Step 1

Setup:

- Option 1: CS Laptop
- Option 2: BYO Laptop

Option 1: Using CS Laptop

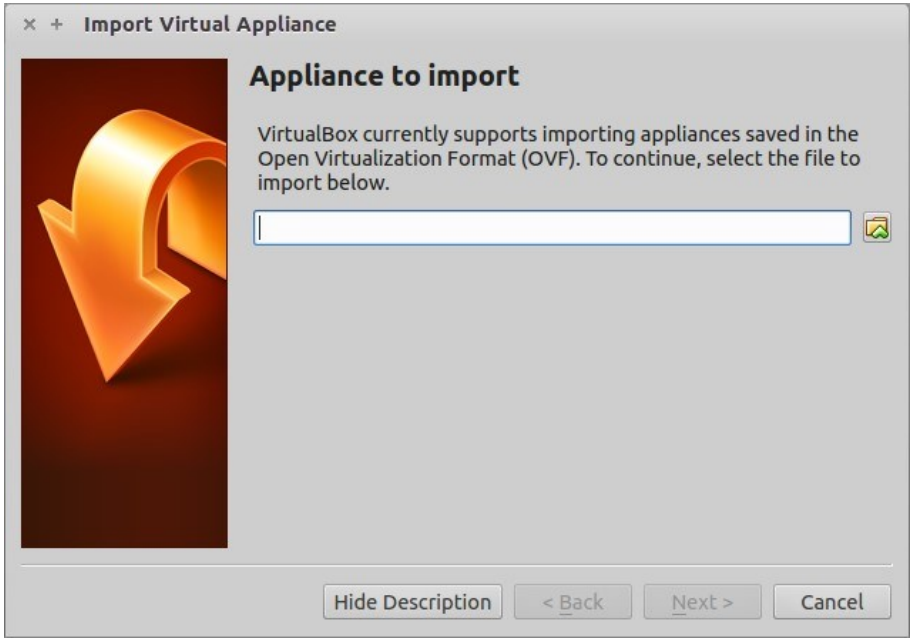
- Open VirtualBox:
  - Click the “VirtualBox 5.0.2 (for CS4480 and CS5488)” item in CSLab Menu
- Skip to Step 2

Note: Make sure that you copy your work to `/home/bitnami/hdrive` before you shutdown the VM

Option 2: BYO Laptop

Import a VM into VirtualBox

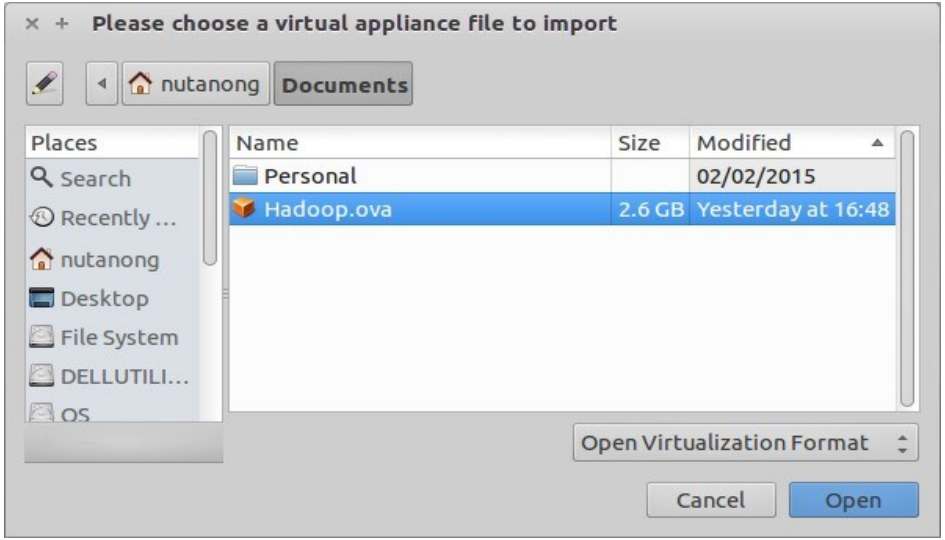
- Download Hadoop.ova from https://drive.google.com/file/d/1r6Rv5hSyD2a9GCylL6h0ljFyhcDXfEus/view?usp=sharing
- Open VirtualBox
- Hit Ctrl+i
- Click the folder logo next to the text box



Option 2: BYO Laptop

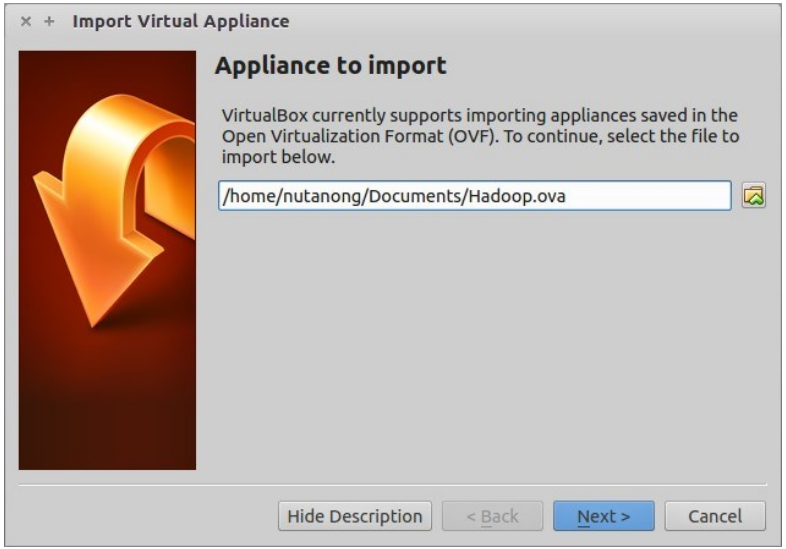
Import a VM into VirtualBox

- Go to the folder containing Hadoop.ova and open the file



Option 2: BYO Laptop

Import a VM into VirtualBox

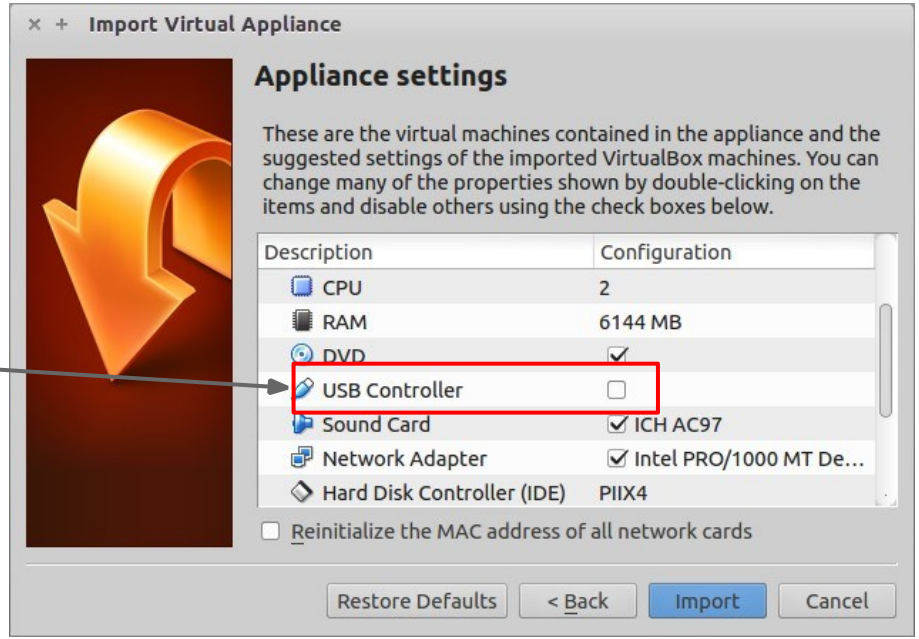


- Click “Next”

Option 2: BYO Laptop

Import a VM into VirtualBox

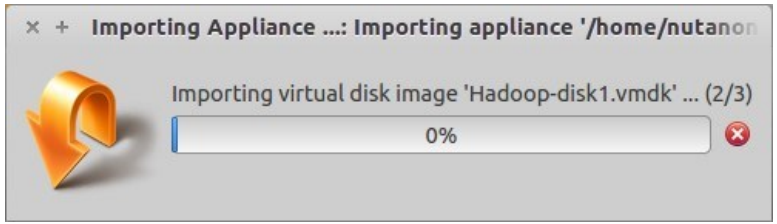
- Make sure that the USB Controller option is ***not*** checked
- Click “Import”



Option 2: BYO Laptop

Import a VM into VirtualBox

- Wait



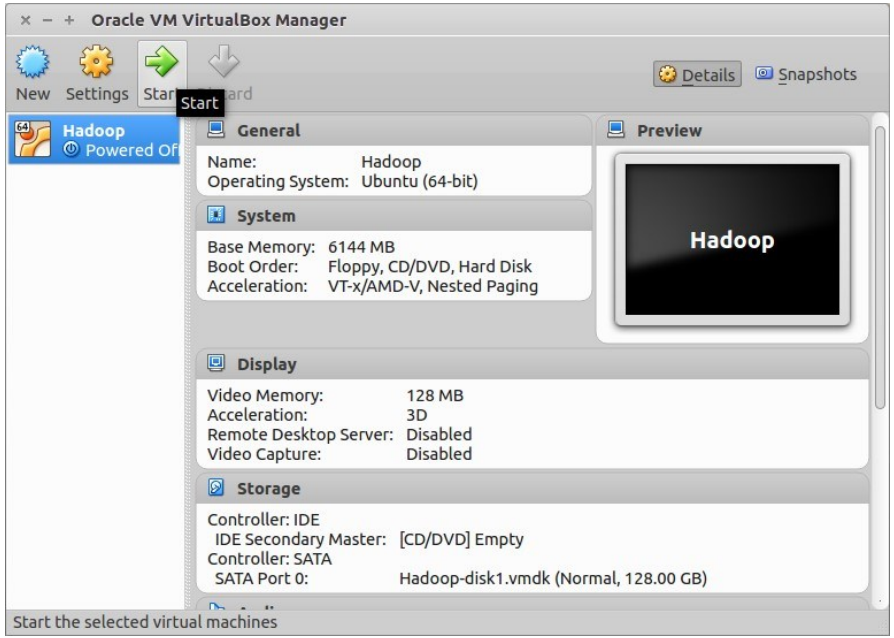
Step 2

Running VM

Setting up VM

Import a VM into VirtualBox

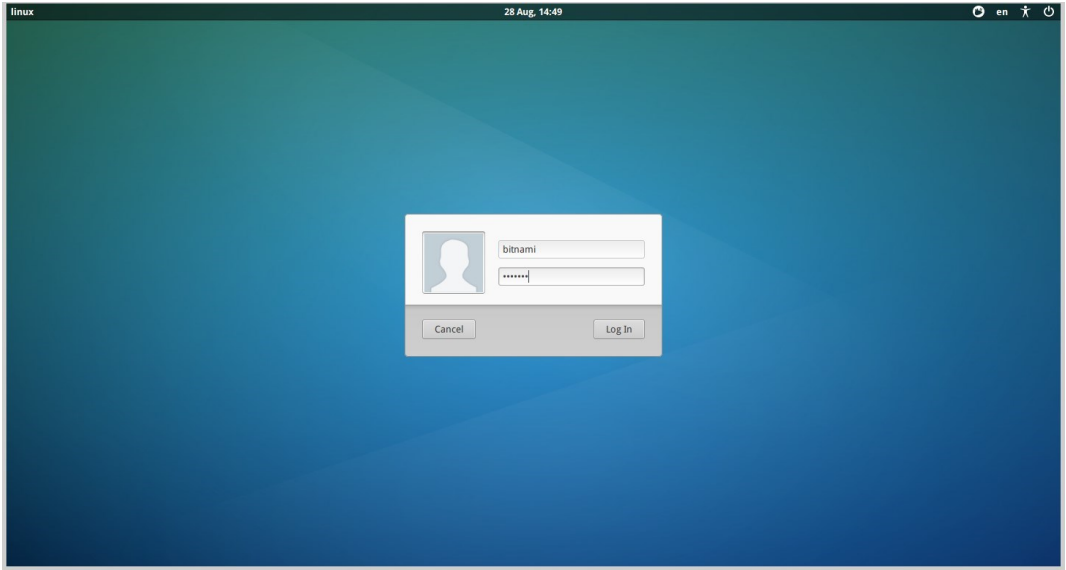
- Click “Start”



Setting up VM

Import a VM into VirtualBox

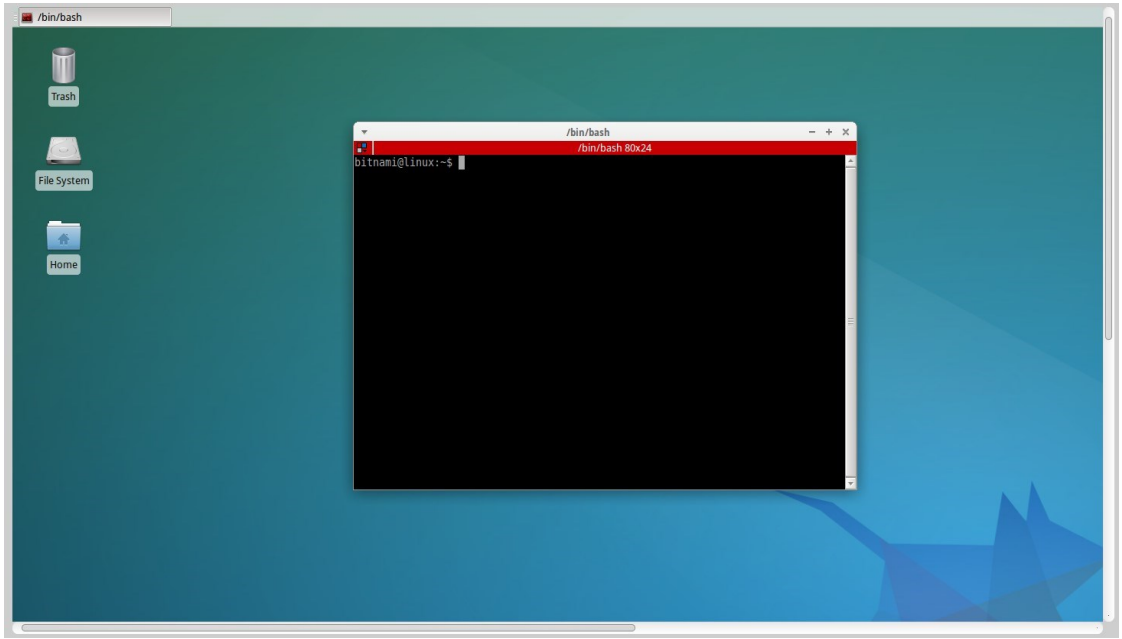
- Login with
  - user: **bitnami**
  - password: **bitnami**



Setting up VM

Import a VM into VirtualBox

- Open a terminal:
  Ctrl+Alt+T



Step 3

Executing Basic HDFS Commands

Hadoop Distributed File System

Start HDFS

Execute start-dfs.sh and jps:

```
bitnami@linux:~$ start-dfs.sh Starting
namenodes on [0.0.0.0]
0.0.0.0: starting namenode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-
bitnami-namenode-linux.out
localhost: starting datanode, logging to /usr/local/hadoop-2.6.0/logs/hadoop-
bitnami-datanode-linux.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop-2.6.0 /logs/hadoop-
bitnami-secondarynamenode-linux.out
bitnami@linux:~$ jps
2451 Jps
2345 SecondaryNameNode
2186 DataNode
2061 NameNode
```

Hadoop Distributed File System

HDFS Commands HDFS

Usage

```
bitnami@linux:~$ hdfs dfs -usage Usage:
hadoop fs [generic options]
     [-appendToFile <localsrc> ... <dst>]
     [-cat [-ignoreCrc] <src> ...]
     [-checksum <src> ...]
     [-chgrp [-R] GROUP PATH...]
     [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
     [-chown [-R] [OWNER][:[GROUP]] PATH...]
     [-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
     [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
     [-count [-q] [-h] <path> ...]
     [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
     [-createSnapshot <snapshotDir> [<snapshotName>]]
     [-deleteSnapshot <snapshotDir> <snapshotName>]
```

Hadoop Distributed File System

HDFS Commands

Transfer data to HDFS

```
bitnami@linux:~$ mkdir test_dir
bitnami@linux:~$ touch test_dir/test_file.txt bitnami@linux:~$
hdfs dfs -put test_dir
bitnami@linux:~$ hdfs dfs -ls
Found 11 items
drwxr-xr-x   - bitnami supergroup          0 2015-05-14 02:39 .sparkStaging
drwxr-xr-x   - bitnami supergroup          0 2015-08-15 23:52 ex_data drwxr-xr-x
- bitnami supergroup          0 2015-04-25 08:36 input drwxr-xr-x   - bitnami
supergroup          0 2015-08-11 18:50 max-temp drwxr-xr-x   - bitnami supergroup
0 2015-05-15 18:38 max-temp-workflow drwxr-xr-x   - bitnami supergroup          0
2015-08-11 18:57 max-temp2 drwxr-xr-x   - bitnami supergroup          0 2015-05-
18 12:30 oozie-bitn -rw-r--r--   1 bitnami supergroup         41 2015-08-13 17:44
sample.txt drwxr-xr-x   - bitnami supergroup          0 2015-05-18 12:27 share
drwxr-xr-x   - bitnami supergroup          0 2015-08-13 20:46 temp drwxr-xr-x   -
bitnami supergroup          0 2015-08-28 15:42 test_dir
bitnami@linux:~$ hdfs dfs -ls test_dir
Found 1 items
-rw-r--r--   1 bitnami supergroup          0 2015-08-28 15:42 test_dir/test_file.txt
```

Your Tasks

- Delete test_dir/test_file.txt from HDFS
- Remove test_dir from HDFS
- Create test_dir2 on HDFS
- Copy test_file.txt to test_dir2 on HDFS

Recaps

- Setting up VM
- Executing basic HDFS commands, e.g.,
  - put
  - copy
  - rm
  - rmdir