

# T1 – Tutorial: VirtualBox, Linux, Python

## Large-Scale Data Analysis

Fabian Gieseke

Image Group  
Department of Computer Science  
University of Copenhagen

Universitetsparken 1, Room 1-1-N110  
[fabian.gieseke@di.ku.dk](mailto:fabian.gieseke@di.ku.dk)

# Outline

① Virtualbox

② Linux

③ Python

# Outline

① Virtualbox

② Linux

③ Python

# VirtualBox: Installation

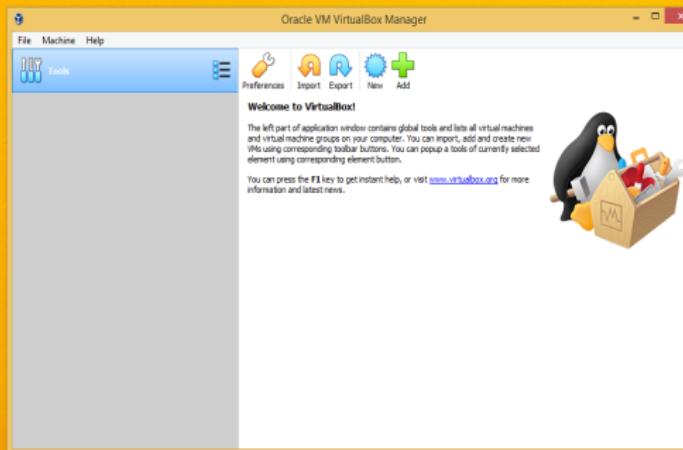
## Procedure

- 1 Visit download webpage: <https://www.virtualbox.org/wiki/Downloads>
  - ▶ Windows/Mac: Download the corresponding VirtualBox platform package for your system. Double click on file and follow instructions ...
  - ▶ Linux: For instance, on Ubuntu 18.04, `sudo apt install virtualbox virtualbox-qt virtualbox-dkms` is enough. Otherwise, check installation procedures for your system (e.g., <https://www.virtualbox.org/manual/ch02.html#install-linux-host>).
- 2 System requirements: VirtualBox does not need many resources. However, for the image below and for the assignments, you will need  $\geq 20\text{GB}$  free disk space!
- 3 Extensions (optional): We don't need the VirtualBox extensions provided on the webpage for the course, but they might be a nice add-on (e.g., they improve the performance of the virtual machine and render copy & paste between host and virtual machine possible).
- 3 We have prepared a VirtualBox image for you, which is based on Ubuntu Mate 18.04 (64bit). Please download the `LSDA.ova` image from here (about 4.1GB):  
[https://sid.elda.dk/cgi-sid/ls.py?share\\_id=FqlW8M6qFm](https://sid.elda.dk/cgi-sid/ls.py?share_id=FqlW8M6qFm)

Hot picks:

- Pre-built virtual machines for developers at [Oracle Tech Network](#)

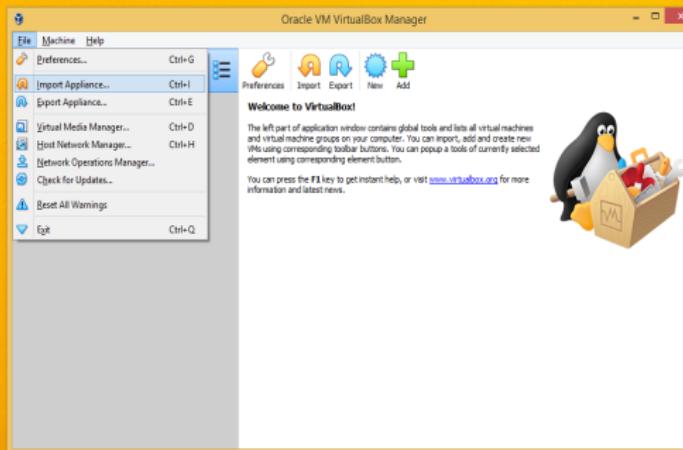
# VirtualBox: Installation



Start VirtualBox ...

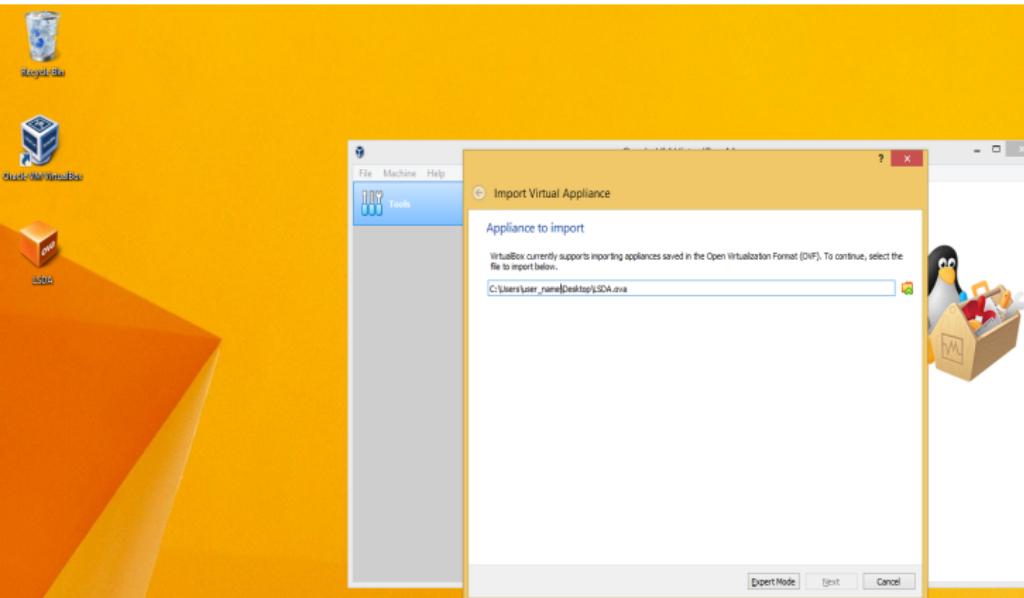


# VirtualBox: Installation



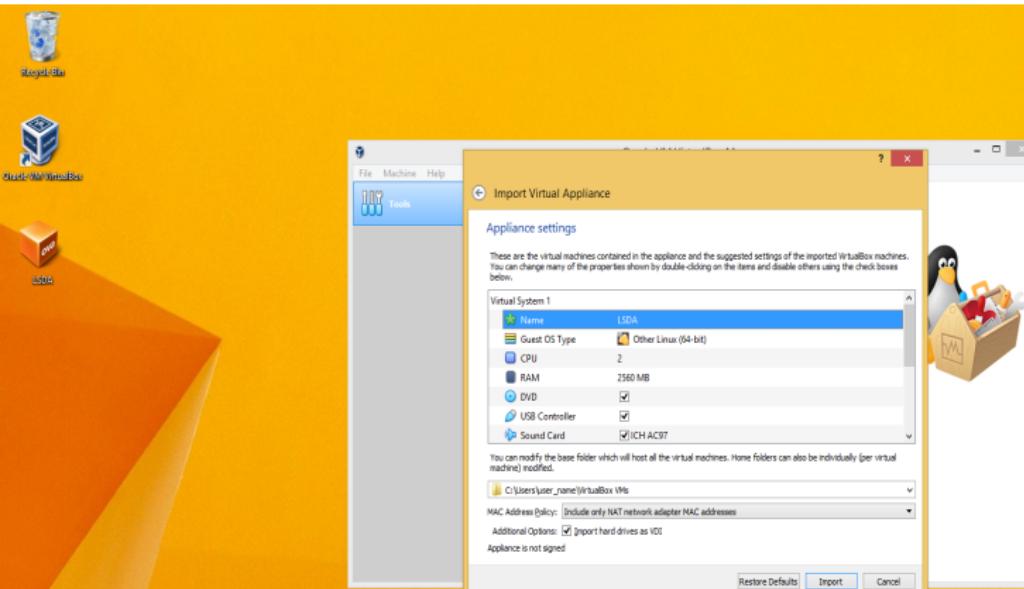
We have to import the image. Click "File → Import Appliance"

# VirtualBox: Installation



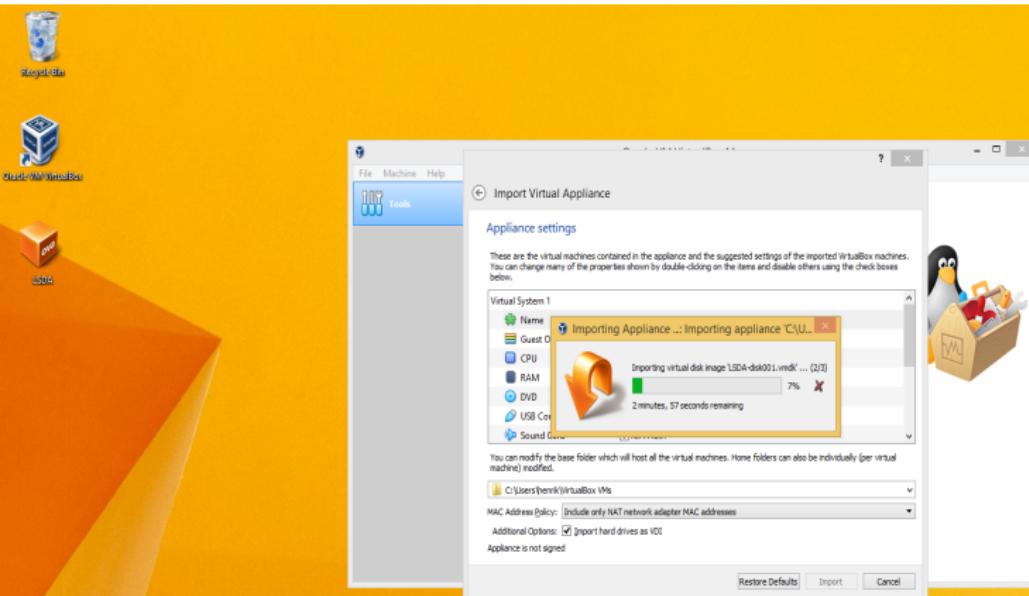
Select the VirtualBox image you have downloaded before ...

# VirtualBox: Installation

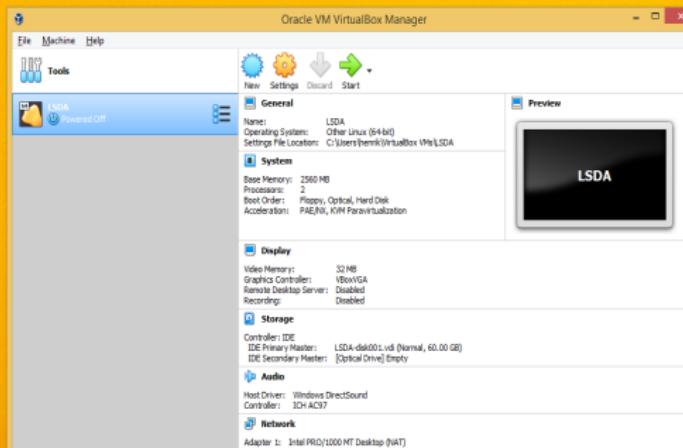


Just import the existing configurations ...

# VirtualBox: Installation

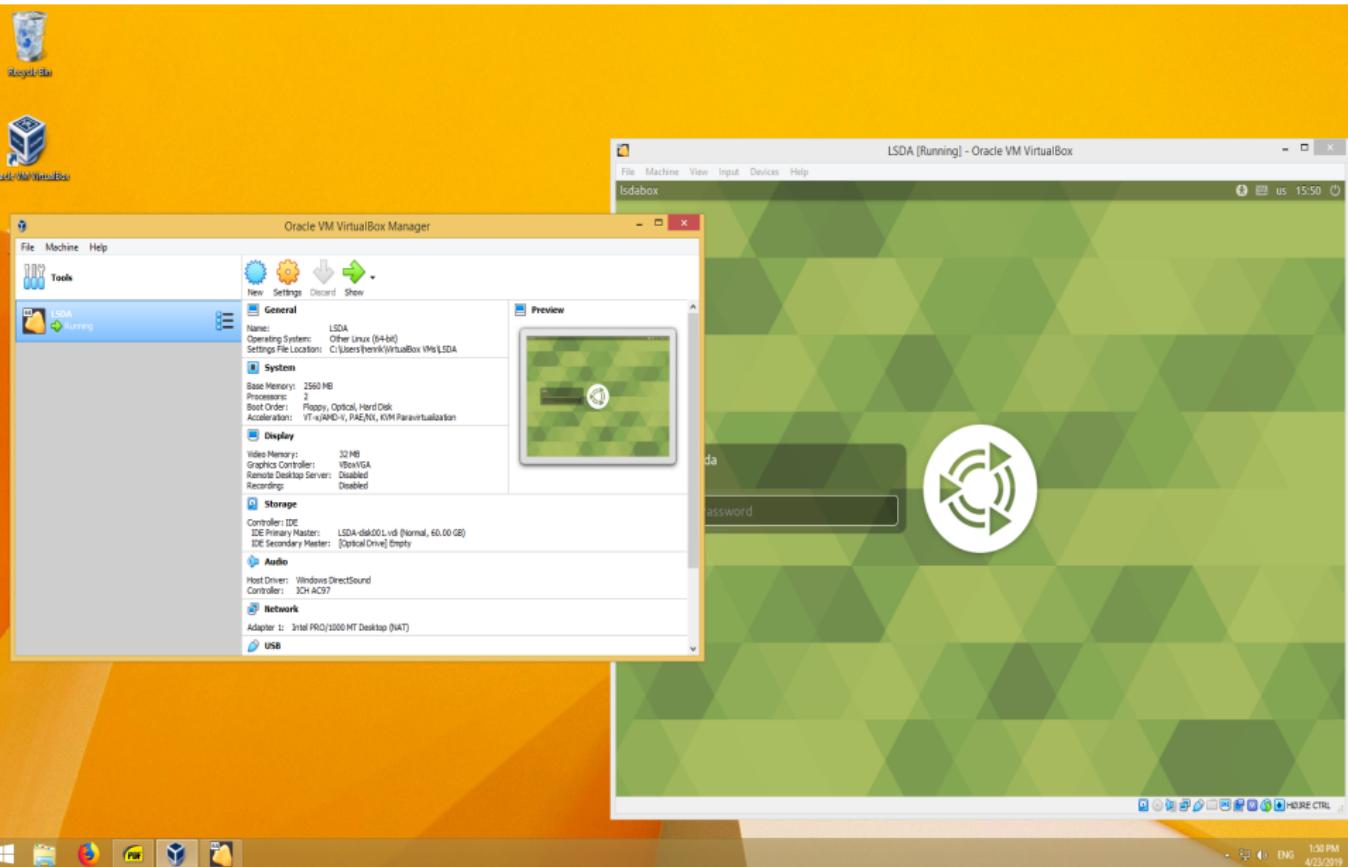


# VirtualBox: Installation

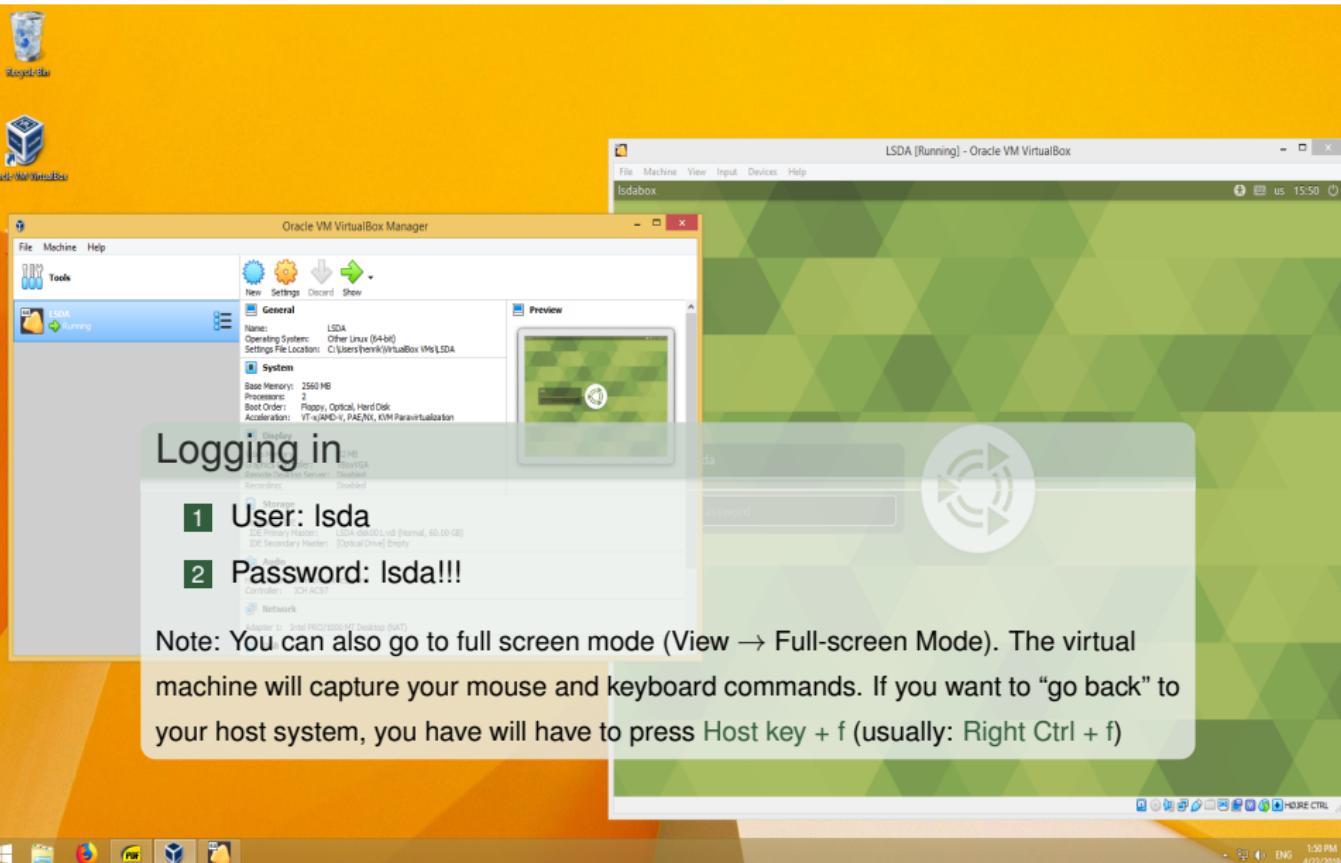


The LSDA image should be available in the left panel.  
Select and click "Start", you should see something like

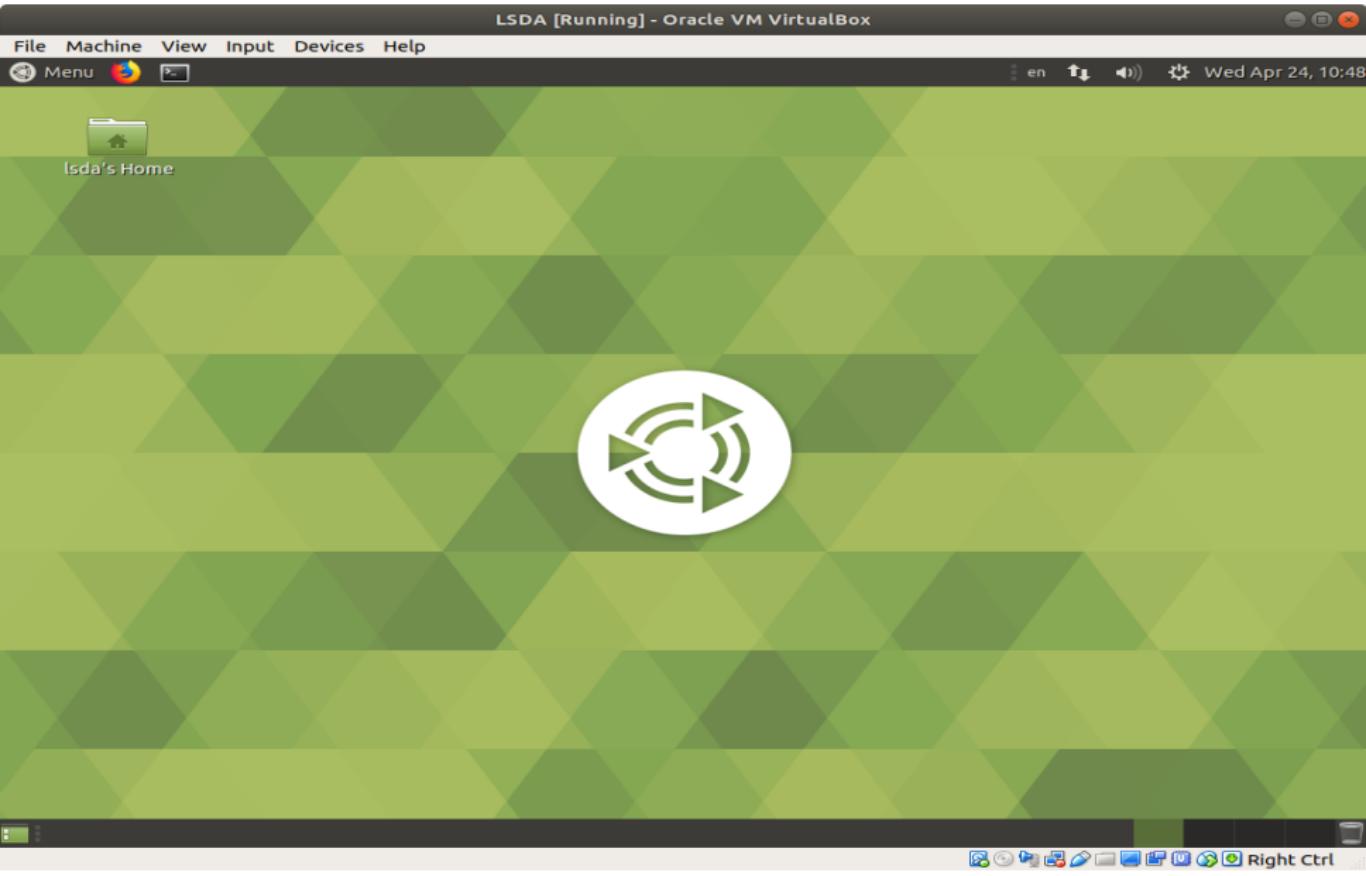
# VirtualBox: Installation



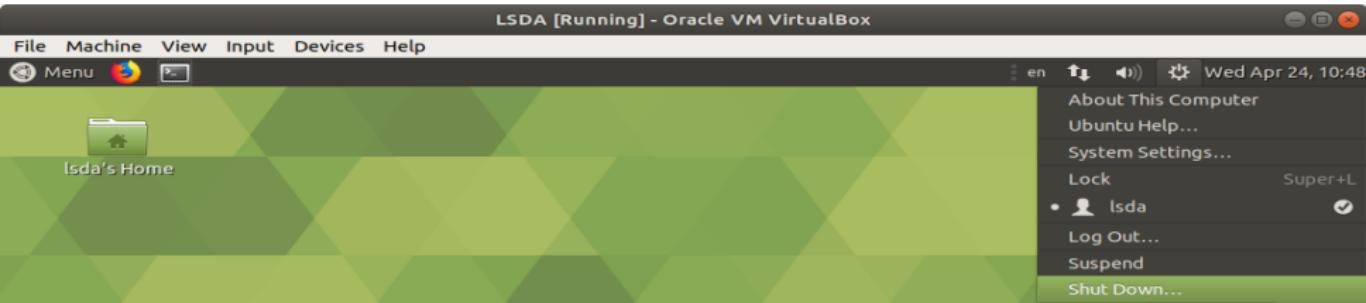
# Logging in . . .



# Ubuntu Mate 18.04



# Shutting down ...



## Shut Down of Virtual Machine

If you want to shut down the virtual machine, you can select “Shut Down” from the dropdown menu in the top right corner of your system.

In case the system does not react anymore, you can also “power off” your virtual machine: Right click on image in the left panel of the Oracle VM VirtualBox Manager and select “Close →Power Off”.

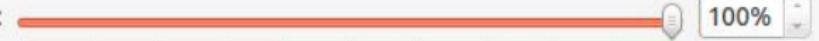
# Optional: Performance

**System**

Motherboard Processor Acceleration

Processor(s):  2

1 CPU 8 CPUs

Execution Cap:  100%

1% 100%

Extended Features:  Enable PAE/NX

**VirtualBox**

Make sure the virtual machine is shut down (see previous slide). In the VirtualBox Manager: Select LSDA image in the left panel and click on “Settings”.

1 Select “System → Processor” and check that the number of processors is set to 2.

2 Feel free to change other settings, but do not change the amount of Base Memory (keep it to 2560 MB).

Note: The availability of additional processors and video memory depends on your system. Get in touch with us if you run into problems.

**Cancel** **OK**

# Optional: SSH Access

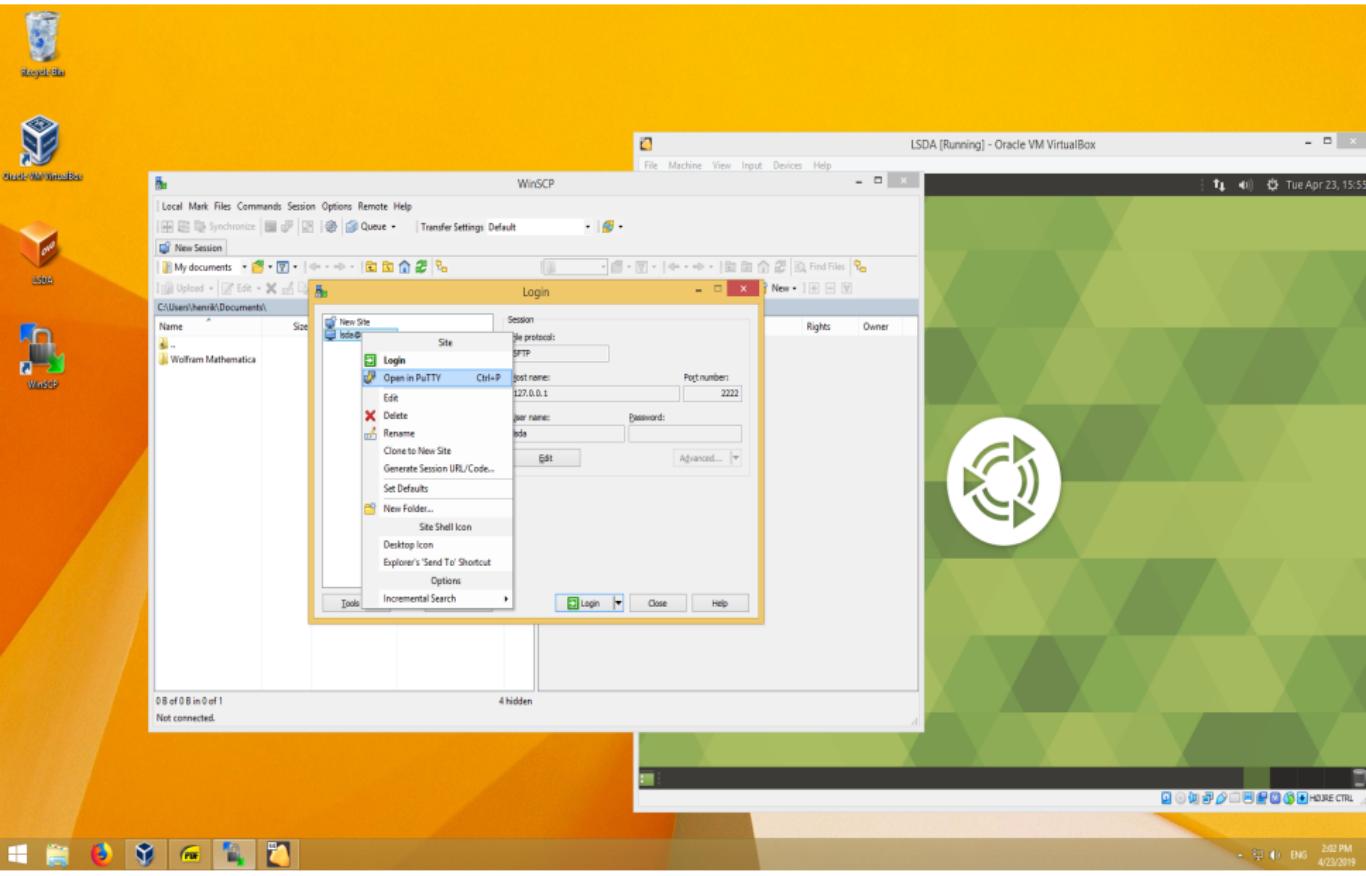
The screenshot shows a Windows desktop environment. In the background, there's a virtual machine window titled "LSDA [Running] - Oracle VM VirtualBox". The desktop icons include "LogMeIn", "Oracle VM VirtualBox", and "LSDA". In the foreground, a WinSCP interface is open, showing a file transfer session between a local folder ("My documents") and a remote site ("lsda"). A "Login" dialog box is overlaid on the WinSCP window, prompting for "Site protocol" (set to "SFTP"), "Host name" (set to "127.0.0.1"), and "Port number" (set to "2222"). A password field is also present. A callout bubble with the title "Remote Login via SSH" points to the "Host name" field.

## Remote Login via SSH

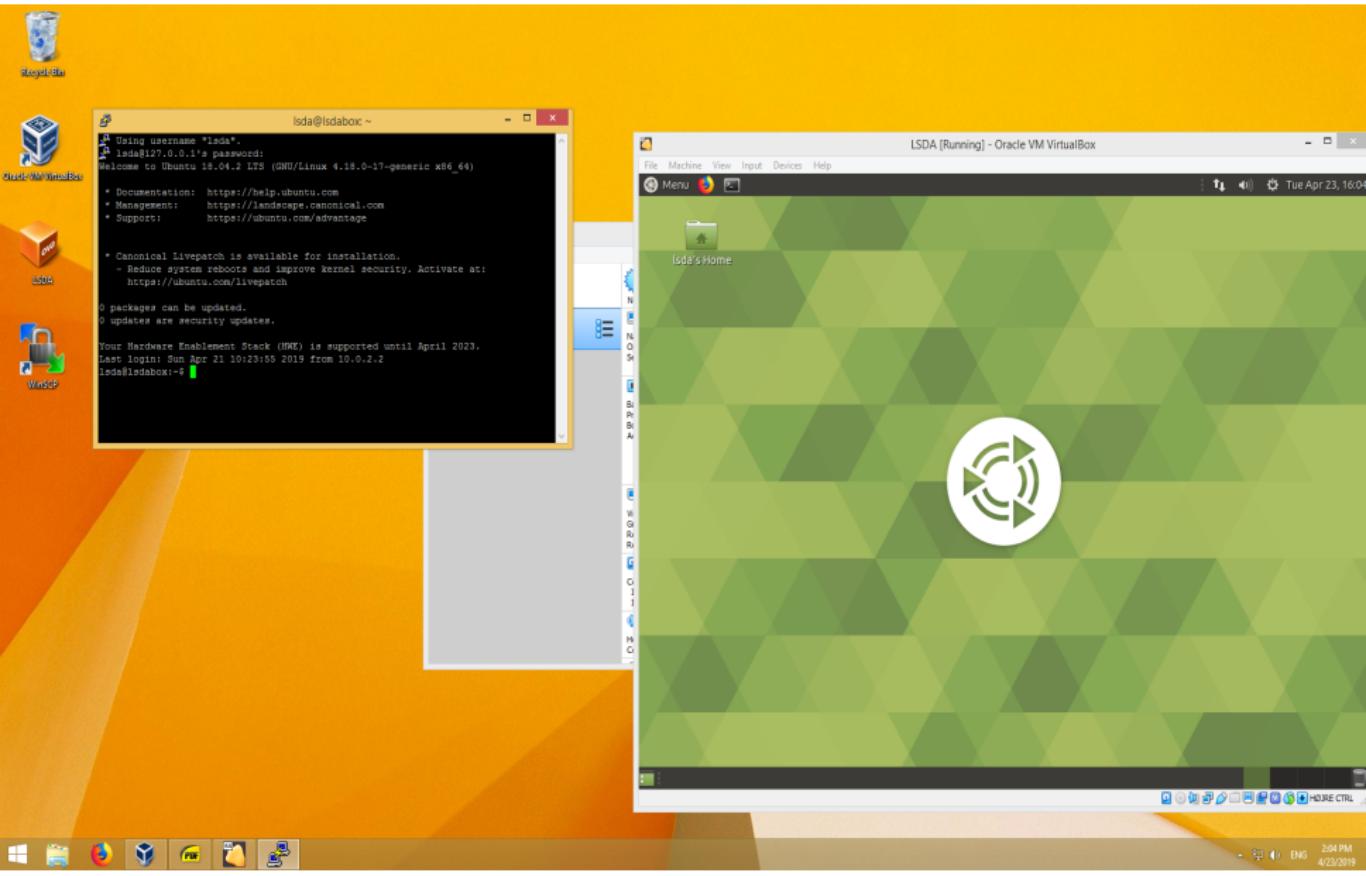
In principle, you can do all steps within the virtual machine. However, if you like, you can also use "ssh" to access the virtual machine from your computer.

- 1 Boot your virtual machine ...
- 2 Mac/Linux: On Mac and Linux hosts, simply start a terminal (on your computer!). Within the terminal, type in: `ssh -p 2222 lsda@127.0.0.1`
- 3 Windows: You will need to install a SSH client first. As an example, we will use WinSCP (<https://winscp.net/download/WinSCP-5.9.4-Setup.exe>). Install and start WinSCP. Then, start a new session (see background for login details).

# Optional: SSH Access



# Optional: SSH Access



# Outline

① Virtualbox

② Linux

③ Python

# Linux and Terminal

The screenshot shows a Linux desktop environment with a green and grey geometric background. In the top right corner, there is a system tray with icons for file transfer, volume control, and system status. The date and time 'Wed Apr 24, 11:03' are also visible. A terminal window titled 'LSDA [Running] - Oracle VM VirtualBox' is open in the background, showing a command-line session. In the foreground, a window titled 'Command Line' is active, displaying a terminal session with the following content:

```
lsda@lsdabox: ~
File Edit View Search Terminal Help
lsda@lsdabox: ~$ pwd
/home/lsda
lsda@lsdabox: ~$ ls -l
total 40
drwxr-Xr-x 2 lsda lsda 4096 Apr 19 09:09 Desktop
drwxr-Xr-x 2 lsda lsda 4096 Apr 19 09:37 Downloads
drwxr-Xr-x 2 lsda lsda 4096 Apr 19 09:10 Host
drwxr-Xr-x 2 lsda lsda 4096 Apr 19 09:09 Pictures
drwxr-Xr-x 3 lsda lsda 4096 Apr 19 09:20 Snap
drwxr-Xr-x 2 lsda lsda 4096 Apr 19 09:09 Templates
drwxr-Xr-x 2 lsda lsda 4096 Apr 19 09:09 Videos
```

Below the terminal window, a list of numbered tasks is displayed:

- 1 Start terminal (see background). You are now in your "home directory". You can use the `pwd` command to check for the absolute path (hit the RETURN key to execute).
- 2 Make use of `ls -l` to show the content of the current directory. Change the command to `ls -la`. What's the difference?  
Hint: The keyboard is set to 'En'. You can change this via the systems settings!
- 3 Make use of further Linux commands to
  - ▶ Type in `man ls`. This shows infos about `ls` (leave with Shift+q).
  - ▶ Check what you can do with the arrow keys (up/down) ...
  - ▶ Create and delete directories via `mkdir` and `rm -r`.
  - ▶ Type in `htop`. What's the output? Leave program via Ctrl+c.
- 4 At home: Play around with other commands (see link on next slide).

# Optional: Terminal

## Learning the shell

This tutorial has been deprecated! Please use the new version at [LinuxCommand.org](http://LinuxCommand.org)

- [LinuxCommand](#)
- [Learning the shell](#)
- [Writing shell scripts](#)
- [Script library](#)
- [SuperMan pages](#)
- [Who, What, Where, Why](#)

[Previous](#) | [Contents](#) | [Next](#)

### Why bother?

Why do you need to learn the command line anyway? Well, let me tell you a story. Not long ago we had a problem where I used to work. There was a shared drive on one of our file servers that kept getting full. I won't mention that this legacy operating system did not support user quotas; that's another story. But the server kept getting full and stopping people from working. One of the software engineers in our company spent the better part of a day writing a C++ program that would look through the directories of all the users and add up the space they were using and make a listing of the results. Since I was forced to use the legacy OS while I was on the job, I installed a version of the bash shell that works on it. When I heard about the problem, I realized I could do all the work this engineer had done with this single line:

```
du -s * | sort -nr > $HOME/space_report.txt
```

Graphical user interfaces (GUIs) are helpful for many tasks, but they are not good for all tasks. I have long felt that most computers today do not use electricity. They instead seem to be powered by the "pumping" motion of the mouse! Computers were supposed to free us from manual labor, but how many times have you performed some task you felt sure the computer should be able to do? You ended up doing the work by tediously working the mouse. Pointing and clicking, pointing and clicking.

I once heard an author remark that when you are a child you use a computer by looking at the pictures. When you grow up, you learn to read and write. Welcome to Computer Literacy 101. Now let's get to work.

### Contents

1. [What is "the shell?"](#)
  1. [What's an xterm, gnome-terminal, konsole, etc.?](#)
  2. [Starting a Terminal](#)
  3. [Testing the Keyboard](#)
  4. [Using the Mouse](#)
2. [Navigation](#)
  1. [File System Organization](#)
  2. [pwd](#)
  3. [cd](#)
3. [Looking Around](#)
  1. [ls](#)
  2. [less](#)
  3. [file](#)
4. [A Guided Tour](#)
  1. [/](#)
  2. [/boot](#)
  3. [/etc](#)
  4. [/bin, /usr/bin](#)
  5. [/sbin, /usr/sbin](#)
  6. [/usr](#)
  7. [/usr/local](#)
  8. [/var](#)
  9. [/lib](#)

# Outline

① Virtualbox

② Linux

③ Python

# Hello Python

LSDA [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Menu examples

File Edit View Go Bookmarks Help

Back Forward Up × C examples

Places Computer lsda Desktop File System Documents Downloads Music Pictures

lsda hadoop spark hello.py

3 items, Free space: 50.5 GB

hello.py (~/LSDA/examples) - Pluma

File Edit View Search Tools Documents Help

Open Save Undo Redo

hello.py \*

1 print("Hello!")

Python Tab Width: 4 Ln 1, Col 16 INS

lsda@lsdabox:~/LSDA/examples

Hello LSDA!

drwxrwxr-x 6 lsda lsda 3096 Apr 19 09:54 LSDA  
drwxr-xr-x 2 lsda lsda 4096 Apr 19 09:09 MUSIC  
drwxr-xr-x 2 lsda lsda 4096 Apr 19 09:09 Public  
drwxr-xr-x 3 lsda lsda 4096 Apr 19 09:20 snap  
drwxr-xr-x 2 lsda lsda 4096 Apr 19 09:09 VIDEOS  
lsda@lsdabox:~\$ cd LSDA  
lsda@lsdabox:~/LSDA\$ total 12  
drwxrwxr-x 3 lsda lsda 12288 Apr 19 11:07 hello.py  
-rw-rw-r-- 1 lsda lsda 16 Apr 24 11:07 hello.py  
drwxrwxr-x 3 lsda lsda 12288 Apr 19 11:07 LSDA  
lsda@lsdabox:~/LSDA\$ Hello!  
lsda@lsdabox:~/LSDA\$

1 Click on “lsda’s Home” and go to “LSDA/examples”.  
2 Within this directory, create a document “hello.py” (right click: “Create Document→Empty Document”). Add the content shown in the background, save the file.  
3 Start terminal, enter the directory and show its content.  
4 Run the Python program by executing `python hello.py`.

lsda@lsdabox:~/LSDA... examples hello.py (~/LSDA/exam... Right Ctrl

# Jupyter Notebooks & Python

The screenshot shows a Linux desktop environment with several windows open:

- A terminal window titled "lsda@lsdabox:~" showing command-line output related to Jupyter Notebook setup.
- A Firefox browser window titled "Home - Mozilla Firefox" showing the URL "0.0.0.0:8888/tree". The page content includes a heading "Jupyter Notebooks: Interactive Python!" and two numbered steps:
  - 1 Activate your Python environment by executing source `./venvs/lsda/bin/activate.`
  - 2 Execute jupyter notebook
- A file manager window titled "lsda@lsdabox:~" showing a list of files and folders, including ".ipynb" files and a ".DS\_Store" file.

At the bottom of the screen, there is a dock with icons for various applications, and the system tray shows the date and time as "Wed Apr 24, 11:15".

By executing these steps, the webbrowser should open automatically with a so-called Jupyter notebook being available in the tab. You can use this notebook to interactively run Python code. For more details, check out the online documentation at <http://jupyter.org>.