LAB 1 - WEEK 1 AND 2

SETTING UP THE DEV ENVIRONMENT - VM

Setup Cloudera Quickstart VM Lab Environment on your machine

1. Machine Pre-requisites

- A machine with 64 bit OS and more than 8 GB RAM* (Total of 16GB+ RAM is preferred)
- At least 30-40 GB free space on hard disk
- Virtualization (VTx) parameter should be enabled in the BIOS settings of your machine otherwise the VM will not start!

2. VM Software Installation (Getting VMWare is strongly recommended)

In the course lecture videos, Cloudera QuickStart VM is used for demonstrating labs. As this is a VM (download link for which is given below in section 3), you need a VM software using which you can run this Cloudera QuickStart VM.

There are many different VM software available depending upon your host operating system. For example, if you've a Windows machine then you can get VMWare Player and there's VMWare Fusion for Mac machines. Virtual box will work for all the major OSs.

Following is a list of 2 such VM tools. You only need to get any one of them. I highly recommend the first one as all the lecture videos are based on VMWare.

1) *VMWare*
   - *Workstation/Player for Windows/Linux*
   - *Fusion for Mac*

   We'll just use the free version.

2) *Virtual box*

3. Download & Install Cloudera QuickStart VM

- *Download the copy of Cloudera QuickStart VM from here that is based on the virtual machine software you installed in the previous step.*
- *Unzip it (preferably with 7-zip) and start Cloudera QuickStart VM*
  - *Start the .vmx file for WorkStation/Player OR*
  - *Start the .ovf file for VirtualBox*

Take help from this video if you face any issues with installing Cloudera QuickStart VM on latest Apple MacBook machines which have M chips instead of Intel .

While working in the VM environment, it's best to give at least 8GB RAM and 2 CPU cores to the VM and leave rest for the host OS. Do not use Cloudera Manager that's available in the VM as it's memory needs are quite intensive.

When the QuickStart VM boots, it configures all the services you might expect on a Cloudera cluster. Obviously, this single-node "pseudo-distributed" simple setup does not represent the performance, scalability, and reliability of a fully-distributed cluster – but it does give you a taste of how easy it is to perform powerful work with your data. The core services that we'll mostly need for the course are already running on this VM.

If you do not plan to use a VM or Docker then you'll need to install Hadoop and other tools on your own. There may be compatibility issues with different versions of different tools, so refer this page to see which version of which tool works well together.

Homework Questions

Questions

Answer and submit the following questions once the above setup is done.
Numbers in square brackets indicate the points for each question.

1) [2] **In the newly set up VM, what does the $HOME environment variable represent?**
2) [2] **What does sudo mean in a Unix/Linux environment, and when is it typically used?**
3) [6] **Write and briefly explain any 6 Unix/Linux commands.**
   Execute each command in the terminal of the new VM and paste **screenshots of the outputs**.
   You may choose from the following commands:
   cd, touch, mkdir, rm, clear, ls, pwd, mv, cat, tail, echo

These questions are intended to help you become familiar with basic Unix/Linux concepts, as Linux is the primary operating system used throughout this course.