Install CM 5.15

You have been provide with 5 nodes that have been prequalified to install CM 5.15 and CDH 5.15. You will still need to complete the following:

1. Enable user / password login for each of the 5 nodes
    a. Create a password for user "centos"
    b. Modify sshd_config to allow password login
    c. Restart the sshd.service
2. Setup /etc/hosts with the following information for each of the 5 hosts
    a. Private_IP   FQDN Shortcut
3. Change the hostname as necessary to the FQDN that you setup above
    a. Reboot the host
4. Install JDK on each of the hosts (you may choose to install on just the host where you will install CM and use the Wizard later to install on the rest
5. On the host that you will install CM:
    a. Configure the repository for CM 5.15.2
    b. Install CM
    c. Install and enable Maria DB (or a DB of your choice)
        i. Don't forget to secure your DB installation
    d. Install the mysql connector or mariadb connector
    e. Create the necessary users and dbs in your database
        i. Grant them the necessary rights
    f. Setup the CM database
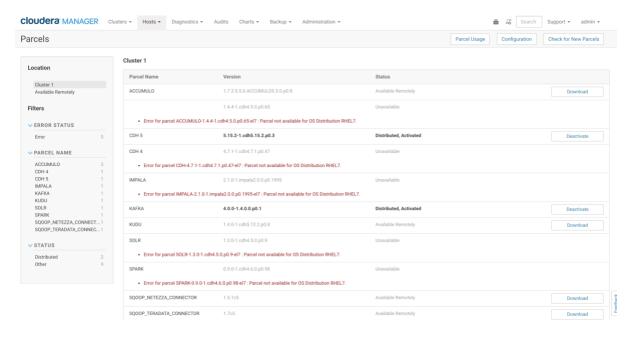   g. Start the CM server and prepare to install the cluster through the CM GUI installation process

Install CDH 5.15.2 Cluster

Go to the CM GUI installer and install CDH 5.15.2

1. Specify hosts for your CDH cluster installation
2. You may choose to use a common user / password
3. Allow CM to install CM Agent on each of the nodes
4. Allow CM to download, distribute and activate CDH on each of the nodes
5. Choose the type of installation you want
    a. We will require Flume, Hive, Impala later on
6. Assign roles to each of the hosts: Please refer here for hints.

Install Sqoop, Spark and Kafka

Installing Kafka requires some special attention. You will need to download, distribute and activate the Kafka package before you can add the Kafka service.

- Create user "training" with password "training" and add to group wheel for sudo access.
- Log into Hue as user "training" with password "training"
  - This will create a HDFS directory for the user
- Now, try ssh'ing into the hosts as user "training"
- From the bit.ly folder, download all.zip and unzip it.
  - Do this in both your CM host and one of the datanode hosts.
  - Go to training_material/devsh/scripts and review the setup.sh
    - See how it works. We are going to use it to create some data for ourselves but it won't work right away. See if you can figure out what needs to be done to make it work.
    - You will need to add user "training" with password "training" to your MySql installation and Grant the necessary rights
- Now, attempt some of the exercises that we did during our previous lessons in our new Hadoop cluster.
  - You should practice the following as a minimum
    - Flume
    - Sqoop
    - Hive
    - Impala
    - Spark