

---

**Some organizations will not allow you to perform some of these tasks without elevated privileges, please budget time to work with your internal IT.**

There is much ground to cover during the tutorial and to maximize the time we are requesting that the following items be prepared prior to arriving at the course.

## “Hadoop in a Box” Virtual Machine (VM) Prep

To use the VM we require the following:

- Install the free Oracle VM VirtualBox
  - <https://www.virtualbox.org/wiki/Downloads>
  - Under the “VirtualBox Platform Package” section, select the option that is appropriate for you.
    - **VirtualBox 5.1.28 platform packages.** The binaries are released under the terms of the GPL version 2.
      - [Windows hosts](#)
      - [OS X hosts](#)
      - [Linux distributions](#)
      - [Solaris hosts](#)
  - After the installation, VirtualBox may need to install additional updates to work properly, please allow this to occur.
- Download test VM from:
  - <https://drive.google.com/file/d/0BxUmDImSfCRPSFRzcTZXeTRqUzg/view?usp=sharing>  
**This is not the VM we are using in the course, rather a small VM for you to get set up**
  - Download “ohdsi\_hw\_vm.zip”
  - Unzip the file to expose “ohdsi\_tutoria\_vm.vdi”
- Turn on Virtualization –
  - In order for the VM to work, virtualization needs to be enabled on your machine. The easiest way to do this is to Google “Enable Virtualization on X” and replace X with the model of your laptop. For example “Enable Virtualization on Lenovo T450” which brings up [these directions](#) to enable virtualization.

- Connect the test VM to Oracle VM VirtualBox
  - Open Oracle VM VirtualBox Manager

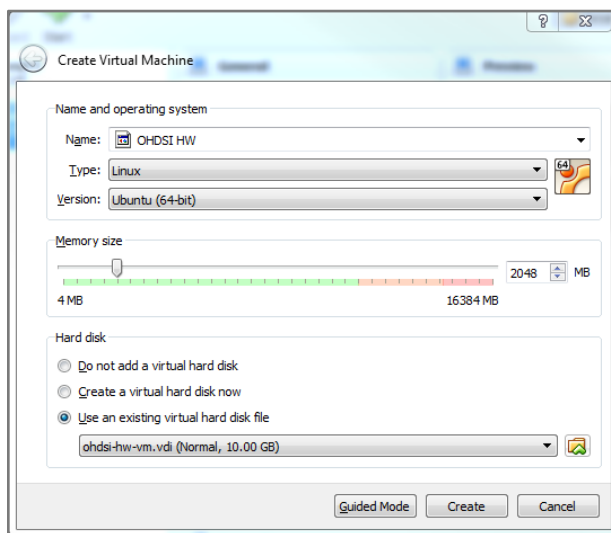


- Click on “new” in the manager window



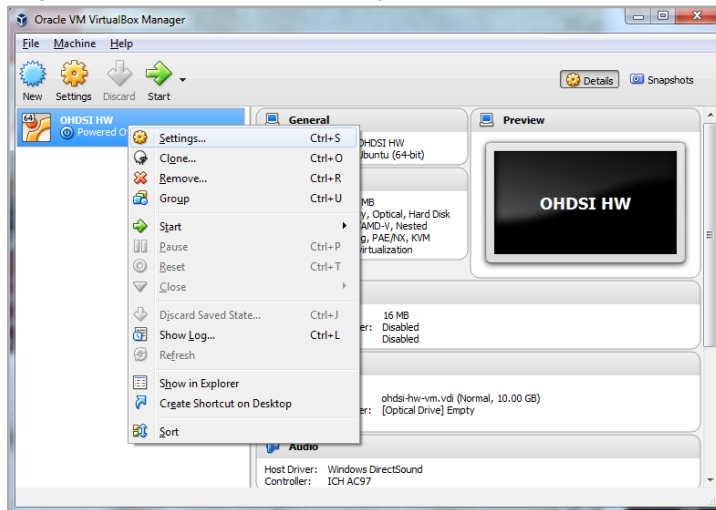
New

- In the Create Virtual Machine window:

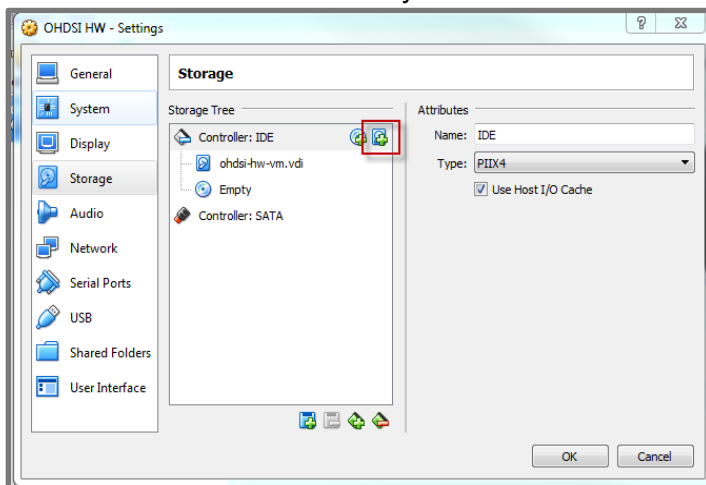


- Name: OHDSI HW
- Type: Linux
- Version:
  - Ubuntu (64-bit) (if you do not see this as an option you probably do not have virtualization set up properly, see step above)
  - If you do not have 64-bit machine use the Ubuntu (32-bit)
- Memory Size: 2048
- Hard Disk: select “Use an existing virtual hard disk file” and use the folder option to navigate to your “ohdsi\_tutoria\_vm.vdi”
- Hit the “Create” button

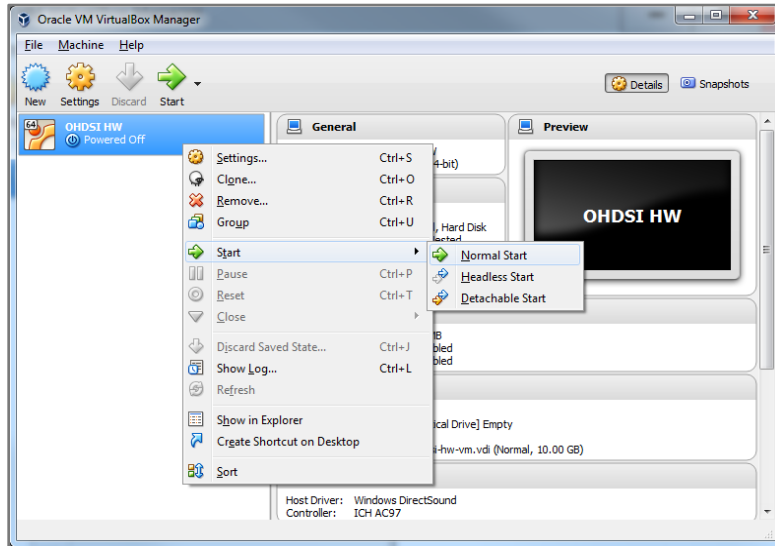
- Right-mouse button click on your new VM and select “Settings”



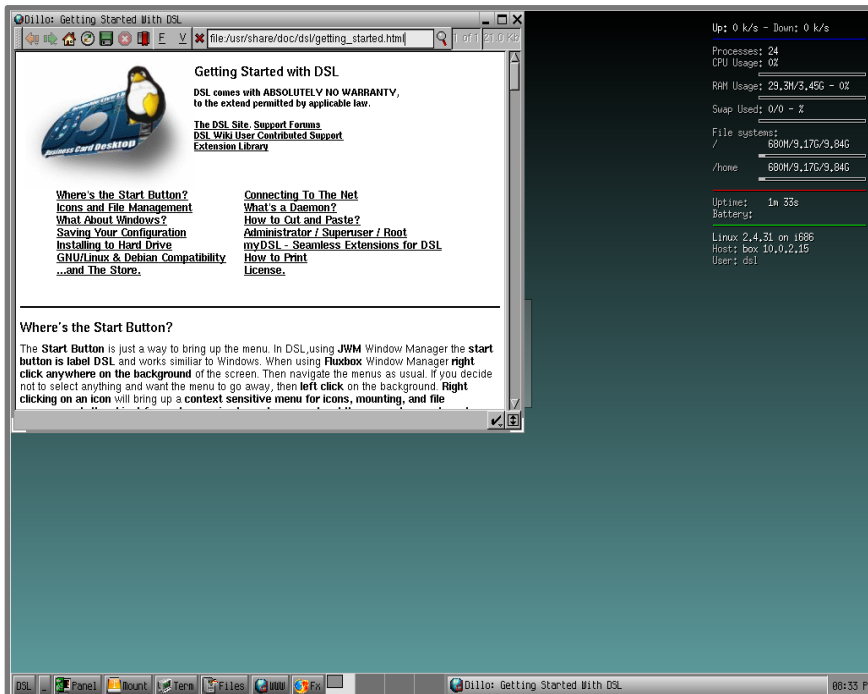
- If you see the “ohdsi\_tutoria\_vm.vdi” under the “Controller: STATA” – if you do, remove it by pressing the disk with a minus sign on it.
- You should only see the “ohdsi\_tutoria\_vm.vdi” under “Controller: IDE”. If you do not, hit the “Add Hard Disk” button and “Choose Existing Disk”. Select the “ohdsi-hw-vm.vdi” from where you saved it. Press the “OK” button.



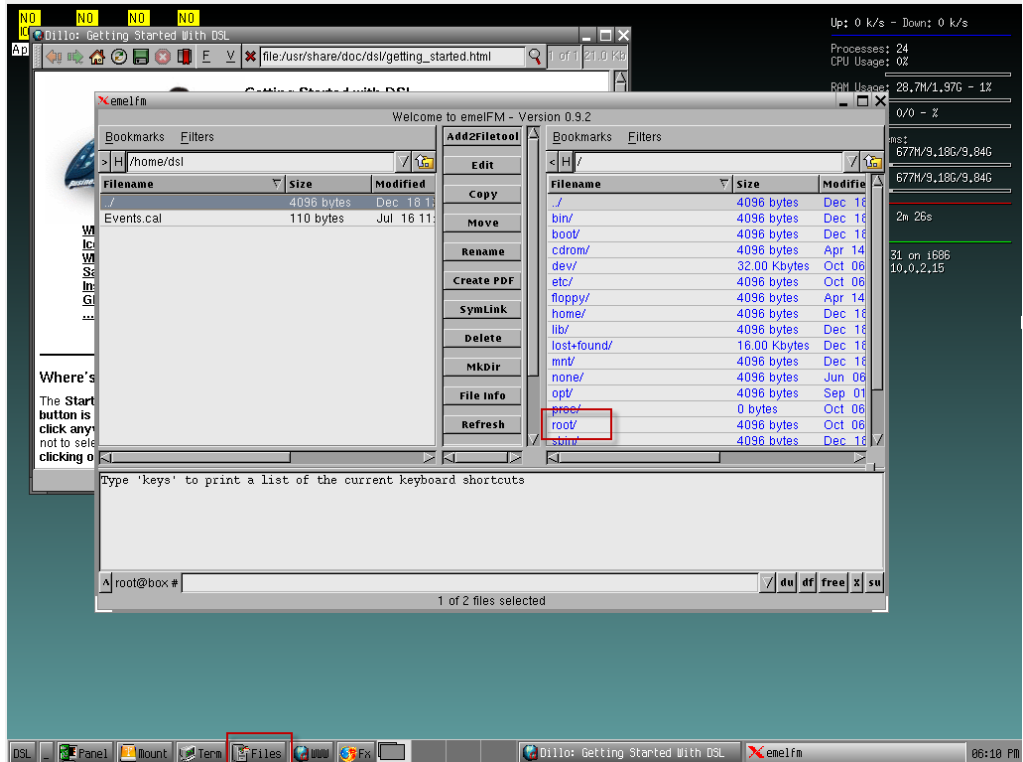
- On the main manager window, right-mouse button click on your new VM and select “Start → Normal Start”



- On the boot up screen, select the default settings and allow it to boot up.
- Login to the box:
  - **Box Login:** root
  - **Password:** toor
- A DSL Linux box will kick off.



- To obtain the secret password:
  - In the bottom left side, click the “Files” button.
  - In the Files screen, navigate to root



- Open the file “secret.txt”
  - Email [evoss3@its.nj.com](mailto:evoss3@its.nj.com) the text inside the file.
- RAM 8GB or more on the laptop you are trying to use to run the VM.
  - While the VM will run on a 4GB RAM box, it is more performant on 8GB or higher

Reminder, this set up is not for the VM we are using the day of, it is to ensure you have your machine set up to run the VM we will be using.

- SynPUF Documentation:
  - For your reference we will be using the Medicare Claims Synthetic Public Use Files (SynPUFs) as our exemplar dataset.
    - For information about the dataset:  
<https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/index.html>
    - CDM SynPUF ETL documentation:  
[https://github.com/OHDSI/ETL-CMS/blob/master/RabbitInAHat/CDMv5\\_CMS\\_MDCR\\_ETL.docx](https://github.com/OHDSI/ETL-CMS/blob/master/RabbitInAHat/CDMv5_CMS_MDCR_ETL.docx)