CARLSON SCHOOL
OF MANAGEMENT
UNIVERSITY OF MINNESOTA

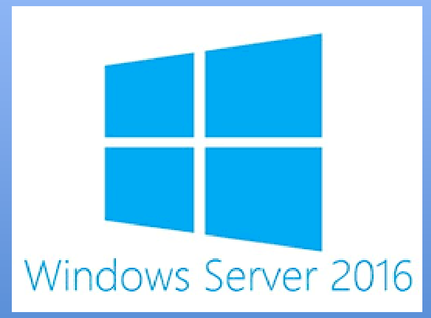
Our Computing Environment

Professor De Liu

MSBA 6330

Computing Architecture

VD - MSBA Virtual Desktop (Windows Server 2016)

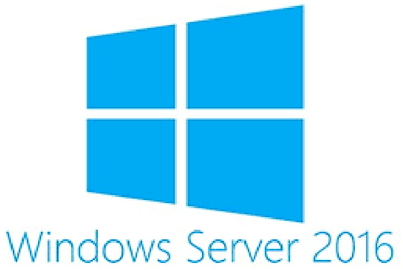


VM – Cloudera Virtual Machine (CentOS 6)

Hadoop Pseudo Cluster – HDFS, Sqoop, Pig, Hive, Spark, HBase, etc

MSBA Virtual Desktop (VD)

- Current configuration – 10GB RAM, 80GB Disk space, with access to the outside Internet
- Requires a **Citrix Client** to connect
  - Occasionally, after idling the VD for a while, the Citrix client may become irresponsive. You'll need to disconnect and reconnect.
- You don't have system admin privilege on VD.
  - If you have suggestions for installing something on this VD, pls email the instructor.
- Files stored in the Documents folder are backed up
  - in case the VD fails and needs to be rebuilt.
  - Save your files in the Documents folder
- Safe to reboot the VD
  - but normally, you probably just want to disconnect.



VM – Cloudera Virtual Machine

- Based on Cloudera QuickStarter VM, customized for the courseware.
- Preinstalled CDH 5.10.x (Cloudera Hadoop distribution).
- By default configured to have 4GB RAM, 1 CPU, and 64GB HDD
  - actual HDD size is much smaller
- It is only visible to the VD (cannot access it outside of VD)
- Its Internet access is limited to the campus network.
- It shares the vagrant folder with VD, which can be used to share files.

`c:/vagrant` ⬅➡ `/vagrant`

VD

VM

- Forwarded ports to VD: 8888 (Hue), 8889 (PySpark), 18088 (Spark History)

How VD interacts with VM

- From VD (using `git bash`)
  - Brings up the VM: `vagrant up` (from `c:/vagrant`)
  - Powers down the VM: `vagrant halt` (from `c:/vagrant`)
  - SSH into the VM (one of the two ways):
    - `vagrant ssh` (from `c:/vagrant`)
    - `ssh cloudera@localhost -p 2222`
  - Play the VM using virtualbox for GUI access into the VM
    - You should not start or terminate the VM using virtual box. If you do that you risk damage the VM. Always uses `vagrant`.
  - the `c:/vagrant` folder can be used to share files between VM and VD.
  - Using browser to access Hue (http://localhost:8888) and PySpark (http://localhost:8889) and Spark History (http://localhost:108088) on VM.
    - Jupyter on VD also uses port 8888. You should shut down Jupyter to use Hue.

If you're inside the VM (via ssh or GUI)

- VM still has access to campus network, which means it has access to:
  - canvas.umn.edu: download and upload stuff
  - gmail.com: send and receive stuff
  - github.umn.edu: you can see or clone repositories on github.umn
  - MySQL servers: e.g. you can download data from csom-idsdl2.oit.umn.edu
  - UMN web servers or ftp servers: files shared via csom-idsdl.oit.umn.edu
- From the command line
  - Clone a repo: `git clone https://your_x500@github.umn.edu/....`
  - Download a file on a web server: `wget umn_url`
  - Upload files to a ftp server (if you have access to one): ftp
  - Share files with VD via the `/vagrant` folder