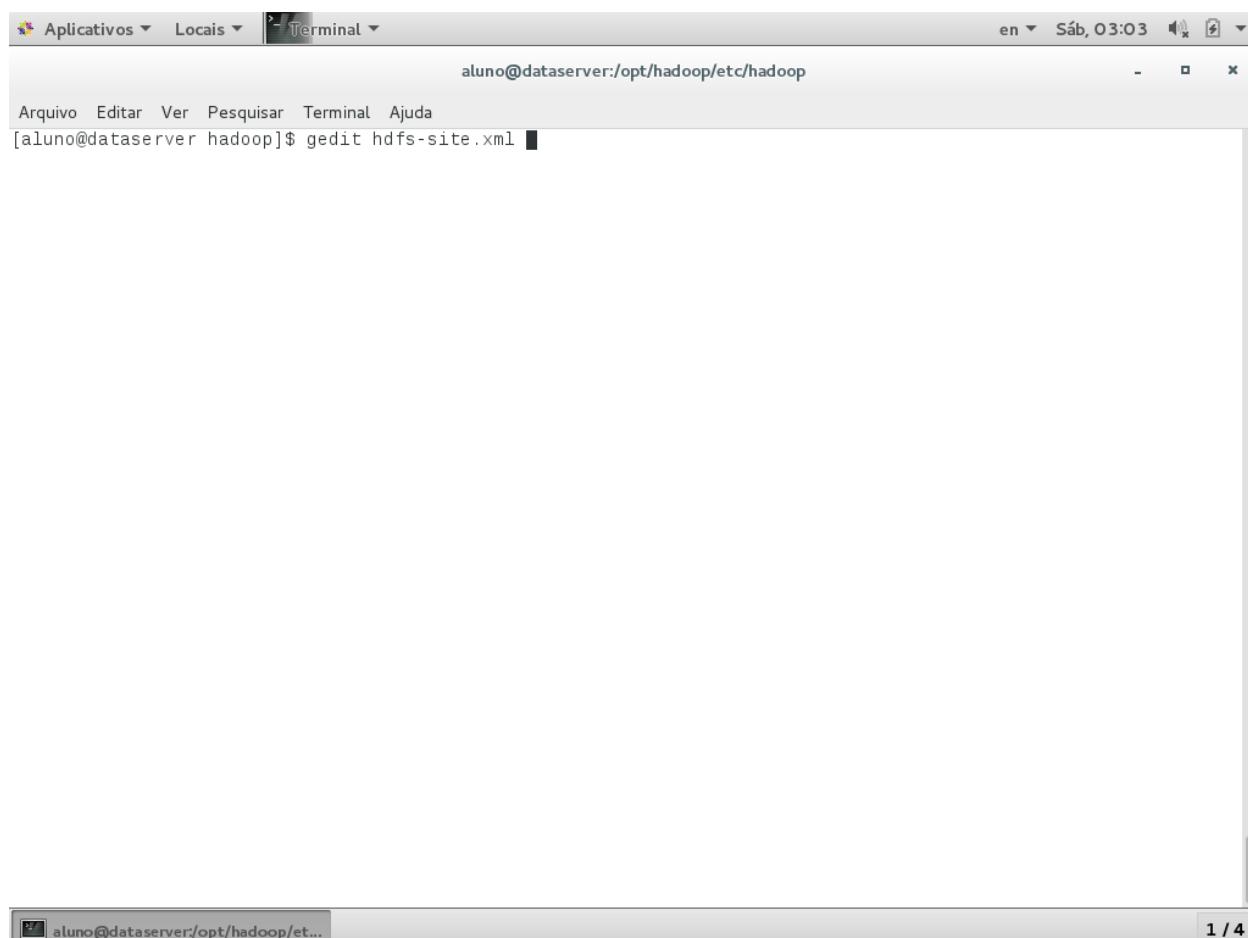
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

At the bottom of the terminal window, there is a status bar with the following information:

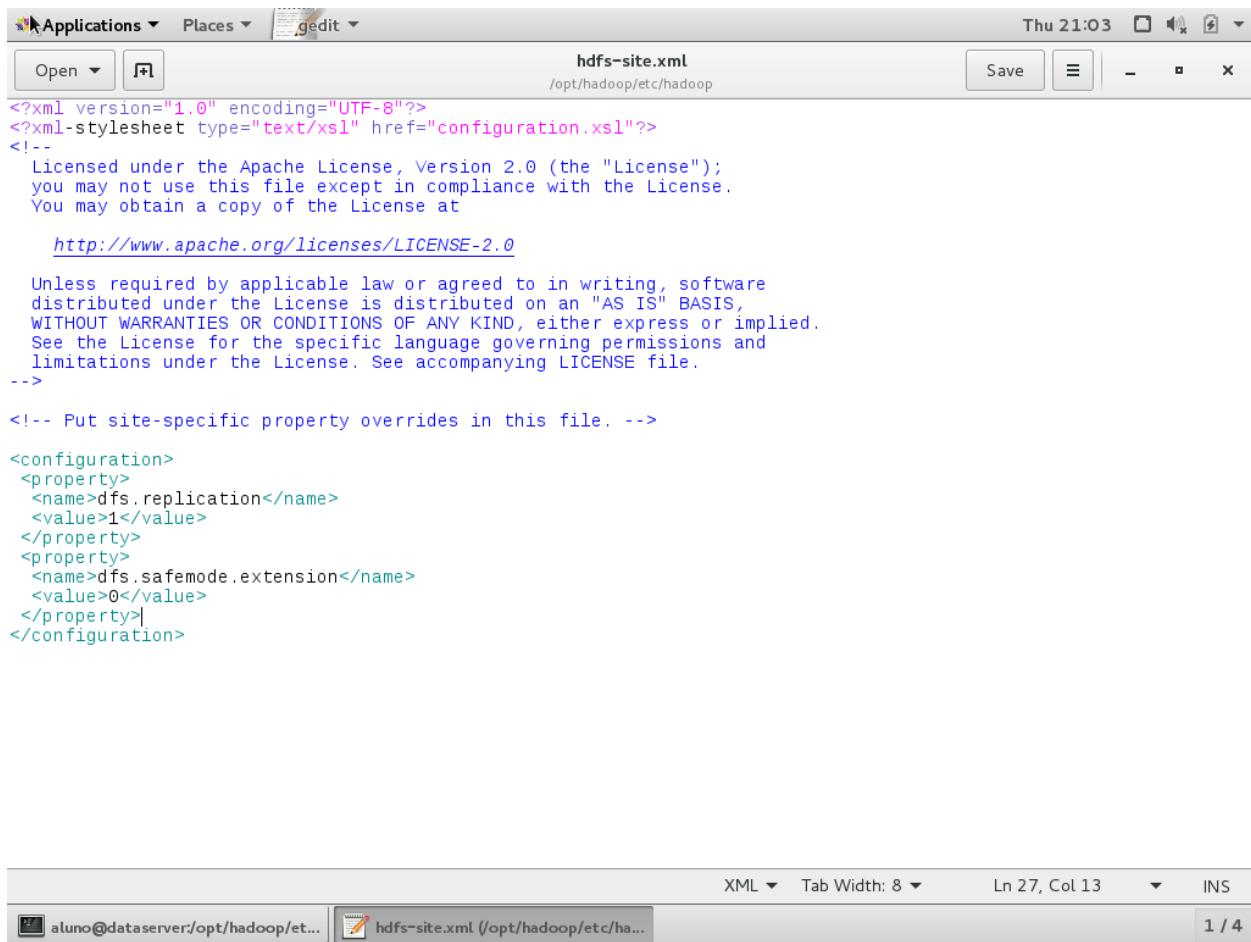
Salvando o arquivo "/opt/hadoop/etc/hadoop/core-site.xml"... XML ▾ Largura da tabulação: 8 ▾ Lin 23, Col 13 ▾ INS aluno@dataserver:/opt/hadoop/etc/et... core-site.xml (/opt/hadoop/etc/ha... 1 / 4

Acrescentar as propriedades conforme acima e salvar o arquivo. Isso permite configurar o Hadoop em modo Pseudo-distribuído

Instalação e Configuração do Ecosistema Hadoop



Editar o arquivo hdfs-site.xml



The screenshot shows a terminal window titled "gedit" with the file "hdfs-site.xml" open. The file path is "/opt/hadoop/etc/hadoop". The content of the file is as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
 Licensed under the Apache License, Version 2.0 (the "License");
 you may not use this file except in compliance with the License.
 You may obtain a copy of the License at
 http://www.apache.org/licenses/LICENSE-2.0

 Unless required by applicable law or agreed to in writing, software
 distributed under the License is distributed on an "AS IS" BASIS,
 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 See the License for the specific language governing permissions and
 limitations under the License. See accompanying LICENSE file.
-->

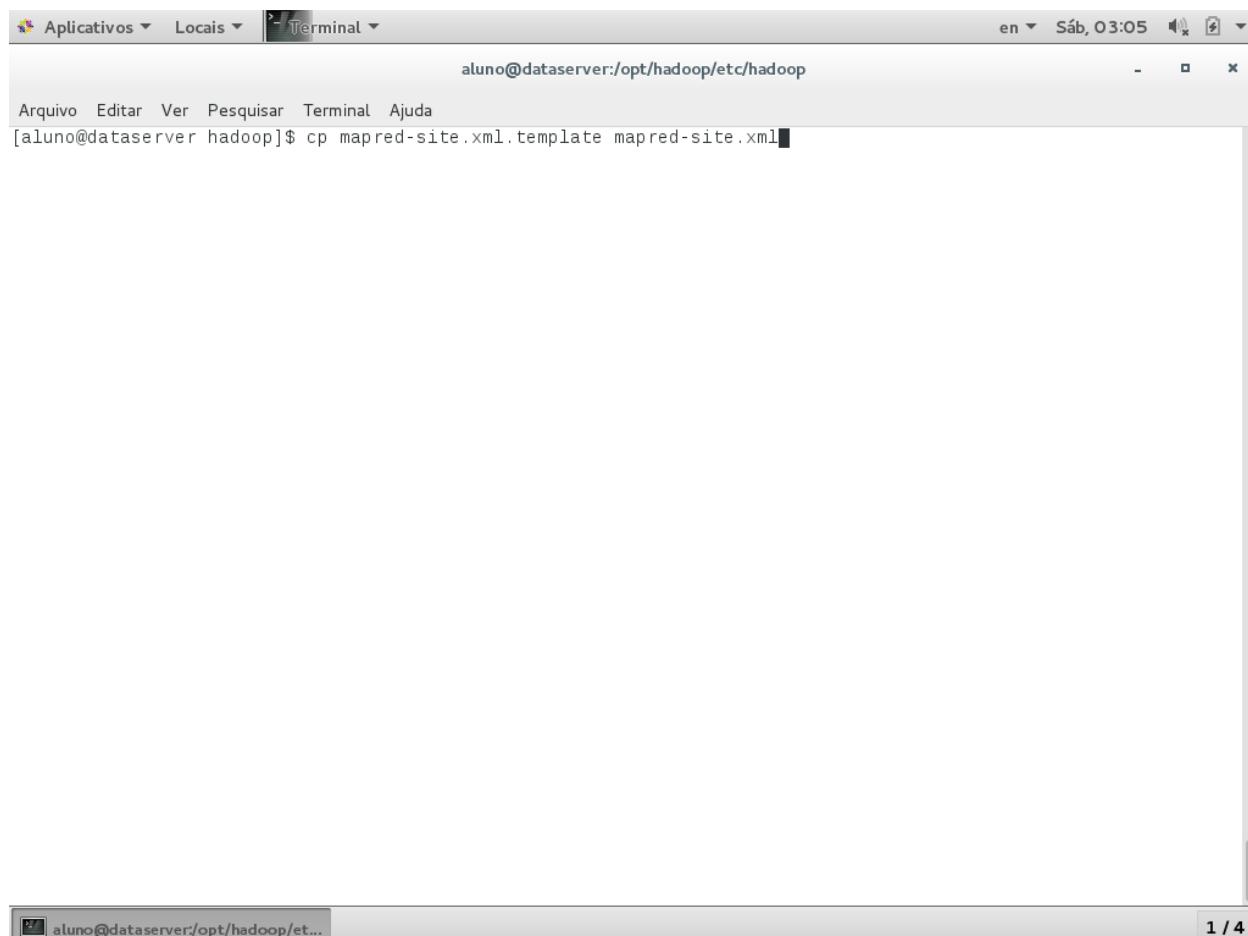
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.safemode.extension</name>
<value>0</value>
</property>
</configuration>
```

The terminal window also shows the command line: aluno@dataserver:/opt/hadoop/etc... and the file name hdfs-site.xml (/opt/hadoop/etc/ha...). The status bar indicates XML, Tab Width: 8, Ln 27, Col 13, and 1 / 4.

Acrescentar as propriedades conforme acima e salvar o arquivo

Instalação e Configuração do Ecosistema Hadoop

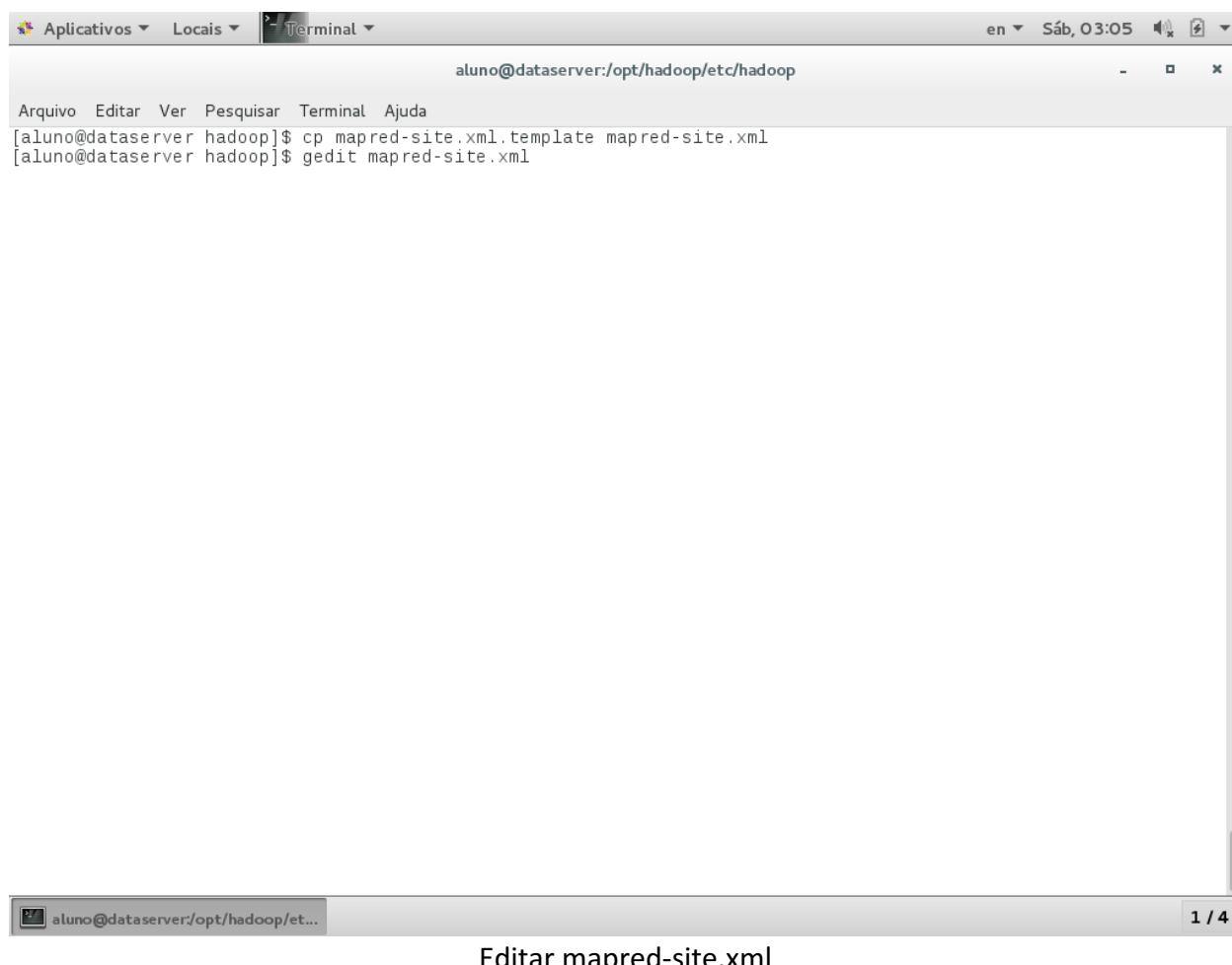


A screenshot of a Linux terminal window titled "Terminal". The window has a menu bar with "Aplicativos", "Locais", and "Terminal". The status bar shows "en" and "Sáb, 03:05". The terminal prompt is "aluno@dataserver:/opt/hadoop/etc/hadoop". The command entered is "[aluno@dataserver hadoop]\$ cp mapred-site.xml.template mapred-site.xml". The window has standard window controls (minimize, maximize, close) in the top right corner.

Copiar o arquivo template e gerar o arquivo mapred-site.xml

1 / 4

Instalação e Configuração do Ecosistema Hadoop



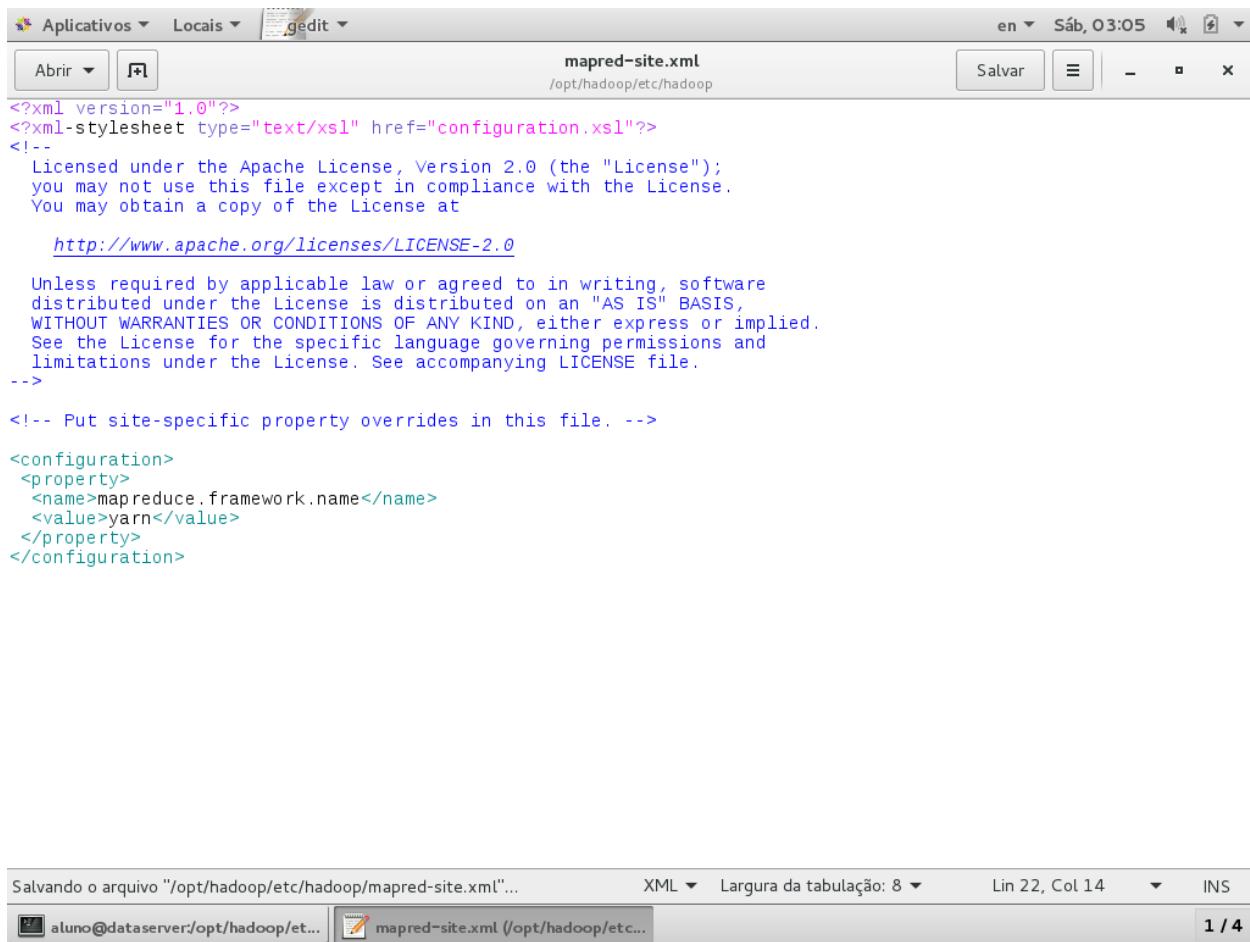
A screenshot of a Linux desktop environment showing a terminal window titled "Terminal". The window title bar also includes "Aplicativos" and "Locais". The status bar at the top right shows "en" and "Sáb, 03:05". The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command line shows the user navigating to the Hadoop configuration directory and performing file operations:

```
aluno@dataserver:/opt/hadoop/etc/hadoop
[aluno@dataserver hadoop]$ cp mapred-site.xml.template mapred-site.xml
[aluno@dataserver hadoop]$ gedit mapred-site.xml
```

The bottom status bar of the terminal window shows "aluno@dataserver:/opt/hadoop/et..." and "1 / 4".

Editar mapred-site.xml

Instalação e Configuração do Ecosistema Hadoop



The screenshot shows a terminal window titled "gedit" displaying the XML configuration file "mapred-site.xml" located at "/opt/hadoop/etc/hadoop". The file contains the Apache License 2.0 header and a section for site-specific overrides. The terminal status bar indicates the file is being saved.

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

Salvando o arquivo "/opt/hadoop/etc/hadoop/mapred-site.xml"...

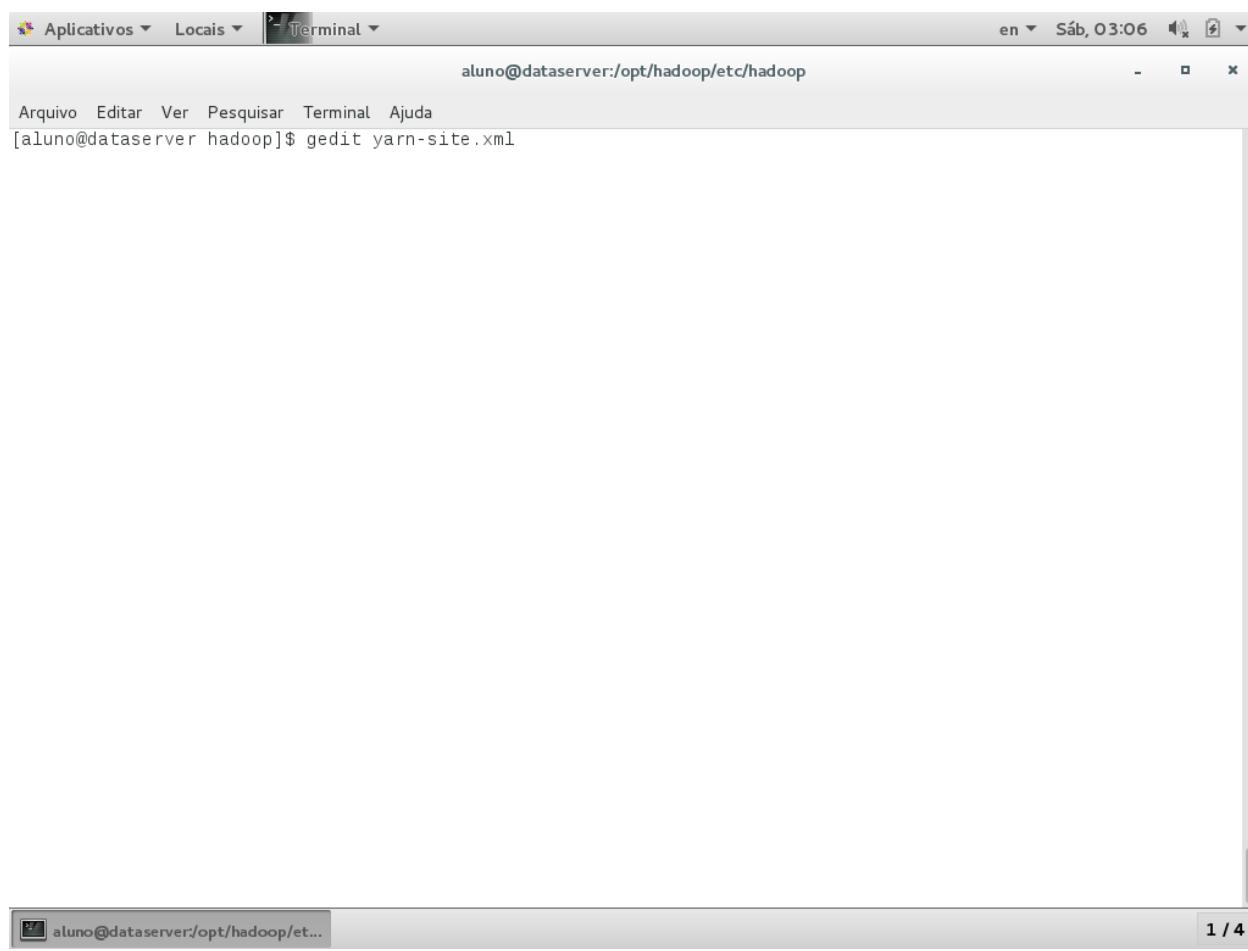
XML ▾ Largura da tabulação: 8 ▾ Lin 22, Col 14 ▾ INS

aluno@dataserver:/opt/hadoop/etc... mapred-site.xml (/opt/hadoop/etc...)

1 / 4

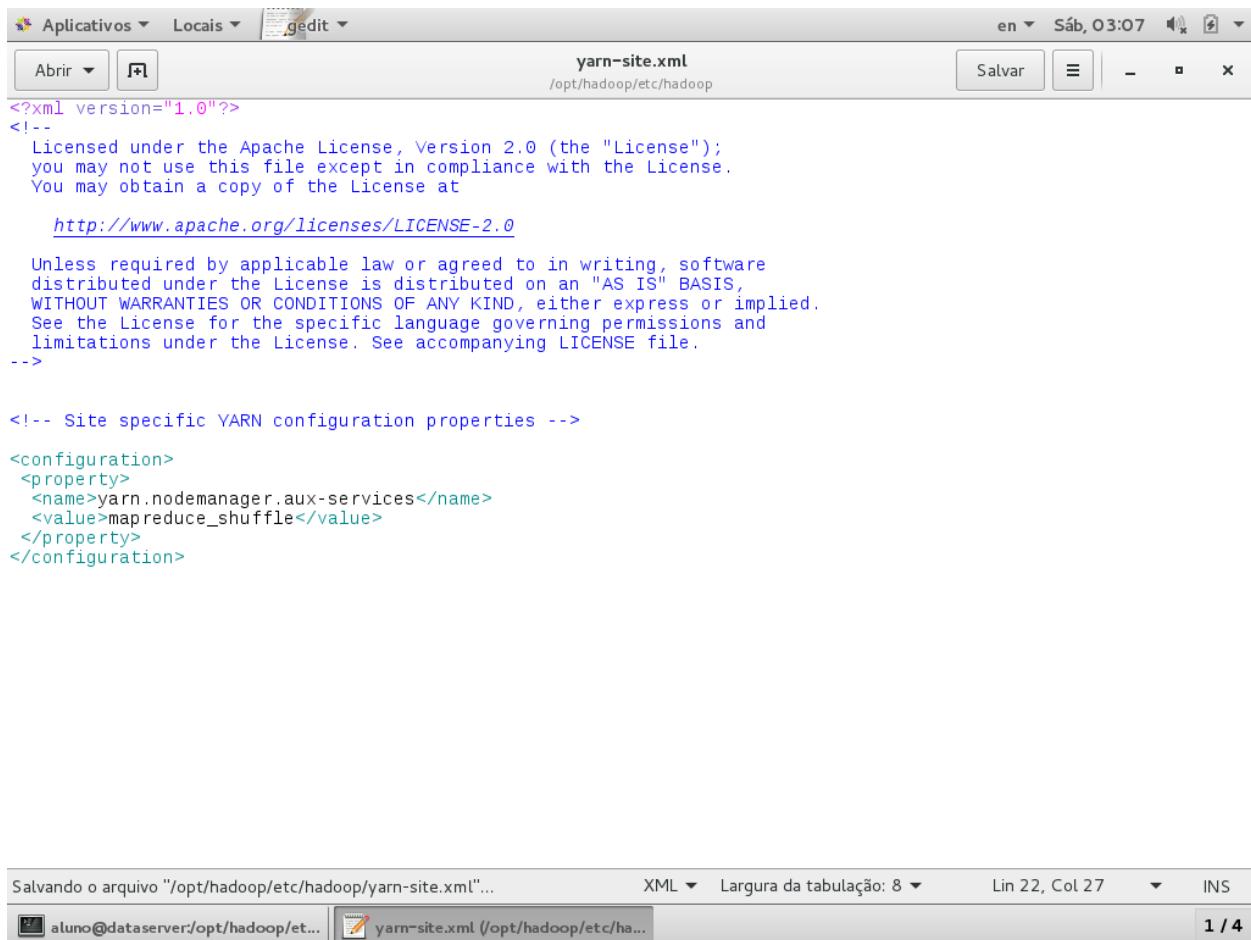
Acrescentar as propriedades conforme acima e salvar o arquivo

Instalação e Configuração do Ecosistema Hadoop



Editar o arquivo yarn-site.xml

Instalação e Configuração do Ecosistema Hadoop



The screenshot shows a terminal window titled "gedit" with the file "yarn-site.xml" open. The file path is "/opt/hadoop/etc/hadoop". The content of the file is as follows:

```
<?xml version="1.0"?>
<!--
 Licensed under the Apache License, Version 2.0 (the "License");
 you may not use this file except in compliance with the License.
 You may obtain a copy of the License at

 http://www.apache.org/licenses/LICENSE-2.0

 Unless required by applicable law or agreed to in writing, software
 distributed under the License is distributed on an "AS IS" BASIS,
 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
 See the License for the specific language governing permissions and
 limitations under the License. See accompanying LICENSE file.
-->

<!-- Site specific YARN configuration properties -->

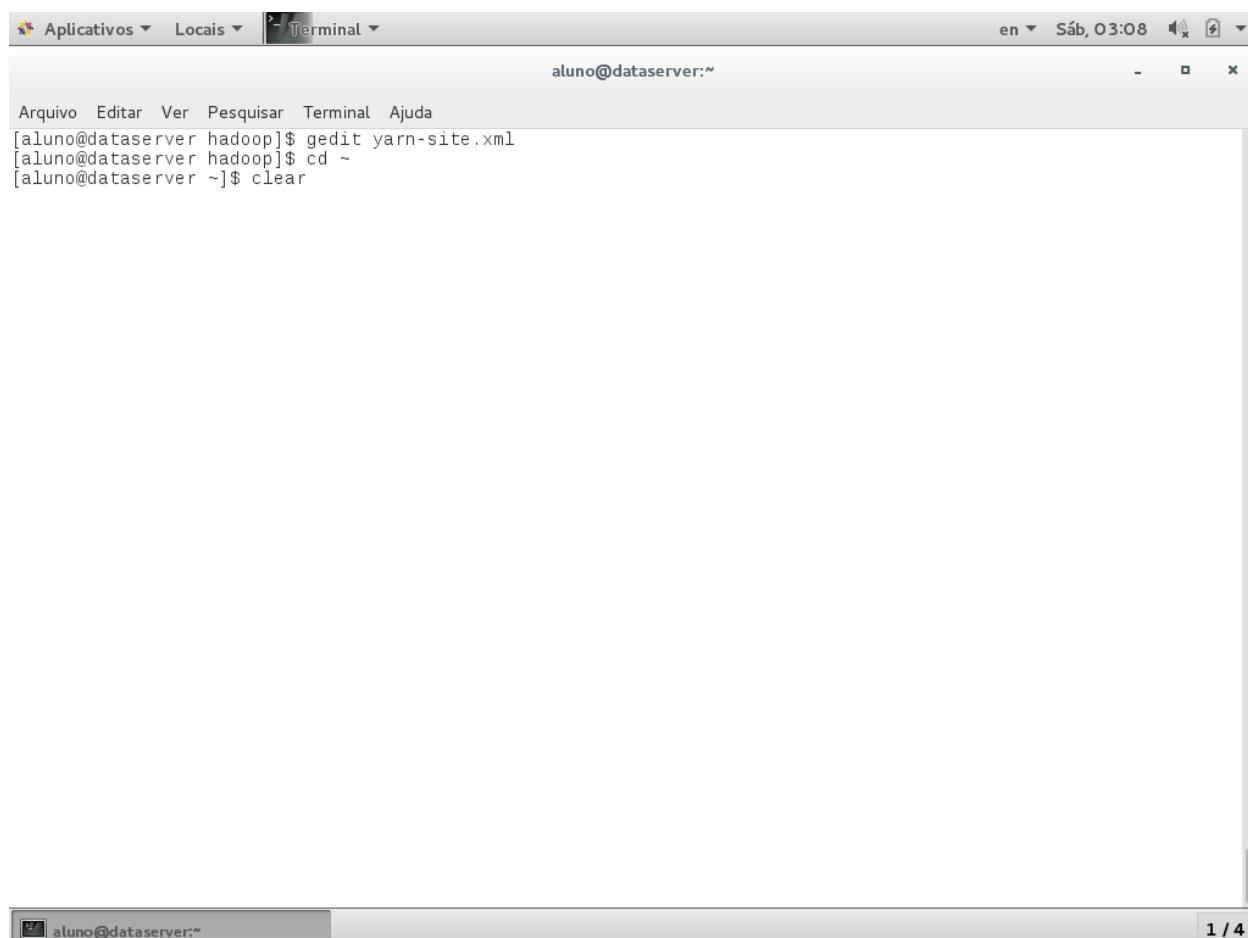
<configuration>
 <property>
 <name>yarn.nodemanager.aux-services</name>
 <value>mapreduce_shuffle</value>
 </property>
</configuration>
```

At the bottom of the terminal window, there is a status bar with the following information:

Salvando o arquivo "/opt/hadoop/etc/hadoop/yarn-site.xml"... XML ▾ Largura da tabulação: 8 ▾ Lin 22, Col 27 ▾ INS
aluno@dataserver:/opt/hadoop/etc... yarn-site.xml (/opt/hadoop/etc/ha... 1 / 4

Acrescentar as propriedades conforme acima e salvar o arquivo

Instalação e Configuração do Ecosistema Hadoop



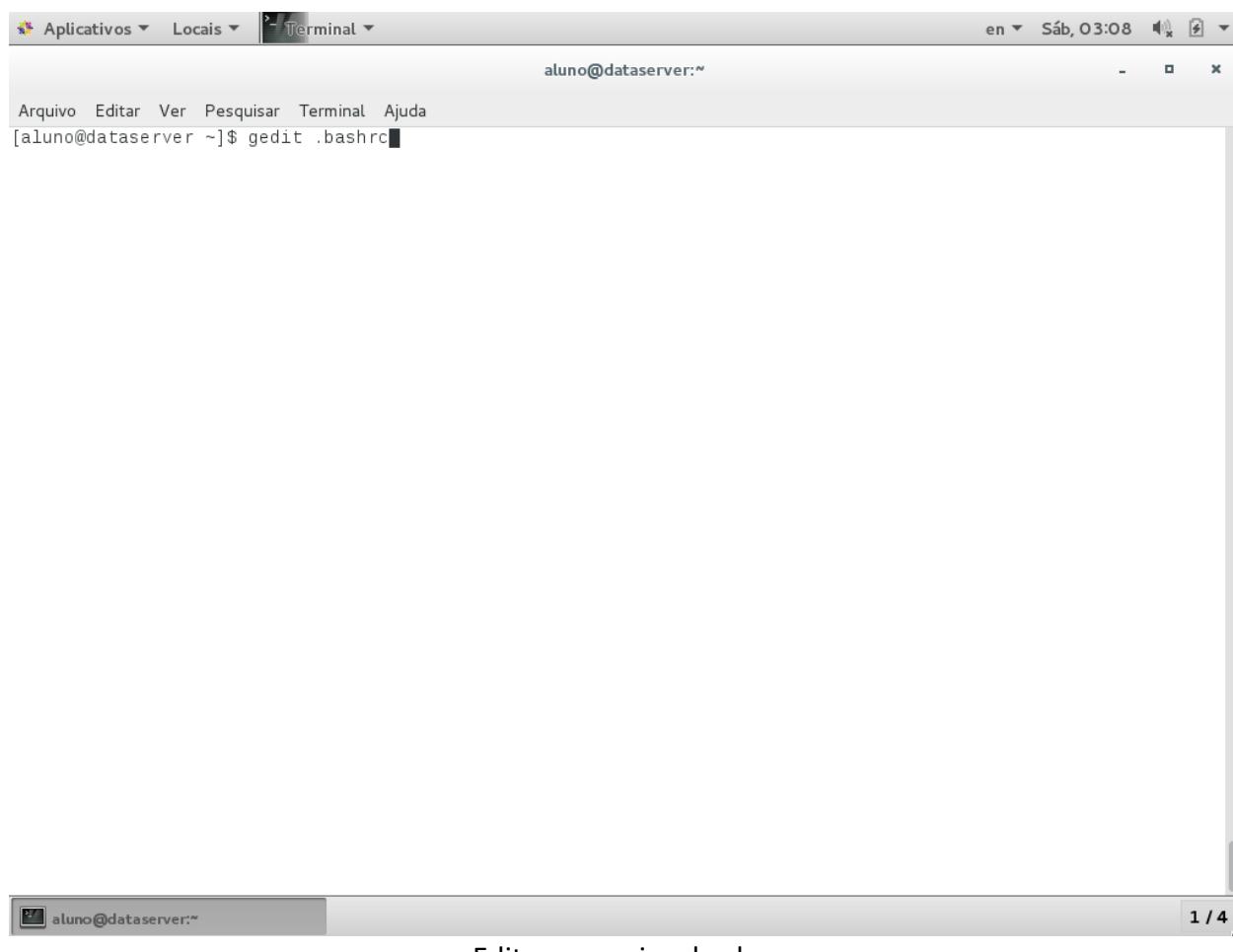
A screenshot of a Linux terminal window titled "Terminal". The window shows the following command history:

```
[aluno@dataserver hadoop]$ gedit yarn-site.xml
[aluno@dataserver hadoop]$ cd ~
[aluno@dataserver ~]$ clear
```

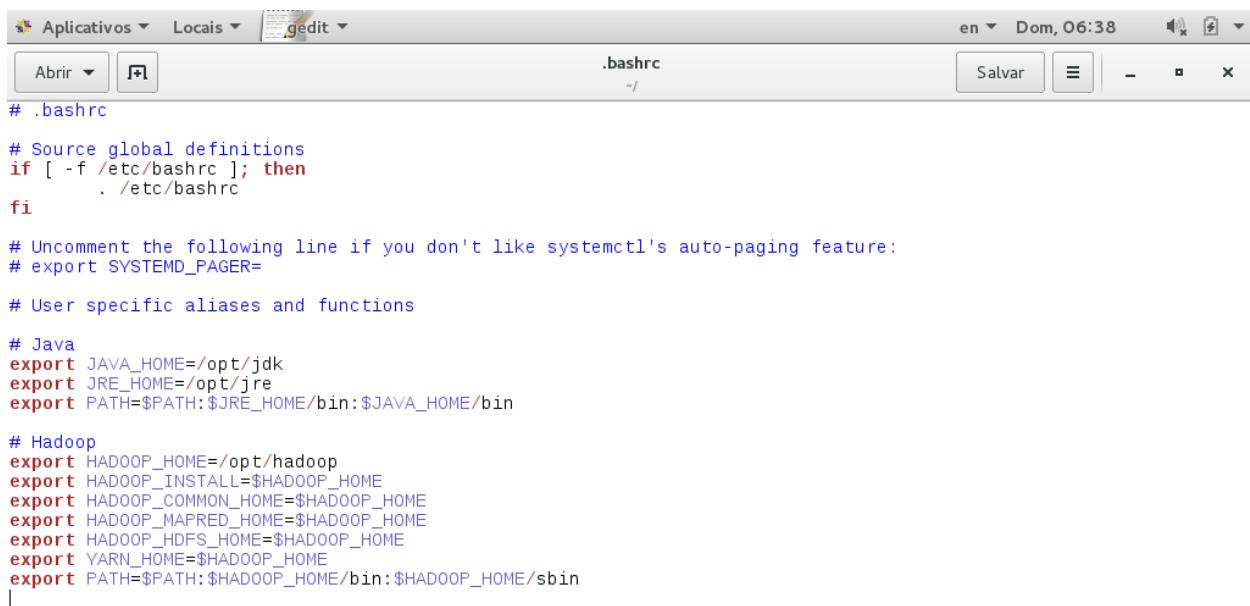
The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows "aluno@dataserver:~". A scroll bar is visible on the right side of the terminal window.

Ir para o diretório home e limpar a tela

1 / 4



Editar o arquivo .bashrc



A screenshot of a terminal window titled "gedit" showing the contents of the ".bashrc" file. The file contains configuration for Java and Hadoop environments. It includes sourcing the global definitions from /etc/bashrc, setting up Java paths, and defining Hadoop environment variables like HADOOP_HOME, HADOOP_INSTALL, etc.

```
# .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
export JAVA_HOME=/opt/jdk
export JRE_HOME=/opt/jre
export PATH=$PATH:$JRE_HOME/bin:$JAVA_HOME/bin

# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```



A screenshot of a terminal window showing the ".bashrc" file loaded in a terminal session. The terminal title is "aluno@dataserver:~". The status bar indicates "sh" as the shell, "Largura da tabulação: 8", "Lin 26, Col 1", and "1 / 4".

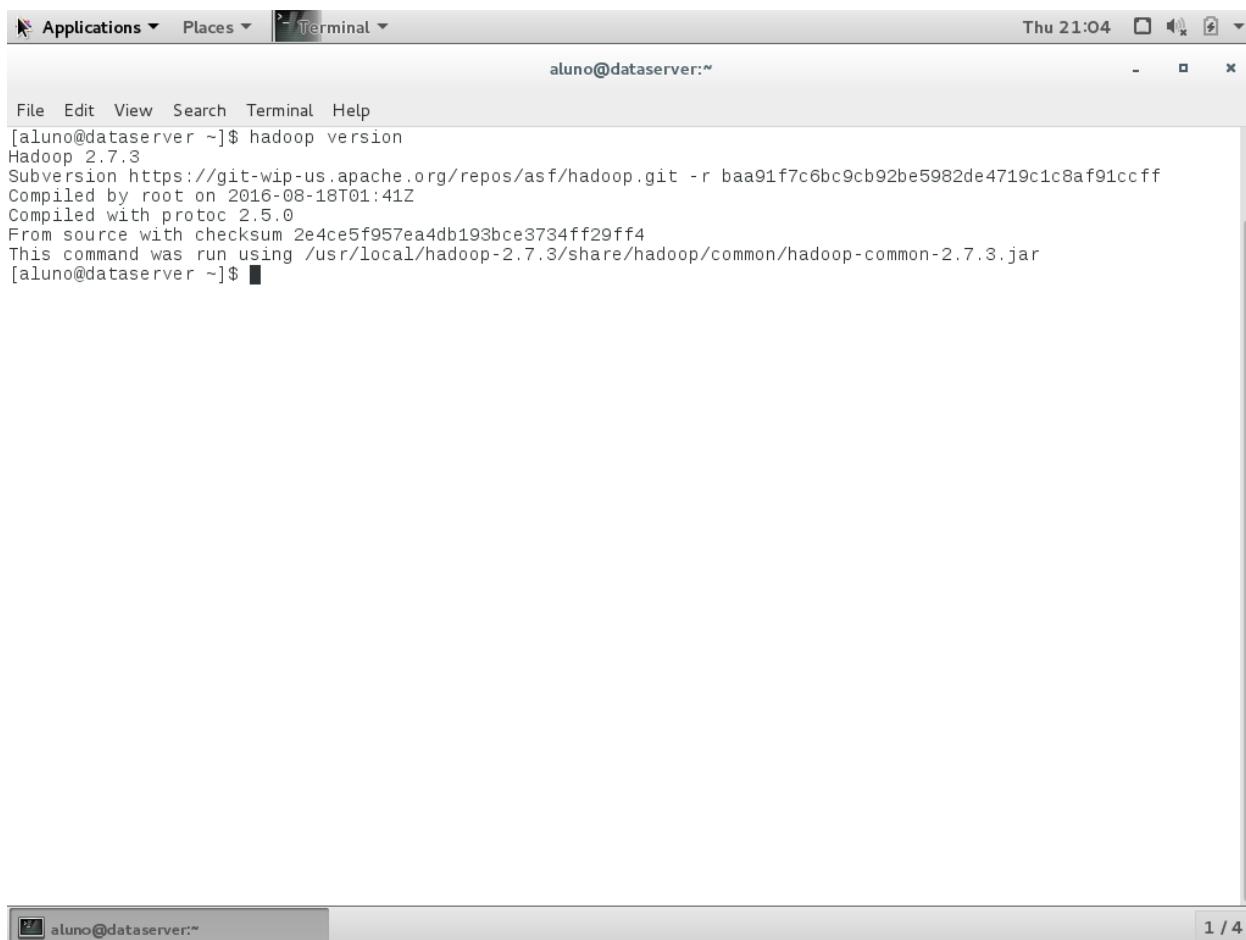
Acrescentar os executáveis do Hadoop no PATH e variáveis de ambiente

The screenshot shows a terminal window titled "Terminal" with the user "aluno@dataserver:~". The window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows "aluno@dataserver:~". The terminal content is as follows:

```
[aluno@dataserver ~]$ gedit .bashrc
[aluno@dataserver ~]$ source .bashrc
[aluno@dataserver ~]$
```

Below the terminal window, there is a status bar with the text "aluno@dataserver:~" and "1 / 4".

source .bashrc

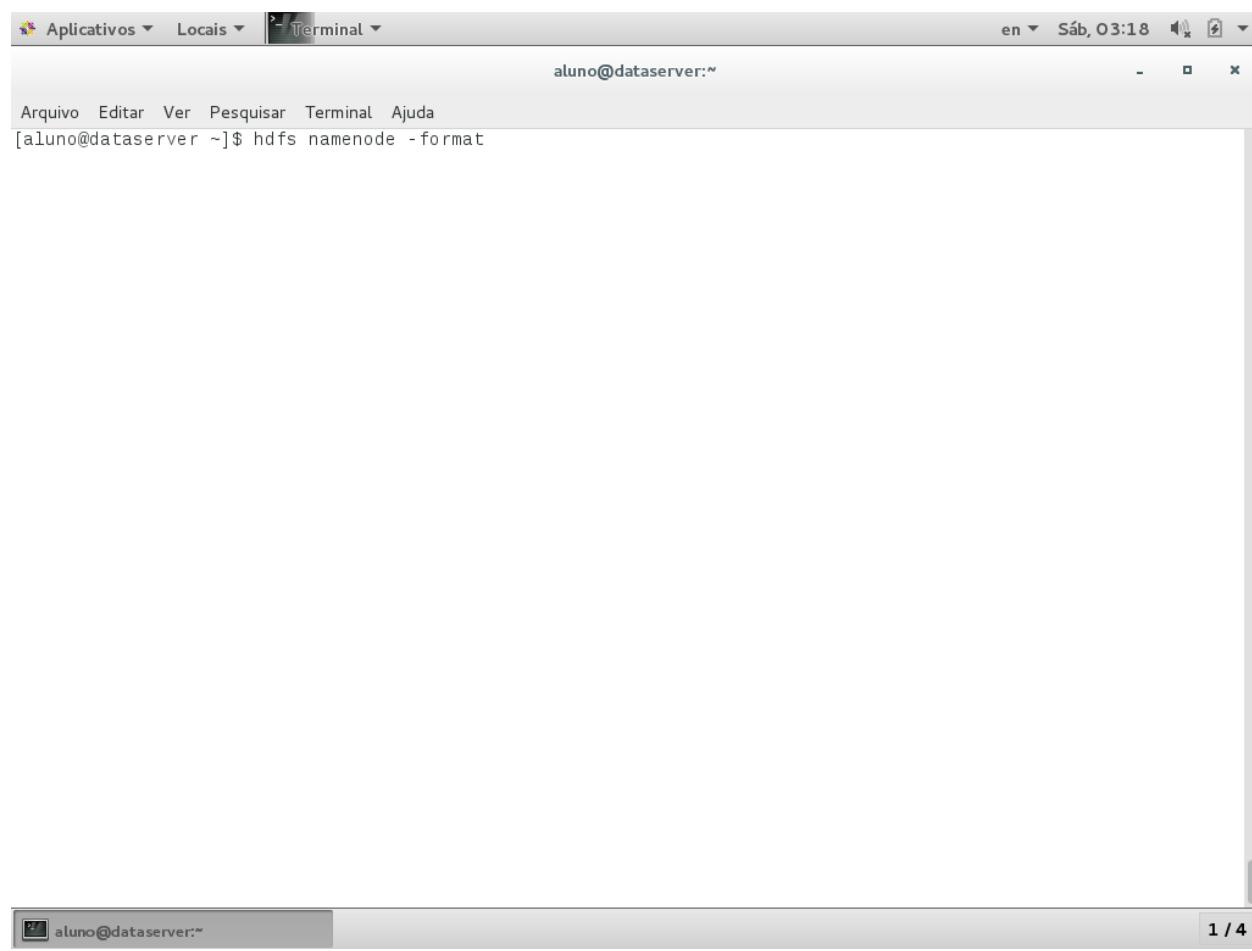


A screenshot of a Linux terminal window titled "Terminal". The window shows the command "hadoop version" being run by a user named "aluno" at a terminal prompt. The output of the command is displayed, indicating that Hadoop 2.7.3 is installed, compiled from source, and running on a specific date and time. The terminal window has a standard title bar with icons for Applications, Places, and Terminal, and a status bar at the bottom right showing the date and time.

```
[aluno@dataserver ~]$ hadoop version
Hadoop 2.7.3
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r baa91f7c6bc9cb92be5982de4719c1c8af91ccff
Compiled by root on 2016-08-18T01:41Z
Compiled with protoc 2.5.0
From source with checksum 2e4ce5f957ea4db193bce3734ff29ff4
This command was run using /usr/local/hadoop-2.7.3/share/hadoop/common/hadoop-common-2.7.3.jar
[aluno@dataserver ~]$
```

hadoop version → Hadoop instalado com sucesso!! Parabéns!

5.4.2. Formatando o Namenode



A screenshot of a terminal window titled "Terminal". The window shows the command `[aluno@dataserver ~]$ hdfs namenode -format` being typed. The terminal interface includes a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom left shows "aluno@dataserver:~" and the status bar at the bottom right shows "1 / 4".

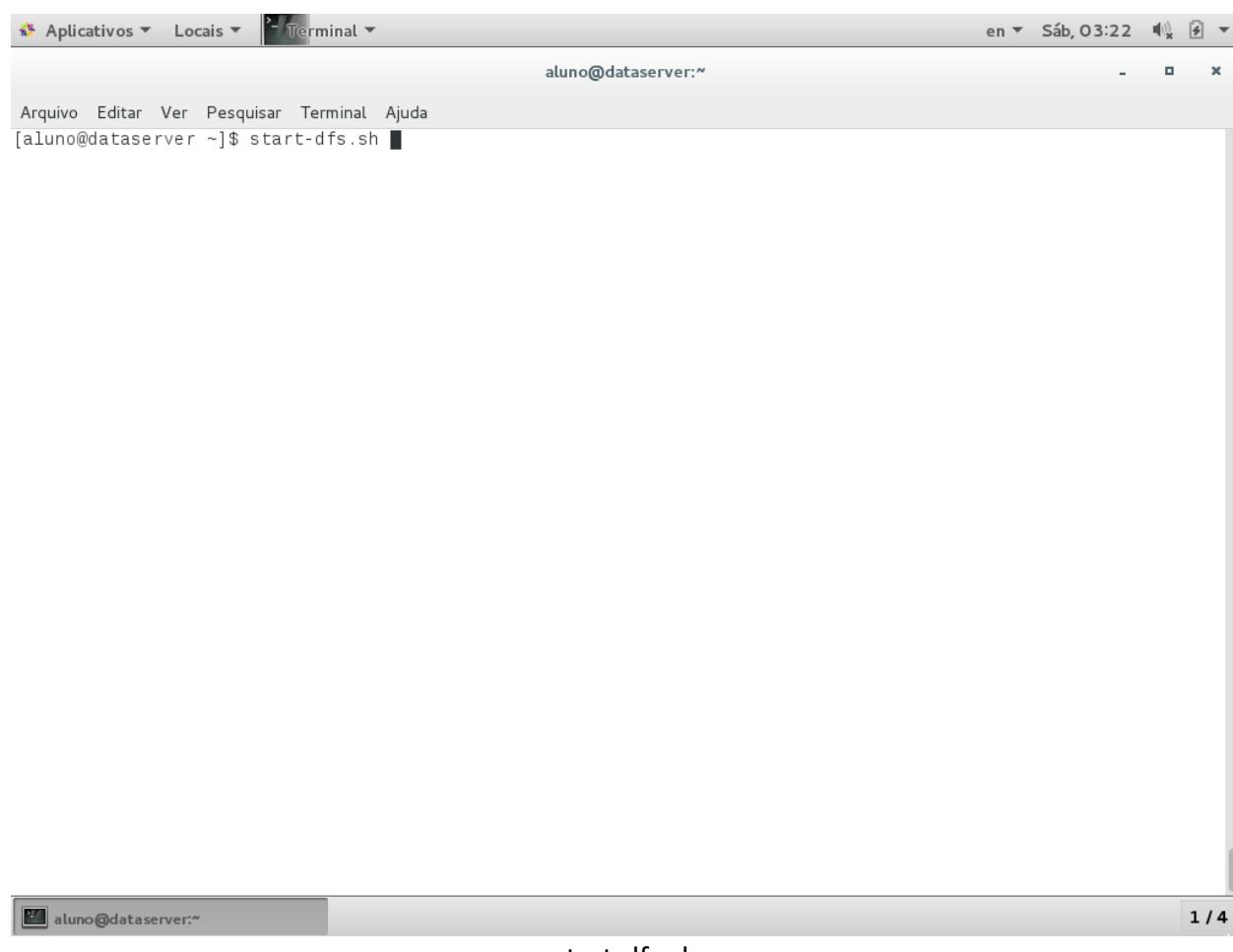
```

aluno@dataserver:~$ hdfs namenode -format
16/01/30 03:18:13 INFO namenode.FSNamesystem: supergroup          = supergroup
16/01/30 03:18:13 INFO namenode.FSNamesystem: isPermissionEnabled = true
16/01/30 03:18:13 INFO namenode.FSNamesystem: HA Enabled: false
16/01/30 03:18:13 INFO namenode.FSNamesystem: Append Enabled: true
16/01/30 03:18:14 INFO util.GSet: Computing capacity for map INodeMap
16/01/30 03:18:14 INFO util.GSet: VM type           = 64-bit
16/01/30 03:18:14 INFO util.GSet: 1.0% max memory 966.7 MB = 9.7 MB
16/01/30 03:18:14 INFO util.GSet: capacity        = 2^20 = 1048576 entries
16/01/30 03:18:14 INFO namenode.FSDirectory: ACLs enabled? false
16/01/30 03:18:14 INFO namenode.FSDirectory: XAttrs enabled? true
16/01/30 03:18:14 INFO namenode.FSDirectory: Maximum size of an xattr: 16384
16/01/30 03:18:14 INFO namenode.NameNode: Caching file names occurring more than 10 times
16/01/30 03:18:14 INFO util.GSet: Computing capacity for map cachedBlocks
16/01/30 03:18:14 INFO util.GSet: VM type           = 64-bit
16/01/30 03:18:14 INFO util.GSet: 0.25% max memory 966.7 MB = 2.4 MB
16/01/30 03:18:14 INFO util.GSet: capacity        = 2^18 = 262144 entries
16/01/30 03:18:14 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
16/01/30 03:18:14 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
16/01/30 03:18:14 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension   = 30000
16/01/30 03:18:14 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
16/01/30 03:18:14 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
16/01/30 03:18:14 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
16/01/30 03:18:14 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
16/01/30 03:18:14 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
16/01/30 03:18:14 INFO util.GSet: Computing capacity for map NameNodeRetryCache
16/01/30 03:18:14 INFO util.GSet: VM type           = 64-bit
16/01/30 03:18:14 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 297.0 KB
16/01/30 03:18:14 INFO util.GSet: capacity        = 2^15 = 32768 entries
16/01/30 03:18:14 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1791171697-127.0.0.1-1454131094108
16/01/30 03:18:14 INFO common.Storage: Storage directory /tmp/hadoop-aluno/dfs/name has been successfully formatted.
16/01/30 03:18:14 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
16/01/30 03:18:14 INFO util.ExitUtil: Exiting with status 0
16/01/30 03:18:14 INFO namenode.NameNode: SHUTDOWN_MSG:
*****SHUTDOWN_MSG: Shutting down NameNode at localhost/127.0.0.1
*****/[aluno@dataserver ~]$ 
```

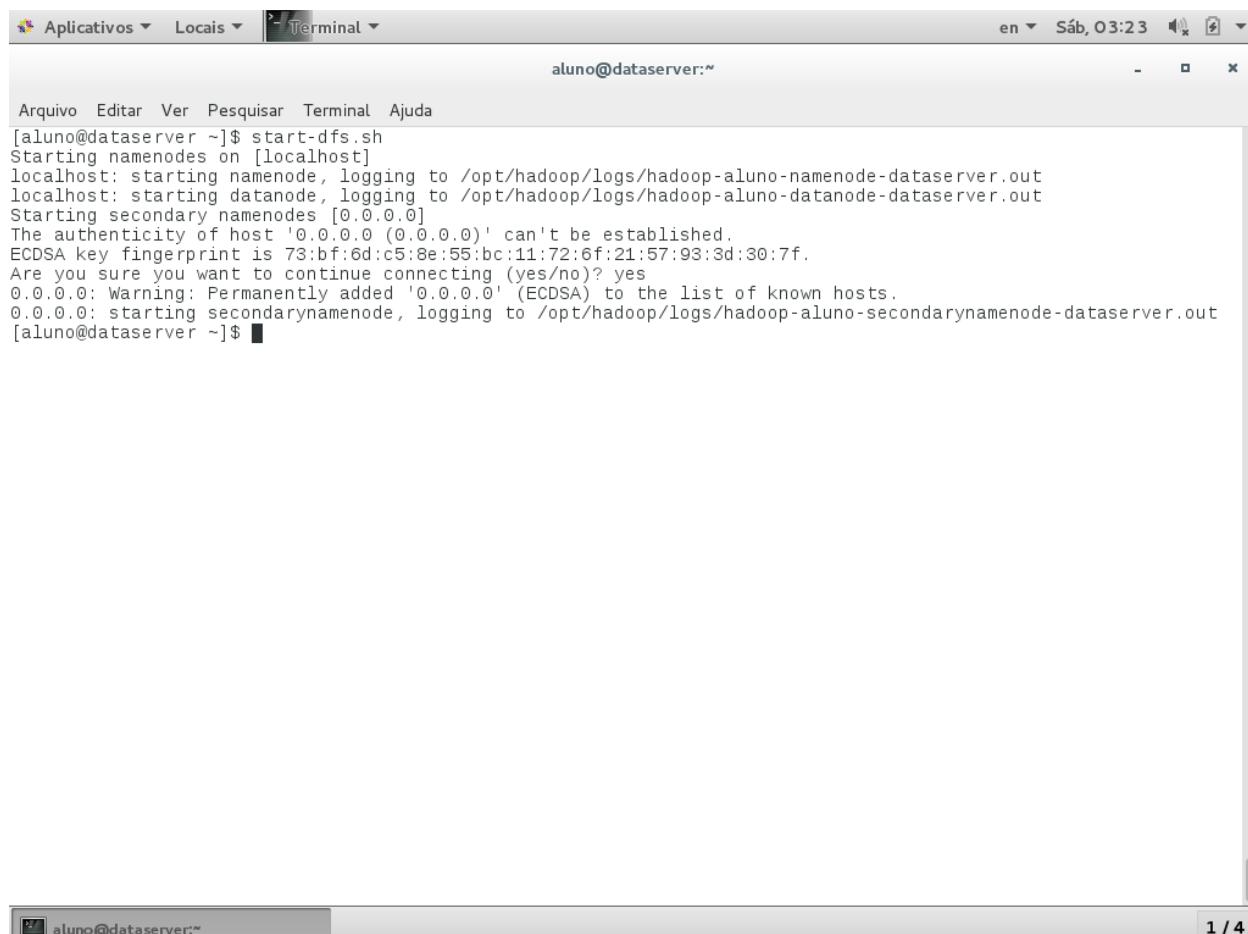
1 / 4

Formatação realizada com sucesso

5.4.3. Iniciando o Hadoop

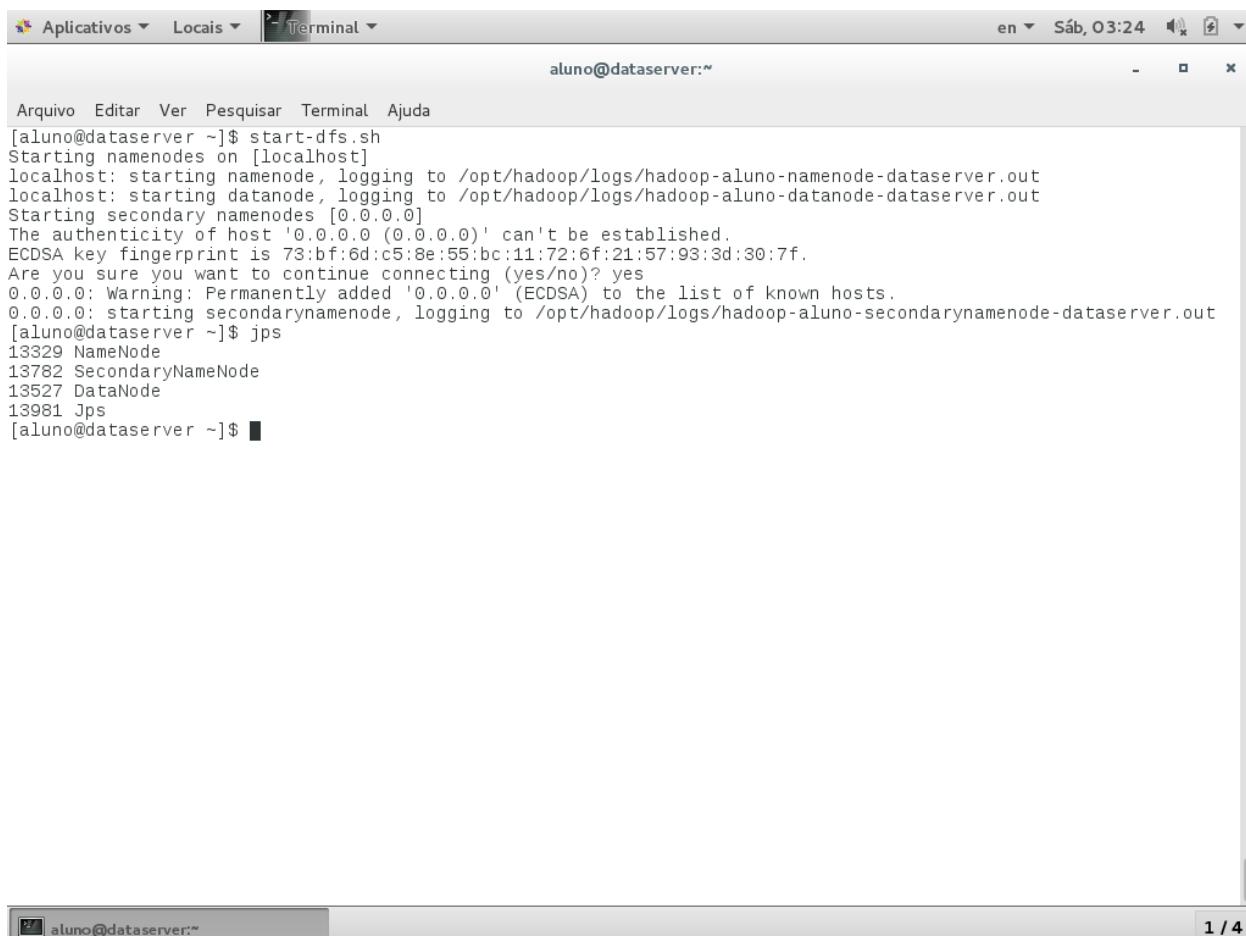


A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "aluno@dataserver:~" and the date/time "Sáb, 03:22". The menu bar includes "Aplicativos", "Locais", and "Terminal". Below the menu is a toolbar with icons for "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The terminal window contains the command "[aluno@dataserver ~]\$ start-dfs.sh" followed by a cursor. At the bottom of the window, there is a status bar with the text "start-dfs.sh" and a page number "1 / 4".



```
[aluno@dataserver ~]$ start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /opt/hadoop/logs/hadoop-aluno-namenode-dataserver.out
localhost: starting datanode, logging to /opt/hadoop/logs/hadoop-aluno-datanode-dataserver.out
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is 73:bf:6d:c5:8e:55:bc:11:72:6f:21:57:93:3d:30:7f.
Are you sure you want to continue connecting (yes/no)? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts.
0.0.0.0: starting secondarynamenode, logging to /opt/hadoop/logs/hadoop-aluno-secondarynamenode-dataserver.out
[aluno@dataserver ~]$ █
```

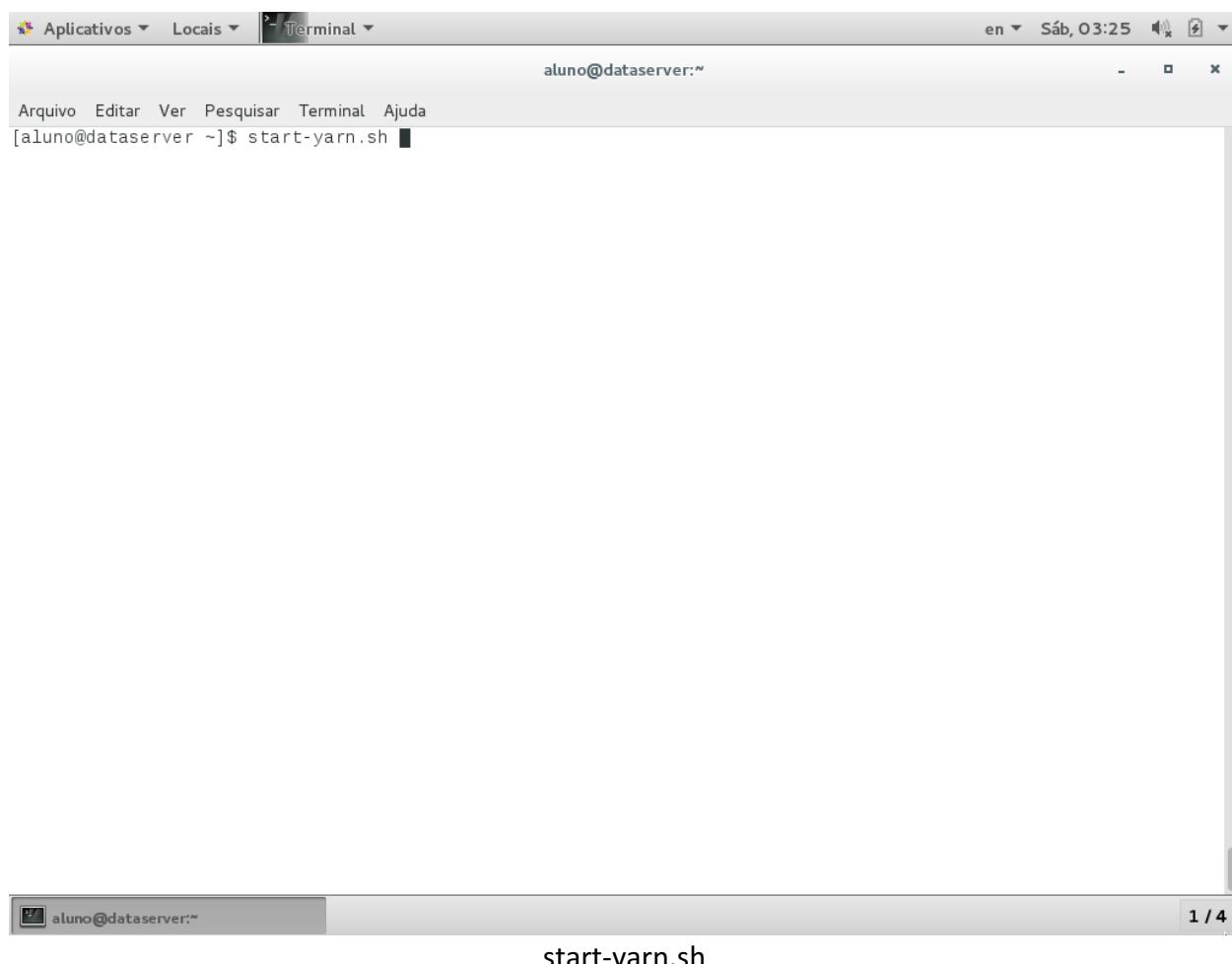
Hadoop iniciado

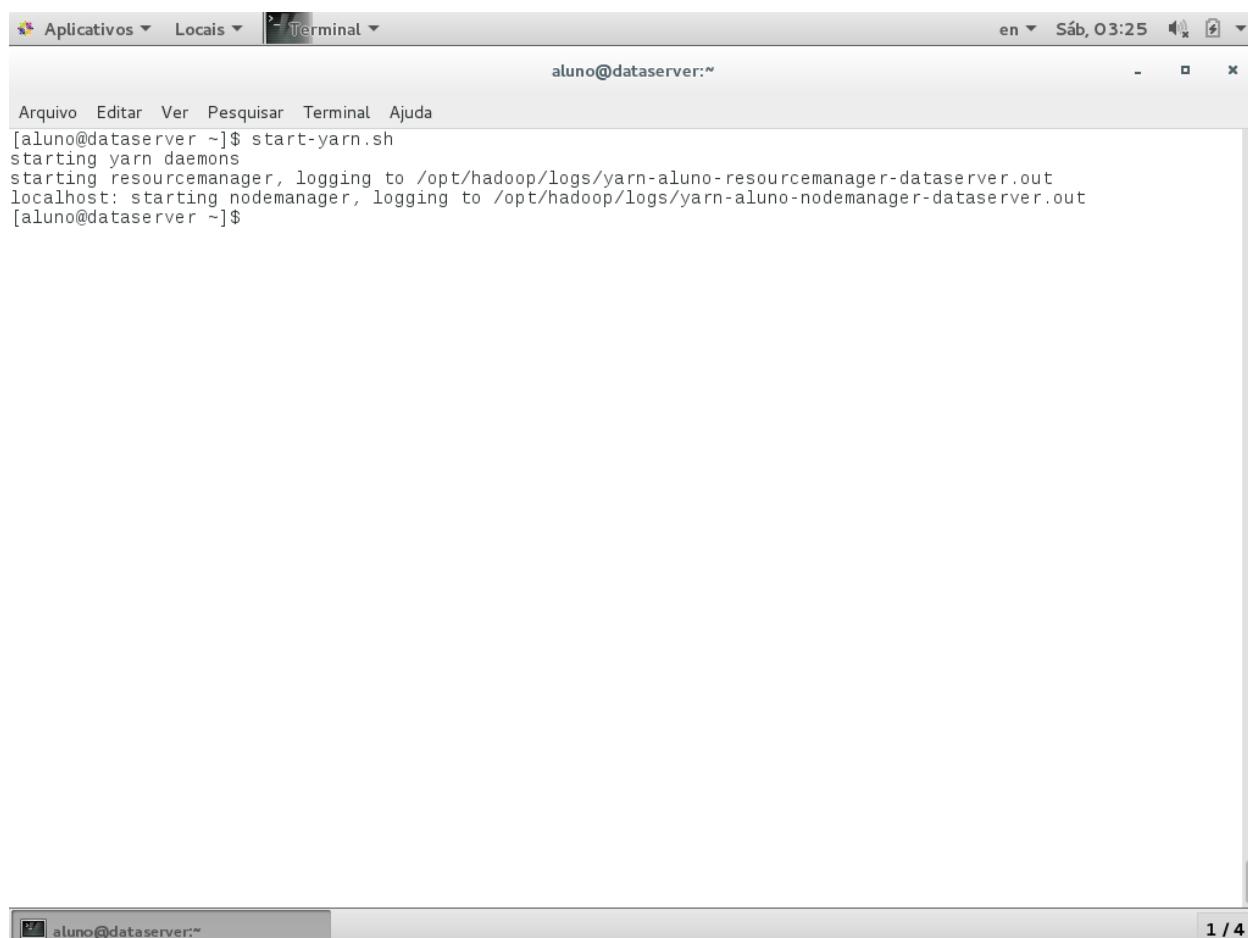


```
[aluno@dataserver ~]$ start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /opt/hadoop/logs/hadoop-aluno-namenode-dataserver.out
localhost: starting datanode, logging to /opt/hadoop/logs/hadoop-aluno-datanode-dataserver.out
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is 73:bf:6d:c5:8e:55:bc:11:72:6f:21:57:93:3d:30:7f.
Are you sure you want to continue connecting (yes/no)? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts.
0.0.0.0: starting secondarynamenode, logging to /opt/hadoop/logs/hadoop-aluno-secondarynamenode-dataserver.out
[aluno@dataserver ~]$ jps
13329 NameNode
13782 SecondaryNameNode
13527 DataNode
13981 Jps
[aluno@dataserver ~]$
```

Checando os serviços inicializados com o comando **jps**

5.4.4. Iniciando o Yarn





```
aluno@dataserver:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /opt/hadoop/logs/yarn-aluno-resourcemanager-dataserver.out
localhost: starting nodemanager, logging to /opt/hadoop/logs/yarn-aluno-nodemanager-dataserver.out
[aluno@dataserver ~]$
```

Yarn iniciado



A screenshot of a Linux terminal window titled "Terminal". The window shows the user "aluno@dataserver:~" at the prompt. The terminal displays the following command-line session:

```
[aluno@dataserver ~]$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /opt/hadoop/logs/yarn-aluno-resourcemanager-dataserver.out
localhost: starting nodemanager, logging to /opt/hadoop/logs/yarn-aluno-nodemanager-dataserver.out
[aluno@dataserver ~]$ jps
13329 NameNode
14037 ResourceManager
13782 SecondaryNameNode
13527 DataNode
14524 Jps
14189 NodeManager
[aluno@dataserver ~]$
```

The terminal window has a standard Linux desktop interface with icons for "Aplicativos" and "Locais" in the top bar, and a status bar showing "en" and "Sáb, 03:25". A scroll bar is visible on the right side of the terminal window.

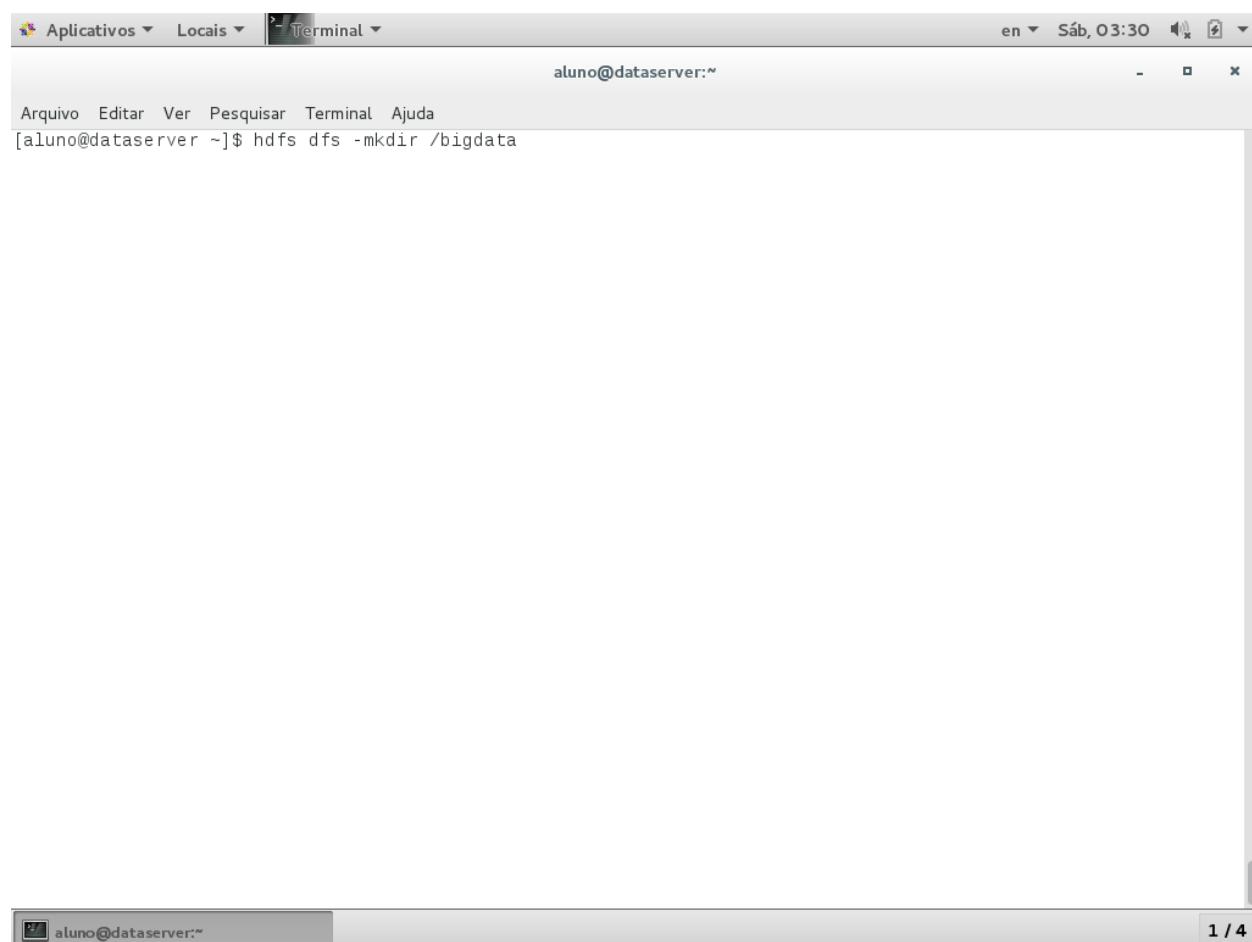
Checando os serviços com o comando **jps**

The screenshot shows the Hadoop Cluster Management UI running in Mozilla Firefox. The URL is <http://localhost:8088/cluster>. The interface includes a sidebar with links like 'About', 'Nodes', 'Node Labels', 'Applications' (with sub-options: NEW, NEW_SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED), and 'Scheduler'. The main area displays 'Cluster Metrics' and 'Scheduler Metrics'. Under 'Scheduler Metrics', it shows the 'Capacity Scheduler' with 'MEMORY' as the scheduling resource type and a minimum allocation of '<memory:1024, vCores:1>'. A table below shows scheduler entries, with a note 'No data available in table'. The status bar at the bottom indicates the user is at 'aluno@dataserver:/opt/hadoop/et...'.

Visualizando jobs – <http://localhost:8088>

1 / 4

5.5. Processando Big Data



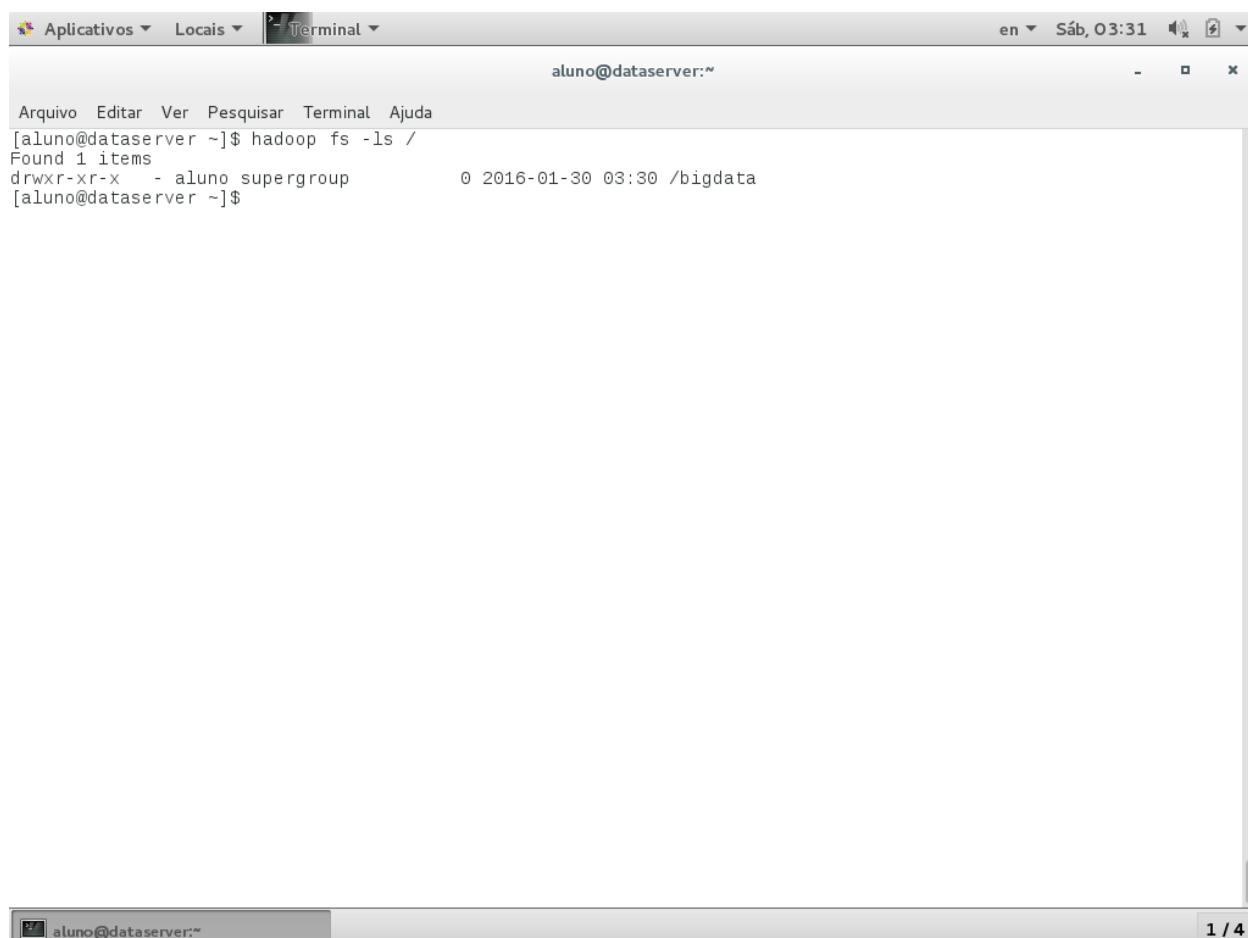
A screenshot of a Linux terminal window titled "Terminal". The window has a menu bar with "Aplicativos", "Locais", and "Terminal". The status bar shows "en" and "Sáb, 03:30". The terminal prompt is "aluno@dataserver:~". Below the prompt, the command "[aluno@dataserver ~]\$ hdfs dfs -mkdir /bigdata" is visible. The window is maximized.

aluno@dataserver:~

Arquivo Editar Ver Pesquisar Terminal Ajuda

[aluno@dataserver ~]\$ hdfs dfs -mkdir /bigdata

Criar o diretório **bigdata** no HDFS



```
aluno@dataserver:~$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - aluno supergroup          0 2016-01-30 03:30 /bigdata
[aluno@dataserver ~]$
```

Listar o HDFS e checar o diretório criado

1 / 4

Instalação e Configuração do Ecosistema Hadoop

The screenshot shows the homepage of the Brazilian Open Data Portal (dados.gov.br). The page is in Portuguese and features a green header with the portal's logo and navigation links for 'Participe', 'Serviços', 'Legislação', and 'Canais'. Below the header, there's a search bar and social media links. The main content area includes a search bar, a section titled 'Dados em destaque' with three items: 'Compras públicas do governo federal', 'Ocorrências Aeronáuticas na Aviação Civil Brasileira', and 'Lista de Eleitores Filiados aos Partidos Políticos'. To the right, there's a sidebar titled 'Publicações mais recentes' listing five datasets with their authors and publication dates.

Conjunto de dados	Autor	Quando
Aeroporto - 1º Balanço do PAC 2015	Sem Responsável	25 Jan
Cidades Digitais - 1º Balanço do PAC 2015	Sem Responsável	25 Jan
Centro de Iniciação ao Esporte - 1º ...	Sem Responsável	25 Jan
Base Cartográfica Contínua do Brasil – ...	Diretoria de ...	15 Dez
REGIÃO DE INFORMAÇÃO DE VOO - FIR	Divisão de Operações	15 Dez

Acessar o portal de dados abertos do governo federal

The screenshot shows the homepage of the Portal Brasileiro de Dados Abertos (dados.gov.br). At the top, there's a navigation bar with links for 'Aplicativos', 'Locais', and 'Navegador Web Firefox'. The main header says 'Bem vindo - Portal Brasileiro de Dados Abertos - Mozilla Firefox'. Below the header, there's a search bar with 'Pesquisar' and a menu bar with 'Participe', 'Serviços', 'Legislação', and 'Canais'. A green banner at the top features the text 'dados.gov.br' and 'PORTAL BRASILEIRO DE DADOS ABERTOS'. On the left, there's a sidebar with a 'Pesquisa ...' input field and a 'PESQUISAR' button. The main content area has a section titled 'Dados em destaque' with three items: 'Compras públicas do governo federal', 'Ocorrências Aeronáuticas na Aviação Civil Brasileira', and 'Lista de Eleitores Filiados aos Partidos Políticos'. To the right, there's a box titled 'Publicações mais recentes' listing five documents with their authors and dates:

Conjunto de dados	Autor	Quando
Aeroporto - 1º Balanço do PAC 2015	Sem Responsável	25 Jan
Cidades Digitais - 1º Balanço do PAC 2015	Sem Responsável	25 Jan
Centro de Iniciação ao Esporte - 1º ...	Sem Responsável	25 Jan
Base Cartográfica Contínua do Brasil - ...	Diretoria de ...	15 Dez
REGIÃO DE INFORMAÇÃO DE VOO - FIR	Divisão de Operações	15 Dez

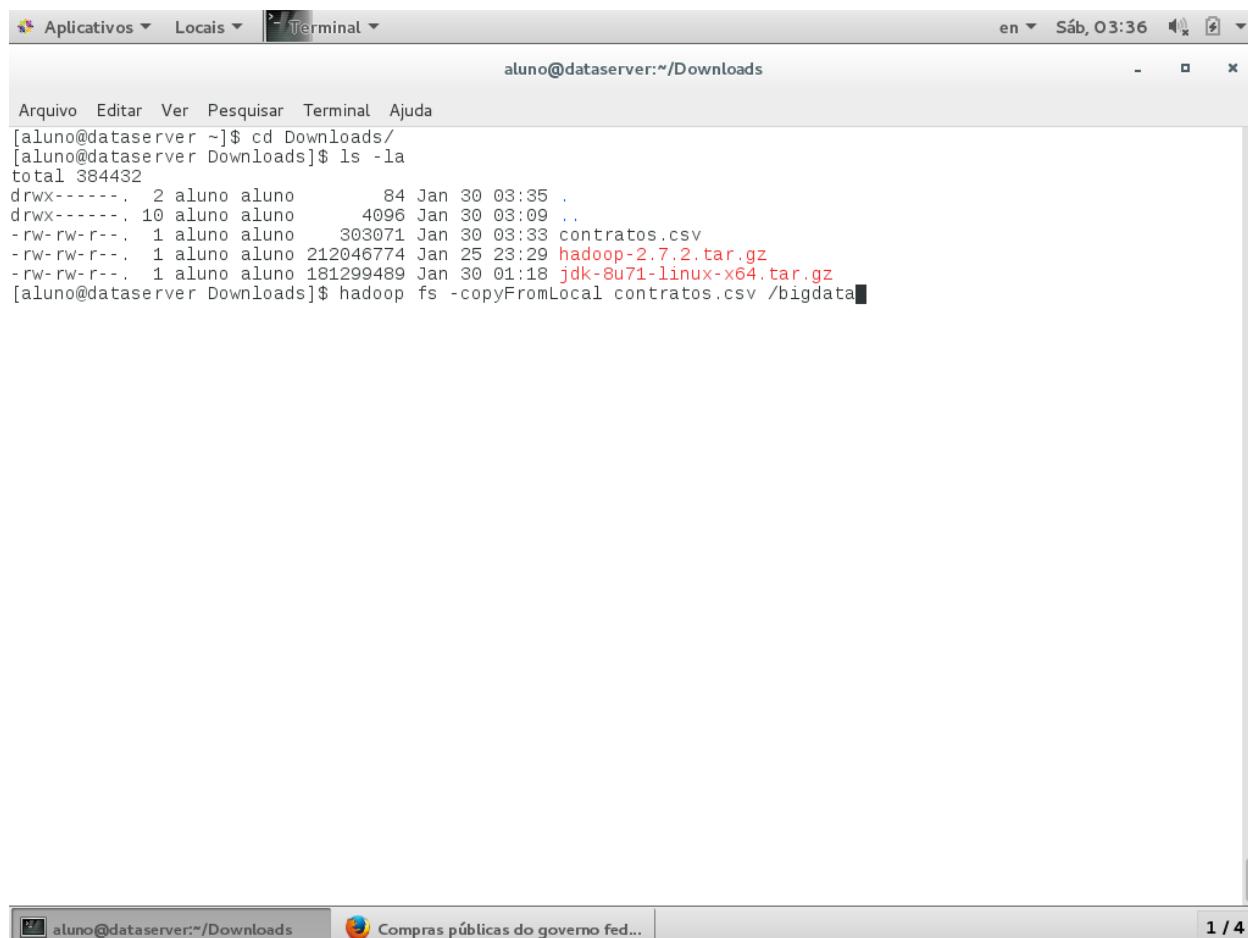
At the bottom, there's a footer with the text 'aluno@dataserver:~' and 'Bem vindo - Portal Brasileiro de D...'. On the far right, it says '1 / 4'.

Clicar no link de compras públicas

Instalação e Configuração do Ecosistema Hadoop

Baixar o arquivo de contratos em formato csv

Instalação e Configuração do Ecosistema Hadoop

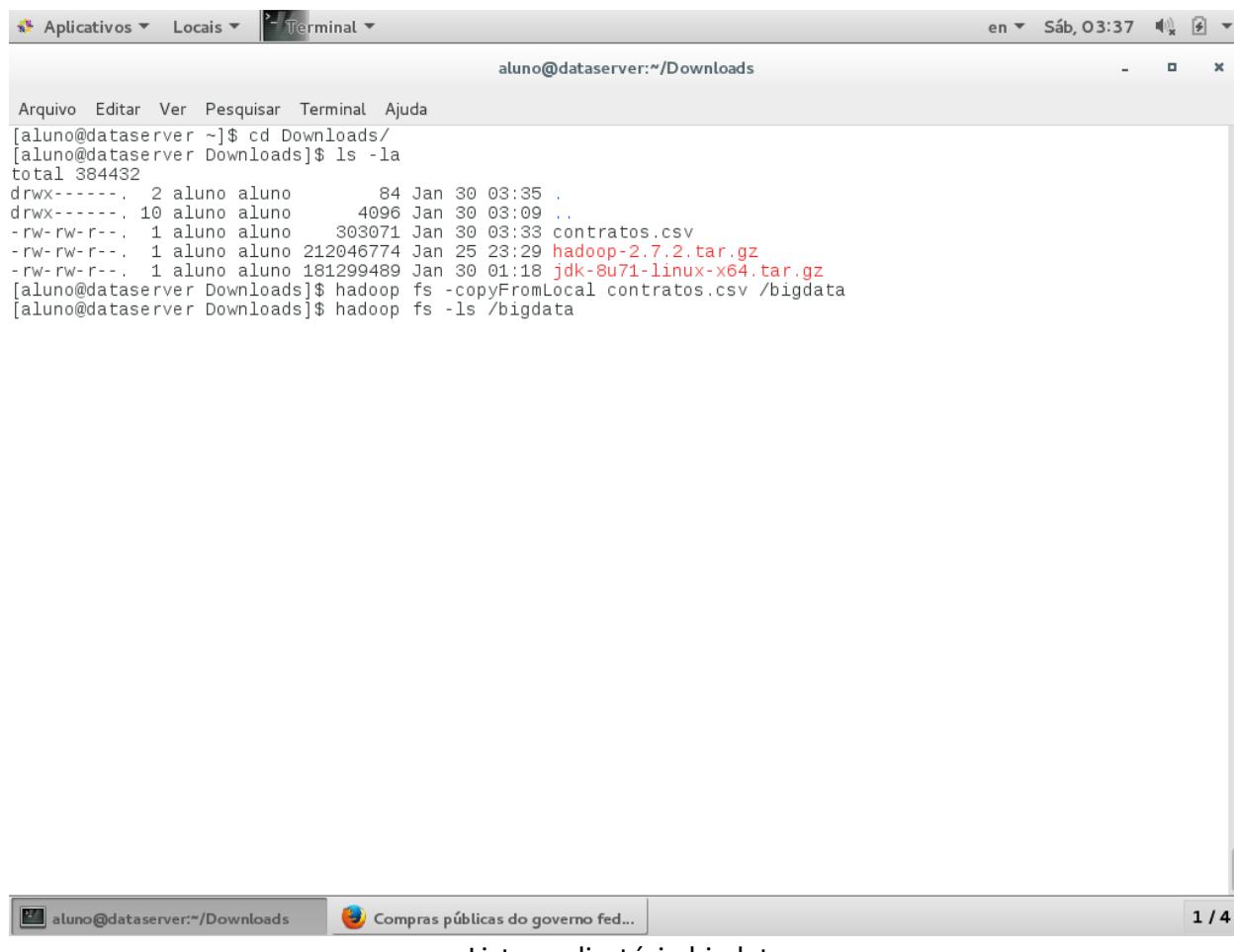


A screenshot of a Linux terminal window titled "Terminal". The window shows the user's command-line session:

```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 384432
drwx----- 2 aluno aluno 84 Jan 30 03:35 .
drwx----- 10 aluno aluno 4096 Jan 30 03:09 ..
-rw-rw-r-- 1 aluno aluno 303071 Jan 30 03:33 contratos.csv
-rw-rw-r-- 1 aluno aluno 212046774 Jan 25 23:29 hadoop-2.7.2.tar.gz
-rw-rw-r-- 1 aluno aluno 181299489 Jan 30 01:18 jdk-8u71-linux-x64.tar.gz
[aluno@dataserver Downloads]$ hadoop fs -copyFromLocal contratos.csv /bigdata
```

The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows "aluno@dataserver:/Downloads" and "1 / 4".

Copiar o arquivo para a pasta bigdata no HDFS



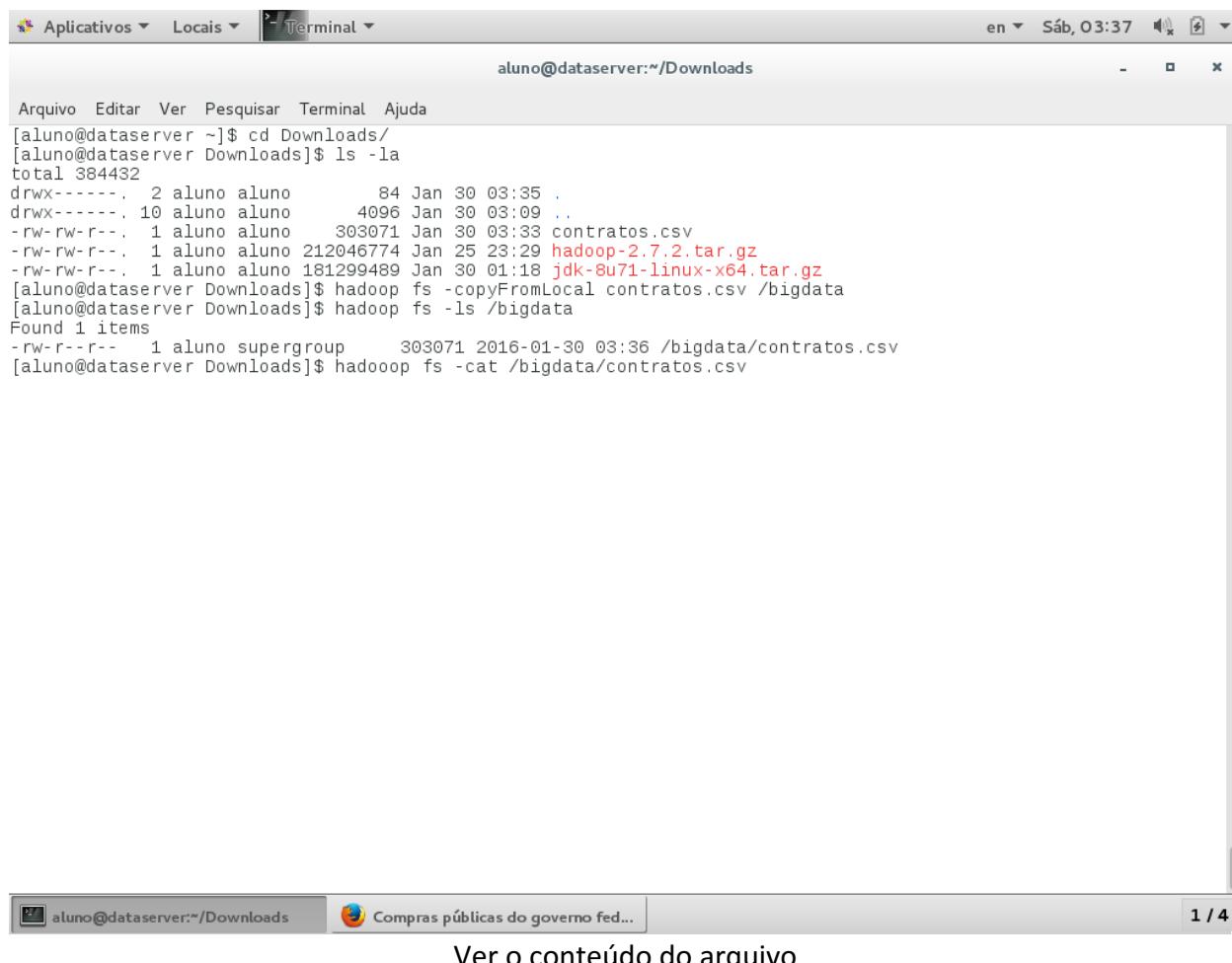
A screenshot of a Linux desktop environment showing a terminal window. The terminal title is "Terminal" and the user is "aluno@dataserver:/Downloads". The terminal content shows the following command-line session:

```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 384432
drwx----- 2 aluno aluno 84 Jan 30 03:35 .
drwx----- 10 aluno aluno 4096 Jan 30 03:09 ..
-rw-rw-r-- 1 aluno aluno 303071 Jan 30 03:33 contratos.csv
-rw-rw-r-- 1 aluno aluno 212046774 Jan 25 23:29 hadoop-2.7.2.tar.gz
-rw-rw-r-- 1 aluno aluno 181299489 Jan 30 01:18 jdk-8u71-linux-x64.tar.gz
[aluno@dataserver Downloads]$ hadoop fs -copyFromLocal contratos.csv /bigdata
[aluno@dataserver Downloads]$ hadoop fs -ls /bigdata
```

The terminal window has a standard Linux interface with a menu bar (Arquivo, Editar, Ver, Pesquisar, Terminal, Ajuda), system tray icons (en, Sáb, 03:37), and window control buttons (minimize, maximize, close).

Listar o diretório bigdata

Instalação e Configuração do Ecosistema Hadoop



The screenshot shows a terminal window titled "Terminal" with the command-line interface "aluno@dataserver:~/Downloads". The terminal displays the following sequence of commands and their output:

```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 384432
drwx----- 2 aluno aluno 84 Jan 30 03:35 .
drwx----- 10 aluno aluno 4096 Jan 30 03:09 ..
-rw-rw-r-- 1 aluno aluno 303071 Jan 30 03:33 contratos.csv
-rw-rw-r-- 1 aluno aluno 212046774 Jan 25 23:29 hadoop-2.7.2.tar.gz
-rw-rw-r-- 1 aluno aluno 181299489 Jan 30 01:18 jdk-8u71-linux-x64.tar.gz
[aluno@dataserver Downloads]$ hadoop fs -copyFromLocal contratos.csv /bigdata
[aluno@dataserver Downloads]$ hadoop fs -ls /bigdata
Found 1 items
-rw-r--r-- 1 aluno supergroup 303071 2016-01-30 03:36 /bigdata/contratos.csv
[aluno@dataserver Downloads]$ hadoop fs -cat /bigdata/contratos.csv
```

Ver o conteúdo do arquivo

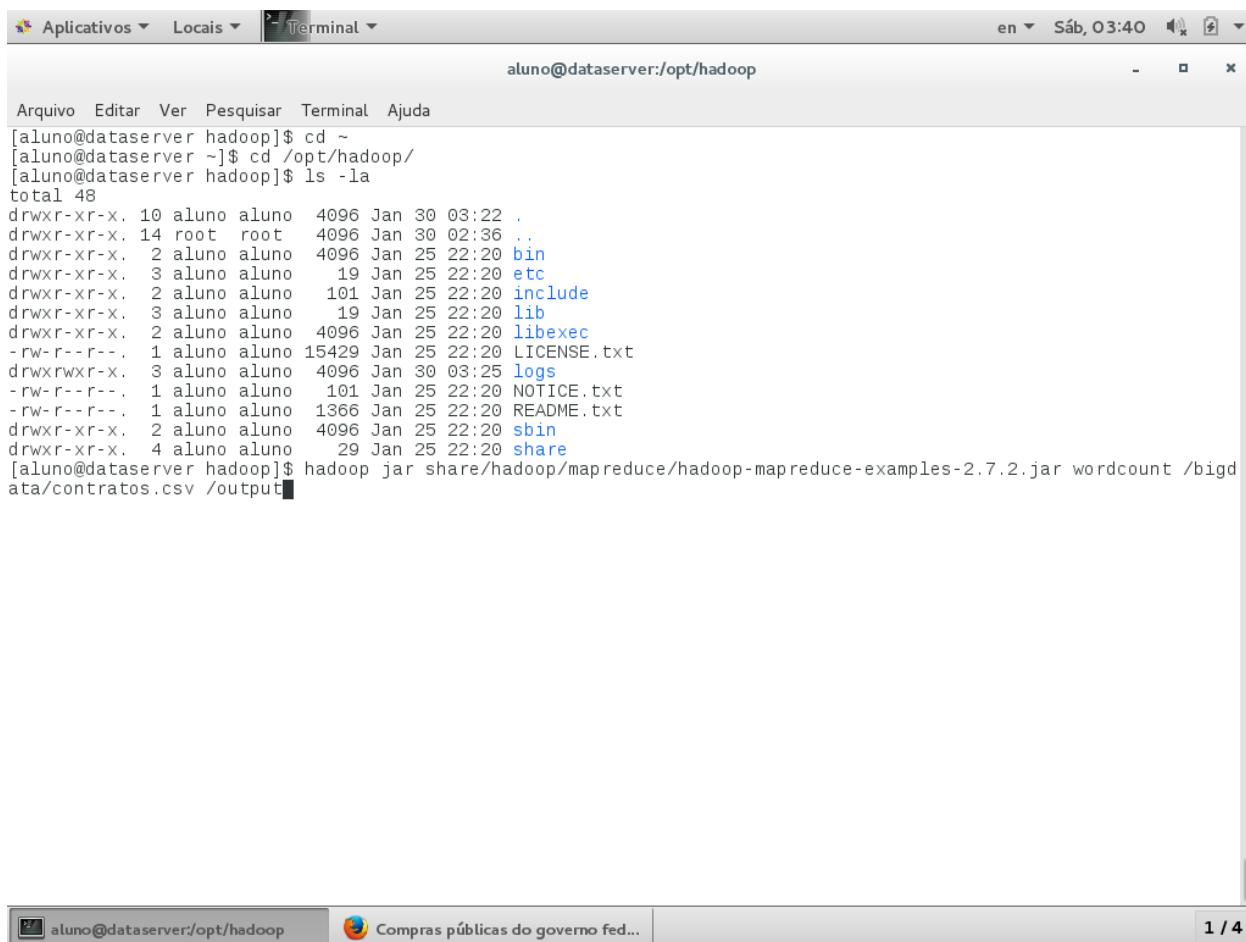
1 / 4

Instalação e Configuração do Ecosistema Hadoop

```
aluno@dataserver:~/Downloads
Arquivo Editar Ver Pesquisar Terminal Ajuda
25 distrito do departamento nacional de producao mineral no estado de alagoas.,1,48425000080/97,,Fornecedor 70.0
05.657/0001-27: DINAMICA SERVICOS GERAIS LTDA,03/11/1997,lei/8.666/93,,03/11/1997,02/11/1998,"R$ 9,708.96"/,cont
ratos/id/contrato/32302750000011997/aditivos,/contratos/id/contrato/32302750000011997/apostilamentos,/contratos/
id/contrato/32302750000011997/eventos
20006350000011997,200063: MJ-DPF-SUPERINTENDENCIA REGIONAL/RS,2: TOMADA DE PREÇOS,000019/1997,50: CONTRATO,Licit
ação 20006302000191997,,00001/1997,Contratação de empresa para realização de serviço de manutenção pre ventiva e
corretiva nos equipamentos de informática pertencentes a SR/DPF/R S e suas Delegacias descentralizadas,2,084300
11006/97-04,,Fornecedor 82.885.112/0001-31: VR COMPUTADORES LTDA,01/01/1998,"Art. 22, II da Lei 8.666/93",01/01/
1998,31/12/1998,"R$ 132,000.00"/,contratos/id/contrato/20006350000011997/aditivos,/contratos/id/contrato/2000635
0000011997/apostilamentos,/contratos/id/contrato/20006350000011997/eventos
20009257000011997,200092: SUPERINTENDENCIA REG.DEP.POLICIA FEDERAL- PE,,0,57: CONVÉNIO,Licitação 200092null00000
0000,,00001/1997,"Prorrogação da vigência do Convênio de Cooperação recíproca entre as partes conveniadas, visan
do o desenvolvimento de atividades conjuntas relacionadas ao estágio de estudantes",0,082000141519643,,24/05/1
999,"Decreto nº 87.497/82 e suas alterações; IN/SAF nº 07/92, alterada pelas IN/SAF nºs 01/93 e 06/94 e Lei nº 8.
666/93.",24/05/1999,23/05/2000,"R$ 150,000.00"/,contratos/id/contrato/20009257000011997/aditivos,/contratos/id/c
ontrato/20009257000011997/apostilamentos,/contratos/id/contrato/20009257000011997/eventos
15301054000011997,153010: MEC-CEFET-CENT.FED.ED.TEC.CELESO S.FONSECA/RJ,3: CONCORRÊNCIA,00003/1997,54: CONCESSÃO,
Licitação 15301003000031997,,00001/1997,"Concessão de uso para instalação de 12 (doze ) outdoors, no tamanho 3m
x 9m mediante remuneração na testada dos muros existentes não incluindo a ocupação interna.",5,23063001327/97-84
,,Fornecedor 29.248.390/0001-03: Klimes Rio Propaganda ao Ar LivreLtda,15/12/1997,lei 8987/85 e lei 8666/93 e su
as atualizações,15/12/1997,15/12/1998,"R$ 3,000.00"/,contratos/id/contrato/15301054000011997/aditivos,/contratos
/id/contrato/15301054000011997/apostilamentos,/contratos/id/contrato/15301054000011997/eventos
25442050000011997,254420: FUNDACAO OSWALDO CRUZ/RJ,4: CONCORRÊNCIA INTERNACIONAL,00008/1996,50: CONTRATO,Licitac
ão 25442004000081996,,00001/1997,Pretação de serviços de operação do Espaço Museu da Vida da Fiocruz,4,25380011
6889663,,Fornecedor 31.880.164/0001-84: HOPE-CONSULTORIA DE RECURSOSHUMANOS LTDA,14/01/1997,Artigo 62 da Lei 8
.666/93,14/01/1997,14/01/1998,"R$ 1,485.188.05"/,contratos/id/contrato/25442050000011997/aditivos,/contratos/id/
contrato/25442050000011997/apostilamentos,/contratos/id/contrato/25442050000011997/eventos
19311250000011997,193112: IBAMA-SUPERINTENDENCIA ESTADUAL/MS,6: DISPENSA DE LICITAÇÃO,0,00001081/1997,50: CONTRATO
,Licitação 19311206010811997,,00001/1997,"Locação de imóvel situado à Rua Paranaiba, 272, centro, Três Lagoas MS
, que a LOCADORA entregará ao LOCATÁRIO em perfeito estado de conservação e asseio, livre e desembaraçado de qua
lquer ônus judicial ou extrajudicial, para sua utilização.",4,02014001081/97-74,***546088***,01/08/1997,Inciso X
do Art. 24 da Lei 8.666/93,01/08/1997,31/07/1998,"R$ 8,400.00"/,contratos/id/contrato/19311250000011997/aditivo
s,/contratos/id/contrato/19311250000011997/apostilamentos,/contratos/id/contrato/19311250000011997/eventos
25003850000011997,250038: GERENCIA ESTADUAL EM SERGIPER/MS/SE,2: TOMADA DE FREÇOS,0,0001/1997,50: CONTRATO,Licitac
ão 25003802000011997,,00001/1997,Prestacao de servicos de fornecimento de passagens aereas domesticas,7,333591/
0009141/97,,Fornecedor 32.705.949/0001-83: PONTAL TURISMO LTDA,23/05/1997,"Lei 8666/93, alterada pela lei 8883/9
3.",23/05/1997,31/12/1997,"R$ 53,353.60"/,contratos/id/contrato/25003850000011997/aditivos,/contratos/id/contrat
o/25003850000011997/apostilamentos,/contratos/id/contrato/25003850000011997/eventos
[aluno@dataserver Downloads]$
```

Conteúdo do arquivo já gravado no HDFS

1 / 4

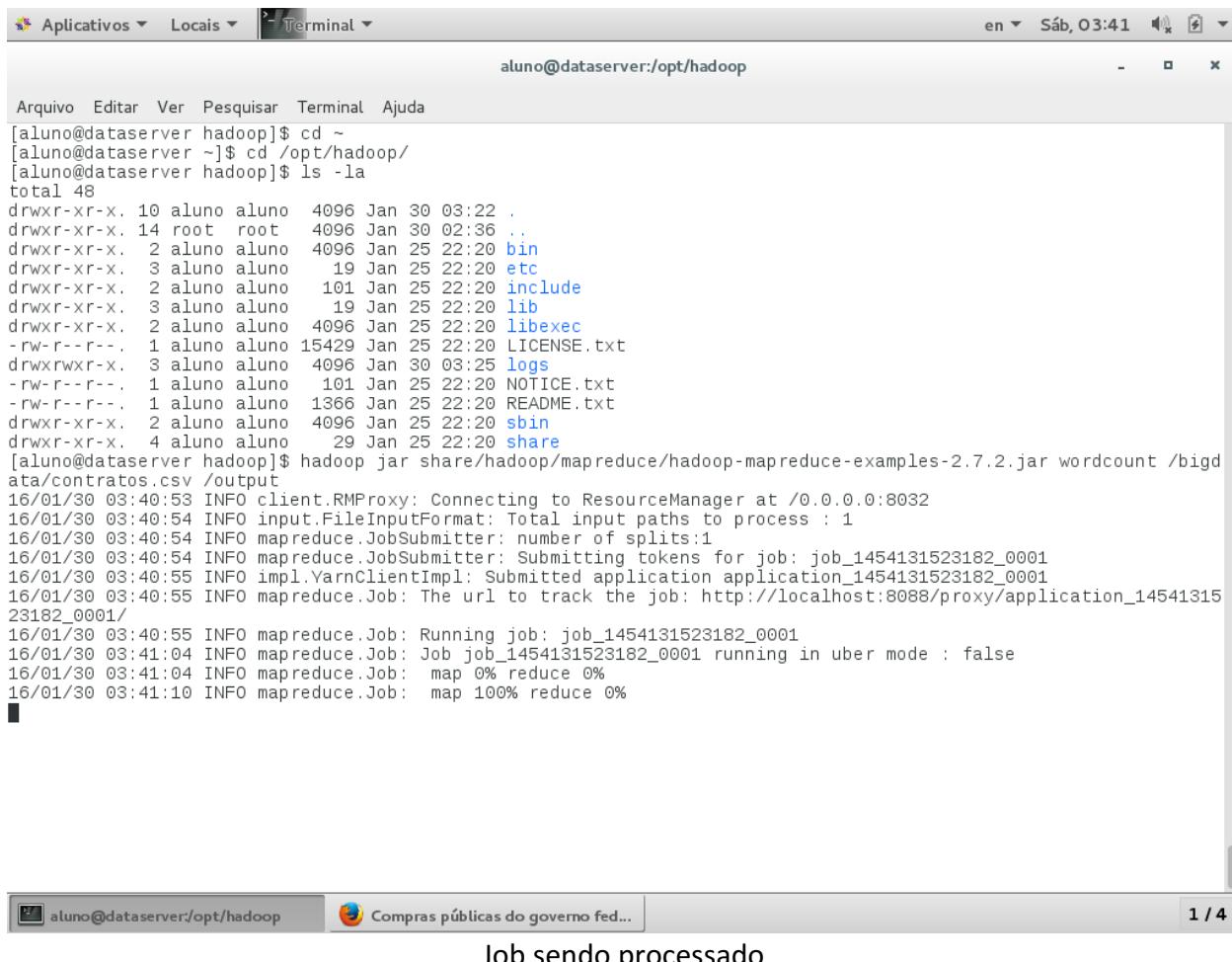


The screenshot shows a terminal window titled "Terminal" with the command prompt "aluno@dataserver:/opt/hadoop". The user has run several commands to navigate to the Hadoop directory and list its contents. They then executed a Hadoop job named "wordcount" on a file named "bigdata/contratos.csv". The output of the job is visible at the bottom of the terminal window.

```
[aluno@dataserver hadoop]$ cd ~  
[aluno@dataserver ~]$ cd /opt/hadoop/  
[aluno@dataserver hadoop]$ ls -la  
total 48  
drwxr-xr-x. 10 aluno aluno 4096 Jan 30 03:22 .  
drwxr-xr-x. 14 root root 4096 Jan 30 02:36 ..  
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 bin  
drwxr-xr-x. 3 aluno aluno 19 Jan 25 22:20 etc  
drwxr-xr-x. 2 aluno aluno 101 Jan 25 22:20 include  
drwxr-xr-x. 3 aluno aluno 19 Jan 25 22:20 lib  
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 libexec  
-rw-r--r--. 1 aluno aluno 15429 Jan 25 22:20 LICENSE.txt  
drwxrwxr-x. 3 aluno aluno 4096 Jan 30 03:25 logs  
-rw-r--r--. 1 aluno aluno 101 Jan 25 22:20 NOTICE.txt  
-rw-r--r--. 1 aluno aluno 1366 Jan 25 22:20 README.txt  
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 sbin  
drwxr-xr-x. 4 aluno aluno 29 Jan 25 22:20 share  
[aluno@dataserver hadoop]$ hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount /bigdata/contratos.csv /output
```

A instalação do Hadoop possui um job chamado wordcount, que pode ser usado como exemplo para processamento de Big Data. Basicamente, o job conta a ocorrência de cada palavra no arquivo. Vamos executar com o comando acima.

Instalação e Configuração do Ecosistema Hadoop

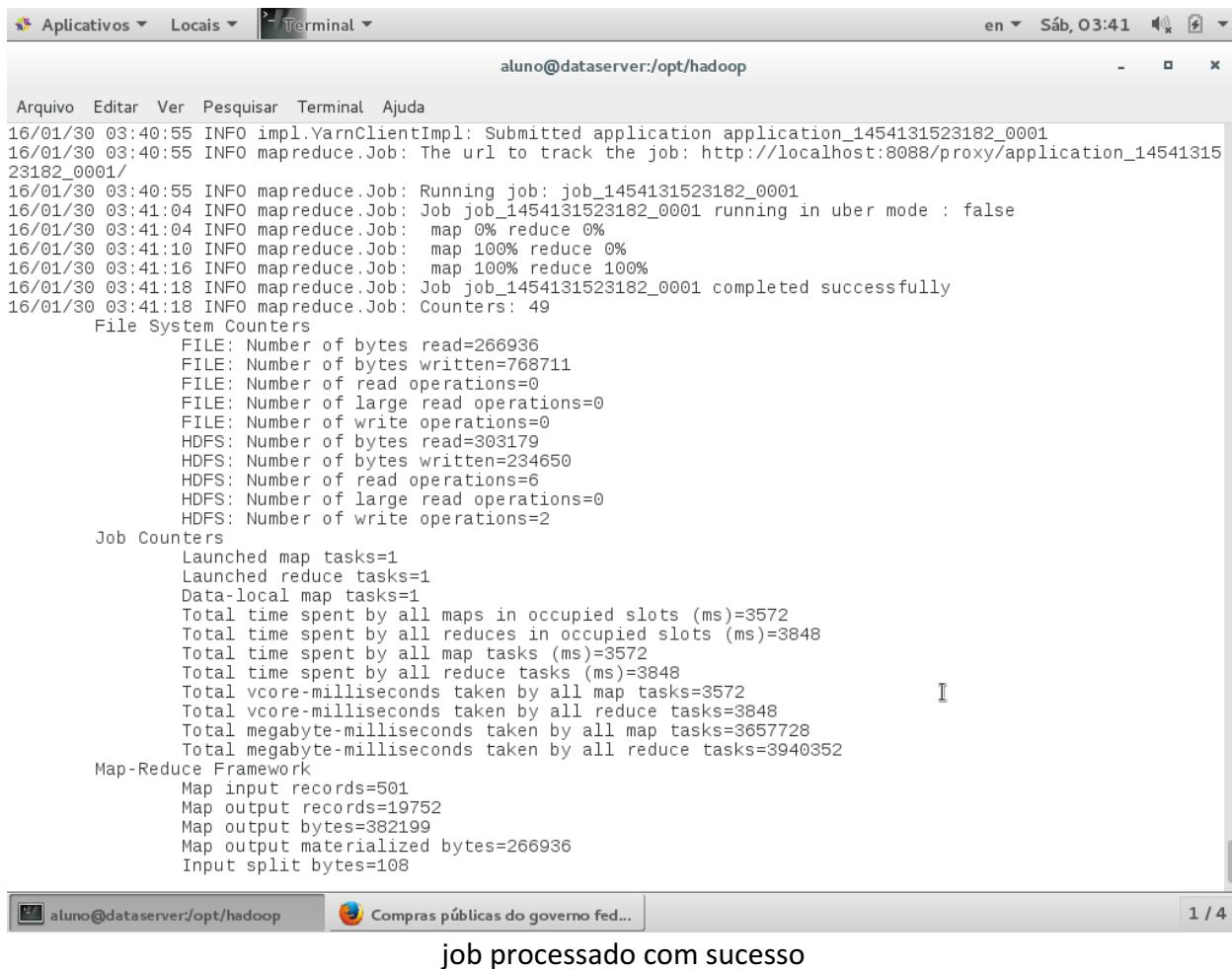


The screenshot shows a terminal window titled "Terminal" with the command prompt "aluno@dataserver:/opt/hadoop". The terminal displays the following command-line session:

```
[aluno@dataserver hadoop]$ cd ~  
[aluno@dataserver ~]$ cd /opt/hadoop/  
[aluno@dataserver hadoop]$ ls -la  
total 48  
drwxr-xr-x. 10 aluno aluno 4096 Jan 30 03:22 .  
drwxr-xr-x. 14 root root 4096 Jan 30 02:36 ..  
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 bin  
drwxr-xr-x. 3 aluno aluno 19 Jan 25 22:20 etc  
drwxr-xr-x. 2 aluno aluno 101 Jan 25 22:20 include  
drwxr-xr-x. 3 aluno aluno 19 Jan 25 22:20 lib  
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 libexec  
-rw-r--r--. 1 aluno aluno 15429 Jan 25 22:20 LICENSE.txt  
drwxrwxr-x. 3 aluno aluno 4096 Jan 30 03:25 logs  
-rw-r--r--. 1 aluno aluno 101 Jan 25 22:20 NOTICE.txt  
-rw-r--r--. 1 aluno aluno 1366 Jan 25 22:20 README.txt  
drwxr-xr-x. 2 aluno aluno 4096 Jan 25 22:20 sbin  
drwxr-xr-x. 4 aluno aluno 29 Jan 25 22:20 share  
[aluno@dataserver hadoop]$ hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount /bigdata/contratos.csv /output  
16/01/30 03:40:53 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
16/01/30 03:40:54 INFO input.FileInputFormat: Total input paths to process : 1  
16/01/30 03:40:54 INFO mapreduce.JobSubmitter: number of splits:1  
16/01/30 03:40:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1454131523182_0001  
16/01/30 03:40:55 INFO impl.YarnClientImpl: Submitted application application_1454131523182_0001  
16/01/30 03:40:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1454131523182_0001/  
16/01/30 03:40:55 INFO mapreduce.Job: Running job: job_1454131523182_0001  
16/01/30 03:41:04 INFO mapreduce.Job: Job job_1454131523182_0001 running in uber mode : false  
16/01/30 03:41:04 INFO mapreduce.Job: map 0% reduce 0%  
16/01/30 03:41:10 INFO mapreduce.Job: map 100% reduce 0%
```

The terminal window has a title bar with "Aplicativos", "Locais", "Terminal", "en", "Sáb, 03:41", and a volume icon. The status bar at the bottom shows "aluno@dataserver:/opt/hadoop" and "1 / 4".

Job sendo processado

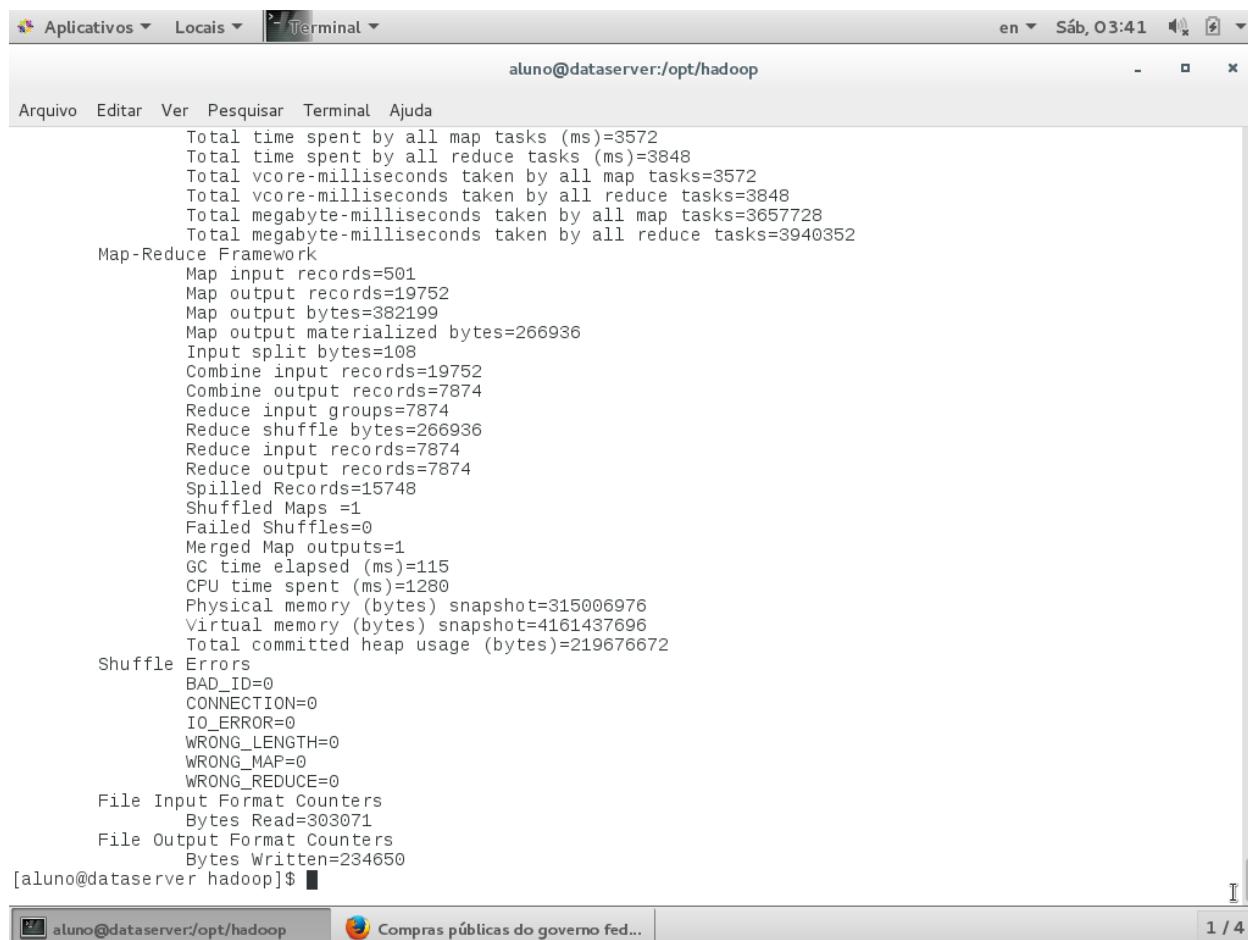


A screenshot of a terminal window titled "Terminal". The window shows command-line output from a Hadoop job. The logs indicate the job was submitted, tracked, and completed successfully. It details various metrics such as file operations, map tasks, reduce tasks, and Map-Reduce Framework statistics.

```
aluno@dataserver:/opt/hadoop
16/01/30 03:40:55 INFO impl.YarnClientImpl: Submitted application application_1454131523182_0001
16/01/30 03:40:55 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1454131523182_0001/
16/01/30 03:40:55 INFO mapreduce.Job: Running job: job_1454131523182_0001
16/01/30 03:41:04 INFO mapreduce.Job: Job job_1454131523182_0001 running in uber mode : false
16/01/30 03:41:04 INFO mapreduce.Job: map 0% reduce 0%
16/01/30 03:41:10 INFO mapreduce.Job: map 100% reduce 0%
16/01/30 03:41:16 INFO mapreduce.Job: map 100% reduce 100%
16/01/30 03:41:18 INFO mapreduce.Job: Job job_1454131523182_0001 completed successfully
16/01/30 03:41:18 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=266936
    FILE: Number of bytes written=768711
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=303179
    HDFS: Number of bytes written=234650
    HDFS: Number of read operations=6
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=3572
    Total time spent by all reduces in occupied slots (ms)=3848
    Total time spent by all map tasks (ms)=3572
    Total time spent by all reduce tasks (ms)=3848
    Total vcore-milliseconds taken by all map tasks=3572
    Total vcore-milliseconds taken by all reduce tasks=3848
    Total megabyte-milliseconds taken by all map tasks=3657728
    Total megabyte-milliseconds taken by all reduce tasks=3940352
  Map-Reduce Framework
    Map input records=501
    Map output records=19752
    Map output bytes=382199
    Map output materialized bytes=266936
    Input split bytes=108
```

job processado com sucesso

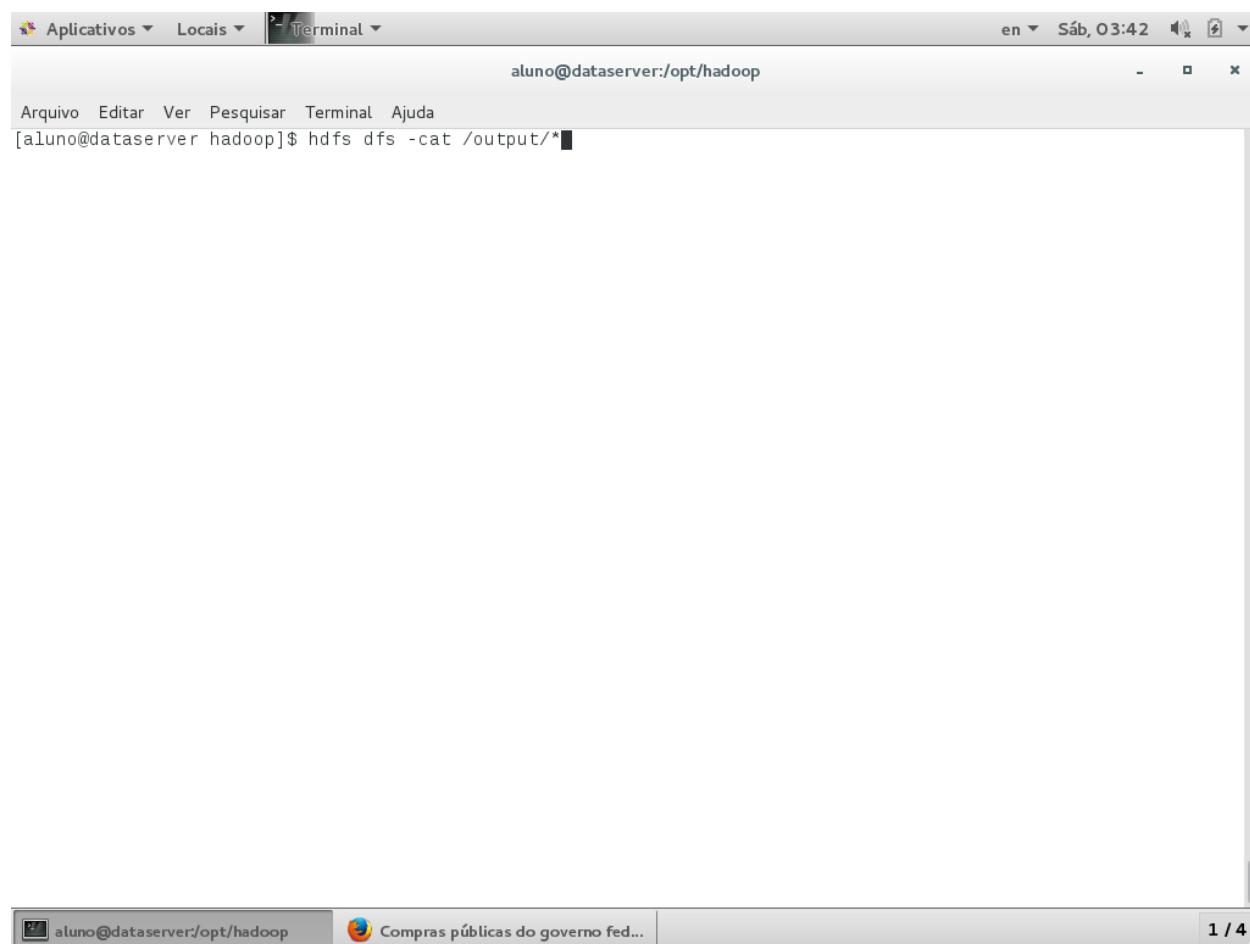
1 / 4



A screenshot of a terminal window titled "Terminal". The window shows command-line output from a Hadoop job. The output includes various performance metrics such as total time spent by map and reduce tasks, map input and output records, shuffle errors, and file format counters. The terminal window has a standard Linux-style interface with tabs for "Aplicativos", "Locais", and "Terminal". The status bar at the top right shows the user "aluno@dataserver:/opt/hadoop" and the date/time "Sáb, 03:41". The bottom of the window shows the URL "aluno@dataserver:/opt/hadoop" and a progress bar indicating the slide is 1/4.

```
Total time spent by all map tasks (ms)=3572
Total time spent by all reduce tasks (ms)=3848
Total vcore-milliseconds taken by all map tasks=3572
Total vcore-milliseconds taken by all reduce tasks=3848
Total megabyte-milliseconds taken by all map tasks=3657728
Total megabyte-milliseconds taken by all reduce tasks=3940352
Map-Reduce Framework
  Map input records=501
  Map output records=19752
  Map output bytes=382199
  Map output materialized bytes=266936
  Input split bytes=108
  Combine input records=19752
  Combine output records=7874
  Reduce input groups=7874
  Reduce shuffle bytes=266936
  Reduce input records=7874
  Reduce output records=7874
  Spilled Records=15748
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=115
  CPU time spent (ms)=1280
  Physical memory (bytes) snapshot=315006976
  Virtual memory (bytes) snapshot=4161437696
  Total committed heap usage (bytes)=219676672
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=303071
File Output Format Counters
  Bytes Written=234650
[aluno@dataserver hadoop]$
```

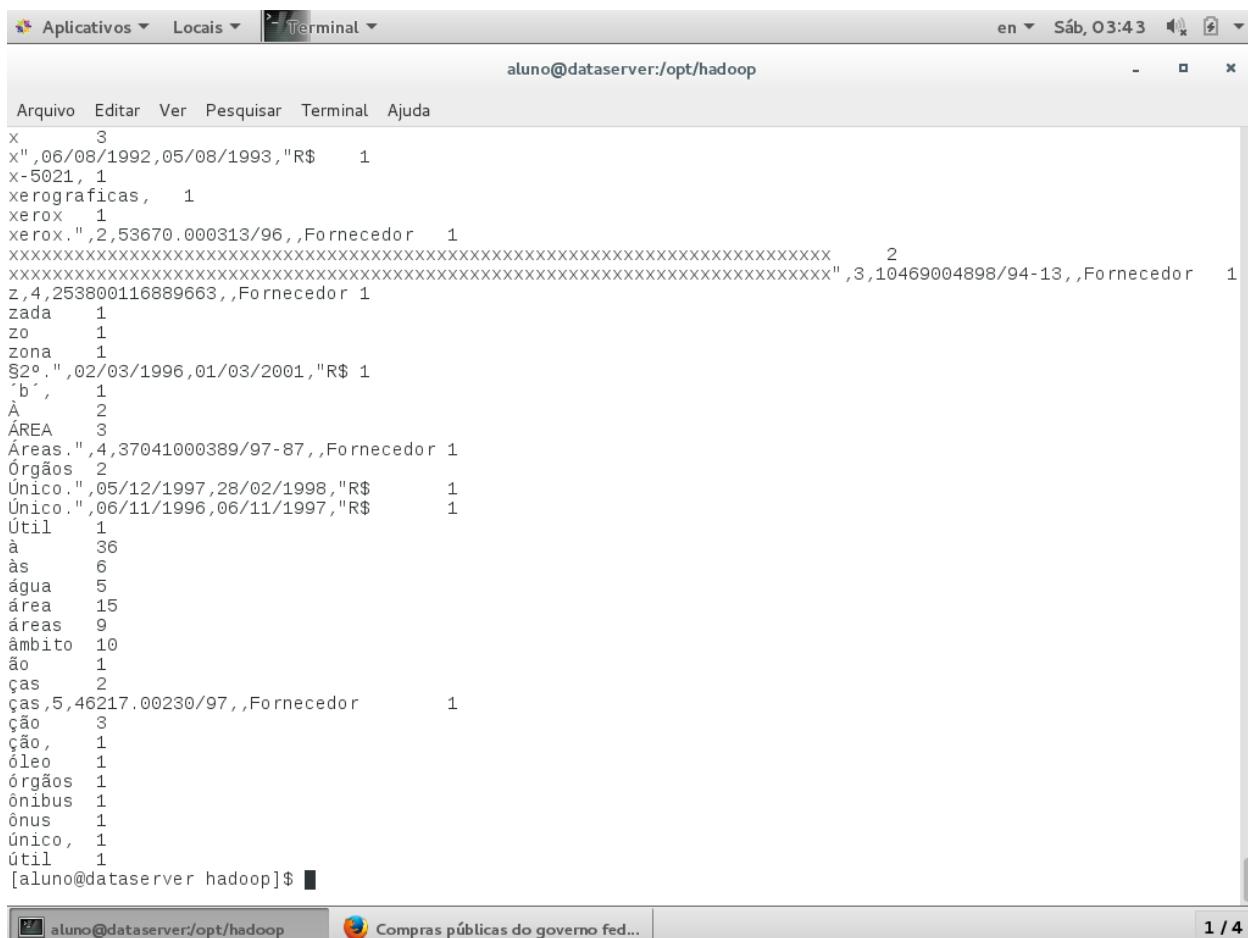
job processado com sucesso



A screenshot of a Linux desktop environment showing a terminal window. The terminal title is "Terminal" and the user is connected to "aluno@dataserver:/opt/hadoop". The menu bar includes "Aplicativos", "Locais", and "Terminal". The system tray shows "en", "Sáb, 03:42", and other icons. The terminal window contains the command "[aluno@dataserver hadoop]\$ hdfs dfs -cat /output/*" followed by a cursor. Below the terminal is a dock with several icons, including one for a browser titled "Compras públicas do governo fed...". A status bar at the bottom shows "1 / 4".

Vamos ver o resultado do processamento

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a terminal window titled "Terminal". The window shows the command "aluno@dataserver:/opt/hadoop" at the top right. The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The terminal content displays a word frequency count for a processed file. The output is as follows:

```
x 3
x", 06/08/1992,05/08/1993,"R$ 1
x-5021, 1
xerograficas, 1
xerox 1
xerox.", 2,53670.000313/96,,Fornecedor 1
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx", 2
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx", 3,10469004898/94-13,,Fornecedor 1
z,4,253800116889663,,Fornecedor 1
zada 1
zo 1
zona 1
$2º.", 02/03/1996,01/03/2001,"R$ 1
'b', 1
ÁREA 3
Áreas.", 4,37041000389/97-87,,Fornecedor 1
Órgãos 2
Único.", 05/12/1997,28/02/1998,"R$ 1
Único.", 06/11/1996,06/11/1997,"R$ 1
Útil 1
à 36
às 6
água 5
área 15
áreas 9
âmbito 10
ão 1
ças 2
cas, 5,46217.00230/97,,Fornecedor 1
ção 3
ção, 1
óleo 1
órgãos 1
ônibus 1
ônus 1
único, 1
útil 1
[aluno@dataserver hadoop]$
```

Arquivo processado. Número de ocorrência de cada palavra/termo no arquivo.

The screenshot shows a Mozilla Firefox browser window with the title "Namenode information - Mozilla Firefox". The address bar shows the URL "dataserver:50070/dfshealth.html#tab-overview". The main content area is titled "Overview 'localhost:9000' (active)". It contains a table with the following configuration details:

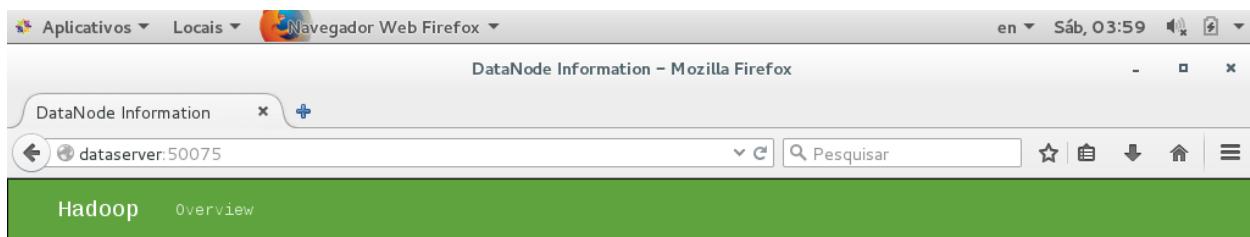
Started:	Sat Jan 30 03:22:42 BRST 2016
Version:	2.7.2, rb165c4fe8a74265c792ce23f546c64604acf0e41
Compiled:	2016-01-26T00:08Z by jenkins from (detached from b165c4f)
Cluster ID:	CID-45278722-9d78-4cc1-a53a-03d72c15a266
Block Pool ID:	BP-1791171697-127.0.0.1-1454131094108

Below the table, there is a summary section with the following status information:

- Security is off.
- Safemode is off.
- 17 files and directories, 5 blocks = 22 total filesystem object(s).

The bottom of the browser window shows the user "aluno@dataserver:/opt/hadoop" and the tab title "Namenode information - Mozilla F...". A navigation bar at the bottom right indicates "1 / 4".

Acesso ao Hadoop pelo browser: <http://dataserver:50070>



DataNode on dataserver:50075

Hadoop, 2015.



Overview

Version	2.7.2
Compiled	2016-01-26T00:08Z by jenkins from (detached from b165c4f)
NameNode Address	localhost:9000
Started	30/01/2016 03:22:54
Last Checkpoint	31/12/1969 22:00:17
Checkpoint Period	3600 seconds
Checkpoint Transactions	1000000

Checkpoint Image URI

- file:///tmp/hadoop-aluno/dfs/namesecondary

Acesso ao Hadoop pelo browser: <http://dataserver:50090>

Terceiro checkpoint:

Clique no meu File – Export Appliance.
Será gerada uma cópia de segurança da sua máquina virtual.

→ VM: DataServer-3.0.ova (Hadoop)

6. Instalação e Configuração do Zookeeper

6.1. Download e Instalação do Zookeeper

The Apache ZooKeeper system for distributed coordination is a high-performance service for building distributed applications.

- [Download](#)
- [Release Notes](#)
- [News](#)

Download

Releases may be downloaded from Apache mirrors: [Download](#)

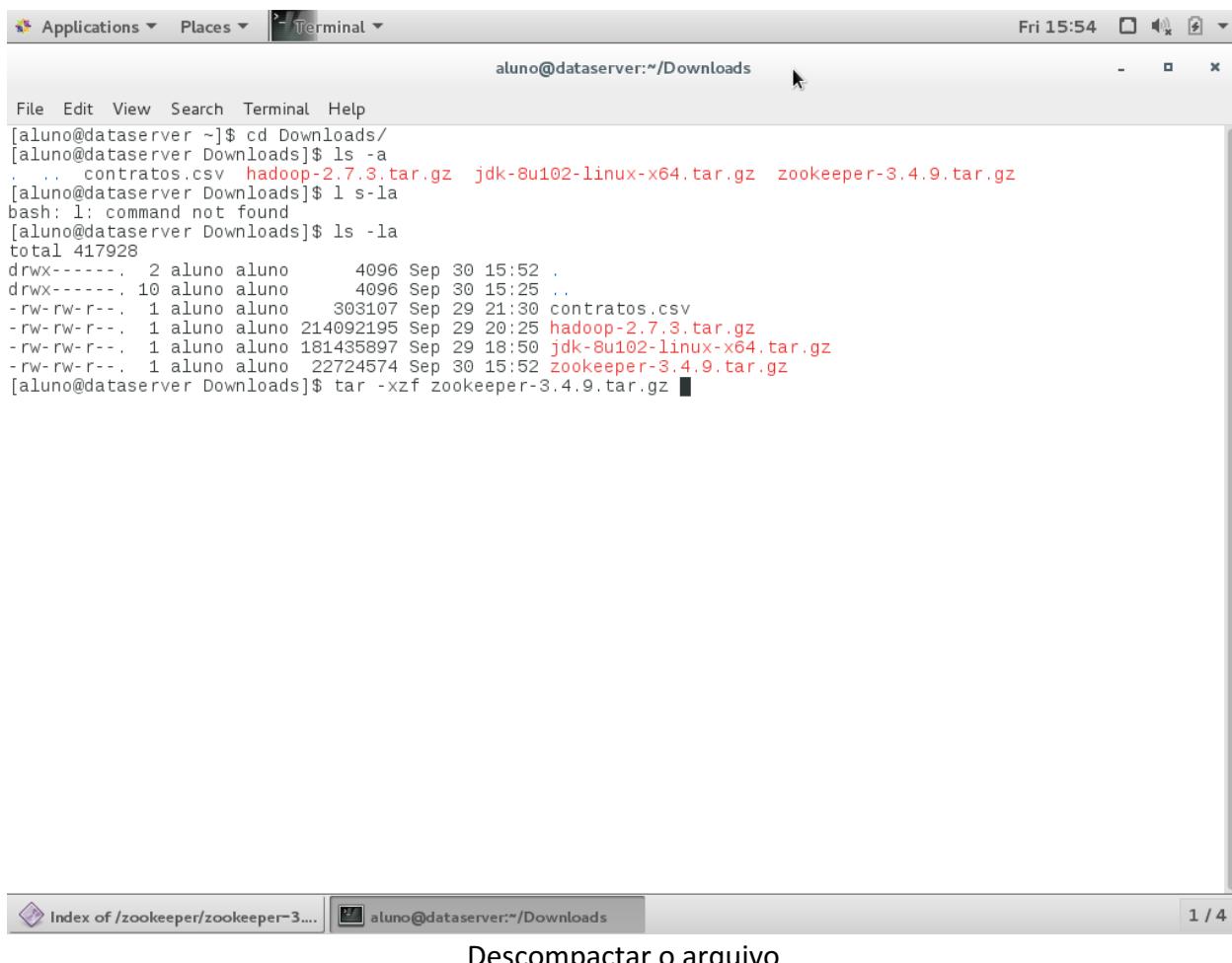
On the mirror, all recent releases are available, but are not guaranteed to be stable. For stable releases, look in the stable directory.

You can verify the integrity of a downloaded release using the PGP signatures and hashes (MD5 or SHA1) hosted at the main [Apache distro site](#). For additional information, refer to the Apache documentation for [verifying the integrity of Apache project releases](#).

Release Notes

1 / 4

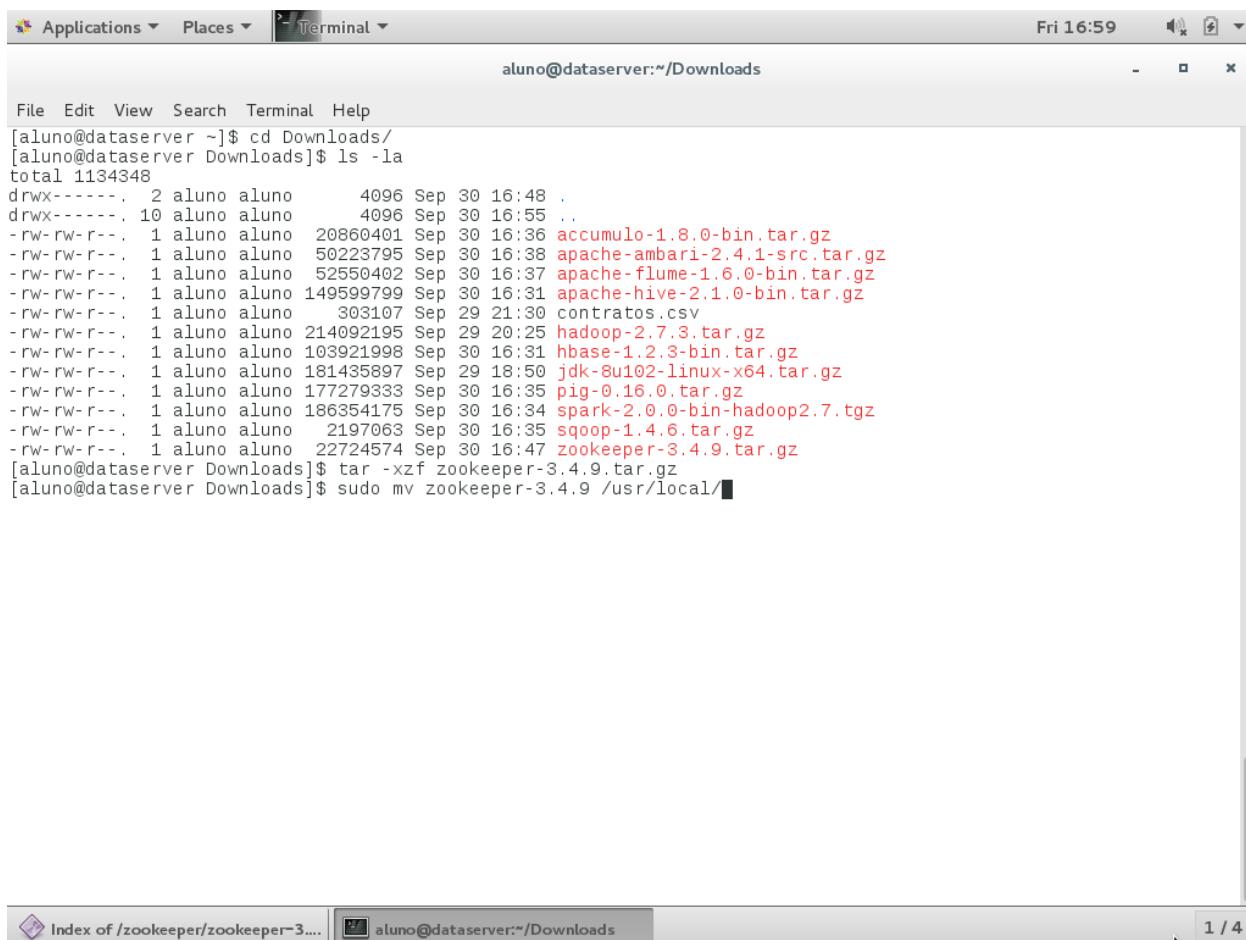
Download do Zookeeper – Versão 3.4.9



The screenshot shows a terminal window titled "Terminal" with the command-line interface. The user is in the "/Downloads" directory and lists files with "ls -a". The terminal shows several files including "contratos.csv", "hadoop-2.7.3.tar.gz", "jdk-8u102-linux-x64.tar.gz", and "zookeeper-3.4.9.tar.gz". The user then runs "tar -xzf zookeeper-3.4.9.tar.gz". Below the terminal window is a status bar with "Index of /zookeeper/zookeeper-3....", "aluno@dataserver:~/Downloads", and "1 / 4".

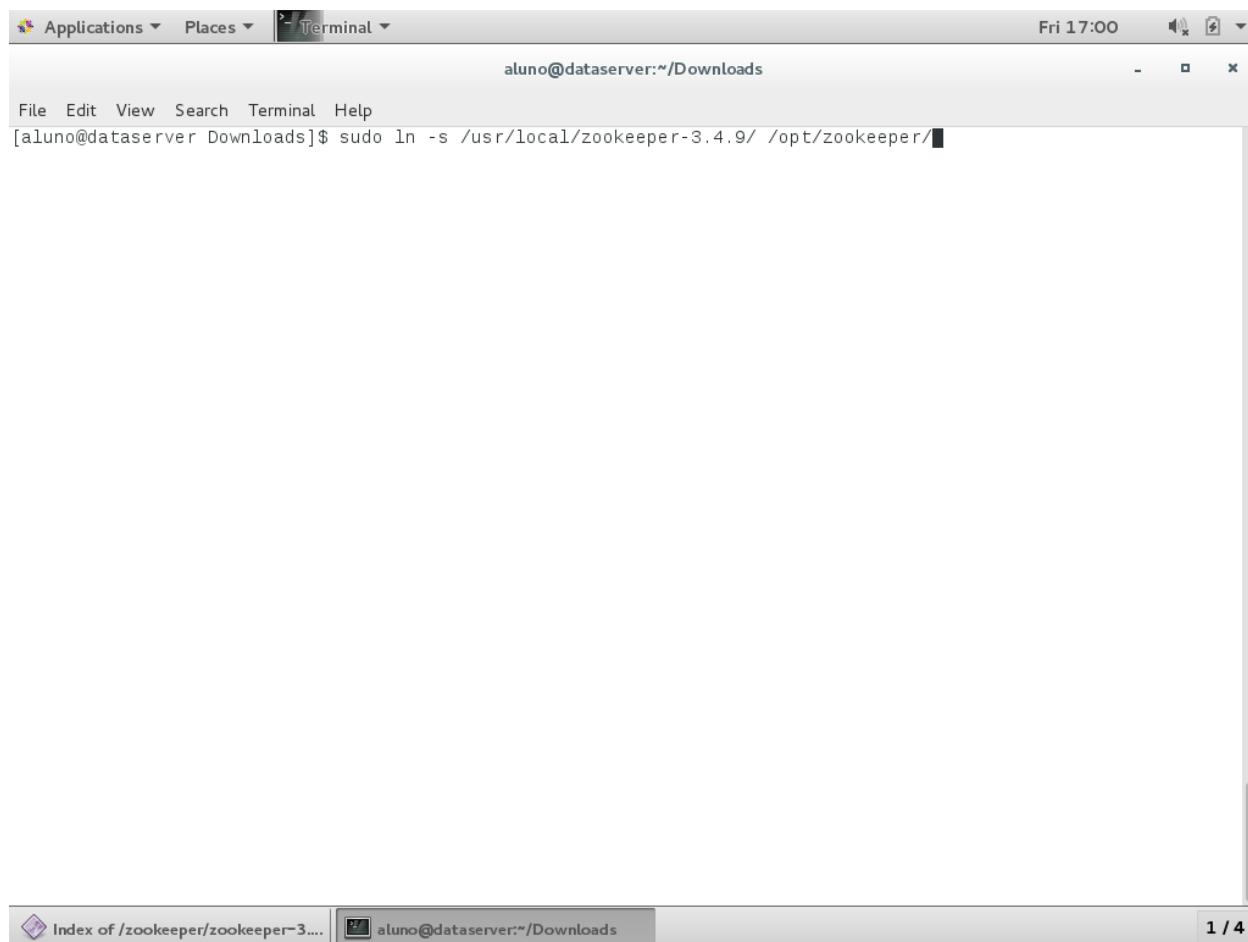
```
[aluno@dataserver ~]$ cd Downloads/  
[aluno@dataserver Downloads]$ ls -a  
. .. contratos.csv hadoop-2.7.3.tar.gz jdk-8u102-linux-x64.tar.gz zookeeper-3.4.9.tar.gz  
[aluno@dataserver Downloads]$ l s-la  
bash: l: command not found  
[aluno@dataserver Downloads]$ ls -la  
total 417928  
drwx----- . 2 aluno aluno 4096 Sep 30 15:52 .  
drwx----- . 10 aluno aluno 4096 Sep 30 15:25 ..  
-rw-rw-r-- . 1 aluno aluno 303107 Sep 29 21:30 contratos.csv  
-rw-rw-r-- . 1 aluno aluno 214092195 Sep 29 20:25 hadoop-2.7.3.tar.gz  
-rw-rw-r-- . 1 aluno aluno 181435897 Sep 29 18:50 jdk-8u102-linux-x64.tar.gz  
-rw-rw-r-- . 1 aluno aluno 22724574 Sep 30 15:52 zookeeper-3.4.9.tar.gz  
[aluno@dataserver Downloads]$ tar -xzf zookeeper-3.4.9.tar.gz
```

Descompactar o arquivo



```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ ls -la
total 1134348
drwx----- 2 aluno aluno 4096 Sep 30 16:48 .
drwx----- 10 aluno aluno 4096 Sep 30 16:55 ..
-rw-rw-r-- 1 aluno aluno 20860401 Sep 30 16:36 accumulo-1.8.0-bin.tar.gz
-rw-rw-r-- 1 aluno aluno 50223795 Sep 30 16:38 apache-ambari-2.4.1-src.tar.gz
-rw-rw-r-- 1 aluno aluno 52550402 Sep 30 16:37 apache-flume-1.6.0-bin.tar.gz
-rw-rw-r-- 1 aluno aluno 149599799 Sep 30 16:31 apache-hive-2.1.0-bin.tar.gz
-rw-rw-r-- 1 aluno aluno 303107 Sep 29 21:30 contratos.csv
-rw-rw-r-- 1 aluno aluno 214092195 Sep 29 20:25 hadoop-2.7.3.tar.gz
-rw-rw-r-- 1 aluno aluno 103921998 Sep 30 16:31 hbase-1.2.3-bin.tar.gz
-rw-rw-r-- 1 aluno aluno 181435897 Sep 29 18:50 jdk-8u102-linux-x64.tar.gz
-rw-rw-r-- 1 aluno aluno 177279333 Sep 30 16:35 pig-0.16.0.tar.gz
-rw-rw-r-- 1 aluno aluno 186354175 Sep 30 16:34 spark-2.0.0-bin-hadoop2.7.tgz
-rw-rw-r-- 1 aluno aluno 2197063 Sep 30 16:35 sqoop-1.4.6.tar.gz
-rw-rw-r-- 1 aluno aluno 22724574 Sep 30 16:47 zookeeper-3.4.9.tar.gz
[aluno@dataserver Downloads]$ tar -xzf zookeeper-3.4.9.tar.gz
[aluno@dataserver Downloads]$ sudo mv zookeeper-3.4.9 /usr/local/
```

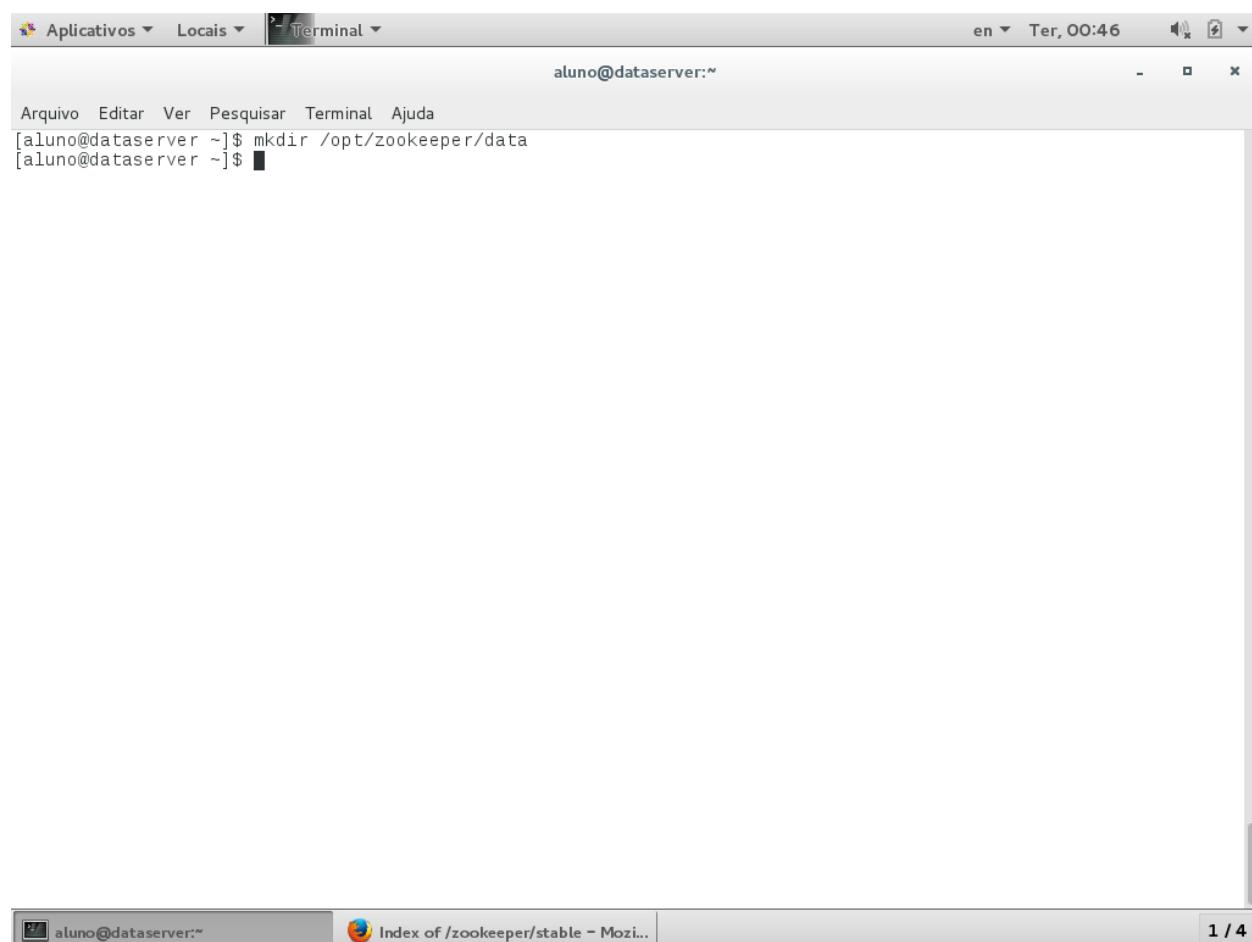
Mover o diretório do Zookeeper para /usr/local



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt shows the user is at "aluno@dataserver:~/Downloads". The command entered is "sudo ln -s /usr/local/zookeeper-3.4.9/ /opt/zookeeper/". Below the terminal window, there is a file browser window titled "Index of /zookeeper/zookeeper-3...." showing the contents of the Zookeeper distribution.

Criar link simbólico no diretório /opt

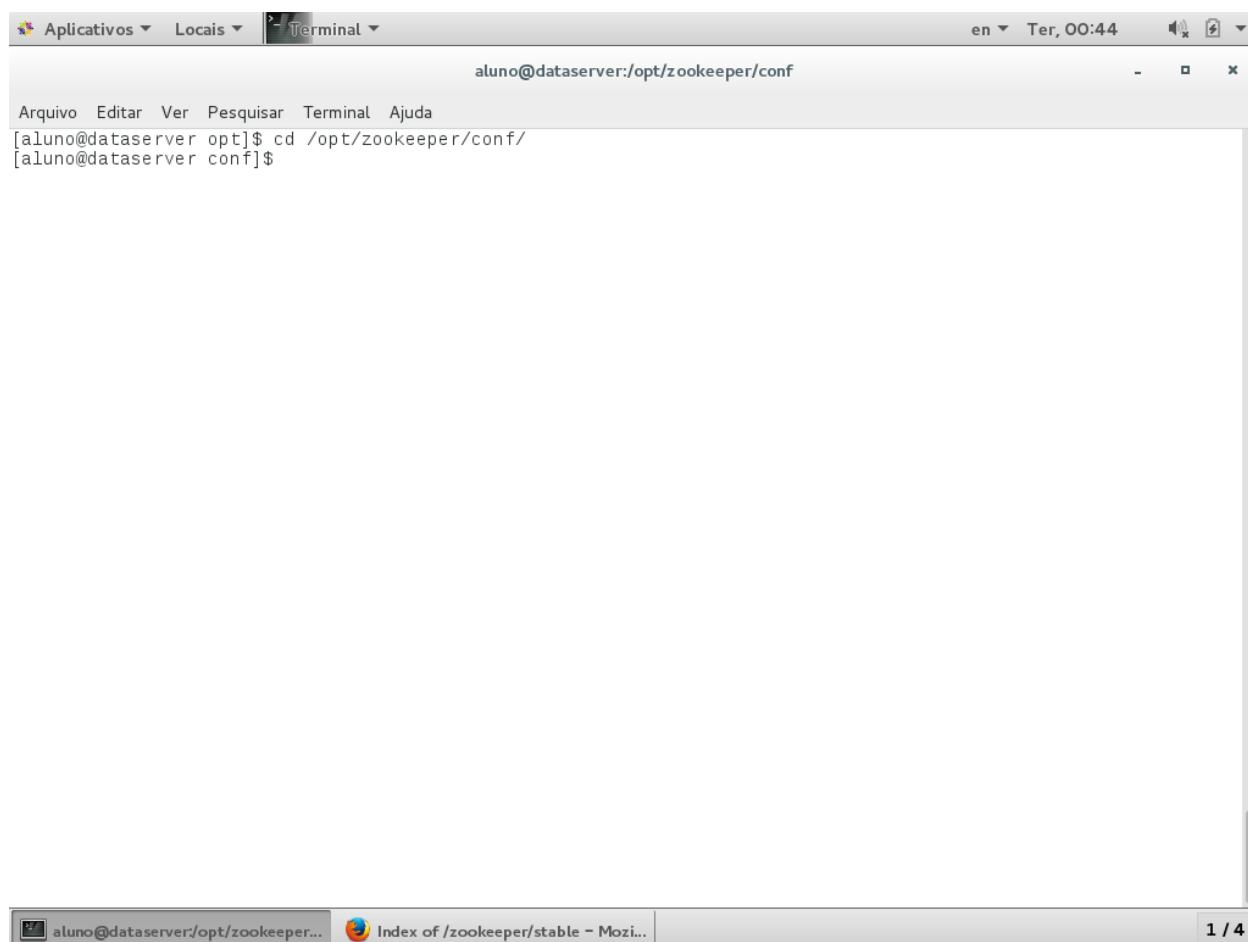
6.2. Configurando do Zookeeper



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~'. The user has run the command 'mkdir /opt/zookeeper/data'. The terminal window is part of a larger desktop interface with other windows visible in the background.

```
aluno@dataserver:~$ mkdir /opt/zookeeper/data
```

Criar o diretório **data** dentro de /opt/zookeeper

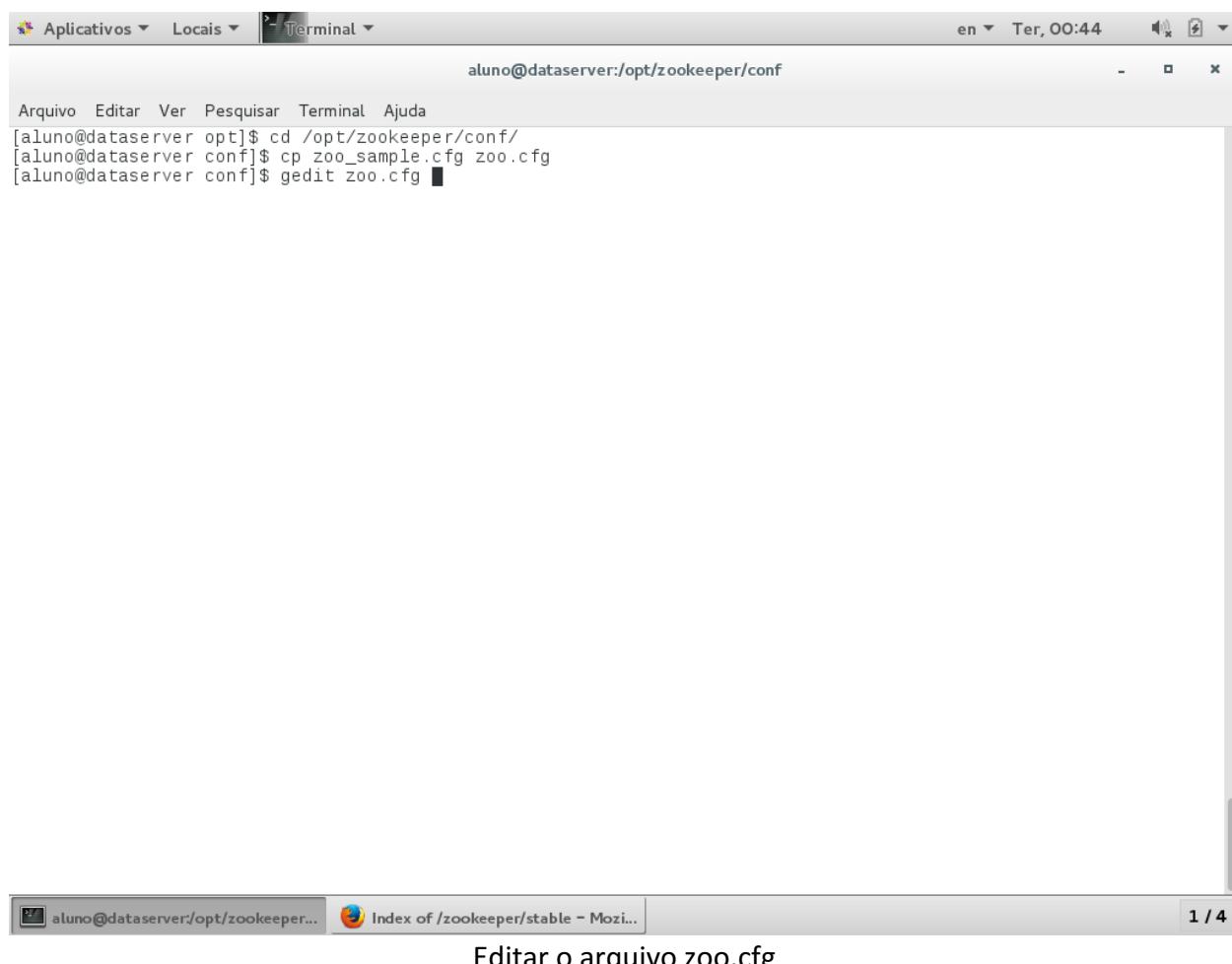


Acessar o diretório /opt/zookeeper/conf



```
aluno@dataserver:~$ cd /opt/zookeeper/conf/
[aluno@dataserver conf]$ cp zoo_sample.cfg zoo.cfg
```

A partir do arquivo template, gerar o arquivo zoo.cfg

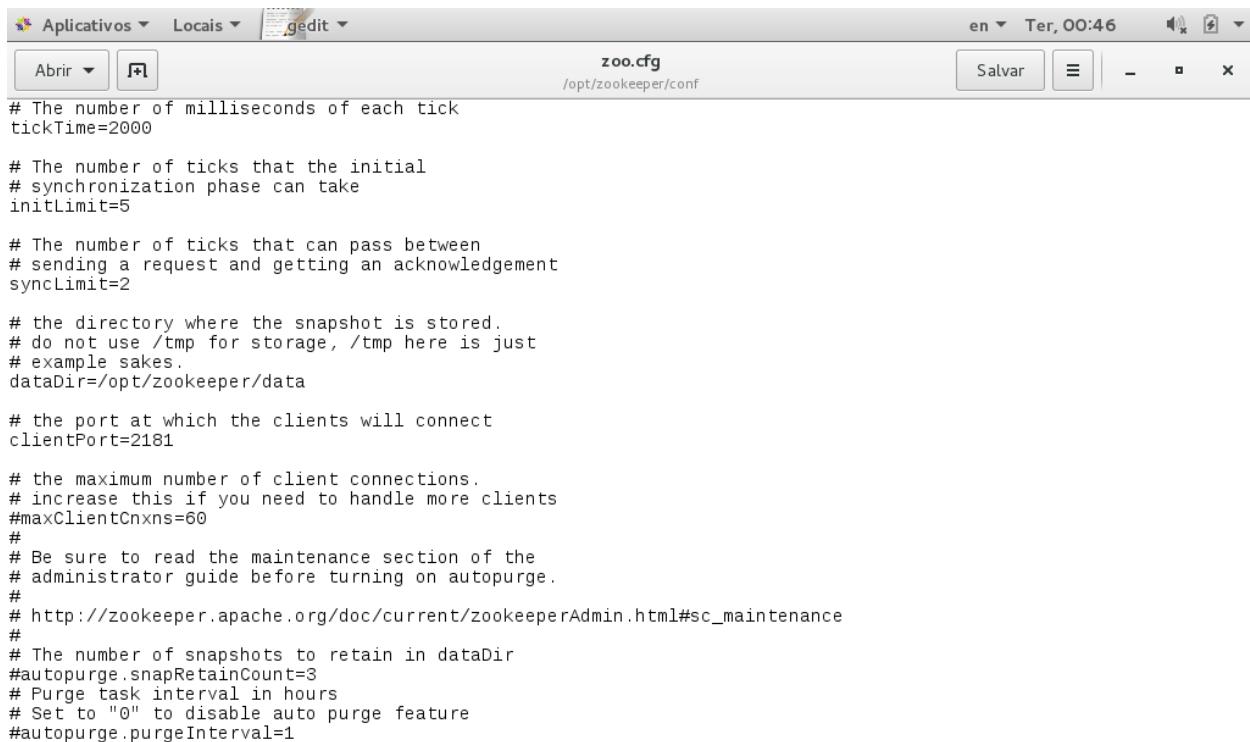


A screenshot of a Linux desktop environment showing a terminal window. The terminal title is "Terminal" and the current directory is "aluno@dataserver:/opt/zookeeper/conf". The terminal window contains the following command history:

```
[aluno@dataserver opt]$ cd /opt/zookeeper/conf/  
[aluno@dataserver conf]$ cp zoo_sample.cfg zoo.cfg  
[aluno@dataserver conf]$ gedit zoo.cfg
```

The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows "aluno@dataserver:/opt/zookeeper..." and "Index of /zookeeper/stable - Mozilla Firefox". A progress bar indicates "1 / 4".

Editar o arquivo zoo.cfg



```
# The number of milliseconds of each tick
tickTime=2000

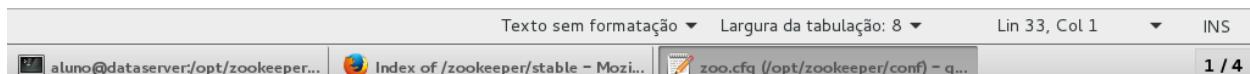
# The number of ticks that the initial
# synchronization phase can take
initLimit=5

# The number of ticks that can pass between
# sending a request and getting an acknowledgement
syncLimit=2

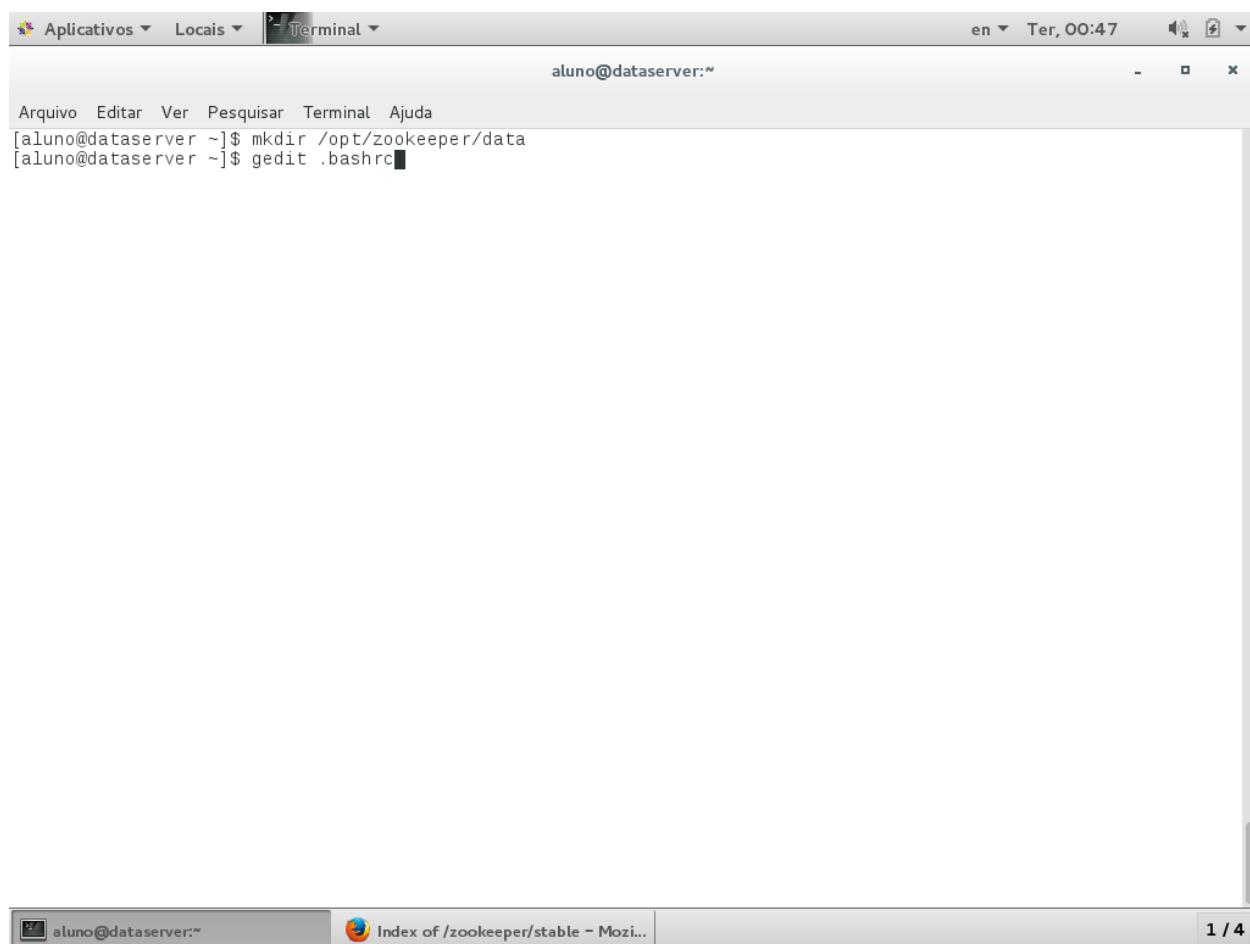
# the directory where the snapshot is stored.
# do not use /tmp for storage, /tmp here is just
# example sakes.
dataDir=/opt/zookeeper/data

# the port at which the clients will connect
clientPort=2181

# the maximum number of client connections.
# increase this if you need to handle more clients
#maxClientCnxns=60
#
# Be sure to read the maintenance section of the
# administrator guide before turning on autopurge.
#
# http://zookeeper.apache.org/doc/current/zookeeperAdmin.html#sc_maintenance
#
# The number of snapshots to retain in dataDir
#autopurge.snapRetainCount=3
# Purge task interval in hours
# Set to "0" to disable auto purge feature
#autopurge.purgeInterval=1
```

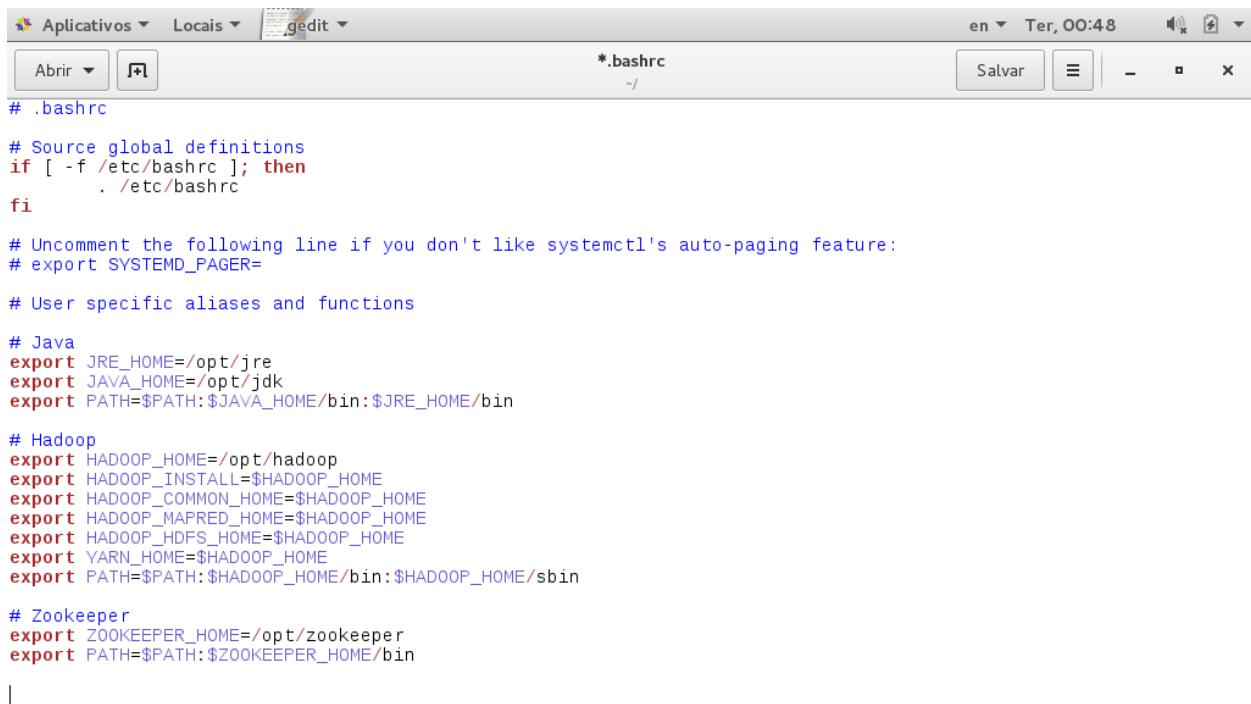


Editar o arquivo conforme tela acima



```
aluno@dataserver:~$ mkdir /opt/zookeeper/data
[aluno@dataserver ~]$ gedit .bashrc
```

Incluir variáveis Zookeeper no /home/aluno/.bashrc



```

Aplicativos Locais gedit
Abrir Salvar
*.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=

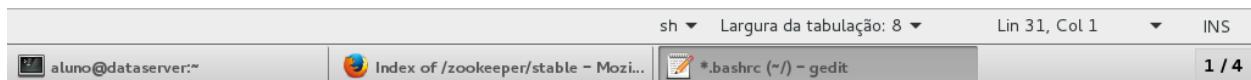
# User specific aliases and functions

# Java
export JRE_HOME=/opt/jre
export JAVA_HOME=/opt/jdk
export PATH=$PATH:$JAVA_HOME/bin:$JRE_HOME/bin

# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Zookeeper
export ZOOKEEPER_HOME=/opt/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin

```



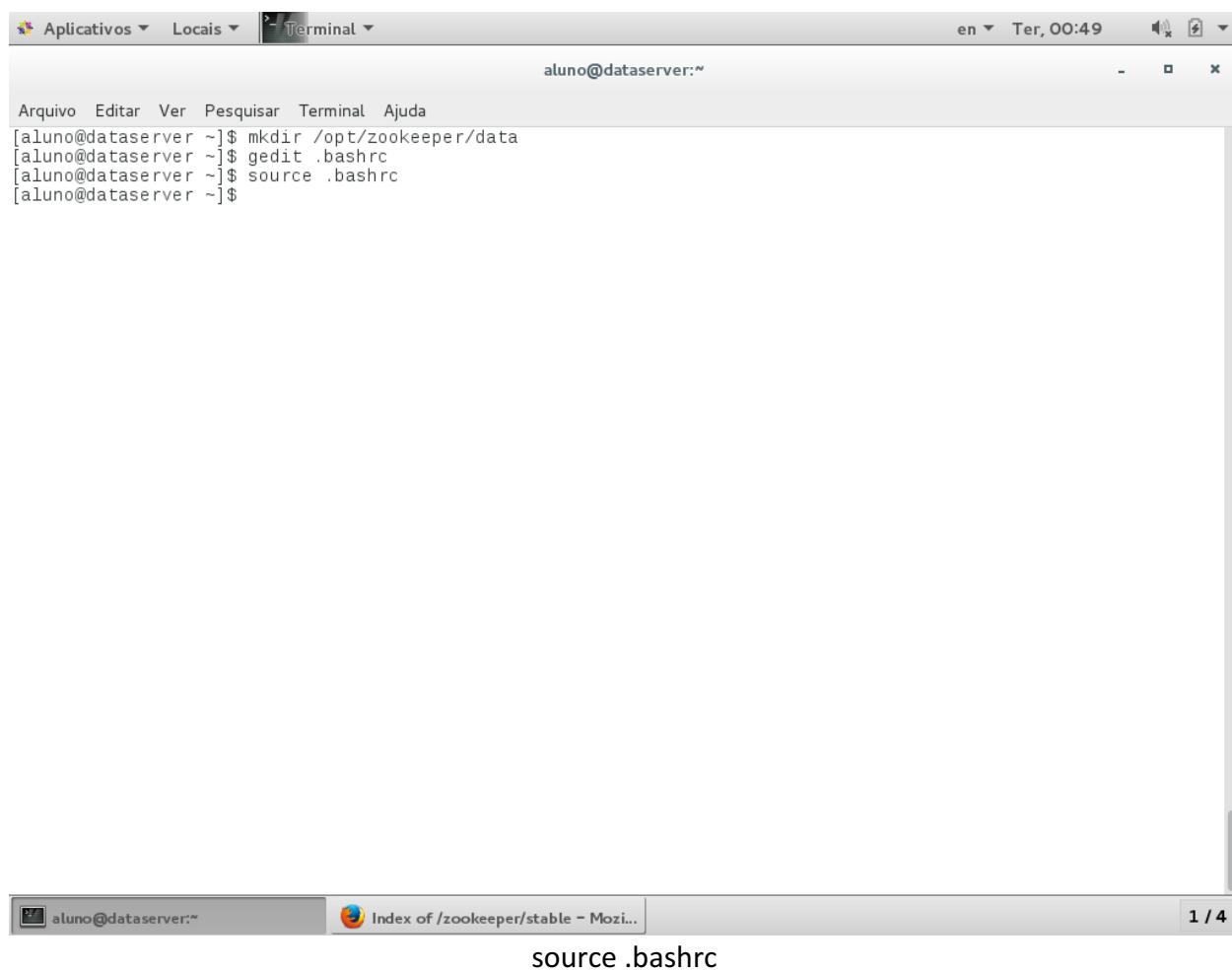
```

sh Largura da tabulação: 8 Lin 31, Col 1 INS
aluno@dataserver:~ Index of /zookeeper/stable - Mozilla/5.0 *.* bashrc (~/) - gedit 1 / 4

```

Variáveis Zookeeper

Instalação e Configuração do Ecosistema Hadoop

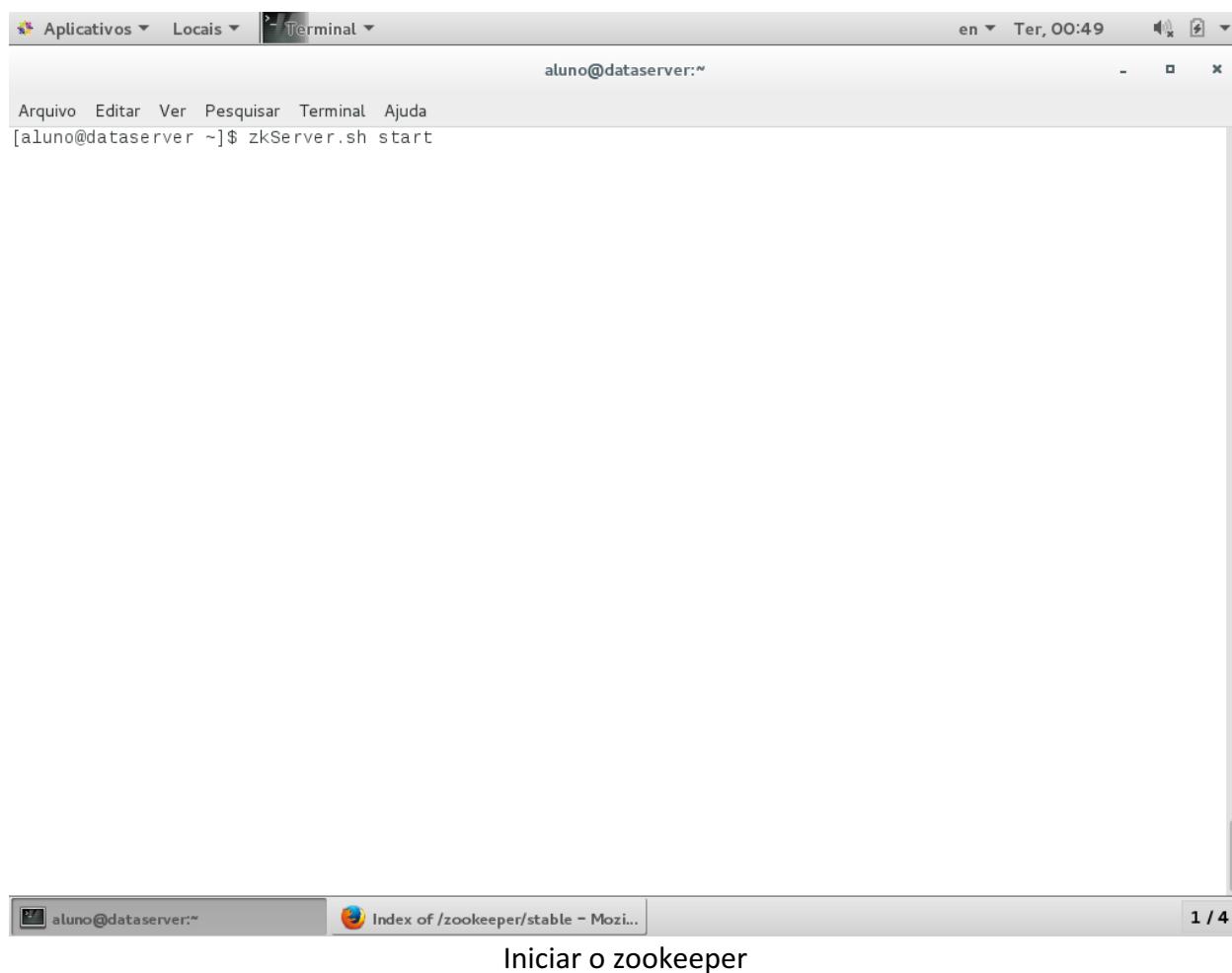


A screenshot of a Linux desktop environment showing a terminal window. The terminal title is "Terminal" and the user is "aluno@dataserver:~". The terminal content shows the following commands being run:

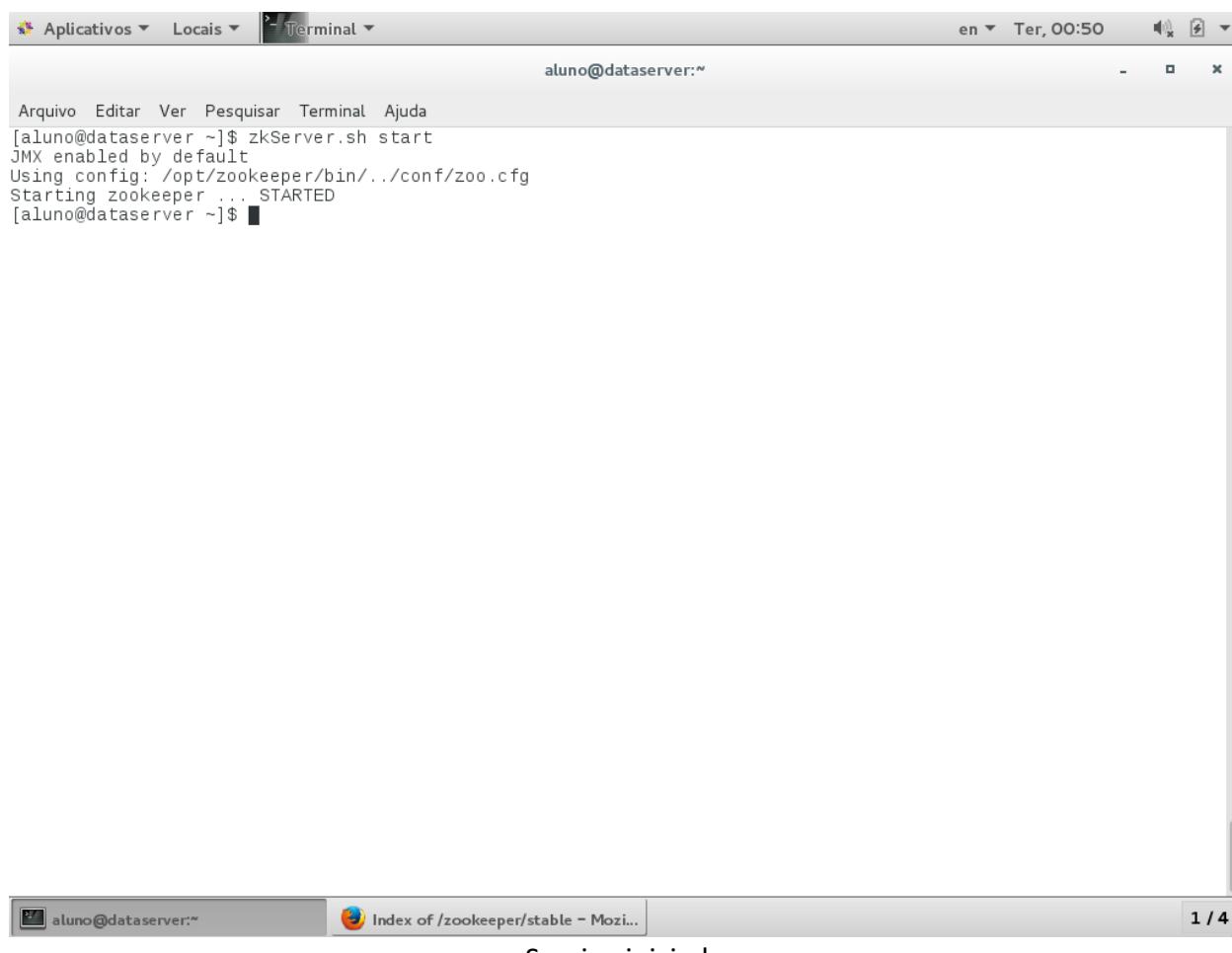
```
[aluno@dataserver ~]$ mkdir /opt/zookeeper/data
[aluno@dataserver ~]$ gedit .bashrc
[aluno@dataserver ~]$ source .bashrc
[aluno@dataserver ~]$
```

The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows "aluno@dataserver:~" and "Index of /zookeeper/stable - Mozilla...". A navigation bar at the bottom indicates "1 / 4".

Instalação e Configuração do Ecosistema Hadoop



Iniciar o zookeeper

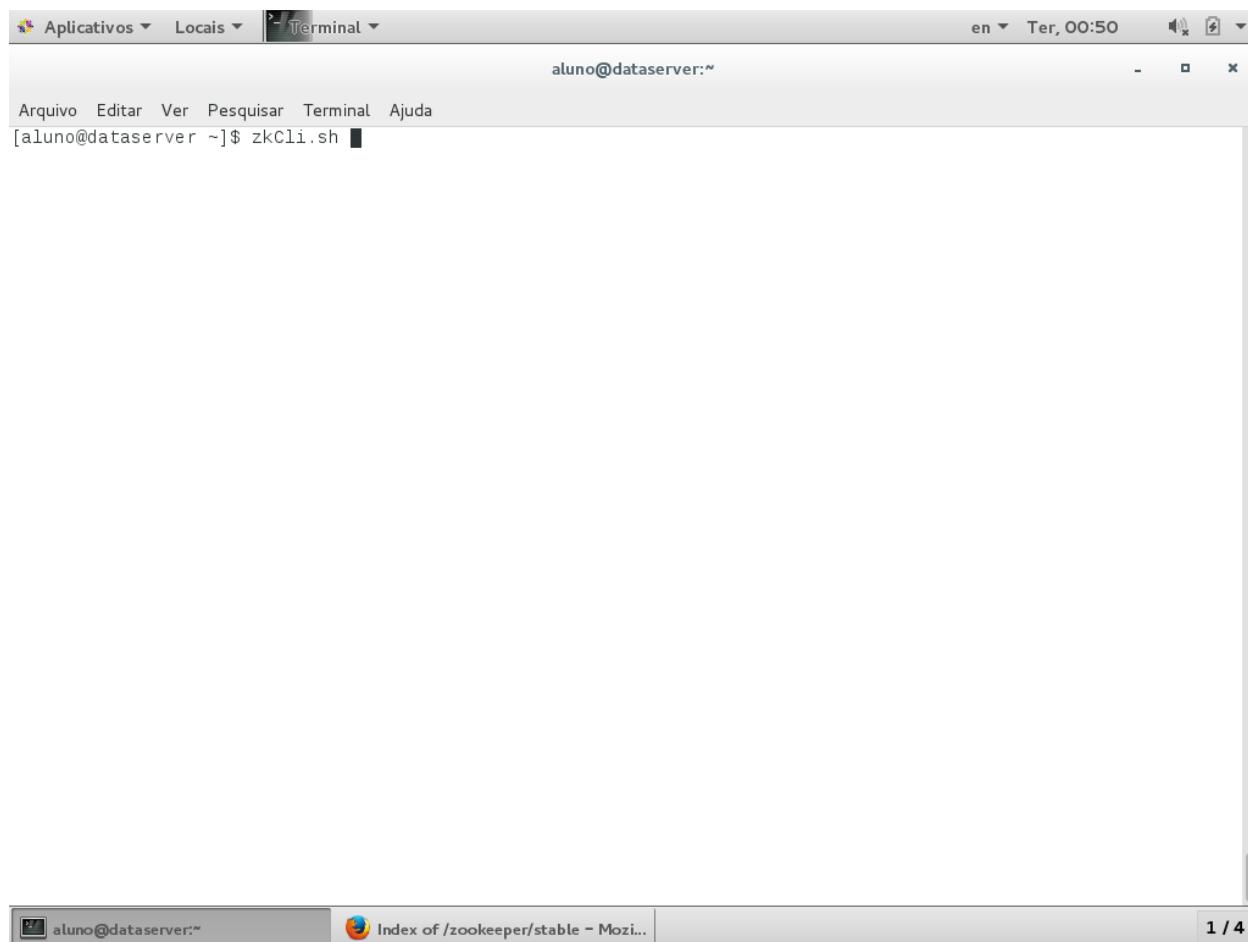


A screenshot of a Linux terminal window titled "Terminal". The window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar shows "aluno@dataserver:~" and "en Ter, 00:50". The terminal content shows the command "zkServer.sh start" being run, followed by output indicating JMX is enabled by default, the configuration file is "/opt/zookeeper/bin/../conf/zoo.cfg", and the zookeeper service has started successfully. The terminal ends with "[aluno@dataserver ~]\$".

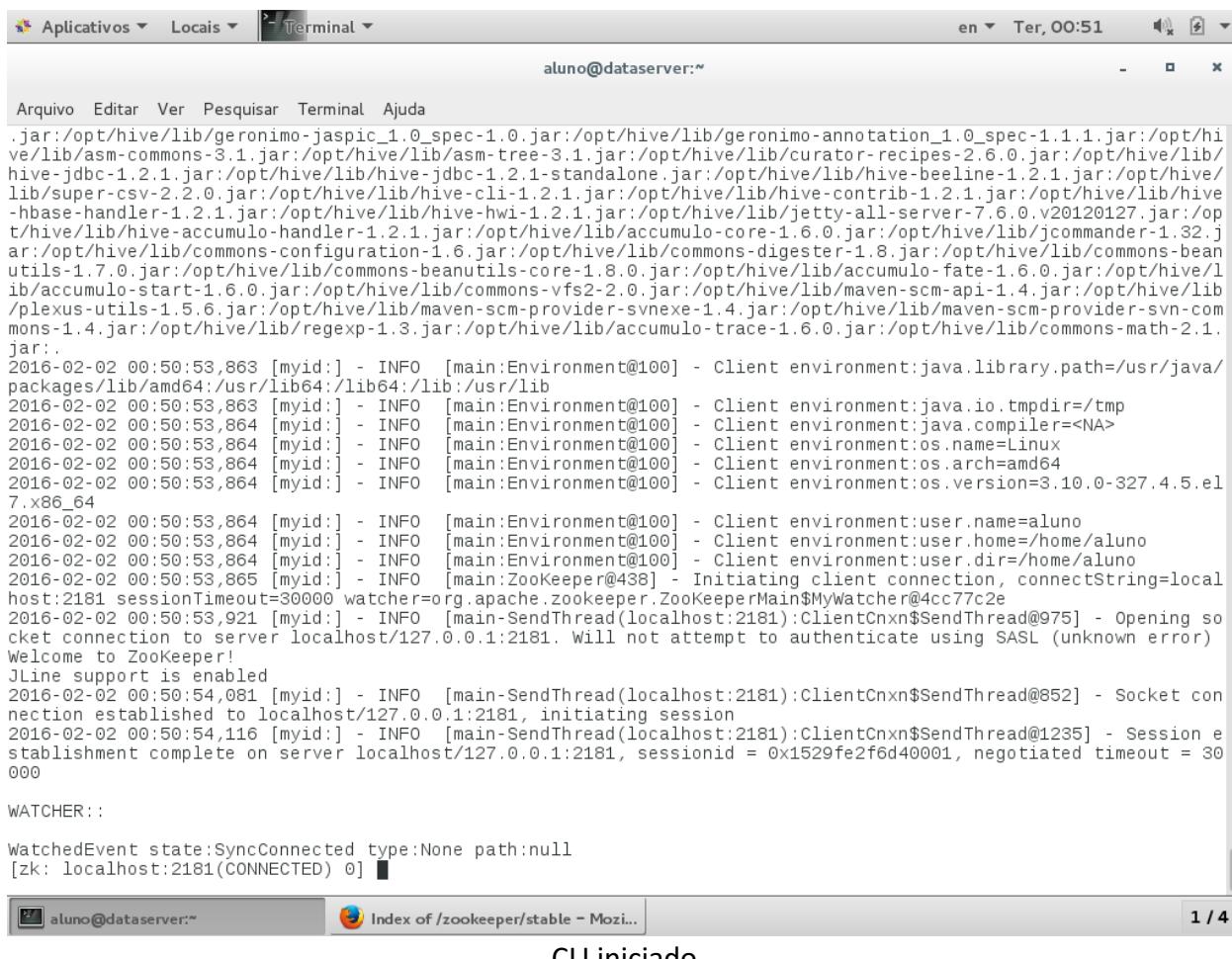
[aluno@dataserver ~]\$ zkServer.sh start
JMX enabled by default
Using config: /opt/zookeeper/bin/../conf/zoo.cfg
Starting zookeeper ... STARTED
[aluno@dataserver ~]\$

The terminal window is part of a larger desktop environment. Below it, a taskbar shows other open applications: a terminal window labeled "aluno@dataserver:~" and a Mozilla Firefox browser window titled "Index of /zookeeper/stable - Mozi...". A status bar at the bottom right indicates "1 / 4".

Serviço iniciado



Iniciar o Zookeeper Command Line Interface (CLI)



```

aluno@dataserver:~$ 
Arquivo Editar Ver Pesquisar Terminal Ajuda
.jar:/opt/hive/lib/geronimo-jaspic_1.0_spec-1.0.jar:/opt/hive/lib/geronimo-annotation_1.0_spec-1.1.1.jar:/opt/hive/lib/asm-commons-3.1.jar:/opt/hive/lib/asm-tree-3.1.jar:/opt/hive/lib/curator-recipes-2.6.0.jar:/opt/hive/lib/hive-jdbc-1.2.1.jar:/opt/hive/lib/hive-jdbc-1.2.1-standalone.jar:/opt/hive/lib/hive-beeline-1.2.1.jar:/opt/hive/lib/lib/super-csv-2.2.0.jar:/opt/hive/lib/hive-cli-1.2.1.jar:/opt/hive/lib/hive-contrib-1.2.1.jar:/opt/hive/lib/hive-hbase-handler-1.2.1.jar:/opt/hive/lib/hive-hwi-1.2.1.jar:/opt/hive/lib/jetty-all-server-7.6.0.v20120127.jar:/opt/hive/lib/hive-accumulo-handler-1.2.1.jar:/opt/hive/lib/accumulo-core-1.6.0.jar:/opt/hive/lib/jcommander-1.32.jar:/opt/hive/lib/commons-configuration-1.6.jar:/opt/hive/lib/commons-digester-1.8.jar:/opt/hive/lib/commons-beanutils-1.7.0.jar:/opt/hive/lib/commons-beanutils-core-1.8.0.jar:/opt/hive/lib/accumulo-fate-1.6.0.jar:/opt/hive/lib/accumulo-start-1.6.0.jar:/opt/hive/lib/commons-vfs2-2.0.jar:/opt/hive/lib/maven-scm-api-1.4.jar:/opt/hive/lib/plexus-utils-1.5.6.jar:/opt/hive/lib/maven-scm-provider-svnexe-1.4.jar:/opt/hive/lib/maven-scm-provider-svn-commons-1.4.jar:/opt/hive/lib/regexp-1.3.jar:/opt/hive/lib/accumulo-trace-1.6.0.jar:/opt/hive/lib/commons-math-2.1.jar:.

2016-02-02 00:50:53,863 [myid:] - INFO  [main:Environment@100] - Client environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib
2016-02-02 00:50:53,863 [myid:] - INFO  [main:Environment@100] - Client environment:java.io.tmpdir=/tmp
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:java.compiler=<NA>
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:os.name=Linux
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:os.arch=amd64
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:os.version=3.10.0-327.45.el7.x86_64
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:user.name=aluno
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:user.home=/home/aluno
2016-02-02 00:50:53,864 [myid:] - INFO  [main:Environment@100] - Client environment:user.dir=/home/aluno
2016-02-02 00:50:53,865 [myid:] - INFO  [main:ZooKeeper@438] - Initiating client connection, connectString=localhost:2181 sessionTimeout=30000 watcher=org.apache.zookeeper.ZooKeeperMain$MyWatcher@4cc77c2e
2016-02-02 00:50:53,921 [myid:] - INFO  [main-SendThread(localhost:2181):ClientCnxn$SendThread@975] - Opening socket connection to server localhost/127.0.0.1:2181. Will not attempt to authenticate using SASL (unknown error)
Welcome to ZooKeeper!
JLine support is enabled
2016-02-02 00:50:54,081 [myid:] - INFO  [main-SendThread(localhost:2181):ClientCnxn$SendThread@852] - Socket connection established to localhost/127.0.0.1:2181, initiating session
2016-02-02 00:50:54,116 [myid:] - INFO  [main-SendThread(localhost:2181):ClientCnxn$SendThread@1235] - Session establishment complete on server localhost/127.0.0.1:2181, sessionid = 0x1529fe2f6d40001, negotiated timeout = 30 000

WATCHER:::

WatchedEvent state:SyncConnected type:None path:null
[zk: localhost:2181(CONNECTED) 0] 
```

CLI iniciado

7. Instalação e Configuração do HBase

Podemos instalar HBase em qualquer um dos três modos: Standalone mode, Pseudo Distributed mode e Fully Distributed mode.

7.1. Download e Instalação do HBase

Welcome to Apache HBase™

Apache HBase™ is the Hadoop database, a distributed, scalable, big data store. Use Apache HBase™ when you need random, realtime read/write access to your Big Data. This project's goal is the hosting of very large tables -- billions of rows X millions of columns -- atop clusters of commodity hardware. Apache HBase is an open-source, distributed, versioned, non-relational database modeled after Google's [Bigtable: A Distributed Storage System for Structured Data](#) by Chang et al. Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and HDFS.

Download

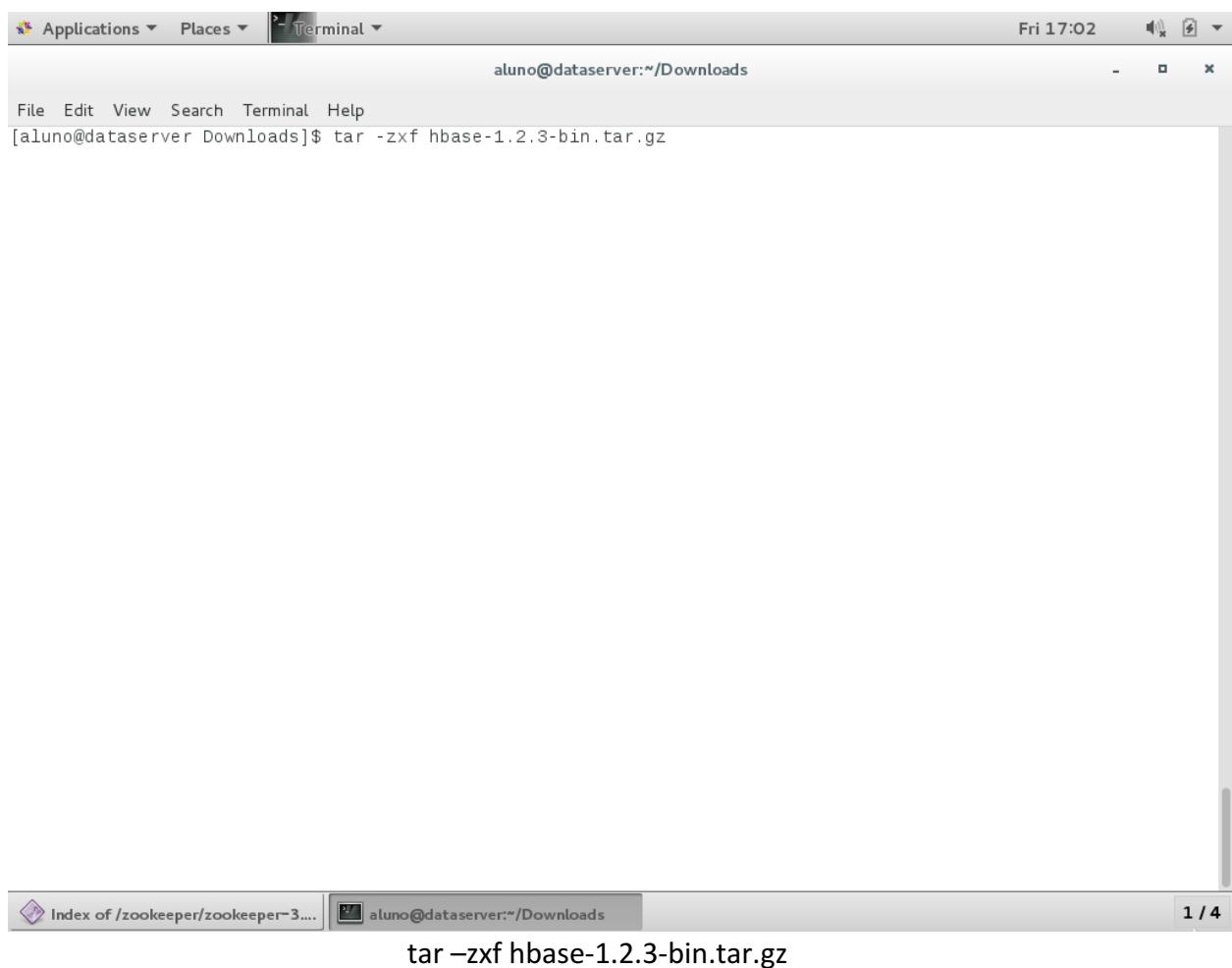
Click [here](#) to download Apache HBase™.

Features

- Linear and modular scalability.
- Strictly consistent reads and writes.
- Automatic and configurable sharding of tables

Download do Hbase – Versão 1.2.3

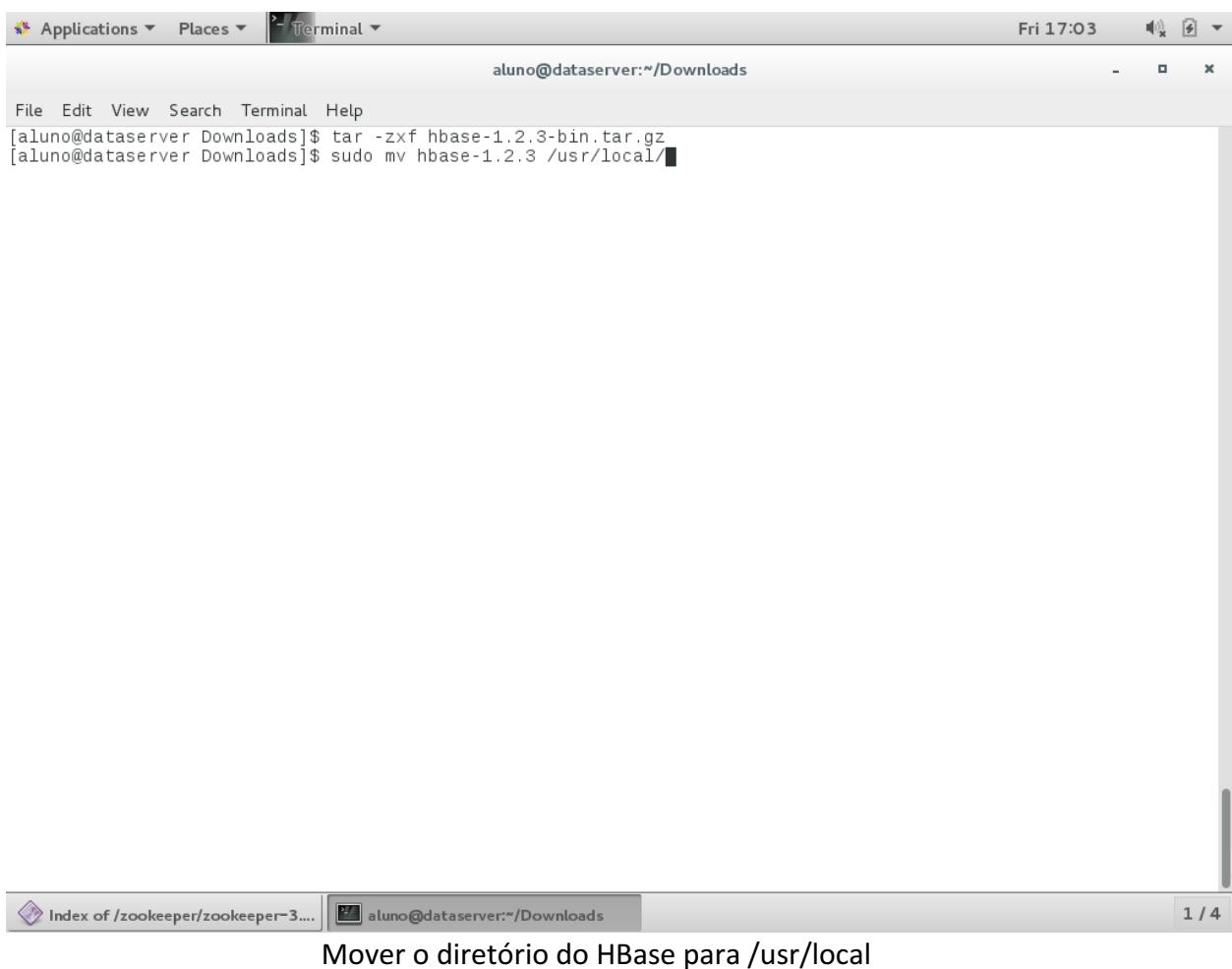
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top right shows "Fri 17:02". The terminal prompt is "aluno@dataserver:~/Downloads". The command entered is "tar -zxf hbase-1.2.3-bin.tar.gz". Below the terminal window, there is a file browser window titled "Index of /zookeeper/zookeeper-3...." showing the same command again.

```
File Edit View Search Terminal Help
[aluno@dataserver Downloads]$ tar -zxf hbase-1.2.3-bin.tar.gz
```

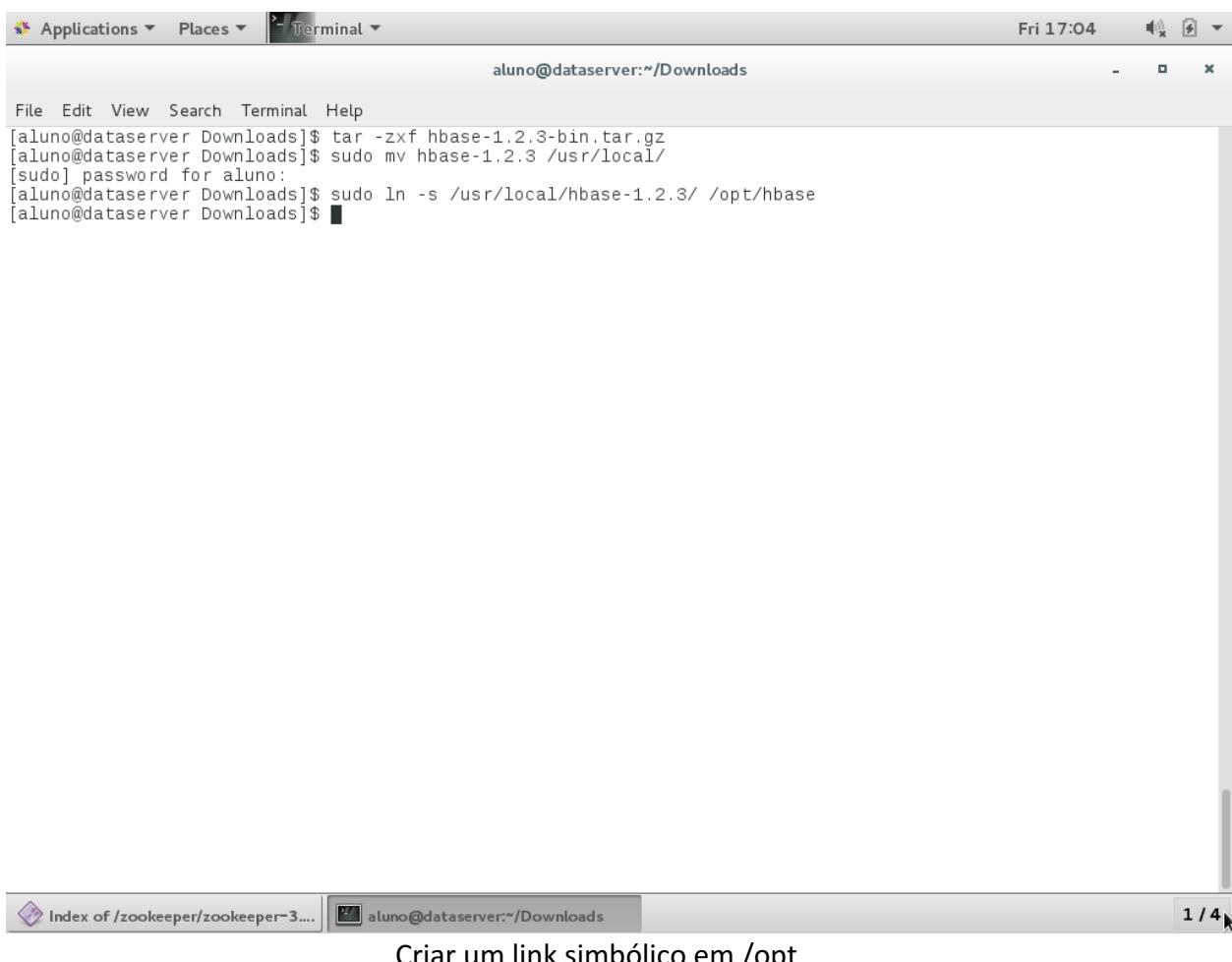
```
Index of /zookeeper/zookeeper-3.... 1 / 4
aluno@dataserver:~/Downloads
tar -zxf hbase-1.2.3-bin.tar.gz
```



```
File Edit View Search Terminal Help
[aluno@dataserver Downloads]$ tar -zxf hbase-1.2.3-bin.tar.gz
[aluno@dataserver Downloads]$ sudo mv hbase-1.2.3 /usr/local/
```

Index of /zookeeper/zookeeper-3.... aluno@dataserver:~/Downloads 1 / 4

Mover o diretório do HBase para /usr/local



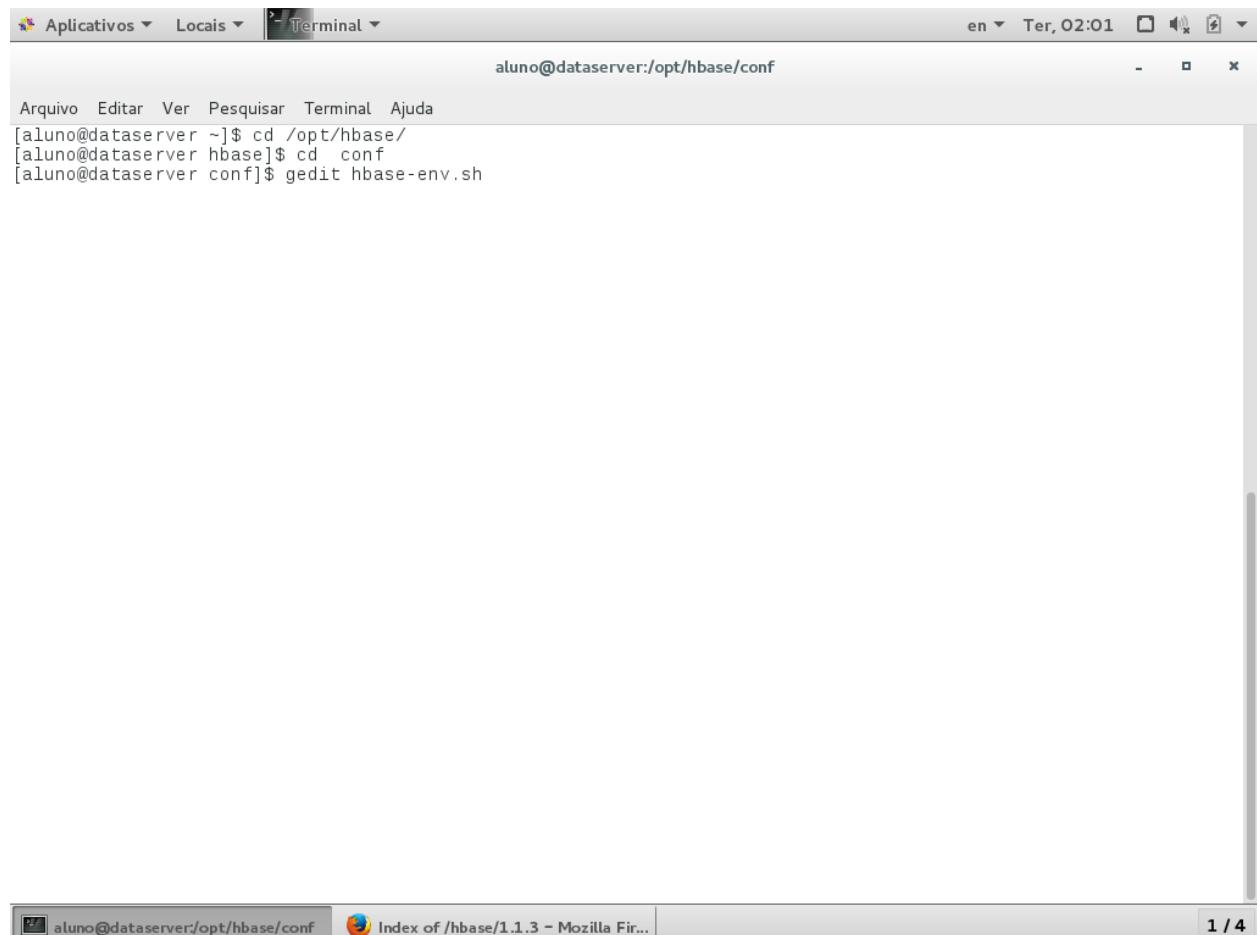
The screenshot shows a terminal window titled "Terminal" with the command line interface "aluno@dataserver:~/Downloads". The terminal displays the following commands:

```
File Edit View Search Terminal Help
[aluno@dataserver Downloads]$ tar -zxf hbase-1.2.3-bin.tar.gz
[aluno@dataserver Downloads]$ sudo mv hbase-1.2.3 /usr/local/
[sudo] password for aluno:
[aluno@dataserver Downloads]$ sudo ln -s /usr/local/hbase-1.2.3/ /opt/hbase
[aluno@dataserver Downloads]$
```

Below the terminal window, there is a small thumbnail image labeled "Index of /zookeeper/zookeeper-3...." with the status "aluno@dataserver:~/Downloads" and the number "1 / 4".

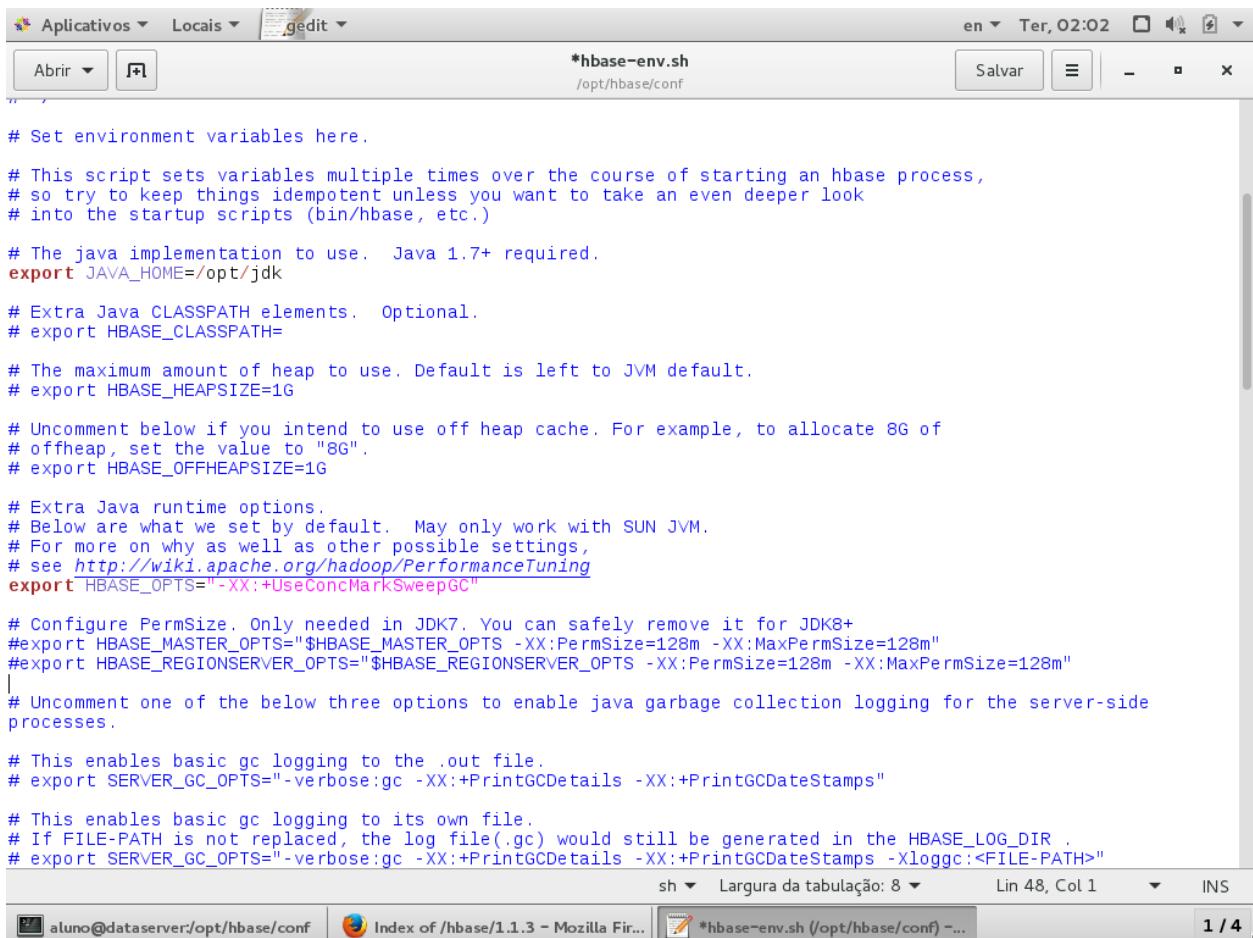
Criar um link simbólico em /opt

7.2. Configurando o HBase



A screenshot of a Linux terminal window titled "Terminal". The window shows the user's session path: "aluno@dataserver:/opt/hbase/conf". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The terminal history shows the commands: [aluno@dataserver ~]\$ cd /opt/hbase/ [aluno@dataserver hbase]\$ cd conf [aluno@dataserver conf]\$ gedit hbase-env.sh

No diretório /opt/hbase/conf, editar o arquivo hbase-env.sh



```

Aplicativos Locais gedit
Abrir + *hbase-env.sh
/opt/hbase/conf
Salvar - x
# Set environment variables here.

# This script sets variables multiple times over the course of starting an hbase process,
# so try to keep things idempotent unless you want to take an even deeper look
# into the startup scripts (bin/hbase, etc.)

# The java implementation to use. Java 1.7+ required.
export JAVA_HOME=/opt/jdk

# Extra Java CLASSPATH elements. Optional.
# export HBASE_CLASSPATH=

# The maximum amount of heap to use. Default is left to JVM default.
# export HBASE_HEAPSIZE=1G

# Uncomment below if you intend to use off heap cache. For example, to allocate 8G of
# offheap, set the value to "8G".
# export HBASE_OFFHEAPSIZE=1G

# Extra Java runtime options.
# Below are what we set by default. May only work with SUN JVM.
# For more on why as well as other possible settings,
# see http://wiki.apache.org/hadoop/PerformanceTuning
export HBASE_OPTS="-XX:+UseConcMarkSweepGC"

# Configure PermSize. Only needed in JDK7. You can safely remove it for JDK8+
#export HBASE_MASTER_OPTS="$HBASE_MASTER_OPTS -XX:PermSize=128m -XX:MaxPermSize=128m"
#export HBASE_REGIONSERVER_OPTS="$HBASE_REGIONSERVER_OPTS -XX:PermSize=128m -XX:MaxPermSize=128m"

# Uncomment one of the below three options to enable java garbage collection logging for the server-side
processes.

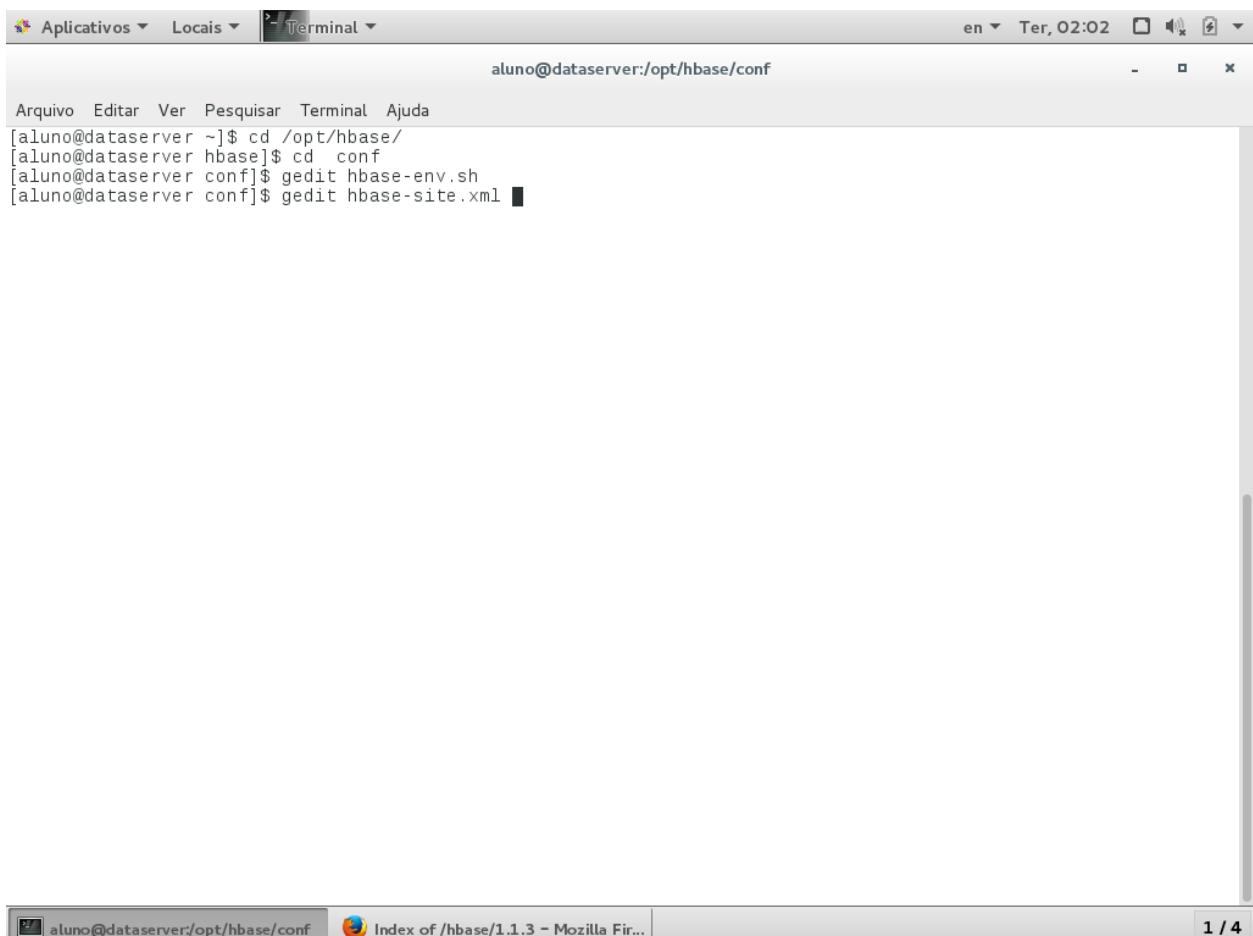
# This enables basic gc logging to the .out file.
# export SERVER_GC_OPTS="-verbose:gc -XX:+PrintGCDetails -XX:+PrintGCDateStamps"

# This enables basic gc logging to its own file.
# If FILE-PATH is not replaced, the log file(.gc) would still be generated in the HBASE_LOG_DIR .
# export SERVER_GC_OPTS="-verbose:gc -XX:+PrintGCDetails -XX:+PrintGCDateStamps -Xloggc:<FILE-PATH>"
```

sh ▾ Largura da tabulação: 8 ▾ Lin 48, Col 1 ▾ INS

aluno@dataserver:/opt/hbase/conf Index of /hbase/1.1.3 – Mozilla Fir... *hbase-env.sh (/opt/hbase/conf) ... 1 / 4

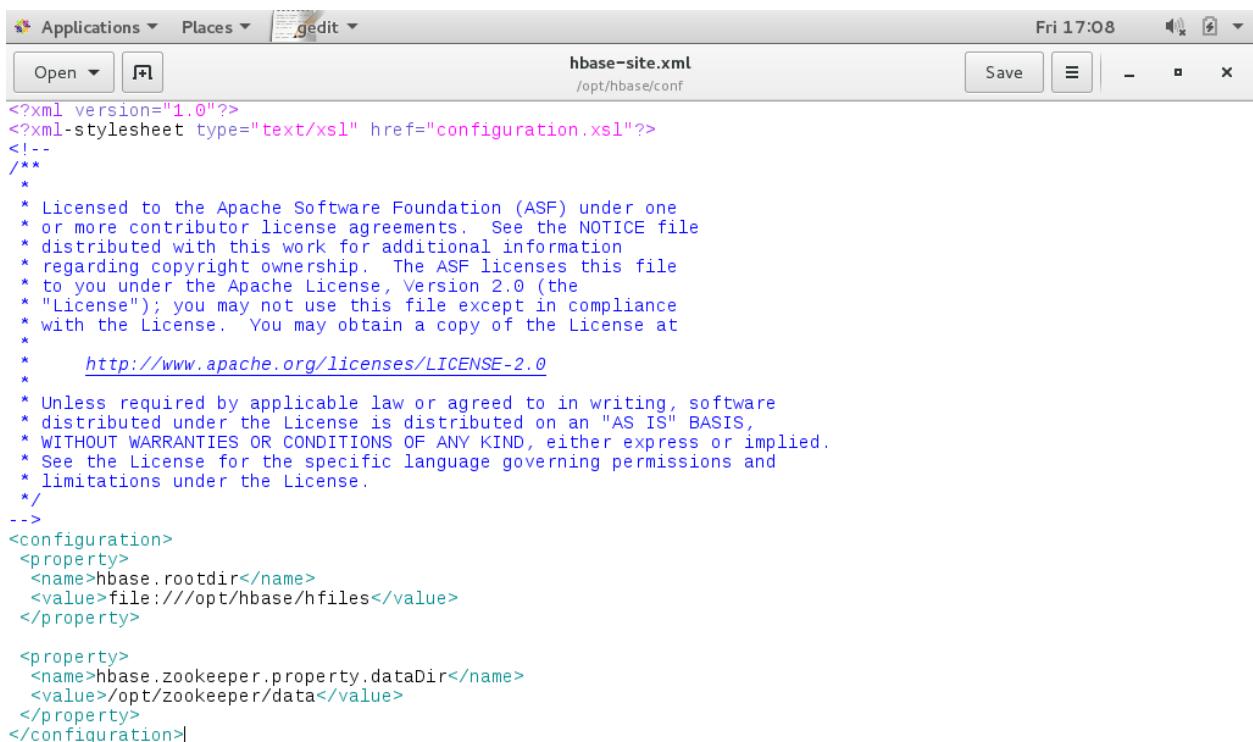
Editar o PATH do Java e comentar as linhas do PermSize



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:/opt/hbase/conf". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command history shows:

```
[aluno@dataserver ~]$ cd /opt/hbase/
[aluno@dataserver hbase]$ cd conf
[aluno@dataserver conf]$ gedit hbase-env.sh
[aluno@dataserver conf]$ gedit hbase-site.xml
```

No mesmo diretório conf, editar o arquivo hbase-site.xml



```

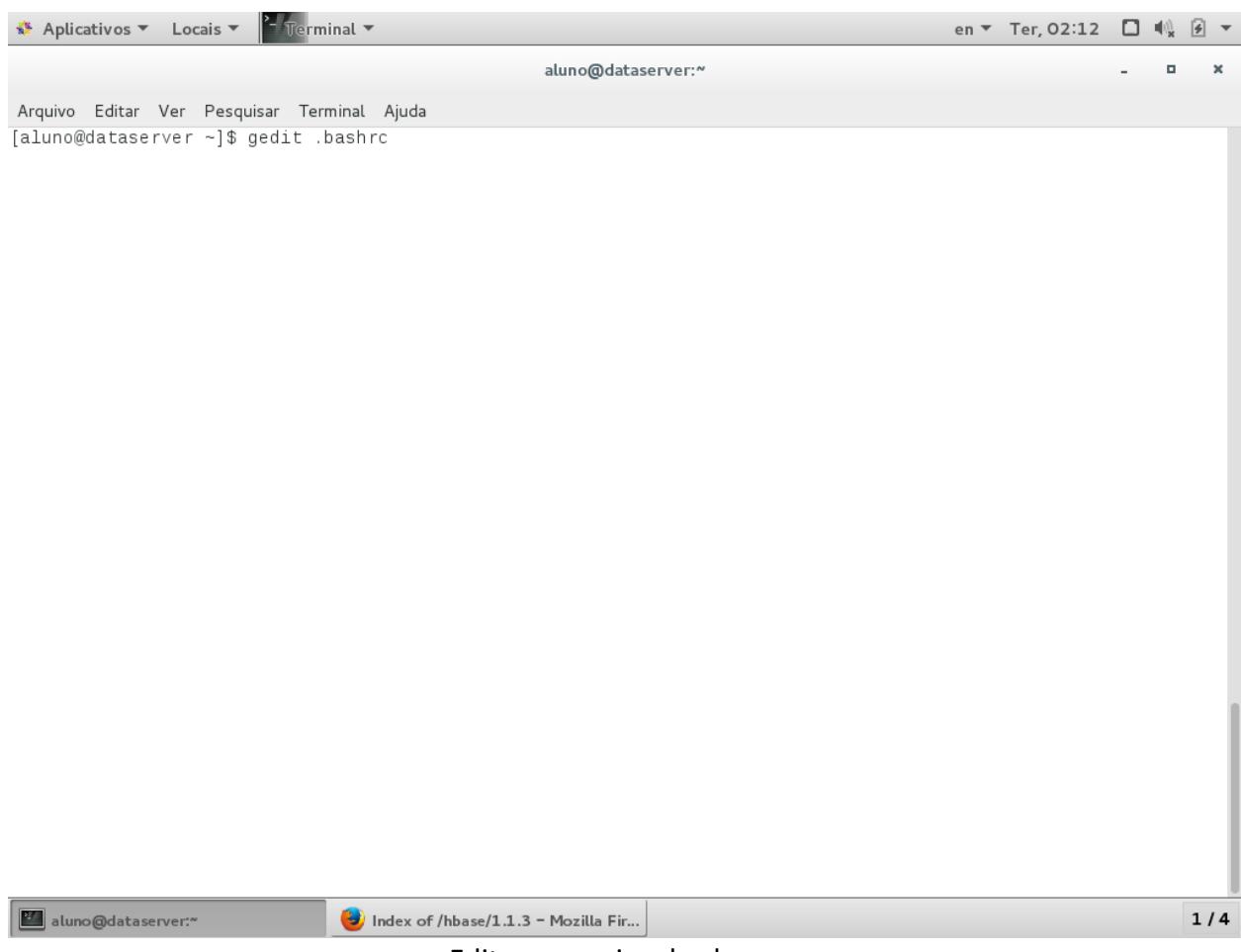
Applications ▾ Places ▾ gedit ▾
Open ⌂ Save Fri 17:08
hbase-site.xml /opt/hbase/conf

<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
/***
*
* Licensed to the Apache Software Foundation (ASF) under one
* or more contributor license agreements. See the NOTICE file
* distributed with this work for additional information
* regarding copyright ownership. The ASF licenses this file
* to you under the Apache License, Version 2.0 (the
* "License"); you may not use this file except in compliance
* with the License. You may obtain a copy of the License at
*
*     http://www.apache.org/licenses/LICENSE-2.0
*
* Unless required by applicable law or agreed to in writing, software
* distributed under the License is distributed on an "AS IS" BASIS,
* WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
* See the License for the specific language governing permissions and
* limitations under the License.
*/
-->
<configuration>
<property>
<name>hbase.rootdir</name>
<value>file:///opt/hbase/hfiles</value>
</property>
<property>
<name>hbase.zookeeper.property.dataDir</name>
<value>/opt/zookeeper/data</value>
</property>
</configuration>

```



Incluir as linhas entre as tags <configuration>



Editar o arquivo .bashrc

1 / 4

```

#.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
export JRE_HOME=/opt/jre
export JAVA_HOME=/opt/jdk
export PATH=$PATH:$JAVA_HOME/bin:$JRE_HOME/bin

# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Zookeeper
export ZOOKEEPER_HOME=/opt/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin

# HBase
export HBASE_HOME=/opt/hbase
export PATH=$PATH:$HBASE_HOME/bin

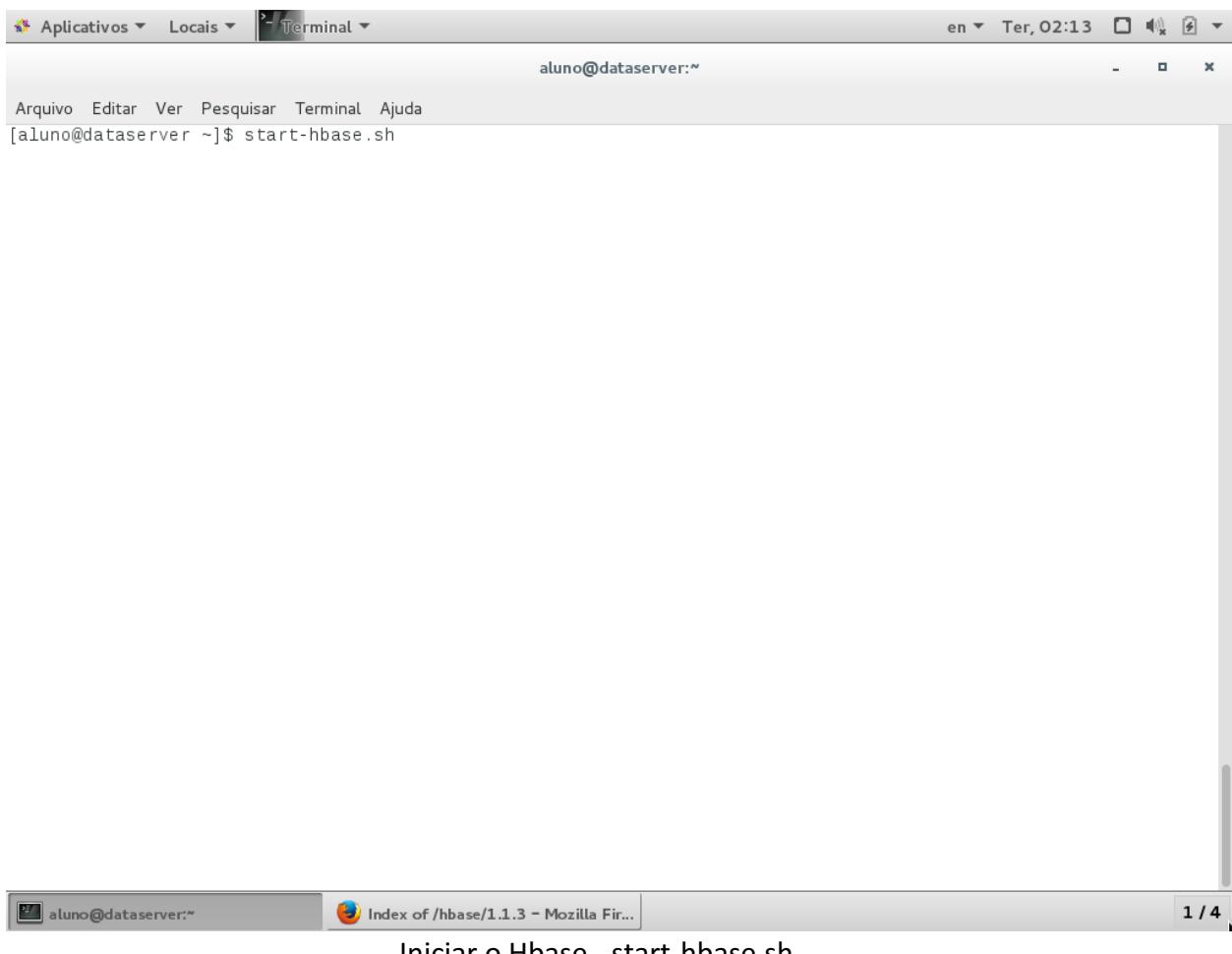
```

Variáveis HBase

The screenshot shows a Linux desktop environment. At the top, there is a panel with icons for 'Aplicativos' (Applications), 'Locais' (Locations), and a 'Terminal' icon. The terminal window is open and titled 'Terminal'. It displays the command line interface with the user 'aluno' at 'datalogger:~'. The commands entered are:

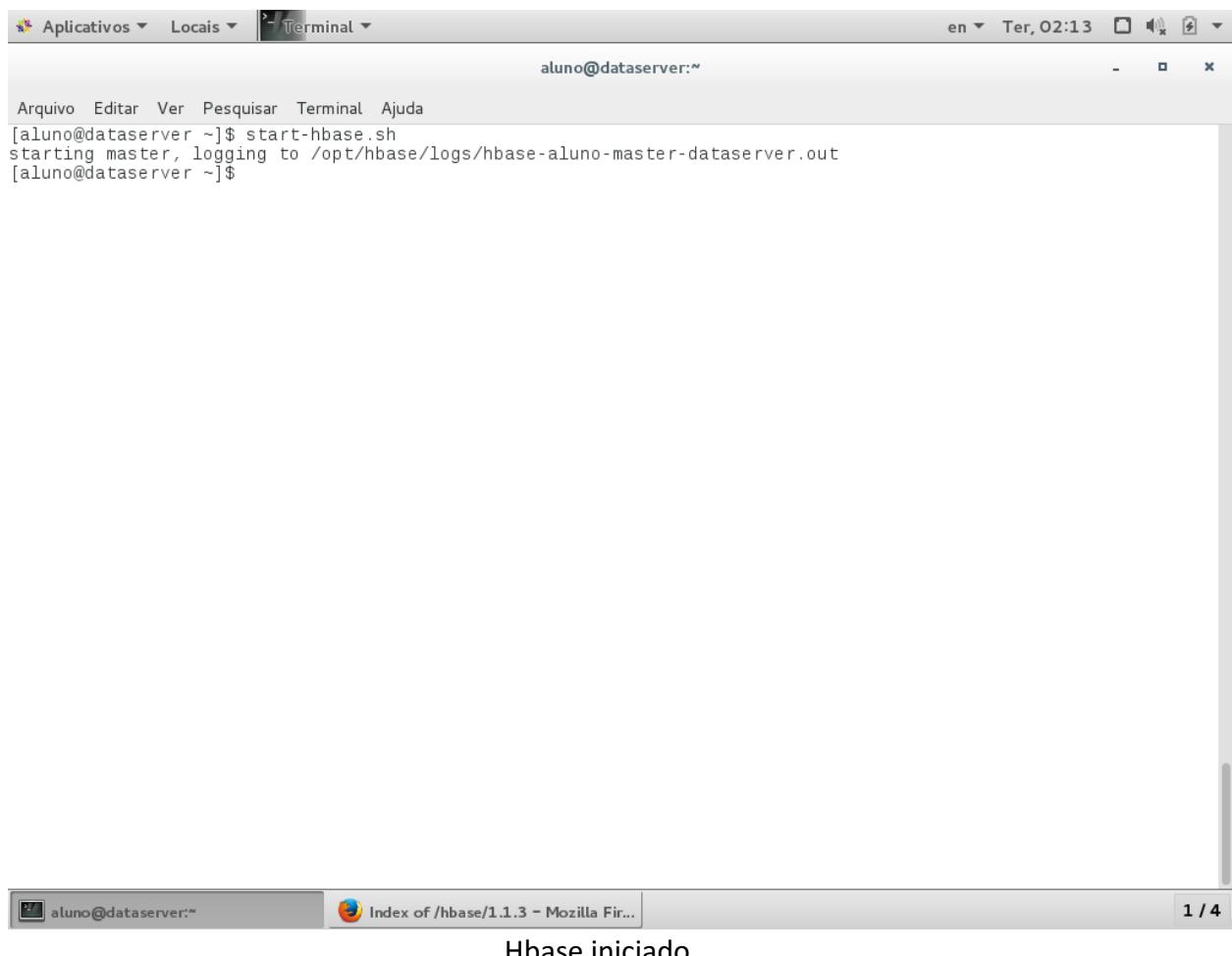
```
aluno@datalogger:~$ gedit .bashrc
[aluno@datalogger ~]$ source .bashrc
[aluno@datalogger ~]$
```

Below the terminal, there is a Firefox browser window showing the 'Index of /hbase/1.1.3 - Mozilla Fir...' page. The browser's address bar also shows the command 'source .bashrc'. The status bar at the bottom of the browser window indicates '1 / 4'.



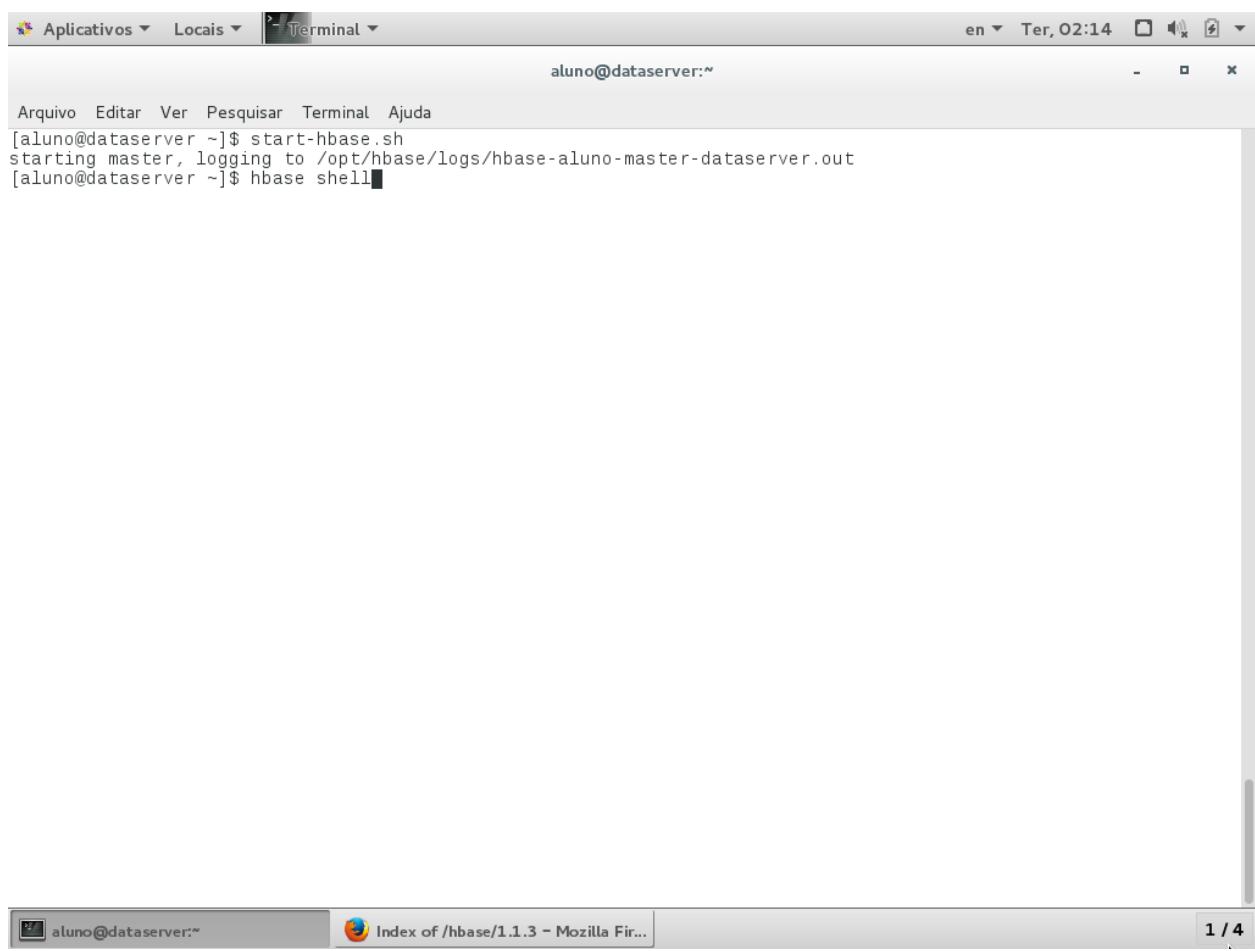
Iniciar o Hbase - start-hbase.sh

1 / 4



```
aluno@dataserver:~$ start-hbase.sh
starting master, logging to /opt/hbase/logs/hbase-aluno-master-dataserver.out
[aluno@dataserver ~]$
```

Hbase iniciado

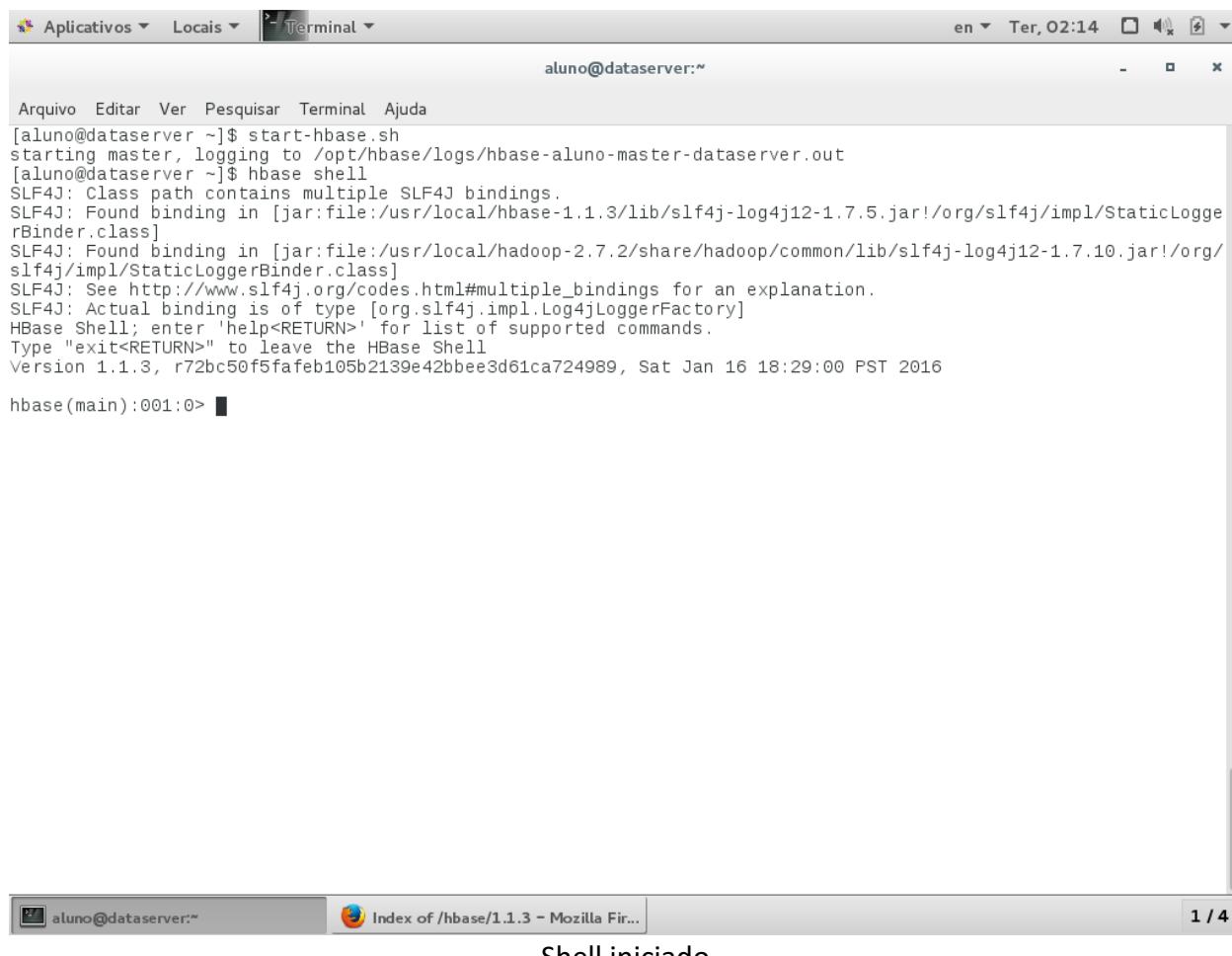


A screenshot of a Linux desktop environment showing a terminal window titled "Terminal". The window title bar also includes "Aplicativos" and "Locais". The status bar at the top right shows "en Ter, 02:14". The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command line shows:

```
aluno@dataserver:~$ start-hbase.sh
starting master, logging to /opt/hbase/logs/hbase-aluno-master-dataserver.out
[aluno@dataserver ~]$ hbase shell
```

Abrir o shell do Hbase

1 / 4



A screenshot of a terminal window titled "Terminal". The window shows the command "start-hbase.sh" being run, followed by the "hbase shell" command. The output includes logs about SLF4J binding and the HBase Shell prompt. The terminal window has a standard OS X interface with a menu bar and status bar.

```
[aluno@dataserver ~]$ start-hbase.sh
starting master, logging to /opt/hbase/logs/hbase-aluno-master-dataserver.out
[aluno@dataserver ~]$ hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase-1.1.3/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.2/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.1.3, r72bc50f5fafec105b2139e42bbe3d61ca724989, Sat Jan 16 18:29:00 PST 2016
hbase(main):001:0> ■
```

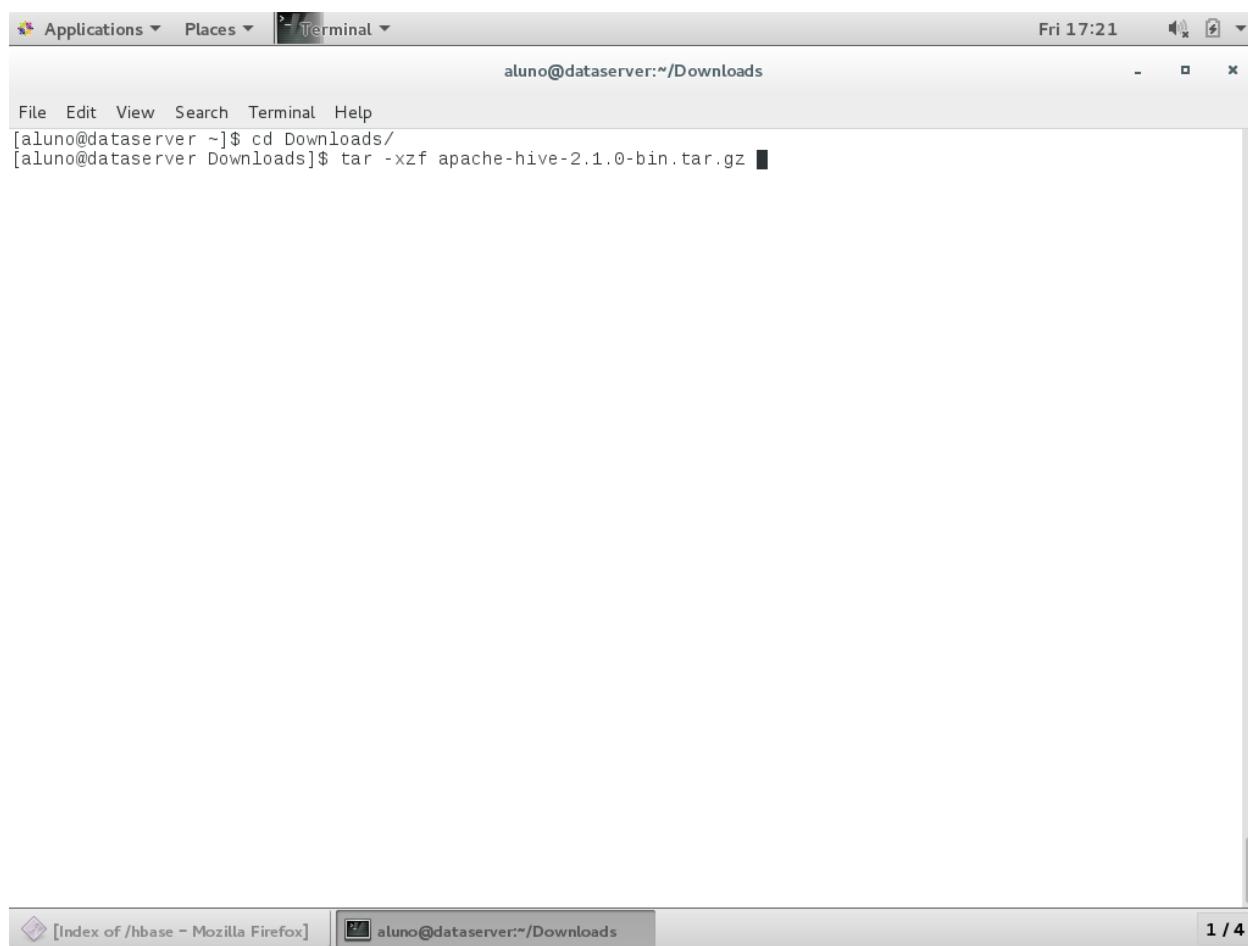
Shell iniciado

8. Instalação e Configuração do Hive

8.1. Download e Instalação do Hive

The screenshot shows a Firefox browser window with the title "Apache Hive TM - Mozilla Firefox". The address bar displays the URL <https://hive.apache.org/index.html>. The page content is the Apache Hive homepage, featuring a yellow bee logo with the word "HIVE" below it. The main heading is "APACHE HIVE TM". A descriptive text block explains that the Apache Hive™ data warehouse software facilitates reading, writing, and managing large datasets using SQL. Below this, sections include "Getting Started With Apache Hive Software" with a bulleted list of links, and "Getting Involved With The Apache Hive Community" with a paragraph about the project's history and contribution guidelines. The left sidebar contains navigation links for "GENERAL" (Home, Downloads, License, Privacy Policy), "DOCUMENTATION" (Language Manual, Javadoc, Wiki), "COMMUNITY" (Becoming a Committer, Edit Website, How to Contribute, Resources for contributors, Issue Tracking, Mailing Lists, People), and "DEVELOPMENT" (Builds, Design Docs, FAQ, Hive JIRA). The bottom right corner of the browser window shows "1 / 4".

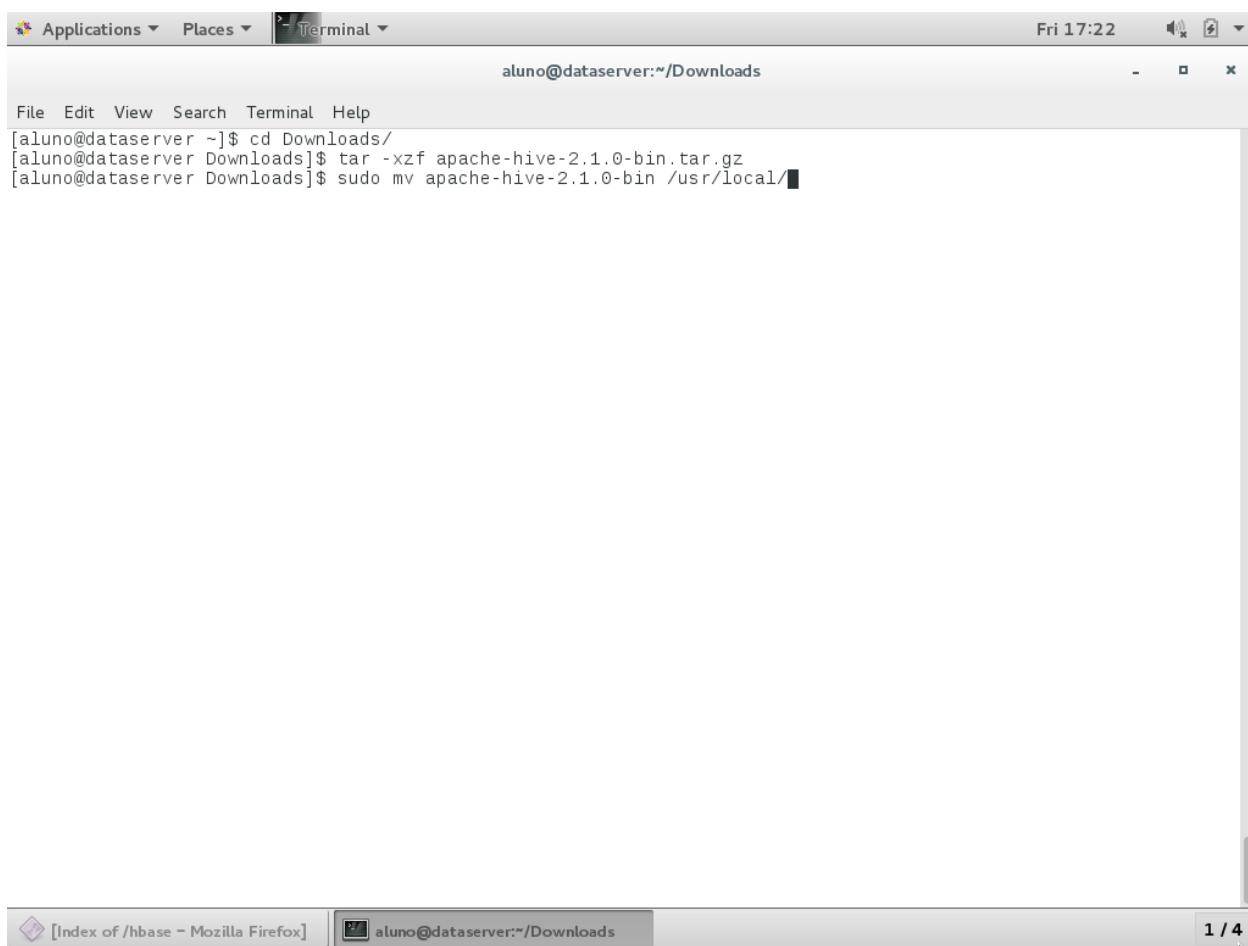
Download do Hive – Versão 2.1.0



```
File Edit View Search Terminal Help
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ tar -xzf apache-hive-2.1.0-bin.tar.gz
```

Descompactando o arquivo

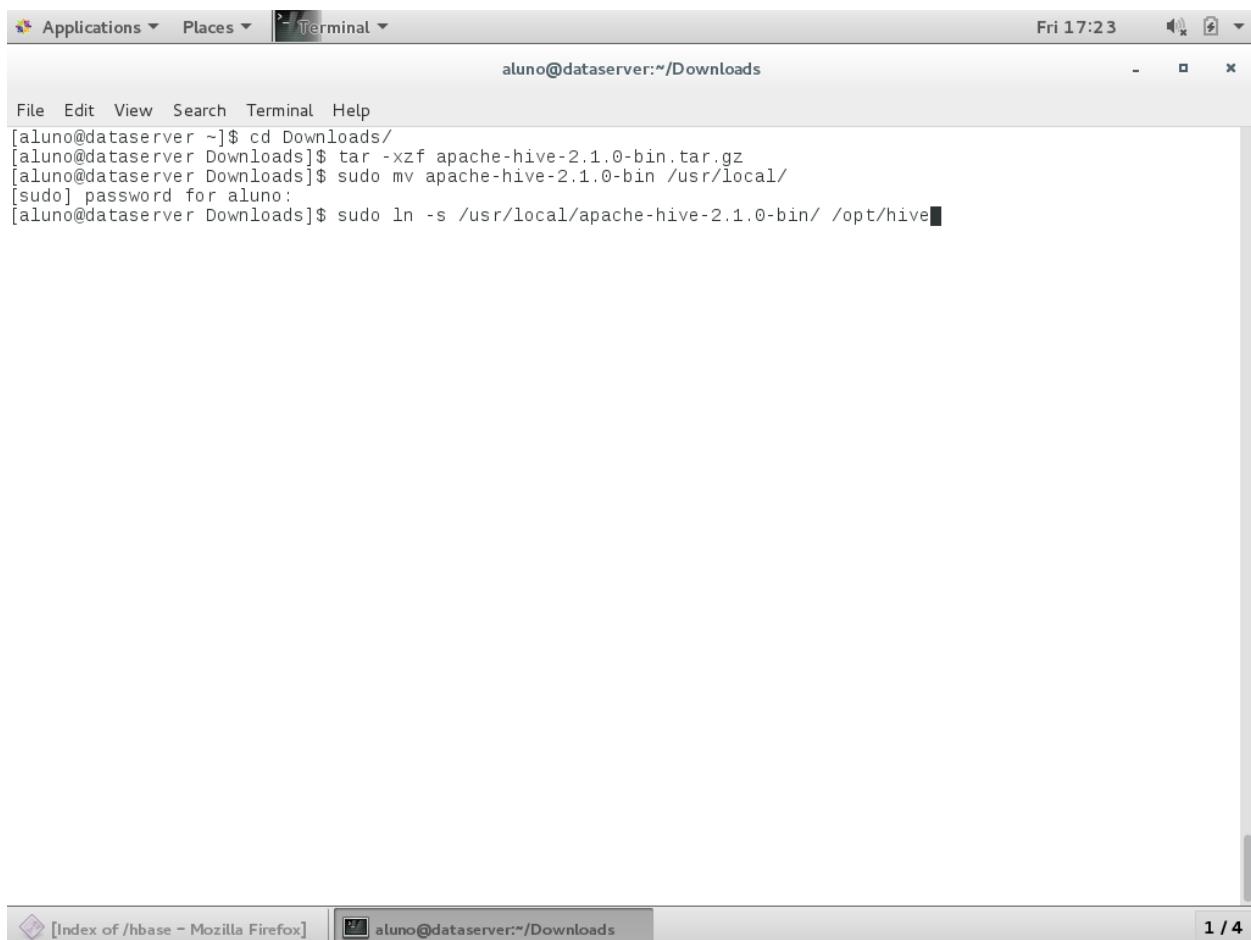
1 / 4



```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ tar -xzf apache-hive-2.1.0-bin.tar.gz
[aluno@dataserver Downloads]$ sudo mv apache-hive-2.1.0-bin /usr/local/
```

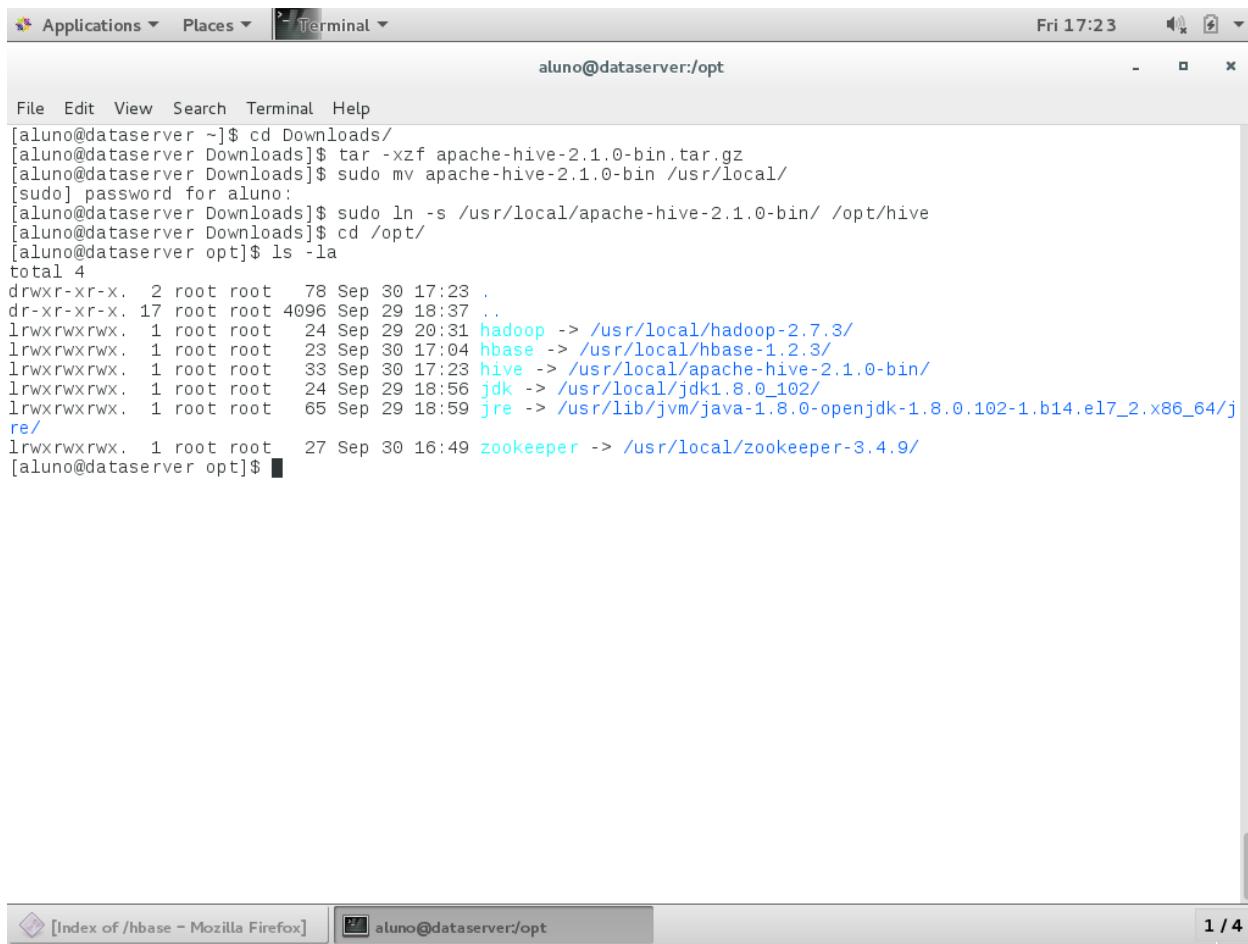
Movendo o diretório do Hive para /usr/local

1 / 4



```
[aluno@dataserver ~]$ cd Downloads/
[aluno@dataserver Downloads]$ tar -xzf apache-hive-2.1.0-bin.tar.gz
[aluno@dataserver Downloads]$ sudo mv apache-hive-2.1.0-bin /usr/local/
[sudo] password for aluno:
[aluno@dataserver Downloads]$ sudo ln -s /usr/local/apache-hive-2.1.0-bin/ /opt/hive
```

Criando o link simbólico para o Hive no diretório /opt



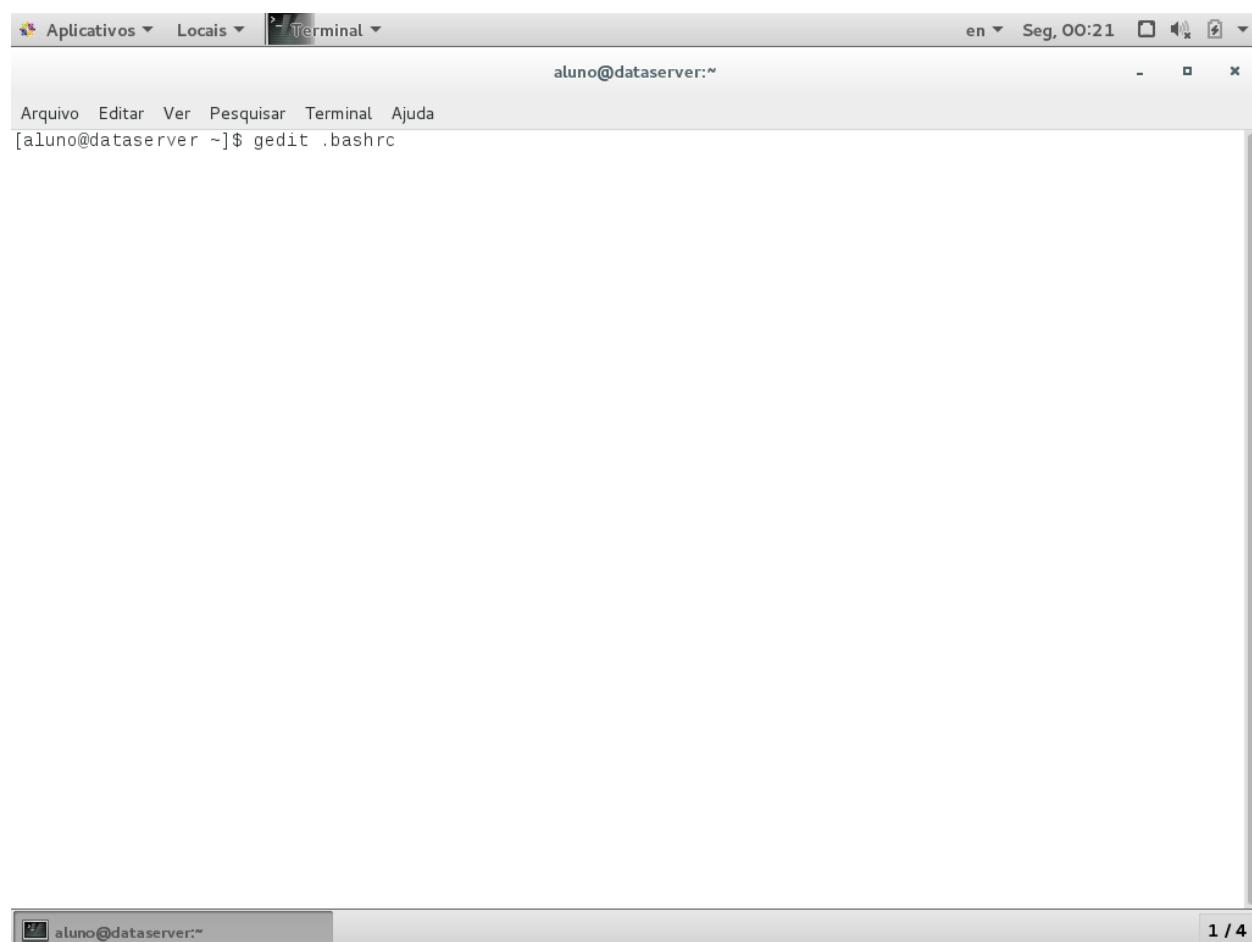
The screenshot shows a terminal window titled "Terminal" with the command-line interface. The session starts with the user navigating to the Downloads directory and extracting the apache-hive-2.1.0-bin.tar.gz file. Then, the user uses sudo to move the extracted files to /usr/local/. Finally, the user creates symbolic links for hadoop, hbase, hive, jdk, and jre in the /opt/hive directory. The terminal also shows the user's password being entered.

```
[aluno@dataserver ~]$ cd Downloads/  
[aluno@dataserver Downloads]$ tar -xzf apache-hive-2.1.0-bin.tar.gz  
[aluno@dataserver Downloads]$ sudo mv apache-hive-2.1.0-bin /usr/local/  
[sudo] password for aluno:  
[aluno@dataserver Downloads]$ sudo ln -s /usr/local/apache-hive-2.1.0-bin/ /opt/hive  
[aluno@dataserver Downloads]$ cd /opt/  
[aluno@dataserver opt]$ ls -la  
total 4  
drwxr-xr-x. 2 root root 78 Sep 30 17:23 .  
dr-xr-xr-x. 17 root root 4096 Sep 29 18:37 ..  
lrwxrwxrwx. 1 root root 24 Sep 29 20:31 hadoop -> /usr/local/hadoop-2.7.3/  
lrwxrwxrwx. 1 root root 23 Sep 30 17:04 hbase -> /usr/local/hbase-1.2.3/  
lrwxrwxrwx. 1 root root 33 Sep 30 17:23 hive -> /usr/local/apache-hive-2.1.0-bin/  
lrwxrwxrwx. 1 root root 24 Sep 29 18:56 jdk -> /usr/local/jdk1.8.0_102/  
lrwxrwxrwx. 1 root root 65 Sep 29 18:59 jre -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/  
[aluno@dataserver opt]$
```

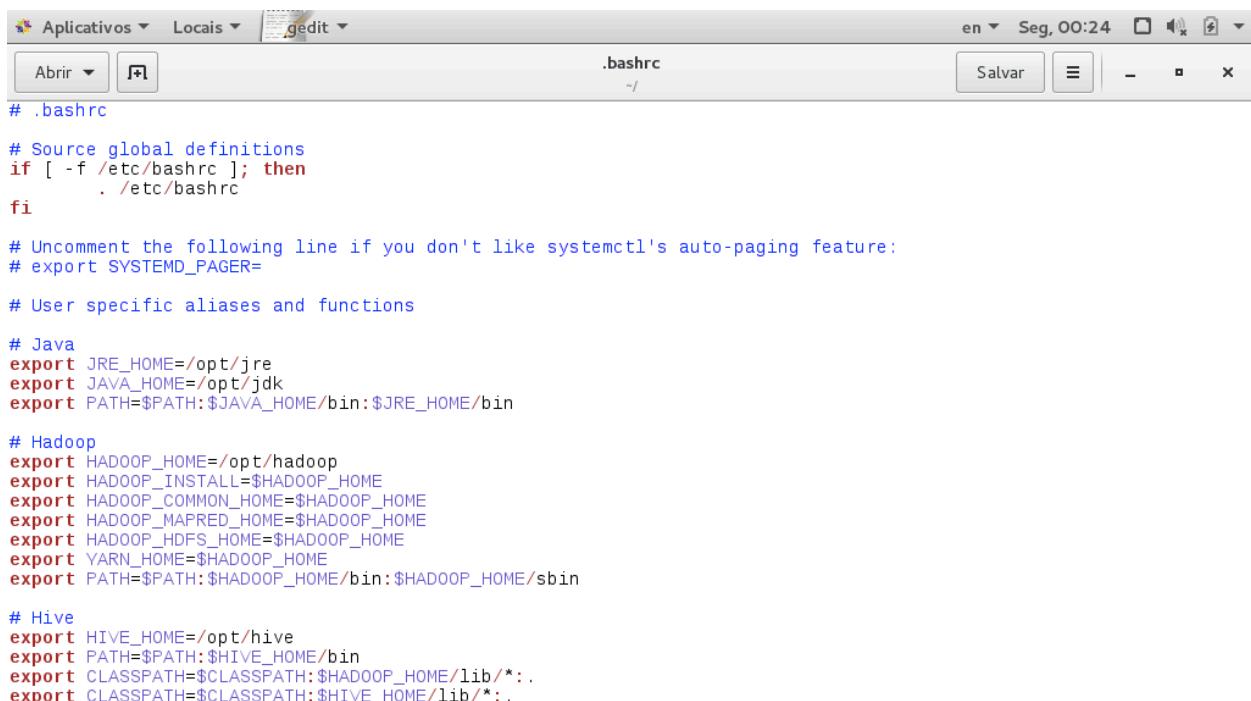
Link criado

1 / 4

8.2. Configurando o Hive



Editando o arquivo .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
export JRE_HOME=/opt/jre
export JAVA_HOME=/opt/jdk
export PATH=$PATH:$JAVA_HOME/bin:$JRE_HOME/bin

# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Hive
export HIVE_HOME=/opt/hive
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:.
export CLASSPATH=$CLASSPATH:$HIVE_HOME/lib/*:.

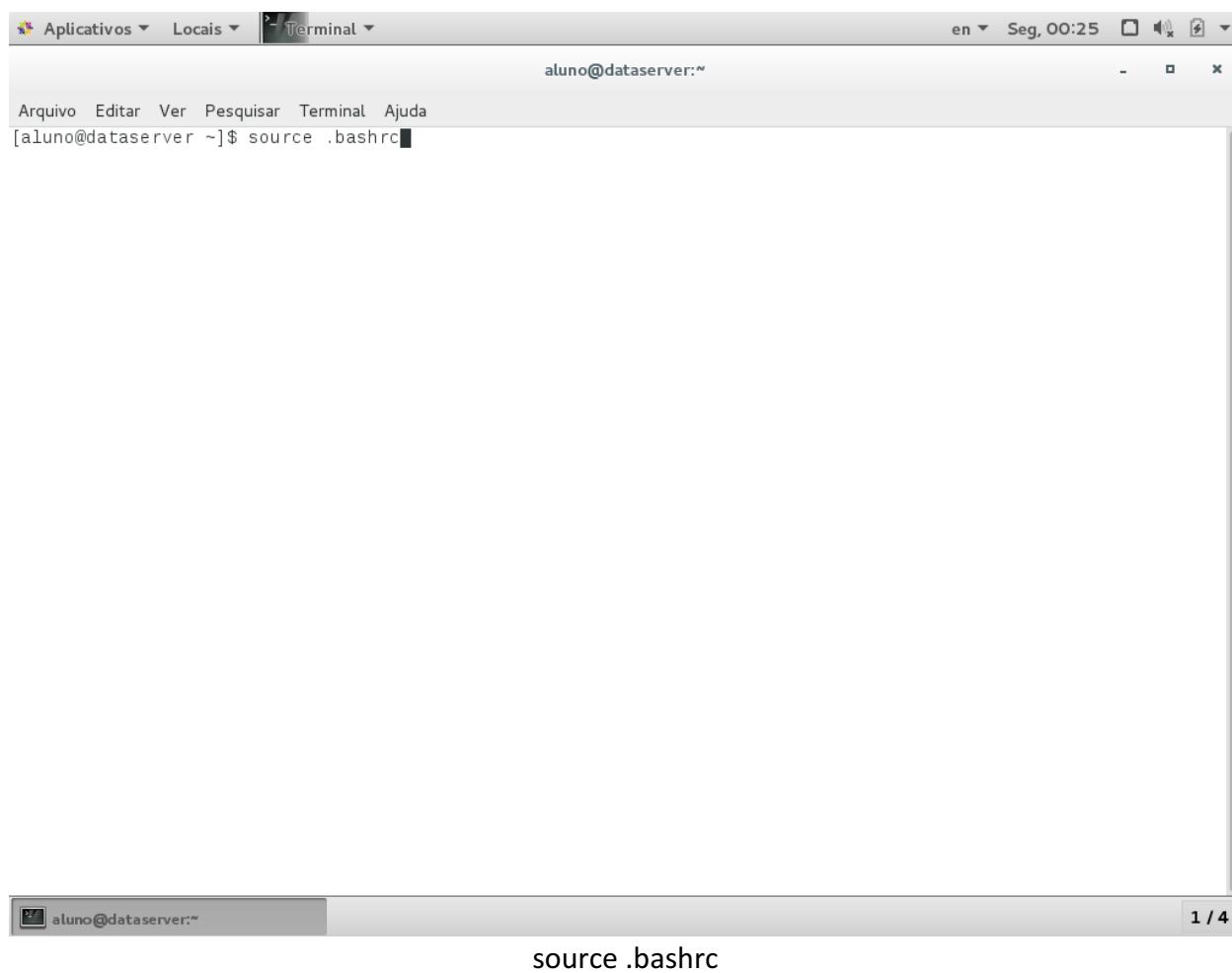
```



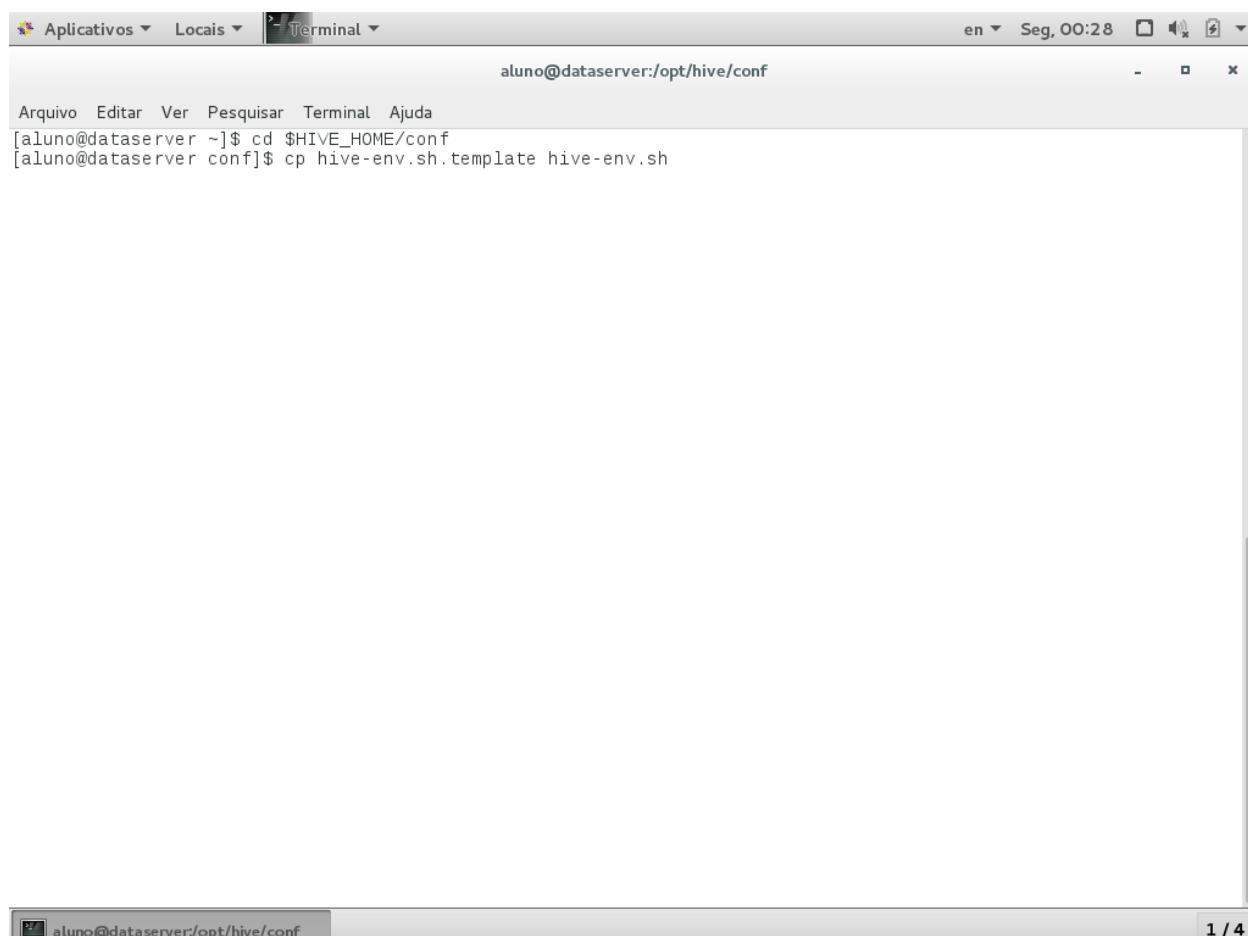
sh ▾ Largura da tabulação: 8 ▾ Lin 32, Col 1 ▾ INS
 aluno@dataserver:~ | .bashrc (~/) - gedit 1 / 4

Variáveis de ambiente do Hive

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top shows "en Seg, 00:25". The terminal prompt is "aluno@dataserver:~". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". A command line input field contains the text "[aluno@dataserver ~]\$ source .bashrc". The bottom status bar shows the user icon, the terminal name, and "1 / 4".



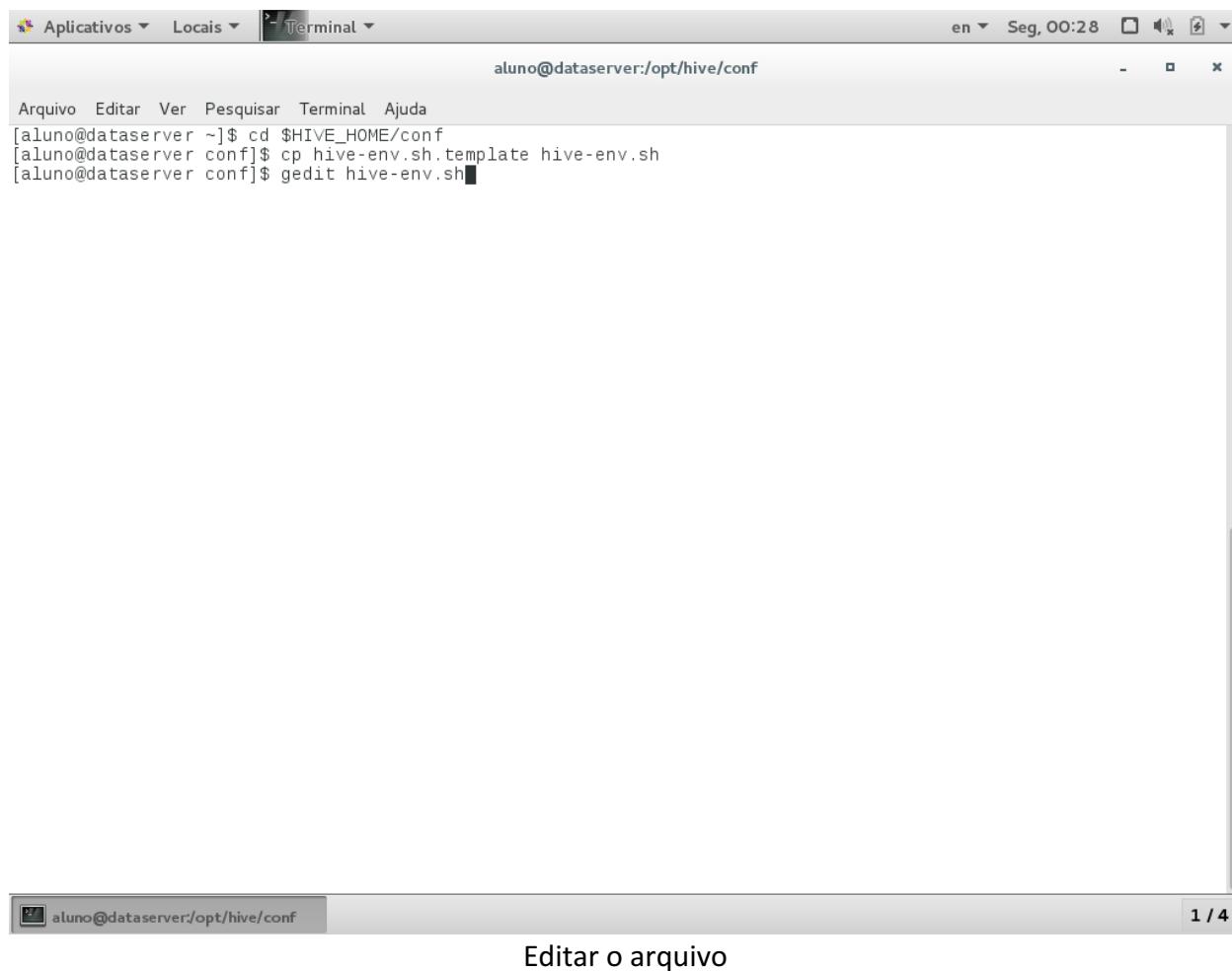
A screenshot of a terminal window titled "Terminal". The window shows the command line interface of a Linux system. The user is in the directory "/opt/hive/conf". The command entered is "cp hive-env.sh.template hive-env.sh". The terminal window has a standard OS X-style title bar with icons for application, location, and window control.

```
aluno@dataserver:~/opt/hive/conf
[aluno@dataserver ~]$ cd $HIVE_HOME/conf
[aluno@dataserver conf]$ cp hive-env.sh.template hive-env.sh
```

A partir do arquivo template, gerar o arquivo hive-env.sh

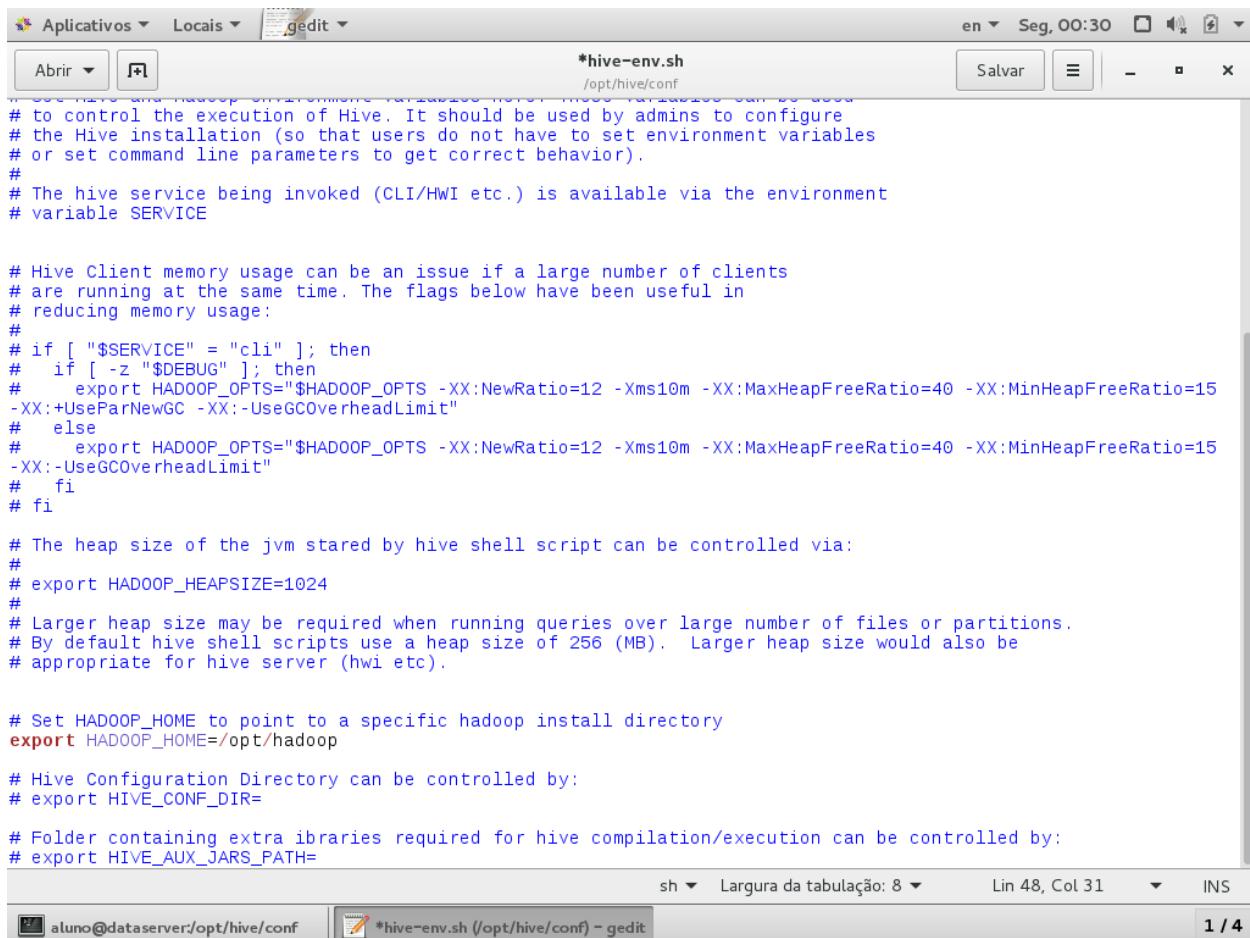
1 / 4

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a terminal window titled "Terminal". The window shows a command-line session on a Linux system. The user, "aluno", is navigating to the Hive configuration directory and copying a template file to a new file. The terminal interface includes a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom indicates "aluno@dataserver:~/opt/hive/conf" and "1 / 4". A button labeled "Editar o arquivo" is visible at the bottom left of the terminal window.

```
[aluno@dataserver ~]$ cd $HIVE_HOME/conf  
[aluno@dataserver conf]$ cp hive-env.sh.template hive-env.sh  
[aluno@dataserver conf]$ gedit hive-env.sh
```



```

Aplicativos Locais gedit
*hive-env.sh
/opt/hive/conf
en ▾ Seg, 00:30 □ 🔍 ⌂ Salvar
# to control the execution of Hive. It should be used by admins to configure
# the Hive installation (so that users do not have to set environment variables
# or set command line parameters to get correct behavior).
#
# The hive service being invoked (CLI/HWI etc.) is available via the environment
# variable SERVICE

# Hive Client memory usage can be an issue if a large number of clients
# are running at the same time. The flags below have been useful in
# reducing memory usage:
#
# if [ "$SERVICE" = "cli" ]; then
#   if [ -z "$DEBUG" ]; then
#     export HADOOP_OPTS="$HADOOP_OPTS -XX:NewRatio=12 -Xms10m -XX:MaxHeapFreeRatio=40 -XX:MinHeapFreeRatio=15
# -XX:+UseParNewGC -XX:-UseGCOverheadLimit"
#   else
#     export HADOOP_OPTS="$HADOOP_OPTS -XX:NewRatio=12 -Xms10m -XX:MaxHeapFreeRatio=40 -XX:MinHeapFreeRatio=15
# -XX:-UseGCOverheadLimit"
#   fi
# fi

# The heap size of the jvm started by hive shell script can be controlled via:
#
# export HADOOP_HEAPSIZE=1024
#
# Larger heap size may be required when running queries over large number of files or partitions.
# By default hive shell scripts use a heap size of 256 (MB). Larger heap size would also be
# appropriate for hive server (hwi etc).

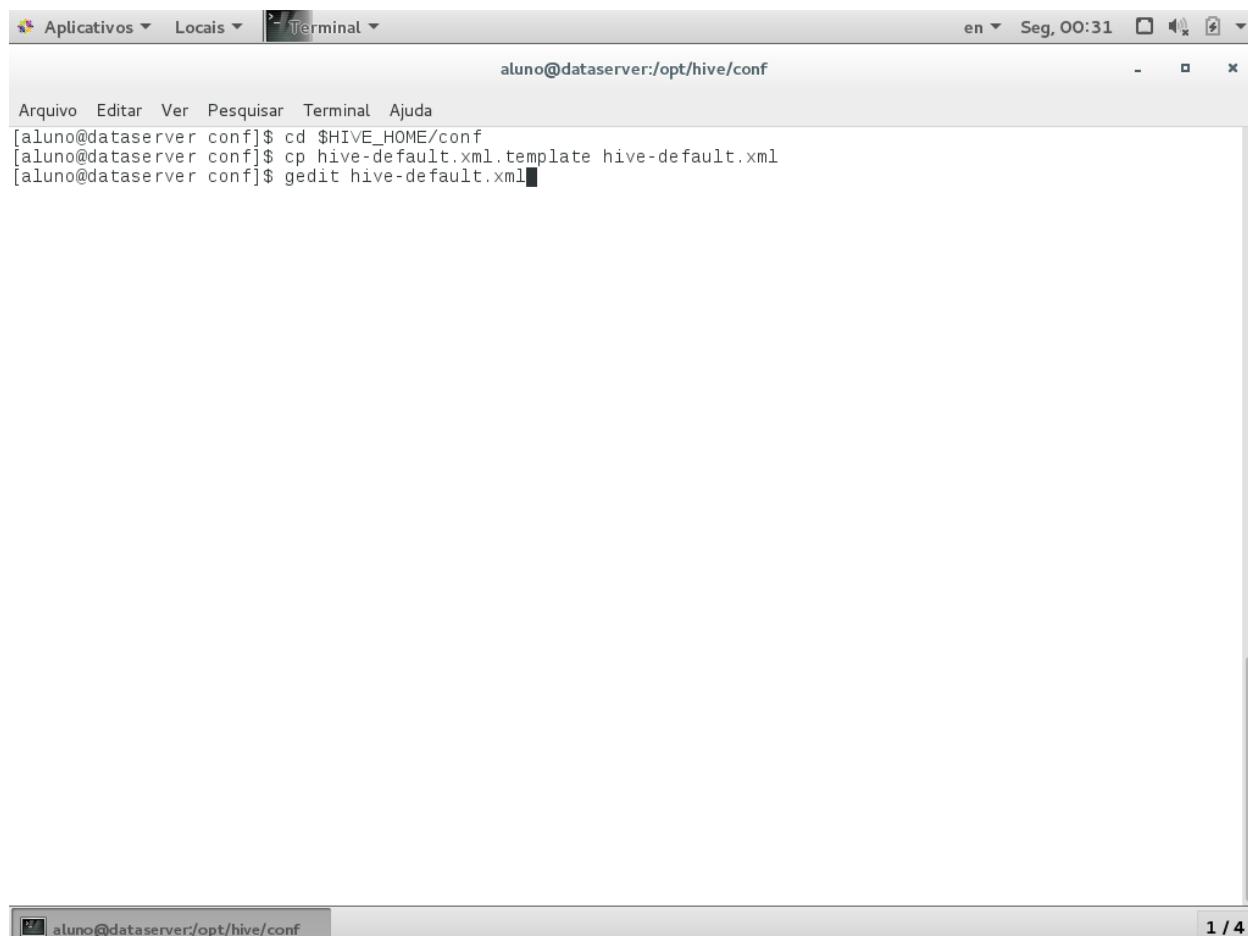
# Set HADOOP_HOME to point to a specific hadoop install directory
export HADOOP_HOME=/opt/hadoop

# Hive Configuration Directory can be controlled by:
# export HIVE_CONF_DIR=

# Folder containing extra libraries required for hive compilation/execution can be controlled by:
# export HIVE_AUX_JARS_PATH=
sh ▾ Largura da tabulação: 8 ▾ Lin 48, Col 31 ▾ INS
aluno@dataserver:/opt/hive/conf *hive-env.sh (/opt/hive/conf) - gedit 1 / 4

```

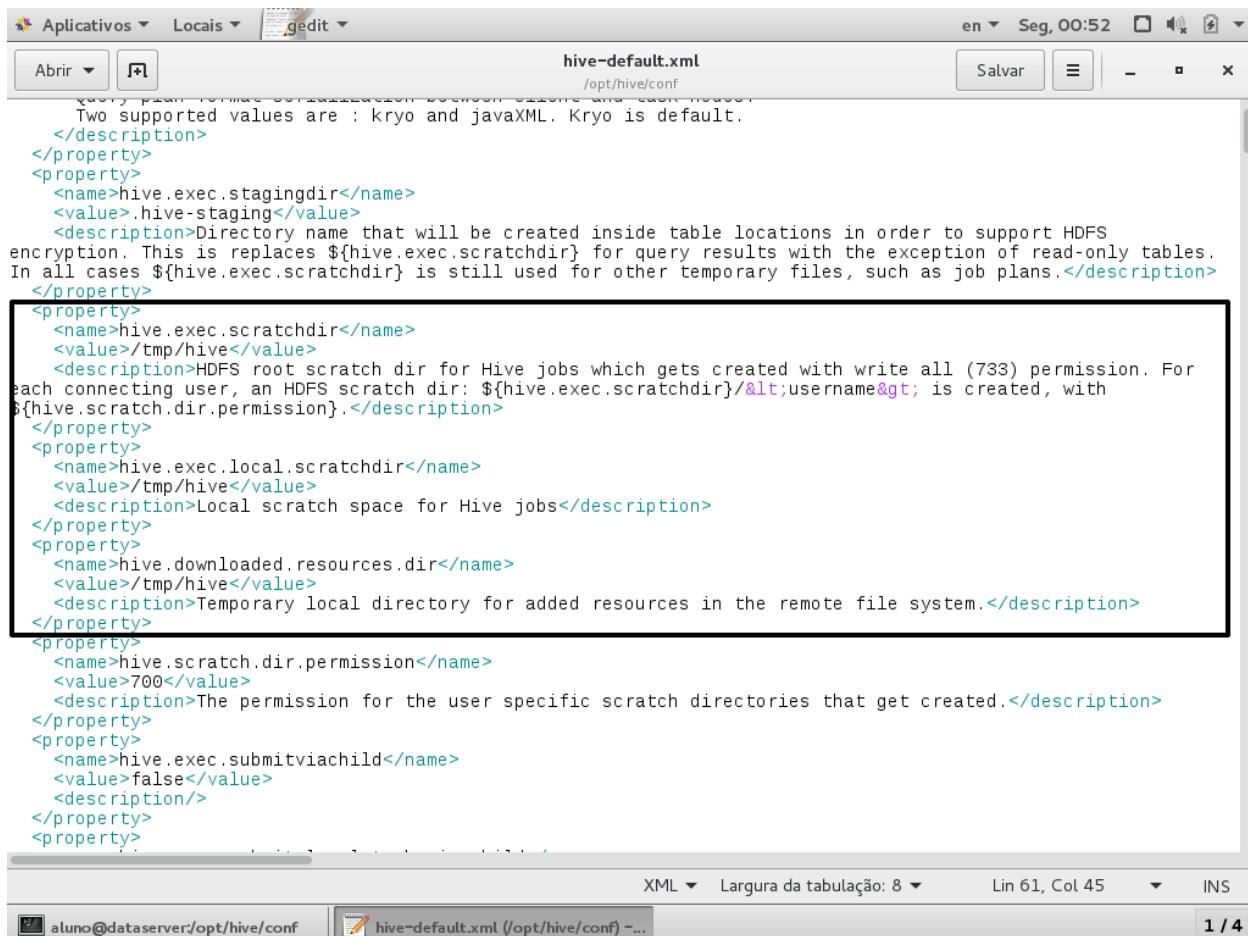
Incluir PATH do Hadoop, conforme tela acima



A screenshot of a Linux terminal window titled "Terminal". The window shows a command-line session where a user named "aluno" is navigating to the Hive configuration directory and copying a template XML file to a new one. The terminal window has a standard title bar with icons for applications, locations, and a search bar. The status bar at the bottom right shows the date and time as "Seg, 00:31". The terminal content is as follows:

```
aluno@dataserver:~$ cd $HIVE_HOME/conf  
aluno@dataserver conf]$ cp hive-default.xml.template hive-default.xml  
aluno@dataserver conf]$ gedit hive-default.xml
```

A partir do template, gerar o arquivo `hive-site.xml`

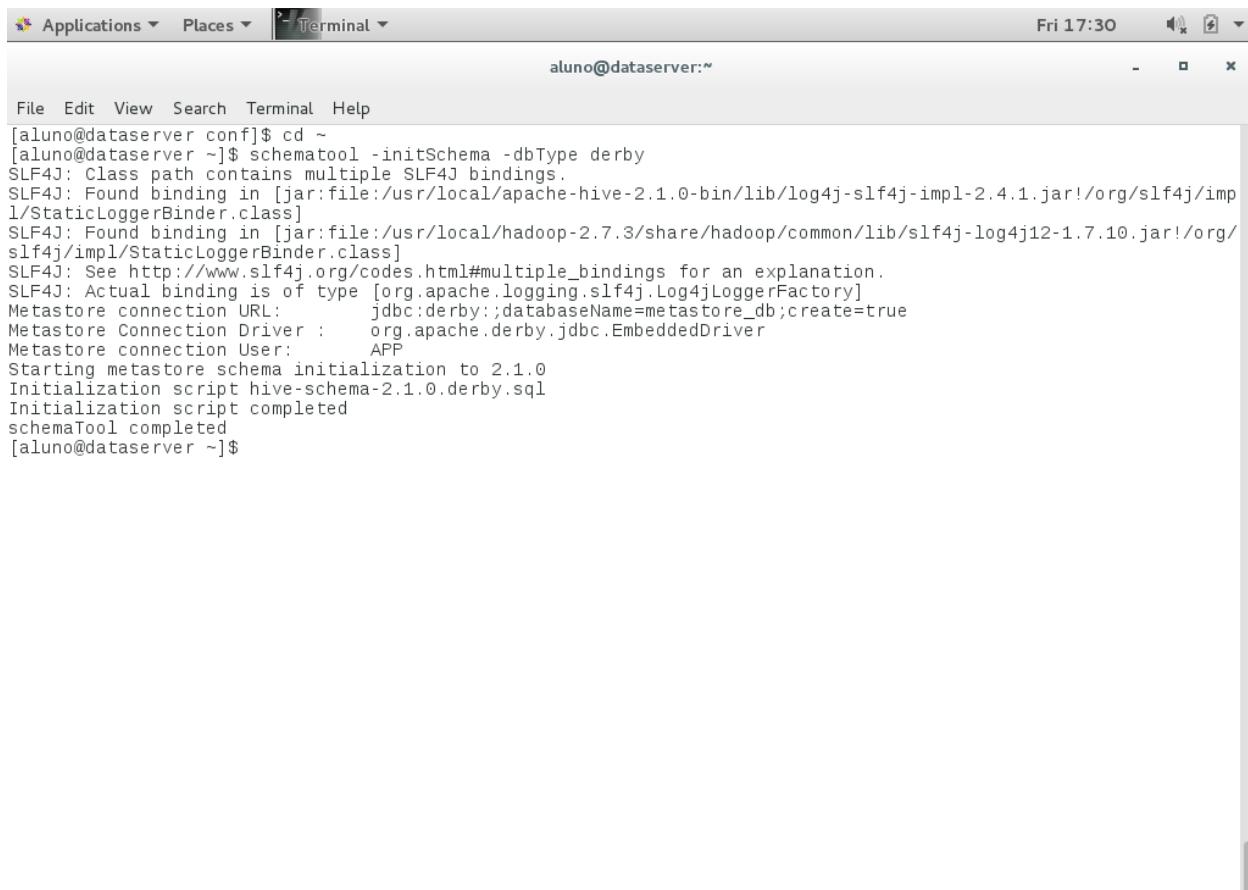


```

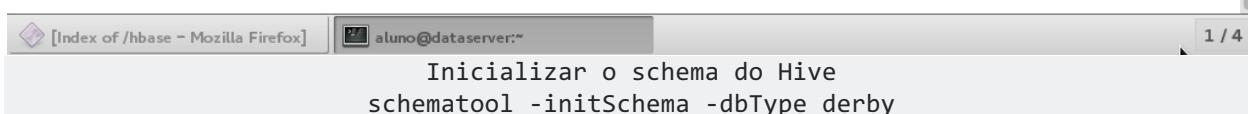
<!-- Two supported values are : kryo and javaXML. Kryo is default. -->
</description>
</property>
<property>
  <name>hive.exec.stagingdir</name>
  <value>.hive-staging</value>
  <description>Directory name that will be created inside table locations in order to support HDFS encryption. This is replaces ${hive.exec.scratchdir} for query results with the exception of read-only tables. In all cases ${hive.exec.scratchdir} is still used for other temporary files, such as job plans.</description>
</property>
<property>
  <name>hive.exec.scratchdir</name>
  <value>/tmp/hive</value>
  <description>HDFS root scratch dir for Hive jobs which gets created with write all (733) permission. For each connecting user, an HDFS scratch dir: ${hive.exec.scratchdir}/&lt;username&gt; is created, with ${hive.scratch.dir.permission}.</description>
</property>
<property>
  <name>hive.exec.local.scratchdir</name>
  <value>/tmp/hive</value>
  <description>Local scratch space for Hive jobs</description>
</property>
<property>
  <name>hive.downloaded.resources.dir</name>
  <value>/tmp/hive</value>
  <description>Temporary local directory for added resources in the remote file system.</description>
</property>
<property>
  <name>hive.scratch.dir.permission</name>
  <value>700</value>
  <description>The permission for the user specific scratch directories that get created.</description>
</property>
<property>
  <name>hive.exec.submitviachild</name>
  <value>false</value>
  <description/>
</property>
<property>

```

Editar as linhas conforme cima



```
[aluno@dataserver conf]$ cd ~
[aluno@dataserver ~]$ schematool -initSchema -dbType derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Metastore connection URL:      jdbc:derby:;databaseName=metastore_db;create=true
Metastore Connection Driver :  org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User:    APP
Starting metastore schema initialization to 2.1.0
Initialization script hive-schema-2.1.0.derby.sql
Initialization script completed
schemaTool completed
[aluno@dataserver ~]$
```

Index of /hbase - Mozilla Firefox

Iniciar o schema do Hive
schematool -initSchema -dbType derby



The screenshot shows a terminal window titled 'Terminal' with the user 'aluno@dataserver:~'. The terminal displays the output of the 'hive' command. It includes several SLF4J binding logs, a warning about Hive-on-MR being deprecated, and ends with a prompt 'hive>'. The window has standard Linux desktop icons at the top.

```
[aluno@dataserver ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/usr/local/apache-hive-2.1.0-bin/lib/hive-common-2.1.0.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

Executando o Hive (execute o comando jps para se certificar que o Hadoop está ativo)



The screenshot shows a terminal window titled 'Terminal' with the user 'aluno@dataserver:~'. The terminal displays the following output:

```
[aluno@dataserver ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/usr/local/apache-hive-2.1.0-bin/lib/hive-common-2.1.0.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show tables;
OK
Time taken: 1.352 seconds
hive>
```

O comando “show tables;” demonstra que o Hive foi instalado com sucesso

9. Instalação e Configuração do Pig

9.1. Download e Instalação do Pig

Welcome to Apache Pig! – Mozilla Firefox

Apache > Pig >

Welcome to Apache Pig!

News

- [Apache Pig 0.16.0 is released!](#)
- [Getting Started](#)
- [Getting Involved](#)

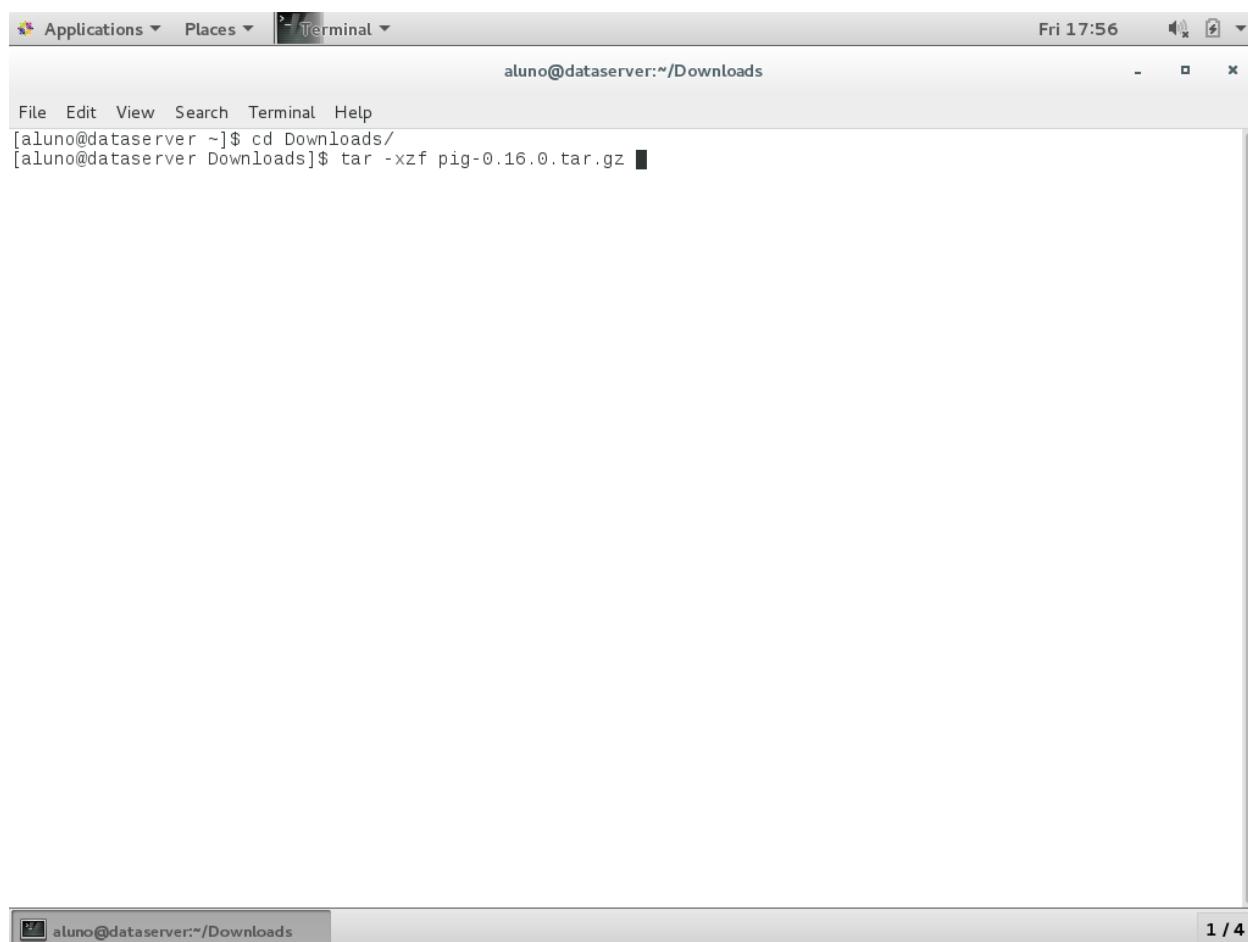
Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets.

At the present time, Pig's infrastructure layer consists of a compiler that produces sequences of Map-Reduce programs, for which large-scale parallel implementations already exist (e.g., the Hadoop subproject). Pig's language layer currently consists of a textual language called Pig Latin, which has the following key properties:

- **Ease of programming.** It is trivial to achieve parallel execution of simple, "embarrassingly parallel" data analysis tasks. Complex tasks comprised of multiple interrelated data transformations are explicitly encoded as data flow sequences, making them easy to write, understand, and maintain.
- **Optimization opportunities.** The way in which tasks are encoded permits the system to optimize their execution automatically, allowing the user to focus on semantics rather than

Download do Pig – Versão 0.16.0

1 / 4



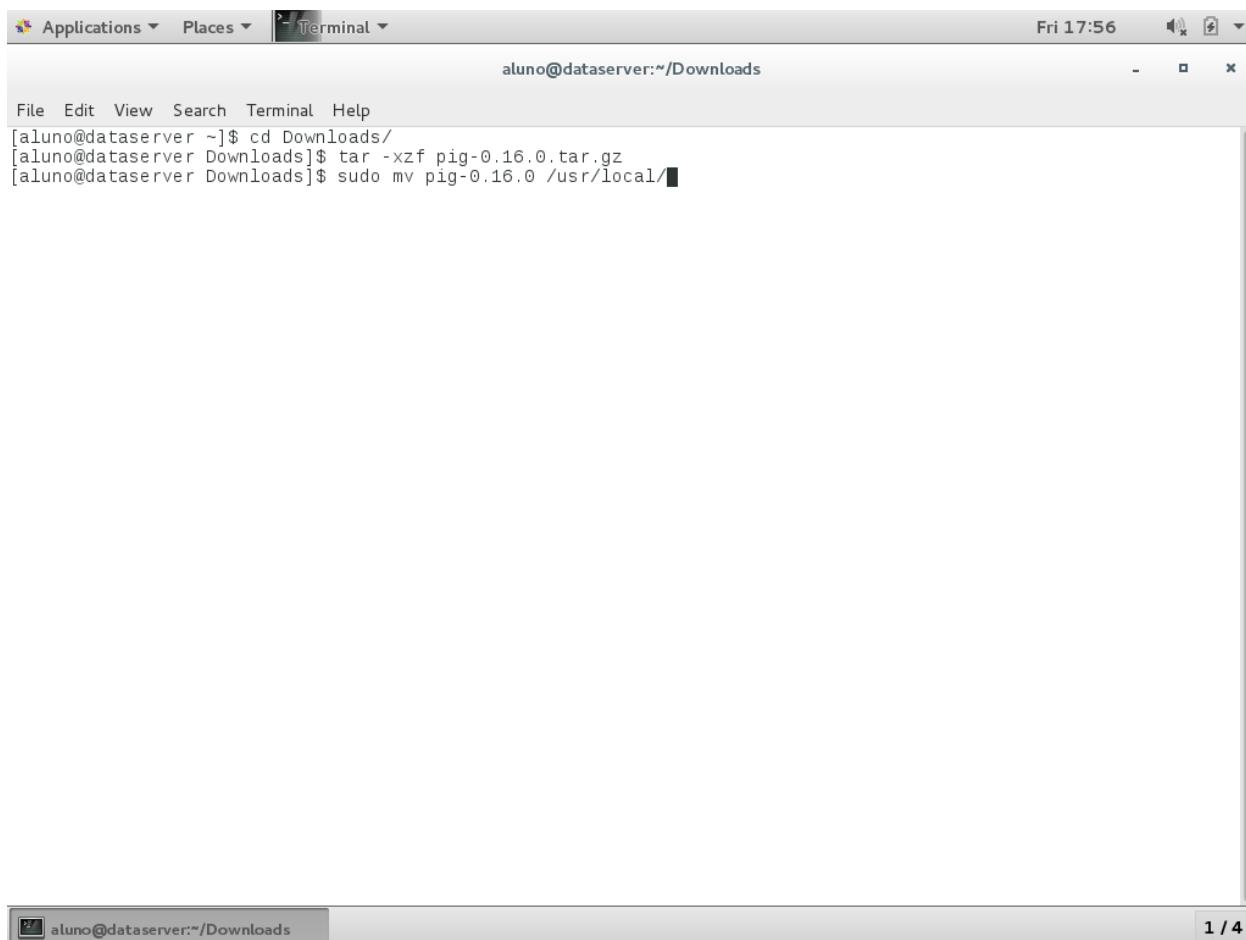
A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~/Downloads'. The user has run the command 'tar -xzf pig-0.16.0.tar.gz', which is shown in the terminal window. The terminal window is part of a larger desktop interface with a menu bar at the top.

```
[aluno@dataserver ~]$ cd Downloads/  
[aluno@dataserver Downloads]$ tar -xzf pig-0.16.0.tar.gz
```

Descompactando o arquivo

1 / 4

Instalação e Configuração do Ecosistema Hadoop

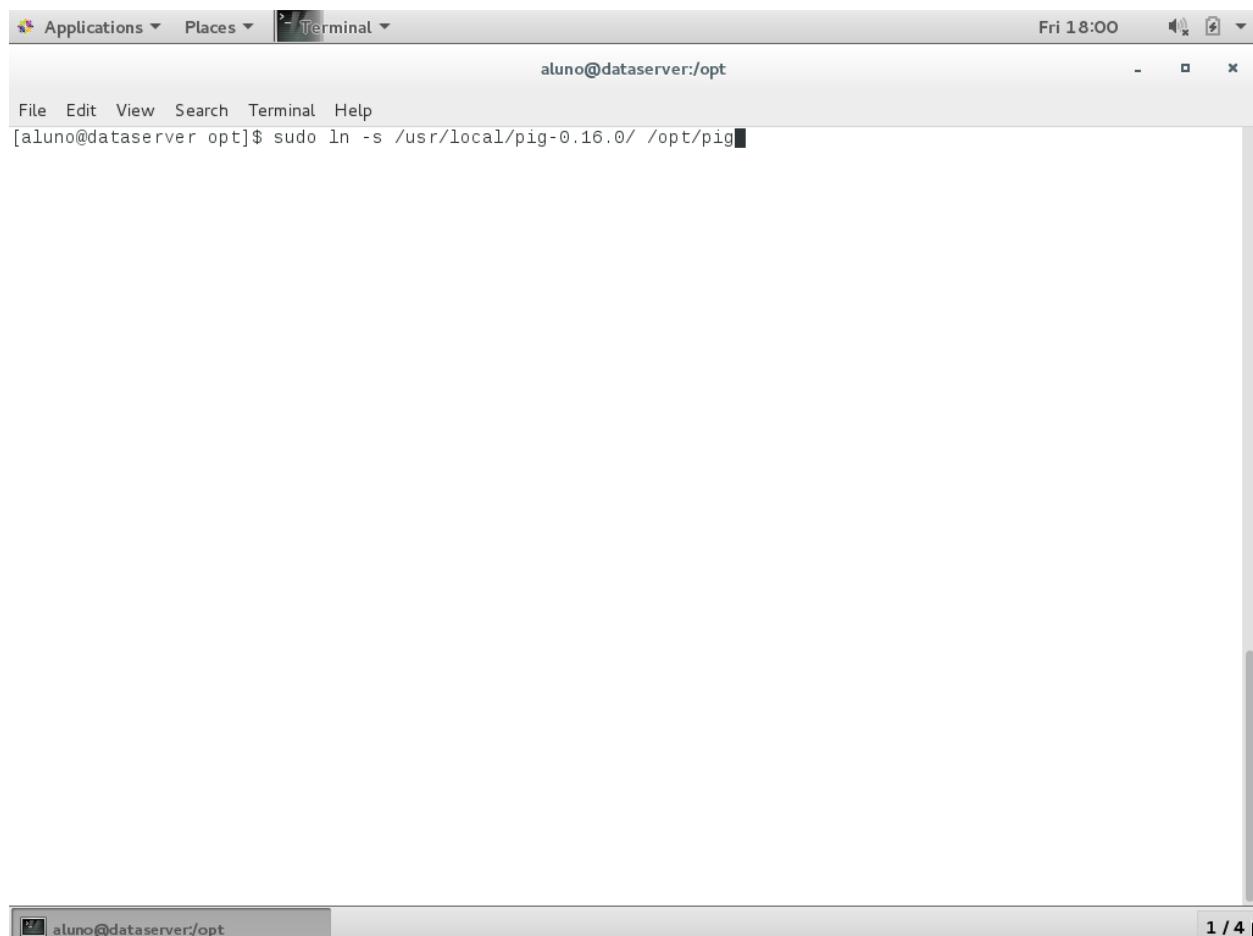


A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~/Downloads'. The user has run the following commands:

```
[aluno@dataserver ~]$ cd Downloads/  
[aluno@dataserver Downloads]$ tar -xzf pig-0.16.0.tar.gz  
[aluno@dataserver Downloads]$ sudo mv pig-0.16.0 /usr/local/■
```

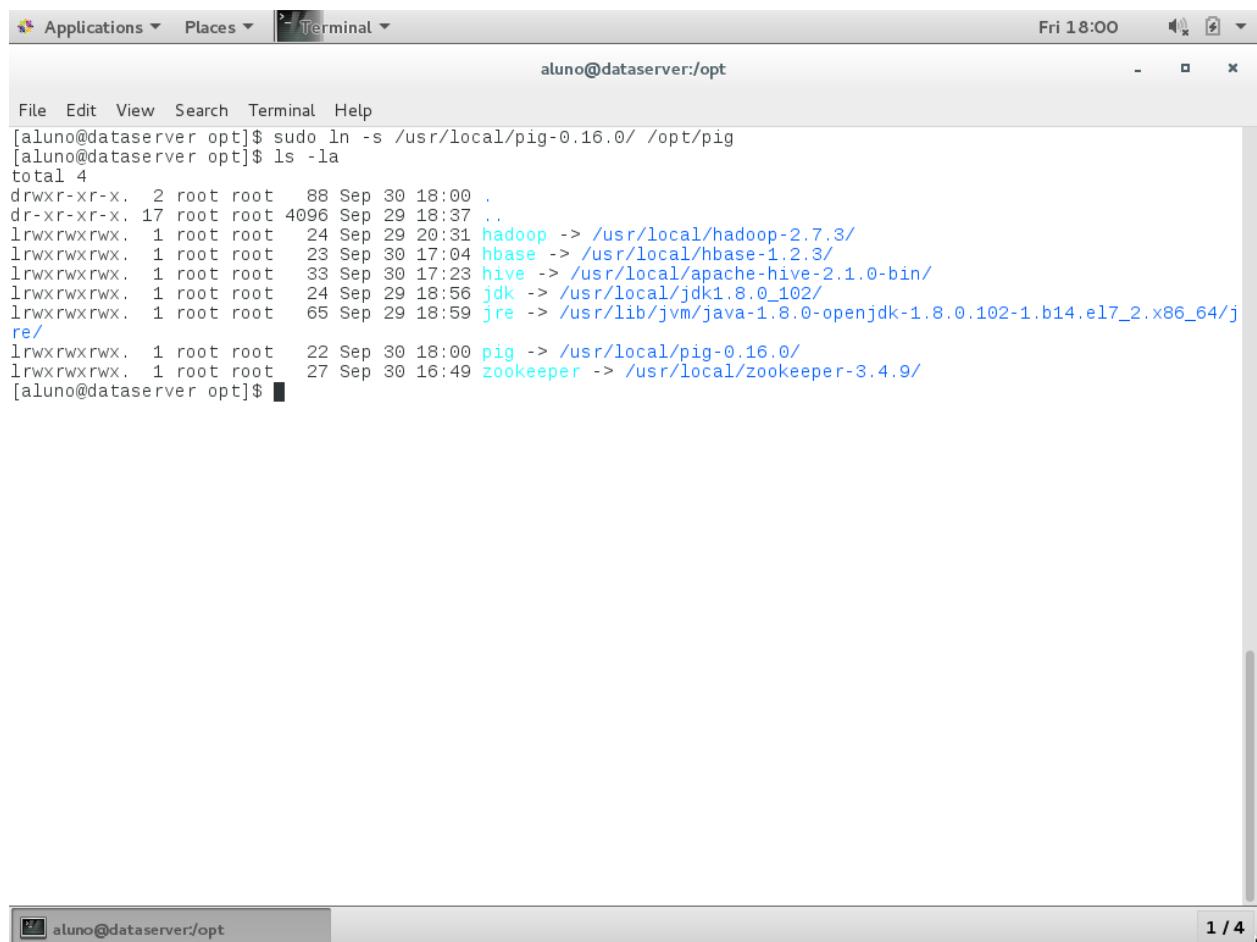
The terminal window has a scroll bar on the right side. At the bottom of the window, there is a status bar with the text 'aluno@dataserver:~/Downloads' on the left and '1 / 4' on the right.

Movendo a pasta do Pig para /usr/local



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~/opt". The user has typed the command "sudo ln -s /usr/local/pig-0.16.0/ /opt/pig" and is pressing the Enter key. The status bar at the bottom shows "1 / 4".

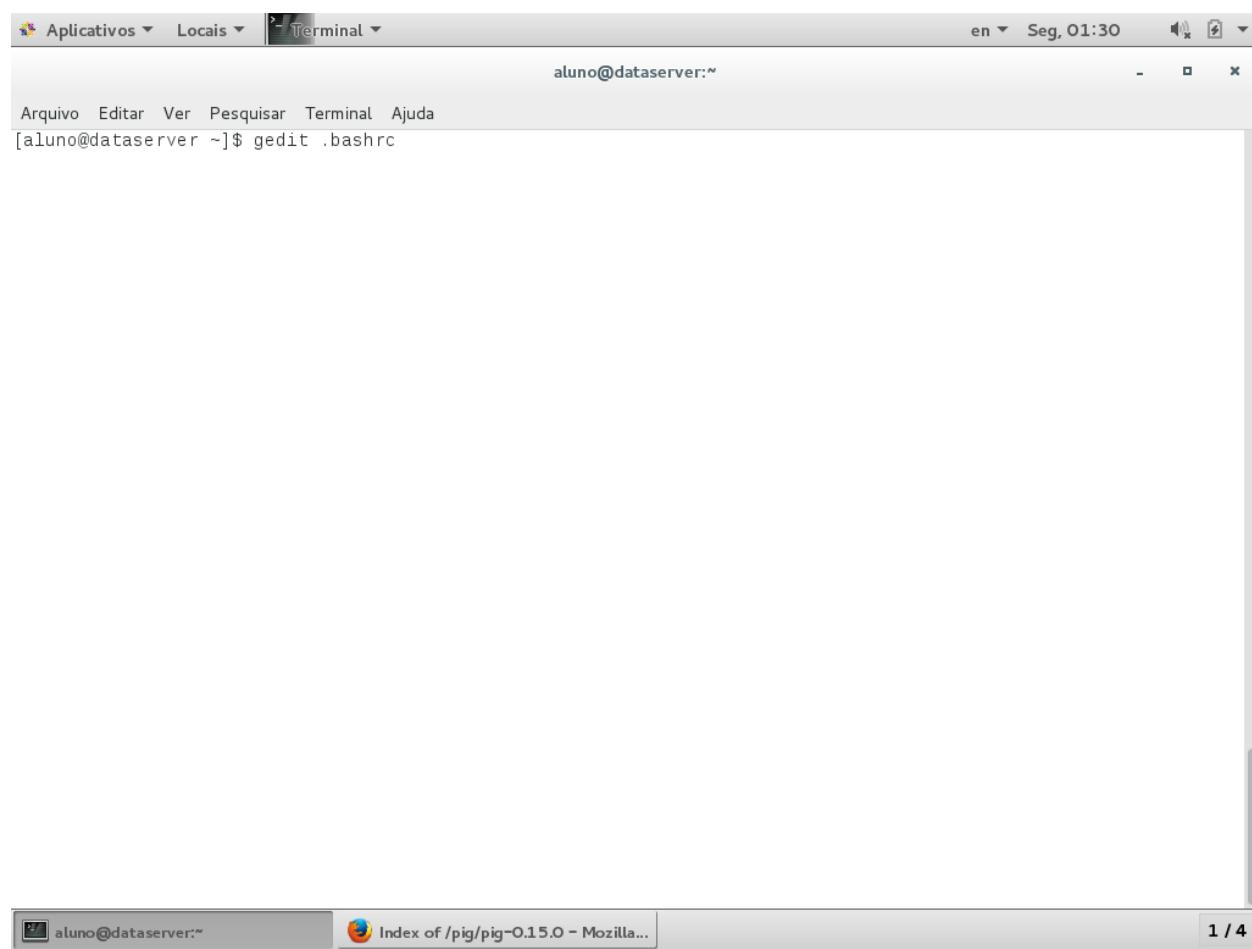
Criar link simbólico para o diretório de instalação do Pig



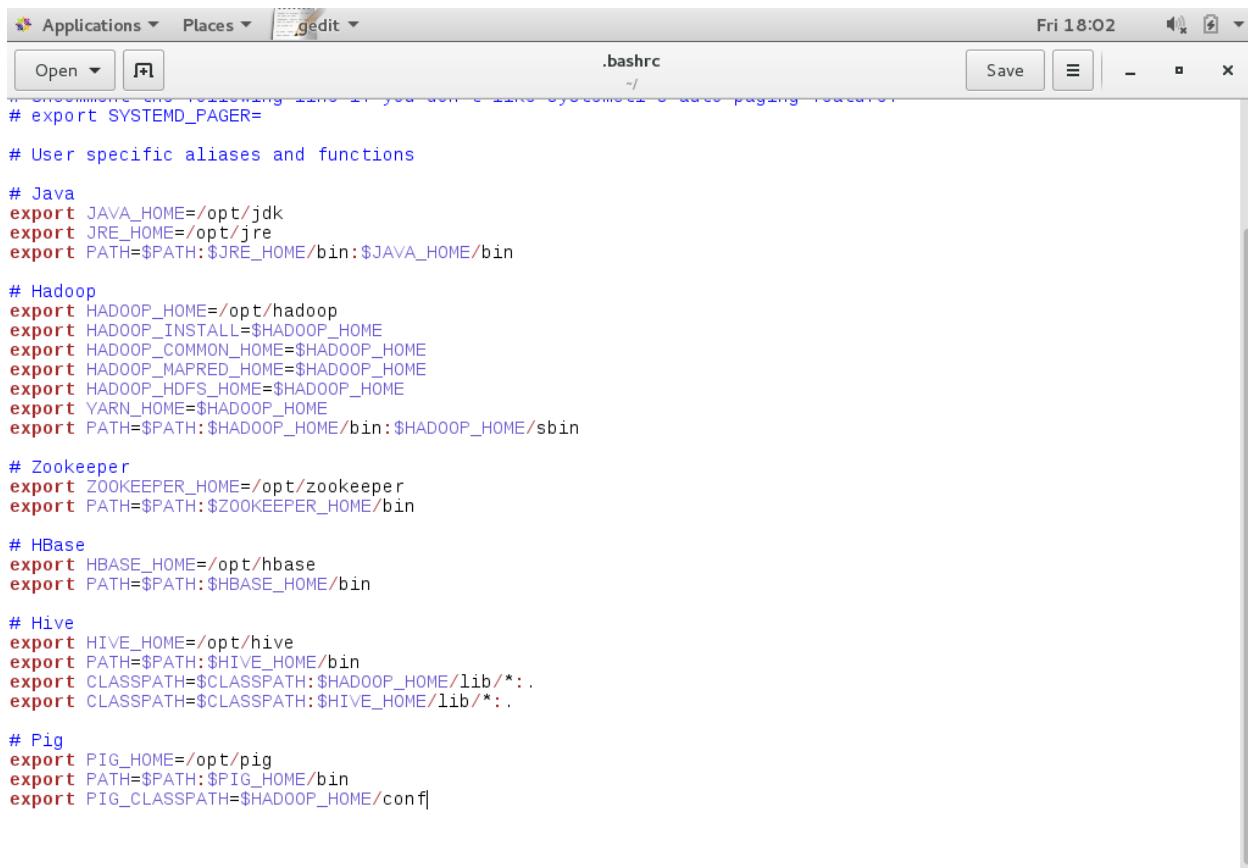
```
[aluno@dataserver opt]$ sudo ln -s /usr/local/pig-0.16.0/ /opt/pig
[aluno@dataserver opt]$ ls -la
total 4
drwxr-xr-x. 2 root root 88 Sep 30 18:00 .
dr-xr-xr-x. 17 root root 4096 Sep 29 18:37 ..
lrwxrwxrwx. 1 root root 24 Sep 29 20:31 hadoop -> /usr/local/hadoop-2.7.3/
lrwxrwxrwx. 1 root root 23 Sep 30 17:04 hbase -> /usr/local/hbase-1.2.3/
lrwxrwxrwx. 1 root root 33 Sep 30 17:23 hive -> /usr/local/apache-hive-2.1.0-bin/
lrwxrwxrwx. 1 root root 24 Sep 29 18:56 jdk -> /usr/local/jdk1.8.0_102/
lrwxrwxrwx. 1 root root 65 Sep 29 18:59 jre -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/
lrwxrwxrwx. 1 root root 22 Sep 30 18:00 pig -> /usr/local/pig-0.16.0/
lrwxrwxrwx. 1 root root 27 Sep 30 16:49 zookeeper -> /usr/local/zookeeper-3.4.9/
[aluno@dataserver opt]$
```

Link criado

9.2. Configurando do Pig



Editando o arquivo .bashrc



```

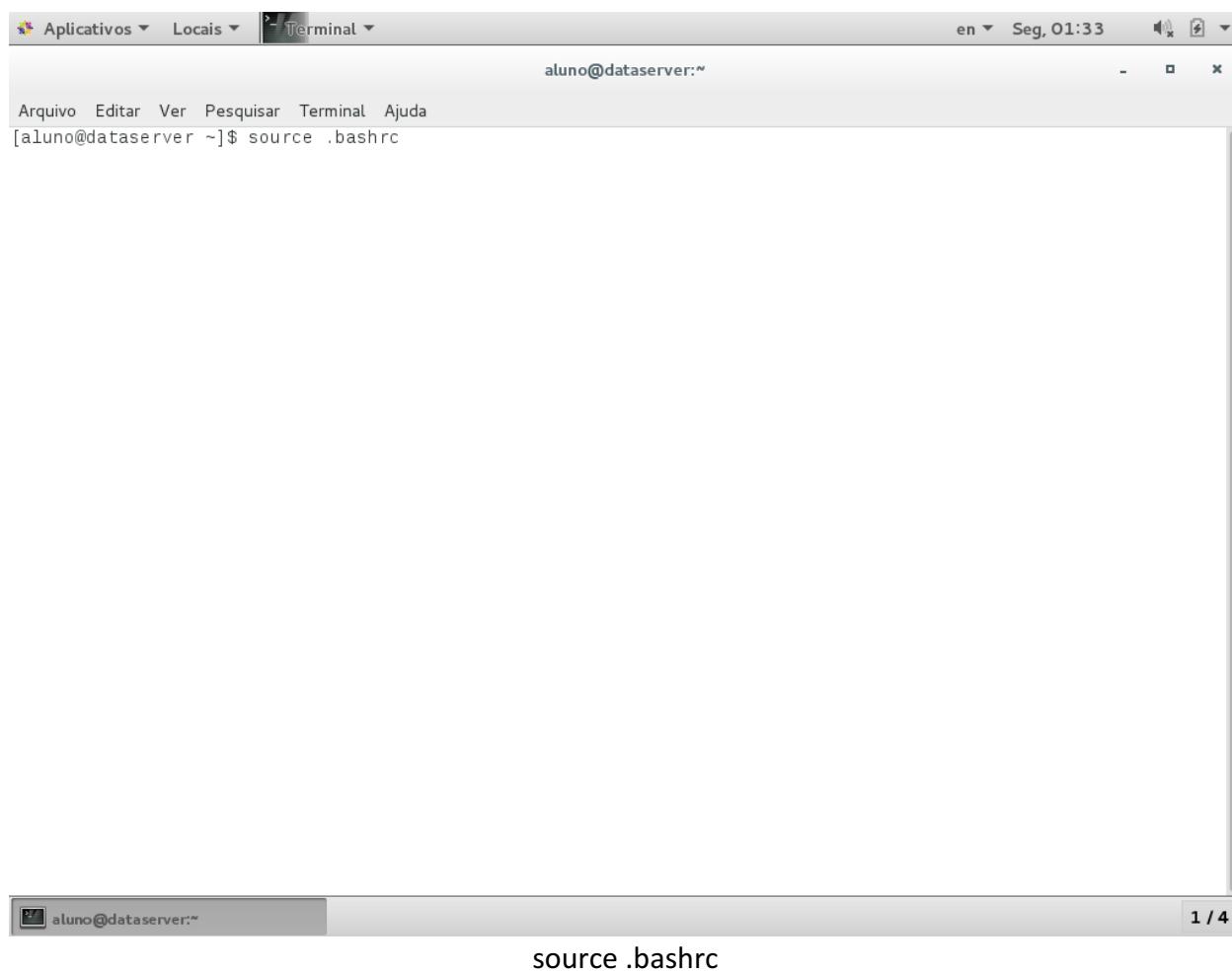
# export SYSTEMD_PAGER=
# User specific aliases and functions
# Java
export JAVA_HOME=/opt/jdk
export JRE_HOME=/opt/jre
export PATH=$PATH:$JRE_HOME/bin:$JAVA_HOME/bin
# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
# Zookeeper
export ZOOKEEPER_HOME=/opt/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin
# HBase
export HBASE_HOME=/opt/hbase
export PATH=$PATH:$HBASE_HOME/bin
# Hive
export HIVE_HOME=/opt/hive
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:.
export CLASSPATH=$CLASSPATH:$HIVE_HOME/lib/*:.
# Pig
export PIG_HOME=/opt/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$HADOOP_HOME/conf

```

The screenshot shows a terminal window with the title "gedit" and the file ".bashrc" open. The code in the file sets environment variables for Java, Hadoop, Zookeeper, HBase, Hive, and Pig. It includes paths for bin and sbin directories, and classpath entries for Hadoop and Hive.

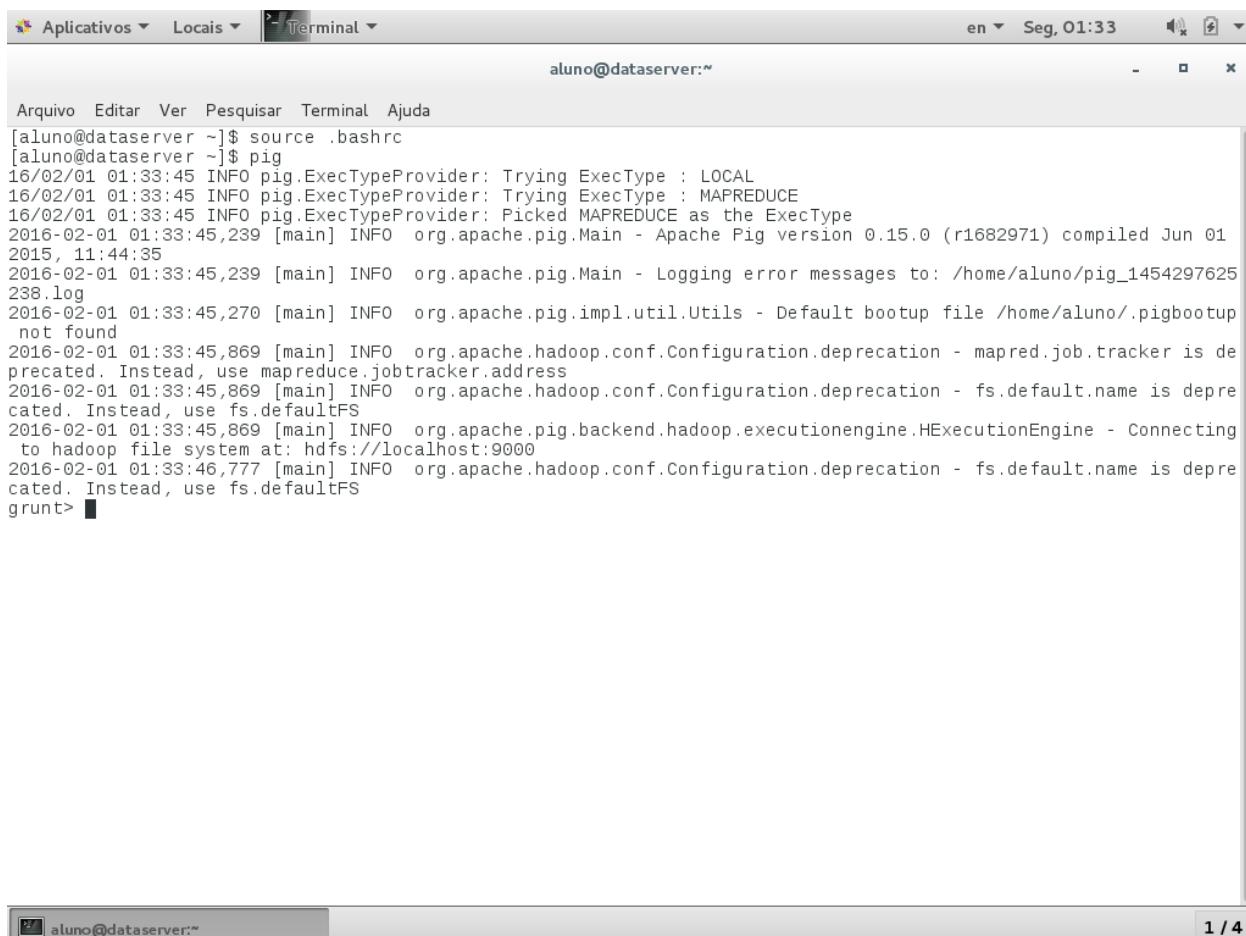
Inserir variáveis de ambiente do Pig

Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The menu bar includes "Aplicativos", "Locais", and "Terminal". The status bar shows "en Seg, 01:33". The terminal prompt is "aluno@dataserver:~". The user has typed the command "[aluno@dataserver ~]\$ source .bashrc". The bottom status bar also shows "aluno@dataserver:~" and "1 / 4".

```
aluno@dataserver:~$ source .bashrc
```

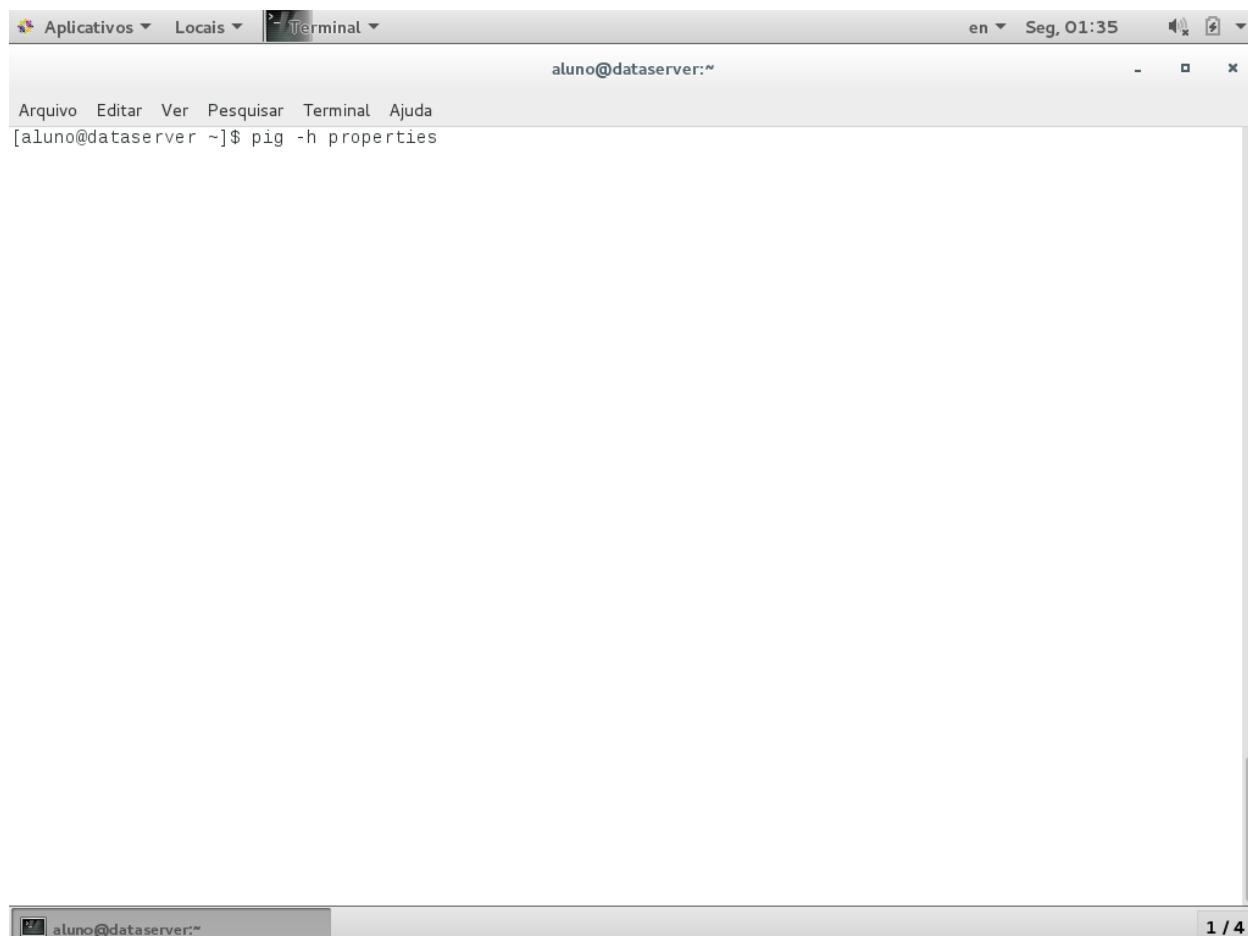


A screenshot of a terminal window titled "Terminal". The window shows a command-line interface with the user "aluno@dataserver:~". The terminal displays the output of running the "pig" command after sourcing ".bashrc". The log output includes several INFO messages from the Apache Pig and Hadoop frameworks, indicating the configuration of the execution type (LOCAL/MAPREDUCE), the version of Pig (0.15.0), and the connection to the local host's HDFS at port 9000. The session ends with the prompt "grunt>".

```
[aluno@dataserver ~]$ source .bashrc
[aluno@dataserver ~]$ pig
16/02/01 01:33:45 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
16/02/01 01:33:45 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
16/02/01 01:33:45 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2016-02-01 01:33:45,239 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01
2015, 11:44:35
2016-02-01 01:33:45,239 [main] INFO org.apache.pig.Main - Logging error messages to: /home/aluno/pig_1454297625
238.log
2016-02-01 01:33:45,270 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/aluno/.pigbootup
not found
2016-02-01 01:33:45,869 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is de
precated. Instead, use mapreduce.jobtracker.address
2016-02-01 01:33:45,869 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is depre
cated. Instead, use fs.defaultFS
2016-02-01 01:33:45,869 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting
to hadoop file system at: hdfs://localhost:9000
2016-02-01 01:33:46,777 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is depre
cated. Instead, use fs.defaultFS
grunt>
```

Pig instalado com sucesso

1 / 4



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The status bar at the top right shows 'en Seg, 01:35'. The terminal menu bar includes 'Aplicativos', 'Locais', and 'Terminal'. The user prompt is 'aluno@dataserver:~'. Below the prompt, the command '[aluno@dataserver ~]\$ pig -h properties' is entered. The terminal window has scroll bars on the right side.

O comando **pig -h properties** lista as variáveis configuradas

```

aluno@dataserver:~$ pig -list
Logging:
verbose=true|false; default is false. This property is the same as -v switch
brief=true|false; default is false. This property is the same as -b switch
debug=OFF|ERROR|WARN|INFO|DEBUG; default is INFO. This property is the same as -d switch
aggregate.warning=true|false; default is true. If true, prints count of warnings
of each type rather than logging each warning.

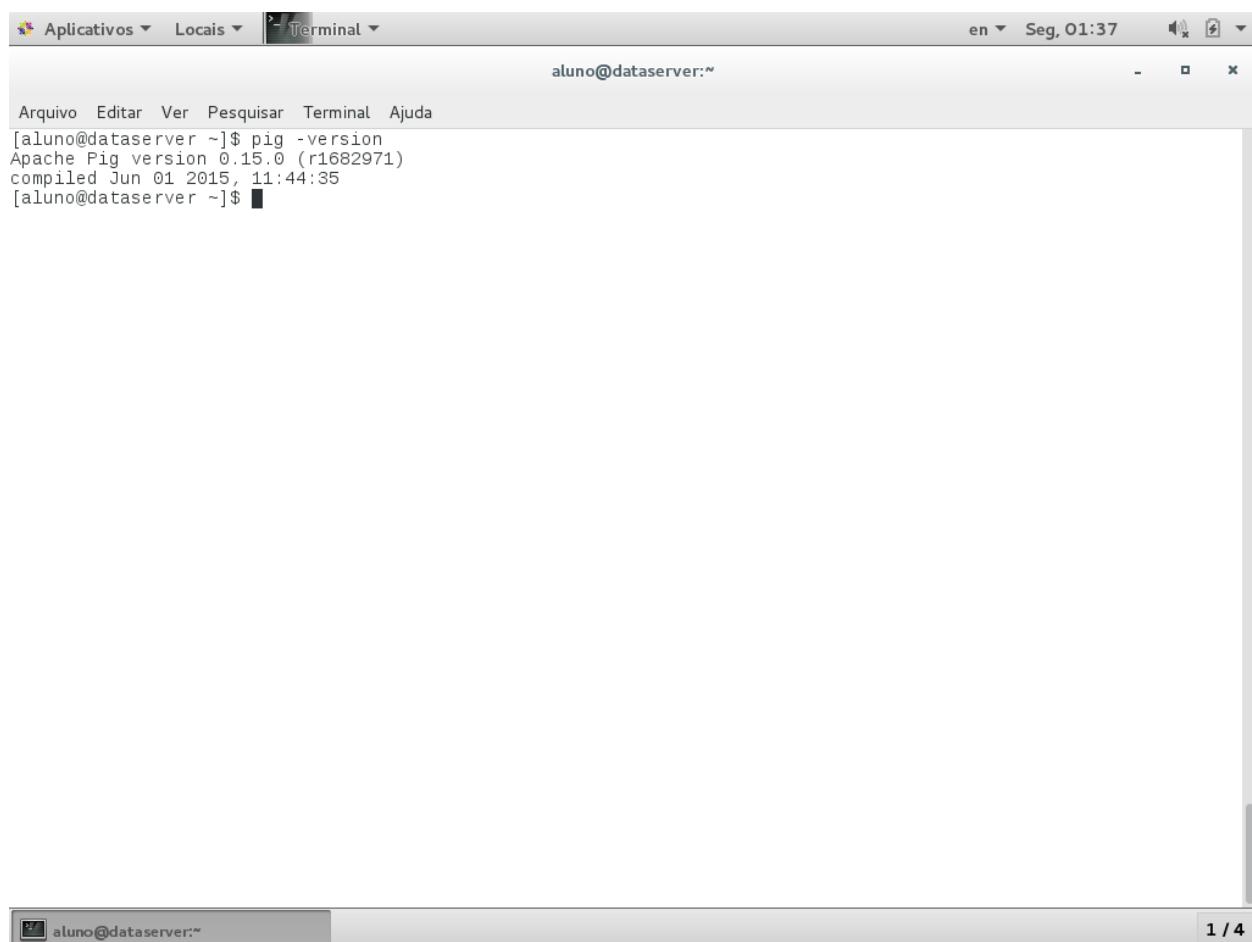
Performance tuning:
pig.cachedbag.memusage=<mem fraction>; default is 0.2 (20% of all memory).
Note that this memory is shared across all large bags used by the application.
pig.skewedjoin.reduce.memusagea=<mem fraction>; default is 0.3 (30% of all memory).
Specifies the fraction of heap available for the reducer to perform the join.
pig.exec.nocombiner=true|false; default is false.
Only disable combiner as a temporary workaround for problems.
opt.multiquery=true|false; multiquery is on by default.
Only disable multiquery as a temporary workaround for problems.
opt.fetch=true|false; fetch is on by default.
Scripts containing Filter, Foreach, Limit, Stream, and Union can be dumped without MR jobs.
pig.tmpfilecompression=true|false; compression is off by default.
Determines whether output of intermediate jobs is compressed.
pig.tmpfilecompression.codec=lzo|gzip; default is gzip.
Used in conjunction with pig.tmpfilecompression. Defines compression type.
pig.noSplitCombination=true|false. Split combination is on by default.
Determines if multiple small files are combined into a single map.
pig.exec.mapPartAgg=true|false. Default is false.
Determines if partial aggregation is done within map phase,
before records are sent to combiner.
pig.exec.mapPartAgg.minReduction=<min aggregation factor>. Default is 10.
If the in-map partial aggregation does not reduce the output num records
by this factor, it gets disabled.

Miscellaneous:
execType=mapreduce|tez|local; default is mapreduce. This property is the same as -x switch
pig.additional.jars.uris=<comma separated list of jars>. Used in place of register command.
udf.import.list=<comma seperated list of imports>. Used to avoid package names in UDF.
stop.on.failure=true|false; default is false. Set to true to terminate on the first error.
pig.datetime.default.tz=<UTC time offset>. e.g. +08:00. Default is the default timezone of the host.
Determines the timezone used to handle datetime datatype and UDFs.

Additionally, any Hadoop property can be specified.
16/02/01 01:36:00 INFO pig.Main: Pig script completed in 224 milliseconds (224 ms)
[aluno@dataserver ~]$ 
```

Variáveis Pig

1 / 4



A screenshot of a terminal window titled "Terminal". The window shows the command "pig -version" being run by the user "aluno" on a server named "dataserver". The output of the command is displayed, indicating the Apache Pig version is 0.15.0 (r1682971), compiled on June 01, 2015, at 11:44:35.

```
aluno@dataserver:~$ pig -version
Apache Pig version 0.15.0 (r1682971)
compiled Jun 01 2015, 11:44:35
[aluno@dataserver ~]$
```

Verificar a versão do Pig

10. Instalação e Configuração do Spark

10.1. Download e Instalação do Spark

Downloads | Apache Spark – Mozilla Firefox

Downloads | Apache ...

spark.apache.org/downloads.html

APACHE *Lightning-fast cluster computing*

Download Libraries Documentation Examples Community FAQ Apache Software Foundation

Download Apache Spark™

Our latest stable version is Apache Spark 2.0.0, released on July 26, 2016 ([release notes](#)) ([git tag](#))

1. Choose a Spark release: [2.0.0 \(Jul 26 2016\)](#)
2. Choose a package type: [Pre-built for Hadoop 2.7 and later](#)
3. Choose a download type: [Direct Download](#)
4. Download Spark: [spark-2.0.0-bin-hadoop2.7.tgz](#)
5. Verify this release using [2.0.0 signatures and checksums](#) and [project release KEYS](#).

Note: Starting version 2.0, Spark is built with Scala 2.11 by default. Scala 2.10 users should download the Spark source package and build [with Scala 2.10 support](#).

Latest News

- Spark 2.0.0 released (Jul 26, 2016)
- Spark 1.6.2 released (Jun 25, 2016)
- Call for Presentations for Spark Summit EU is Open (Jun 16, 2016)
- Preview release of Spark 2.0 (May 26, 2016)

Archive

Download Spark

Link with Spark

Spark artifacts are [hosted in Maven Central](#). You can add a Maven dependency with the following coordinates:

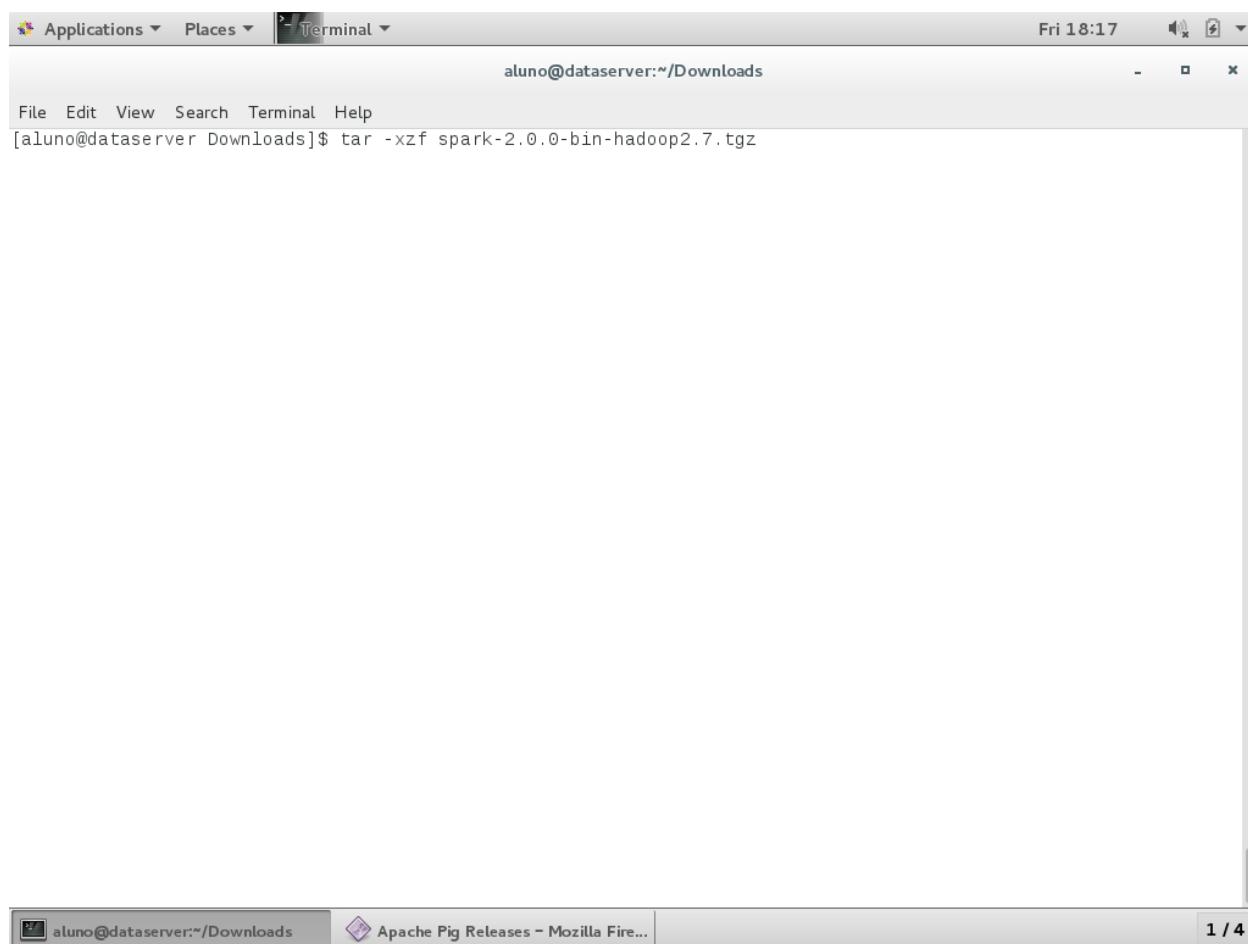
Downloads | Apache Spark – Mozilla Firefox

Download do Spark – Versão 2.0.0

Built-in Libraries

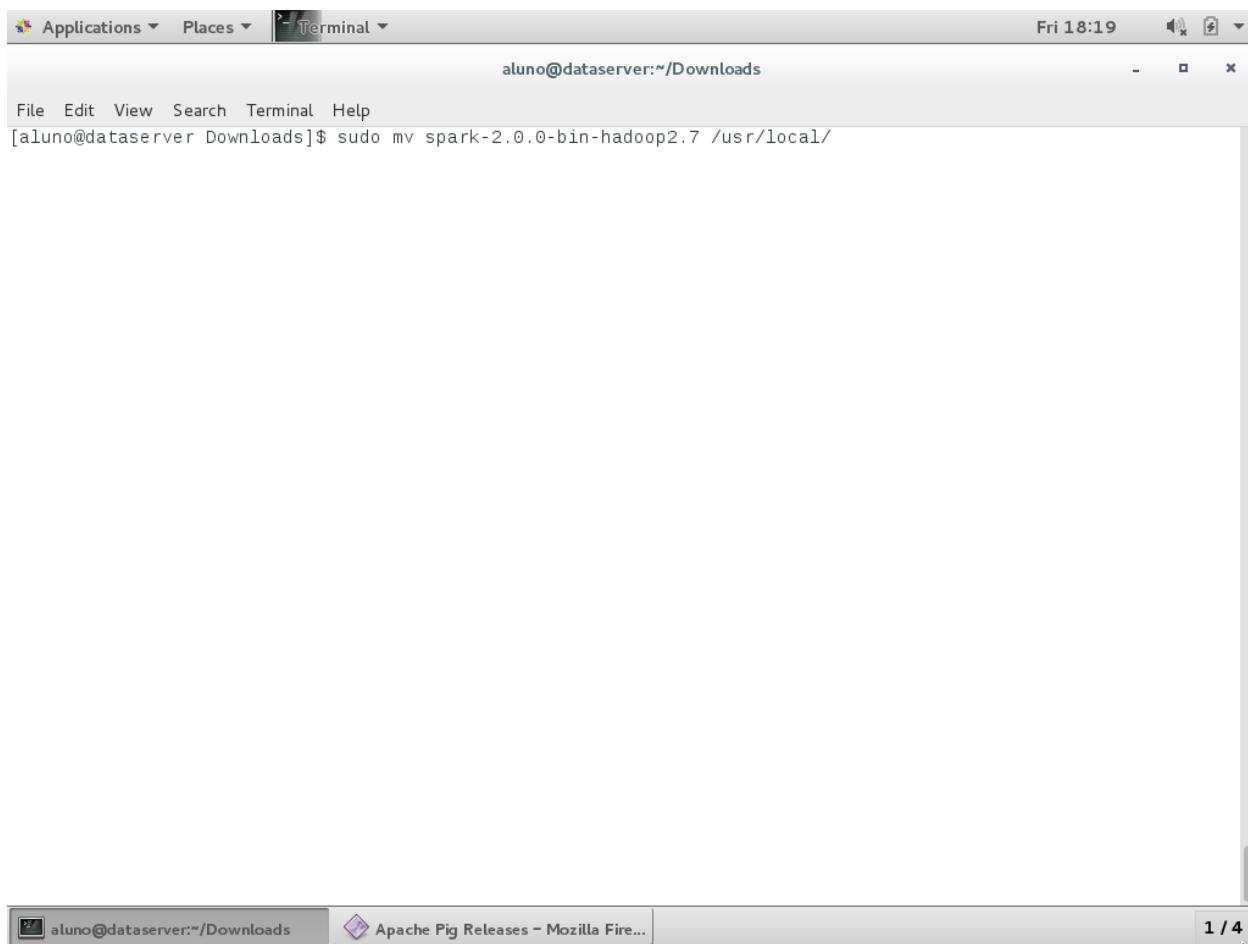
1 / 4

Instalação e Configuração do Ecosistema Hadoop



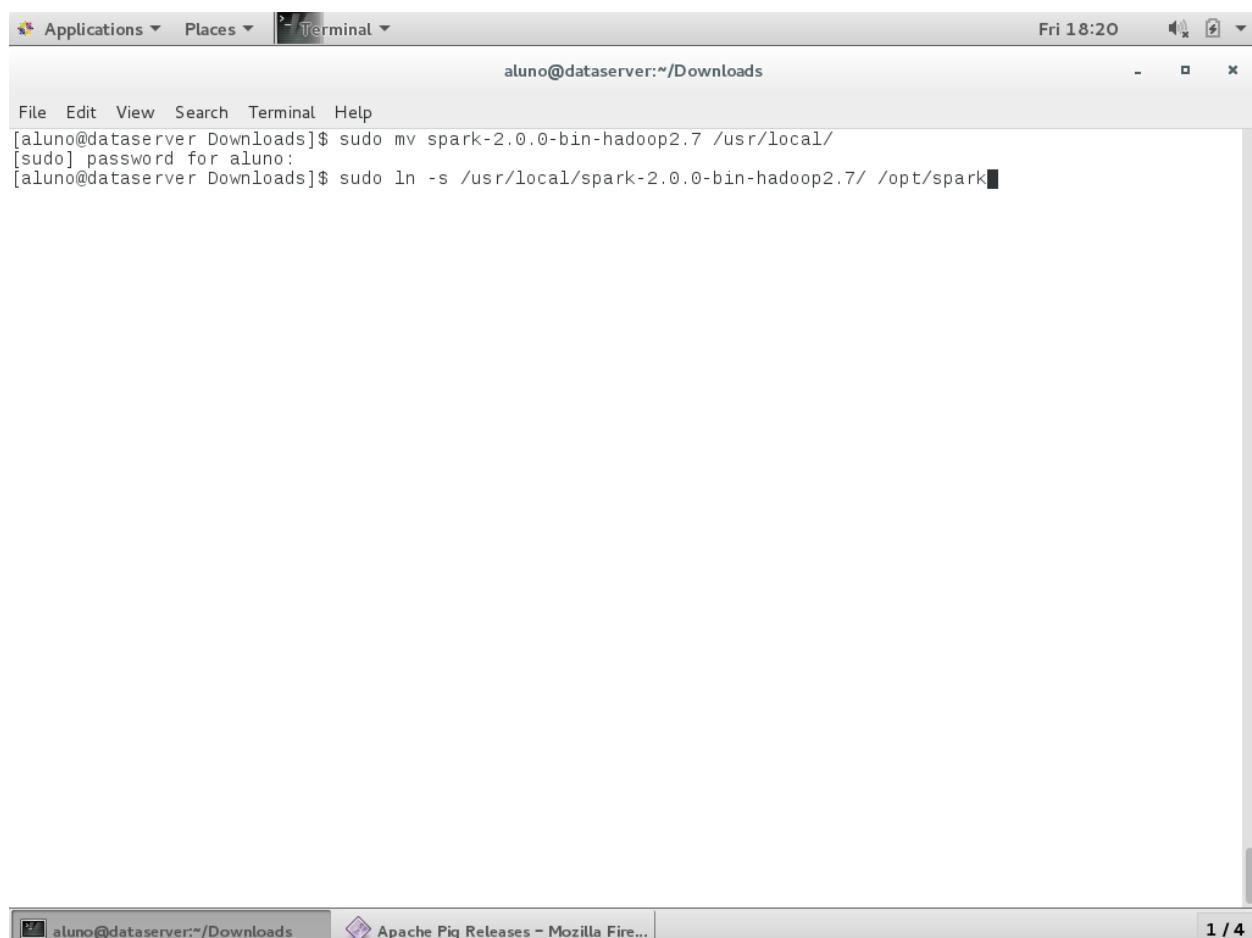
Descompactando o arquivo

Instalação e Configuração do Ecosistema Hadoop



Copiando o diretório do Spark para /usr/local

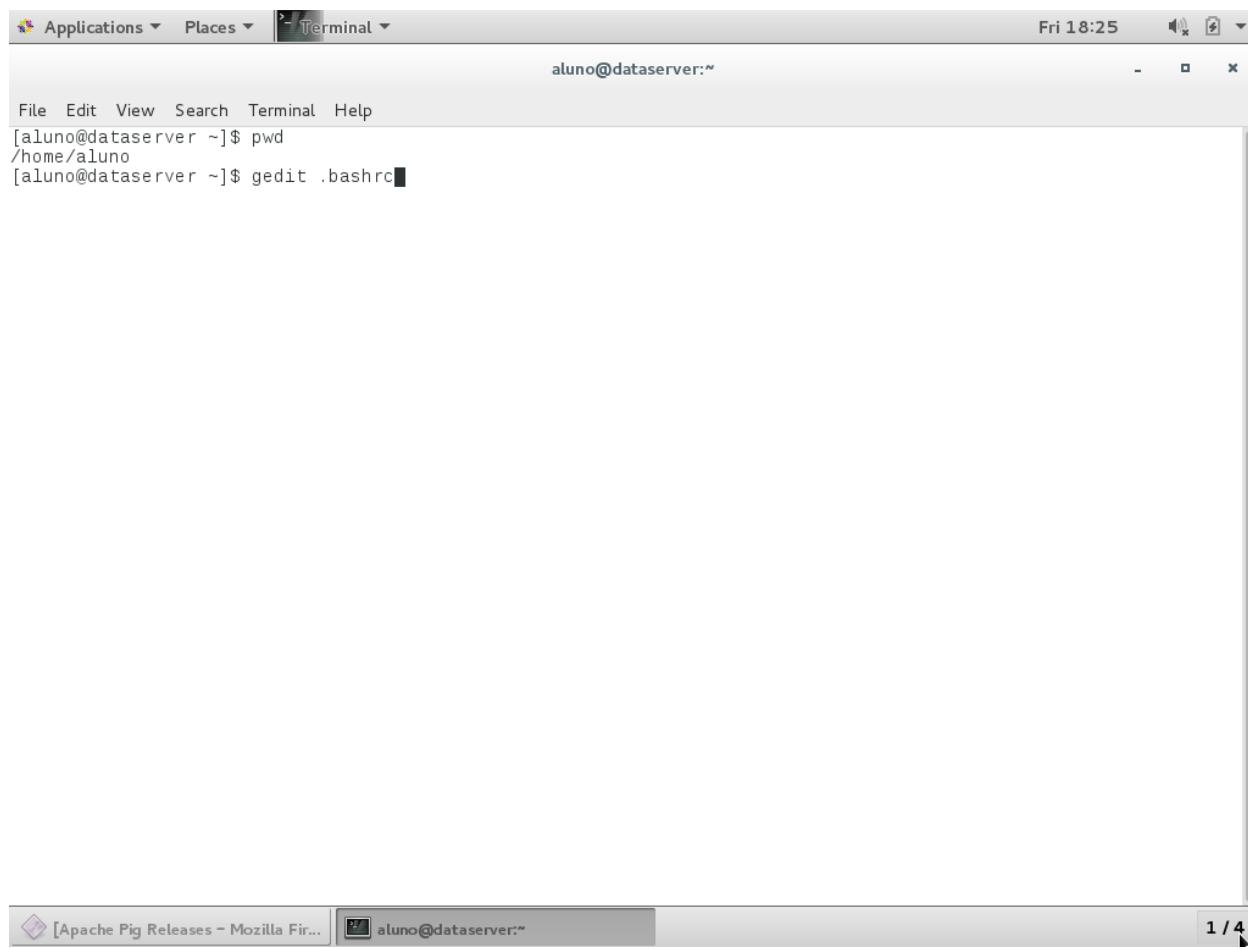
1 / 4



The screenshot shows a terminal window titled "Terminal" with the command line interface. The terminal window has a title bar with "Applications", "Places", and "Terminal". The status bar at the top right shows "Fri 18:20". The terminal window itself has a header bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The main area of the terminal shows the following command sequence:

```
[aluno@dataserver Downloads]$ sudo mv spark-2.0.0-bin-hadoop2.7 /usr/local/  
[sudo] password for aluno:  
[aluno@dataserver Downloads]$ sudo ln -s /usr/local/spark-2.0.0-bin-hadoop2.7/ /opt/spark
```

Criando o link simbólico



Editando o arquivo .bashrc

The screenshot shows a Linux desktop interface. At the top, there's a panel with 'Applications', 'Places', and a search bar. The date and time 'Fri 18:24' are also visible. Below the panel, a window titled '.bashrc' is open in the 'gedit' text editor. The file contains configuration code for various Hadoop components like Java, Hadoop, Zookeeper, HBase, Hive, Pig, and Spark. A terminal window is also visible at the bottom, showing the command 'Incluir variáveis Spark'.

```
# User specific aliases and functions

# Java
export JAVA_HOME=/opt/jdk
export JRE_HOME=/opt/jre
export PATH=$PATH:$JRE_HOME/bin:$JAVA_HOME/bin

# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Zookeeper
export ZOOKEEPER_HOME=/opt/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin

# HBase
export HBASE_HOME=/opt/hbase
export PATH=$PATH:$HBASE_HOME/bin

# Hive
export HIVE_HOME=/opt/hive
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:.
export CLASSPATH=$CLASSPATH:$HIVE_HOME/lib/*:.

# Pig
export PIG_HOME=/opt/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$HADOOP_HOME/conf

# Spark
export SPARK_HOME=/opt/spark
export PATH=$PATH:$SPARK_HOME/bin
```

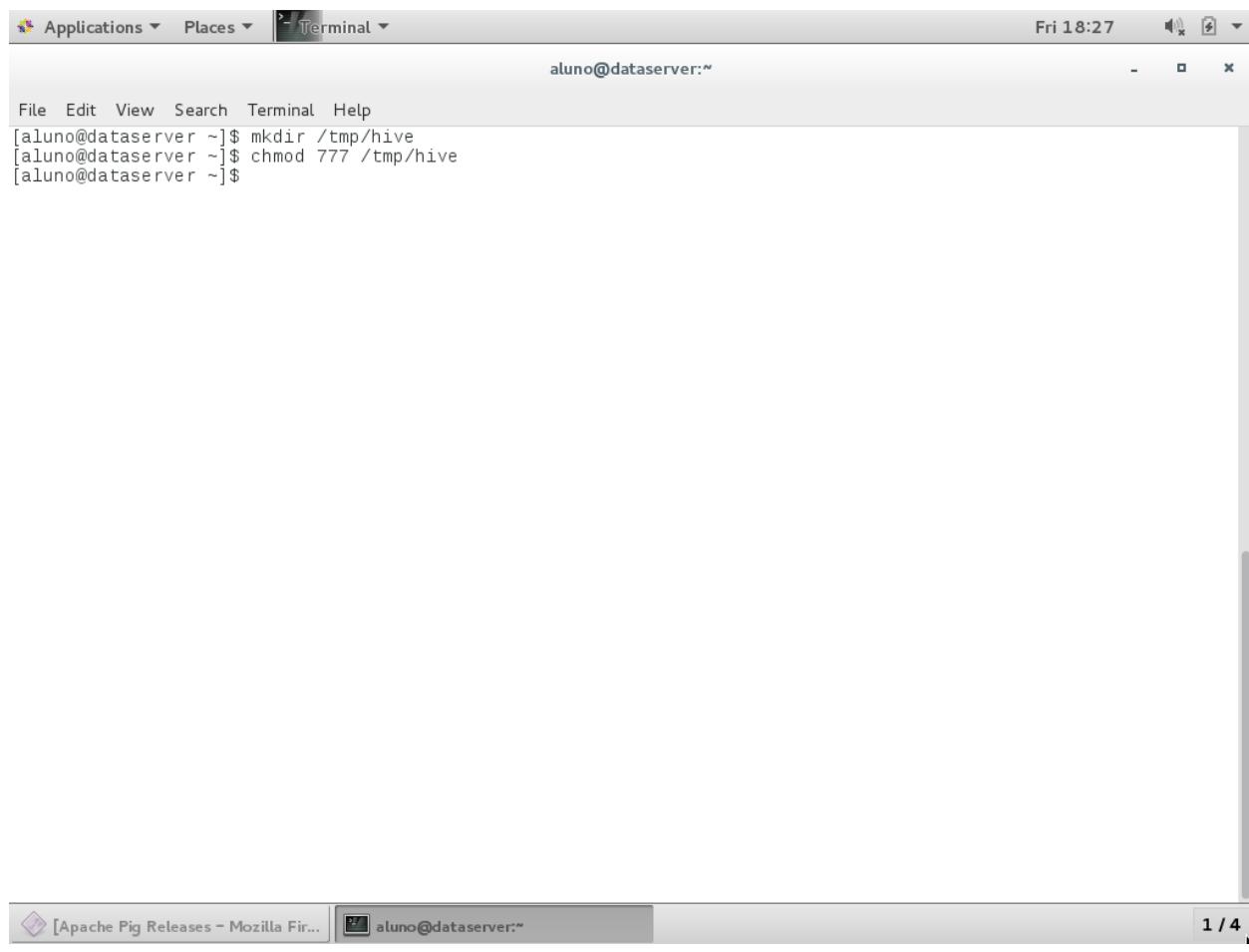
Instalação e Configuração do Ecosistema Hadoop



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top right shows "Fri 18:26". The terminal prompt is "aluno@dataserver:~". The menu bar includes "File Edit View Search Terminal Help". The command history shows:

```
[aluno@dataserver ~]$ pwd  
/home/aluno  
[aluno@dataserver ~]$ source .bashrc
```

The terminal window is part of a larger desktop interface. Below it, another window titled "[Apache Pig Releases - Mozilla Fir..." is visible. The bottom right corner of the screen shows "1 / 4".



A screenshot of a Linux desktop environment showing a terminal window. The terminal title is "Terminal" and the user is "aluno@dataserver:~". The terminal shows the following commands being run:

```
[aluno@dataserver ~]$ mkdir /tmp/hive  
[aluno@dataserver ~]$ chmod 777 /tmp/hive  
[aluno@dataserver ~]$
```

The terminal window has a standard top bar with "Applications", "Places", and "Terminal" buttons. The date and time "Fri 18:27" are shown in the top right corner. Below the terminal window, there is a browser tab titled "[Apache Pig Releases - Mozilla Fir...]" and a status bar indicating "1 / 4".

Se necessário chmod 777 /tmp/hive



The screenshot shows a terminal window titled "Terminal" with the user "aluno@dataserver:~". The terminal displays the output of the "spark-shell" command. It includes log messages from Spark's default log4j profile, Java native library loading, and host resolution. It also shows the Scala version 2.11.8 startup message and the Spark shell prompt "scala>".

```
[aluno@dataserver ~]$ spark-shell
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
16/09/30 18:28:04 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
-java classes where applicable
16/09/30 18:28:05 WARN Utils: Your hostname, localhost resolves to a loopback address: 127.0.0.1; using 10.0.2.1
5 instead (on interface enp0s3)
16/09/30 18:28:05 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
16/09/30 18:28:07 WARN SparkContext: Use an existing SparkContext, some configuration may not take effect.
Spark context Web UI available at http://10.0.2.15:4040
Spark context available as 'sc' (master = local[*], app id = local-1475285286549).
Spark session available as 'spark'.
Welcome to
    / \ \
   /   \ \
  /     \ \
 /       \ \
/         \ \
 \       / \
  \     / \
   \   / \
    \ / \
     \ /
      \
version 2.0.0

Using Scala version 2.11.8 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_102)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

Spark shell

1 / 4

The screenshot shows the Apache Spark 2.0.0 web UI interface. At the top, there's a header bar with the Apache logo, the version '2.0.0', and tabs for 'Jobs', 'Stages', 'Storage', 'Environment', 'Executors', and 'SQL'. To the right, it says 'Spark shell application UI'. Below the header, the title 'Spark Jobs (?)' is displayed. Underneath, user information is shown: 'User: aluno', 'Total Uptime: 38 s', and 'Scheduling Mode: FIFO'. A section titled 'Event Timeline' is expanded, showing a grid where the first column represents time from Saturday, October 1st to Friday, October 7th. The first row, 'Executors', has two columns: 'Added' (blue square) and 'Removed' (red square). The second row, 'Jobs', has three columns: 'Succeeded' (blue square), 'Failed' (red square), and 'Running' (green square). A tooltip 'river added' is visible over the 'Added' executor cell for Saturday. The bottom of the timeline grid shows the date 'October 2016'.



Acessando o Apache Spark pelo browser em <http://localhost:4040>

11. Instalação e Configuração do Sqoop

11.1. Download do Sqoop

The screenshot shows a Firefox browser window with the title "Sqoop - Mozilla Firefox". The address bar contains "sqoop.apache.org". The main content area displays the Apache Sqoop project page. At the top left is the "sqoop" logo. To its right is the "The Apache Software Foundation" logo with the URL "http://www.apache.org/". Below the logos, the page header reads "Apache Sqoop". Underneath, there is a brief description: "Apache Sqoop(TM) is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases." It mentions that Sqoop successfully graduated from the Incubator in March of 2012 and is now a Top-Level Apache project. It also notes the latest stable release is 1.4.6 and the latest cut of Sqoop2 is 1.99.7. A "Download" section provides links to download the software and browse the repository online via Git.

Apache Sqoop

Apache Sqoop(TM) is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases.

Sqoop successfully graduated from the Incubator in March of 2012 and is now a Top-Level Apache project: [More information](#)

Latest stable release is 1.4.6 ([download](#), [documentation](#)). Latest cut of Sqoop2 is 1.99.7 ([download](#), [documentation](#)). Note that 1.99.7 is not compatible with 1.4.6 and not feature complete, it is not intended for production deployment.

Download

Download a release of Sqoop from a [nearby mirror](#).

Sqoop source code is held in the Apache GIT repository.

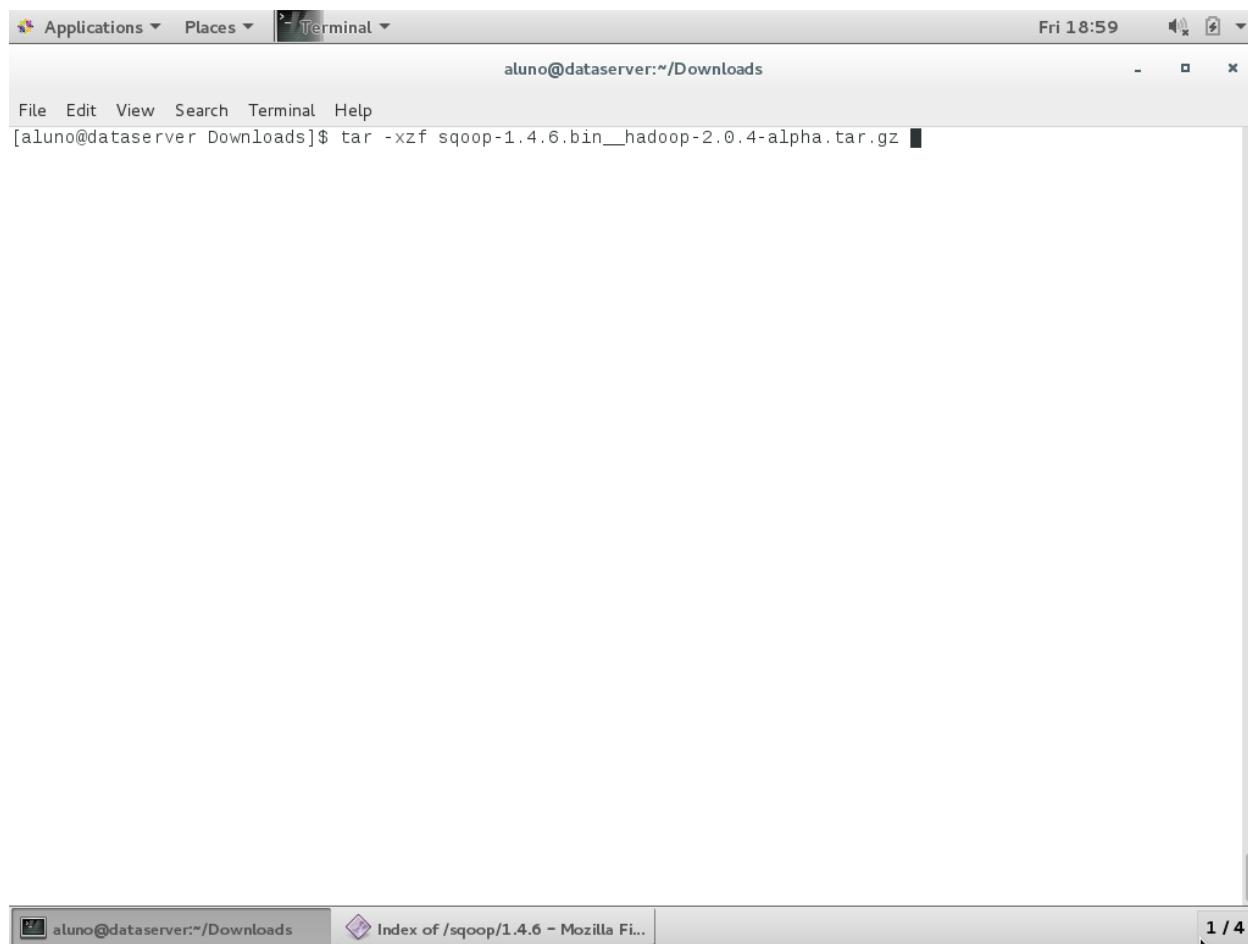
You might clone the repository using following command:

```
git clone https://git-wip-us.apache.org/repos/asf/sqoop.git
```

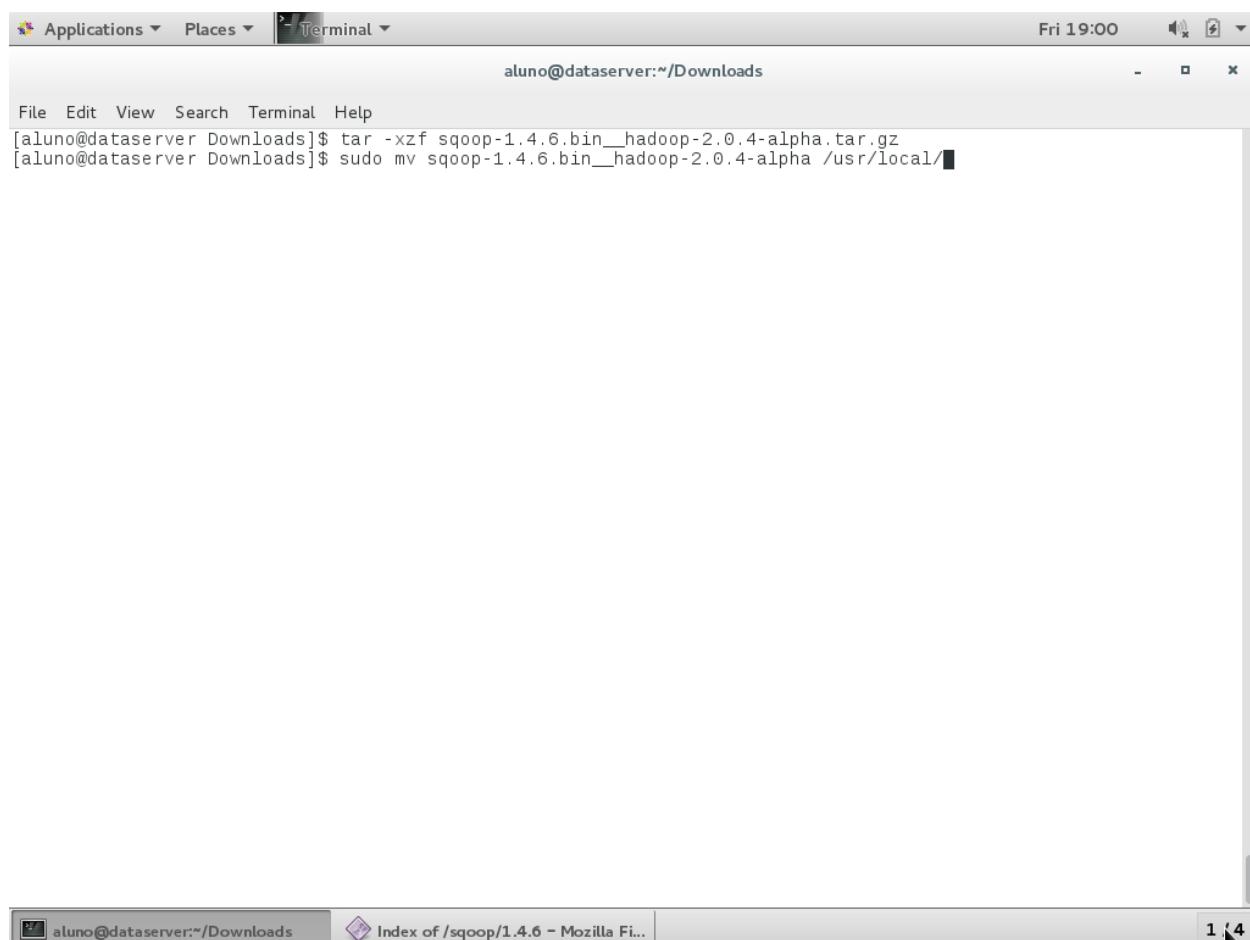
Use following link to browse the repository online:

Download do Sqoop – Versão 1.4.6-hadoop-2.0.4-alpha

Instalação e Configuração do Ecosistema Hadoop



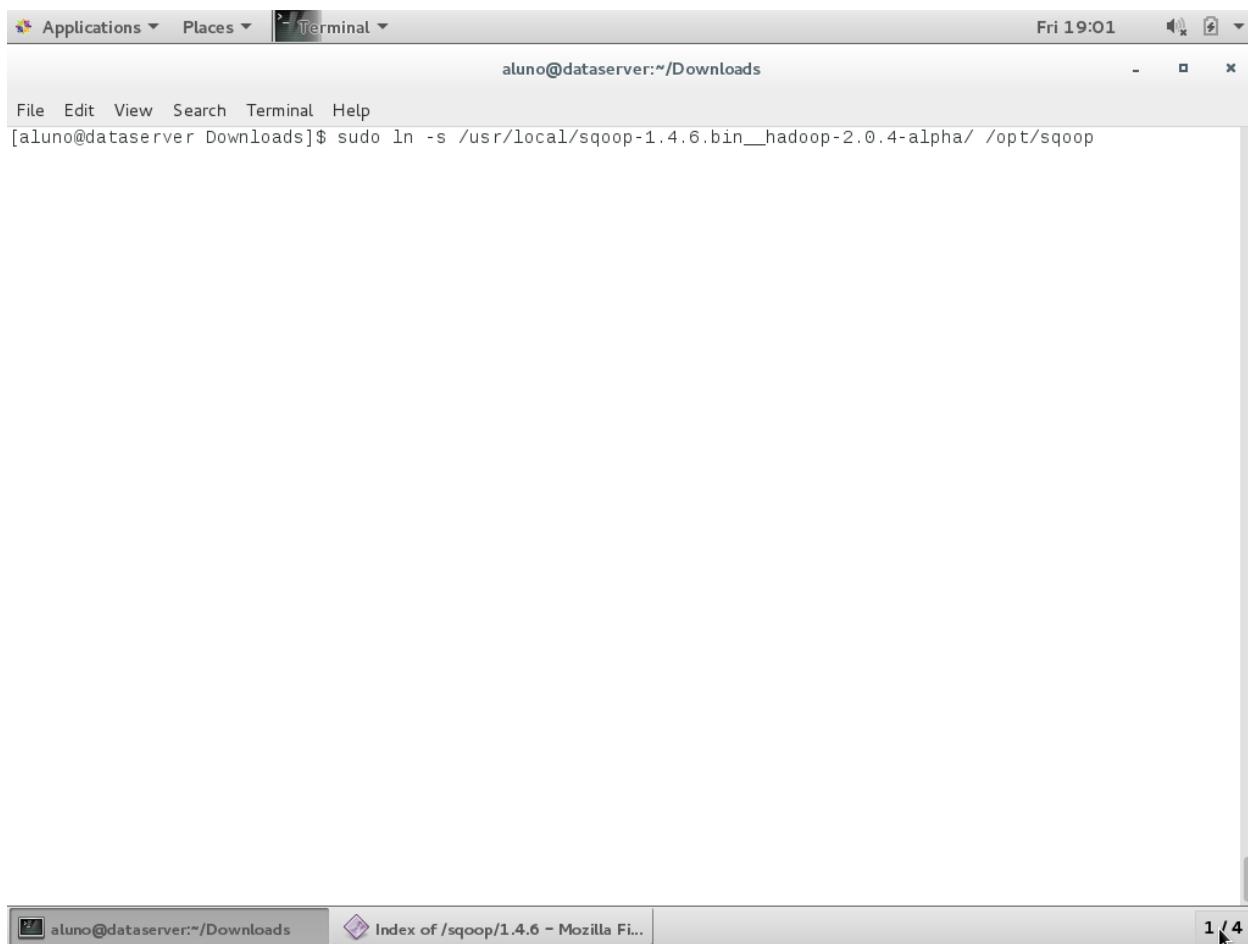
Download concluído



The screenshot shows a terminal window titled 'Terminal' with the command line interface. The terminal window is part of a desktop environment with a menu bar at the top. The terminal content shows two commands being run:

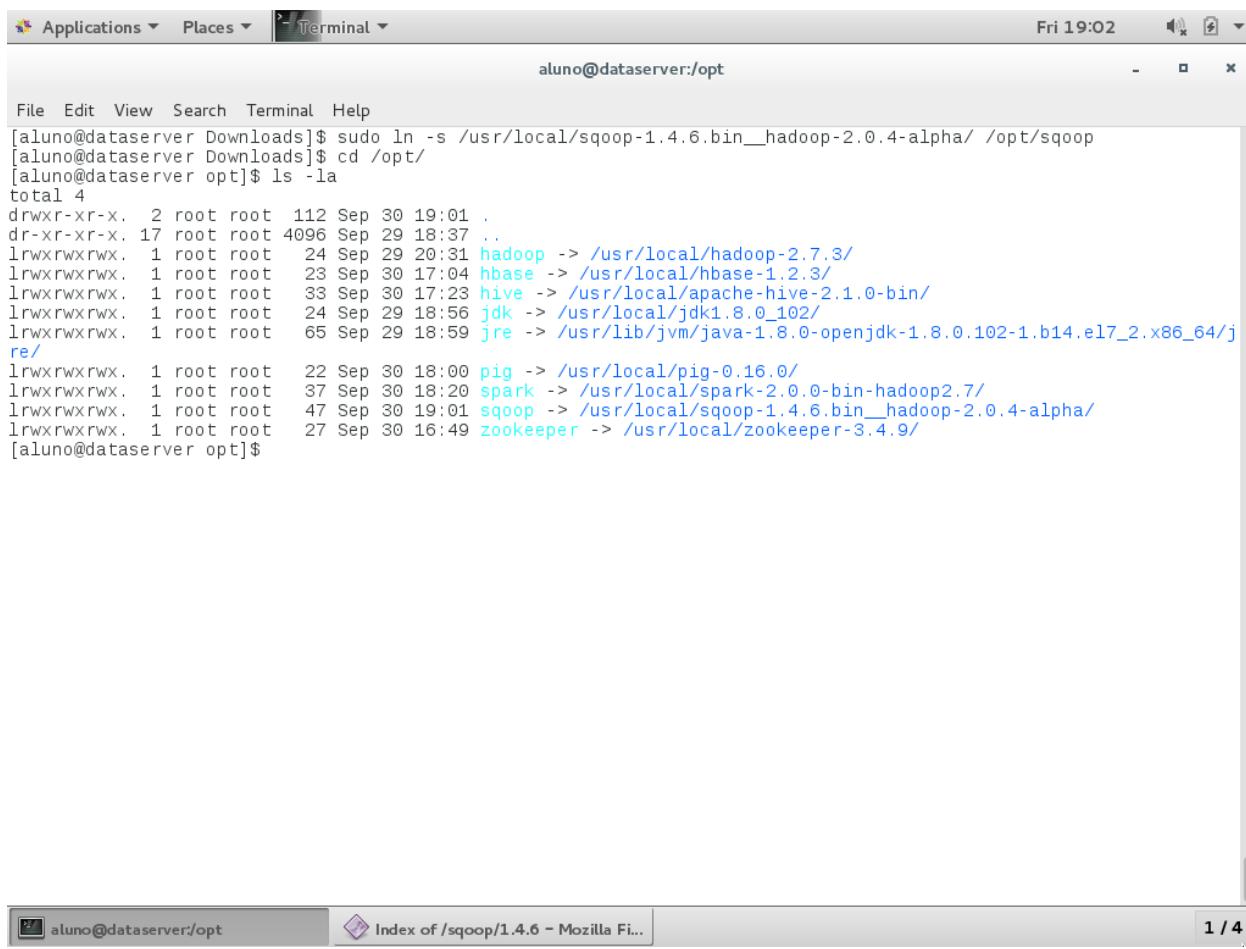
```
[aluno@dataserver Downloads]$ tar -xzf sqoop-1.4.6.bin__hadoop-2.0.4-alpha.tar.gz  
[aluno@dataserver Downloads]$ sudo mv sqoop-1.4.6.bin__hadoop-2.0.4-alpha /usr/local/
```

Descompactar o arquivo e mover a pasta do Sqoop para /usr/local



A screenshot of a Linux desktop environment showing a terminal window. The terminal title bar says "Terminal". The window title is "aluno@dataserver:~/Downloads". The terminal shows the command: [aluno@dataserver Downloads]\$ sudo ln -s /usr/local/sqoop-1.4.6-bin__hadoop-2.0.4-alpha/ /opt/sqoop. The status bar at the bottom of the terminal window shows "aluno@dataserver:~/Downloads" and "Index of /sqoop/1.4.6 - Mozilla Fi...". The status bar also indicates "1 / 4".

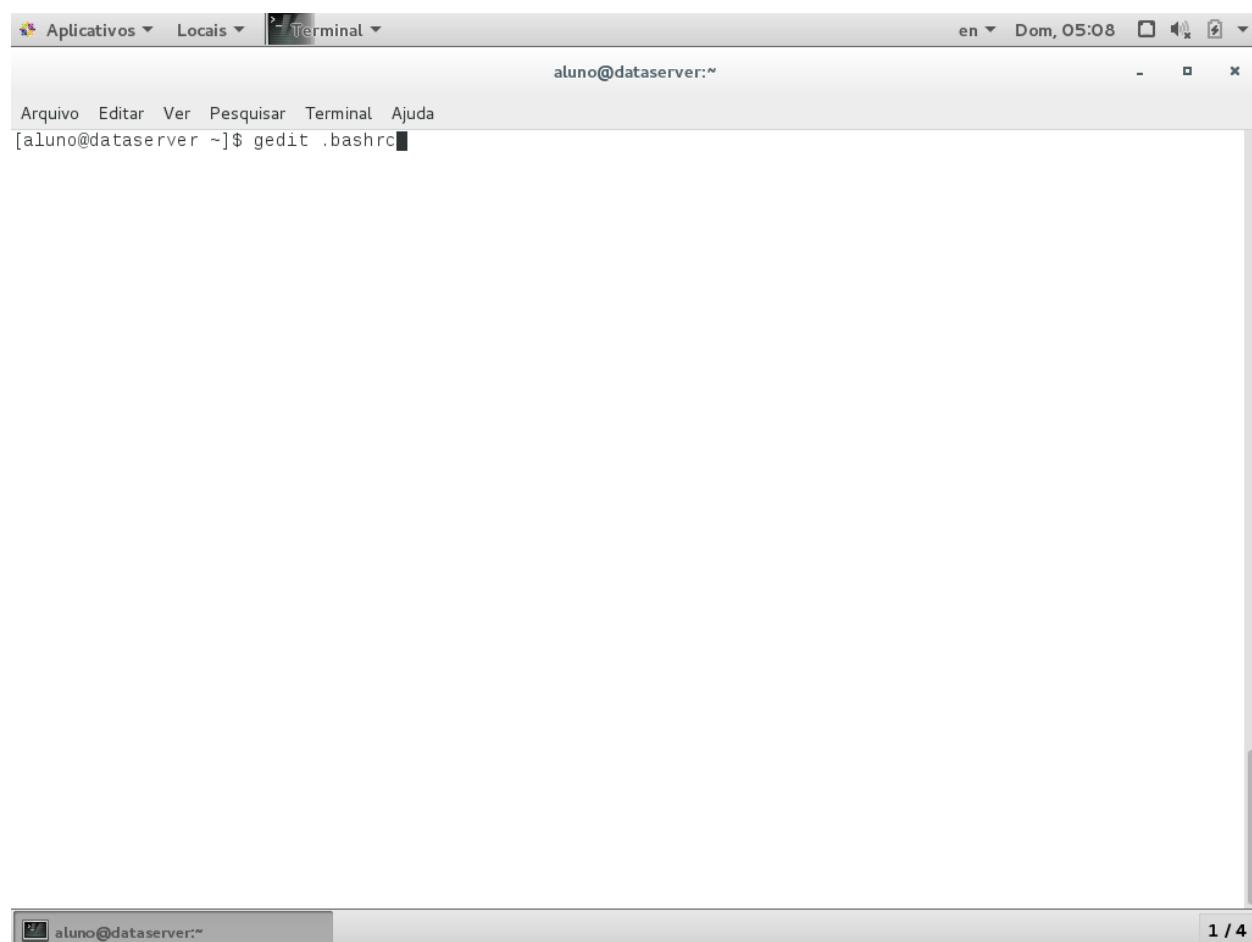
Criar um link simbólico na pasta /opt

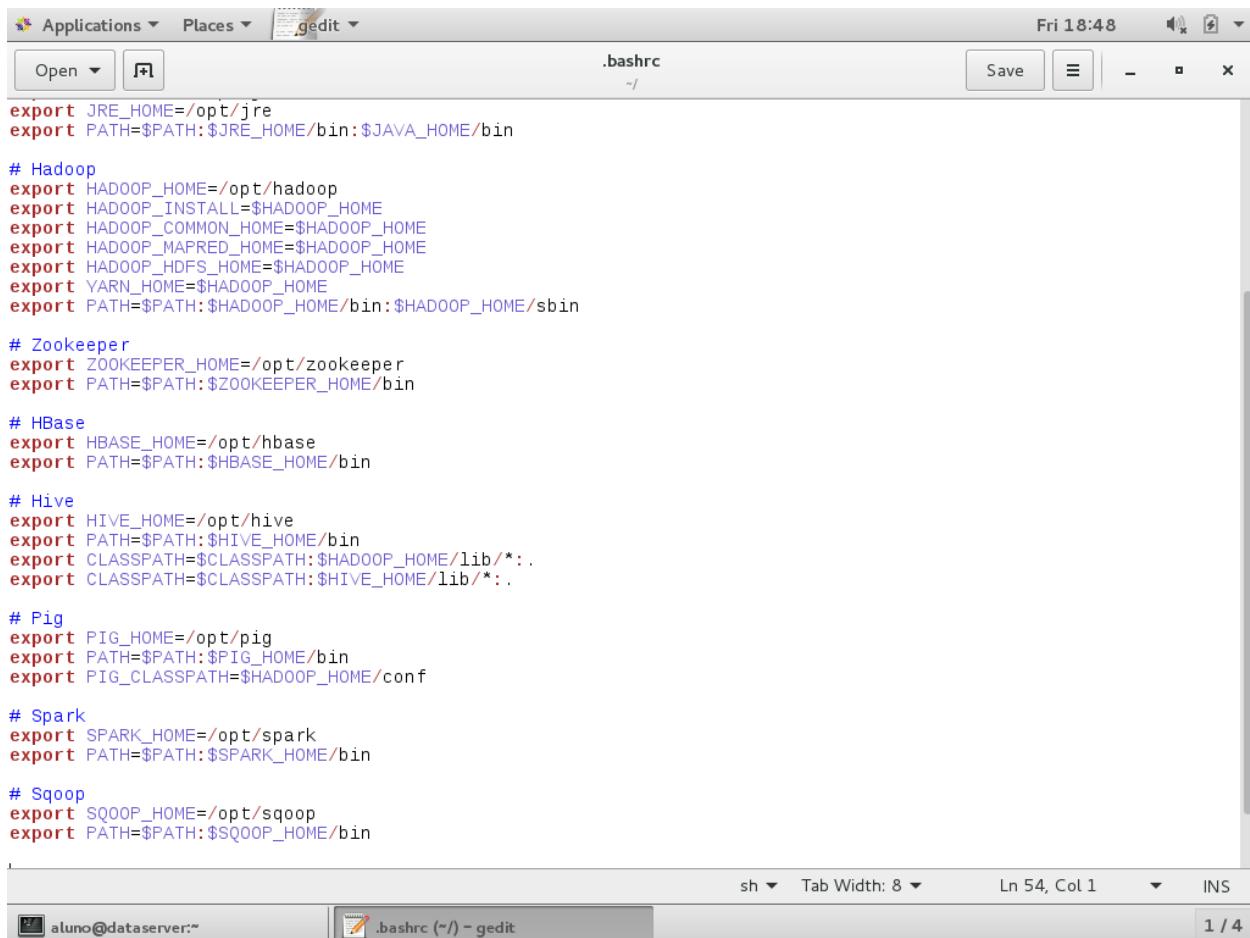


```
[aluno@dataserver Downloads]$ sudo ln -s /usr/local/sqoop-1.4.6-bin__hadoop-2.0.4-alpha/ /opt/sqoop
[aluno@dataserver Downloads]$ cd /opt/
[aluno@dataserver opt]$ ls -la
total 4
drwxr-xr-x. 2 root root 112 Sep 30 19:01 .
dr-xr-xr-x. 17 root root 4096 Sep 29 18:37 ..
lrwxrwxrwx. 1 root root 24 Sep 29 20:31 hadoop -> /usr/local/hadoop-2.7.3/
lrwxrwxrwx. 1 root root 23 Sep 30 17:04 hbase -> /usr/local/hbase-1.2.3/
lrwxrwxrwx. 1 root root 33 Sep 30 17:23 hive -> /usr/local/apache-hive-2.1.0-bin/
lrwxrwxrwx. 1 root root 24 Sep 29 18:56 jdk -> /usr/local/jdk1.8.0_102/
lrwxrwxrwx. 1 root root 65 Sep 29 18:59 jre -> /usr/lib/jvm/java-1.8.0-openjdk-1.8.0.102-1.b14.e17_2.x86_64/jre/
lrwxrwxrwx. 1 root root 22 Sep 30 18:00 pig -> /usr/local/pig-0.16.0/
lrwxrwxrwx. 1 root root 37 Sep 30 18:20 spark -> /usr/local/spark-2.0.0-bin-hadoop2.7/
lrwxrwxrwx. 1 root root 47 Sep 30 19:01 sqoop -> /usr/local/sqoop-1.4.6-bin__hadoop-2.0.4-alpha/
lrwxrwxrwx. 1 root root 27 Sep 30 16:49 zookeeper -> /usr/local/zookeeper-3.4.9/
[aluno@dataserver opt]$
```

Link criado

11.2. Configuração do Sqoop





```
Applications ▾ Places ▾ gedit ▾
Fri 18:48
Open + .bashrc
Save - x
export JRE_HOME=/opt/jre
export PATH=$PATH:$JRE_HOME/bin:$JAVA_HOME/bin

# Hadoop
export HADOOP_HOME=/opt/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Zookeeper
export ZOOKEEPER_HOME=/opt/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin

# HBase
export HBASE_HOME=/opt/hbase
export PATH=$PATH:$HBASE_HOME/bin

# Hive
export HIVE_HOME=/opt/hive
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:.
export CLASSPATH=$CLASSPATH:$HIVE_HOME/lib/*:.

# Pig
export PIG_HOME=/opt/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$HADOOP_HOME/conf

# Spark
export SPARK_HOME=/opt/spark
export PATH=$PATH:$SPARK_HOME/bin

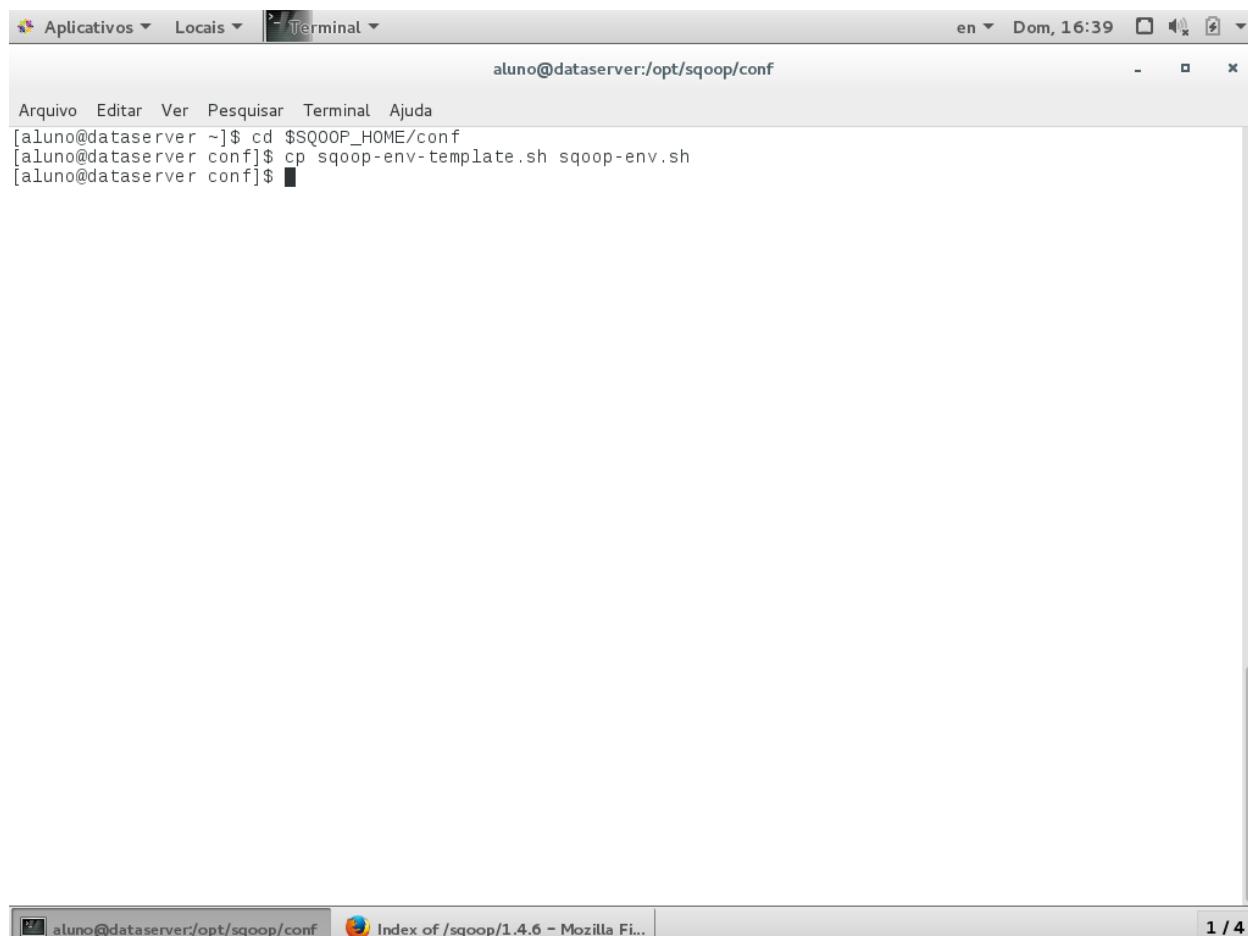
# Sqoop
export SQOOP_HOME=/opt/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
```

Incluir variáveis Sqoop

A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~'. The user has run the commands:

```
[aluno@dataserver ~]$ gedit .bashrc
[aluno@dataserver ~]$ source .bashrc
[aluno@dataserver ~]$
```

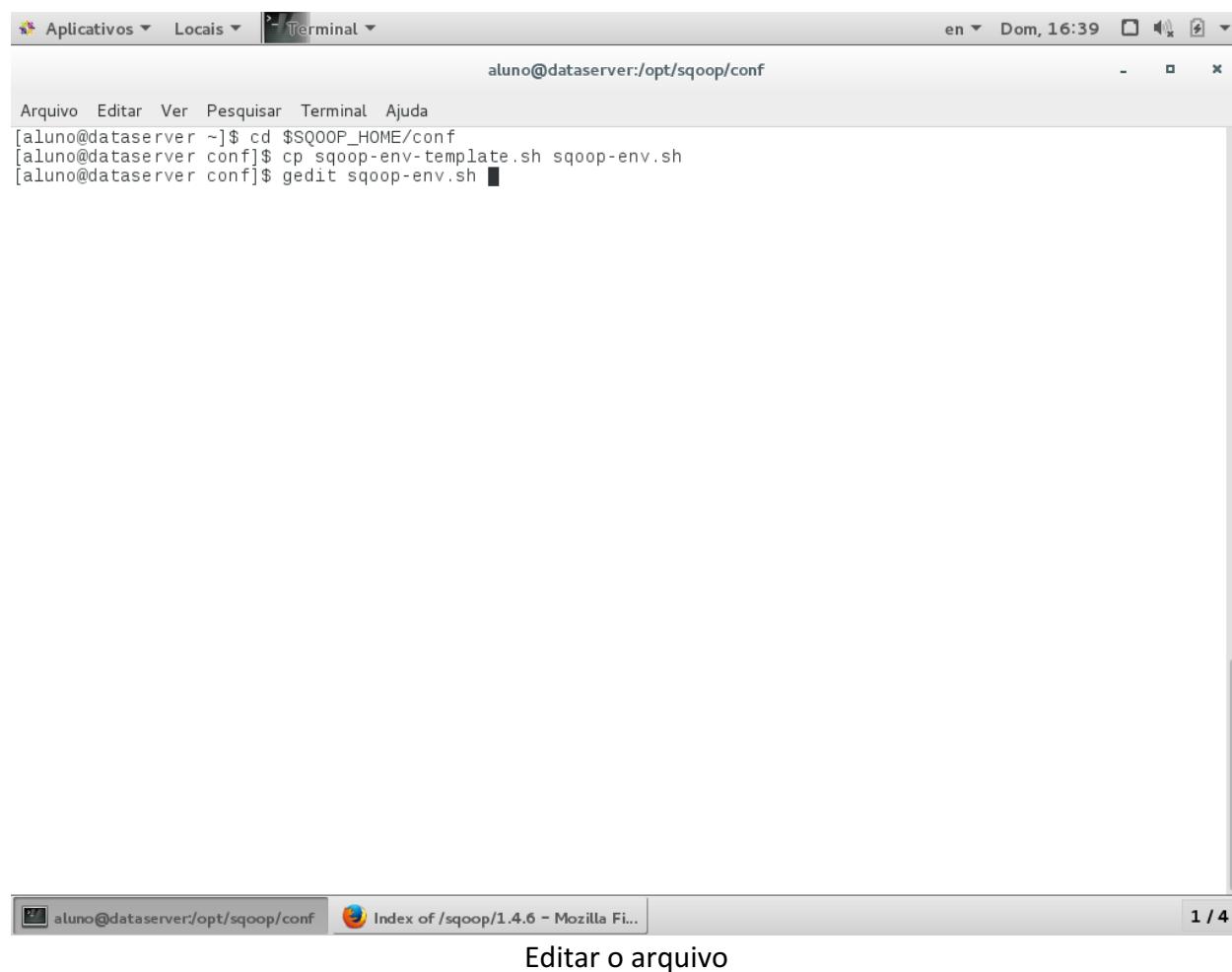
The terminal window is part of a larger desktop interface with icons for 'Aplicativos' and 'Locais' in the top bar. The status bar at the bottom shows 'aluno@dataserver:~' and '1 / 4'.



```
aluno@dataserver:~$ cd $SQOOP_HOME/conf
[aluno@dataserver conf]$ cp sqoop-env-template.sh sqoop-env.sh
[aluno@dataserver conf]$
```

A partir do template, criar o arquivo sqoop-env.sh e editá-lo

Instalação e Configuração do Ecosistema Hadoop

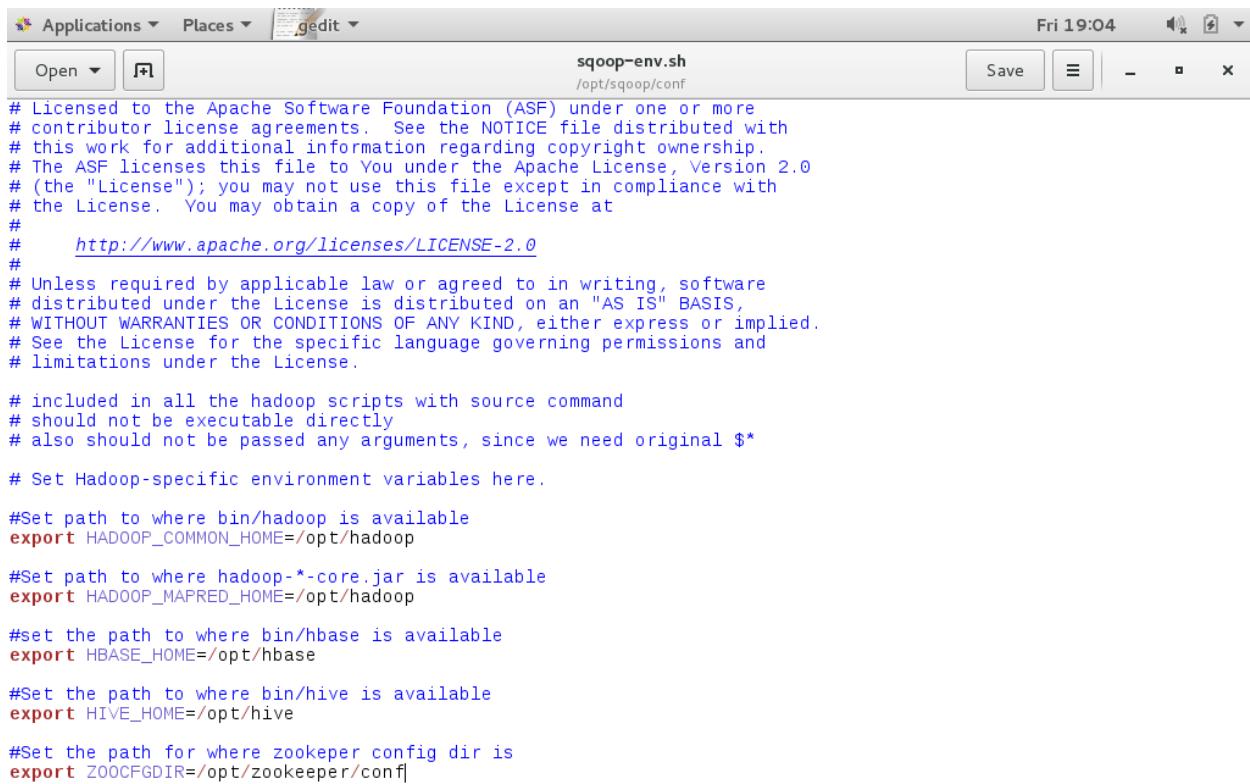


A screenshot of a Linux terminal window titled "Terminal". The window shows a command-line session:

```
[aluno@dataserver ~]$ cd $SQOOP_HOME/conf  
[aluno@dataserver conf]$ cp sqoop-env-template.sh sqoop-env.sh  
[aluno@dataserver conf]$ gedit sqoop-env.sh
```

The terminal window has a menu bar with "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The status bar at the bottom shows the path "aluno@dataserver:/opt/sqoop/conf" and the file name "Index of /sqoop/1.4.6 – Mozilla Fi...". A progress bar at the bottom right indicates "1 / 4".

Editar o arquivo



```

# Licensed to the Apache Software Foundation (ASF) under one or more
# contributor license agreements. See the NOTICE file distributed with
# this work for additional information regarding copyright ownership.
# The ASF licenses this file to You under the Apache License, Version 2.0
# (the "License"); you may not use this file except in compliance with
# the License. You may obtain a copy of the License at
#
#     http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# included in all the hadoop scripts with source command
# should not be executable directly
# also should not be passed any arguments, since we need original $*
# Set Hadoop-specific environment variables here.

#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME=/opt/hadoop

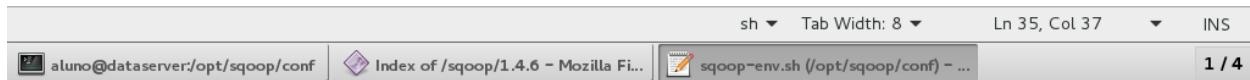
#Set path to where hadoop-*core.jar is available
export HADOOP_MAPRED_HOME=/opt/hadoop

#set the path to where bin/hbase is available
export HBASE_HOME=/opt/hbase

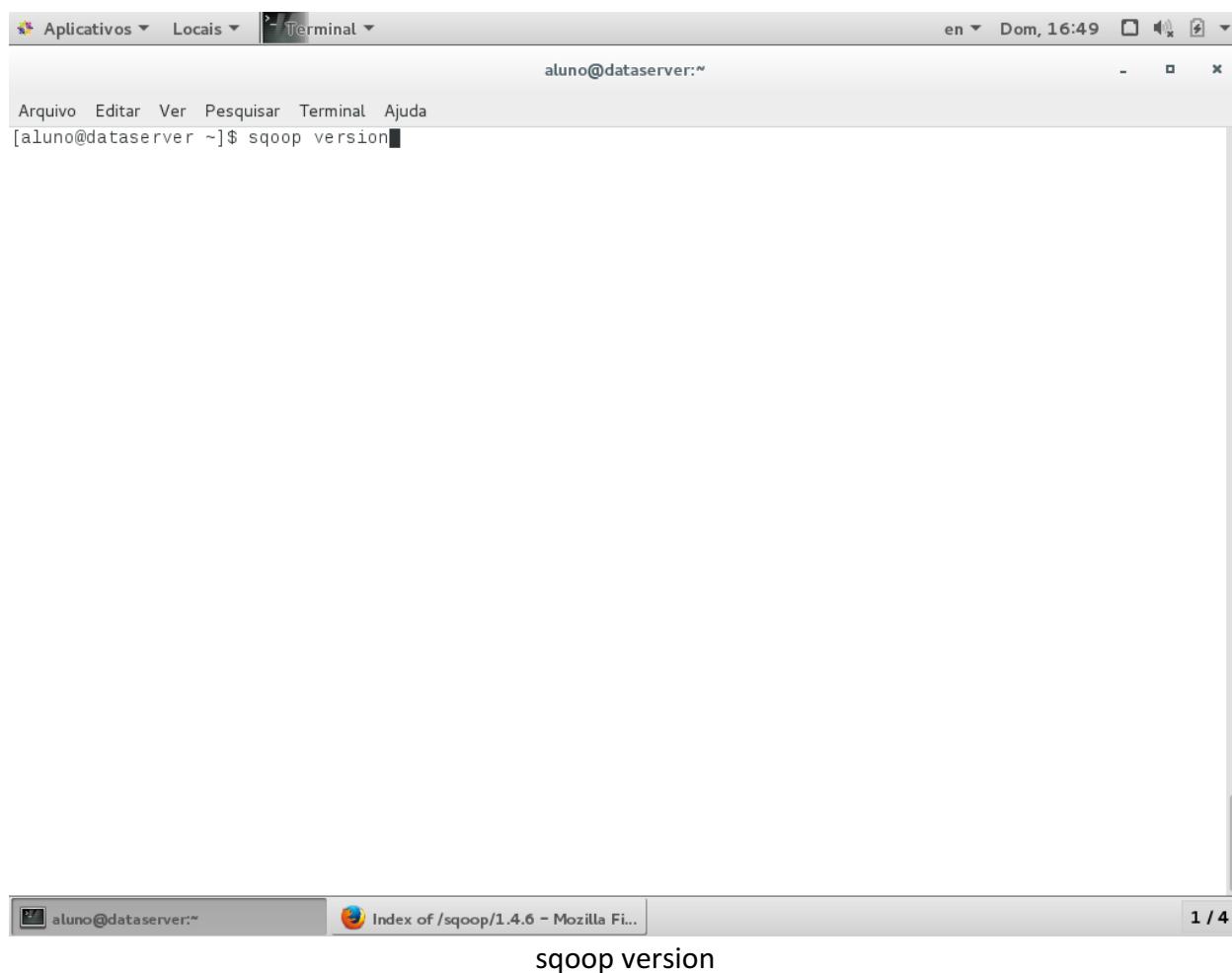
#Set the path to where bin/hive is available
export HIVE_HOME=/opt/hive

#Set the path for where zookeeper config dir is
export ZOOCFGDIR=/opt/zookeeper/conf

```

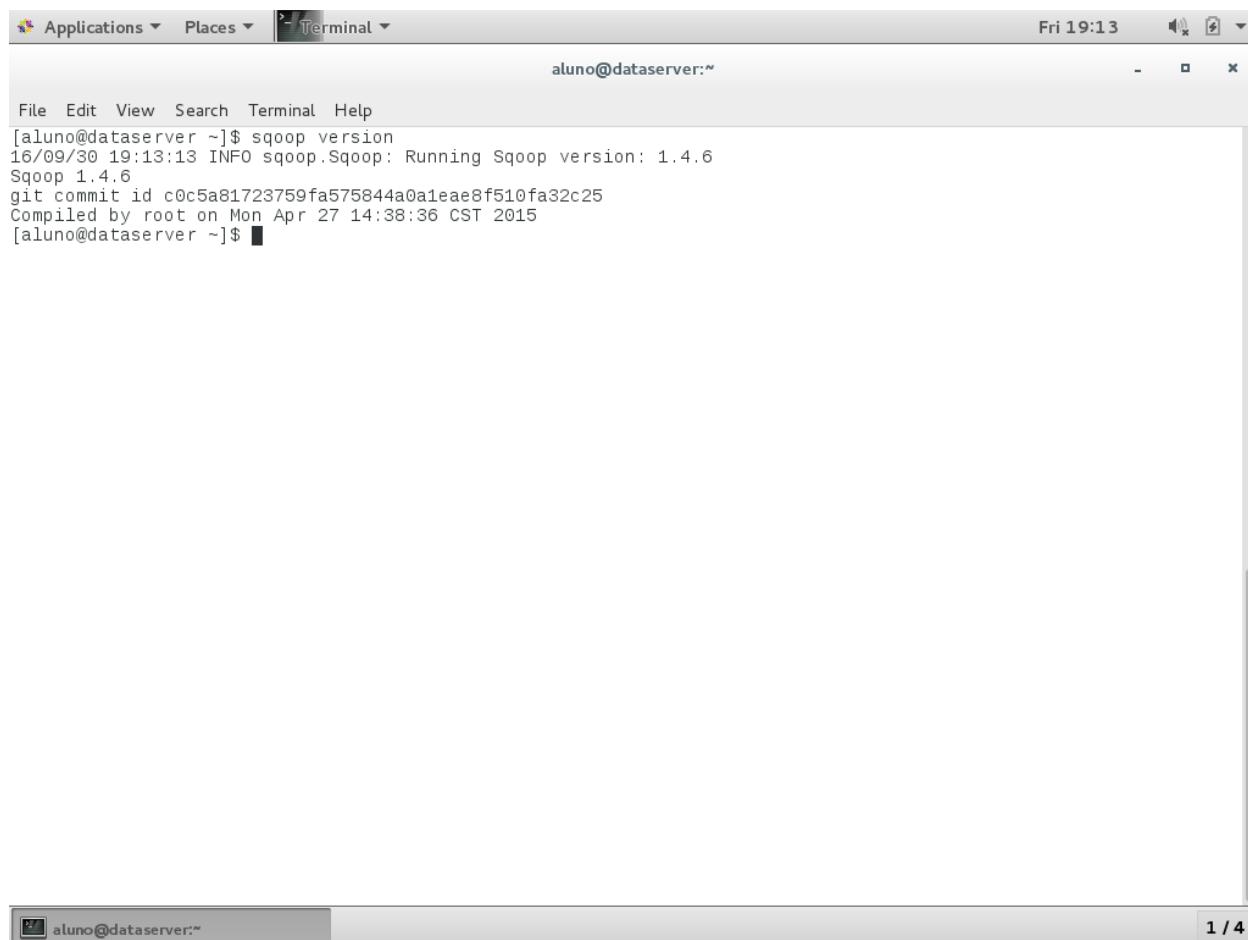


Editar variáveis conforme tela acima



A screenshot of a Linux desktop environment showing a terminal window. The terminal title bar says "Terminal". The window title is "aluno@dataserver:~". The menu bar includes "Arquivo", "Editar", "Ver", "Pesquisar", "Terminal", and "Ajuda". The command "sqoop version" is typed at the prompt. Below the terminal window, the desktop taskbar shows other open applications: a file manager window titled "aluno@dataserver:~" and a Mozilla Firefox browser window titled "Index of /sqoop/1.4.6 – Mozilla Fi...". A status bar at the bottom right indicates "1 / 4".

```
aluno@dataserver:~$ sqoop version
```



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:~'. The user has run the command 'sqoop version', which outputs the following information:

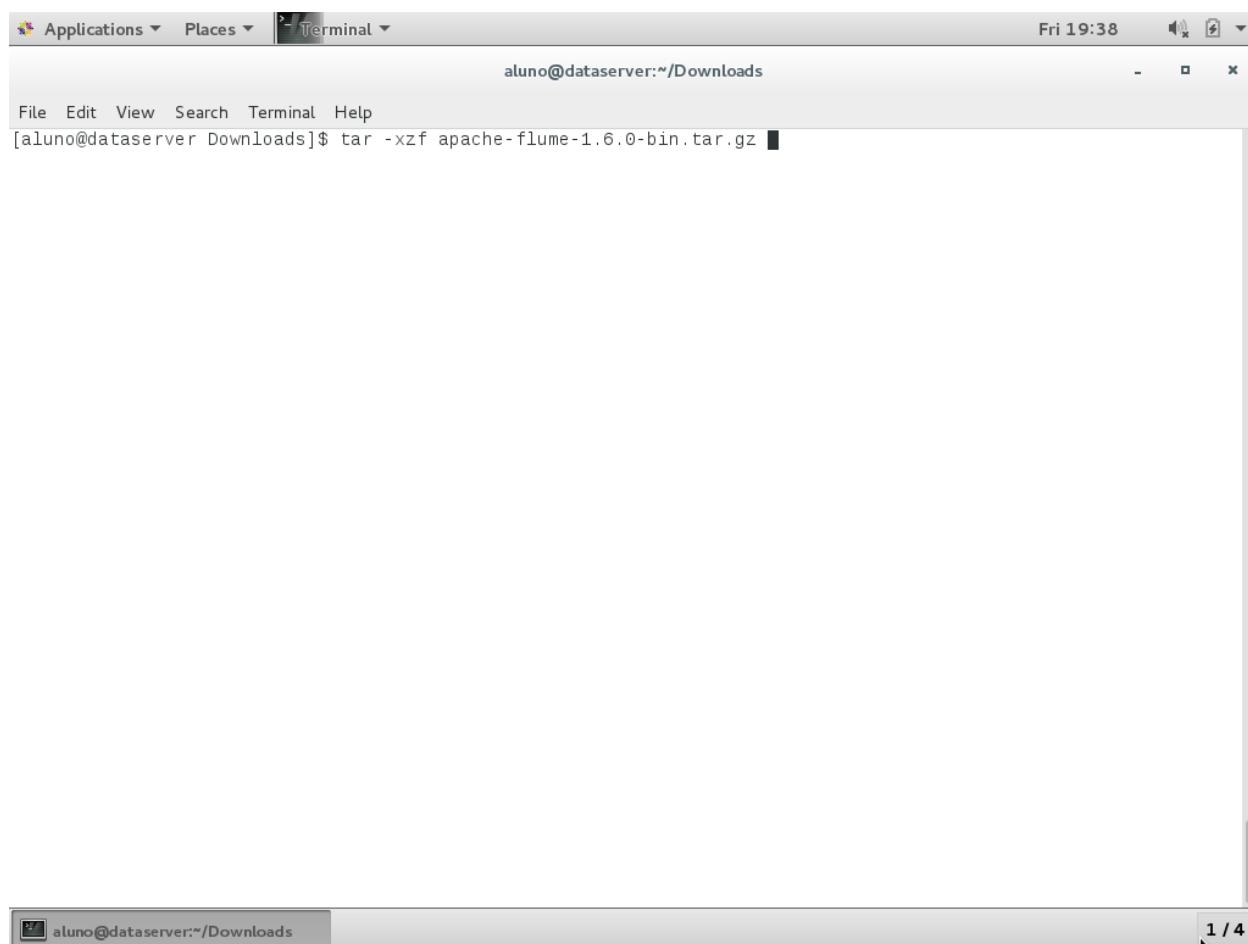
```
[aluno@dataserver ~]$ sqoop version
16/09/30 19:13:13 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Sqoop 1.4.6
git commit id c0c5a81723759fa575844a0a1eae8f510fa32c25
Compiled by root on Mon Apr 27 14:38:36 CST 2015
[aluno@dataserver ~]$
```

Sqoop version

12. Instalação e Configuração do Apache Flume

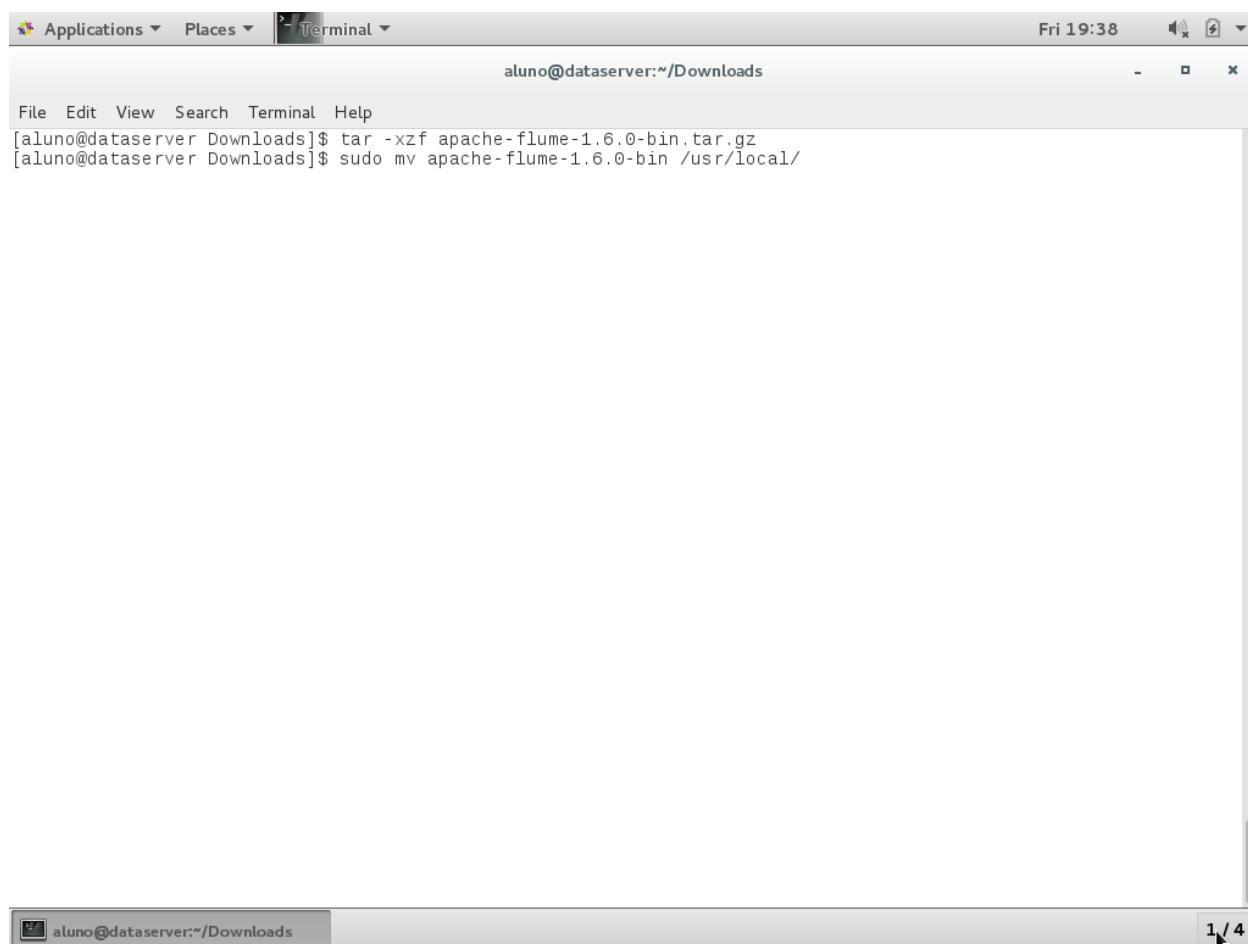
The screenshot shows the Apache Flume homepage. The URL in the address bar is <https://flume.apache.org>. The page features a large green title "Apache Flume™" with a feather icon. A sidebar on the left contains links for "How to Get Involved", "Documentation", "Releases", "Mailing lists", "Team", "Source Repository", and "Apache License". Under "Resources", there are links for "Flume Issue Tracking (Jira)", "Flume Wiki", "Getting Started Guide", "Jenkins Continuous Integration Server", "Sonar Code Quality Reports", and "Apache". At the bottom of the sidebar is a "Home" link. The main content area has a heading "Welcome to Apache Flume" and a description of what Flume is. It includes a diagram showing the Flume architecture: a "Web Server" sending data to an "Agent" (represented by a rectangle containing "Source", "Channel", and "Sink"). The "Agent" then sends data to an "HDFS" storage system.

Download do Apache Flume – Versão 1.6



A screenshot of a terminal window titled "Terminal". The window shows the command `tar -xzf apache-flume-1.6.0-bin.tar.gz` being run by the user `aluno` at the prompt `[aluno@dataserver Downloads]$`. The terminal interface includes a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The status bar at the bottom indicates "1 / 4".

Decomprimir o arquivo



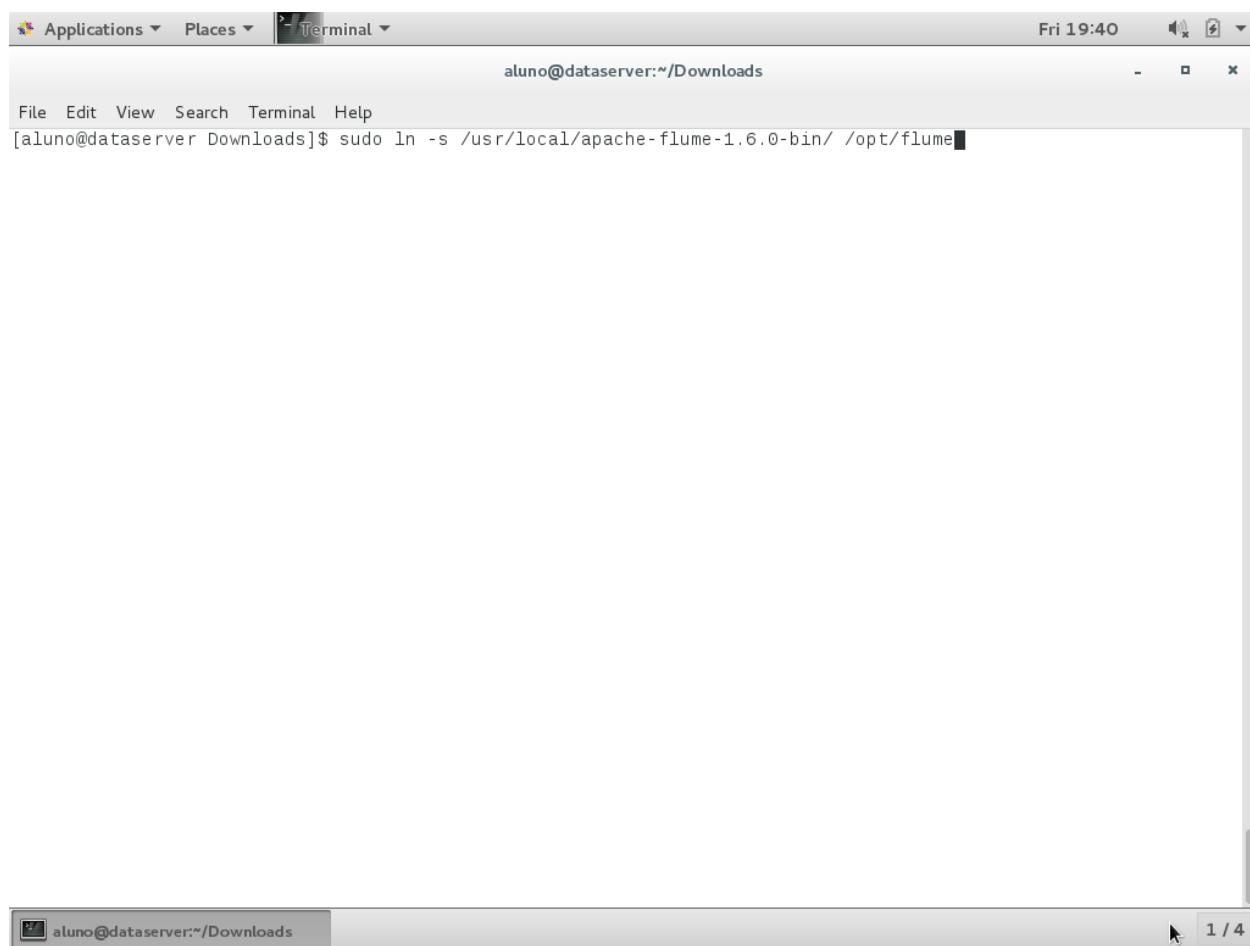
A screenshot of a terminal window titled "Terminal". The window shows a command-line session with the following commands:

```
[aluno@dataserver Downloads]$ tar -xzf apache-flume-1.6.0-bin.tar.gz  
[aluno@dataserver Downloads]$ sudo mv apache-flume-1.6.0-bin /usr/local/
```

The terminal window has a standard Linux desktop interface with a title bar, menu bar, and status bar indicating the date and time.

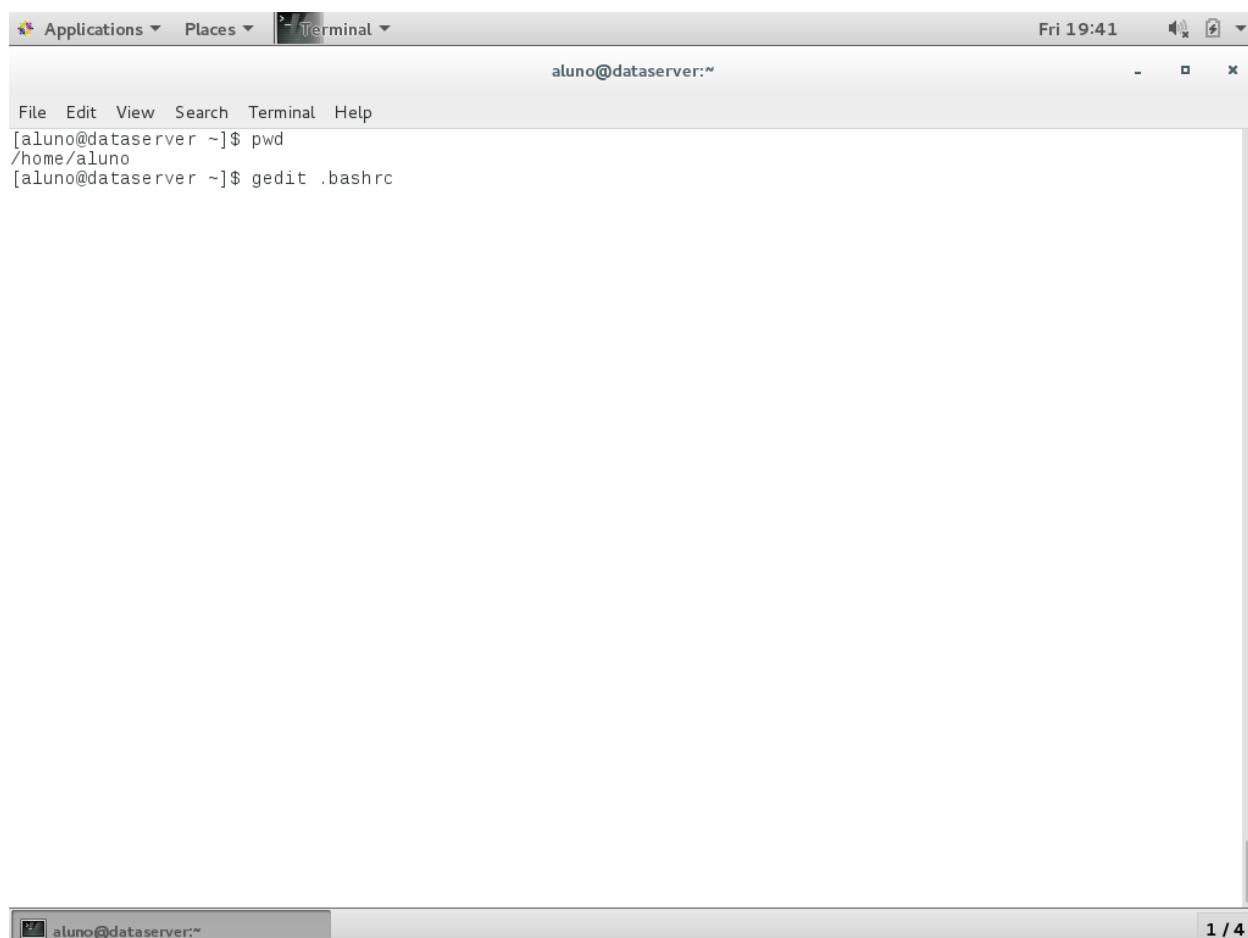
Mover o diretório para /usr/local

1 / 4



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The terminal prompt is "aluno@dataserver:~/Downloads". The user has typed the command "sudo ln -s /usr/local/apache-flume-1.6.0-bin/ /opt/flume" and is pressing the Enter key. The status bar at the bottom of the terminal shows "1 / 4".

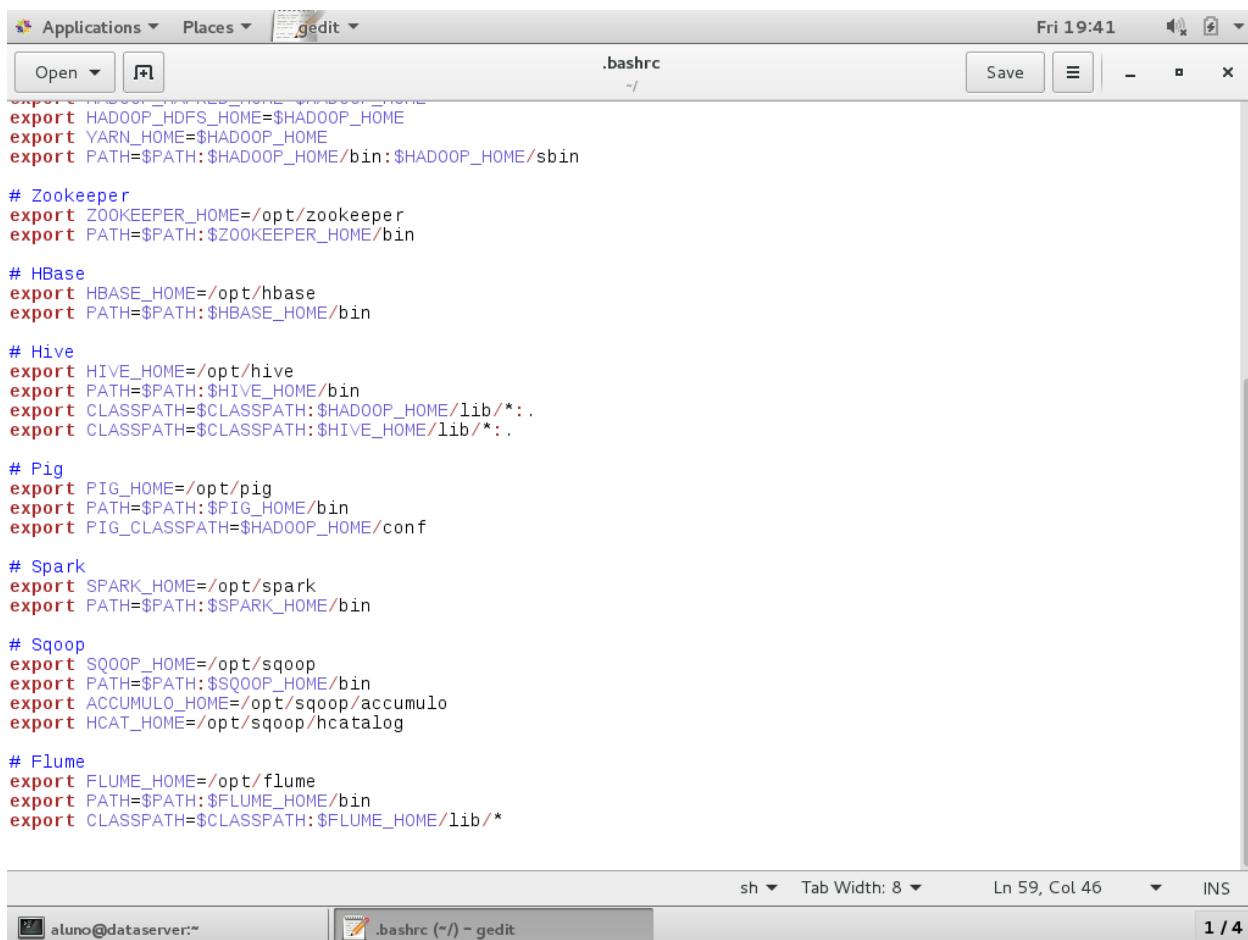
Criar o link simbólico



A screenshot of a Linux desktop environment showing a terminal window. The window title is "Terminal". The status bar at the top right shows "Fri 19:41". The terminal window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The command line shows the user's session: [aluno@dataserver ~]\$ pwd /home/aluno [aluno@dataserver ~]\$ gedit .bashrc

Editar as variáveis de ambiente

1 / 4



The screenshot shows a terminal window titled "gedit" with the file ".bashrc" open. The file contains configuration code for various Hadoop components. The code includes export statements for HDFS, YARN, PATH, Zookeeper, HBase, Hive, Pig, Spark, Sqoop, and Flume. The code is color-coded, with keywords in red and paths in blue.

```
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

# Zookeeper
export ZOOKEEPER_HOME=/opt/zookeeper
export PATH=$PATH:$ZOOKEEPER_HOME/bin

# HBase
export HBASE_HOME=/opt/hbase
export PATH=$PATH:$HBASE_HOME/bin

# Hive
export HIVE_HOME=/opt/hive
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:$HADOOP_HOME/lib/*:.
export CLASSPATH=$CLASSPATH:$HIVE_HOME/lib/*:.

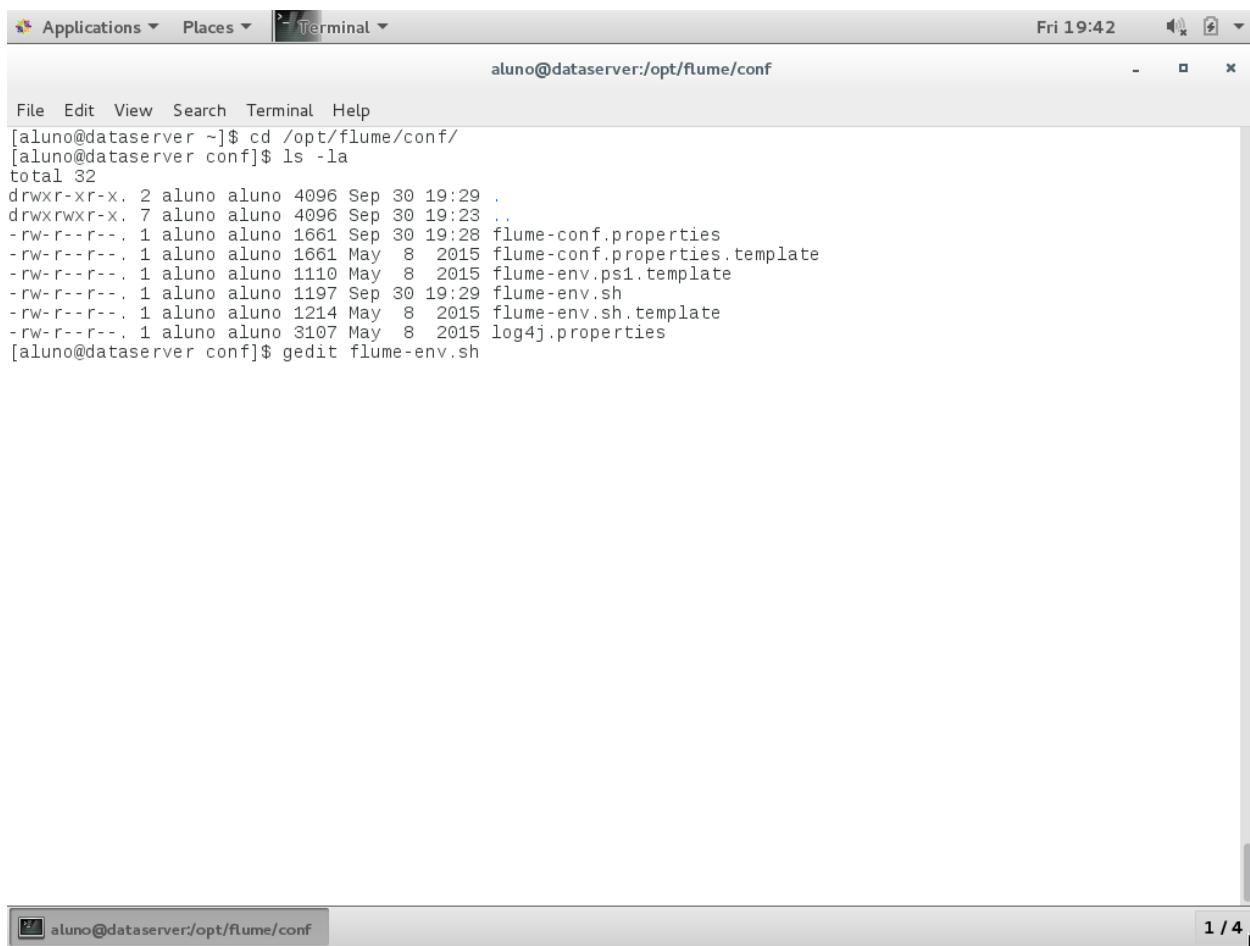
# Pig
export PIG_HOME=/opt/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$HADOOP_HOME/conf

# Spark
export SPARK_HOME=/opt/spark
export PATH=$PATH:$SPARK_HOME/bin

# Sqoop
export SQOOP_HOME=/opt/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
export ACCUMULO_HOME=/opt/sqoop/accumulo
export HCAT_HOME=/opt/sqoop/hcatalog

# Flume
export FLUME_HOME=/opt/flume
export PATH=$PATH:$FLUME_HOME/bin
export CLASSPATH=$CLASSPATH:$FLUME_HOME/lib/*
```

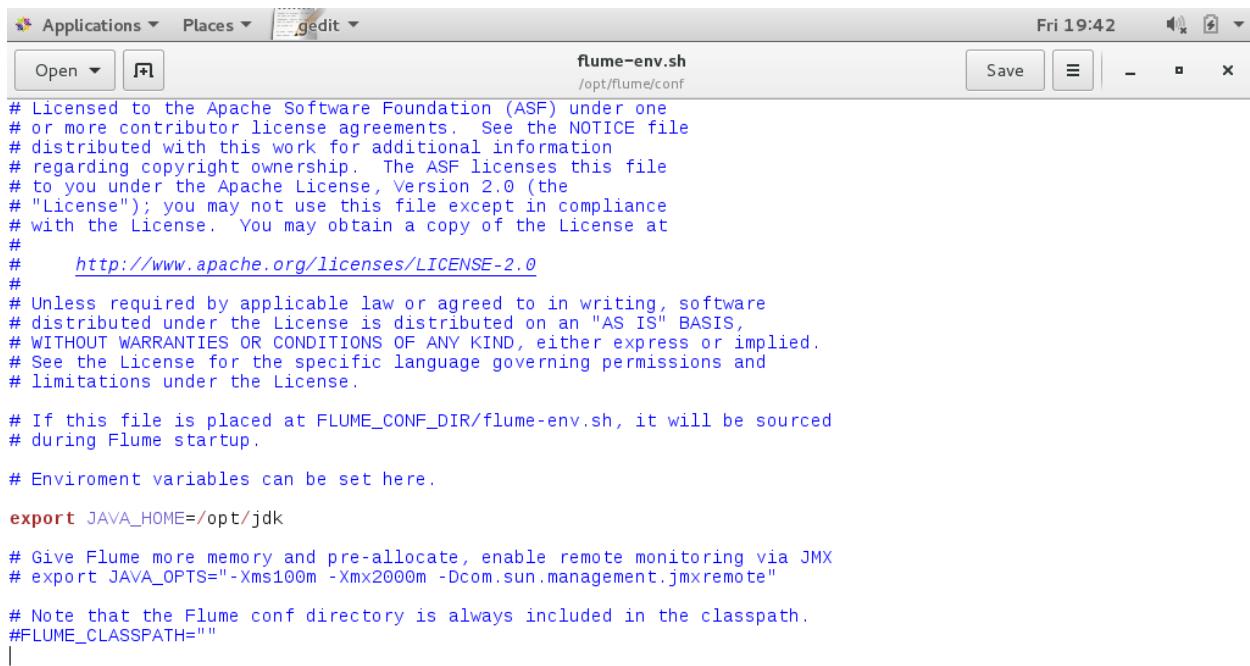
Variáveis de ambiente para o Flume



The screenshot shows a terminal window titled 'Terminal' with the command prompt 'aluno@dataserver:~/opt/flume/conf'. The user has run the command 'ls -la' to list files in the directory. The output shows several files including 'flume-conf.properties', 'flume-env.ps1.template', and 'log4j.properties'. The user then runs 'gedit flume-env.sh' to edit the configuration file.

```
[aluno@dataserver ~]$ cd /opt/flume/conf/
[aluno@dataserver conf]$ ls -la
total 32
drwxr-xr-x. 2 aluno aluno 4096 Sep 30 19:29 .
drwxrwxr-x. 7 aluno aluno 4096 Sep 30 19:23 ..
-rw-r--r--. 1 aluno aluno 1661 Sep 30 19:28 flume-conf.properties
-rw-r--r--. 1 aluno aluno 1661 May  8  2015 flume-conf.properties.template
-rw-r--r--. 1 aluno aluno 1110 May  8  2015 flume-env.ps1.template
-rw-r--r--. 1 aluno aluno 1197 Sep 30 19:29 flume-env.sh
-rw-r--r--. 1 aluno aluno 1214 May  8  2015 flume-env.sh.template
-rw-r--r--. 1 aluno aluno 3107 May  8  2015 log4j.properties
[aluno@dataserver conf]$ gedit flume-env.sh
```

Editar o arquivo flume-env.sh



```

Applications ▾ Places ▾ gedit ▾
Open ▾ flume-env.sh
flume-env.sh
/opt/flume/conf
Fri 19:42
Save | - | x
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License. You may obtain a copy of the License at
#
#     http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# If this file is placed at FLUME_CONF_DIR/flume-env.sh, it will be sourced
# during Flume startup.

# Environment variables can be set here.

export JAVA_HOME=/opt/jdk

# Give Flume more memory and pre-allocate, enable remote monitoring via JMX
# export JAVA_OPTS="-Xms100m -Xmx2000m -Dcom.sun.management.jmxremote"

# Note that the Flume conf directory is always included in the classpath.
#FLUME_CLASSPATH=""
|

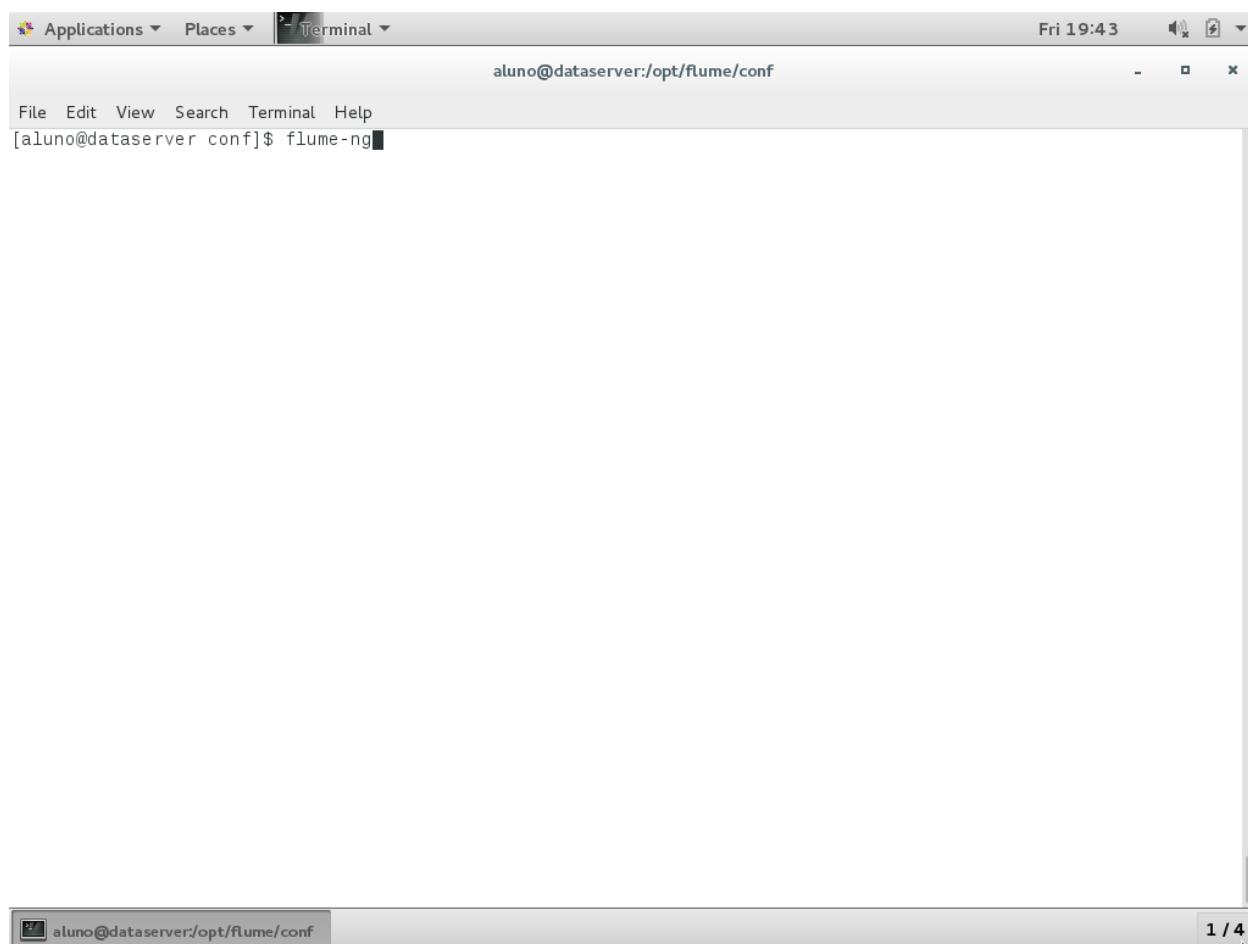
```



```

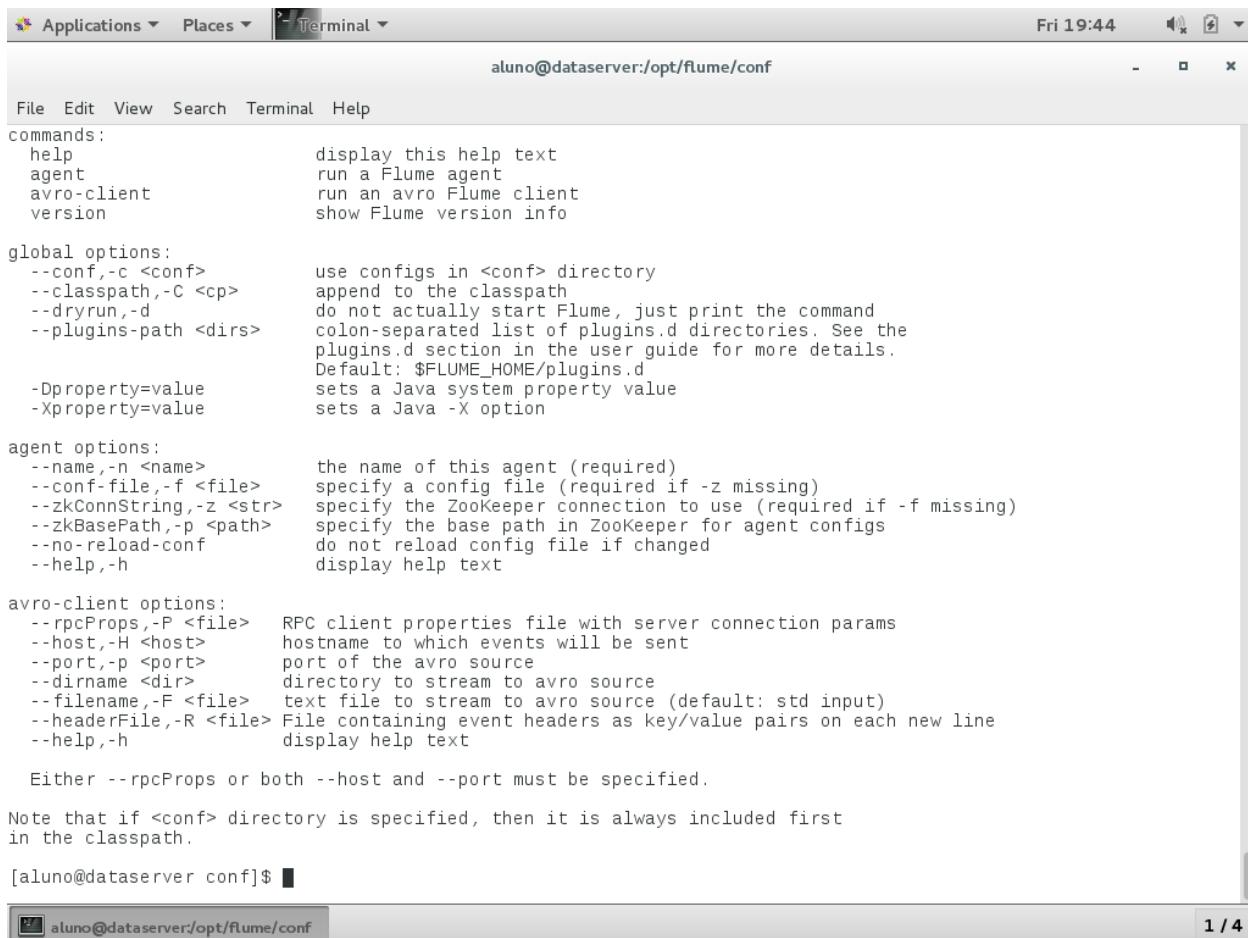
sh ▾ Tab Width: 8 ▾ Ln 29, Col 1 ▾ INS
aluno@dataserver:/opt/flume/conf | flume-env.sh (/opt/flume/conf) - ...
1 / 4
Acrecentar o JAVA_HOME

```



A screenshot of a terminal window titled 'Terminal'. The window is part of a desktop environment with a menu bar at the top showing 'Applications', 'Places', and 'Terminal'. The status bar at the top right shows the date and time as 'Fri 19:43'. The terminal window itself has a title bar with the text 'aluno@dataserver:/opt/flume/conf'. The menu bar within the terminal includes 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The main area of the terminal shows the command '[aluno@dataserver conf]\$ flume-ng' followed by a cursor. At the bottom of the terminal window, there is a status bar with a small icon, the text 'aluno@dataserver:/opt/flume/conf', and a page number '1 / 4'.

Testar a instalação



```

Applications ▾ Places ▾ Terminal ▾
Fri 19:44
aluno@dataserver:/opt/flume/conf
File Edit View Search Terminal Help
commands:
  help           display this help text
  agent          run a Flume agent
  avro-client    run an avro Flume client
  version        show Flume version info

global options:
  --conf,-c <conf>      use configs in <conf> directory
  --classpath,-C <cp>    append to the classpath
  --dryrun,-d            do not actually start Flume, just print the command
  --plugins-path <dirs>  colon-separated list of plugins.d directories. See the
                        plugins.d section in the user guide for more details.
                        Default: $FLUME_HOME/plugins.d
  -Dproperty=value     sets a Java system property value
  -Xproperty=value     sets a Java -X option

agent options:
  --name,-n <name>       the name of this agent (required)
  --conf-file,-f <file>   specify a config file (required if -z missing)
  --zkConnString,-z <str> specify the ZooKeeper connection to use (required if -f missing)
  --zkBasePath,-p <path>  specify the base path in ZooKeeper for agent configs
  --no-reload-conf       do not reload config file if changed
  --help,-h              display help text

avro-client options:
  --rpcProps,-P <file>   RPC client properties file with server connection params
  --host,-H <host>        hostname to which events will be sent
  --port,-p <sport>       port of the avro source
  --dirname <dir>         directory to stream to avro source
  --filename,-F <file>    text file to stream to avro source (default: std input)
  --headerFile,-R <file>  File containing event headers as key/value pairs on each new line
  --help,-h              display help text

Either --rpcProps or both --host and --port must be specified.

Note that if <conf> directory is specified, then it is always included first
in the classpath.

[aluno@dataserver conf]$ 

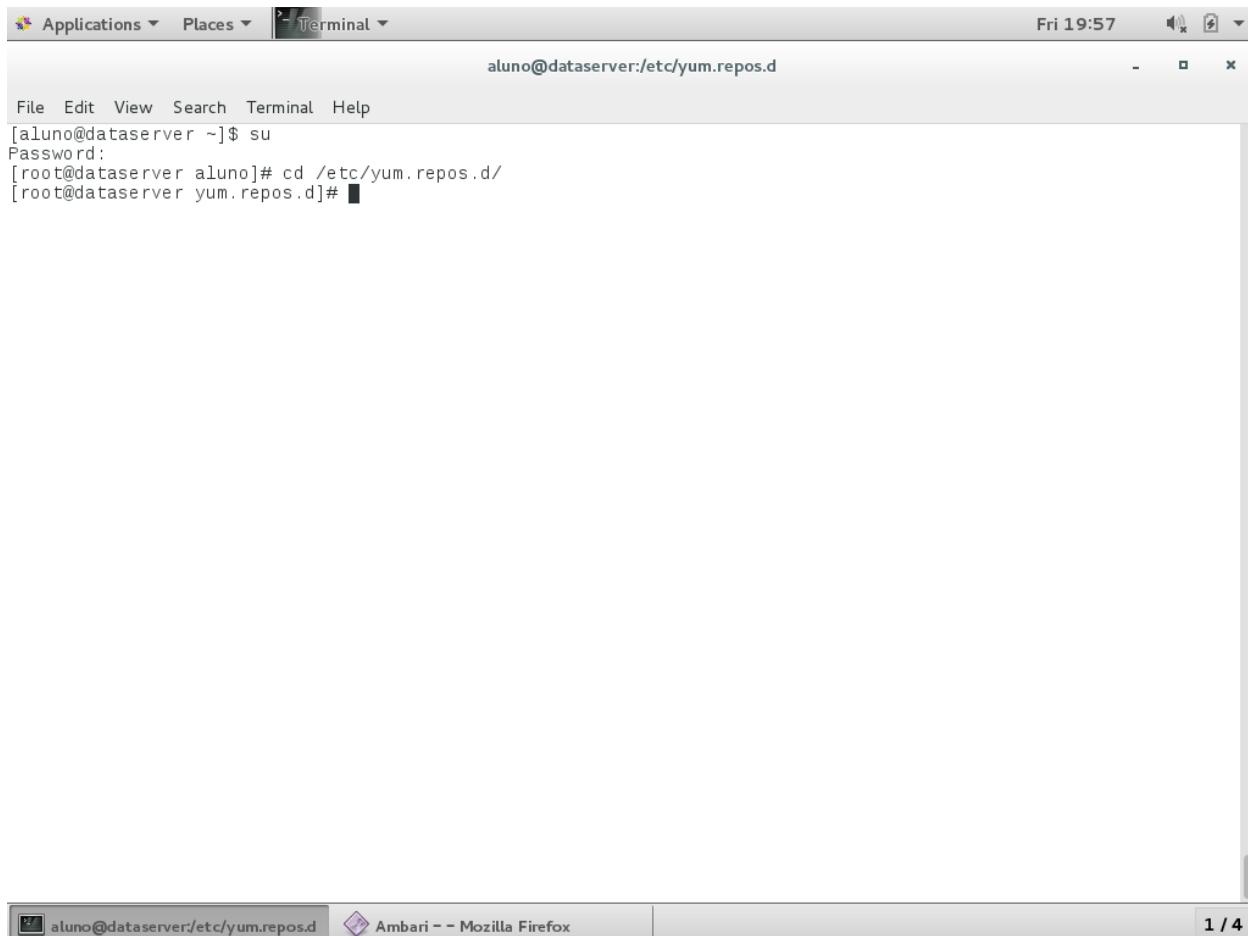
```

Flume instalado com sucesso

1 / 4

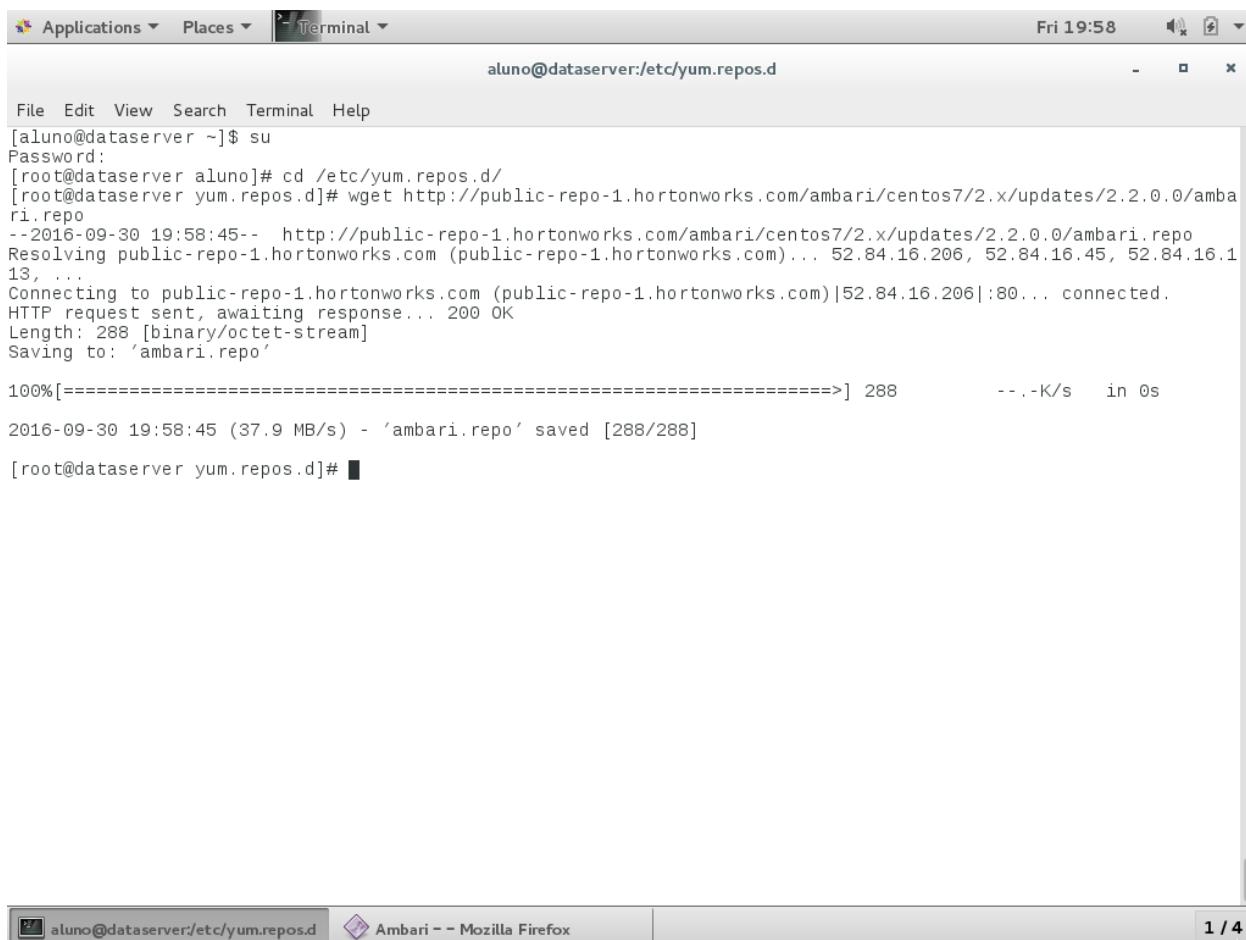
13. Instalação e Configuração do Ambari (Opcional)

Nota: No CentOS, o Ambari pode ser instalado mais facilmente através do gerenciador de pacotes yum.



```
File Edit View Search Terminal Help
[aluno@dataserver ~]$ su
Password:
[root@dataserver alumno]# cd /etc/yum.repos.d/
[root@dataserver yum.repos.d]#
```

Conectado como root, acessar o diretório de repositórios do CentOS



The screenshot shows a terminal window titled 'Terminal' with the command line interface. The user is root ('aluno@dataserver') and is navigating to the '/etc/yum.repos.d' directory. They run 'wget' to download the 'ambari.repo' file from a public repository. The download is completed at 19:58:45 on 2016-09-30, with a speed of 37.9 MB/s and a total size of 288 KB.

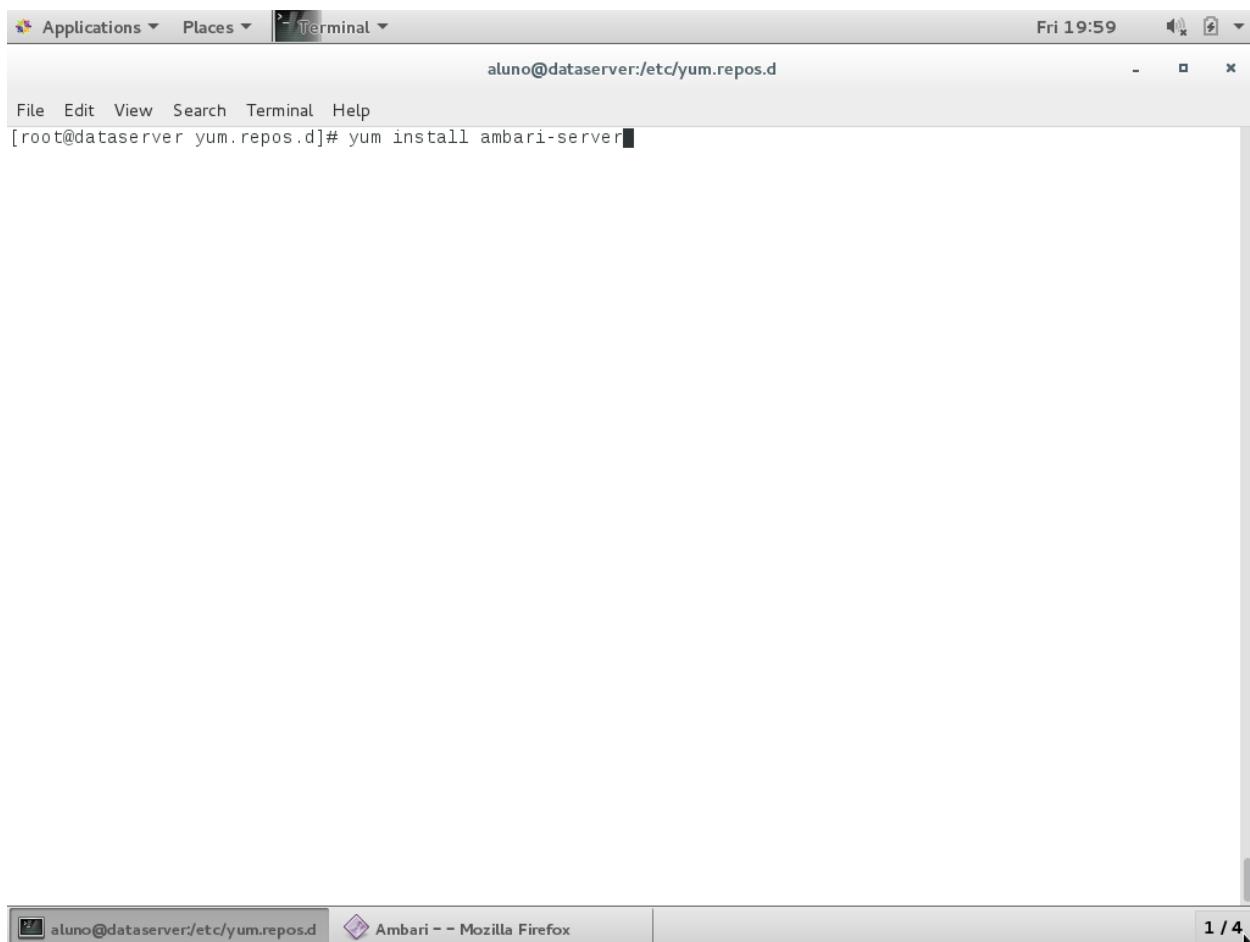
```
File Edit View Search Terminal Help
[aluno@dataserver ~]$ su
Password:
[root@dataserver aluno]# cd /etc/yum.repos.d/
[root@dataserver yum.repos.d]# wget http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.2.0.0/ambari.repo
--2016-09-30 19:58:45-- http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.2.0.0/ambari.repo
Resolving public-repo-1.hortonworks.com (public-repo-1.hortonworks.com)... 52.84.16.206, 52.84.16.45, 52.84.16.1
13, ...
Connecting to public-repo-1.hortonworks.com (public-repo-1.hortonworks.com)|52.84.16.206|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 288 [binary/octet-stream]
Saving to: 'ambari.repo'

100%[=====] 288 --.-K/s in 0s

2016-09-30 19:58:45 (37.9 MB/s) - 'ambari.repo' saved [288/288]

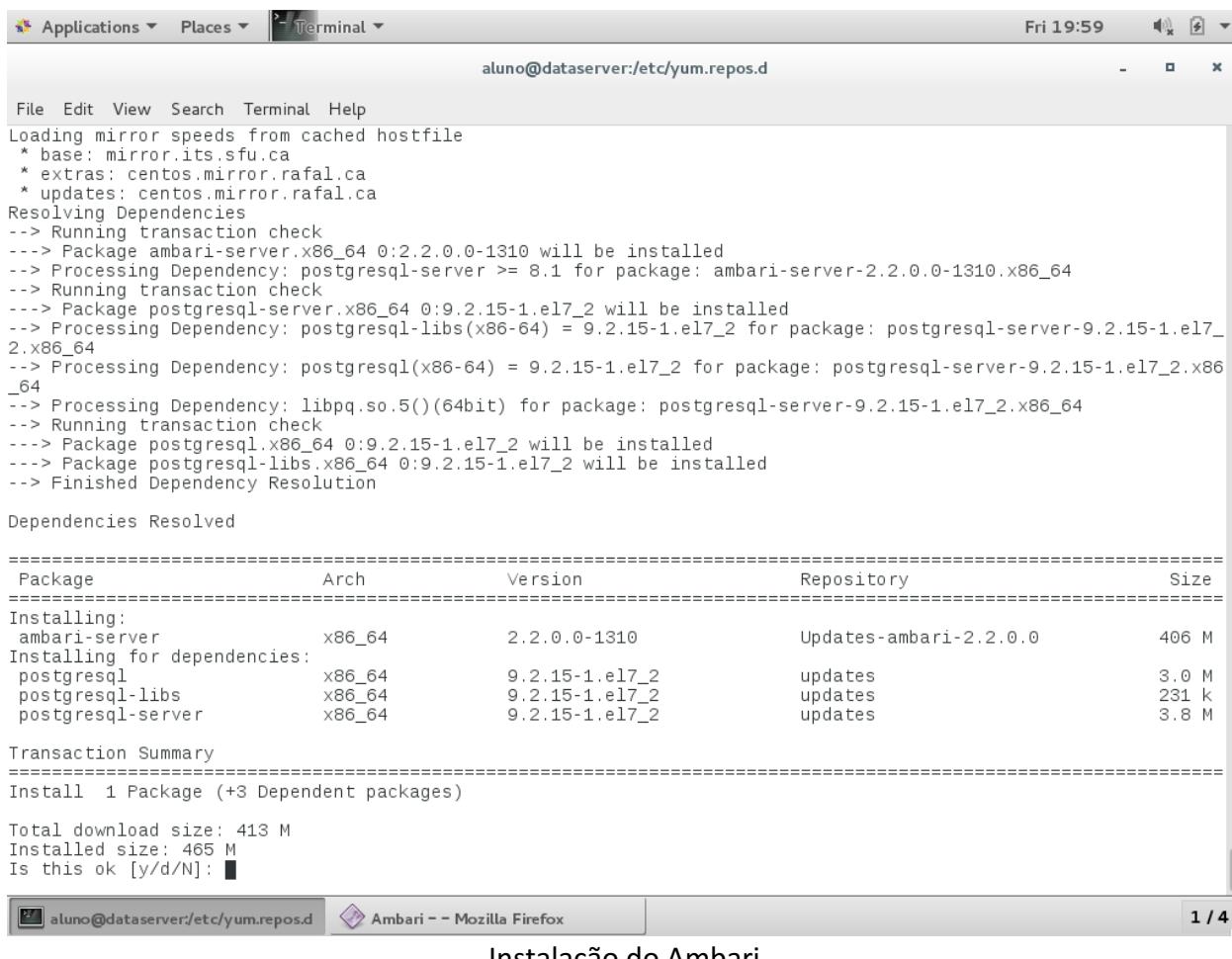
[root@dataserver yum.repos.d]#
```

Download do arquivo de repositório do Ambari



A screenshot of a Linux desktop environment showing a terminal window. The window title is 'Terminal'. The terminal prompt is 'aluno@dataserver:/etc/yum.repos.d'. The user has typed the command '[root@dataserver yum.repos.d]# yum install ambari-server' and is waiting for the output.

Como root, executar: yum install ambari-server



```

Applications ▾ Places ▾ Terminal ▾
Fri 19:59 - x ×
aluno@dataserver:/etc/yum.repos.d

File Edit View Search Terminal Help
Loading mirror speeds from cached hostfile
 * base: mirror.its.sfu.ca
 * extras: centos.mirror.rafal.ca
 * updates: centos.mirror.rafal.ca
Resolving Dependencies
--> Running transaction check
--> Package ambari-server.x86_64 0:2.2.0.0-1310 will be installed
--> Processing Dependency: postgresql-server >= 8.1 for package: ambari-server-2.2.0.0-1310.x86_64
--> Running transaction check
--> Package postgresql-server.x86_64 0:9.2.15-1.el7_2 will be installed
--> Processing Dependency: postgresql-libs(x86-64) = 9.2.15-1.el7_2 for package: postgresql-server-9.2.15-1.el7_2.x86_64
--> Processing Dependency: postgresql(x86-64) = 9.2.15-1.el7_2 for package: postgresql-server-9.2.15-1.el7_2.x86_64
--> Processing Dependency: libpq.so.5()(64bit) for package: postgresql-server-9.2.15-1.el7_2.x86_64
--> Running transaction check
--> Package postgresql.x86_64 0:9.2.15-1.el7_2 will be installed
--> Package postgresql-libs.x86_64 0:9.2.15-1.el7_2 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

=====
Package           Arch      Version       Repository      Size
=====
Installing:
ambari-server    x86_64   2.2.0.0-1310  Updates-ambari-2.2.0.0  406 M
Installing for dependencies:
postgresql        x86_64   9.2.15-1.el7_2  updates          3.0 M
postgresql-libs   x86_64   9.2.15-1.el7_2  updates          231 k
postgresql-server x86_64   9.2.15-1.el7_2  updates          3.8 M

Transaction Summary
=====
Install 1 Package (+3 Dependent packages)

Total download size: 413 M
Installed size: 465 M
Is this ok [y/d/N]: ■

```

aluno@dataserver:/etc/yum.repos.d | Ambari -- Mozilla Firefox | 1 / 4

Instalação do Ambari

```

aluno@dataserver:/etc/yum.repos.d
File Edit View Search Terminal Help
Is this ok [y/d/N]: y
Downloading packages:
(1/4): postgresql-libs-9.2.15-1.el7_2.x86_64.rpm | 231 kB 00:00:00
(2/4): postgresql-server-9.2.15-1.el7_2.x86_64.rpm | 3.8 MB 00:00:02
(3/4): postgresql-9.2.15-1.el7_2.x86_64.rpm | 3.0 MB 00:00:02
warning: /var/cache/yum/x86_64/7/Updates-ambari-2.2.0.0/packages/ambari-server-2.2.0.0-1310.x86_64.rpm: Header V4 RSA/SHA1 Signature, key ID 07513cad: NOKEY
Public key for ambari-server-2.2.0.0-1310.x86_64.rpm is not installed
(4/4): ambari-server-2.2.0.0-1310.x86_64.rpm | 406 MB 00:03:52
-----
Total                                         1.8 MB/s | 413 MB 00:03:52
Retrieving key from http://public-repo-1.hortonworks.com/ambari/centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
Importing GPG key 0x07513CAD:
  Userid   : "Jenkins (HDP Builds) <jenkin@hortonworks.com>"
  Fingerprint: df52 ed4f 7a3a 5882 c099 4c66 b973 3a7a 0751 3cad
  From     : http://public-repo-1.hortonworks.com/ambari/centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
Is this ok [y/N]: y
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Installing : postgresql-libs-9.2.15-1.el7_2.x86_64                               1/4
  Installing : postgresql-9.2.15-1.el7_2.x86_64                                2/4
  Installing : postgresql-server-9.2.15-1.el7_2.x86_64                            3/4
  Installing : ambari-server-2.2.0.0-1310.x86_64                                4/4
  Verifying  : postgresql-libs-9.2.15-1.el7_2.x86_64                               1/4
  Verifying  : postgresql-server-9.2.15-1.el7_2.x86_64                            2/4
  Verifying  : ambari-server-2.2.0.0-1310.x86_64                                3/4
  Verifying  : postgresql-9.2.15-1.el7_2.x86_64                                4/4

Installed:
  ambari-server.x86_64 0:2.2.0.0-1310

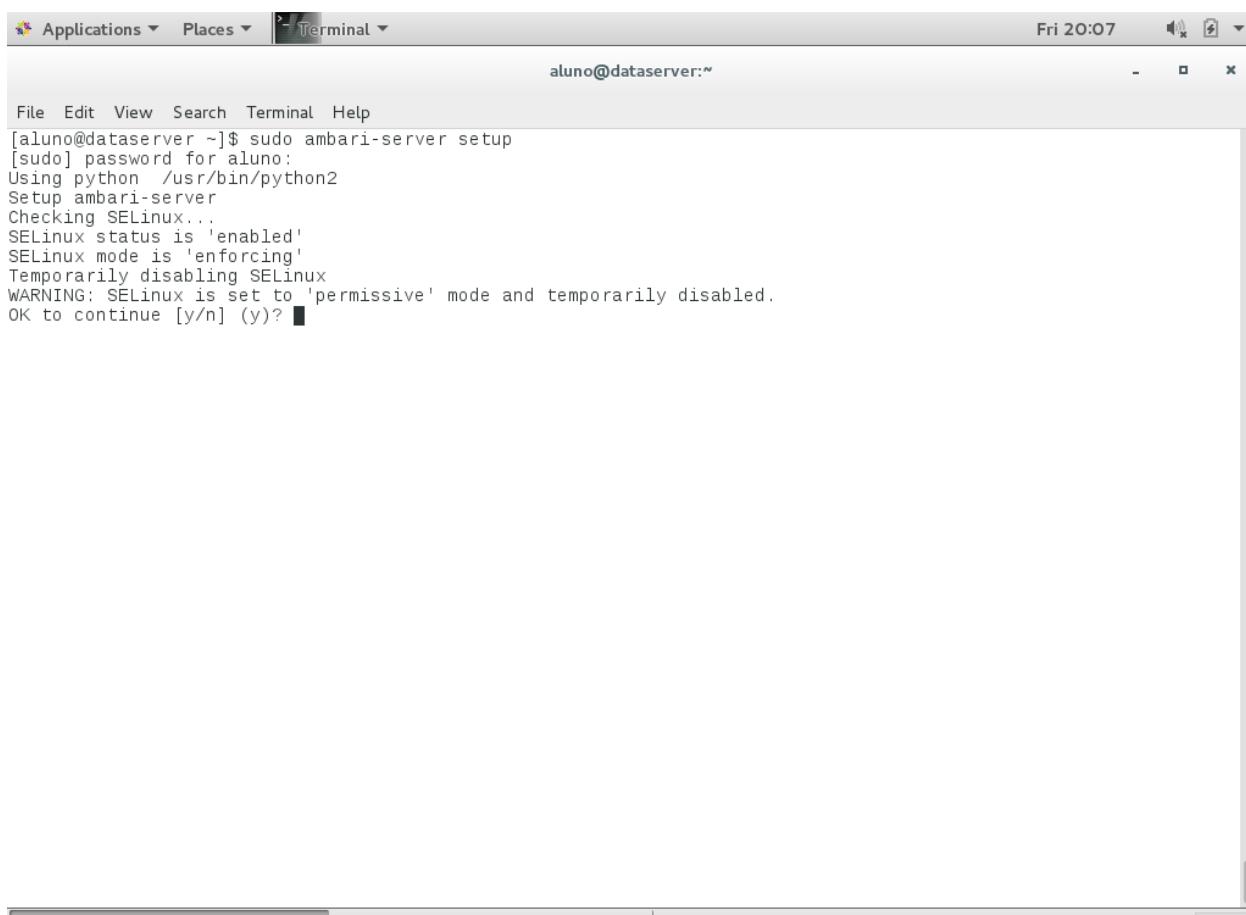
Dependency Installed:
  postgresql.x86_64 0:9.2.15-1.el7_2                                     postgresql-libs.x86_64 0:9.2.15-1.el7_2
  postgresql-server.x86_64 0:9.2.15-1.el7_2

Complete!
[root@dataserver yum.repos.d]#

```

Instalação concluída com sucesso

1 / 4



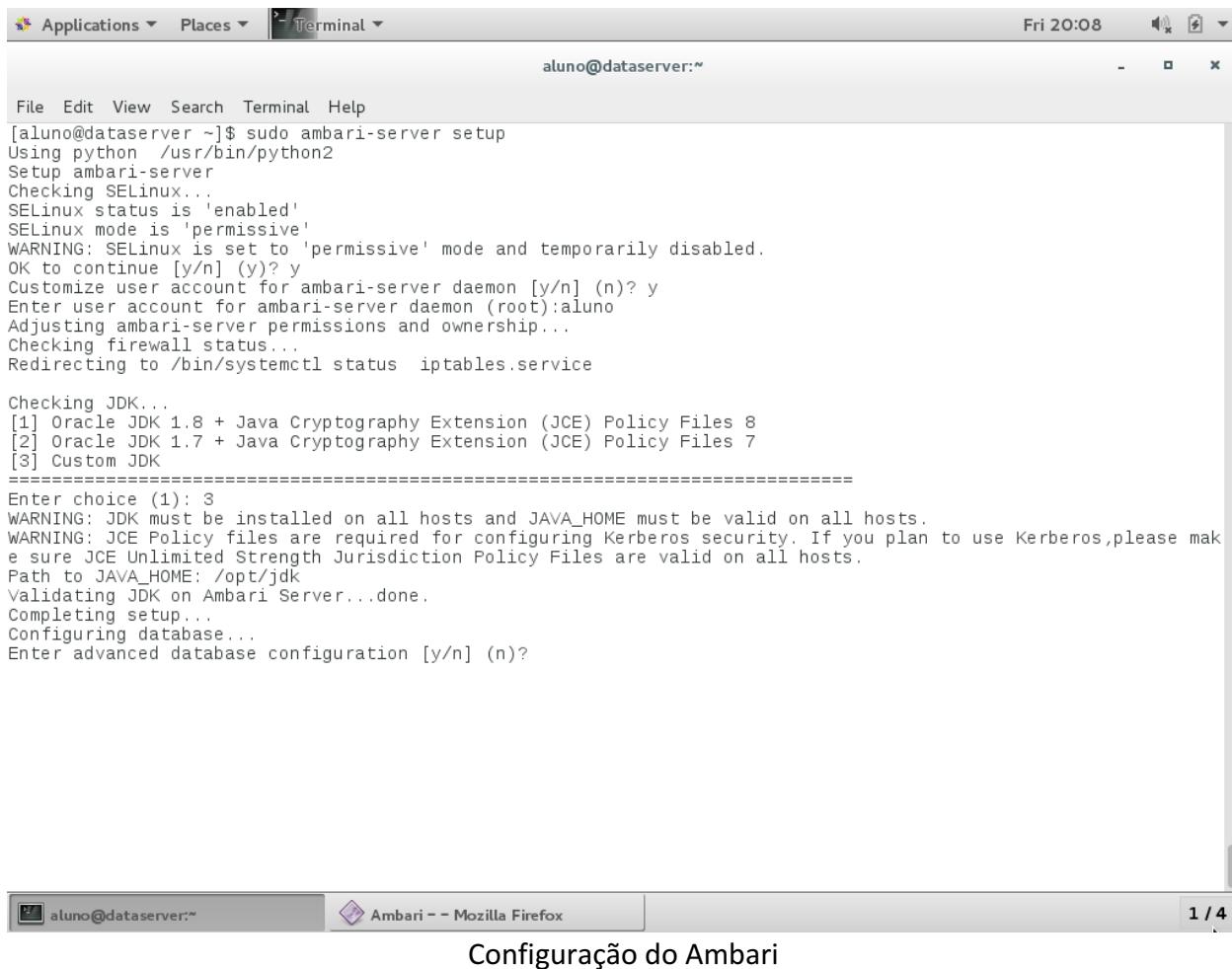
```
[aluno@dataserver ~]$ sudo ambari-server setup
[sudo] password for aluno:
Using python /usr/bin/python2
Setup ambari-server
Checking SELinux...
SELinux status is 'enabled'
SELinux mode is 'enforcing'
Temporarily disabling SELinux
WARNING: SELinux is set to 'permissive' mode and temporarily disabled.
OK to continue [y/n] (y)? █
```

The terminal window title is "Terminal". The user is "aluno" at "dataserver". The command run is "sudo ambari-server setup". It asks for the password "aluno". It uses Python 2. It checks SELinux, noting it is enabled and enforcing. It then temporarily disables SELinux. A warning message states SELinux is set to 'permissive' mode and is temporarily disabled. The user is prompted with "OK to continue [y/n] (y)?".

Below the terminal window, there is a navigation bar with icons for Applications, Places, Terminal, and a search bar. The terminal tab is active. To the right of the terminal window, there is a status bar showing "Fri 20:07". At the bottom of the screen, there is a dock with icons for the terminal, Ambari (Mozilla Firefox), and a file manager. The Ambari icon is highlighted.

Configuração do Ambari

1 / 4



The screenshot shows a terminal window titled 'Terminal' with the user 'aluno@dataserver:~'. The terminal displays the command 'sudo ambari-server setup' and its execution. It handles SELinux settings, customized user accounts, and Java JDK selection. It also asks about advanced database configuration.

```
[aluno@dataserver ~]$ sudo ambari-server setup
Using python /usr/bin/python2
Setup ambari-server
Checking SELinux...
SELinux status is 'enabled'
SELinux mode is 'permissive'
WARNING: SELinux is set to 'permissive' mode and temporarily disabled.
OK to continue [y/n] (y)? y
Customize user account for ambari-server daemon [y/n] (n)? y
Enter user account for ambari-server daemon (root):aluno
Adjusting ambari-server permissions and ownership...
Checking firewall status...
Redirecting to /bin/systemctl status iptables.service

Checking JDK...
[1] Oracle JDK 1.8 + Java Cryptography Extension (JCE) Policy Files 8
[2] Oracle JDK 1.7 + Java Cryptography Extension (JCE) Policy Files 7
[3] Custom JDK
=====
Enter choice (1): 3
WARNING: JDK must be installed on all hosts and JAVA_HOME must be valid on all hosts.
WARNING: JCE Policy files are required for configuring Kerberos security. If you plan to use Kerberos, please make sure JCE Unlimited Strength Jurisdiction Policy Files are valid on all hosts.
Path to JAVA_HOME: /opt/jdk
Validating JDK on Ambari Server...done.
Completing setup...
Configuring database...
Enter advanced database configuration [y/n] (n)?
```

Configuração do Ambari

```

[aluno@dataserver ~]$ sudo ambari-server setup
Using python /usr/bin/python2
Setup ambari-server
Checking SELinux...
SELinux status is 'enabled'
SELinux mode is 'permissive'
WARNING: SELinux is set to 'permissive' mode and temporarily disabled.
OK to continue [y/n] (y)? y
Customize user account for ambari-server daemon [y/n] (n)? y
Enter user account for ambari-server daemon (root):aluno
Adjusting ambari-server permissions and ownership...
Checking firewall status...
Redirecting to /bin/systemctl status iptables.service

Checking JDK...
[1] Oracle JDK 1.8 + Java Cryptography Extension (JCE) Policy Files 8
[2] Oracle JDK 1.7 + Java Cryptography Extension (JCE) Policy Files 7
[3] Custom JDK
=====
Enter choice (1): 3
WARNING: JDK must be installed on all hosts and JAVA_HOME must be valid on all hosts.
WARNING: JCE Policy files are required for configuring Kerberos security. If you plan to use Kerberos, please make sure JCE Unlimited Strength Jurisdiction Policy Files are valid on all hosts.
Path to JAVA_HOME: /opt/jdk
Validating JDK on Ambari Server...done.
Completing setup...
Configuring database...
Enter advanced database configuration [y/n] (n)? y
Configuring database...
=====
Choose one of the following options:
[1] - PostgreSQL (Embedded)
[2] - Oracle
[3] - MySQL
[4] - PostgreSQL
[5] - Microsoft SQL Server (Tech Preview)
[6] - SQL Anywhere
=====
Enter choice (1): 1

```

The terminal window is titled "Terminal" and shows the command-line interface for setting up the Ambari-server. It includes prompts for SELinux, Java permissions, user accounts, firewalls, and JDK validation. It then moves on to database configuration, specifically choosing PostgreSQL as the embedded database.

```
[1] Oracle JDK 1.8 + Java Cryptography Extension (JCE) Policy Files 8
[2] Oracle JDK 1.7 + Java Cryptography Extension (JCE) Policy Files 7
[3] Custom JDK
=====
Enter choice (1): 3
WARNING: JDK must be installed on all hosts and JAVA_HOME must be valid on all hosts.
WARNING: JCE Policy files are required for configuring Kerberos security. If you plan to use Kerberos, please make sure JCE Unlimited Strength Jurisdiction Policy Files are valid on all hosts.
Path to JAVA_HOME: /opt/jdk
Validating JDK on Ambari Server...done.
Completing setup...
Configuring database...
Enter advanced database configuration [y/n] (n)? y
Configuring database...
=====
Choose one of the following options:
[1] - PostgreSQL (Embedded)
[2] - Oracle
[3] - MySQL
[4] - PostgreSQL
[5] - Microsoft SQL Server (Tech Preview)
[6] - SQL Anywhere
=====
Enter choice (1): 1
Database name (ambari): ambari
Postgres schema (ambari): ambari
Username (ambari): ambari
Enter Database Password (bigdata):
Re-enter password:
Default properties detected. Using built-in database.
Configuring ambari database...
Checking PostgreSQL...
Running initdb: This may take upto a minute.
Initializing database ... OK

About to start PostgreSQL
Configuring local database...

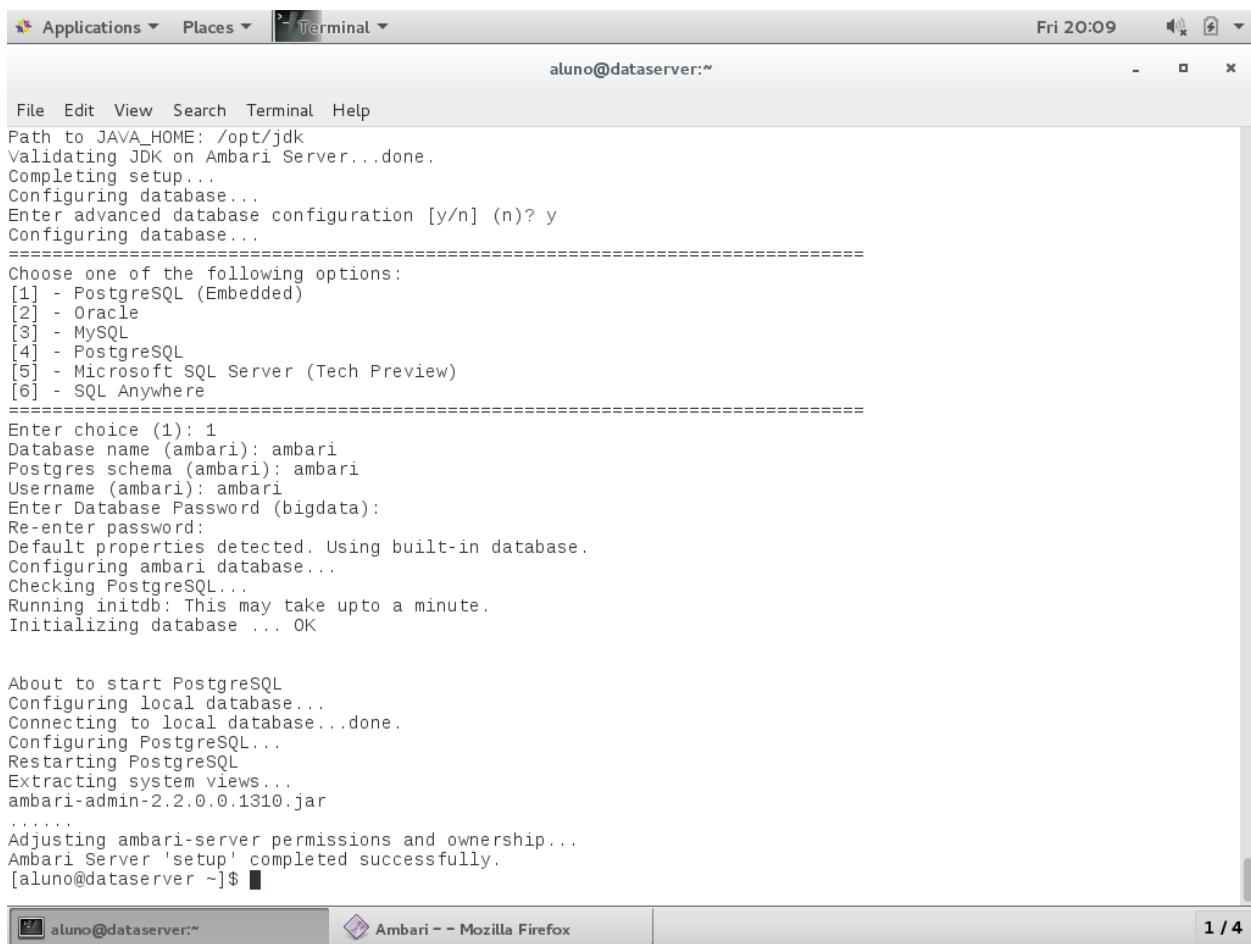
```

aluno@dataserver:~

Ambari -- Mozilla Firefox

1 / 4

Configuração em andamento

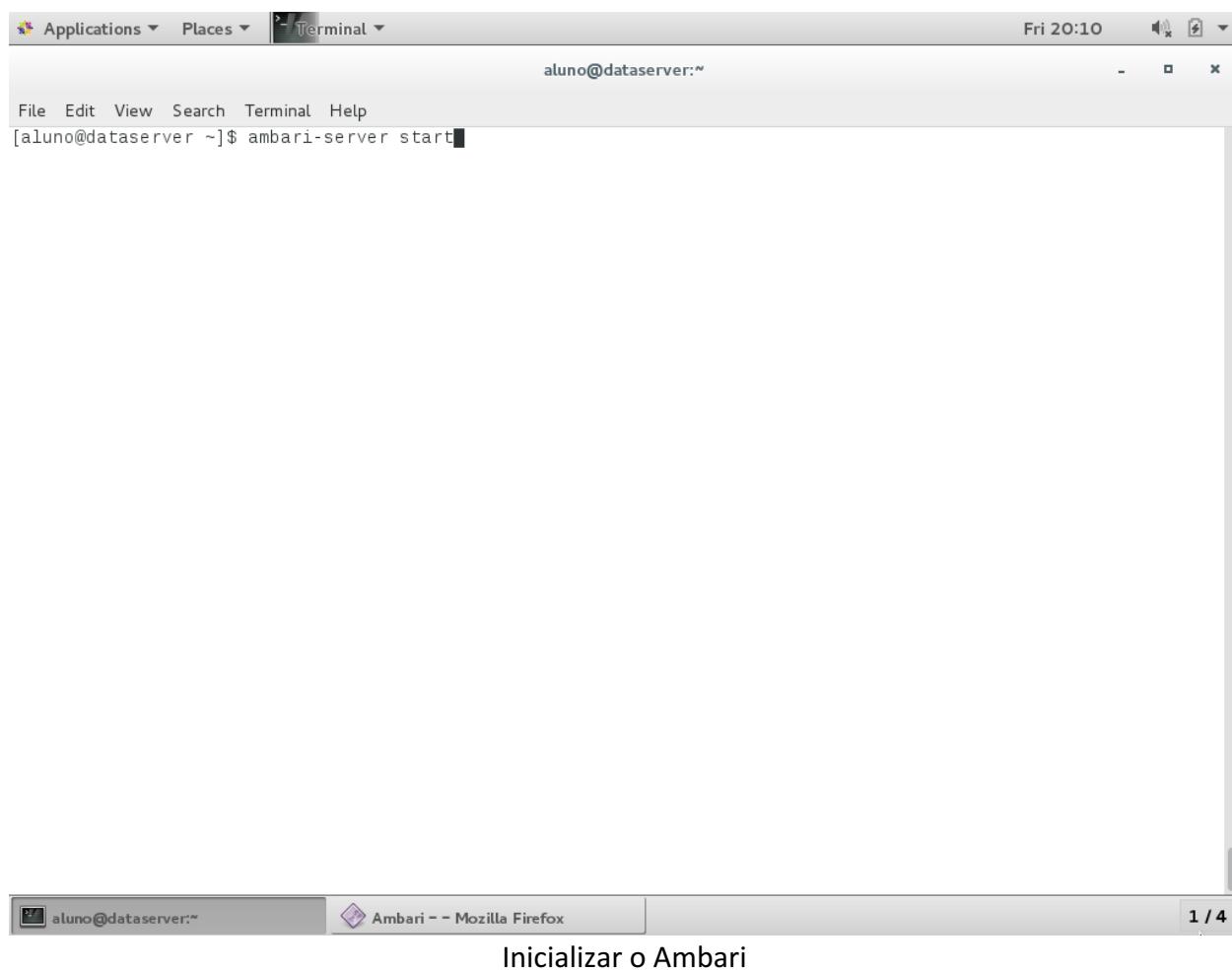


The screenshot shows a terminal window titled 'Terminal' with the command prompt 'aluno@dataserver:~'. The window displays the configuration process for the Ambari database. It starts by validating the JDK and completing setup. It then asks if advanced database configuration is desired, which is confirmed with 'y'. It then lists options for choosing a database system, with option 1 selected (PostgreSQL). It prompts for the database name ('ambari'), schema ('ambari'), and username ('ambari'). It then asks for the database password ('bigdata') and re-enters it. It detects default properties and configures the ambari database. It checks PostgreSQL and initializes the database, noting it may take a minute. It then starts PostgreSQL, extracts system views, and initializes the ambari-admin jar. Finally, it adjusts permissions and ownership for the ambari-server. The process concludes successfully.

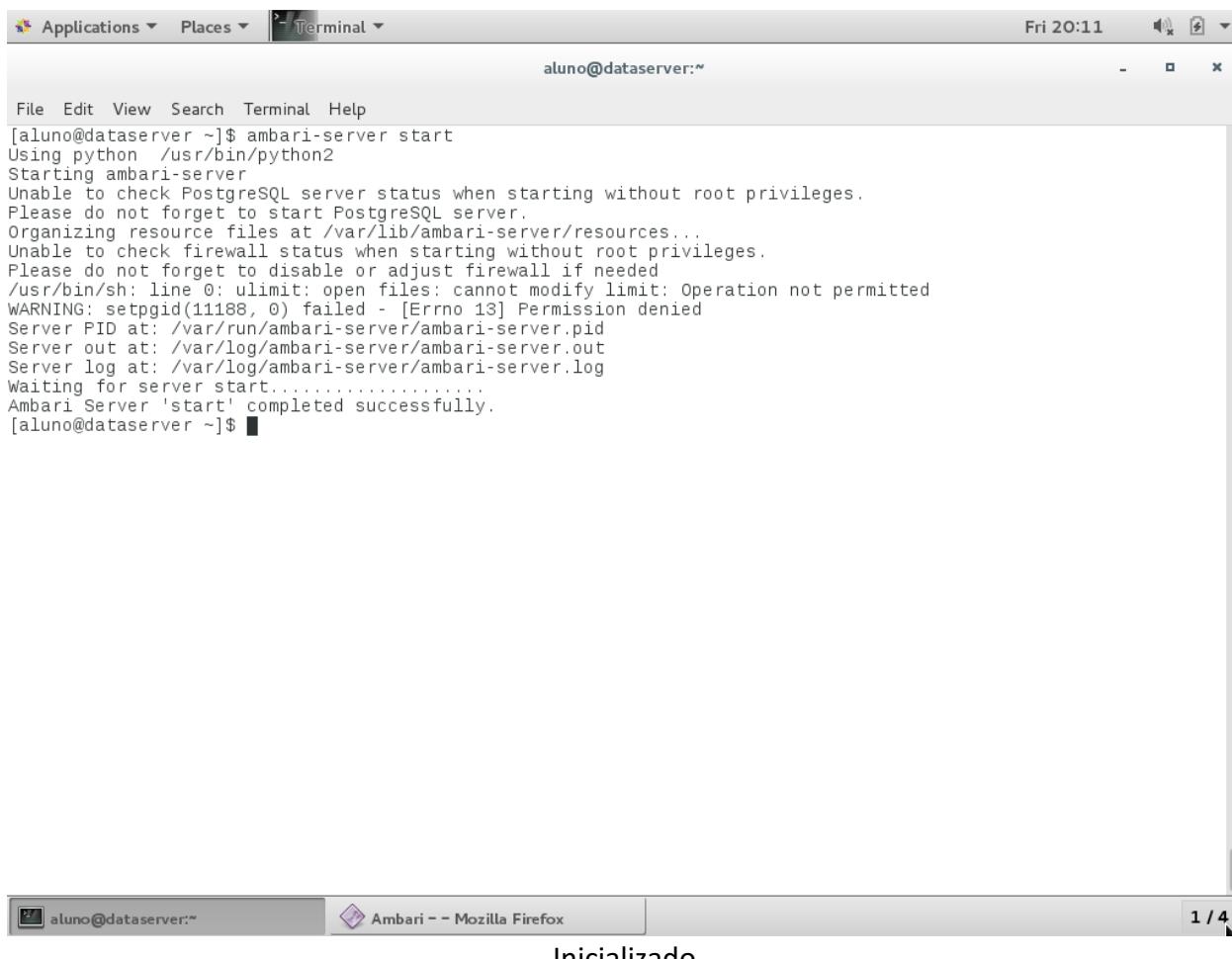
```
Path to JAVA_HOME: /opt/jdk
Validating JDK on Ambari Server...done.
Completing setup...
Configuring database...
Enter advanced database configuration [y/n] (n)? y
Configuring database...
=====
Choose one of the following options:
[1] - PostgreSQL (Embedded)
[2] - Oracle
[3] - MySQL
[4] - PostgreSQL
[5] - Microsoft SQL Server (Tech Preview)
[6] - SQL Anywhere
=====
Enter choice (1): 1
Database name (ambari): ambari
Postgres schema (ambari): ambari
Username (ambari): ambari
Enter Database Password (bigdata):
Re-enter password:
Default properties detected. Using built-in database.
Configuring ambari database...
Checking PostgreSQL...
Running initdb: This may take upto a minute.
Initializing database ... OK

About to start PostgreSQL
Configuring local database...
Connecting to local database...done.
Configuring PostgreSQL...
Restarting PostgreSQL
Extracting system views...
ambari-admin-2.2.0.0.1310.jar
...
Adjusting ambari-server permissions and ownership...
Ambari Server 'setup' completed successfully.
[aluno@dataserver ~]$
```

Configuração concluída



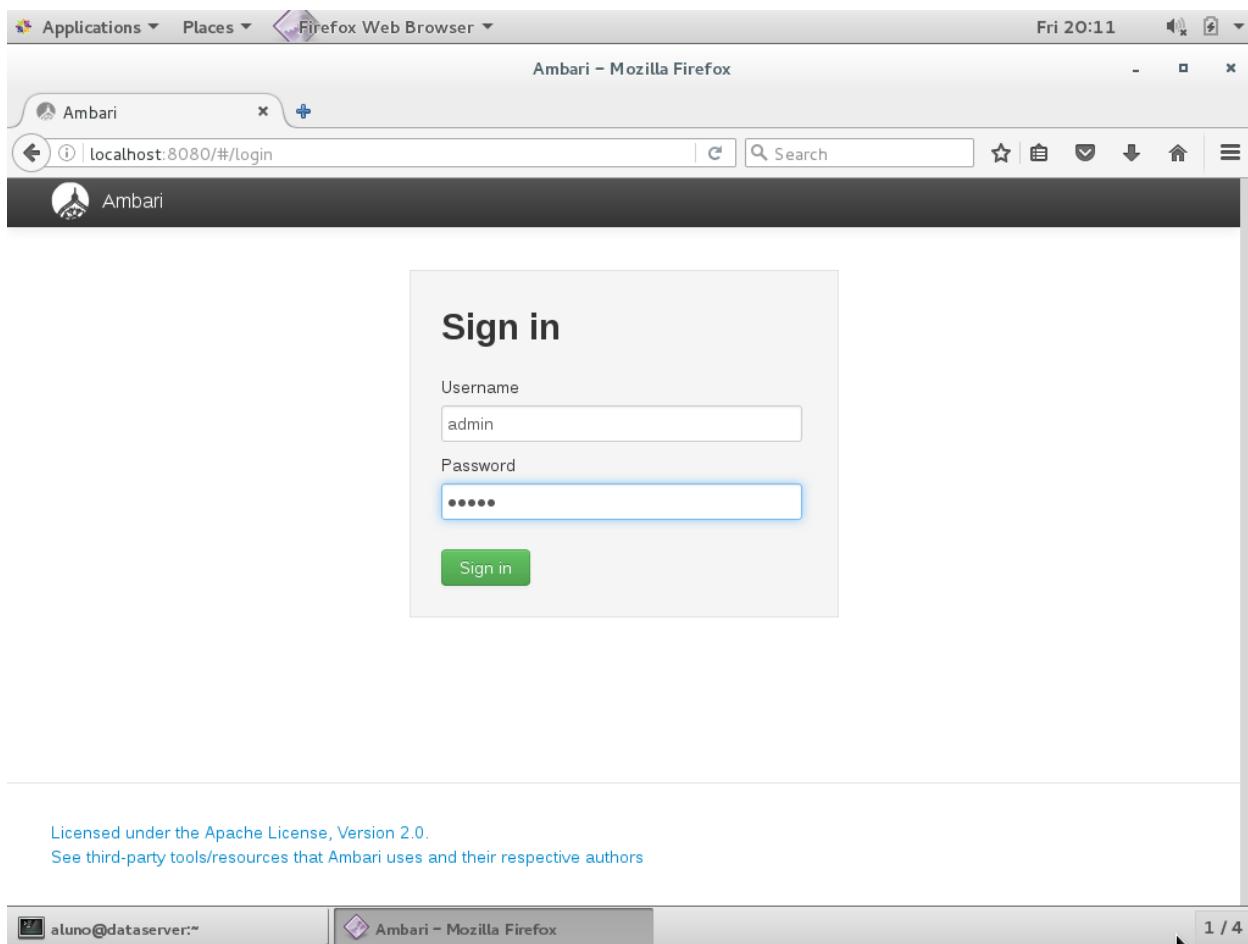
Iniciar o Ambari



A screenshot of a terminal window titled "Terminal". The window shows the command `ambari-server start` being run by the user `aluno`. The output indicates several errors related to PostgreSQL server status, firewall permissions, and ulimit settings. It also shows the creation of a PID file at `/var/run/ambari-server/ambari-server.pid`, an output log at `/var/log/ambari-server/ambari-server.out`, and a log file at `/var/log/ambari-server/ambari-server.log`. The process ends with a message indicating the server has started and the Ambari Server has completed successfully.

```
[aluno@dataserver ~]$ ambari-server start
Using python /usr/bin/python2
Starting ambari-server
Unable to check PostgreSQL server status when starting without root privileges.
Please do not forget to start PostgreSQL server.
Organizing resource files at /var/lib/ambari-server/resources...
Unable to check firewall status when starting without root privileges.
Please do not forget to disable or adjust firewall if needed
/usr/bin/sh: line 0: ulimit: open files: cannot modify limit: Operation not permitted
WARNING: setpgid(11188, 0) failed - [Errno 13] Permission denied
Server PID at: /var/run/ambari-server/ambari-server.pid
Server out at: /var/log/ambari-server/ambari-server.out
Server log at: /var/log/ambari-server/ambari-server.log
Waiting for server start.....
Ambari Server 'start' completed successfully.
[aluno@dataserver ~]$
```

Inicializado



Acessar o browser – <http://dataserver:8080> - usuário: admin / senha: admin

The screenshot shows the Apache Ambari web interface running in Mozilla Firefox. The title bar reads "Ambari - Mozilla Firefox". The left sidebar has three main sections: "Clusters" (No clusters), "Views" (Views), and "User + Group Management" (Users, Groups). The main content area is titled "Welcome to Apache Ambari" with the sub-instruction "Provision a cluster, manage who can access the cluster, and customize views for Ambari users.". It features a "Create a Cluster" section with a "Launch Install Wizard" button, a "Manage Users + Groups" section, and a "Deploy Views" section. The bottom status bar shows "aluno@dataserver:~" and "Ambari - Mozilla Firefox". A page navigation bar at the bottom indicates "1 / 4".

Pronto para configuração do cluster

Quarto checkpoint:

Clique no meu File – Export Appliance.
Será gerada uma cópia de segurança da sua máquina virtual.

→ VM: DataServer-vFinal.ova (Completa)

Download disponível em :

<http://datascienceacademy.com.br/blog/aluno/EngenhariaHadoopSpark/VMs>

Parabéns!

Você tem um ambiente de testes para
armazenar e processar Big Data!