Clairvoyant

Cloudera Quickstart VM Setup

General Notes

The training uses the VM CDH5.X (Cloudera's Distribution, including Apache Hadoop 2.X) installed in Pseudo-Distributed mode. Pseudo-Distributed mode is a method of running Hadoop whereby all Hadoop Daemons run on the same machine. It is a cluster consisting of a single machine. It works just like a larger Hadoop cluster, except the block replication factor is set to 1, since there is only a single DataNode available.

Instructions to Install the VM

1. If you're using a Windows machine, ensure that Virtualization is enabled in your BIOS

2. Download VirtualBox from:
   https://www.virtualbox.org/wiki/Downloads

   a. Pick the appropriate binary for your operating system.

   b. Then follow the prompts for installing VirtualBox.

3. Once the installation is complete, download the CDH 5.X VM from the URL: http://www.cloudera.com/content/cloudera/en/documentation/core/latest/topics/cloudera_quickstart_vm.html

   a. Page down until you find the Downloading the Quickstart VM section and click the link in that section

   b. Select VirtualBox under the Platform Dropdown

   c. Click "Get it Now"

   You may need to download 7-zip from www.7-zip.org to extract the Cloudera VM. After uncompressing, you will get a file called "cloudera-quickstart-vm-x.x.x.x-virtualbox.ovf." Move it to location of your choice in File System.

4. Once you have extracted the VM, we will load the VM into VirtualBox.

   a. Open Virtual Box and click on **File -> Import Appliance...**

   b. From the file dialog open **cloudera-quickstart-vm-x.x.x-x-virtualbox.ovf** located in the decompressed (or unzipped) Cloudera VM download.

5. Add Port Forwarding to allow SSH and SCP from your Local Machine to the VM (note: this can be done while the machine is running)

    a. Go to the VirtualBox Manager and Select the VM from the list on the left

    b. Click Settings

    c. Click the "Network" tab on the window the pops up

    d. Expand "Advanced"

    e. Click on the "Port Forwarding" button

    f. Click the "+" button and add a port forwarding section with this information:

        i. Name: 2222, Protocol: TCP, Host Port: 2222, Guest Port: 22

    g. Click OK

6. Enable Bidirectional Copy

    a. Go to the VirtualBox Manager and Select the VM from the list on the left

    b. Click Settings

    c. Under "General", select the "Advanced" Tab

    d. Set "Shared Clipboard" to Bidirectional

7. Start the VM.

8. Once the VM starts up, you should see the Desktop within VirtualBox. This is your sandbox to play with Hadoop.

9. Open Terminal and run the below command. If the setup is successful, it will print the current version.

```
$ hadoop version
```

10. Download the "com.clairvoyant.workshop.zip" from the Setup: Exercise Data Google Drive.

a. You can do this from the VM by clicking on the icon at the top left of the VMs Desktop and navigating to the below link

https://drive.google.com/open?id=1TljwnmecqiavU39nymOlGSX56QnKU7c6

11. Open a Terminal and navigate to where you downloaded the zip file to

a. To start a Terminal go to the VM and click the icon at the top of the interface

12. Unzip the file

```
$ unzip com.clairvoyantsoft.workshop.zip
```

13. Navigate to the parent directory

```
$ cd com.clairvoyantsoft.workshop
```

14. Run Maven Clean Compile

```
$ mvn clean compile
```

Points to note while working in the VM

1. The VM is set to automatically log in as the user cloudera. Should you logout at any time, log back in as the user cloudera with the password cloudera. Root password is cloudera.

Points to note during the exercises

1. There are additional challenges for most of the Hands-On Exercises. If you finish the main exercise, please attempt the additional exercise.

2. If you are working in eclipse, you can load the project by clicking on **File → Import → Existing Maven Projects → Next.**   From the dialog select the **Root directory** and the project that was downloaded and extracted.

How To's

<u>Copy Files from your Local Machine to the VM</u>
1. Open a terminal window on your Local Machine

2. Navigate to the directory that contains the files you want to copy to your VM

3. Execute the SCP command (the file will be copied to the cloudera users home directory):

```
$ scp -P 2222 {local_file_name} cloudera@localhost:
```

<u>Login as Root on the VM</u>

1. Open a terminal window on the VM

2. Login as Root:

```
$ sudo su
```

Additional Documentation on the Quickstart VM

● http://www.cloudera.com/content/cloudera/en/documentation/core/latest/topics/cloudera_quickstart_vm.html