



Odisee
DE CO-HOGESCHOOL

Big Data – VM



Jens Baetens

Installatie

▣ Install virtual box

- <https://www.virtualbox.org/wiki/Downloads>

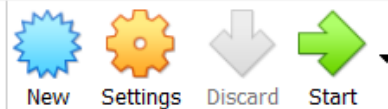
▣ Download ubuntu image via teams

- https://hubkaho.sharepoint.com/:u:/r/sites/Section_OBI64a-2122/Lesmateriaal/Big%20Data%20Students.ova?csf=1&web=1&e=gSamRL
- Deze image heeft reeds een aantal geïnstalleerde programma's
 - ▣ Hadoop, spark, notebooks, mongodb, ...
- Administrator met naam bigdata en paswoord bigdata



File Machine Help

- Preferences... Ctrl+G
- Import Appliance... Ctrl+I
- Export Appliance... Ctrl+E
- New Cloud VM...
- Virtual Media Manager... Ctrl+D
- Host Network Manager... Ctrl+H
- Network Operations Manager...
- Check for Updates...
- Reset All Warnings
- Exit Ctrl+Q



General

Name: Big Data
Operating System: Ubuntu (64-bit)

System

Base Memory: 8192 MB
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, KVM Paravirtualization

Display

Video Memory: 16 MB
Graphics Controller: VMSVGA
Remote Desktop Server: Disabled
Recording: Disabled

Storage

Controller: IDE
IDE Secondary Device 0: [Optical Drive] Empty
Controller: SATA
SATA Port 0: Big Data.vdi (Normal, 30.00 GB)

Audio

Host Driver: Windows DirectSound
Controller: ICH AC97

Network

Adapter 1: Intel PRO/1000 MT Desktop (NAT)

USB

USB Controller: OHCI
Device Filters: 0 (0 active)

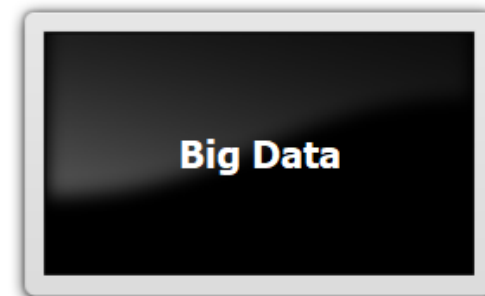
Shared folders

None

Description

None

Preview





Tools



Big Data Insights

Saved



Big Data Insights

Powered



Big Data Insights

Powered

← Import Virtual Appliance

Appliance to import

Please choose the source to import appliance from. This can be a local file system to import OVF archive or one of known cloud service providers to import cloud VM from.

Source: Local File System

Please choose a file to import the virtual appliance from. VirtualBox currently supports importing appliances saved in the Open Virtualization Format (OVF). To continue, select the file to import below.

File: C:\Users\jens.baetens3\Documents\Big Data.ova



Klik hier en selecteer de file

Appliance settings

These are the virtual machines contained in the appliance and the suggested settings of the imported VirtualBox machines. You can change many of the properties shown by double-clicking on the items and disable others using the check boxes below.

Virtual System 1	
 Name	Big Data 1
 Guest OS Type	Ubuntu (64-bit)
 CPU	1
 RAM	8192 MB
 DVD	<input checked="" type="checkbox"/>
 USB Controller	<input checked="" type="checkbox"/>
 Sound Card	<input checked="" type="checkbox"/> ICH AC97
 Network Adapter	<input checked="" type="checkbox"/> Intel PRO/1000 MT Desktop (82540EM)
 Storage Controller (IDE)	PIIX4
 Storage Controller (IDE)	PIIX4
 Storage Controller (SATA)	AHCI
 Virtual Disk Image	Big Data-disk001.vmdk
 Base Folder	C:\Users\jens.baetens3\VirtualBox VMs
 Primary Group	/

Kies hier een naam

Klik hier om het importeren te starten

Machine Base Folder: C:\Users\jens.baetens3\VirtualBox VMs

MAC Address Policy: Include only NAT network adapter MAC addresses

Additional Options: ☒ Import hard drives as VDI

Appliance is not signed

Restore Defaults

Import

Cancel



Tools

**Big Data**

Powered Off

**Big Data For Students**

Powered Off

**Big Data Student Test**

Powered Off

Selecteer de geïmporteerde VM



New



Settings



Discard



Start



Start de VM

**General**

Name: Big Data
Operating System: Ubuntu (64-bit)

**System**

Base Memory: 8192 MB
Boot Order: Floppy, Optical, Hard Disk
Acceleration: VT-x/AMD-V, Nested Paging, KVM Paravirtualization

**Display**

Video Memory: 16 MB
Graphics Controller: VMSVGA
Remote Desktop Server: Disabled
Recording: Disabled

**Storage**

Controller: IDE
IDE Secondary Device 0: [Optical Drive] Empty
Controller: SATA
SATA Port 0: Big Data.vdi (Normal, 30.00 GB)

**Audio**

Host Driver: Windows DirectSound
Controller: ICH AC97

**Network**

Adapter 1: Intel PRO/1000 MT Desktop (NAT)

**USB**

USB Controller: OHCI
Device Filters: 0 (0 active)

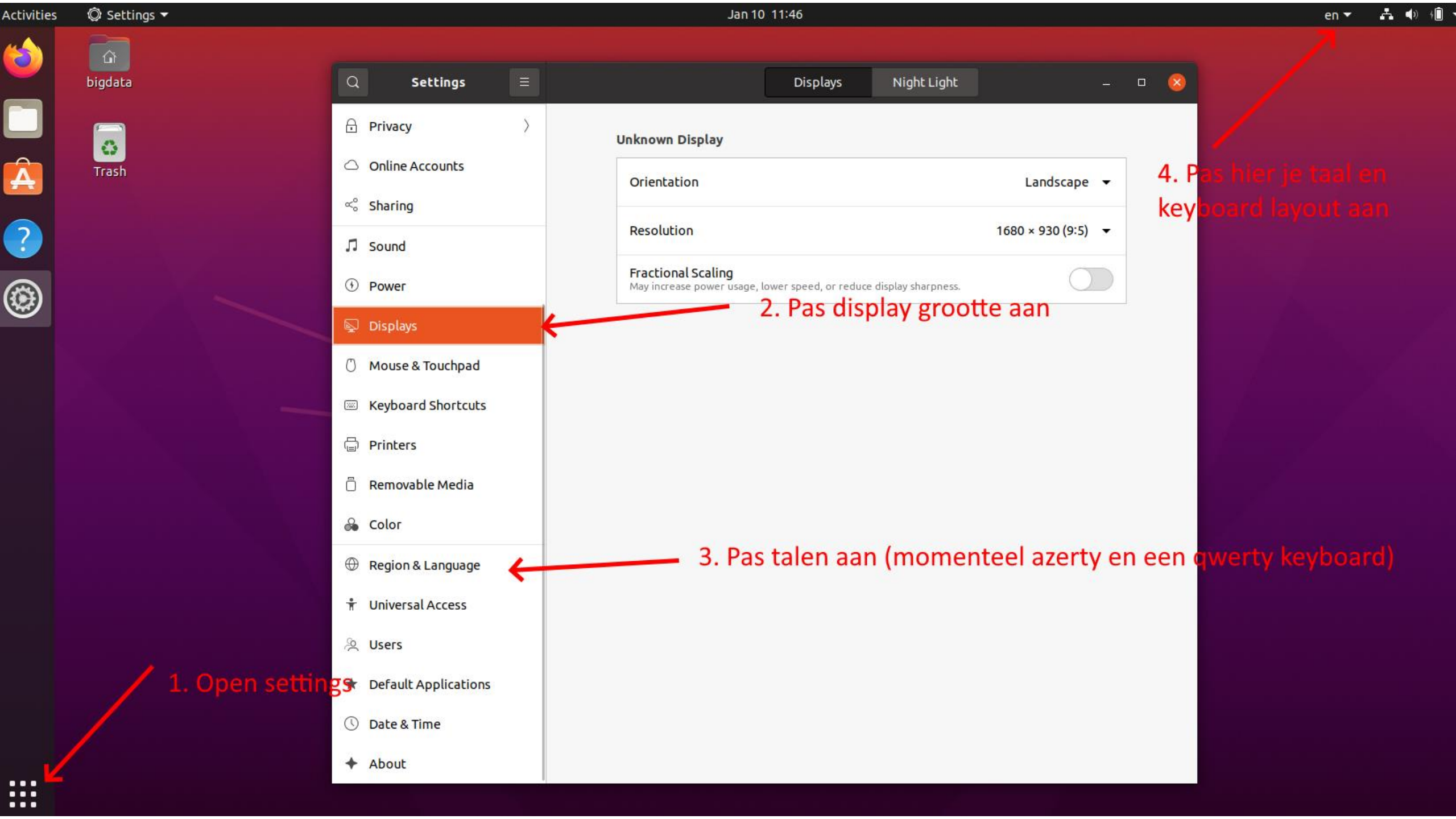
**Shared folders**

None

**Description**

None

**Preview**



Settings

DisplaysNight Light

Privacy

Online Accounts

Sharing

Sound

Power

Displays

Mouse & Touchpad

Keyboard Shortcuts

Printers

Removable Media

Color

Region & Language

Universal Access

Users

Default Applications

Date & Time

About

Unknown Display

Orientation

Landscape

Resolution

1680 × 930 (9:5)

Fractional Scaling

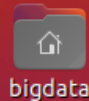
May increase power usage, lower speed, or reduce display sharpness.

1. Open settings

2. Pas display grootte aan

3. Pas talen aan (momenteel azerty en een qwerty keyboard)

4. Pas hier je taal en keyboard layout aan



bigdata



Trash



bigdata@bigdata-VirtualBox: ~



```
(base) bigdata@bigdata-VirtualBox:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/bigdata/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/bigdata/.ssh/id_rsa
Your public key has been saved in /home/bigdata/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:N4KUJtK6FiUhCNK6vQL1q2aFfCKOm4fK68g0ZEDDI9U bigdata@bigdata-VirtualBox
The key's randomart image is:
+---[RSA 3072]-----+
|B=o.                  |
|+=+ooE .             |
|o.+ + +              |
|o . = + .            |
| Boo . S o           |
|* =o+   o .          |
|+=o= .               |
|O+* .                |
|B@o.                 |
+-----[SHA256]-----+
(base) bigdata@bigdata-VirtualBox:~$ cat .ssh/id_rsa.pub >> .ssh/authorized_keys
(base) bigdata@bigdata-VirtualBox:~$ chmod 640 .ssh/authorized_keys
(base) bigdata@bigdata-VirtualBox:~$ sudo service ssh restart
(base) bigdata@bigdata-VirtualBox:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:qJF9AMlgbLr5xqE9L8g1XLu5V/LCXJZVaFexdw3GDo.
Are you sure you want to continue connecting (yes/no/[fingerprint])? y
Please type 'yes', 'no' or the fingerprint: yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-43-generic x86_64)

* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage

0 updates can be applied immediately.

The list of available updates is more than a week old.
To check for new updates run: sudo apt update
```

Maak een nieuwe ssh key aan

Voer deze commando's uit

1. Start terminal (Control + Alt + T)

Opstarten Hadoop cluster

```
bigdata@bigdata-VirtualBox: ~  
(base) bigdata@bigdata-VirtualBox:~$ start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as bigdata in 10 seconds.  
WARNING: This is not a recommended production deployment configuration.  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [bigdata-VirtualBox]  
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.  
Starting resourcemanager  
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.  
Starting nodemanagers  
WARNING: YARN_CONF_DIR has been replaced by HADOOP_CONF_DIR. Using value of YARN_CONF_DIR.  
(base) bigdata@bigdata-VirtualBox:~$  
(base) bigdata@bigdata-VirtualBox:~$  
(base) bigdata@bigdata-VirtualBox:~$ jps  
7841 NodeManager  
7157 NameNode  
7287 DataNode  
8184 Jps  
7706 ResourceManager  
7500 SecondaryNameNode  
(base) bigdata@bigdata-VirtualBox:~$
```

Start de hadoop cluster

Controleer de lopende java processen

Deze moeten actief zijn indien het goed gelukt is

Controller opstarten cluster

Activities Firefox Web Browser Jan 10 12:15 en

Namenode information x +

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

2. Surf naar localhost:9870 voor de Web interface van het HDFS

Overview 'localhost:9000' (✓active)

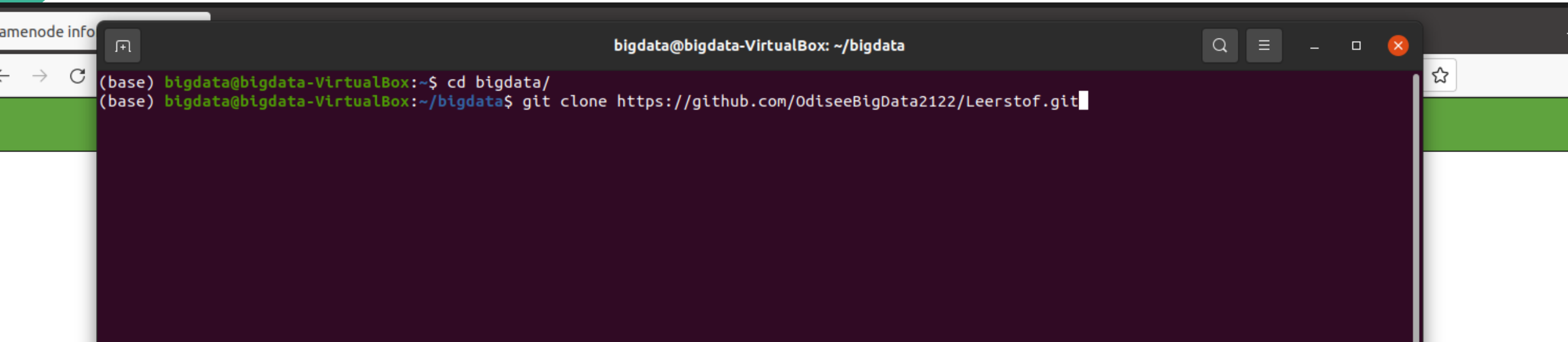
1. Start firefox

Started:	Mon Jan 10 12:07:32 +0100 2022
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Mon Jul 06 20:44:00 +0200 2020 by brahma from branch-3.3.0
Cluster ID:	CID-88b9300a-d090-486e-bab1-55ad0de1de03
Block Pool ID:	BP-1472610732-127.0.1.1-1640348735344

Summary

Security is off.

Clone repository leerstof

A screenshot of a terminal window titled 'bigdata@bigdata-VirtualBox: ~/bigdata'. The terminal shows two commands: first, '(base) bigdata@bigdata-VirtualBox:~\$ cd bigdata/' and second, '(base) bigdata@bigdata-VirtualBox:~/bigdata\$ git clone https://github.com/OdisseeBigData2122/Leerstof.git'. The cursor is at the end of the second command. The terminal has a dark purple background and white text. The window is part of a larger interface with a green header bar and a search icon in the top right corner.

```
bigdata@bigdata-VirtualBox: ~/bigdata
(base) bigdata@bigdata-VirtualBox:~$ cd bigdata/
(base) bigdata@bigdata-VirtualBox:~/bigdata$ git clone https://github.com/OdisseeBigData2122/Leerstof.git
```

- ▣ Ik raad aan om alles in de bigdata folder te plaatsen
- ▣ De publieke repository met de slides/voorbeeldcode/... is:
 - <https://github.com/OdisseeBigData2122/Leerstof.git>