# 

**EPx.0 – Cloudera Cluster installation Guide**

Contents

[1. Introduction 3](#_Toc462928003)

[1.1 Purpose 3](#_Toc462928004)

[1.2 Scope 3](#_Toc462928005)

[1.3 Definitions, Acronyms, and Abbreviations 3](#_Toc462928006)

[2. Functional Requirements 4](#_Toc462928007)

[3. Kafka Upgrade 4](#_Toc462928008)

[4. System Configuration 4](#_Toc462928009)

[5. Operating System Installation 5](#_Toc462928010)

[5.1 Out of the Box Installations: 5](#_Toc462928011)

[5.1.1 **JAVA** 5](#_Toc462928012)

[5.1.2 **MySQL** 6](#_Toc462928013)

[6. Cloudera Installation: 7](#_Toc462928014)

[6.1 Following will be installed as part of this step: 7](#_Toc462928015)

[6.2 Cloudera Machine Installation 7](#_Toc462928016)

[6.3 Cloudera Parcels Install & Configure 7](#_Toc462928017)

[6.4 Failure in Installation of any Node 21](#_Toc462928018)

[6.5 Auto-Restart (Optional) 23](#_Toc462928019)

[6.6 Change Database for Cloudera (In case if the Installer didn’t asked for it) 24](#_Toc462928020)

[6.6.1 Stop the services on all the nodes for Server and Agent both. 24](#_Toc462928021)

[6.6.2 Go to the MYSQL server and create a blank database: 24](#_Toc462928022)

[6.6.3 Then edit the file on Namenode and update the following parameters, For ex for MySQL: 24](#_Toc462928023)

[6.6.4 Then just restart the Cloudera services on all nodes 25](#_Toc462928024)

[6.7 In Box Installations: 25](#_Toc462928025)

[6.7.1 Phoenix 25](#_Toc462928026)

[6.8 Spark Mode: 25](#_Toc462928027)

# 

# Introduction

Objective of the Document is to provide steps for the Installation of Cloudera cluster.

## Purpose

This document intends to guide on the Installation part of the Cloudera Cluster

## Scope

The scope of this document is limited to installation and configuration of the Cloudera Cluster

## Definitions, Acronyms, and Abbreviations

|  |  |
| --- | --- |
| **Abbreviations / Acronyms** | **Definitions** |
| EP | Entity Profiling 3.0 |
| NameNode | NM |
| DataNode | DN |
| ClearInsight | CI |

# Functional Requirements

Following are functional requirements-

1. Creating Cloudera distribution based Cluster with required technologies and services.
2. Configuring the cluster.

# Kafka Upgrade

Existing Kafka deployment should be upgraded to version 0.9.0.1.

# System Configuration

1. Suggested System Configuration

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S. No | No Of Node | Node | Recommended System Configuration | | | | Components |
| Core | RAM | OS Volume GB i.e. (“/”) | Data Volume GB i.e. (“/data”) |
| 1 | 1 | Name Node | 8 | 12 | 40 | 75 | NameNode, SecondaryNameNode, SparkJobHistory, Hbase Master, ZookeeperServer, Yarn ResourceManager, Yarn JobHistoryServer |
| 2 | 2 | Spark Node | 12 | 32 | 40 | 50 | ZookeeperServer, Yarn NodeManager |
| 3 | 3 | Data Node | 12 | 32 | 40 | 100 | Hbase RegionServer, Hbase Master(any 1 node), DataNode, ZookeeperServer (any 2 nodes) |

# Operating System Installation

**Note**: Below Steps from 1 to 6 are recommended to be configured by NMG Team.

1. Install OEL Linux Version 7.1 with all default packages on all Node (NameNode, Data Nodes & Spark Nodes)
2. Install following packages on all hosts for dependency resolutions:

**sudo yum install chkconfig bind-utils psmisc libxslt zlib sqlite cyrus-sasl-plain cyrus-sasl-gssapi fuse portmap fuse-libs redhat-lsb ntp -y**

1. Change ‘vm.*swappiness’* to 5 (default is 60)

Command to change vm.swappiness is:

**sysctl vm.swappiness=5**

**cat /proc/sys/vm/swappiness**

1. Static IP entries should be configured for all systems, Remove localhost entries from the host files.

**vi /etc/hosts and comment out as below**

#127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4

#::1 localhost localhost.localdomain localhost6 localhost6.localdomain6

1. Disable SELinux, Iptables and IP6Table

**systemctl stop firewalld**

**systemctl disable firewalld**

1. NTPD service must be started

**systemctl start ntpd**

**systemctl enable ntpd**

1. MOST IMPORTANT

<https://www.cloudera.com/documentation/enterprise/5-9-x/topics/cm_ig_install_path_a.html>

## Out of the Box Installations:

### **JAVA**

Remove the OpenJDK installation if present and Install Oracle JDK on all nodes.

#### Execute below command to uninstall OpenJDK

**sudo yum remove openjdk-6-jre default-jre default-jre-headless**

**or**

yum -y remove java\*

**Note** that Cloudera discourages the use of OpenJDK.

#### Download the JDK from the SVN folder or internet and run the RPM:

**jdk7.rpm**

Upload JDK on all nodes:

Download the RPM and SCP it over to all Hosts:

**scp jdk7.rpm <YourIP>:/root/jdk7.rpm**

run below command on all nodes:

**rpm -ivh /root/jdk7.rpm**

Oracle JDK installation is completed.

### **MySQL**

To install the MySQL server 5.6, follow the given below steps on **Name Node**

#### Download the mysql rpm file from SVN location or from internet.

**mysql-community-release-el6-5.noarch.rpm**

#### Install the downloaded rpm package.

**rpm -ivh mysql-community-release-el6-5.noarch.rpm**

#### Install the mysql server by using yum command.

**yum install mysql-server**

Note: The yum command by-default also install the dependencies

#### Start the mysql server

**/etc/init.d/mysqld start**

#### Login into mysql server. Because it is freshly installed hence root password is blank

**mysql -uroot**

#### After login mysql prompt will come-up. Run the given commands:

**mysql> use mysql;**

**mysql> update user set password=PASSWORD("<GIVE-NEW-ROOT-**

**PASSWORD>") where User='root';**

**mysql> GRANT ALL ON \*.\* TO 'root'@'%' with GRANT OPTION;**

**mysql> flush privileges;**

**mysql> quit**

#### Run the following query on the MYSQL server:

#### Add the connector jars in the Classpath for MySQL

#### Download the jar from SVN and put it on:

#### **cd /usr/share/java**

**Download jar from internet from svn://172.50.40.37/setups/KyvosInt/QARelease/EntityProfiling/Prerequisites/binaries/mysql-connector-java.jar**

Or

**yum install mysql-connector-java**

Note: We always setup external database in Cloudera Manager to retain details in failover or reboot case of Cloudera.

# **Cloudera Installation**:

## Following will be installed as part of this step:

1. Cloudera Manager 5.8.1 parcel version
2. Zookepeer 3.4.5
3. Hbase 1.2.0
4. Hadoop 2.6
5. Phoenix-4.7.0-clabs-phoenix1.3.0
6. Spark 1.6.0

## Cloudera Machine Installation

Download the Cloudera Manager Installer for 5.8.x version from SVN location on **Namenode**:

**svn://172.50.40.37/setups/KyvosInt/QARelease/EntityProfiling/Prerequisites/binaries/cloudera-manager-installer.bin**

Run the installer on Namenode, Change the permissions and run the file using:

**chmod u+x /<path-to-file>/cloudera-manager-installer.bin**

**cd /<path-to-file>/**

**./cloudera-manager-installer.bin**

Login to the Cloudera Manager using below link:

http://<NameNodehost>:7180/cmf

**Note:** If any issue is there in accessing the webpage check that the hosts

firewalls are turned off. Firewall blockage can cause cloudera agents fail and other kind of hosts problems**.**

## Cloudera Parcels Install & Configure

Cloudera Manager will prompt to select a CDH version. Choose Express version and continue with the Installation.

Select CDH-5.8.10 and on the Next Parcels option will be presented to install.

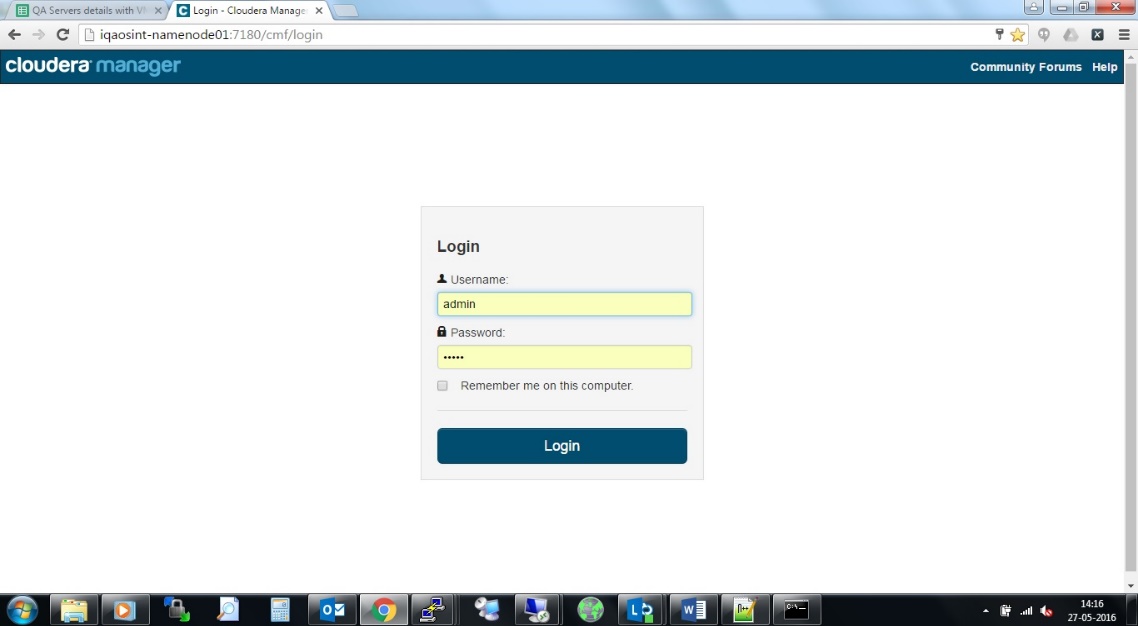
Select the default ones like MR, Yarn and HDFS or go with the custom selection and select the following:

* Yarn (MR2)
* HDFS
* Spark
* HBase
* Zookeeper

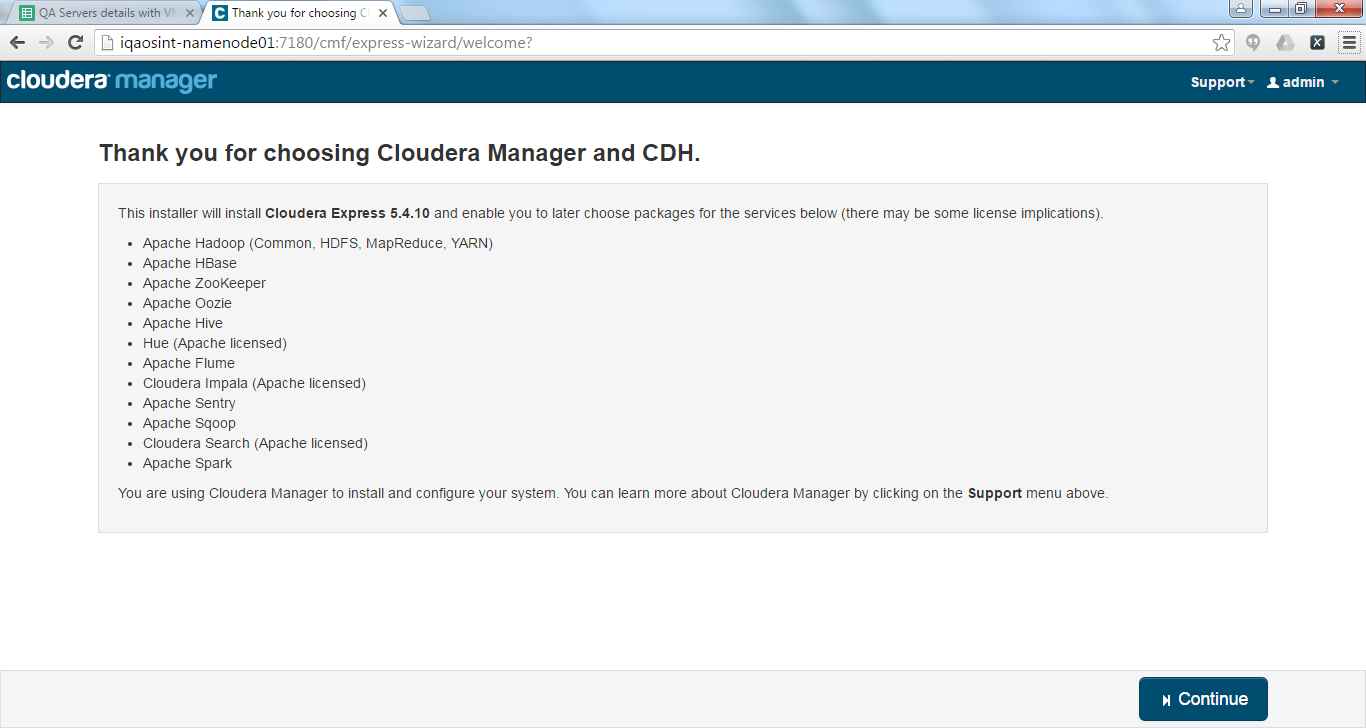
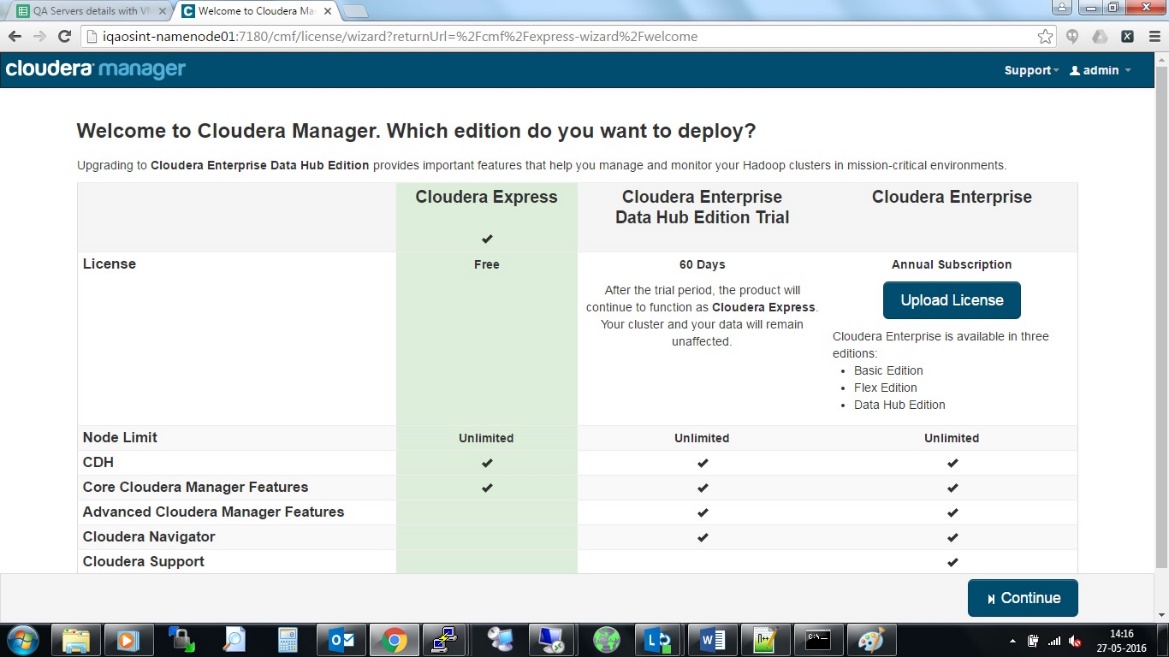
Use below credentails:

**admin / admin**

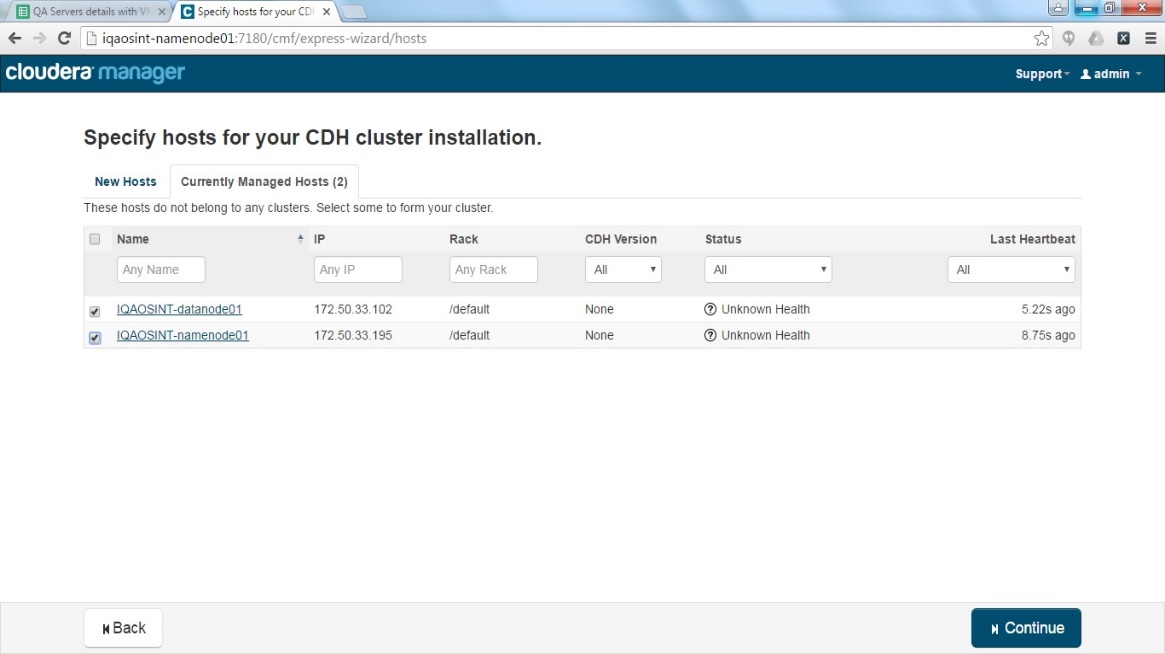
And follow the instructions on the screen.



**Login and Select the Cloudera Express version.** Screenshot are having version 5.4.10 but the current version installed will be 5.8.1. Process is same for all of the installations.



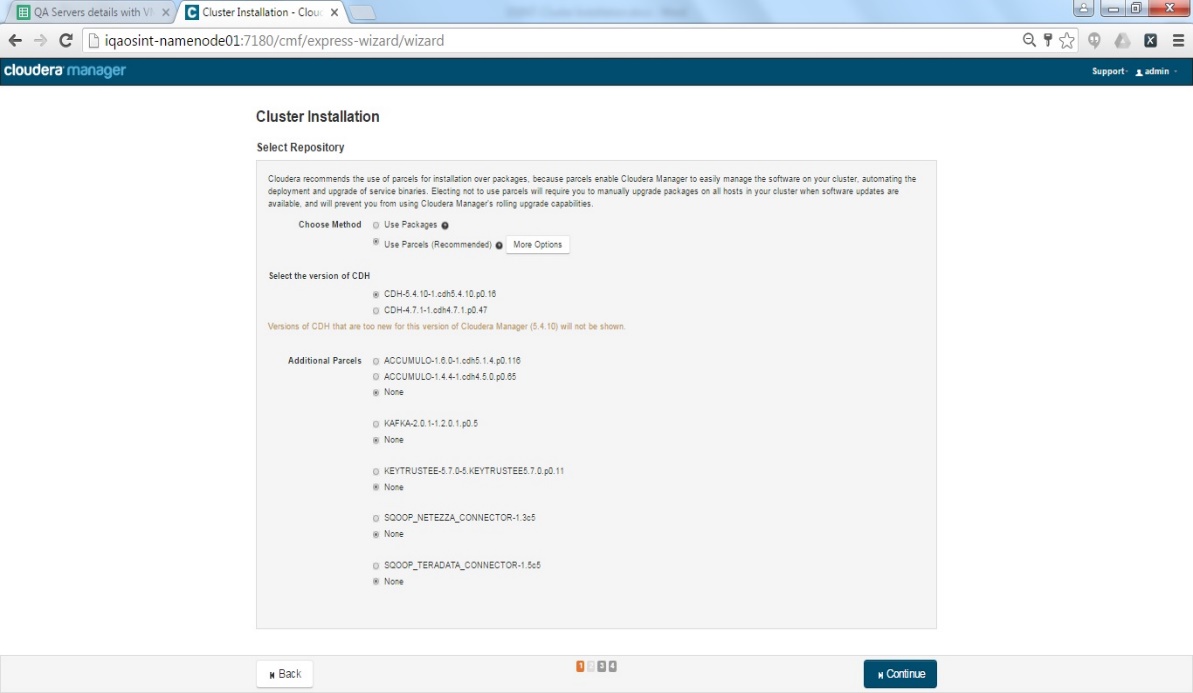
Select the Hosts or enter the Host details which will be used in the Setup:



**Note**: After following above points CDH 5.8.1 will appear instead of 5.4.10

in screenshot.

Select the Parcels (instead of packages) and select desired CDH version and select next/continue as below screenshots:

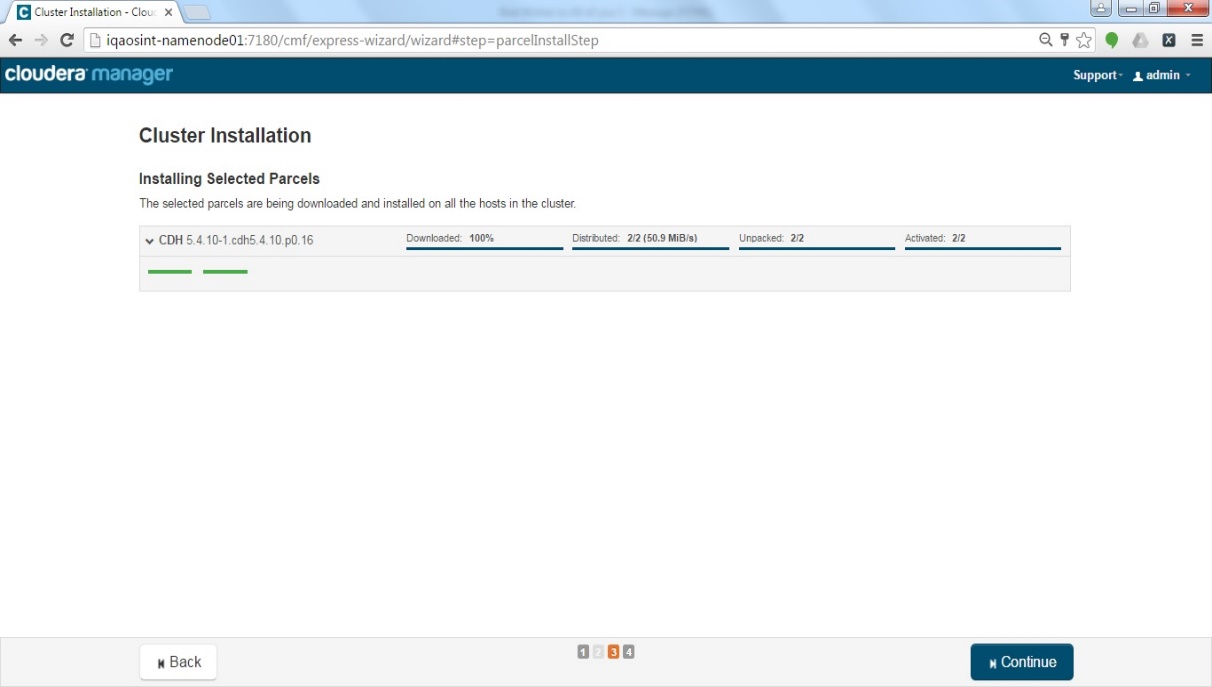
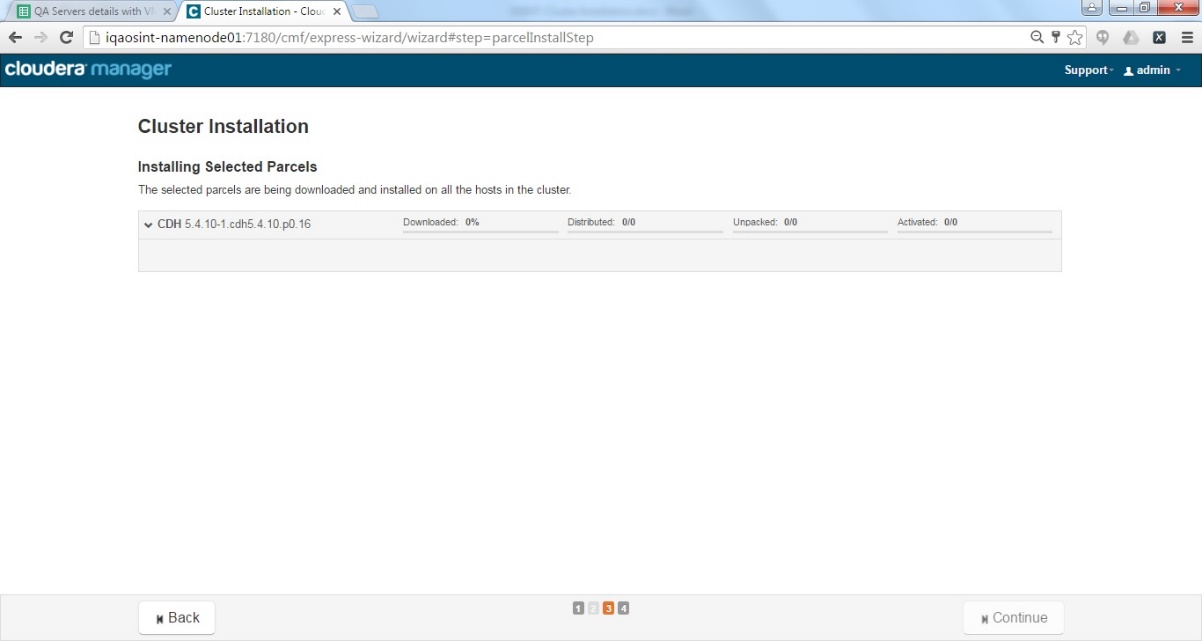


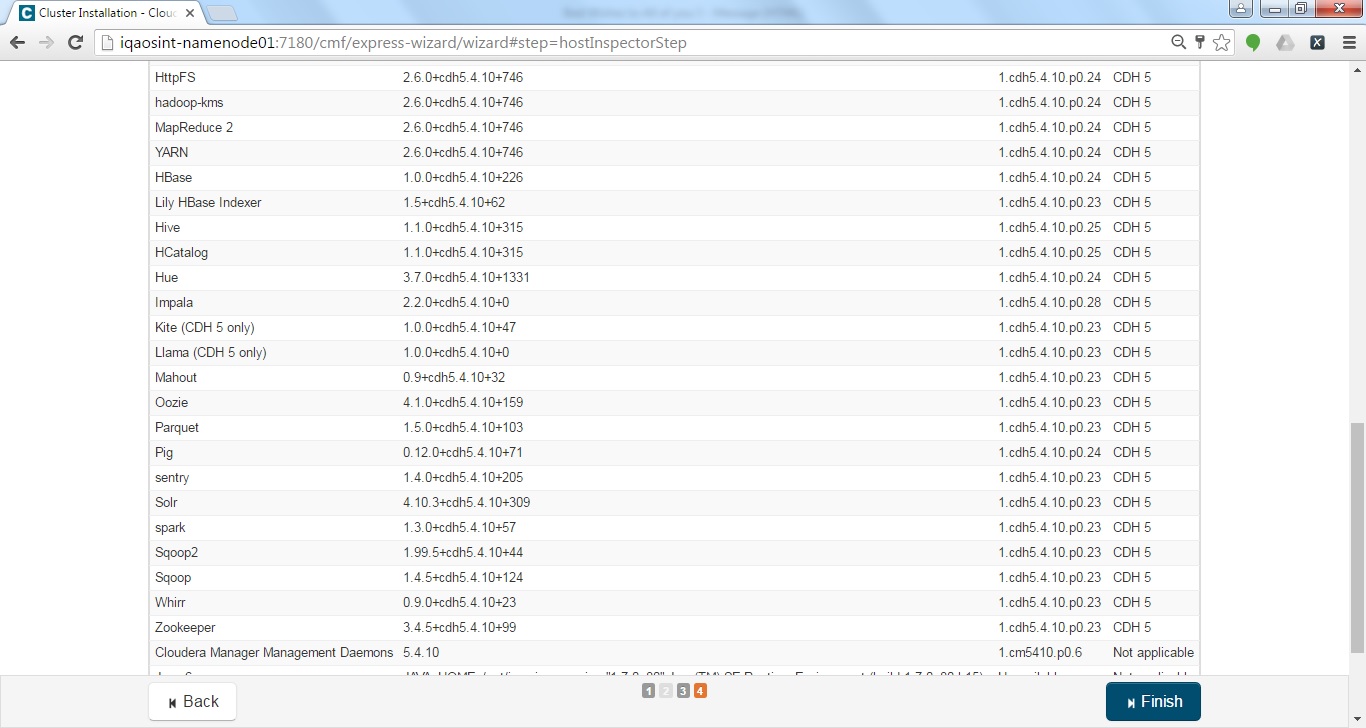
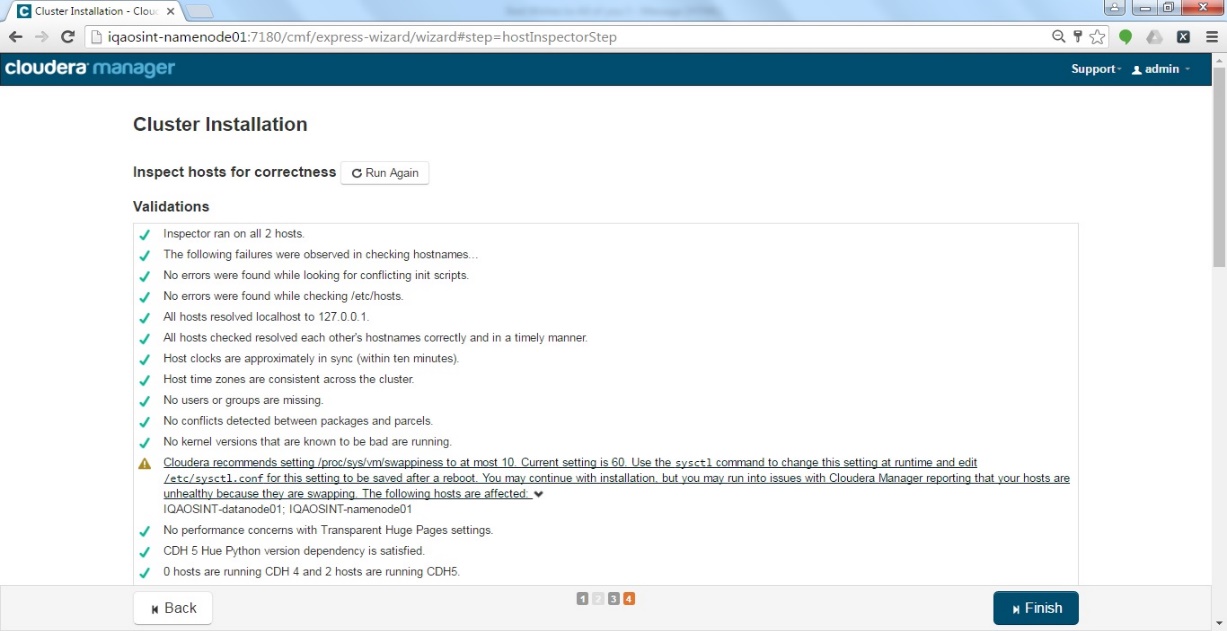
Status progress in below screenshot:

DOWNLOADED: Will download the Cloudera Manager, CDH components and selected service parcels

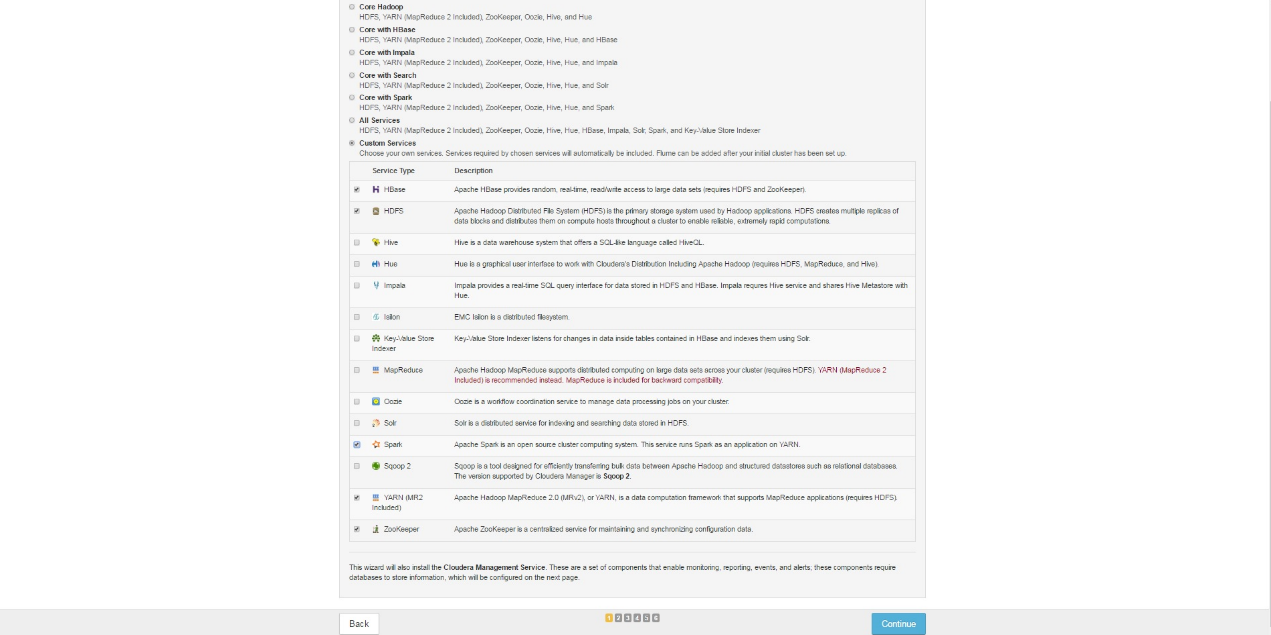
DISTRIBUTED: Will distribute the Parcels over the Cluster

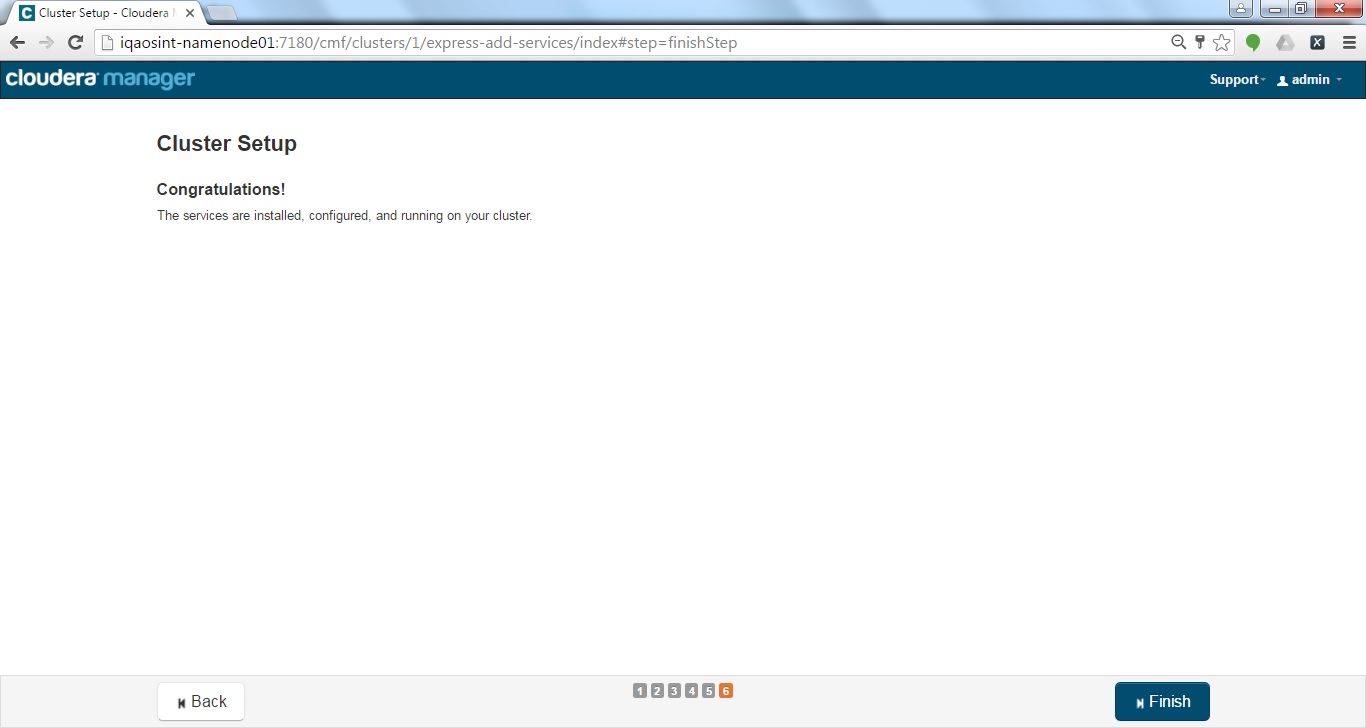
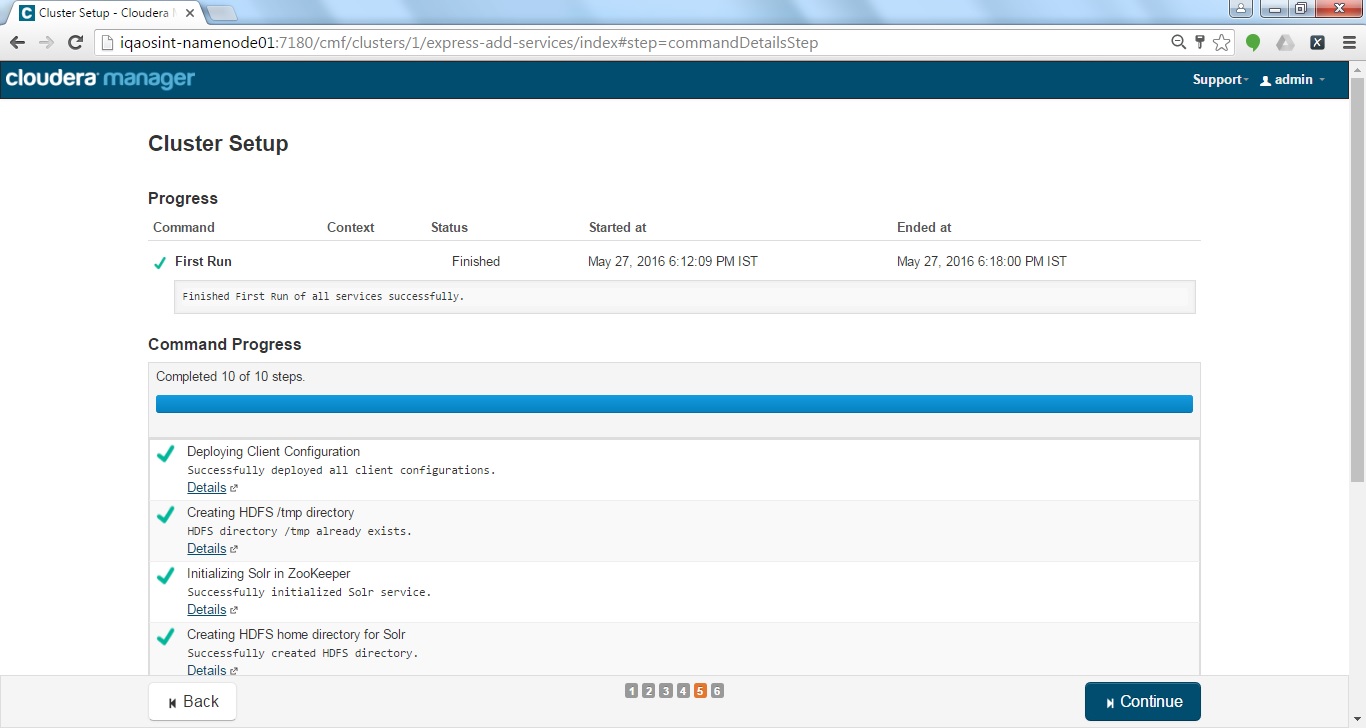
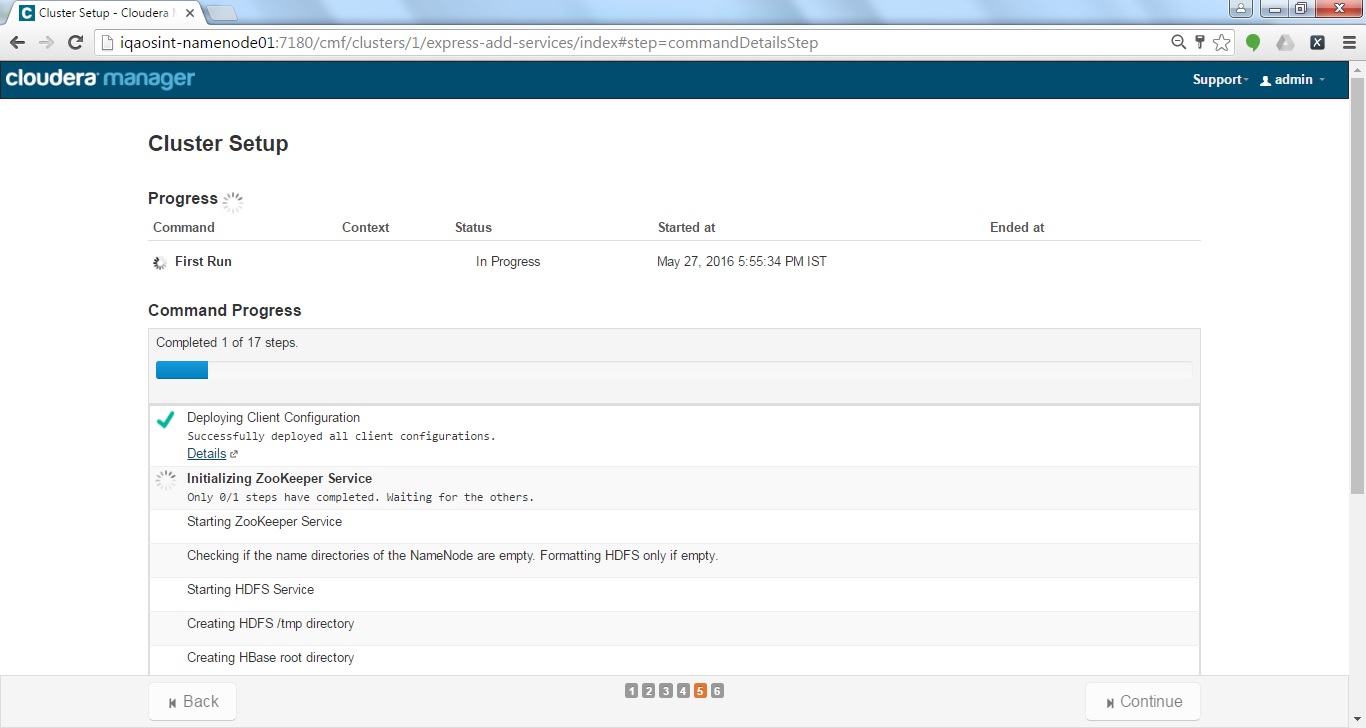
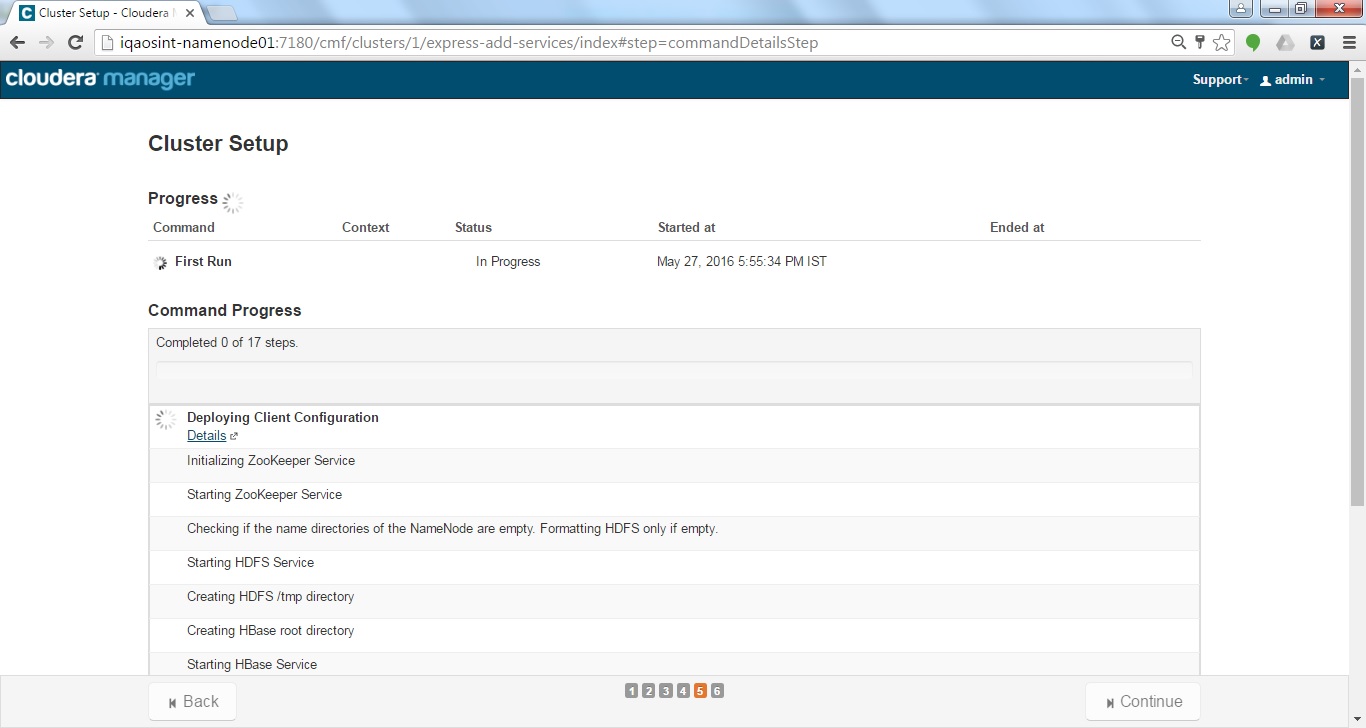
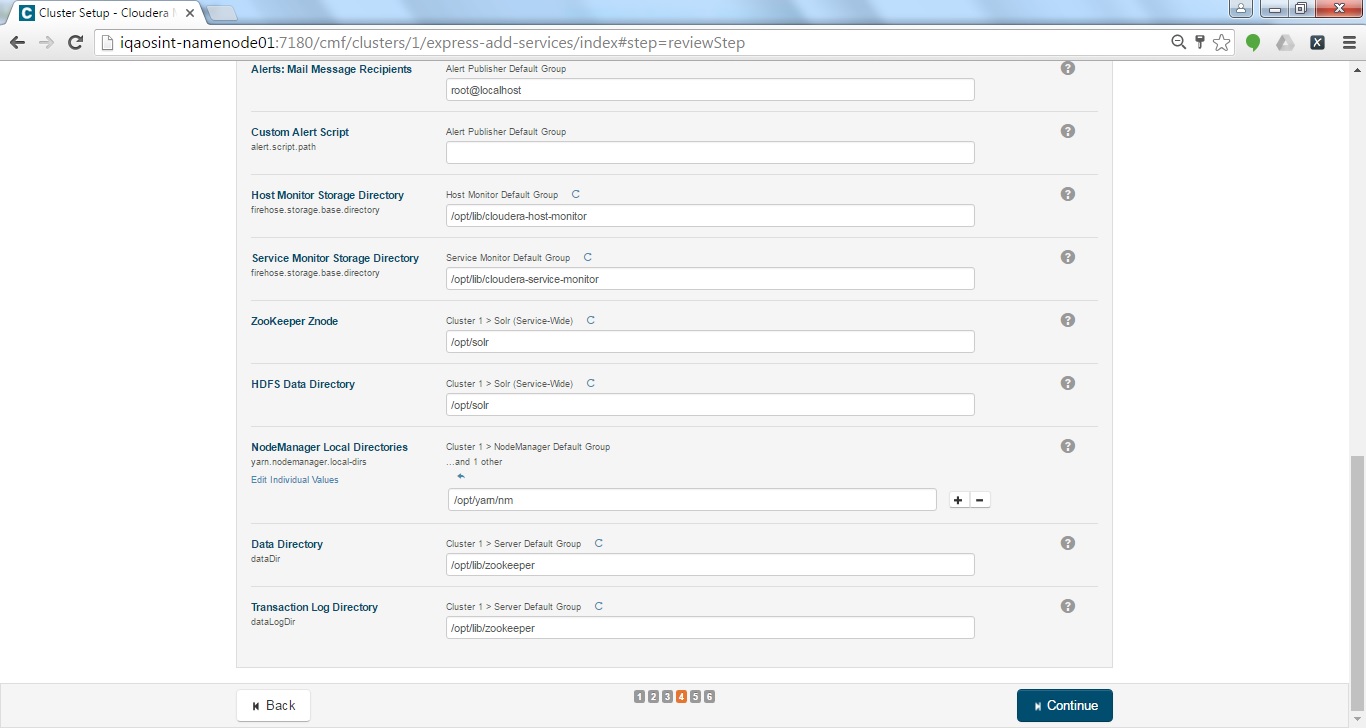
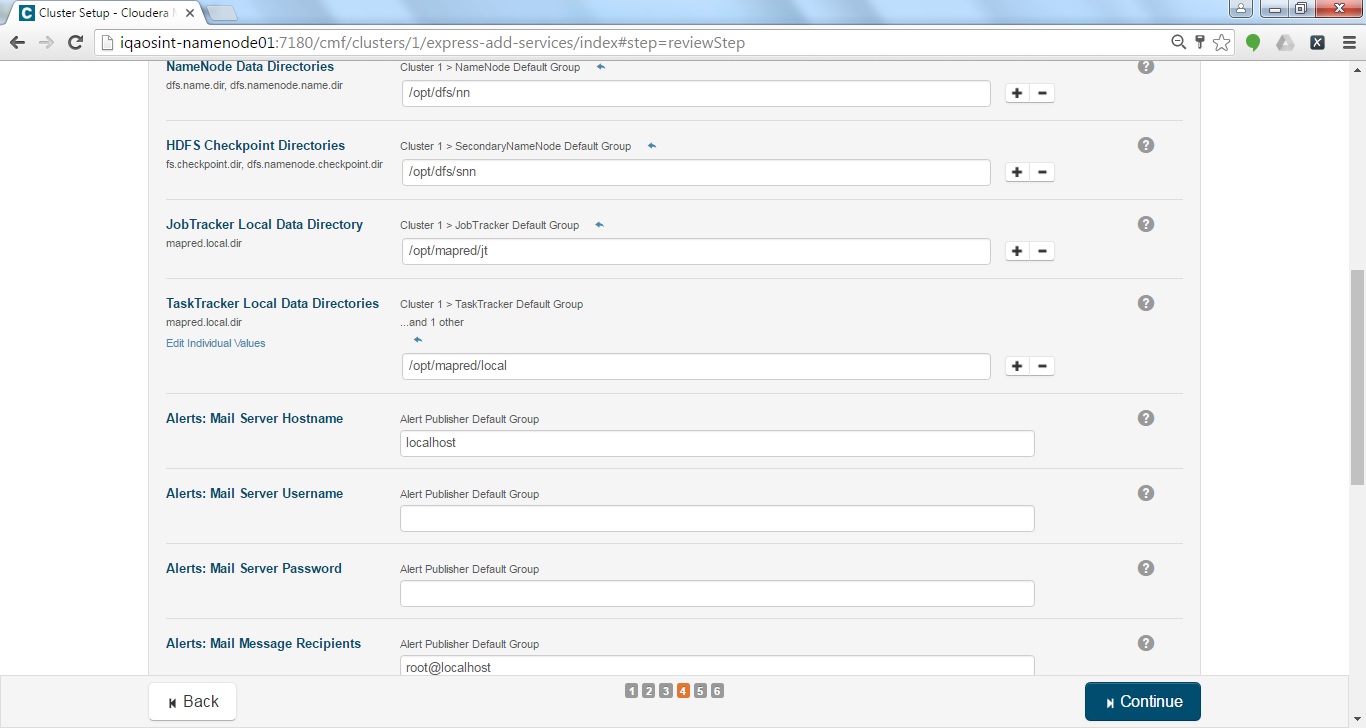
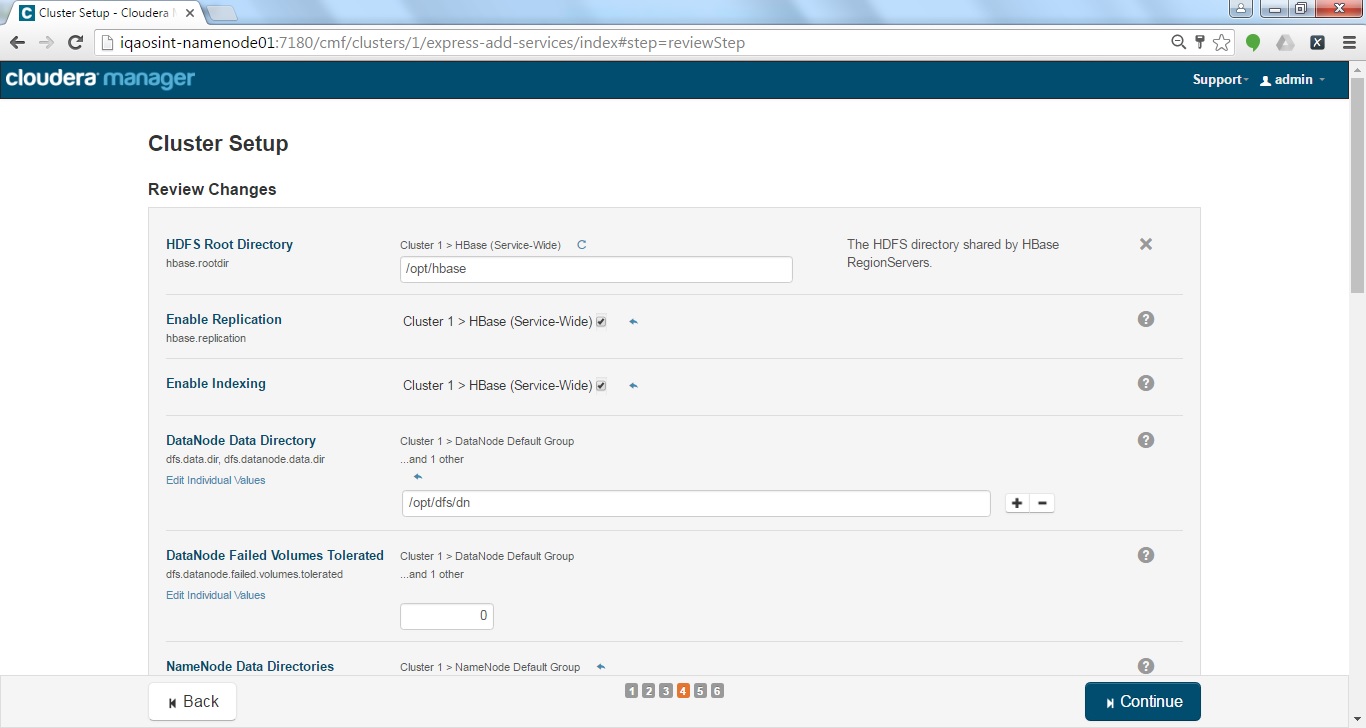
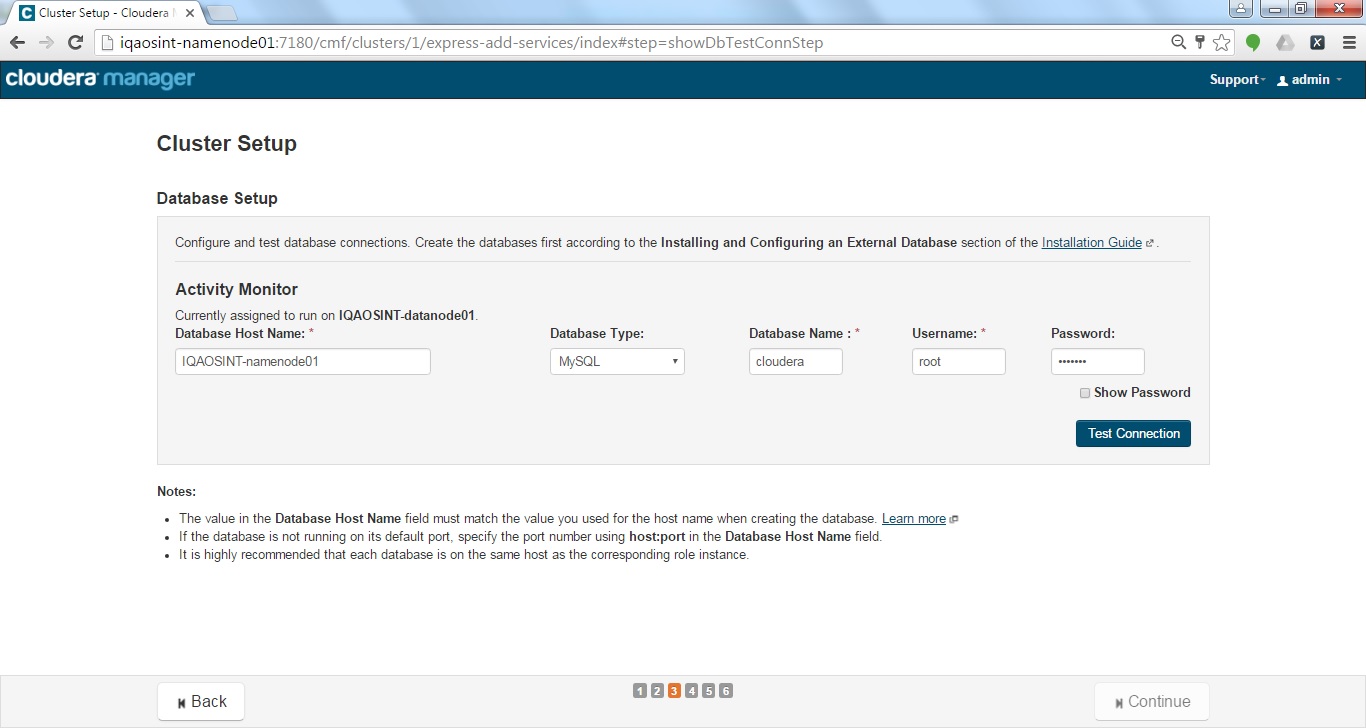
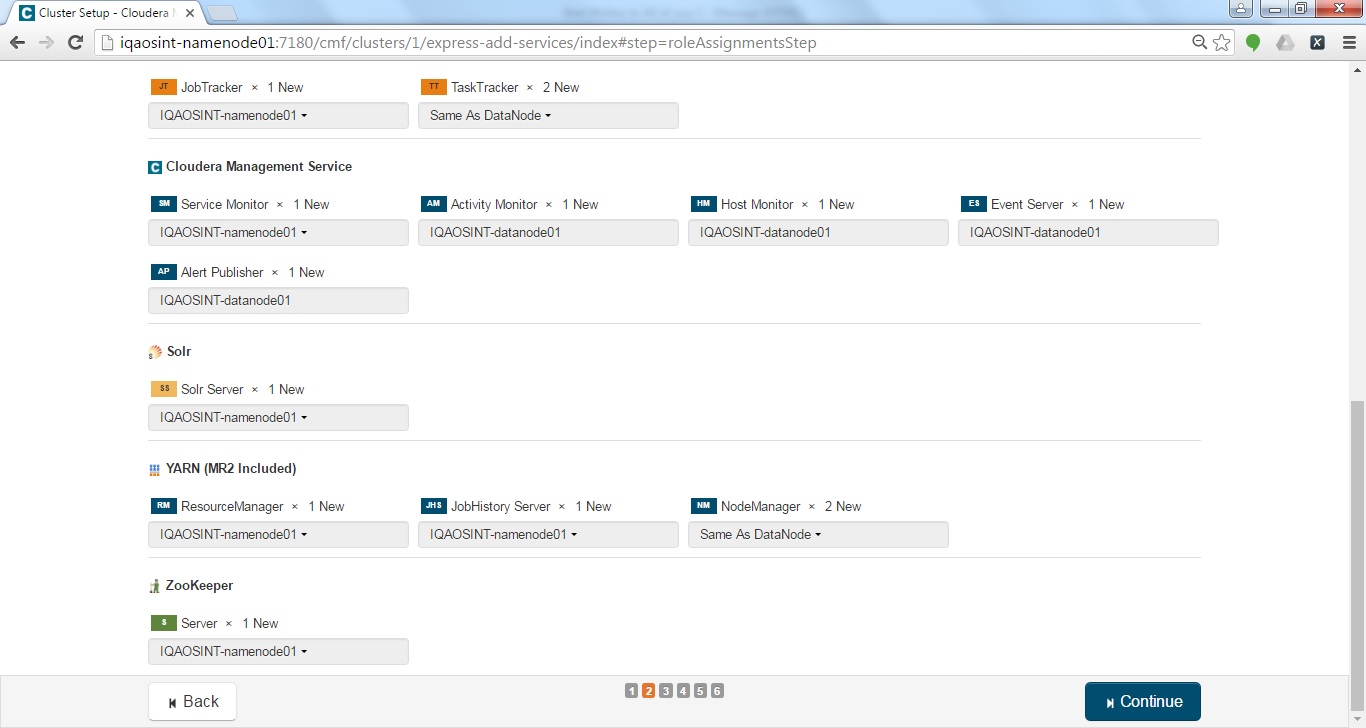
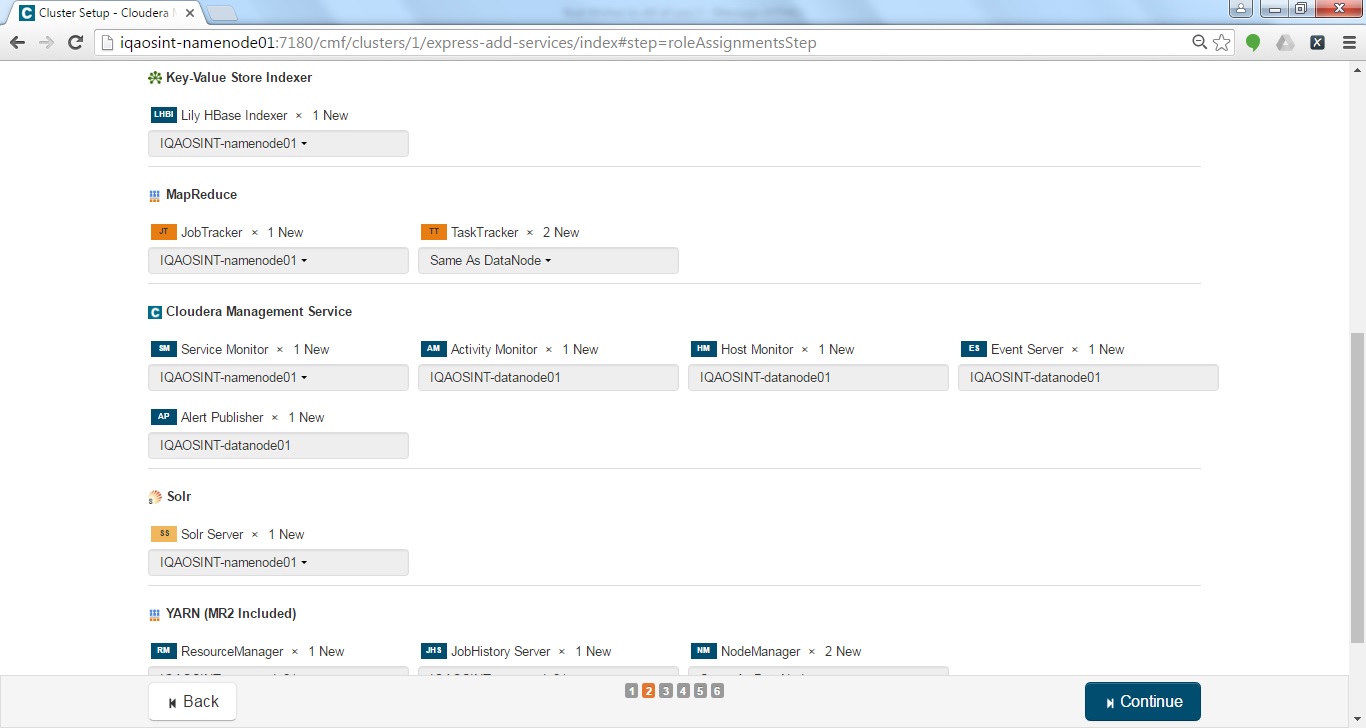
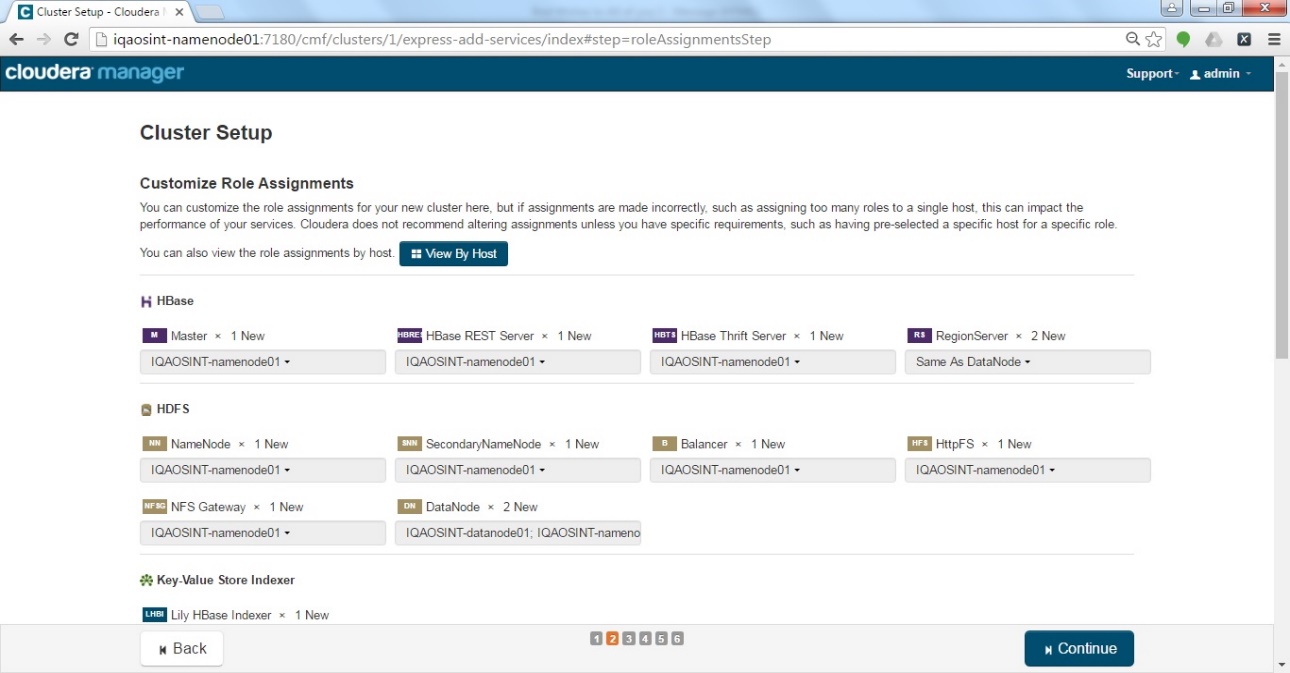
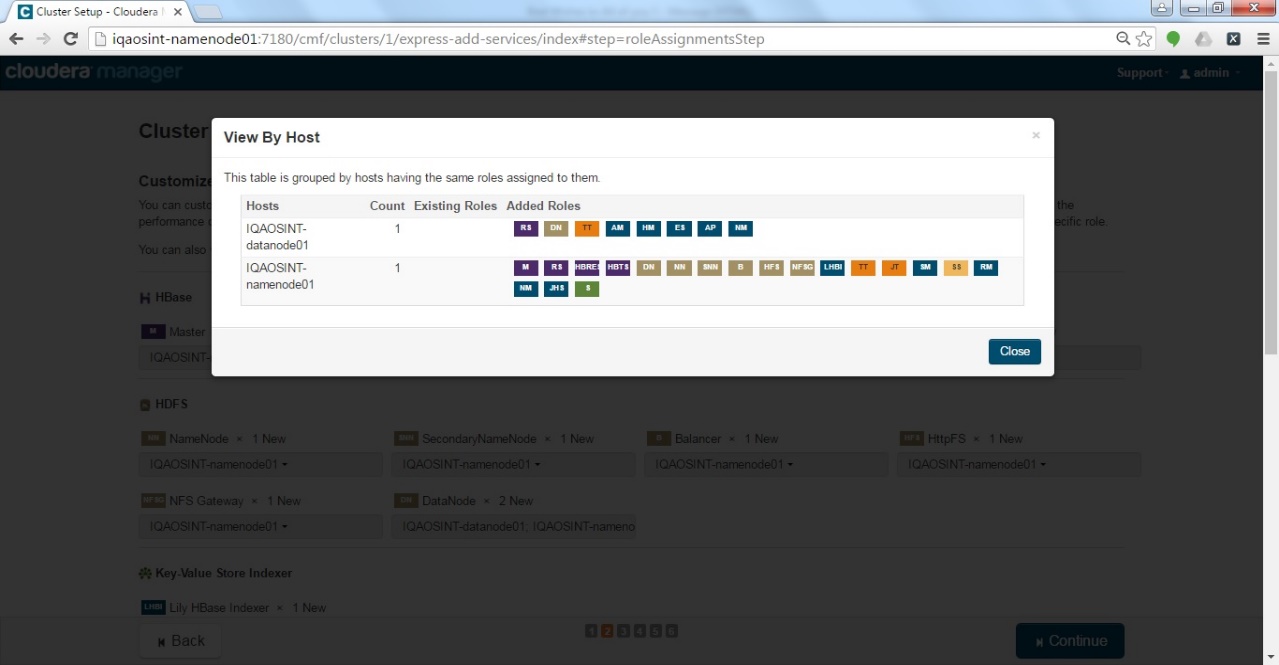
Remember this process depends on Internet and Network Speed and it could take a day. Make sure this part should be done in the time when the Network is on high availability and does not get disconnected frequently.











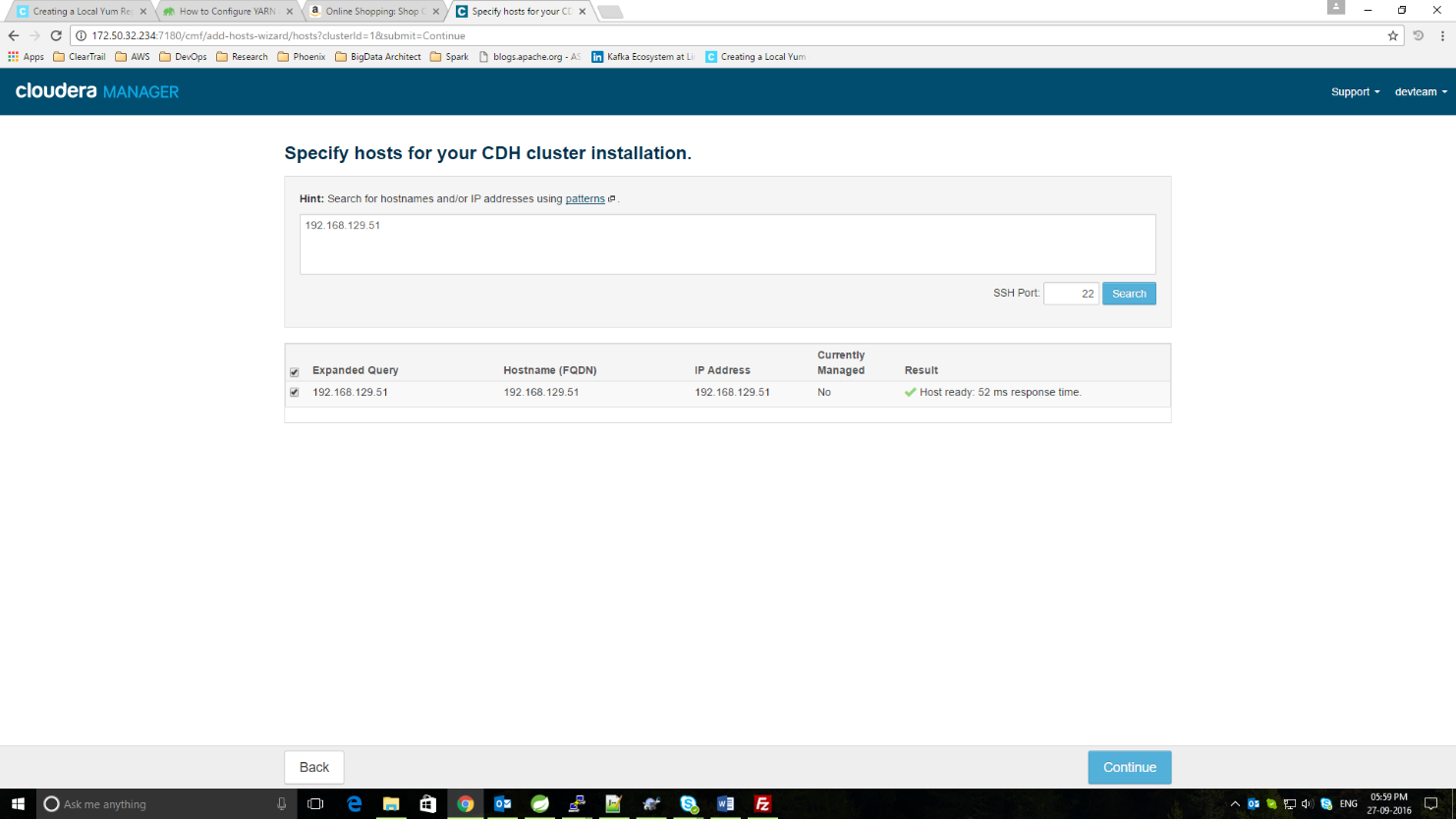
After the Installation fix the Java Heap Spaces for Cloudera Services, keep the Java heap space as required only not as per guidelines of Cloudera. For Agent and Host monitor 768MB is enough

## Failure in Installation of any Node

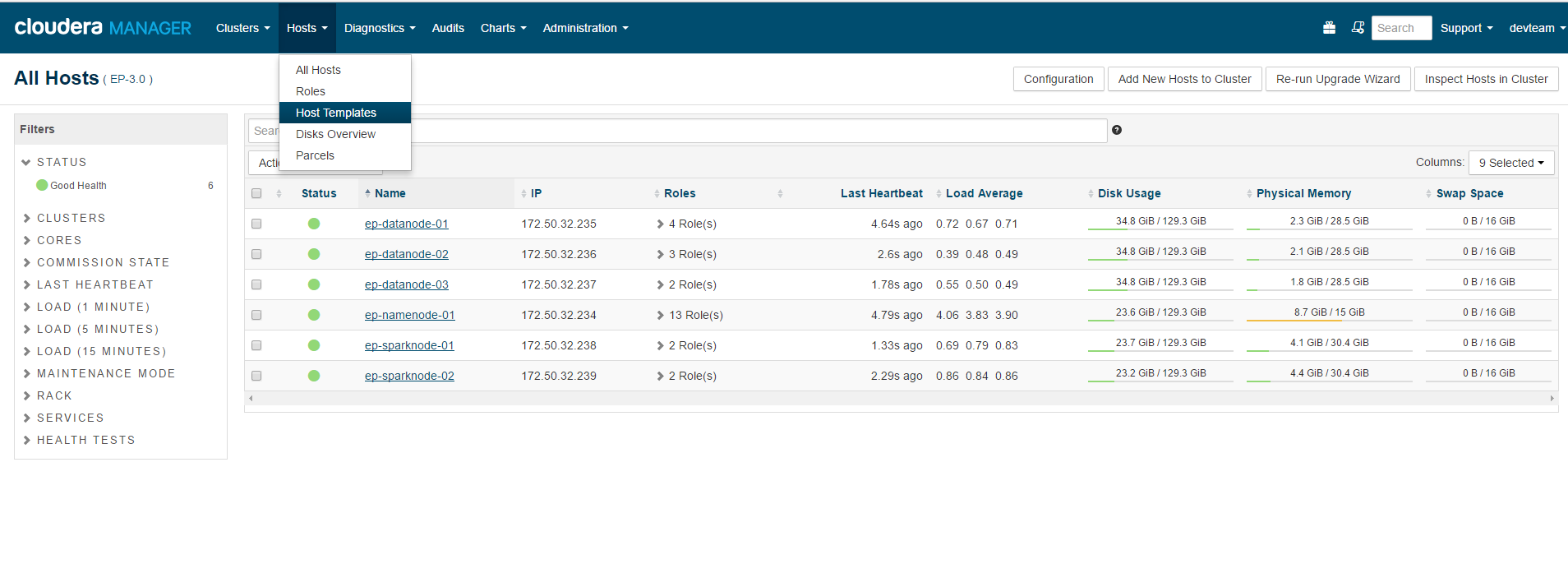
In case of failure of any specific node on the cluster then follow the below steps to add a Single/Multiple node(s) using Host Template.

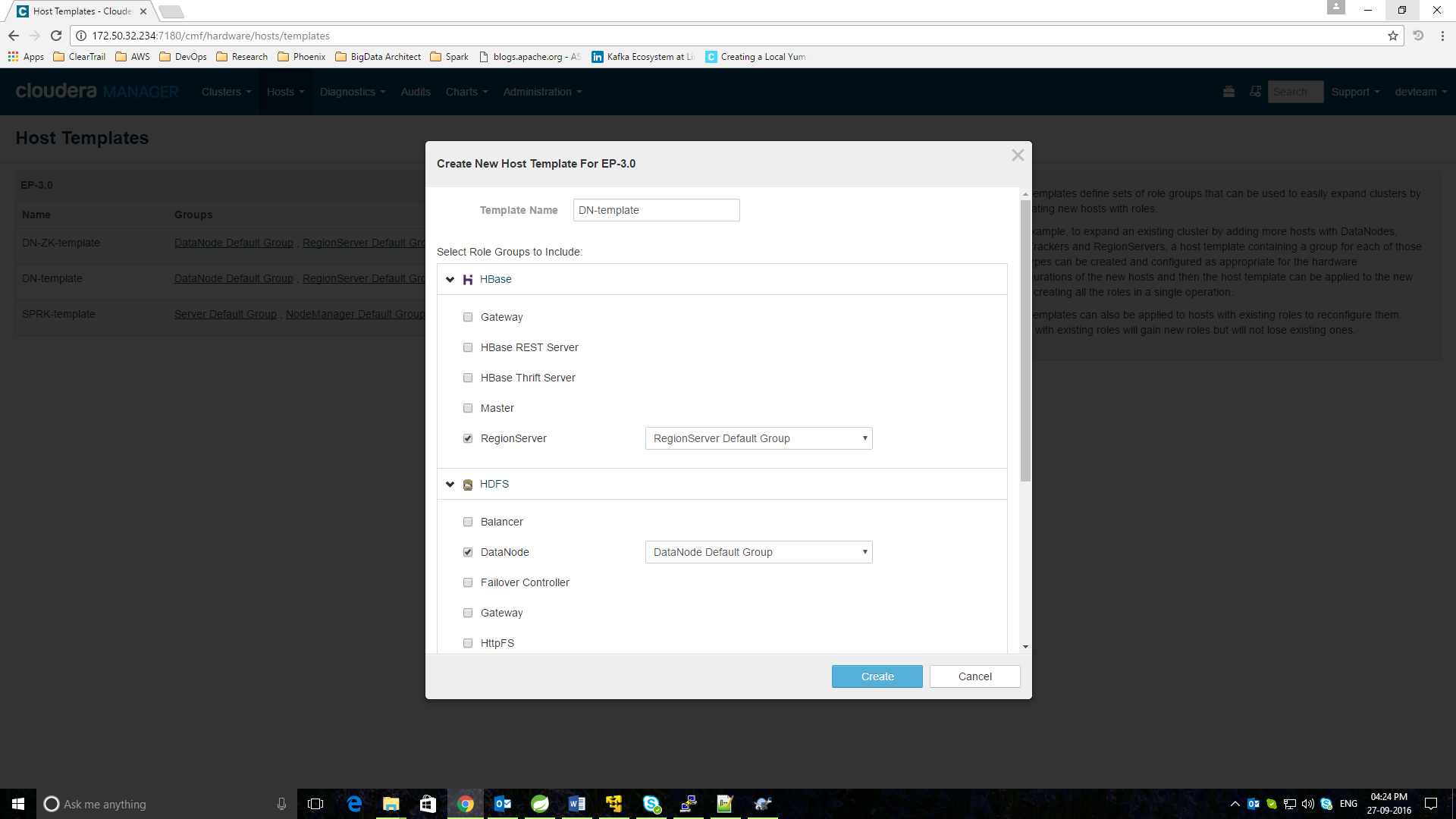
Note: This Method can be used for later to add new nodes on requirement.

Click on add new host(s) and Search for the IP you want to add and select the following, after that click on Continue button to add.

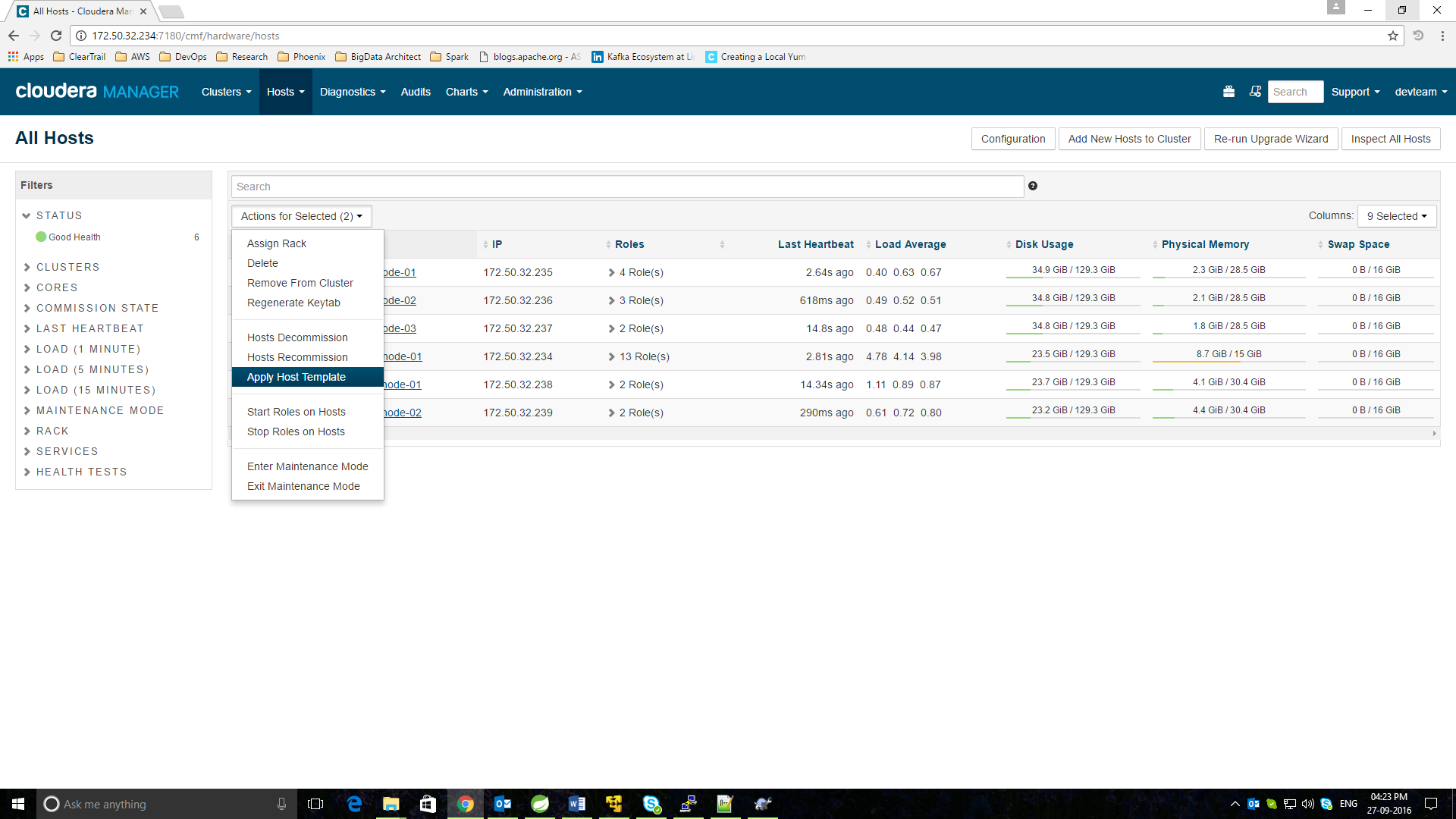


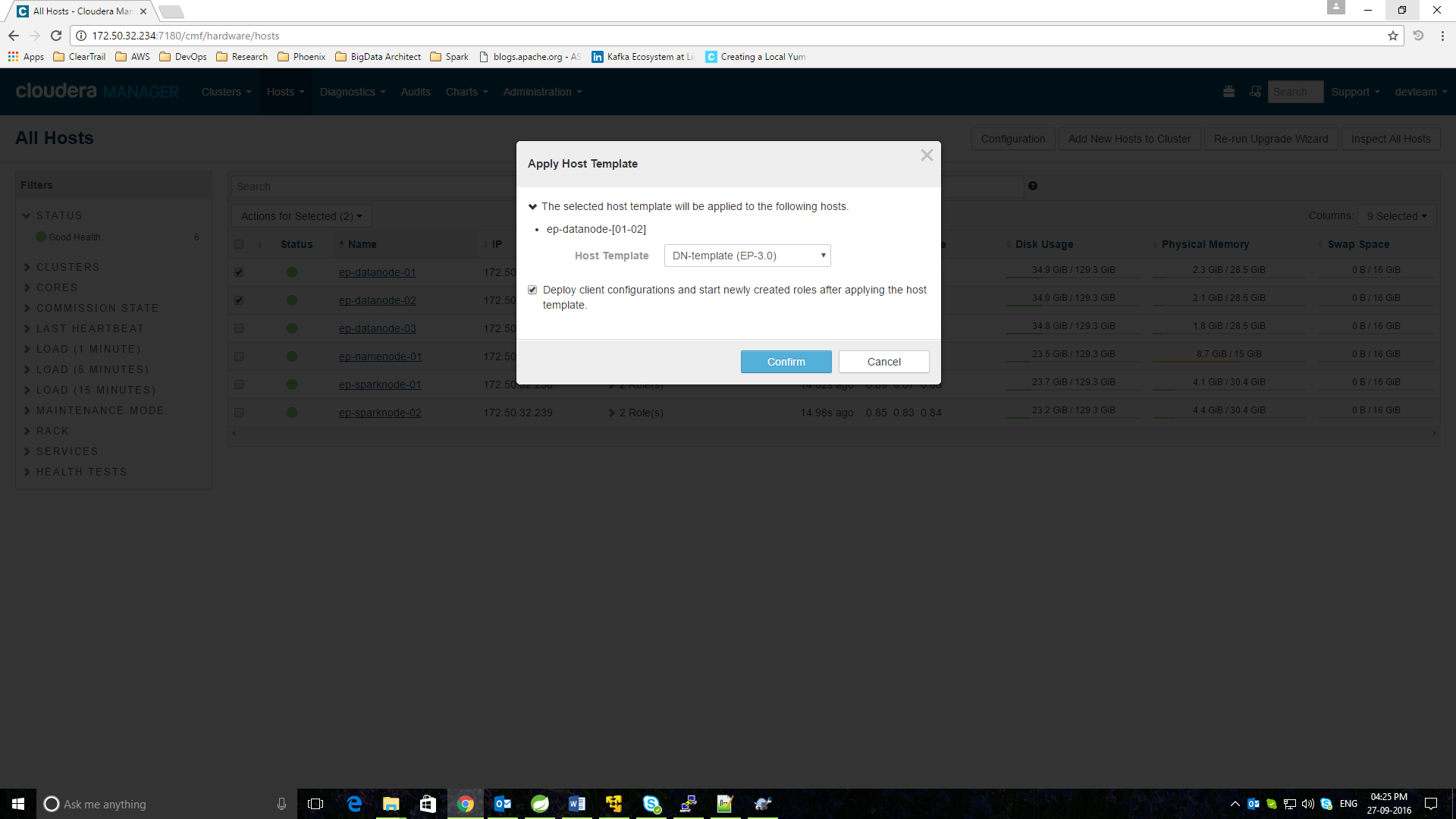
Then apply Host Template to the following node you required.



Click on the Create button and a popup will appear, select the Roles you want to attach with it and click on Create button.  
  


After that move to the Hosts screen, Select the nodes you want to apply the template and click on Apply Host Template.



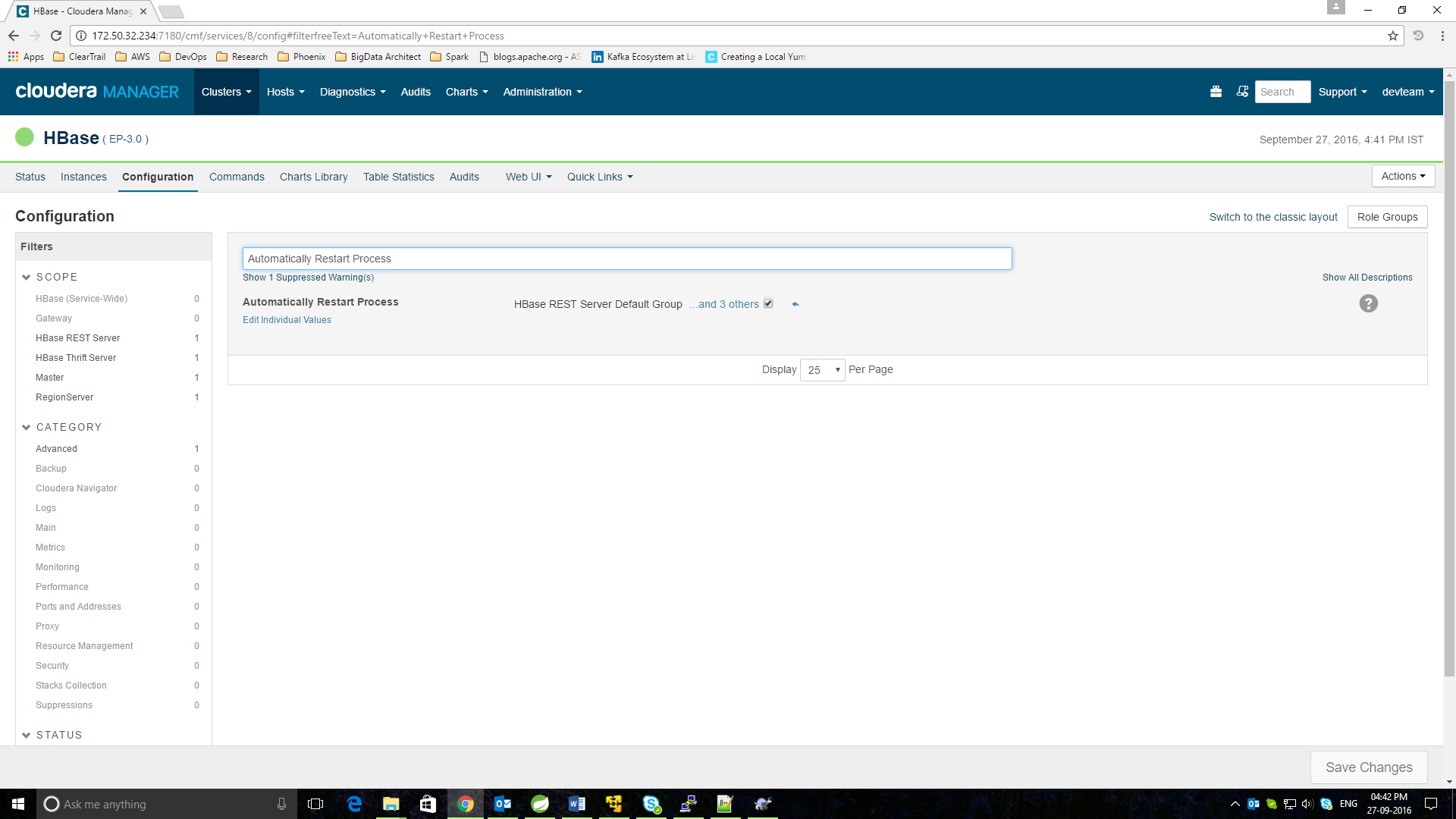


Select the Deploy configuration checkbox and click on Confirm button to apply.

Check for the Roles, If started or not. In case if any of the services failed after applying this. You just need to restart the whole service for it.  
  
**Note: This approach is best when you need to distribute your environment in different roles. And also have a template available to apply the Roles quickly to the new nodes**.

## Auto-Restart (Optional)

To enable Auto-Restart just go in the **Service > Configuration** and Search for **“Automatically Restart Process”** and check all of the groups to auto-restart.



Follow these steps for each of the Service.

**Note: This is a good practice while needed the Cloudera to Auto-Restart on hard reboots.**

## Change Database for Cloudera (In case if the Installer didn’t asked for it)

In case if you ever needed to change the database engine for Cloudera, then just follow these simple steps

### Stop the services on all the nodes for Server and Agent both.

**sudo service cloudera-scm-server stop # For Namenode only**

**sudo service cloudera-scm-server agent stop # For All Nodes**

### Go to the MYSQL server and create a blank database:

**mysql -uroot -p**

**create database cloudera;**

### Then edit the file on Namenode and update the following parameters, For ex for MySQL:

**vi /etc/cloudera-scm-server/db.properties**

**com.cloudera.cmf.db.type=mysql #<Your Database engine>**

**com.cloudera.cmf.db.host=<YourNamenodeHost>:3306**

**com.cloudera.cmf.db.name=cloudera #<YourDatabaseName>**

**com.cloudera.cmf.db.user=root #<MYSQL User>**

**com.cloudera.cmf.db.password=ctadmin #<MYSQL Password>**

### Then just restart the Cloudera services on all nodes

**sudo service cloudera-scm-server start # For Namenode only**

**sudo service cloudera-scm-server agent start # For All Nodes**

Cloudera will automatically import the scripts and create all its required tables.

Note: This is required only after Installation steps and this is a one-time step. If you configured your whole cluster and will try to change the database. Then you’ll lose all of your configurations and settings.

## In Box Installations:

### Phoenix

1) Install and activate the parcel:

a.) In Cloudera Manager, go to Hosts, then Parcels.

b.) Select Edit Settings.

c.) Click the + sign next to an existing Remote Parcel Repository URL, and add the appropriate URL

**CDH 5.8.1 - >** [**http://archive.cloudera.com/cloudera-labs/phoenix/parcels/latest/**](http://archive.cloudera.com/cloudera-labs/phoenix/parcels/latest/)

Click Save Changes.

d.) Select Hosts, then Parcels.

e.) In the list of Parcel Names, CLABS\_PHOENIX should be available. Select it and

choose Download.

f.) The first cluster is selected by default. To choose a different cluster for distribution,

select it. Find CLABS\_PHOENIX in the list, and click Distribute.

g.) Click Actions > Restart.

This will open the phoenix shell, execute

**phoenix-sqlline.py <host>:<port> (e.g. phoenix-sqlline.py <your-host>:2181)**

## Spark Mode:

Click on Spark Via Cloudera UI

Spark->Configuration->Serach Default Application Deploy Mode

Select cluster as per below screenshot

