November 10, 2017

Vinith SiNGEDI & Jagdish Saripella

Zaloni

Clouder Manager & CDH Installation

version 1.0

Table of Contents

[Document Revision: 3](#_Toc498076829)

[Cloudera Manager & CDH Installation 4](#_Toc498076830)

[General Pre-Requisites: 4](#_Toc498076831)

[Before Installaing Cloudera Manager : 4](#_Toc498076832)

[Install and configure Mysql Server 4](#_Toc498076833)

[Configure and Tune Mysql 4](#_Toc498076834)

[Configure Cloudera Manager to use Mysql datastore 6](#_Toc498076835)

[Establish Cloudera Manager Repository Stratergy 7](#_Toc498076836)

[Make the rpms available through webserver: 7](#_Toc498076837)

[Install Cloudera Manager 7](#_Toc498076838)

[Prepare Cloudera Manager server to use external Database: 7](#_Toc498076839)

[Start Cloudera Manager server: 7](#_Toc498076840)

[On cloudera host execute the below 7](#_Toc498076841)

[Login to Cloudera Manager : 7](#_Toc498076842)

# Document Revision:

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Revision Number | Comment | Author |
| 10-Nov-2017 | V1.0 | Initial Document | Jagdish Saripella |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Cloudera Manager & CDH Installation

Cloudera Manager is an administrative tool by Cloudera that eases the installation and management of CDH ( Cloudera’s Hadoop Distribution ). Below literature will cover the necessary pre-requistes needed to install Cloudera Manager and CDH

General Pre-Requisites:

Below are some general pre-requisites before installing cloudera manager on the host . This pre-requiste stand same for all the hosts participating in Hadoop cluster.

* Disable Selinux
* Setup NTP
* Disable firewall
* Define host names

Cloudera Manager Host : <clouderamanagerserver>

# Before Installaing Cloudera Manager :

Install and configure Mysql Server ([reference link](https://www.cloudera.com/documentation/enterprise/5-8-x/topics/cm_ig_mysql.html#cmig_topic_5_5))

Cloudera Manager uses RDBMS as a backend to store metadata information about the cluster and cluster management . By default it comes with an embedded RDBMS , but for production use it is recommended to use and external RDBMS . In current deployment MYSQL is being used to store all the necessary metadata .

* yum -y install mysql-server mysql
* chkconfig mysqld on
* service mysqld start
* Download mysql connector and extract it and copy to the below location
  + wget <https://cdn.mysql.com//Downloads/Connector-J/mysql-connector-java-5.1.44.tar.gz>
  + tar zxvf mysql-connector-java-5.1.44.tar.gz
  + cp mysql-connector-java-5.1.44/mysql-connector-java-5.1.44-bin.jar /usr/share/java/mysql-connector-java.jar

## Configure and tune MYSQL

After starting mysql set the mysql root password Set mysql root password . “root” password for mysql will be shared separately.

Stop the running mysql instance, backup the existing my.cnf file and then add the below configuration as recommended by Cloudera

[mysqld]

transaction-isolation = READ-COMMITTED

# Disabling symbolic-links is recommended to prevent assorted security risks;

# to do so, uncomment this line:

# symbolic-links = 0

key\_buffer\_size = 32M

max\_allowed\_packet = 32M

thread\_stack = 256K

thread\_cache\_size = 64

query\_cache\_limit = 8M

query\_cache\_size = 64M

query\_cache\_type = 1

max\_connections = 550

#expire\_logs\_days = 10

#max\_binlog\_size = 100M

#log\_bin should be on a disk with enough free space. Replace '/var/lib/mysql/mysql\_binary\_log' with an appropriate path for your system

#and chown the specified folder to the mysql user.

log\_bin=/var/lib/mysql/mysql\_binary\_log

# For MySQL version 5.1.8 or later. For older versions, reference MySQL documentation for configuration help.

binlog\_format = mixed

read\_buffer\_size = 2M

read\_rnd\_buffer\_size = 16M

sort\_buffer\_size = 8M

join\_buffer\_size = 8M

# InnoDB settings

innodb\_file\_per\_table = 1

innodb\_flush\_log\_at\_trx\_commit = 2

innodb\_log\_buffer\_size = 64M

innodb\_buffer\_pool\_size = 4G

innodb\_thread\_concurrency = 8

innodb\_flush\_method = O\_DIRECT

innodb\_log\_file\_size = 512M

[mysqld\_safe]

log-error=/var/log/mysqld.log

pid-file=/var/run/mysqld/mysqld.pid

sql\_mode=STRICT\_ALL\_TABLES

Start mysql after the configurations are modified .

## Configure Cloudera Manager to use Mysql datastore

Below are the databases that are required for different components within hadoop and cloudera .

| **Role** | **Database** | **User** | **Password** |
| --- | --- | --- | --- |
| Activity Monitor | amon | amon | Passwords to be shared separately |
| Reports Manager | rman | rman | Passwords to be shared separately |
| Hive Metastore Server | metastore | hive | Passwords to be shared separately |
| Sentry Server | sentry | sentry | Passwords to be shared separately |
| Cloudera Navigator Audit Server | nav | nav | Passwords to be shared separately |
| Cloudera Navigator Metadata Server | navms | navms | Passwords to be shared separately |
| Hue | Hue | Hue | Passwords to be shared separately |
| Oozie | Oozie | oozie | Passwords to be shared separately |

Please note all the above databases are created per “utf8” character set per cloudera recommendation

Below is an example of how the database was created :

* CREATE DATABASE amon DEFAULT CHARACTER SET utf8;
* GRANT ALL on amon.\* TO 'amon'@'%' IDENTIFIED BY 'password';

# Establish Cloudera Manager Repository Stratergy

Necessary software rpm’s need to be download to local repository from cloudera archive website . In this case we have used 5.8.4 version (<http://archive.cloudera.com/cm5/repo-as-tarball/5.8.4/)>

Below steps provide information of all the steps that were executed to create the repository and install the package .

## Make the rpms available through webserver:

The complete repository as a tar ball was used in this case . So untarring the repository into a webserver and making it available through the webserver is necessary .

# Install Cloudera Manager

* yum -y install oracle-j2sdk1.7 cloudera-manager-server cloudera-manager-daemons

## Prepare Cloudera Manager server to use external Database:

Before starting the cloudera manager server , run the below command to prepare cloudera manager to write into external database

* /usr/share/cmf/schema/scm\_prepare\_database.sh mysql –uroot –P scm scm password

password for scm database shared separately

## Start Cloudera Manager server:

## On cloudera host execute the below

* service cloudera-scm-server start

## Login to Cloudera Manager :

* Cloudera Manager can be access from this url : <clouderamanagerurl>
* Once Logged in accept the license agreement , upload the necessary license file and proceed adding nodes that will participate in the Hadoop cluster .
* The steps are self descriptive and will guide in adding different Hadoop technologies
* Mysql database username and password will be required when installing components such as Hive, hue ,oozie . These were created in prior steps and having passwords handy will help.

Reference :

<https://www.cloudera.com/documentation/enterprise/5-8-x/topics/cm_ig_install_path_b.html#cmig_topic_6_6>