



**Odisee**  
DE CO-HOGESCHOOL

# Big Data – VM

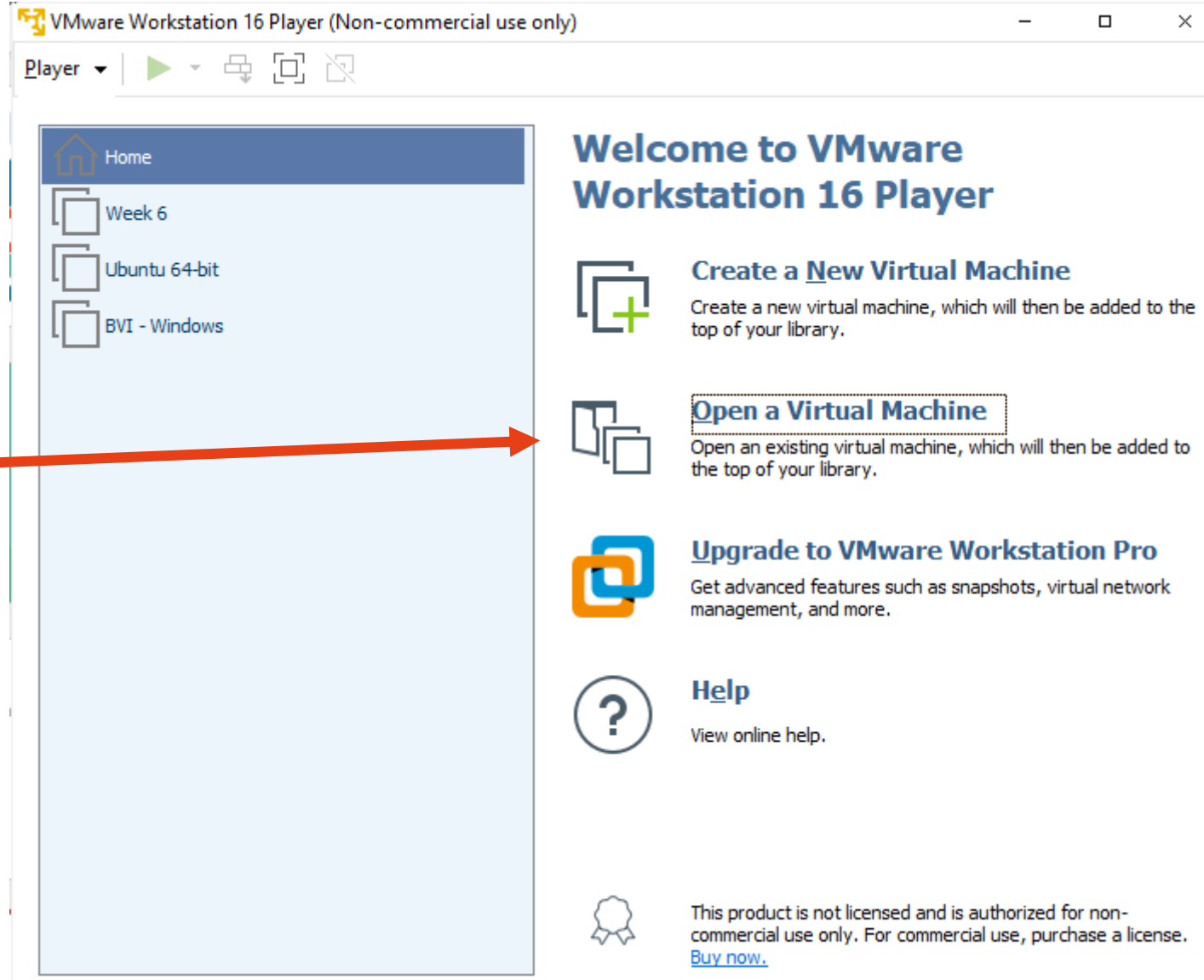


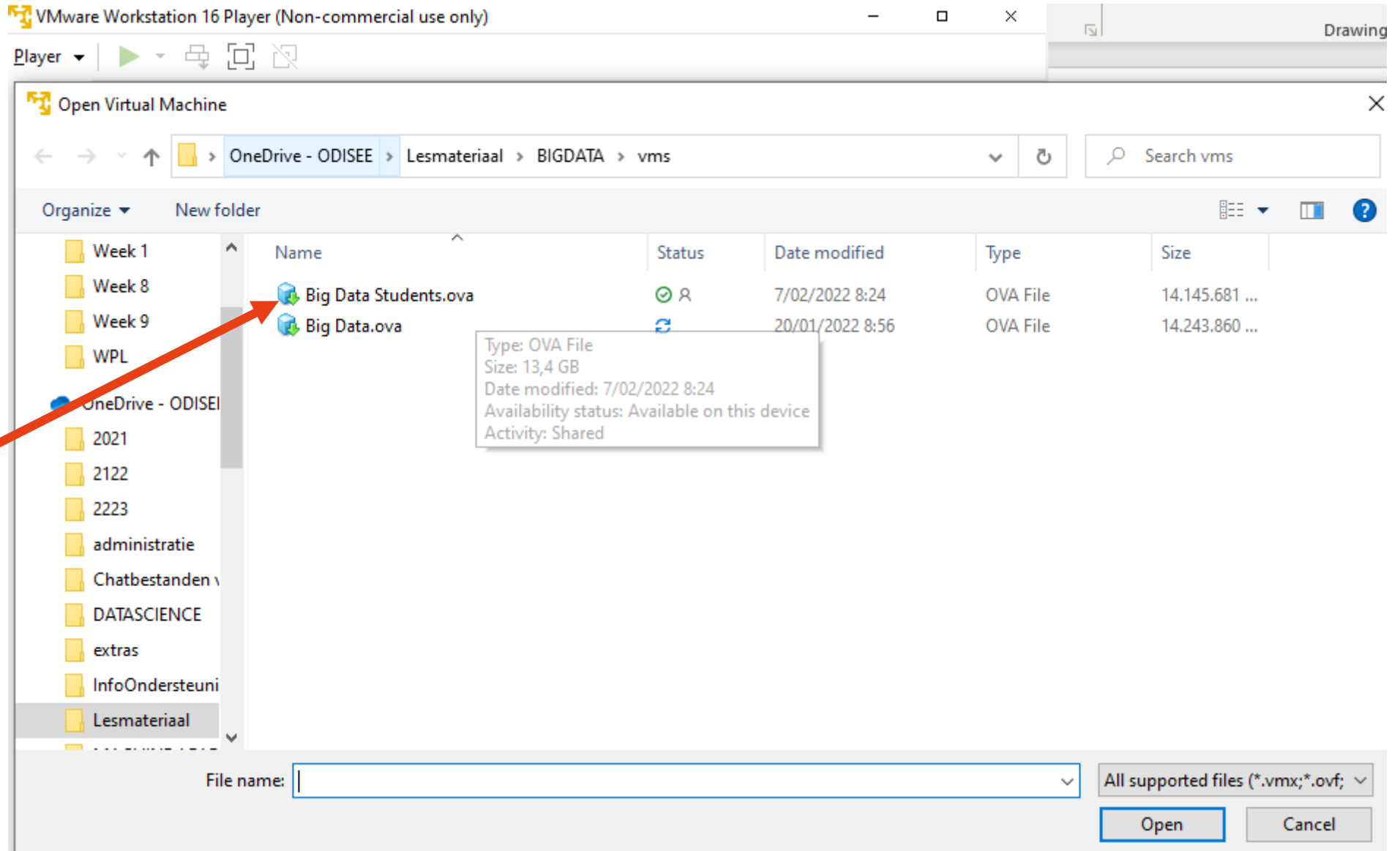
Jens Baetens

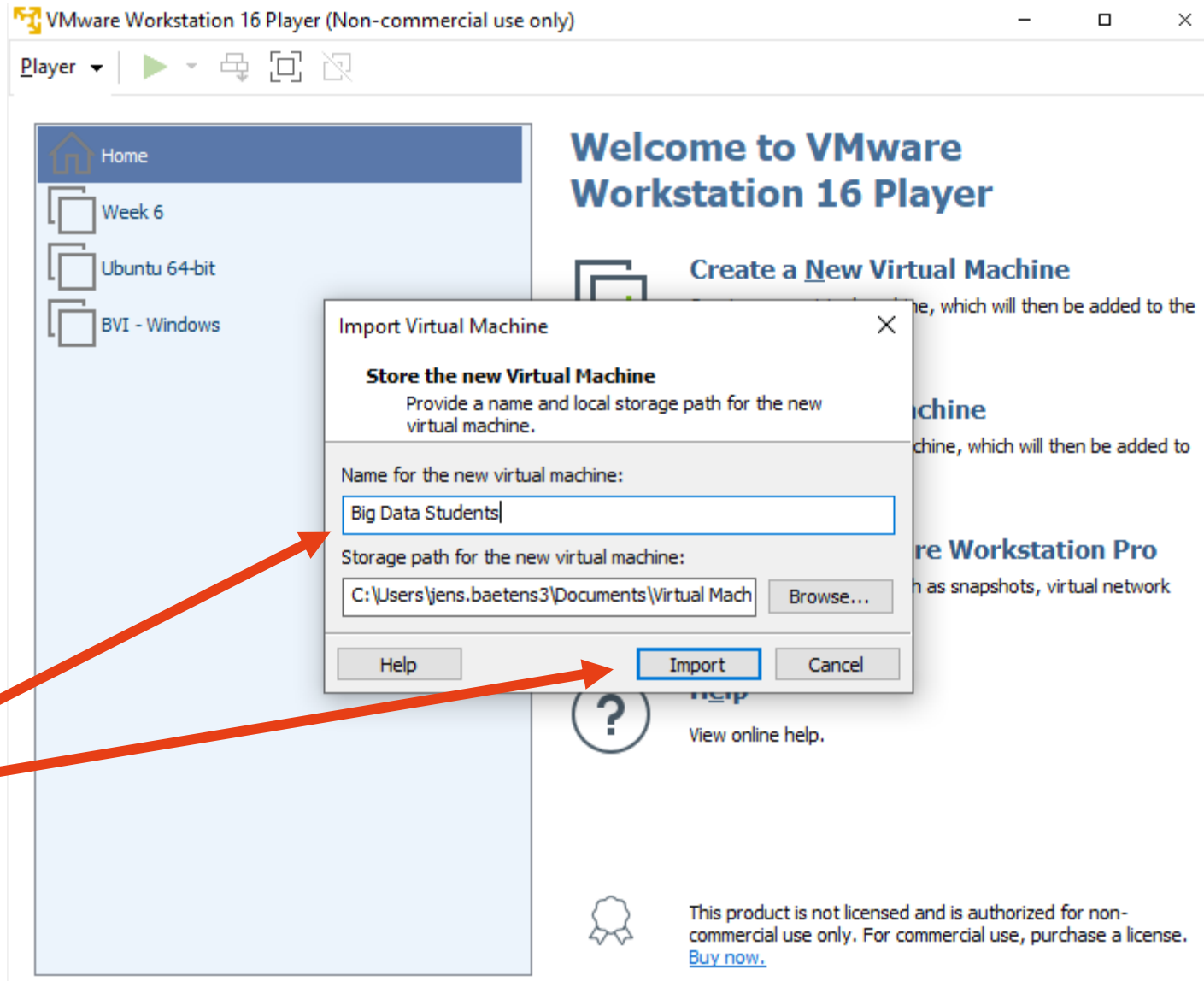
# Installatie

- ▣ Install VMware workstation player (indien nog niet geïnstalleerd)
  - ▬ <https://www.vmware.com/go/getplayer-win>
- ▣ Download ubuntu image via one-drive link
  - ▬ [Big Data Students.ova](#)
  - ▬ Deze image heeft reeds een aantal geïnstalleerde programmas
    - ▣ Hadoop, spark, notebooks, mongodb, ...
  - ▬ Administrator met naam bigdata en paswoord bigdata

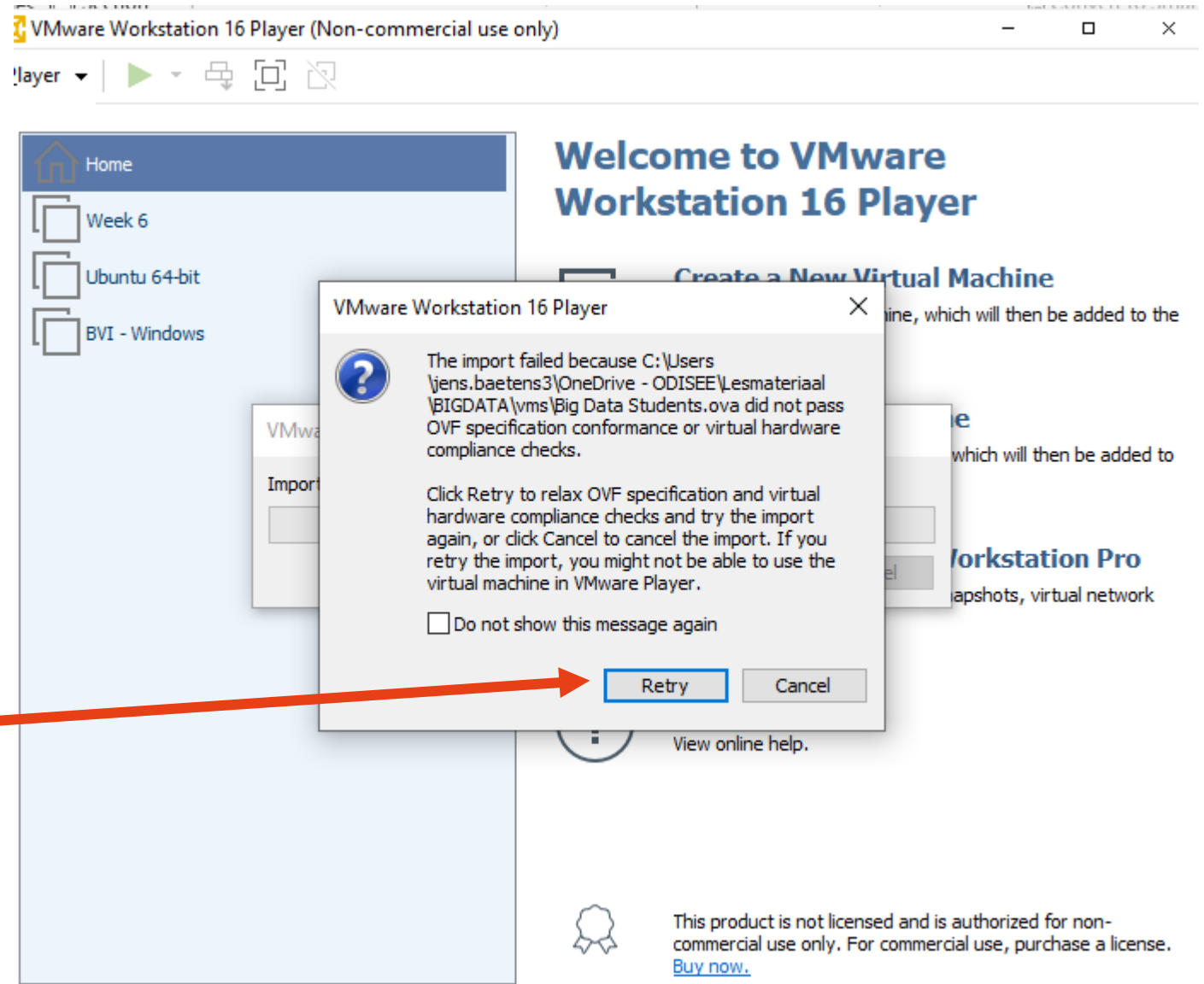






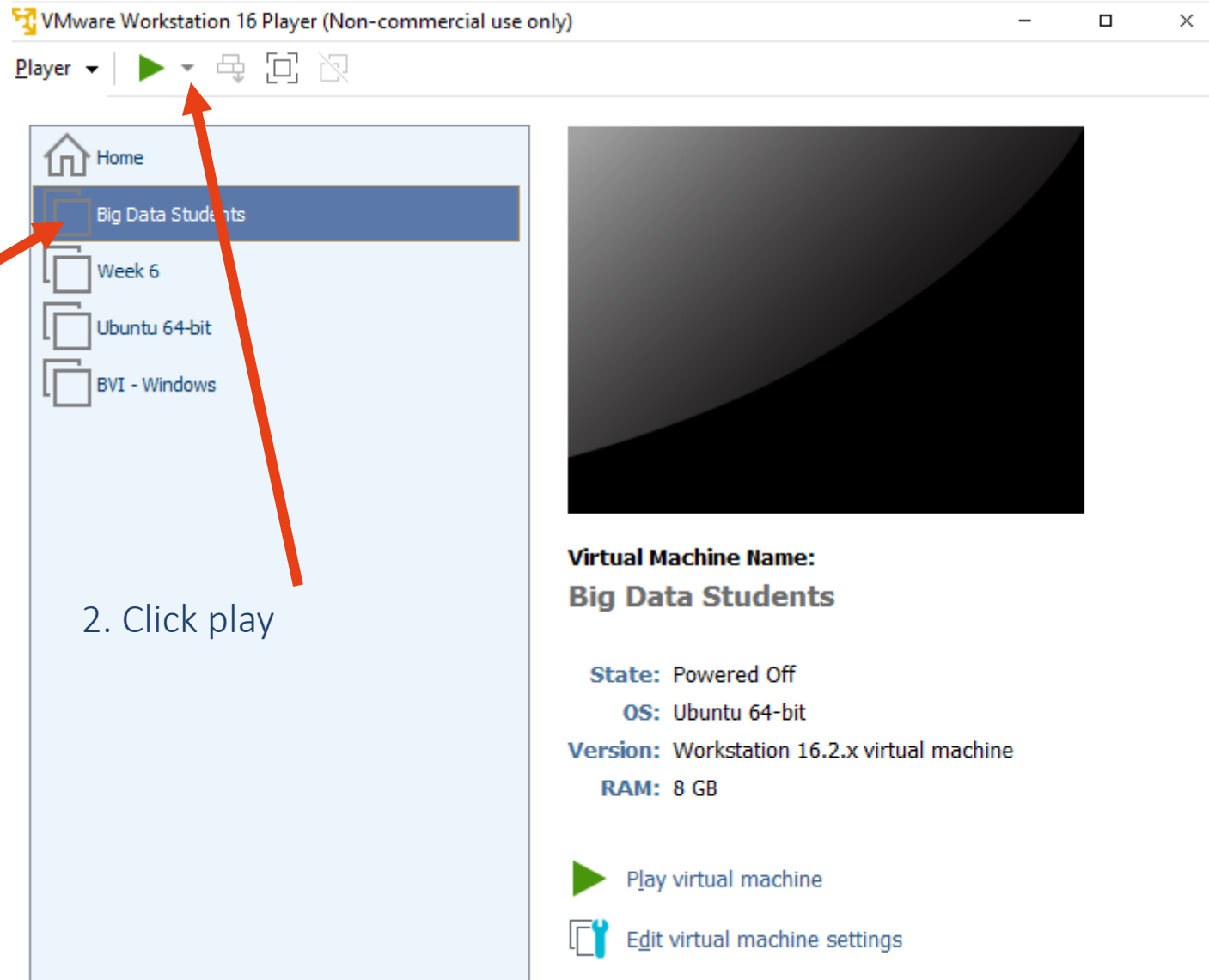


Pick a name  
Click import



Click retry if this pops up

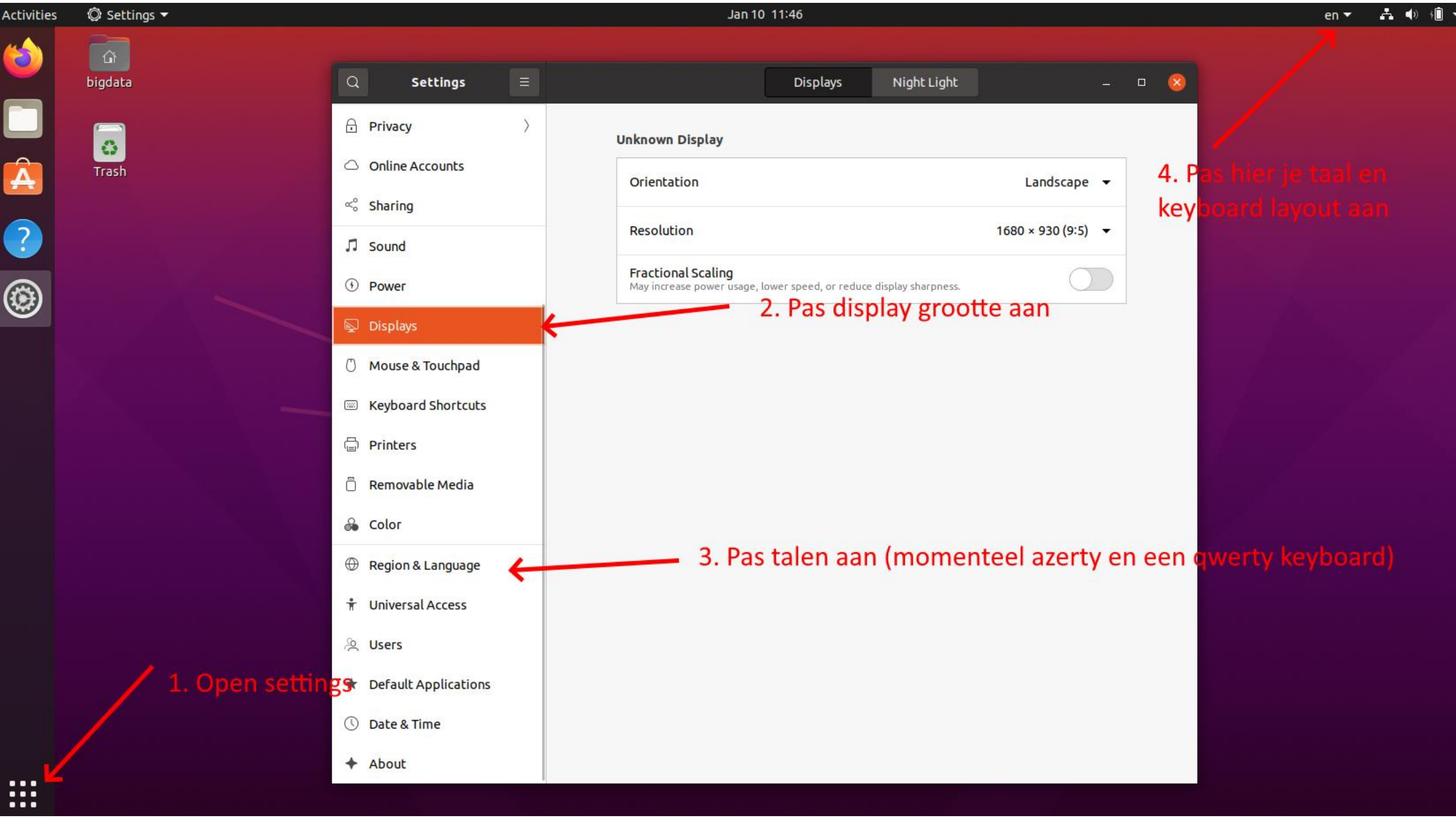


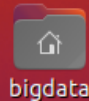


1. Select VM

2. Click play







bigdata



Trash



&gt;\_

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=0.                  |
```

```
|=+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
```

```
1. Start terminal (Control + Alt + T)
```

Maak een nieuwe ssh key aan

Voer deze commando's uit

1

2

3

4

# 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)

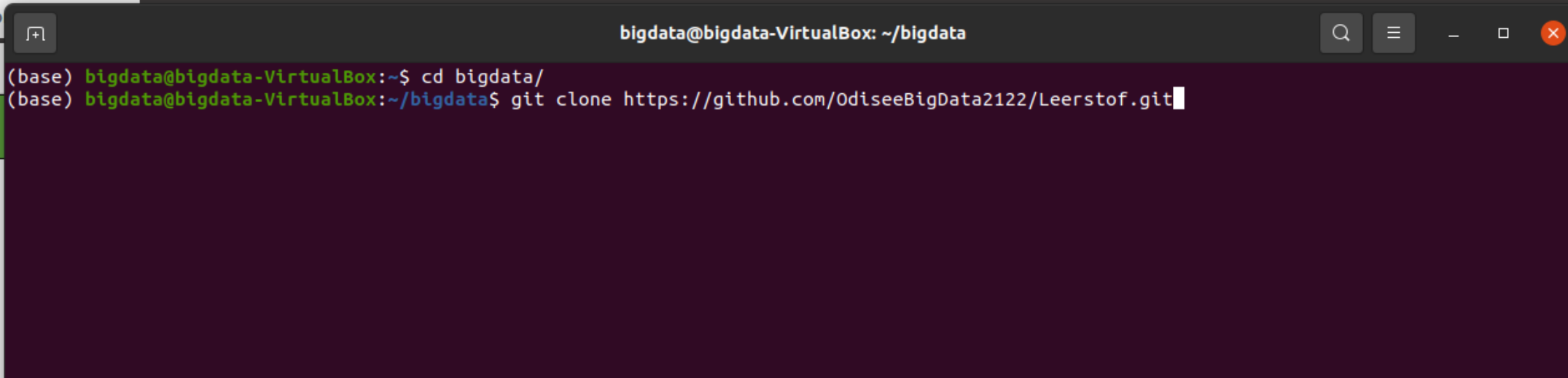
1. Start firefox

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



The screenshot shows a terminal window titled "bigdata@bigdata-VirtualBox: ~/bigdata". The prompt is "(base) bigdata@bigdata-VirtualBox:~\$". The user enters "cd bigdata/" and the prompt changes to "(base) bigdata@bigdata-VirtualBox:~/bigdata\$". The user then enters "git clone https://github.com/OdiseeBigData2122/Leerstof.git".

```
bigdata@bigdata-VirtualBox: ~/bigdata
(base) bigdata@bigdata-VirtualBox:~$ cd bigdata/
(base) bigdata@bigdata-VirtualBox:~/bigdata$ git clone https://github.com/OdiseeBigData2122/Leerstof.git
```

- ▣ Ik raad aan om alles in de bigdata folder te plaatsen
- ▣ De publieke repository met de slides/voorbeeldcode/... is:
  - <https://github.com/OdiseeBigData2122/Leerstof.git>