Cloudwick Professional services

for Brattle Group

Table of Contents

Contents

[Introduction - 3 -](#_Toc480282413)

[1. Upgrading Cloudera Manager from CM 5.6.0 to 5.10.1 - 3 -](#_Toc480282414)

[1.1 Required information before performing the upgrade - 3 -](#_Toc480282415)

[2. Upgrading CDH from 5.6.0 to 5.10.1 - 5 -](#_Toc480282416)

[3. Installing Spark2 (beta version) - 5 -](#_Toc480282417)

[4. Adding Anaconda parcel using Cloudera Manager - 5 -](#_Toc480282418)

[5. Adding Edge nodes using Cloudera Manager - 6 -](#_Toc480282419)

[6. Installing SSSD service on New nodes - 6 -](#_Toc480282420)

[7. Integrating RStudio server with Spark - 7 -](#_Toc480282421)

# 

# Introduction

This document describes the activities/tasks performed for the Brattle Group

which involves the upgrading of cluster from 5.6.0 to 5.10.1.

## Upgrading Cloudera Manager from CM 5.6.0 to 5.10.1

## 1.1 Required information before performing the upgrade

* Find out the current version of your Cloudera Manager and Jdk deployed. Go to Support > About
* The version of CDH can be found on the Home page of Cloudera Manager.
* Take a note of your Operating system type and Version.
* Review the [CDH 5 and Cloudera Manager 5 Requirements and Supported Versions](https://www.cloudera.com/documentation/enterprise/latest/topics/rn_consolidated_pcm.html#xd_583c10bfdbd326ba--5a52cca-1476e7473cd--7f8d) for the new versions you are upgrading to

<https://www.cloudera.com/documentation/enterprise/latest/topics/rn_consolidated_pcm.html#xd_583c10bfdbd326ba--5a52cca-1476e7473cd--7f8d>

* Read the [Cloudera Manager 5 Release Notes](http://www.cloudera.com/documentation/enterprise/release-notes/topics/rg_release_notes_cm.html" \t "_blank).
* Back up Cloudera Manager Databases
  + Stop the Cloudera Management Service
  + Select Clusters -> Cloudera Management Service
  + Select Actions -> Stop
* To locate the information of Databases for Cloudera manager server go to /etc/cloudera-scm-server/db.properties file.
* For remaining services Databases information go to Clusters > Cloudera Management Service > Configuration and select the Database category.
* Obtain the required passwords from your Database Administrator
* Customized Cloudera Manager scripts backup can be found in
  + /usr/share/cmf/bin/import\_credentials.sh
  + /usr/share/cmf/bin/gen\_credentials\_ad.sh

**Note:** The Scripts should be added back to the same location before starting the

Cloudera Manager scm server

* Stop Cloudera Manager Server and Agent on the host that is running that is running the above services
  + Service cloudera-scm-server stop
  + Service cloudera-scm-agent stop
* Back up the following directories
  + /etc/cloudera-scm-server
  + /etc/cloudera-scm-agent
* Commands for backing up the Databases:
  + mysqldump -hcmbhdpnn002.brattle.net -uamon -p amon > /root/backup/database-backup.amon
  + mysqldump -hcmbhdpnn002.brattle.net -urman -p rman > /root/backup/database-backup.rman
  + mysqldump -hcmbhdpnn002.brattle.net -unavms -p navms > /root/backup/databackup.navms
  + mysqldump -hcmbhdpnn002.brattle.net -unav -p nav > /root/backup/databackup.nav
* Back up the current repo from
  + /etc/yum.repos.d
* Download the Cloudera Manager repo from <https://archive.cloudera.com/cm5/redhat/6/x86_64/cm/>.
* Copy the repo file to /etc/yum.repos.d/
* Perform the following commands for the upgrade
  + Yum clean all
  + Yum upgrade cloudera-manager-server cloudera-manager-daemons cloudera-manager-agent
* Make sure the above Packages are upgraded
* Copy the Customized Scripts (import\_credentials.sh, gen\_credentials\_ad.sh) to the following location
  + /usr/share/cmf/bin/
* Start the Cloudera Manager server
  + Service cloudera-manager-server start
  + For Agent Upgrade
* Stop all the Agents using
  + Service cloudera-scm-agent stop
* Copy the repo file from the CM server to all the agents (/etc/yum.repos.d)
* Perform the following commands for the Agents upgrade on all the hosts
  + Yum clean all
  + Yum upgrade cloudera-manager-daemons cloudera-manager-agent
  + Service cloudera-scm-agent start
* Login to the Cloudera Manager console
* Skip the Upgrade Agents using Parcels since you already upgraded them using the packages
* The Host inspector verifies the current Agent verisons with CM version and everything is matched then it prompts to Continue and Finish.
* To verify the upgrade go to Hosts tab and click inspect all hosts and Verify the last heartbeat for each host has occurred within one minute.

## Upgrading CDH from 5.6.0 to 5.10.1

* Go to CM UI and stop all the services
* For performing the CDH upgrade you will need to take the following backups
* mysqldump -hcmbhdpnn002.brattle.net -uhive -p metastore > /root/backup/metasotre-backup.sql -hive
* mysqldump -hcmbhdpnn002.brattle.net -uhue -pLiverp00l2015 hue > /root/backup/hue-backup.sql -hue
* oozie server uses derby database go to /var/lib/oozie/data/seg0 whole directory -oozie
* mysqldump -hcmbhdpnn002.brattle.net -usentry -p sentry > /root/backup/sentry-backup.sql -sentry
* namenode directory backup & Journal nodes directories
* krb5.conf, sssd.conf, nscd.conf
* On Home tab you will find Upgrade cluster and then the wizard will prompt to perform steps accordingly and on the Home tab you will be able to see the new CDH parcels as 5.10.1
* Smoke test the cluster by running some sample Map reduce jobs, Hive queries and Spark jobs to see if all the services are running as expected

## Installing Spark2 (beta version)

* Download the Package from <http://archive.cloudera.com/beta/spark2/csd/>
* On the Cloudera Manager host
* Make the required changes on the file to Cloudera-scm:Cloudera-scm and 644
* Restart the Cloudera-scm-server and login to the Web Ui and select Clusters-> Cloudera Management Service and then Select Actions -> Restart
* Now on the Parcels page you will be able to see Spark2.0 to download -> distribute -> Activate
* It can be added as a Service and you will be able to see the Spark2 on the stack along with all the services on the Cloudera Manager

<https://blog.cloudera.com/blog/2016/09/apache-spark-2-0-beta-now-available-for-cdh/>

## Adding Anaconda parcel using Cloudera Manager

* Get the Remote url and paste it on the Parcles -> configurations page and save it
* You will be able to see the Anaconda Parcel and then Download -> Distribute -> Activate.

<https://docs.continuum.io/anaconda/cloudera>

## Adding Edge nodes using Cloudera Manager

* On the Cloudera Manager tab click to Add hosts and then enter the IP addresses of the hosts you want to add to the Cluster
* CM will prompt to install required Oracle JDK and Cloudera Manager Agent Packages and start the Hosts.
* Since Cloudera Manager manages the krb5.conf (kerberos) for that you will need to install packages like workstation and libs on the hosts for CM to deploy the Krb5.conf file automatically
* Then perform Kinit command and verify if you have got the ticket using command klist

<https://www.cloudera.com/documentation/enterprise/5-6-x/topics/cm_mc_adding_hosts.html#cmug_topic_7_5_1__section_ny1_bxv_ls>

## Installing SSSD service on New nodes

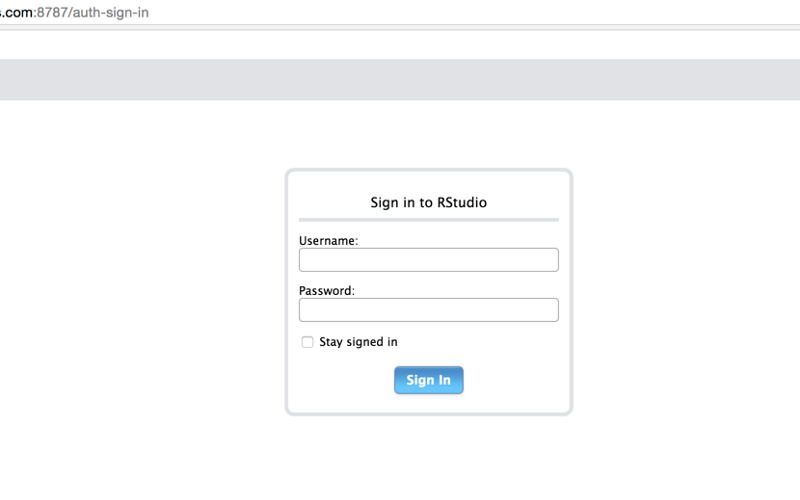
* yum install -y samba4-common pam\_krb5 sssd adcli authconfig
* ntpdate -u cmbdc1.brattle.net
* chkconfig ntpd on
* systemctl enable ntpd.service
* rpm -qa |grep authconfig
* authconfig --enablesssd --enablesssdauth --enablelocauthorize --enableldap --enableldapauth --ldapserver=ldap://cmbdc1.brattle.net --disableldaptls --ldapbasedn=dc=brattle,dc=net --enablemkhomedir --enablecachecreds –update
* copy both Krb5.keytab and sssd.conf file from Cloudera Manager server
* Place krb5.keytab in /etc/ and sssd.conf in /etc/sssd/
* Service sssd restart
* Make sure any one of the user in AD is able to SSH into the newly added nodes and similarly do the same from new nodes to the existing nodes.
* Then do id username to see if you are able to get the output like this

uid=10008(drobinson)gid=10003(prod\_ssh\_login\_edge) groups=10003(prod\_ssh\_login\_edge),10004(hadoop\_admins),10005(hue\_admins)

## Integrating RStudio server with Spark

* yum install epel-release
* yum install R
* Next we move to install of the server IDE, rstudio-server. We will download it from the site
* wget <https://download2.rstudio.org/rstudio-server-rhel-1.0.44-x86_64.rpm>
* yum install --nogpgcheck rstudio-server-rhel-1.0.44-x86\_64.rpm
* systemctl status rstudio-server.service
* Go to Web UI and <http://youriporhosname:8787/>
* You log in with your non-root Linux user
* Since we are using AD for user Authentication there are no users on the host that RStudio server will be able to authenticate, it uses PAM authentication and for that some customization should be done by moving the login file to rstudio
* cp /etc/pam.d/login /etc/pam.d/rstudio
* To open a session you need to connect using Kerberos principal, each user must have their own keytab created to open a session
* Steps for creating a keytab open ktutil:
  + - ktutil: addent -password -p bkvarda@CLOUDERA.INTERNAL -k 1 -e rc4-hmac
    - Password for bkvarda@CLOUDERA.INTERNAL: [enter your password]
    - ktutil: addent -password -p bkvarda@CLOUDERA.INTERNAL -k 1 -e aes256-cts
    - Password for bkvarda@CLOUDERA.INTERNAL: [enter your password]
    - ktutil:wkt bkvarda.keytab

ktutil: quit

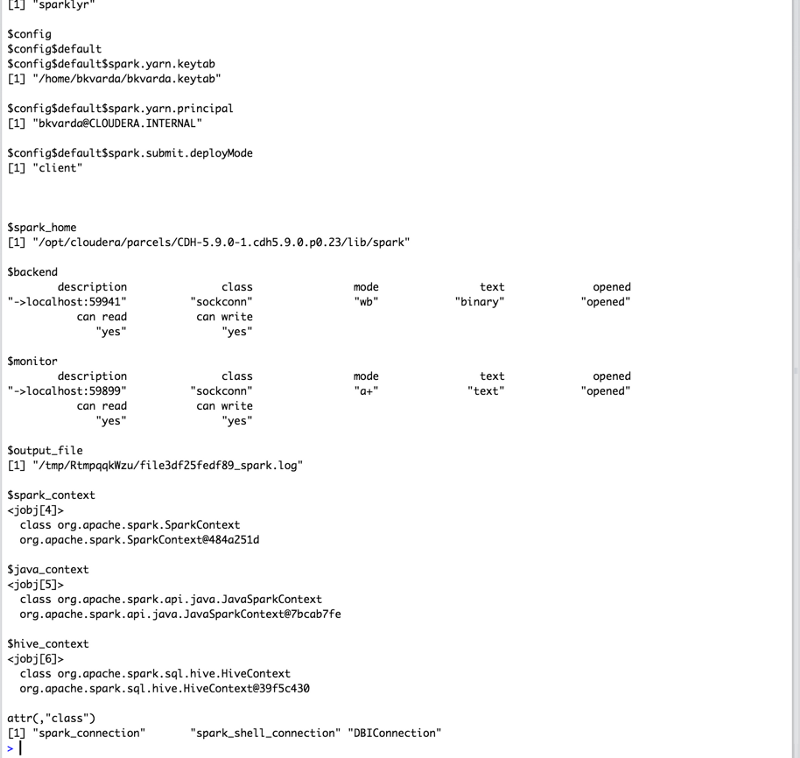
* Using a username and password you will be able to login on the Web UI page
* 
* Try opening up a terminal from within the browser (tools -> shell?) and run kinit and enter the password then close the shell, then enter the following steps one by one
* library(sparklyr)
* readRenviron("/usr/lib64/R/etc/Renviron")
* sc <- spark\_connect(master = "yarn-client",version = "1.6.0", config = list(default = list(spark.yarn.keytab = "/home/drobinson/drobinson.keytab", spark.yarn.principal = "drobinson@BRATTLE.NET"))

Note: You must use your own Principal and keytab in the above command

* If this is successful then is should work pretty quickly, otherwise there is something wrong with Kerberos ticket, assuming everything went well type

Sc

* You must be able to see some output like this:

’

https://medium.com/@bkvarda/sparklyr-r-interface-for-spark-and-kerberos-on-cloudera-80abf5f6b4ad

<http://docs.rstudio.com/ide/server-pro/authenticating-users.html>

## Installing Python 2.6.6 and making it as a default version

* + wget [**https://www.python.org/ftp/python/2.6.6/Python-2.6.6.tgz**](https://www.python.org/ftp/python/2.6.6/Python-2.6.6.tgz)  
    --> unzip/untar this package
* cd Python-2.6.6 and do   
  ./configure && make altinstall
* Repoint the /usr/bin/python2 to point to /usr/local/bin/python2.6.6  
  ln -s /usr/local/bin/python2.6 /usr/bin/python2