How to install SMT tools on a virtual machine

This documentation shows how to install various SMT software on a virtual machine.

Here we focus on a Windows host running "Oracle Virtual box", and the guest machine being on Ubuntu 16.

Status: 09/12/2016 BP V0.03

Contents

Contents 1

Preliminaries 1

Prepare Oracle VM and Ubuntu ISO disk 2

Create the "smt" machine 3

- Install Ubuntu OS on the virtual server: 6
- Ubuntu installation 8
- Booting the new server 13
- Adding "guest additions" 14
- Install the necessary packages for the course 15
- Network configuration (optional) 16

Preliminaries

1. Your host computer has at least 50Gb of Disk available and minimum 2Gb RAM to allocate to the virtual machine (be aware that actually in order to compile certain software you might have to install it on a 4Gb computer (and allocate ~ 3 Gb to the virtual machine)
2. Your host is able to connect to Internet without "heavy" firewall rules

Note that it takes approximately about one hour to download / install all.

Prepare Oracle VM and Ubuntu ISO disk

Install “Oracle VM VirtualBox” on your PC, google “Oracle VM” then choose to download “Oracle VM VirtualBox”, choose the “Windows (64-bit)” file, that’s a windows installer.

The latest release is virtualbox-5.1 (21/nov/2016)

e.g.: VirtualBox-5.1.10_112026_win.exe

Get Ubuntu 16.04 (From Ubuntu website) => create an ISO disk.

For example, connect to

e.g. ubuntu-16.04.1-desktop-amd64.iso, save it somewhere on your PC.

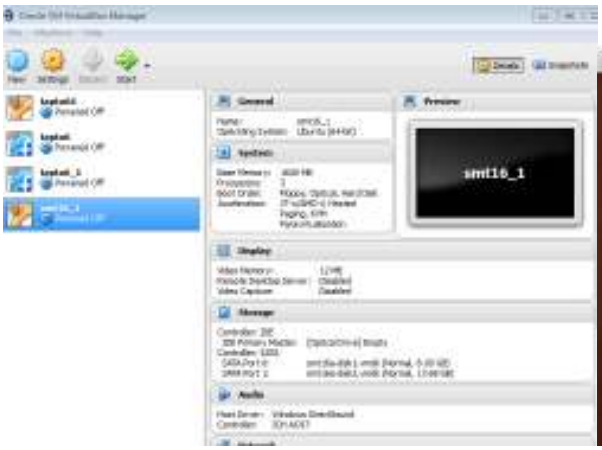
Launch the application “Oracle Virtual Box”

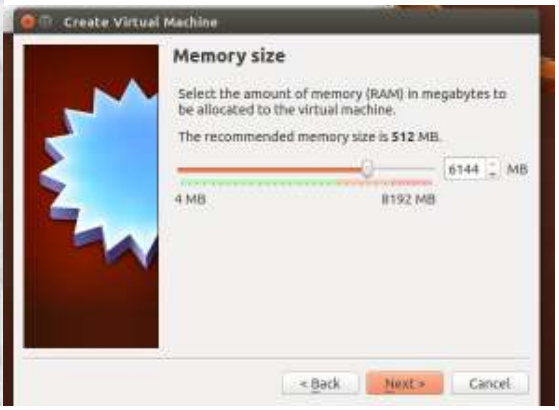
Create the “smt” machine

Create a new machine with Ubutu 64bits

Click on « new »

Give at the very least 4 Gb of memory for the virtual server (6 should be fine, more does not harm)

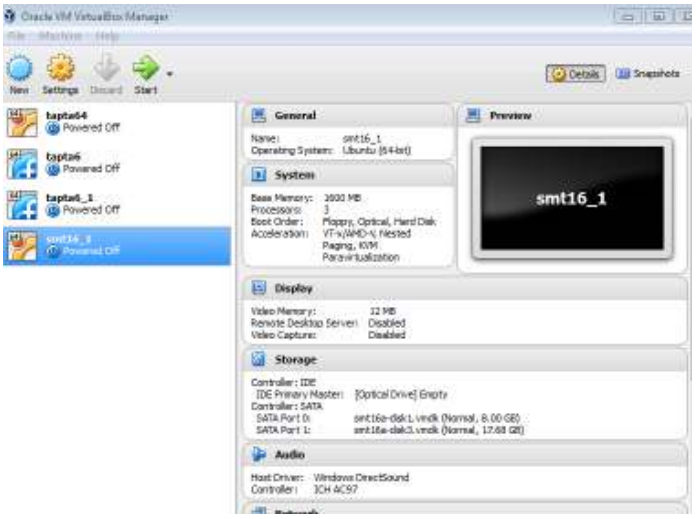




Then create a hard disk:

Give a size of 50Gb (note that Tapta will not use this size at the beginning, only about 20Gb will be necessary until you train a model on it).

Then your system should look like something like this:

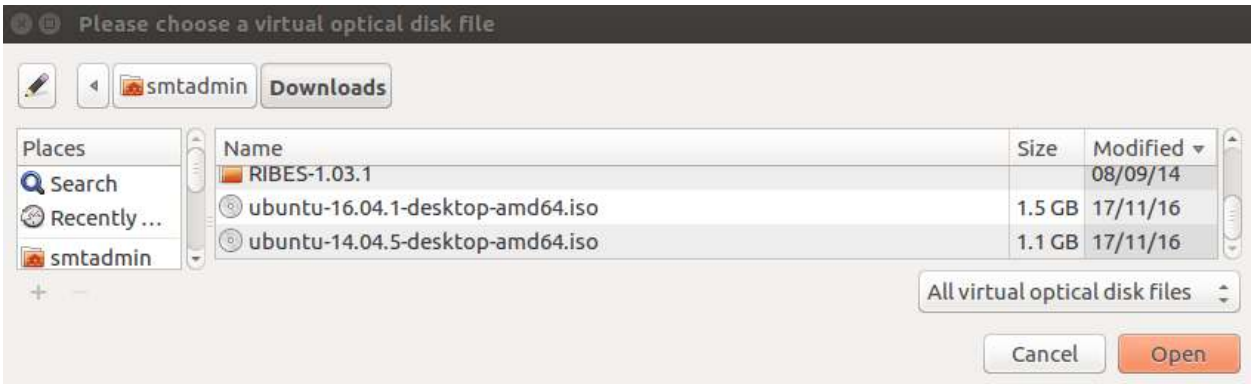


Install Ubuntu OS on the virtual server:

Click on “start” for the newly created server.

Choose the ISO image you have downloaded…

Here select the Ubuntu 16_04 you downloaded

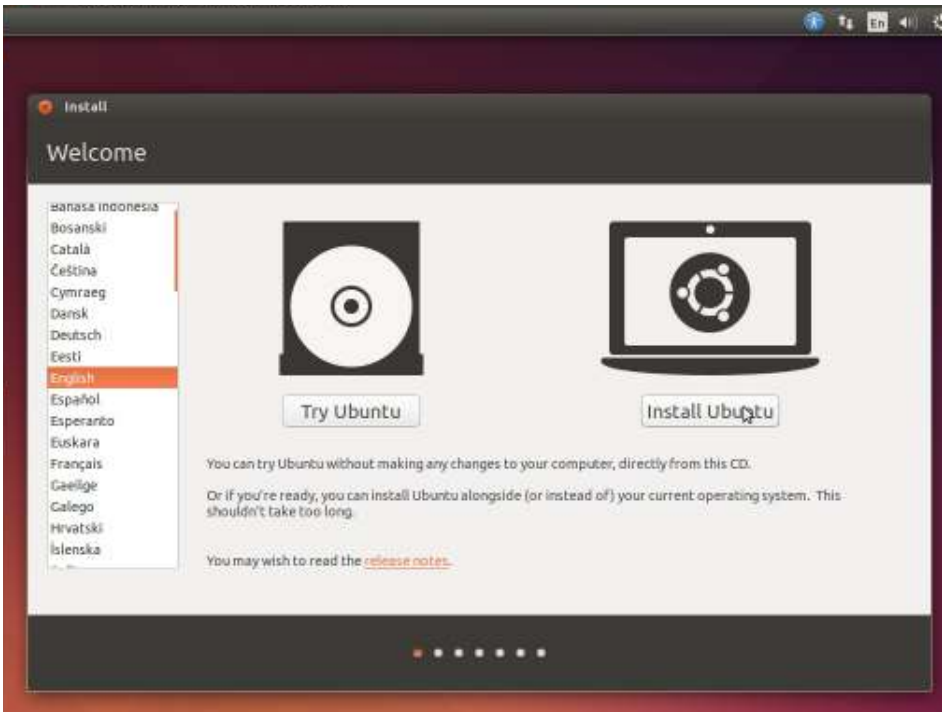


Your server is now ready to start

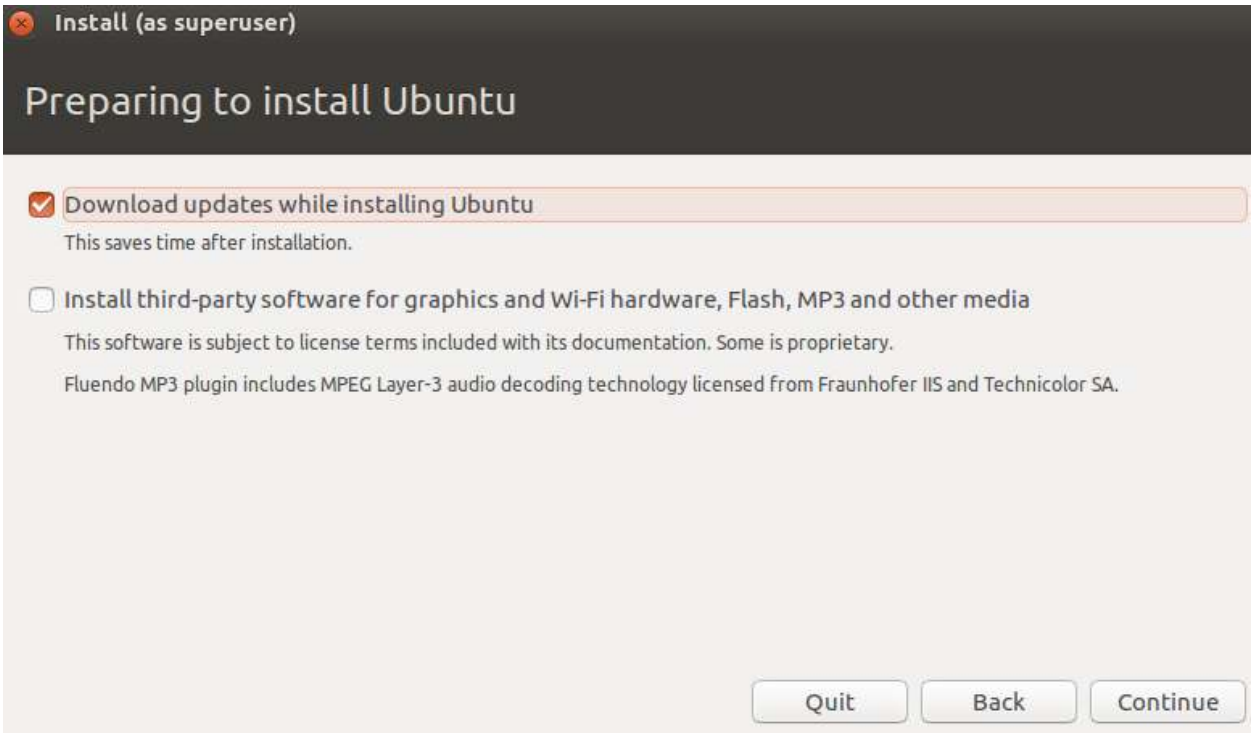
Click “start”

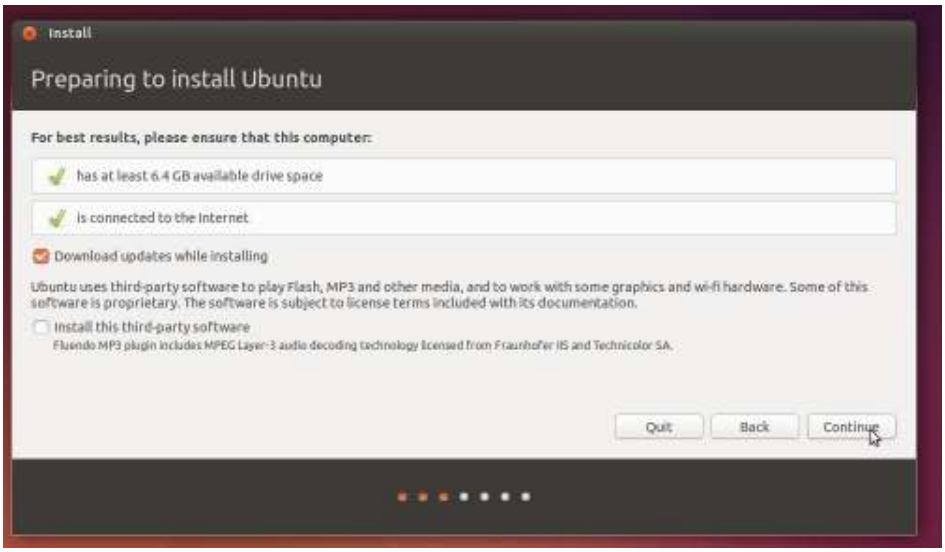
Ubuntu installation

After starting the server, should boot on Ubuntu and show you the following screen:

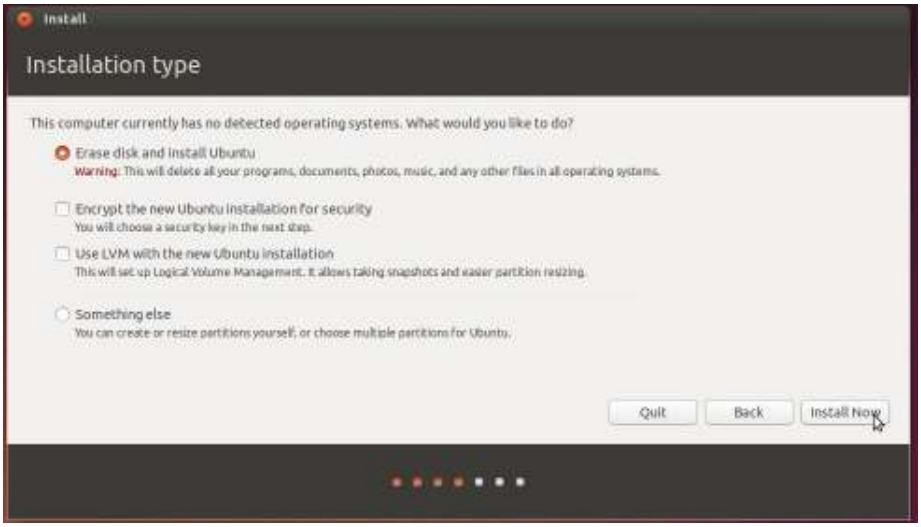


Then be sure you are connected to internet, click on « install Ubuntu »

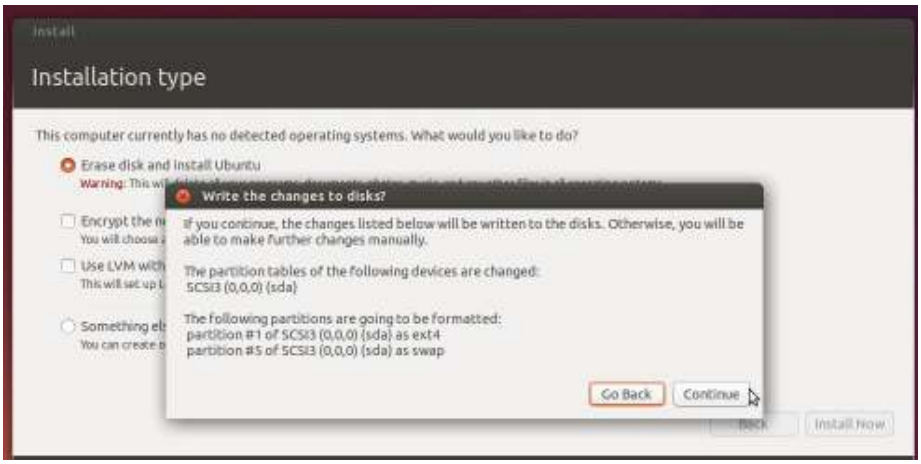

Install (as superuser)
Preparing to install Ubuntu
Download updates while installing Ubuntu
This saves time after installation.
Install third-party software for graphics and Wi-Fi hardware, Flash, MP3 and other media
This software is subject to license terms included with its documentation. Some is proprietary.
Fluendo MP3 plugin includes MPEG Layer-3 audio decoding technology licensed from Fraunhofer IIS and Technicolor SA.
Quit
Back
Continue


Install
Preparing to install Ubuntu
For best results, please ensure that this computer:
has at least 6.4 GB available drive space
is connected to the Internet
Download updates while installing
Ubuntu uses third-party software to play Flash, MP3 and other media, and to work with some graphics and wi-fi hardware. Some of this software is proprietary. The software is subject to license terms included with its documentation.
Install this third-party software
Fluendo MP3 plugin includes MPEG Layer-3 audio decoding technology licensed from Fraunhofer IIS and Technicolor SA.
Quit
Back
Continue

The following screen asks to format the disk

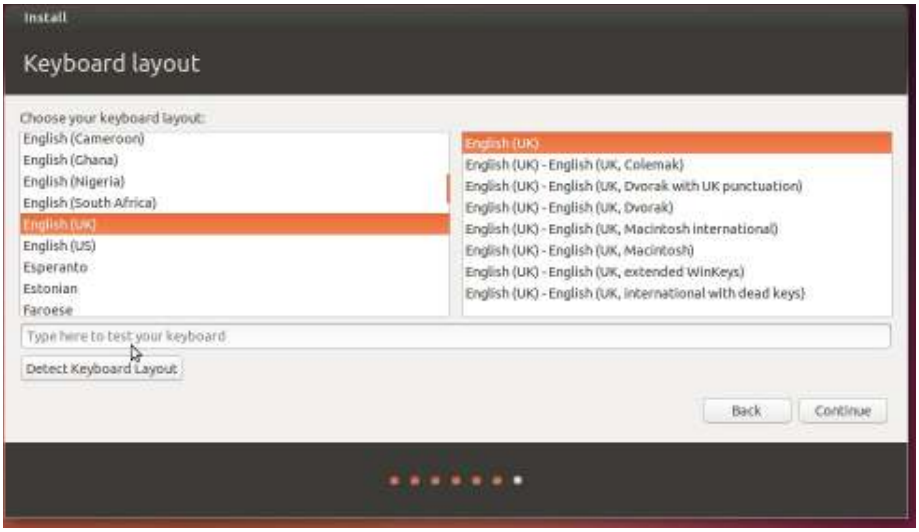


(don’t worry this is the virtual disk !) you can click continue



Select your time zone…

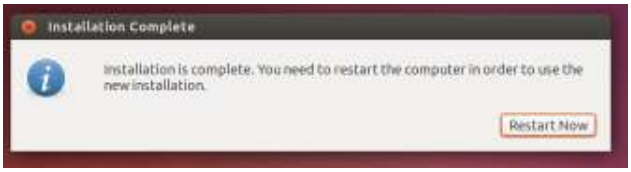
Select your keyboard:



Give a name to your server (here smt)

(your computer name, user name, password will be used by the virtual machine only)

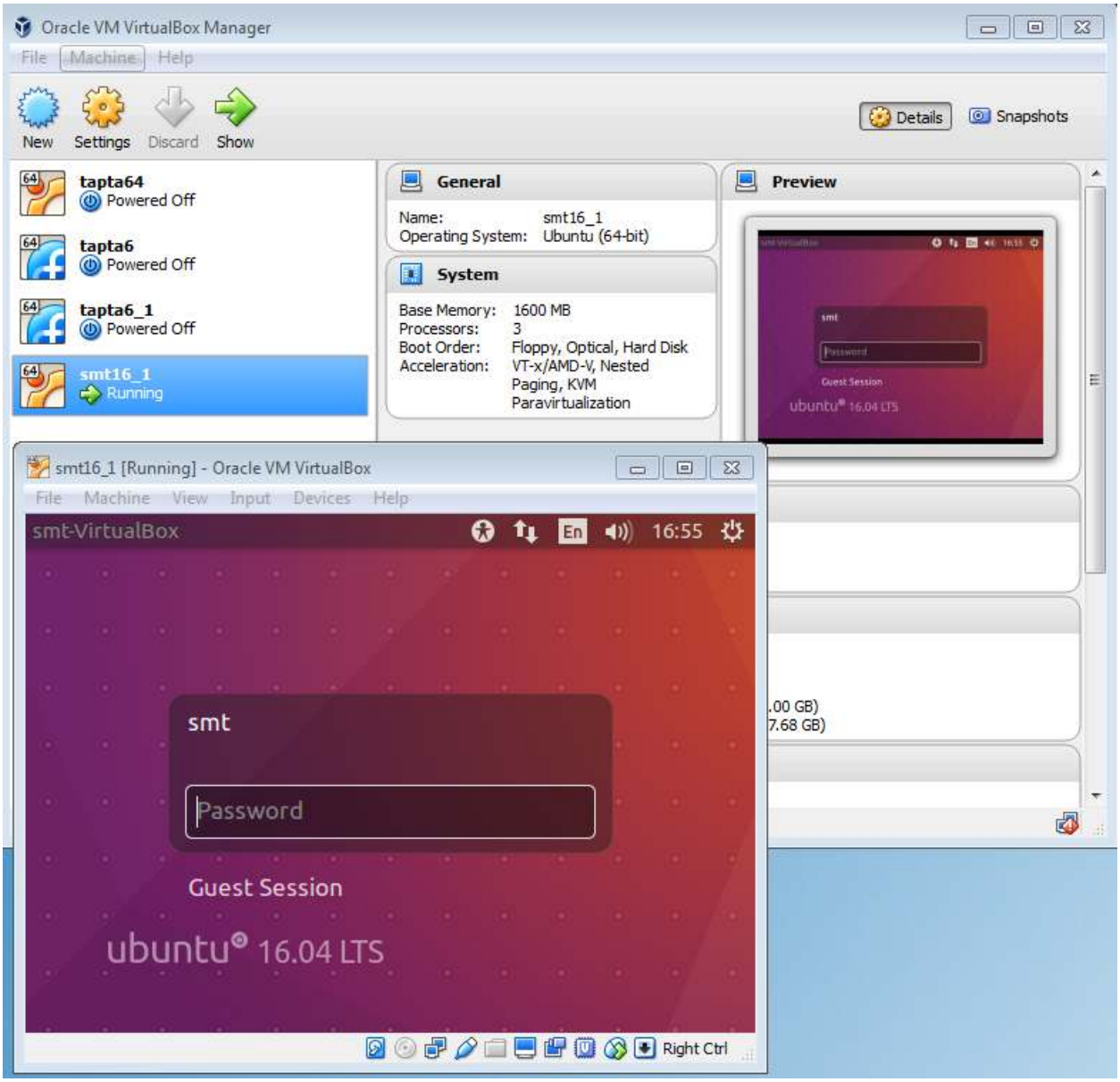
After few screens, it takes about 5 minutes, you will see:



Then click on “Restart now” and the server will reboot.

(it will first show a message saying “please remove the installation medium, then press ENTER:”, ignore it and press “enter”).

Booting the new server



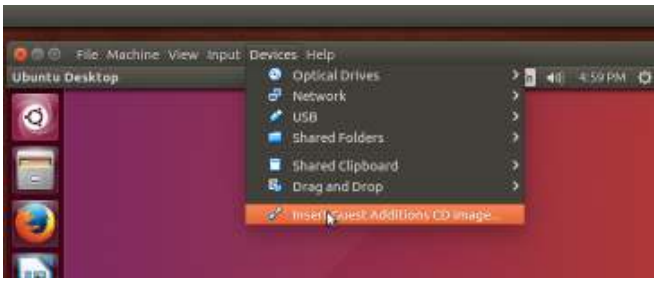
The username is the one you provided (usually smt), enter the corresponding password

Adding “guest additions”

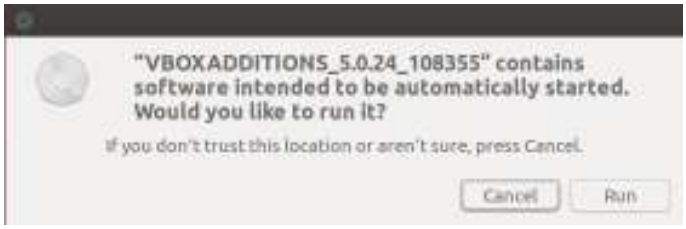
You can see that the server is not “user friendly” (cannot be resized, etc.) you need to install the guest additions for that.

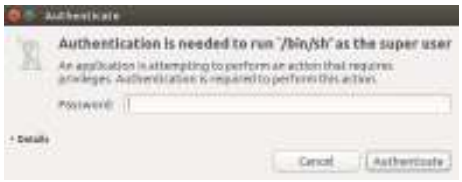
In another browser, get Ubuntu guest additions and save the iso image somewhere on your PC: eg. VBoxGuestAdditions.iso

On the main menu (not inside the Ubuntu, but the main menu of the Virtual Machine), choose “Devices” then “insert guest additions CD Image...”



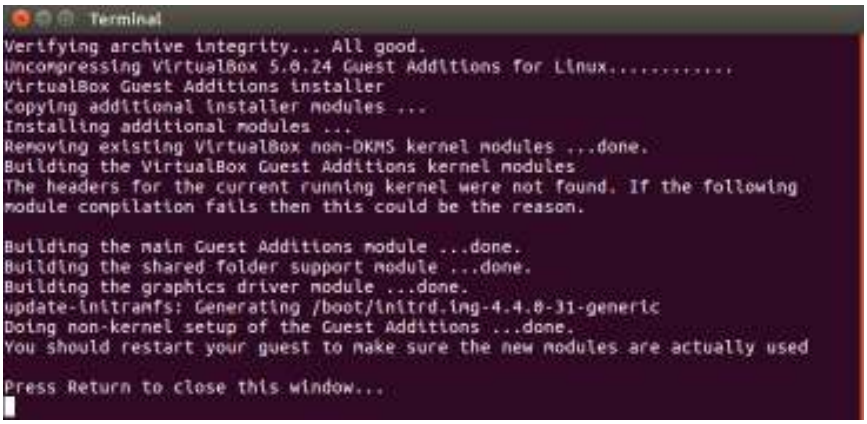
Then click “run”, it will ask for the “smt” password, and will install guest addition CD





followed by the “smt” password:

then the “guest” addition is running…



press return in the window, then reboot the server.

Then start the virtual server

The server boots.

Then you can restart the server, you can now resize the Ubuntu window

Install the necessary packages for the course

Install SMT software, get the « install » bash from Bruno:

Open a terminal and type:

```
git clone https://github.com/mlearningbruno/giancourse.git Icon2016
```

it will display something like:

```
Cloning into 'Icon2016'...
remote: Counting objects: 27, done.
remote: Compressing objects: 100% (24/24), done.
remote: Total 27 (delta 4), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (27/27), done.
Checking connectivity... done.
```

You can now launch the install script:

Note that two install scripts are provided: a “full” install and a “minimal” install

The full-install tries to install these additional packages:

- Tensorflow playground (so that you can use it locally)
- Cdec: another MT engine
- Haddoop: so that you can try to proceed with using Joshua MT
- Char-rnn: (lua torch based) for generating Shakespear texts

- Indic-corpora: a bigger Hindi corpus (https://ufal.mff.cuni.cz/hindencorp)

Now proceed with the installation you want (Note that you can start with the « minimal » and later launch the « full install »

1. Minimal installation script:
   `bash Icon2016/labs/installSmtLabs-min.sh`

   ➔it might take around half an hour for the script to download all software and compile

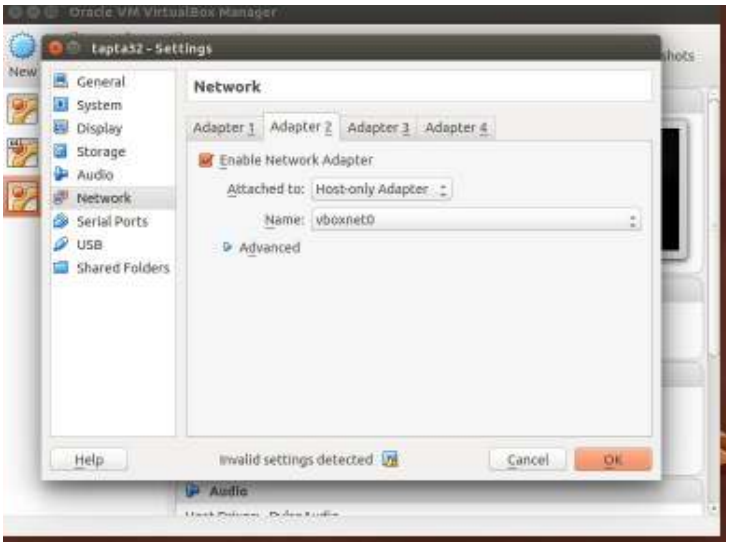
2. Full install:
   `bash Icon2016/labs/installSmtLabs.sh`

   ➔it might take around one hour for the script to download all software and compile

Please DO read the file Icon2016/labs/WelcomePleaseRead.txt which briefly describe the components installed and provides some information about copyright/terms-of-use issues of these components.

Network configuration (optional)

Configure network…



Note that if you are running the VM, you can get its IPaddress launching

Ifconfig

And getting the eth1 address as the VM address, which you can then use on the host machine