HADOOP ADMINISTRATION

Project/Assignment



Hadoop Administration

Table of Contents

[1. Introduction 3](#_Toc425777070)

[2. Problem Solution 3](#_Toc425777071)

[3. Output 4](#_Toc425777072)

# Introduction

You need to use VMWARE player or any virtualization tool to set up the environment

CentOS 6.5 is preferred

You need to create a regular user in CentOS

You may either configure a DNS server or have the appropriate entries in hosts file

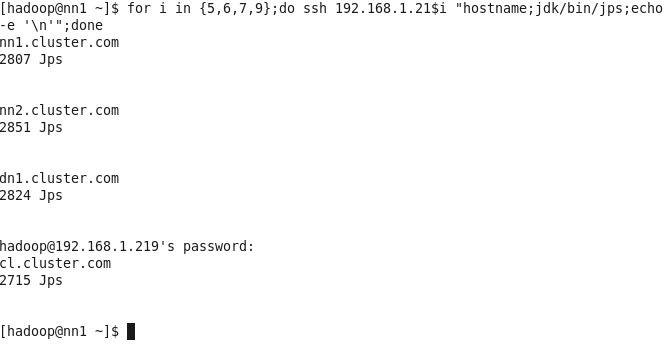
# Problem Solution

Create a 4 node Hadoop cluster with Hadoop 2.6.0 installed. Please make a note of the following points

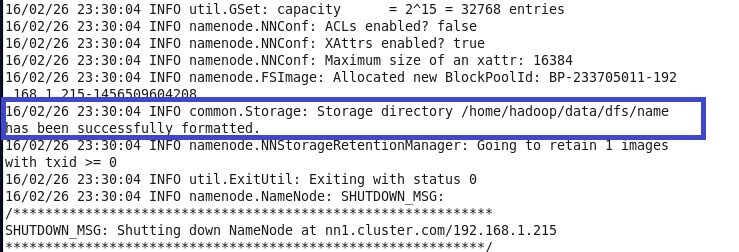
1. You need to use the Federation feature in this cluster
2. Make One machine as Namenode 1 and another as Namenode 2, one machine as Datanode and the last one as client
3. Namenode 1 should hold the namespace for Marketing department and Namenode 2 should hold the namespace for production department
4. Use viewfs on client
5. Create a folder called test in Marketing. Make sure that the client has access only to Marketing department. Use the client to input some data in to test folder
6. Now, grant access to the production department as well. Make sure that the client can access both the Namenodes independently
7. Take down one of the Namenode. Make sure that the client still has access to the other Namenode
8. Install SQOOP in this cluster
9. Install MySQL and create a Database called test
10. Inside the database create a table called students.
11. The table has two columns, student ID and Name. Input some values in to this table
12. Using the SQOOP import tool, selectively import the data from student table in to HDFS. Make sure that we are importing the student names having id>4

Result

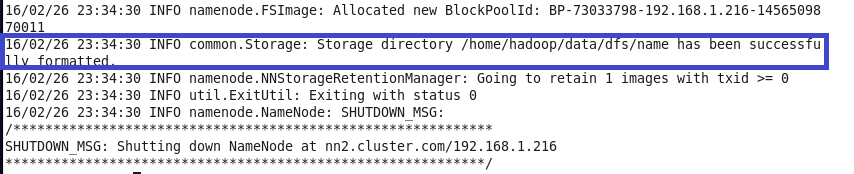
**Nodes:**



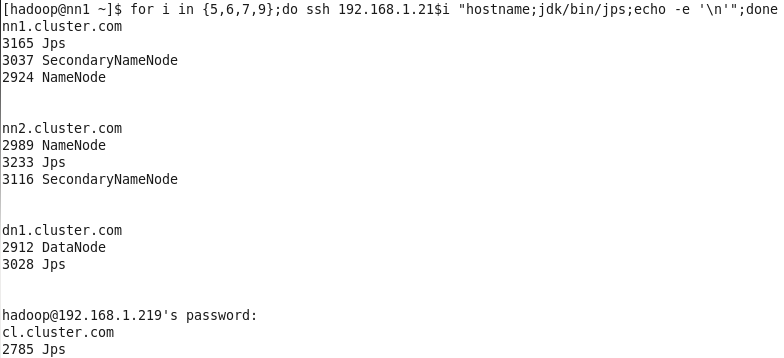
**Namenode-1 Format**



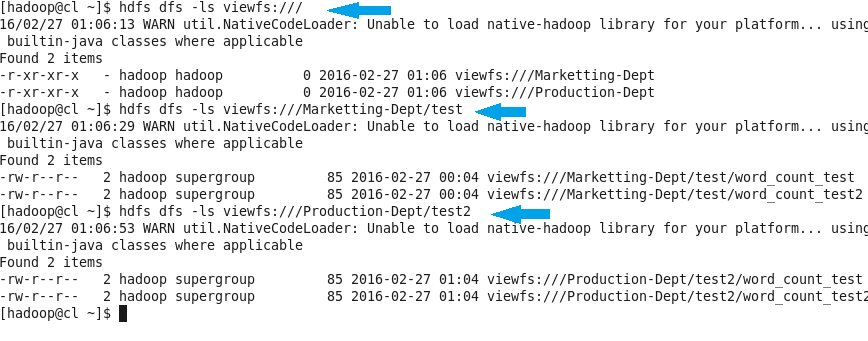
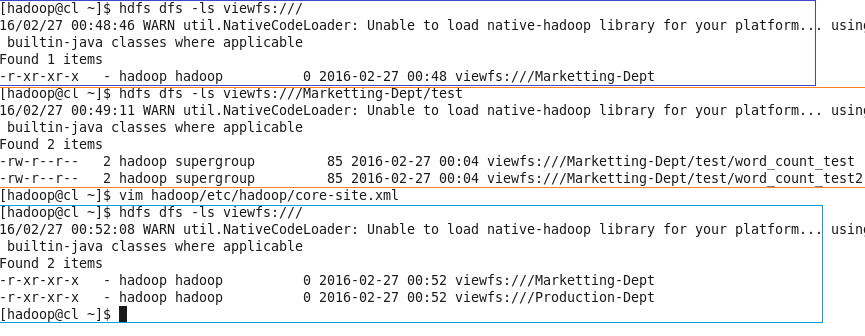
**Namenode-2 Format**



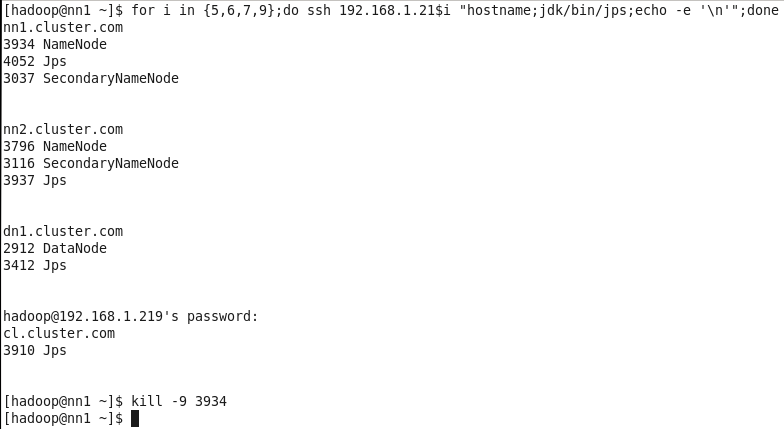
**Nodes/Daemons**



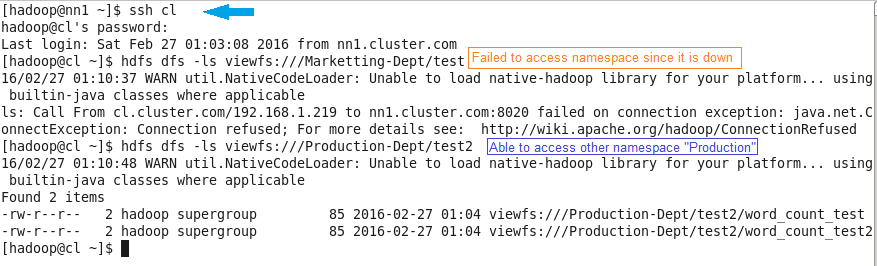
**Client view: (limited & overall access)**



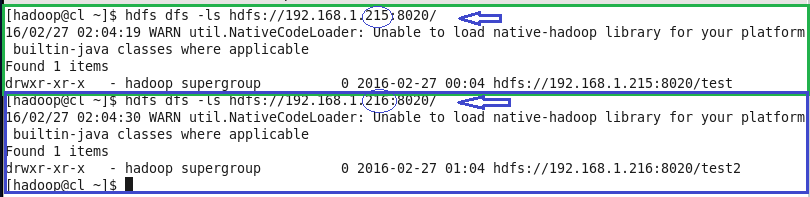
**Killing Namenode-1**



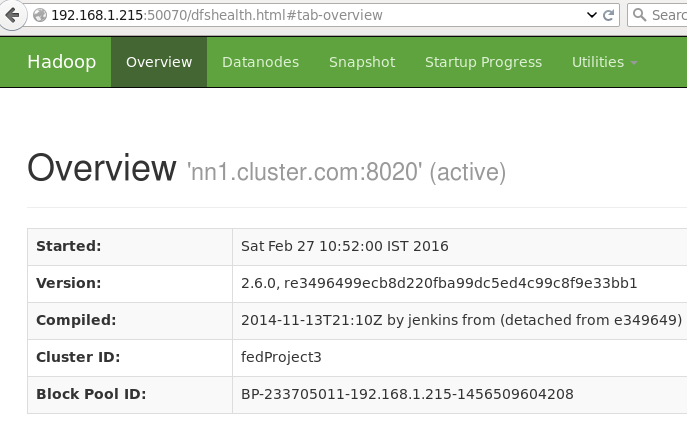
**Client can see only Namenode-2**



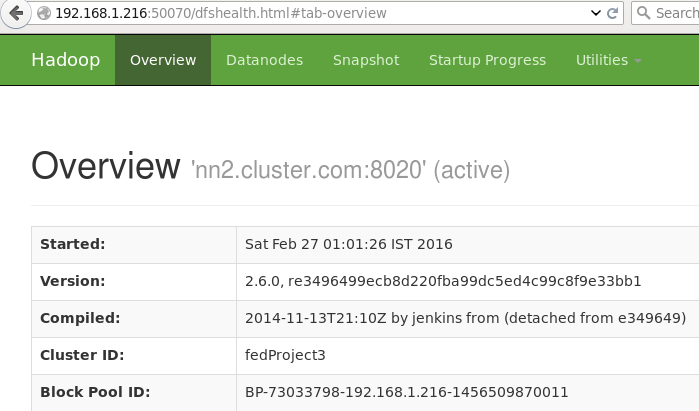
**Both Namenodes are up and running**



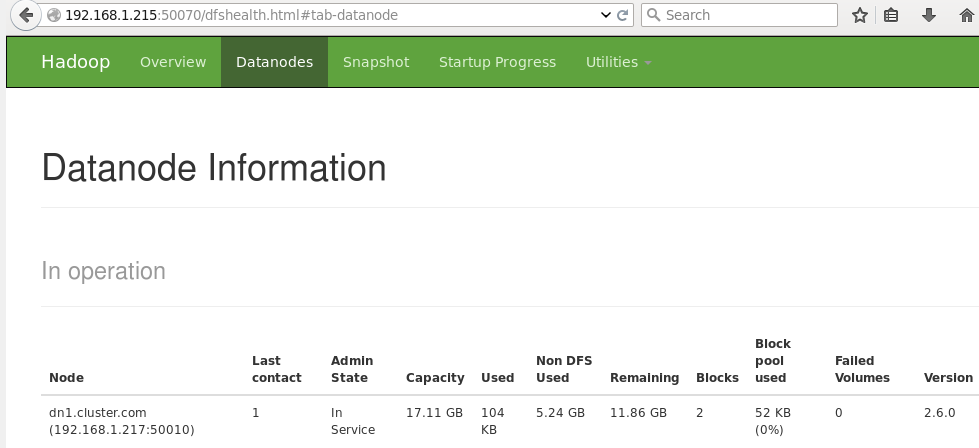
**Web Interface View - NN1**



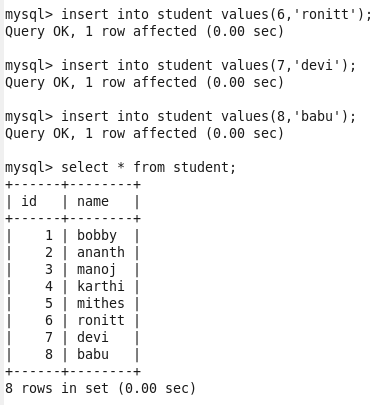
**Web Interface View - NN2**



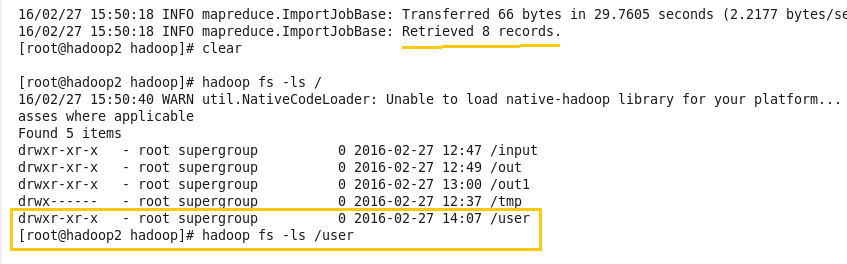
**Datanode**



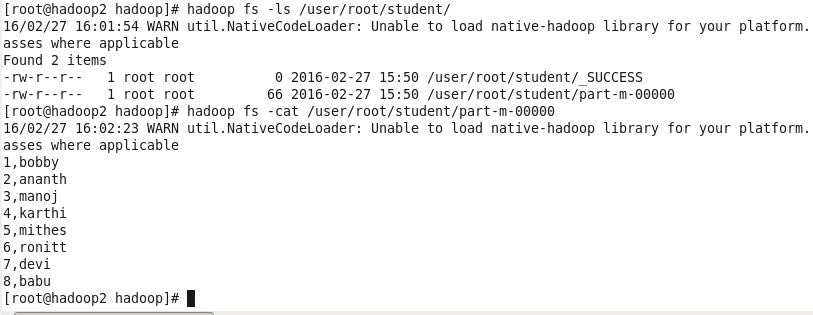
**Sqoop**



**Full table import**



**Full table result via output file**



**With filter result (i.e id > 4)**

