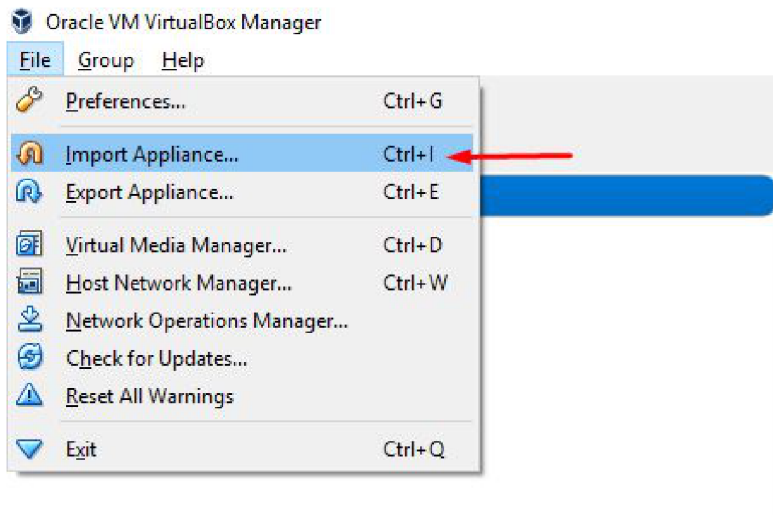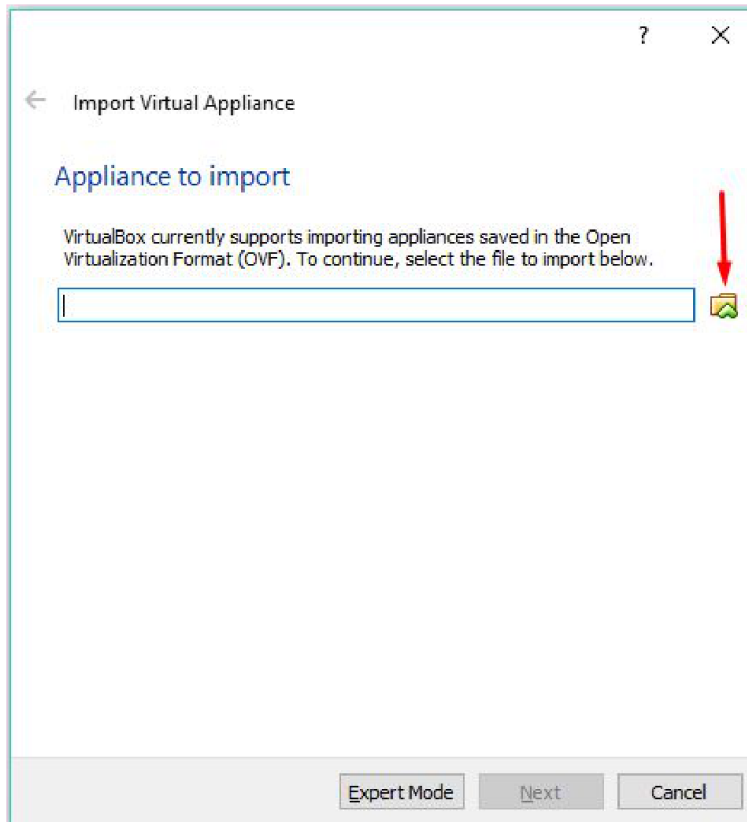## Pyspark VM Installation Document

1. **Important Note** - If you have already done my Big Data training (20th - 28th April 2021), then you don't need to do this. The VM for Big Data training should work for this training as well.
2. Make sure that you have enabled Virtualization in your machine as per this link before proceeding to the next step.
3. Download and Install Oracle Virtual Box from this link.
4. Download the VM from this link and save it on your system. (Big file, about 6 GB so it will take time to download it. Make sure that you have a good internet speed before you do this.)
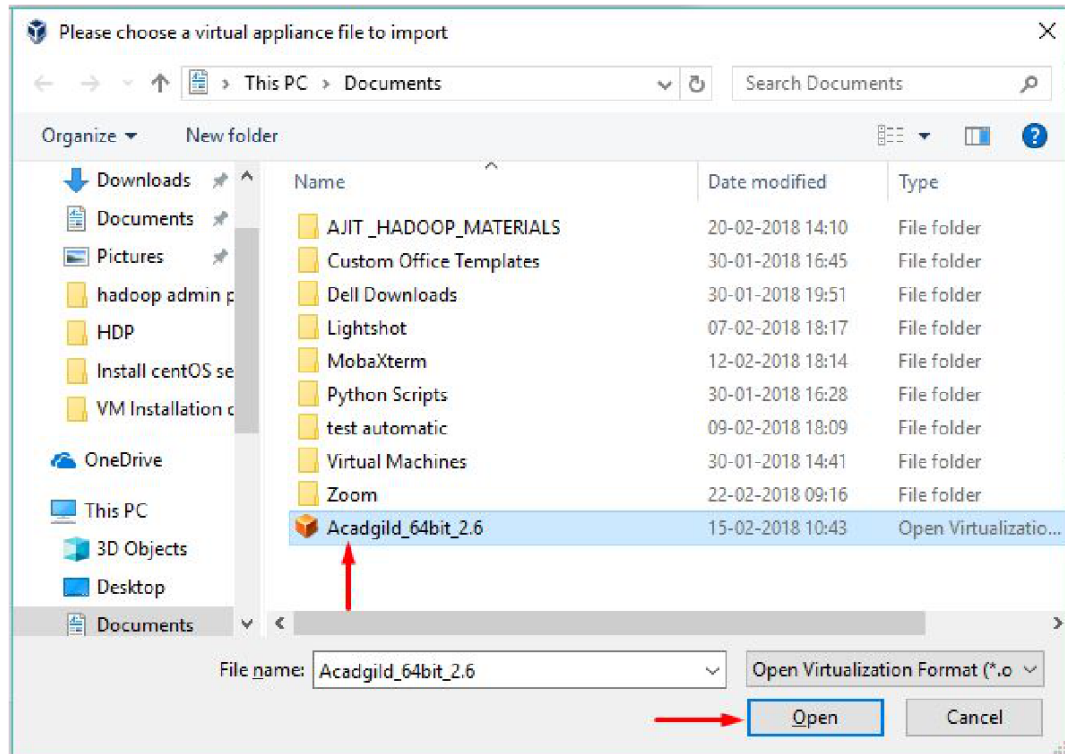5. Open the virtual box, Click on File and then select Import Appliance as below :-

6. Click on the Folder option as shown below and browse to the location where you have saved the VM in step 4 above
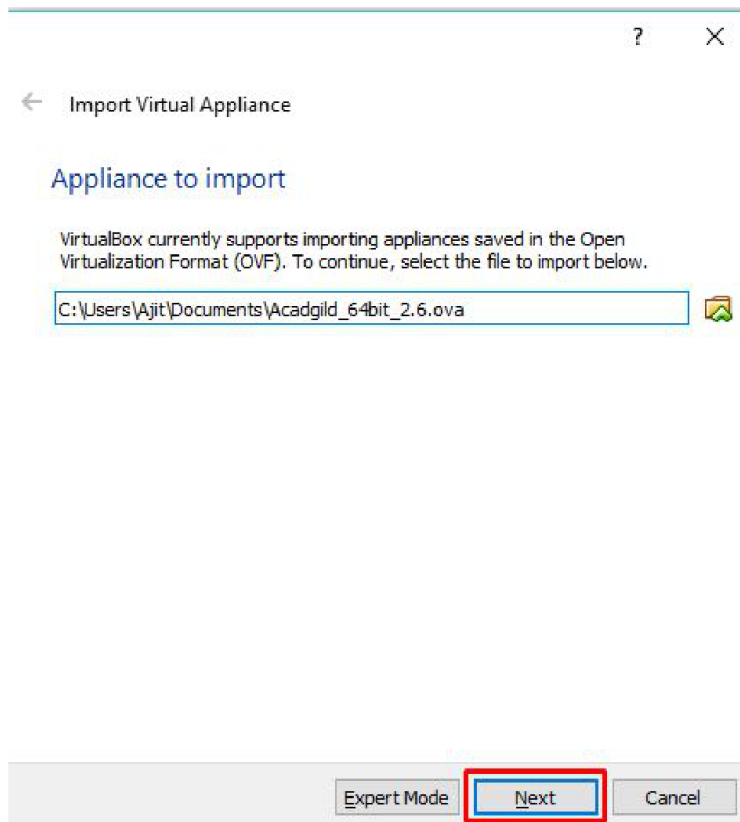
7.

Select the VM to be imported and then click on Open as shown below (note that your VM name will be MultiVM.ova)
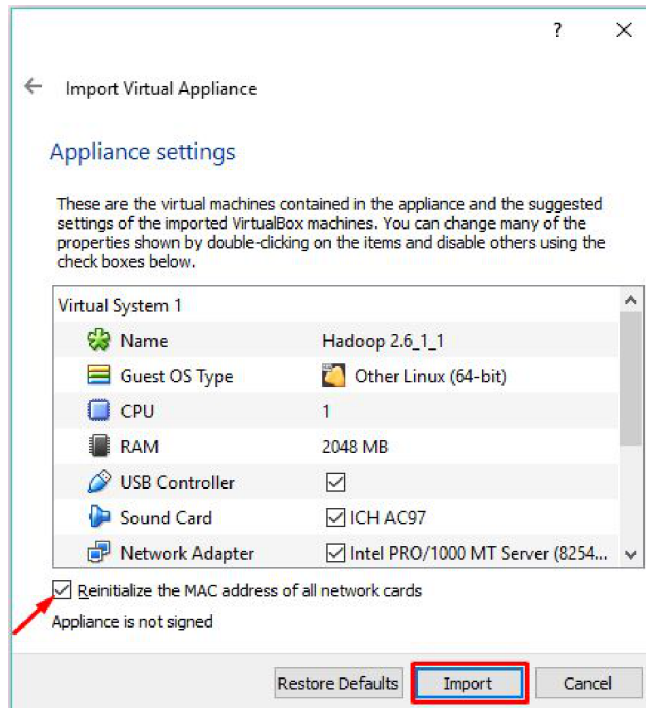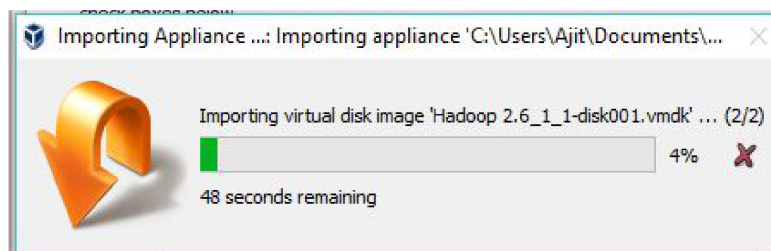
Click on Next to proceed further
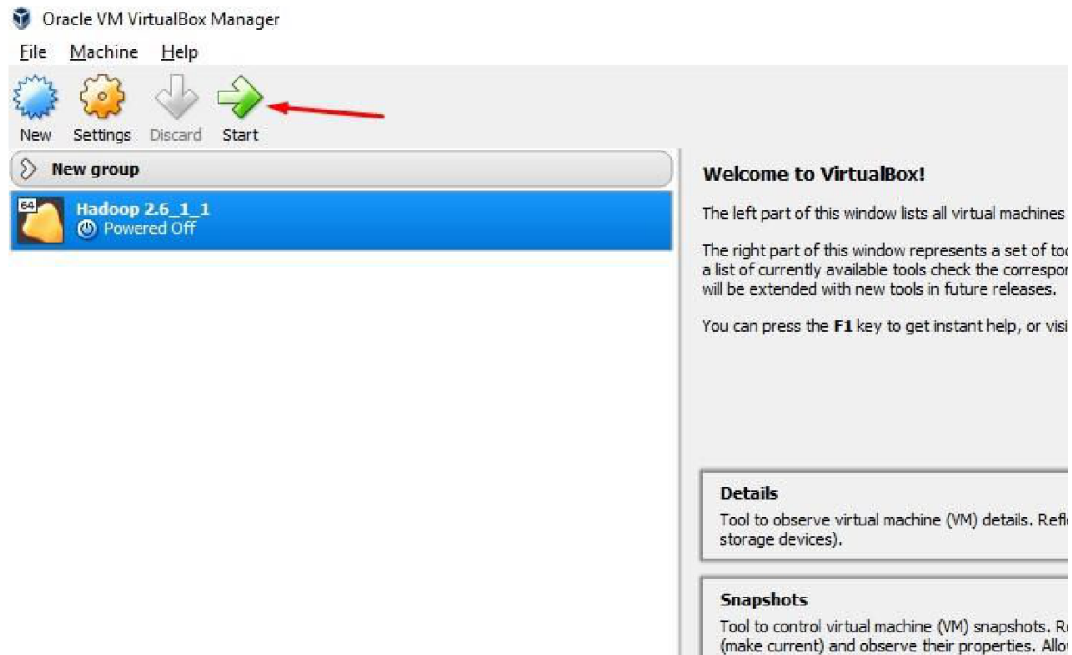
Click on Import to initiate importing the MultiVM.

Note : Every system has their unique identity, thust to avoid network conflicts during VM

import,The MAC address must be reinitialized. Make sure the box is ticked as shown below

Once the importing is initiated, there's a prompt with following message.

Once the import is complete, click on start as shown in the below screenshot:

If it asked for the password, enter it as "Citic@rp1234 Note :The first "C" in the password is in capital.