

CS585/DS503: Downloading the Virtual Machine

Option 1: Using the Already-Built VM

- This VM has almost all needed software, which you will use for your projects.
- It does not contain Spark. I will provide you with another VM that has Spark later in the semester.
- **Pros: Everything is built for you, and you are ready to code your projects**
- **Cons: The VM is more than a year old, so newer versions of the software is already out there.**

1- You need to download the Virtual Machine from this link:
<http://web.cs.wpi.edu/~meltabakh/VM/Hadoop-VM-2015-V3.vdi>

2- To run the machine, download the VirtualBox tool. It is free and can be obtained from:
<https://www.virtualbox.org/wiki/Downloads>

3- When open VirtualBox → Click the “New” icon to start setting up the machine.
OS Type: Linux
Version: Ubuntu (64 bits)
memory: Recommended to have between 4 to 8 GBs (The higher the better)
Hard Drive: Use existing virtual hard drive <<and select the downloaded file>>

4- When the machine is setup and you start it, the OS username is “Hadoop”, password is “hadoop@12”, and root password is “root@12”

5- You should find a Readme file on the desktop with useful instructions on how to start.. Make sure to read this file.

6- If you have problems, we will go through the steps in class together

CS585/DS503: Downloading the Virtual Machine

Option 2: Build your own VM

- **Pros:** It is a good experience to do it yourself. You will have the most recent versions
- **Cons:** It may take you some time to be able to install things correctly and make it work

- 1- Download Ubuntu OS VM. It is available on several sites. Here is one:
<http://www.osboxes.org/ubuntu/>
- 2- To run the machine, download the VirtualBox tool. It is free and can be obtained from:
<https://www.virtualbox.org/wiki/Downloads>
- 3- When open VirtualBox → Click the “New” icon to start setting up the machine.
OS Type: Linux
Version: Ubuntu (64 bits)
memory: Recommended to have between 4 to 8 GBs (The higher the better)
Hard Drive: Use existing virtual hard drive <<and select the downloaded file>>
- 4- Start installing the needed software for the course, which includes:
 - Apache Hadoop (MapReduce)
 - Hadoop Streaming library
 - Apache Pig Latin
 - MongoDB
 - Apache Spark