Table of Contents

[1. Infrastructure 4](#_Toc482897499)

[2. System Architecture 5](#_Toc482897500)

[a. Cluster Sizing 7](#_Toc482897501)

[b. Network Design 7](#_Toc482897502)

[3. Operating System 8](#_Toc482897503)

[a. Version 8](#_Toc482897504)

[b. Hostname convention 8](#_Toc482897505)

[c. DNS resolution 8](#_Toc482897506)

[d. OS Base Repo Access 8](#_Toc482897507)

[e. Java Version 8](#_Toc482897508)

[f. NTP Configuration 8](#_Toc482897509)

[g. Kernel and OS Tuning 8](#_Toc482897510)

[i. Swappiness 8](#_Toc482897511)

[ii. iptables 9](#_Toc482897512)

[iii. SELinux 9](#_Toc482897513)

[h. Disks layout, format and mount options 9](#_Toc482897514)

[4. Backend Database 10](#_Toc482897515)

[5. Hadoop Deployment Strategy 10](#_Toc482897516)

[a. Master Nodes 10](#_Toc482897517)

[b. Utility Nodes 10](#_Toc482897518)

[c. Slave Nodes 10](#_Toc482897519)

[d. Edge Nodes 10](#_Toc482897520)

[e. Component placement 10](#_Toc482897521)

[f. Server and Network Topology 11](#_Toc482897522)

[g. Fault Tolerance / High Availability 11](#_Toc482897523)

[h. Cluster configuration 11](#_Toc482897524)

[i. Memory Allocation 11](#_Toc482897525)

[ii. HDFS 11](#_Toc482897526)

[iii. YARN 11](#_Toc482897527)

[iv. Map Reduce 11](#_Toc482897528)

[v. Impala 11](#_Toc482897529)

[vi. HBase 11](#_Toc482897530)

[vii. Resource Scheduling 11](#_Toc482897531)

[viii. Static Service Pools 11](#_Toc482897532)

[ix. Dynamic Resource Pools 12](#_Toc482897533)

[6. Security 12](#_Toc482897534)

[a. Secure Access / Firewalls 12](#_Toc482897535)

[b. Authentication 13](#_Toc482897536)

[i. Users and Actors 13](#_Toc482897537)

[ii. Server Access 13](#_Toc482897538)

[iii. Web Interfaces 13](#_Toc482897539)

[c. Kerberos 13](#_Toc482897540)

[d. Authorization 13](#_Toc482897541)

[i. HDFS ACLs 13](#_Toc482897542)

[ii. Linux Filesystem ACLs 13](#_Toc482897543)

[e. Accounting 13](#_Toc482897544)

[i. Access and Audit Logs 13](#_Toc482897545)

[f. Cloudera Navigator 14](#_Toc482897546)

[g. Encryption 14](#_Toc482897547)

[i. TLS for Cloudera Management Traffic 14](#_Toc482897548)

[iii. HDFS Transparent Encryption 14](#_Toc482897549)

[7. Monitoring 14](#_Toc482897550)

[a. Cloudera Manager 14](#_Toc482897551)

[b. Alerts 15](#_Toc482897552)

[8. Backup Strategy 15](#_Toc482897553)

[a. Databases 15](#_Toc482897554)

[b. HDFS FSImage 15](#_Toc482897555)

[c. OS/Bedrock/Config files/Scripts 15](#_Toc482897556)

[9. Roles & Responsibility 15](#_Toc482897557)

[10. Appendix 16](#_Toc482897558)

[a. Filesystem & Partition Recommendation: 16](#_Toc482897559)

[i. Production 16](#_Toc482897560)

[ii. Dev : 17](#_Toc482897561)

[b. High Availability : 19](#_Toc482897562)

[c. Software Packages : 19](#_Toc482897563)

[11. Checklist 20](#_Toc482897564)

[12. WEB INTERFACES ON HADOOP ENVIRONMENT: 21](#_Toc482897565)

[a. Cloudera Manager: 21](#_Toc482897566)

[b. Resource Manager: 21](#_Toc482897567)

[c. HUE: 21](#_Toc482897568)

[13. CLOUDERA MANAGER: 22](#_Toc482897569)

[Architecture 22](#_Toc482897570)

[Cloudera manger URL : 24](#_Toc482897571)

[Start/Stop, check status of cloudera manager server/agents: 24](#_Toc482897572)

[Cloudera-Manager-Server 24](#_Toc482897573)

[Cloudera-Agent 24](#_Toc482897574)

[Logs for Cloudera manager server/agents: 25](#_Toc482897575)

[Cloudera-Manager-Server 25](#_Toc482897576)

[Cloudera-Agent Logs 25](#_Toc482897577)

[Integrating Cloudera Manager with Active Directory 25](#_Toc482897578)

[14. Server Hostname: 26](#_Toc482897579)

[15. Integrating servers to active directory 26](#_Toc482897580)

[a. Creating a DNS entry 26](#_Toc482897581)

[b. Joining Linux servers to AD 30](#_Toc482897582)

[16. Active Directory structure for Big data OU 33](#_Toc482897583)

[Service Account for Kerberos Integration to Active Directory 34](#_Toc482897584)

[17. Web Addresess from where Packages Needs to be Downloaded 38](#_Toc482897585)

[18. Hosting an Internal Repo using Apache Web Server 39](#_Toc482897586)

[Starting a Apache http web-server: 40](#_Toc482897587)

[19. Temporary Network re-architecture around 1G compatible TOR 41](#_Toc482897588)

# Infrastructure

Below hardware specs are provided by Millicom team :

|  |  |
| --- | --- |
| **MASTER NODE** | |
| **Base Configuration** | |
| BC1M22HGSC | RH1288 V3 (8\*2.5inch HDD Chassis)(Only for oversea,except Japan)H12M-03 |
| BC1M01FGEB | SM211 Onboard NIC,2xGE Electrical Interface(I350),RJ45 |
| BC1M13RISE | PCIe Riser Card,1 slot(x16),RISER1,used for RH1288 V3 |
| BC1M14RISE | PCIe Riser Card,1 slot(x16),RISER2,used for RH1288 V3 |
| EN4MCACC | 800W DC Power Supply Module (including DC power line) |
| **Haswell EP CPU** | |
| BC1M28CPU | Intel Xeon E5-2640 v3(2.6GHz/8-core/20MB/90W) Processor (with heatsink) |
| **DDR4 Memory** | |
| N24DDR403 | DDR4 RDIMM Memory,32GB,2400MT/s,2Rank(2G\*4bit),1.2V,ECC |
| **Hard Disk(include Front Panel)** | |
| N900S1210W2 | HDD,900GB,SAS 12Gb/s,10K rpm,128MB or above,2.5inch(2.5inch Drive Bay) |
| **RAID Card and Other Accessories** | |
| BC1M52ESMQ | SR320BC(LSI2208) SAS/SATA RAID Card,RAID0,1,5,6,10,50,60,6Gb/s,1GB Cache,used for RH1288 V3's 8HDD chassis |
| BC1M2BAT | LSI2208 RAID Card SuperCap(include cable,bracket),used for rack servers |
| **PCI-E Card** | |
| CN2ITGAA20 | Ethernet Adapter,10Gb Optical Interface(Intel 82599),2-Port,SFP+(with 2x Multi-mode Optical Transceiver),PCIe 2.0 x8 |
| **Cables and Optional Equipments** | |
| B1UBBRK | 1U Ball Bearing Rail Kit |
| B1U2UCMA | 1U/2U Cable Management Arm |

|  |  |
| --- | --- |
| **DATA NODE** | |
| **Base Configuration** | |
| BC4M20HGSB | RH2288 V3 (12\*3.5inch HDD Chassis)(Only for oversea,except Japan)H22M-03 |
| BC1M01FGEB | SM211 Onboard NIC,2xGE Electrical Interface(I350),RJ45 |
| BC1M06FAN | 8038 Fan module |
| BC1M15IHDD | 2\*2.5" Rear Hard Disk Backplane+2\*3.5" Rear Hard Disk Backplane Module |
| EN4MCACC | 800W DC Power Supply Module (including DC power line) |
| **Haswell EP CPU** | |
| BC1M13CPU | Intel Xeon E5-2660 v3(2.6GHz/10-core/25MB/105W) Processor (with heatsink) |
| **DDR4 Memory** | |
| N24DDR402 | DDR4 RDIMM Memory,16GB,2400MT/s,2Rank(1G\*8bit),1.2V,ECC |
| **Hard Disk(include Front Panel)** | |
| N4000ST7W3 | HDD,4000GB,SATA 6Gb/s,7.2K rpm,64MB,3.5inch(3.5inch Drive Bay) |
| N600S1210W2 | HDD,600GB,SAS 12Gb/s,10K rpm,128MB or above,2.5inch(2.5inch Drive Bay) |
| **RAID Card and Other Accessories** | |
| BC1M01ESMQ | SR430C(LSI3108) SAS/SATA RAID Card,RAID0,1,5,6,10,50,60,12Gb/s,1GB Cache,used for RH2288 V3/RH2288H V3's 12&25HDD chassis/5288 V3 |
| BC1M01TFM | LSI3108 1GB RAID Card SuperCap(4GB,include cable,bracket),used for rack servers/X6800 |
| **PCI-E Card** | |
| CN2ITGAA20 | Ethernet Adapter,10Gb Optical Interface(Intel 82599),2-Port,SFP+(with 2x Multi-mode Optical Transceiver),PCIe 2.0 x8 |
| **Cables and Optional Equipments** | |
| E2USRAIL01 | 2U Ball Bearing Rail Kit |
| B1U2UCMA | 1U/2U Cable Management Arm |

Network equipment and other cluster related accessories are called out in the below reference document shared by Millicom team.

Reference Document : <https://docs.google.com/spreadsheets/d/14dywjrxvFJtH9sOCOXAwwY4oOF6wQ1kc_jK9ZkP3JcA/edit#gid=710765837>

# System Architecture

This section describes, at a high level, Hadoop cluster system architecture.

In order to meet the initial infrastructure requirements of holding 26 TB for a 13 month period Zaloni’s recommendation for the infrastructure is a cluster of servers referred to as the Master pod which has:

1 Admin/Edge Pod consisting of 2 master/edge nodes

1 Data Pod consisting of 4 data/worker nodes.

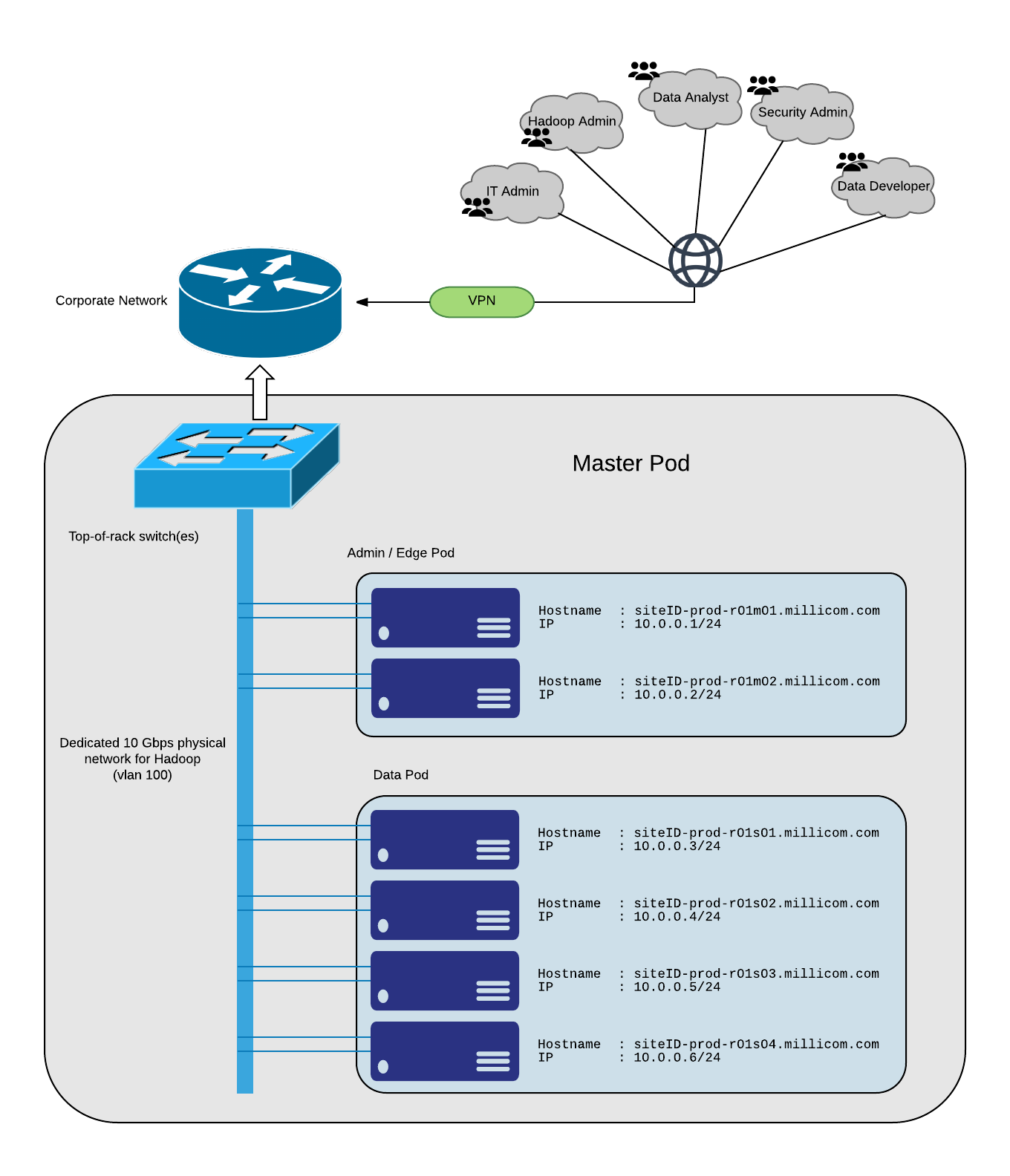
The Master pod is recommended as the starting point for each country implementation. As the data storage needs and ingestion rates increase, scaling is achieved by adding Data pods and/or Admin/Edge pods. Note that high availability is currently not a requirement. But if high availability is needed, a third master node would be added.

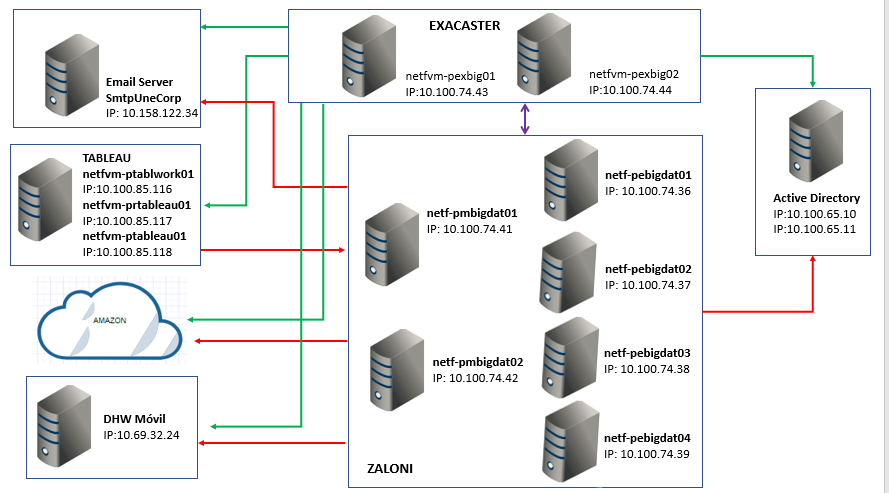
Scaling in increments of a pod provides the pre-designed black box view of infrastructure and helps achieve implementation and cost efficiency.

Cloudera Hadoop distribution will be the Hadoop platform software running on the servers.

Bedrock application software will reside on the edge nodes.The edge nodes are also referred to as landing nodes since incoming data may be dropped on them.

The following diagram illustrates the concept of an infrastructure Master pod specification.





Cluster sizing is done to meet the current data storage requirement and also taking into consideration:

* additional storage needed for compute purposes such as shuffling, temporary or intermediate data
* room for growth, while still being cost efficient over the time horizon
* better parallelization of jobs is ensured with at least 4 nodes

Assuming the default replication of each block of storage in a Hadoop cluster is 3 times,

the initial pod capacity will be (4 nodes x 12 disks x 4 TB per disk) / 3x replication factor = 64 TB

This sizing meets the current requirement of around 26 TB for a 13 month period while also taking into consideration:

- additional storage needed for compute purposes such as shuffling, temporary or intermediate data

- room for growth, while still being cost efficient over the time horizon

better parallelization of jobs is ensured with at least 4 nodes

## Cluster Sizing

The Data Lake is designed to hold 9 month’s worth of structured data needed for the data science team, with a current data window of 13 months. Data storage requirement (raw data and any derived tables) is calculated as:

15 TB in 9 months => 23 TB in 13 months (based on current QuickWins project)

The cluster contains 4 data/slave nodes. The table below outlines available Hadoop storage based on the number of data nodes:

|  |  |
| --- | --- |
| **Total Raw Storage over 4 data nodes** | 4 nodes x 12 disks x 3.6\* TB each =~ 172 TB |
| Total Intermediate Storage (Configurable) | 0.25 x 172 TB = 43 TB |
| HDFS Available Raw Storage (Configurable) | 0.75 x 172 TB = 129 |
| HDFS Usable Storage (Configurable)  (with a replication factor of 3) | 129 TB / 3 = 43 TB |

\* Like with any HDDs, entire advertised space of a disk is not usable. A small fraction becomes unusable/reserved after partitioning and formatting. Typically, only 3.6 TB is usable on a 4 TB disk.

Note:

The data could be accumulated beyond 13 months and the Data Lake scaled as needed.

## Network Design

Network is a critical resource for a Hadoop cluster. High network bandwidth is used during the shuffle phase of MapReduce, during data replication etc. For this reason, we chose to have a 10 Gbps network for the Hadoop cluster with a dedicated Top-of-Rack switch. Servers are configured with NIC bonding to provide increased bandwidth and fault tolerance .The ToR Switch also has redundant power supplies for fault tolerance.

# Operating System

## Version

Based on Cloudera’s recommendation for CDH X.X TBD, and Millicom corporate server OS policy, Red Hat Enterprise Linux Server release 6.9is installed on all the cluster nodes.

## Hostname convention

For easy recognition of roles and physical connectivity a hostname convention comprising of the project name identifier, followed by the rack ID, the machine class, and a machine ID is used.

For example, siteID-prod-r01m01 represents a particular site’s production cluster, rack 1, master node 1

This convention is also useful when configuring rack awareness within Cloudera Manager

## DNS resolution

All hosts in the cluster must have DNS resolution configured and be able perform both forward and reverse lookups on other nodes in the cluster. DNS resolution is important to avoid running into complex issues when using hosts file based name resolution.

TBD: Add DNS server info

## OS Base Repo Access

Several enterprise linux administration packages and Cloudera Software depend on RPMs available from both the CentOS Base as well as the EPEL (Extra Packages for Enterprise Linux) Repositories.

TBD: Add repo approach - use internal satellite vs setup offline repo vs proxy to internet repo

Proxy Server : Proxy allows to download the necessary packages that will be required for the initial setup and or for any future updates .

## Java Version

Both Cloudera software and Bedrock Data Management Platform require JDK to run. Oracle JDK version 1.7.0\_67 is installed in all nodes at /usr/java/jdk1.7.0\_67/

## NTP Configuration

All machines in the cluster need to have the same time and date settings, including time zones. Network Time Protocol will be used to ensure all servers are in time sync.

TBD: Add Millicom NTP server info

## Kernel and OS Tuning

### Swappiness

Hadoop depends on heavy usage of RAM resources on machines in the cluster. High system swapping has very undesirable effects when there are huge chunks of data being swapped in and out of RAM. Some services can potentially be marked as down if delays are present due to excessive paging. So, the kernel swappiness will be set to 1 (lowest on a scale of 1 to 100)

### iptables

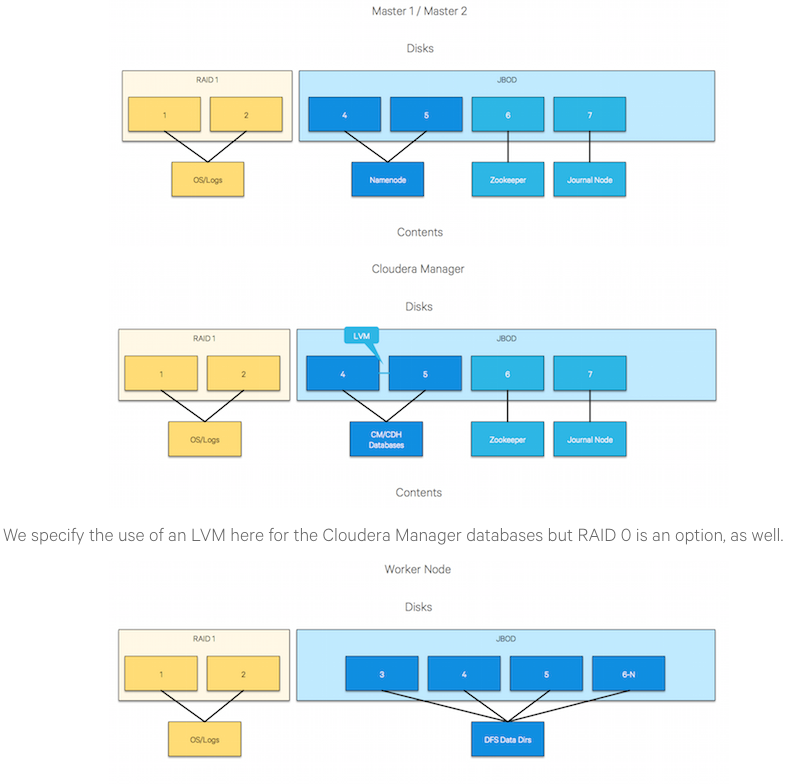
Host based firewalls (iptables) will be disabled to reduce computational load arising from policy enforcement. Hardware based network firewalls can be leveraged for restricting access to the cluster

### SELinux

SELinux must be disabled on each host before deploying CDH on your cluster as there are no SELinux Policies corresponding to Hadoop’s access patterns

## Disks layout, format and mount options

Apendix 10.1 provides details of recommended formatting & partitioning for Millicom .



# Backend Database

MySQL or TBD is used as backend for Cloudera Manager and various other CDH components that store metadata in relational databases. Bedrock also leverages Mysql as its backend.

TBD: can include custom config for cdh, bedrock

# Hadoop Deployment Strategy

All of the cluster nodes will be running Cloudera’s Distribution of Hadoop (CDH). CDH is the most complete, tested, and popular distribution of Apache Hadoop and related projects. CDH delivers the core elements of Hadoop – scalable storage and distributed computing – along with a Web-based user interface and vital enterprise capabilities

CDH version TBD XX will be installed, administered and configured by Cloudera Manager version TBD XX. As such, all nodes in the cluster will be running the Cloudera Manager Agent process.

## Master Nodes

Master hosts run Hadoop master processes such as the HDFS NameNode and YARN Resource Manager. Master2 is used as our Master node

## Utility Nodes

Utility hosts run other cluster processes that are not master processes such as Cloudera Manager and the Hive Metastore. Master1 plays dual role in our cluster running both master processes and utility processes.

## Slave Nodes

Worker hosts primarily run DataNodes and other distributed processes such as Impalad or HBase Region Servers

## Edge Nodes

Edge hosts are client access points for launching jobs in the cluster.

## Component placement

|  |  |  |
| --- | --- | --- |
| Server | Cloudera Components | Other software |
| Master1/Utility/Edge? | Cloudera Manager + Other management services, Hue, Impala Catalog Server,  Spark History Server, Oozie, Sentry, Solr, Sqoop, Zookeeper,  HBase (Master, Thrift server, REST server),  Secondary Namenode, HDFS Balancer, NFS Gateway,  Hive Server2, Hive Metastore, Hive WebHCAT,  Client Gateways | MySQL |
| Master2/Edge? | Namenode, Impala State Store, Resource Manager, Job History Server, SCM Agent | Bedrock, MySQL |
|  |  |  |
| Slave1 | SCM Agent, Region Server, Impala Daemon, Data Node, Node Manager, Client Gateways |  |
| Slave2 | SCM Agent, Region Server, Impala Daemon, Data Node, Node Manager, Client Gateways |  |
| Slave3 | SCM Agent, Region Server, Impala Daemon, Data Node, Node Manager, Client Gateways |  |
| Slave4/Edge? | SCM Agent, Region Server, Impala Daemon, Data Node, Node Manager, Client Gateways | VNC? |

## Server and Network Topology

All servers are physically stacked in one rack in the datacenter with a not dedicated 10 Gbps capable Nexus5k Switch. All cluster nodes will be in a single subnet / vlan. Each server has redundant uplinks to the switch for fault tolerance and load sharing.

TBD: Insert diagram ( Final Diagram to be inserted by Millicom/Huawei team)

## Fault Tolerance / High Availability

The current deployment does not implement any HA functionality but has provisions to be fault tolerant with the addition of subsequent worker pods when scaling out the data lake infrastructure.

## Cluster configuration

### Memory Allocation

### HDFS

FSImage storage

NN and DN Heap sizes

Replication Factor

### YARN

Container allocation

### Map Reduce

Map and Reduce Tasks’ allocation

### Impala

### HBase

## Resource Scheduling

### Static Service Pools

Static service pools serve to isolate services from each other, so that high load on one service has bounded impact on other services.

|  |  |
| --- | --- |
| Service | Allocation % |
| HBase | 10 |
| HDFS | 10 |
| Impala | 20 |
| Solr | 0 |
| YARN (MR2 Included) | 60 |

### Dynamic Resource Pools

YARN and Impala Static Service Pools can each have sub-pools, that can provide different amount of limits, and can be configured to allow only a certain set of users and groups to access the pool.

#### YARN

|  |  |  |
| --- | --- | --- |
| Resource Pool | Weight | Scheduling Policy |
| default | 1 | Dominant Resource Fairness (recommended) |
| users | 1 |  |
| tenant1 | 5 |  |
| tenant 2 | 5 |  |

#### Impala Admission Control

Queues and resource allocation

|  |  |  |
| --- | --- | --- |
| Resource Pool | Max memory | Max memory per query |
| default | 10 GB | 5 GB |
| tenant1 | 50 GB | 10 GB |

# Security

## Secure Access / Firewalls

Hadoop Web interfaces that need to be accessible externally:

* Cloudera Manager
* Cloudera Navigator
* Bedrock
* Namenode
* Resource Manager
* Job History
* Application Master
* Hue
* Spark History
* Oozie Workflow Manager

Other communication channels required

* SSH
* AD
* NTP
* Kerberos
* Data source systems

## Authentication

### Users and Actors

TBD: Define user categories and roles

### Server Access

TBD: Ssh/vnc solution; centrify/sssd

### Web Interfaces

Web interfaces like Cloudera Manager, Cloudera Navigator, Hue and Bedrock will be integrated to an LDAP server. Users can login to these services using their enterprise account credentials

## Kerberos

Hadoop services will be configured with Kerberos which is a strong authentication mechanism and is the only authentication process supported by all CDH services

## Authorization

### HDFS ACLs

### Linux Filesystem ACLs

## Accounting

### Access and Audit Logs

The Logs page on Cloudera Manager presents log information for Hadoop services, filtered by service, role, host, or search phrase as well log level (severity). Logs are also available on disk under /var/log directory on each server. The servers has log dirs specific to each CDH service running on them.

* + 1. List of log dirs:
* /var/log/catalogd
* /var/log/cloudera-scm-alertpublisher
* /var/log/cloudera-scm-eventserver
* /var/log/cloudera-scm-firehose
* /var/log/cloudera-scm-headlamp
* /var/log/cloudera-scm-navigator
* /var/log/hadoop-hdfs
* /var/log/hadoop-httpfs
* /var/log/hadoop-mapreduce
* /var/log/hadoop-yarn
* /var/log/hbase
* /var/log/hcatalog
* /var/log/hive
* /var/log/hue
* /var/log/hue-httpd
* /var/log/impala-llama
* /var/log/impalad
* /var/log/oozie
* /var/log/sentry
* /var/log/solr
* /var/log/spark
* /var/log/sqoop2
* /var/log/statestore
* /var/log/zookeeper

### Cloudera Navigator

Cloudera Navigator is a fully integrated data-management and security system for the Hadoop platform. It provides visibility into and control over the data in Hadoop datastores, and the computations performed on that data. Hadoop administrators, data stewards, and data scientists can use Cloudera Navigator to Audit data access and verify access privileges , Search metadata and visualize lineage, examine data usage patterns and create policies based on those patterns etc.

## Encryption

### TLS for Cloudera Management Traffic

Min level1

* + 1. TLS for CDH Services

For API calls, inter service comm etc

### HDFS Transparent Encryption

Needs KMS license

# Monitoring

## Cloudera Manager

Cloudera Manager Admin Console displays cluster health, metrics, and usage, view processing activities, and view events, logs, and reports to troubleshoot problems and monitor compliance.

NOTE : As the cluster is deployed using Cloudera Manager, admins cannot use use standard Hadoop command-line utilities to start and stop services or change config files directly.

TBD: Add Screenshot

## Alerts

Service instances of type HDFS, MapReduce, and HBase (and their associated roles) can generate alerts if so configured. Alerts can also be configured for the monitoring roles that are a part of the Cloudera Management Service.Alert Publisher can be configured to send alert notifications by email or by SNMP trap to a trap receiver.

TBD: Screenshot and SMTP server info

# Backup Strategy

## Databases

## HDFS FSImage

## OS/Bedrock/Config files/Scripts

# Roles & Responsibility

|  |  |  |
| --- | --- | --- |
| **Task** | **Responsibility** | **Point of Contact** |
| Virtual Machine(s) Setup for Dev Cluster | ? |  |
| Hardware/Cabling | ? |  |
| Network Configuration & Setup | ? |  |
| OS Setup | ? |  |
| Hadoop & Datalake Setup | Zaloni | Yin/Ranjith |

# Appendix

## Filesystem & Partition Recommendation:

Below are recommendations and partition sizes and/or mount points might get modified after design discovery phase .

### Production

Below are recommended partitions and partition size per server . This does not account for any mandates or recommendation that Hardware Vendor might require .

**Table 1 : Master Node:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Mount Point** | **Size(GB)** | **RAW Disks** | **RAID** |
| / | 50 | 2 | RAID 1 |
| /dev | 32 |
| /boot | 1 |
| /opt | 250 |
| /var | 250 |
| /tmp | 200 |
| /grid/0 | 1800 | 4 | RAID 10 |
| /master/0 | 900 GB | 1 | ~~n/a~~  **RAID 0** |
| /master/1 | 900 GB | 1 | ~~n/a~~ **RAID 0** |
| NOTES:  Use ext4 formatting for all partitions.  For /grid/\* and /master/\* partitions :  Do not use Linux Logical Volume Manager (LVM) | | | |

**Table 2 : Slave Node:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Mount Point** | **Size(GB)** | **RAW Disks** | **RAID** |
| / | 30 | 2 | RAID 1 |
| /dev | 40 |
| /boot | 1 |
| /opt | 120 |
| /var | 150 |
| /tmp | 175 |
| /grid/0 | 4 TB | 1 | n/a |
| /grid/1 | 4 TB | 1 | n/a |
| /grid/2 | 4 TB | 1 | n/a |
| /grid/3 | 4 TB | 1 | n/a |
| /grid/4 | 4 TB | 1 | n/a |
| /grid/5 | 4 TB | 1 | n/a |
| /grid/6 | 4 TB | 1 | n/a |
| /grid/7 | 4 TB | 1 | n/a |
| /grid/8 | 4 TB | 1 | n/a |
| /grid/9 | 4 TB | 1 | n/a |
| /grid/10 | 4 TB | 1 | n/a |
| /grid/11 | 4 TB | 1 | n/a |
| NOTES:  Use ext4 formatting for all partitions.  Since fdisk cannot partition disks 2TB and higher , use “parted” command to partition the 4TB disks, below are the instructions :   * On a new drive :   + “parted /dev/<disk-name> mklabel gpt” { eg . parted /dev/sda mklabel gpt } * Partition the complete disk as 1 single partition :   + “parted –a opt /dev/<disk-name> mkpart primary ext4 0% 100%” * Verify if the partition got created using :   + “lsblk” * Format the partition using ext4 fs :   + “mkfs.ext4 /dev/<partition-name>” { make sure to use partition name, usually a partition has a number to the end of disk name eg. /dev/sda1} * Verify if the format went fine using :   + lsblk --fs command . It should show file-system format next to partition.   For /grid/\* partitions :  Do not use Linux Logical Volume Manager (LVM)  Use noatime mount option | | | |
|  | | | |

### Dev :

Below is the recommended partitions and their corresponding size for virtual setup.

|  |  |
| --- | --- |
| Node | Count |
| Master Nodes | 1 |
| Slave Nodes | 4 |
| Edge Node | 1 |

|  |  |
| --- | --- |
| **Master Node : File-System Sizes** | |
| **Mount Point** | **Size(GB)** |
| / | 50 |
| /opt | 1 TB |
| /var | 1 TB |
| /tmp | 200 |
| /grid/0 | 200 |
|  |  |
| RAM | 48 |
| V-CORES | 6 |

|  |  |
| --- | --- |
| **Slave Node : File-System Sizes** | |
| **Mount Point** | **Size(GB)** |
| **/** | **50** |
| /opt | 150 |
| /var | 150 |
| /tmp | 200 |
| /grid/0 | 4 TB |
|  |  |
| RAM | 32 |
| V-CORES | 6 |

|  |  |
| --- | --- |
| **Edge Node : File-System Sizes** | |
| **Mount Point** | **Size(GB)** |
| **/** | **50** |
| /opt | 1 TB |
| /var | 1 TB |
| /tmp | 200 |
| /Landing | 2 TB |
|  |  |
| RAM | 56 |
| V-CORE | 8 |

## High Availability :

Present deployment does not required deployment of a highly available cluster . As the platform matures need for HA will arise .

To enable HA one more master node needs to be deployed and to have better fault tolerance all the servers Master & Slave nodes need to be balanced across 3 Racks ( Each rack hosting 1 Master node) .

Software Configuration : Below are few components that can be configured for HA, below list is not exhaustive , other components can be made HA and can be discussed when HA is pursued .

* HDFS
* YARN(MRv2)
* Hiveserver2

## Software Packages :

Below is the recommended OS version and packages that can be deployed on the servers :

1. Need CentOS 6.7 on all servers
2. Need “Desktop” package group on Edge Node
3. Need “Basic Server” package group on all other nodes

Download link for reference ( any mirror hosting CentOS 6.7 will work as well )

<http://archive.kernel.org/centos-vault/6.7/isos/x86_64/CentOS-6.7-x86_64-bin-DVD1.iso>

# Checklist

Below are the expected set of tasks/items to be completed by Millicom-IT/Huawei before CDH installation.

1. CentOS 6.7 is installed
2. Disk formatted and partitioned per recommendations per Apendix (Section 10.1 FileSystem & Partition Recommendation)
   1. For Slave/Data Nodes , /grid/\* partitions are mounted using “noatime” mount
3. After OS installation is complete :
   1. Set Swappiness to 1
   2. Iptables turned off
   3. SE-Linux turned off
   4. Hostnames & FQDN need to be setup , IP addresses, Routing tables setup
   5. Base and EPEL repos are available ( or proxy need to be setup to access the external repo’s )
   6. DNS ( forward and reverse look up’s should be functional) and NTP setup
   7. AD users added to Linux

# WEB INTERFACES ON HADOOP ENVIRONMENT:

Below are few Web interfaces that Hadoop Environment provides . These web interfaces can be broadly divided into 2 categories, one that is useful for “administrator” and second that is useful for “developers/analyst”

* Cloudera Manager [administrator]
* Resource Manager [administrator/developer/analyst]
* Hue [administrator/developer/analyst]
* Bedrock [administrator/developer/analyst]
* Exacaster

## Cloudera Manager:

Cloudera manager is cluster management tool and is primary tool to administrator the cluster . Developers or analyst have no use from this tool , hence no access needs to be given to developers/analyst .

Below is the URL where Cloudera Manager can be accessed :

URL for Cloudera Manager : <http://cm-pr-bgd-m01.telemovil.com.sv:7180/cmf/login>

## Resource Manager:

YARN resource manager is responsible for tracking resources in a cluster and scheduling applications (example, MR jobs).

Resource manager provides a web URL that can be access on port 8088 on Master 2 node .

URL for Resource Manager : <http://cm-pr-bgd-m02.telemovil.com.sv:8088/cluster>

## HUE:

Hue provides a web interface for many Hadoop related tools , such as :

1. Hive
2. Impala
3. Oozie … etc

Hue should be the primary tool to perform ad-hoc queries . Developers/Analyst can use various tools within Hue to interact with the environment . Once a model has been tested, the queries can be made part of a workflow using Bedrock tool .

URL for Hue Interface : <http://cm-pr-bgd-m01.telemovil.com.sv:8888/accounts/login/?next=/>

# CLOUDERA MANAGER:

Cloudera manager is cluster management tool and is primary tool for hadoop administrator on the cluster .

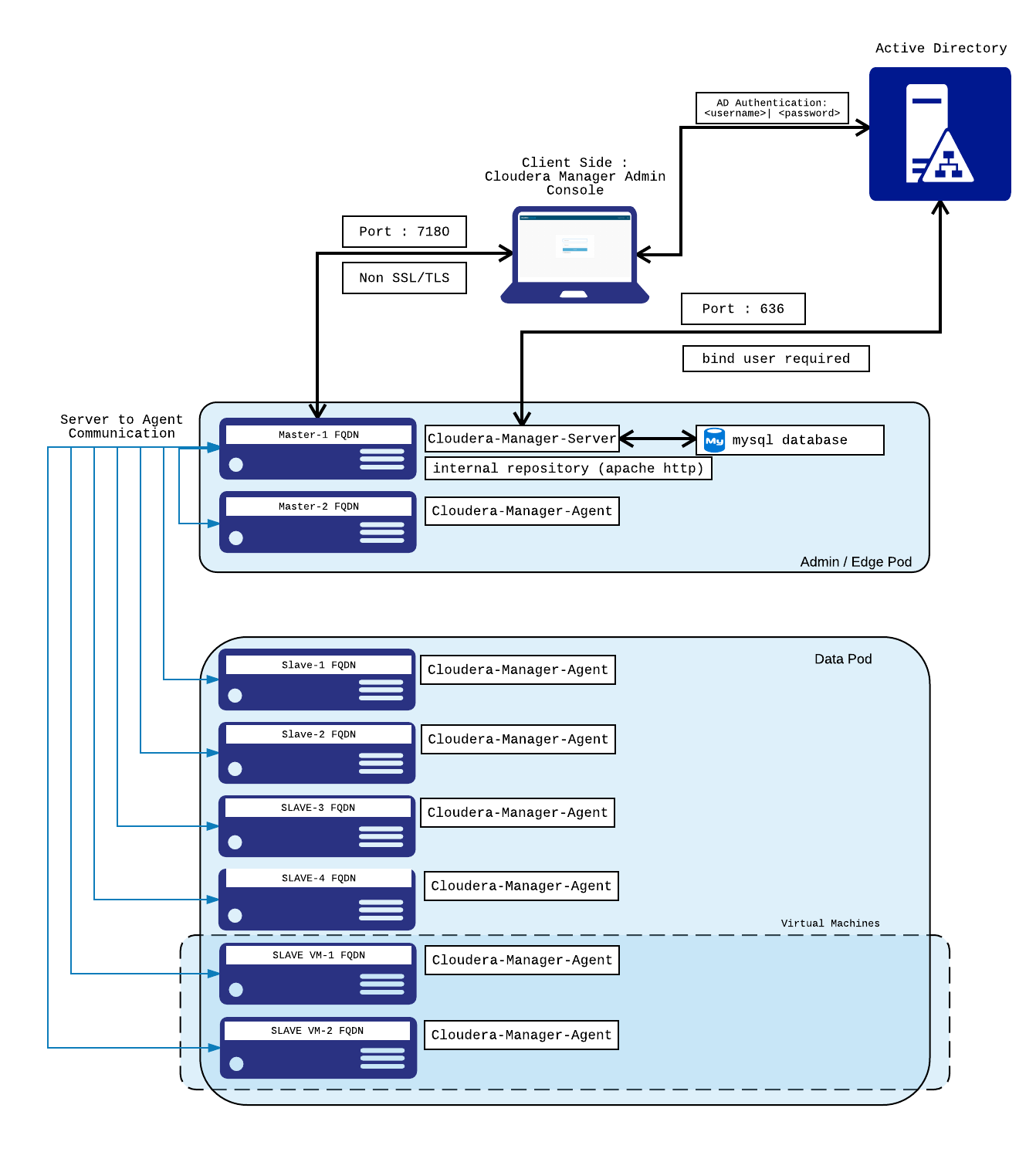
Cloudera Manager performs the below functions :

1. Tracks the Cloudera Manager data model, which is stored in the Cloudera Manager Server database. The data model is a catalog of the available host machines in the cluster, and the services, roles, and configurations assigned to each host.
2. Communicates with Agents to send configuration instructions and track Agents' heartbeats
3. Performs command execution to do tasks
4. Provides an Admin console for the operator to perform management and configuration tasks
5. Creates, reads, validates, updates, and deletes configuration settings
6. Calculates and displays the health of the cluster and its components
7. Tracks host metrics such as disk usage, CPU, and RAM
8. Provides a comprehensive set of APIs for the various features supported in Cloudera Manager
9. Manages Kerberos credentials
10. Monitors the health of Hadoop daemons, and dozens of service performance metrics, and alerts you when you approach critical thresholds.
11. Keeps a history of activity monitoring data and configuration changes

Architecture :



As detailed in the image above , following components will be deployed as below on Tigo BO infrastructure :



## Cloudera manger URL :

URL : TBD

## Start/Stop, check status of cloudera manager server/agents:

### Cloudera-Manager-Server

Start cloudera-manager-server :

Starting cloudera manager takes a bit of time even though after passing the below command it says ok the cloudera manager service is working in the background to prepare the “jetty” server. You can tail logs {as mentioned in log section below} and check if “Jetty” Server is ready. Once “Jetty” is up , you should now be able to access the cloudera-manager-admin console.

1. Login to cm-pr-bgd-m01.telemovil.com.sv host as “root” user
2. On Shell execute the below command

> service cloudera-scm-server start

Stop cloudera-manager-server:

1. Login to cm-pr-bgd-m01.telemovil.com.sv host as “root” user
2. On Shell execute the below command

> service cloudera-scm-server stop

Check cloudera-manager-server Status:

1. Login to cm-pr-bgd-m01.telemovil.com.sv host as “root” user
2. On Shell execute the below command

> service cloudera-scm-server status

### Cloudera-Agent

Start cloudera-agent:

1. Login to cm-pr-bgd-m02/ cm-pr-bgd-s[01-05].telemovil.com.sv.telemovil.com.sv/ biexacaster[1-2].telemovil.com.sv host as “root” user
2. On Shell execute the below command

> service cloudera-scm-agent start

Stop cloudera-agent:

1. Login to cm-pr-bgd-m02/ cm-pr-bgd-s[01-05].telemovil.com.sv.telemovil.com.sv/ biexacaster[1-2].telemovil.com.sv host as “root” user
2. On Shell execute the below command

> service cloudera-scm-agent stop

Check Cloudera-agent Status:

1. Login to cm-pr-bgd-m02/ cm-pr-bgd-s[01-05].telemovil.com.sv.telemovil.com.sv/ biexacaster[1-2].telemovil.com.sv host as “root” user
2. On Shell execute the below command

> service cloudera-scm-agent status

## Logs for Cloudera manager server/agents:

### Cloudera-Manager-Server

On cm-pr-bgd-m01.telemovil.com.sv path is as below:

> /var/log/cloudera-scm-server/ cloudera-scm-server.log

### Cloudera-Agent Logs

On cm-pr-bgd-m02/ cm-pr-bgd-s[01-05].telemovil.com.sv.telemovil.com.sv/ biexacaster[1-2].telemovil.com.sv path is as below :

> /var/log/cloudera-scm-agent/ cloudera-scm-agent.log

## Integrating Cloudera Manager with Active Directory

For Cloudera to connect to Active Directory, below are the pre-requisites:

1. Bind user
2. Security groups for different roles in cloudera manager
3. ldap URL with port ( usually if it’s a non secure protocol port defaults to 389 and if it’s a secured protocol it defaults to 636 ) . AD administrator should be able to provide this information.

# Server Hostname:

Hostname should be alphanumeric and in lower case. “\_” underscore is not allowed.

# Integrating servers to active directory

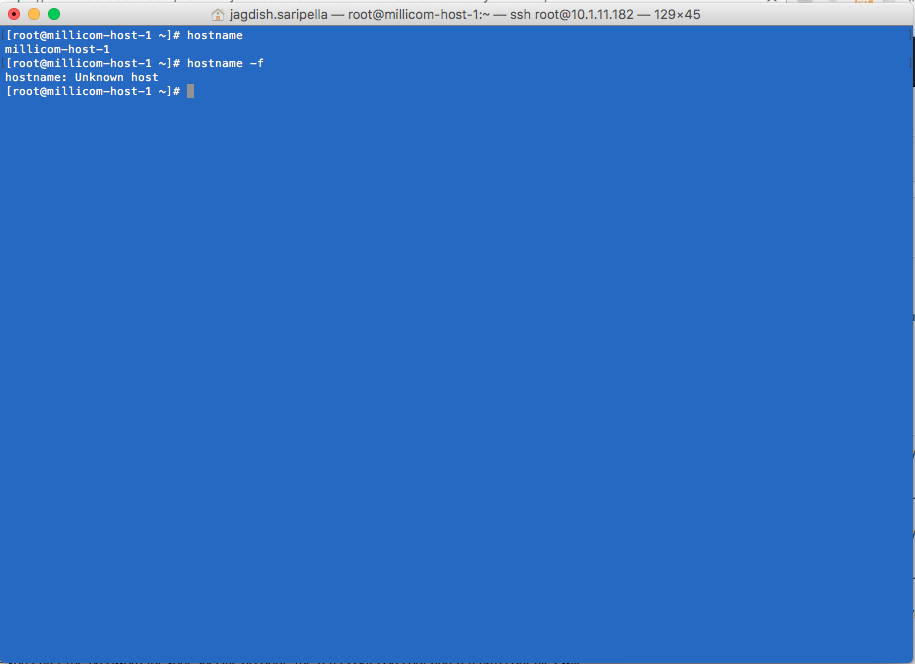
Below are few activities that are needed to declare a server has joined an Active Directory realm . Below are

* Creating DNS Entries : This enables a server to have a forward and a reverse lookup based on its fqdn .
* Integrating Linux Servers to Active Directory : Joining servers to corporate AD , allows users to authenticate against the corporate network .

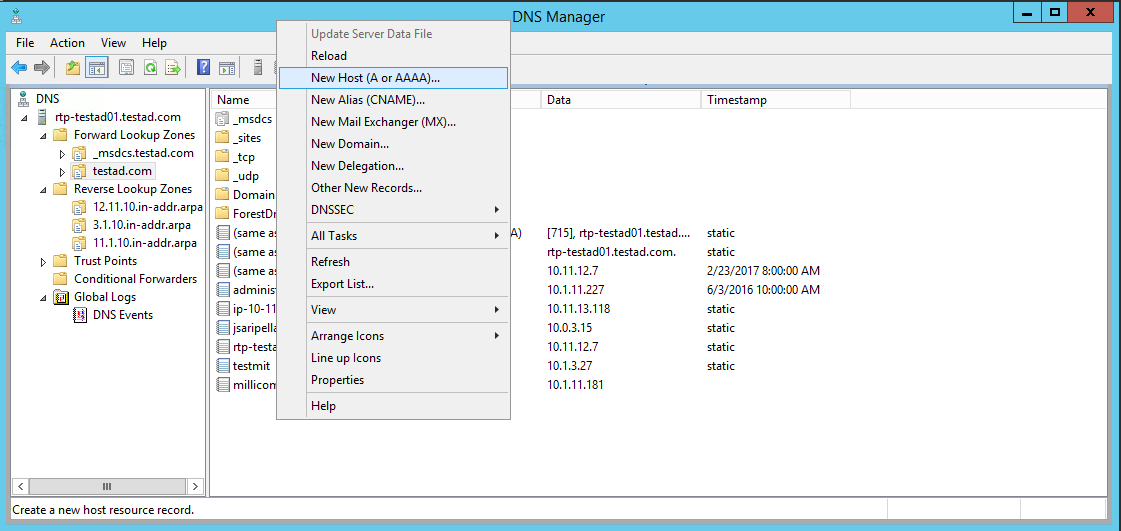
## Creating a DNS entry

Below example runs through the steps need to add a linux server to DNS server.

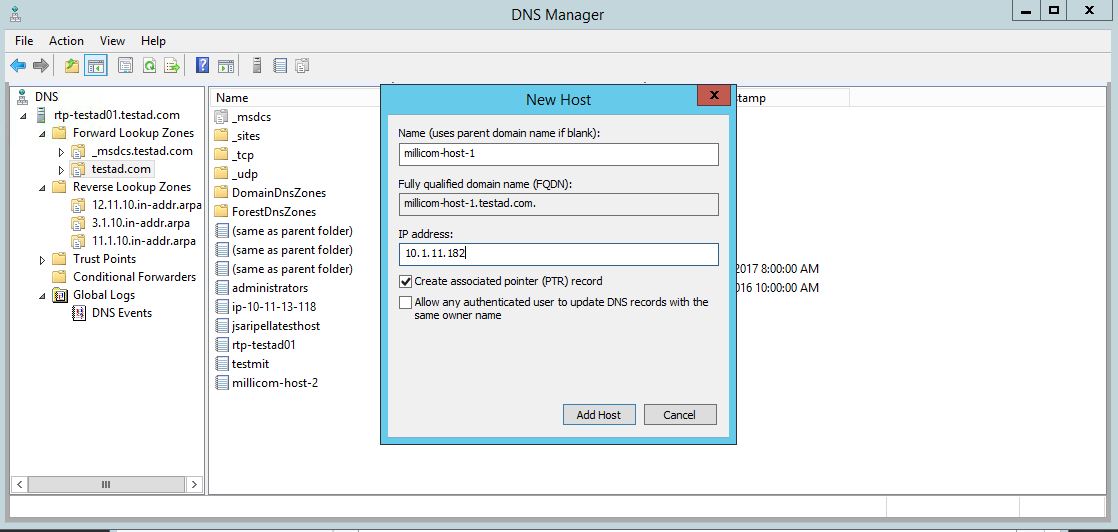
1. On a freshly installed CentOS host run the below commands :
2. “hostname” returns the hostname of the server
3. “hostname -f” does not return any entry . Chances are the server does not have an entry in DNS yet



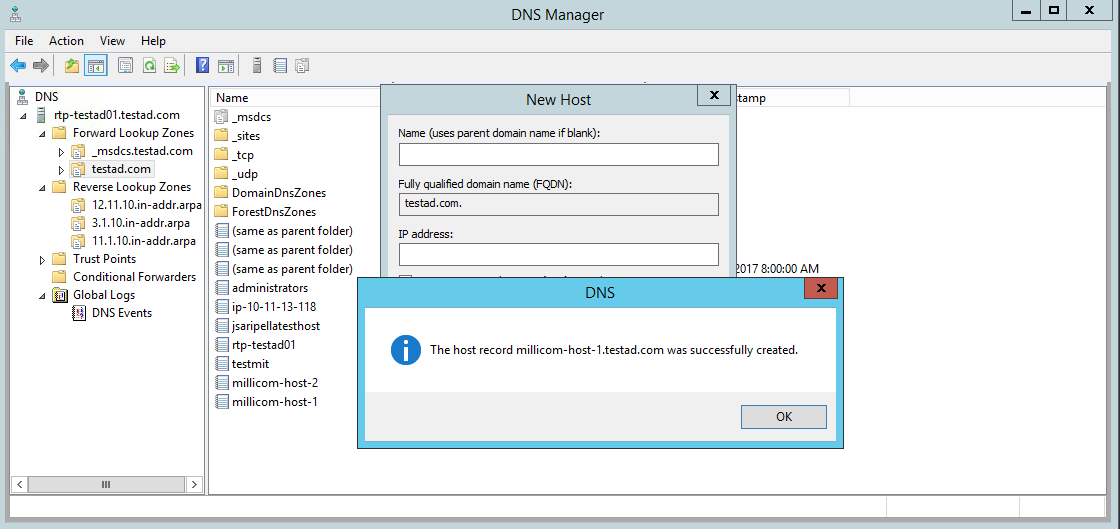
1. Login to the AD server and access the DNS manager console .
2. Under the domain , add a New Host as shown below



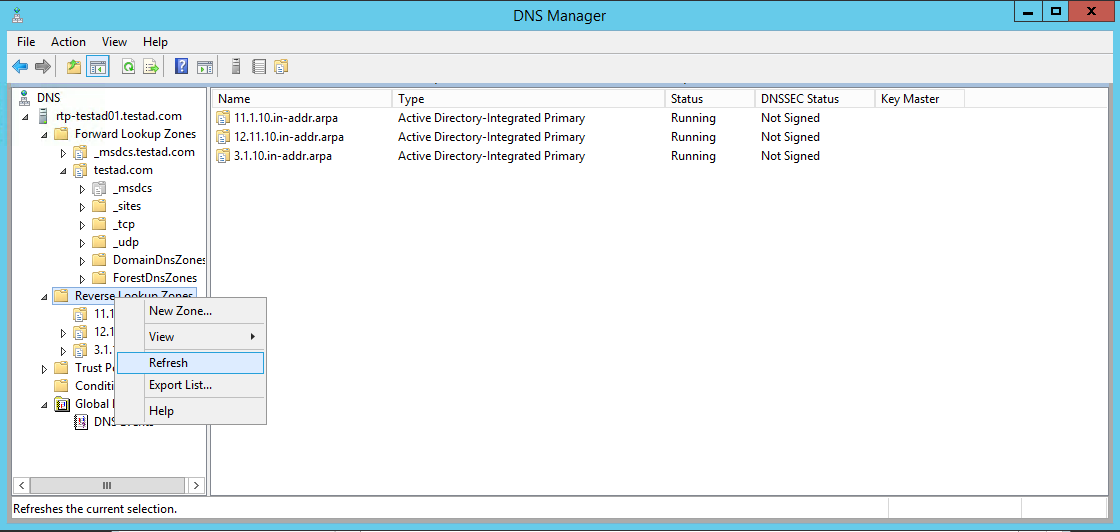
1. Enter the server details as shown below
2. Check the box "Create associated pointer (PTR) record

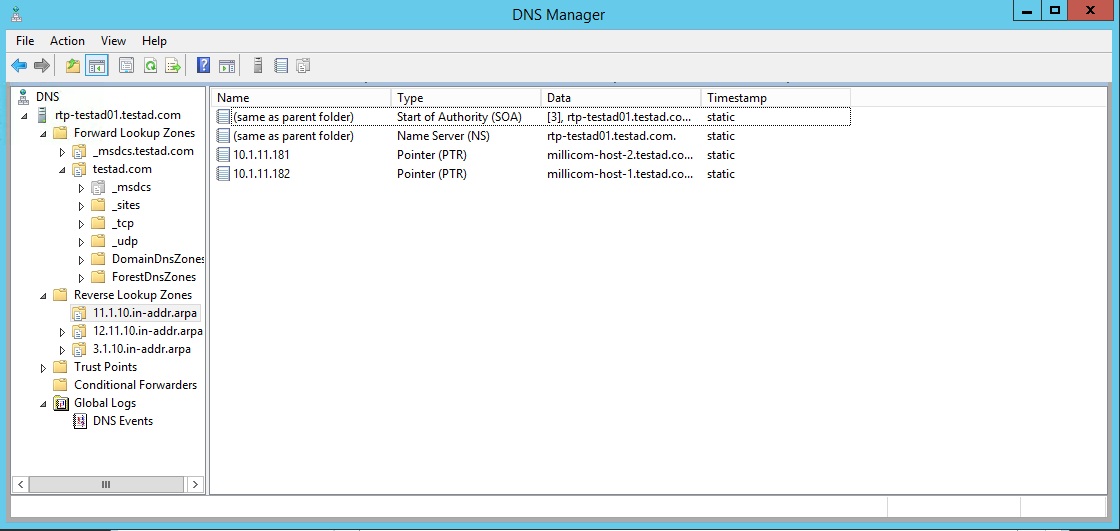


* 1. If the reverse lookup zone is configured properly, you should see the below success message

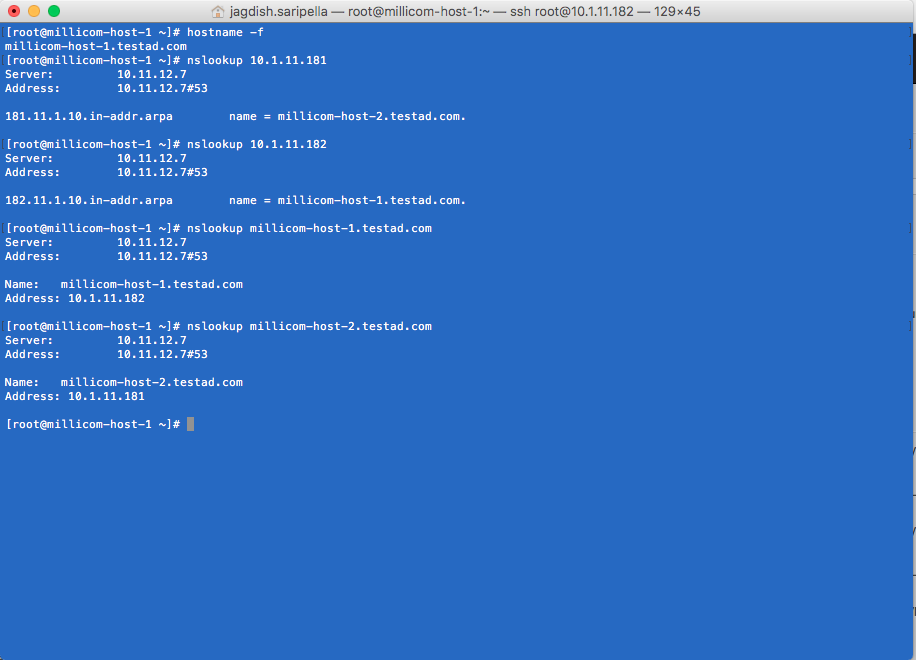


* 1. Refresh the reverse  lookup Zone to see the below server entry





* 1. Once this is done the server you should verify the fqdn of the server
  2. Also you will be able to do a forward and reverse lookup on the servers in your subnet



## Joining Linux servers to AD

Install the adcli package to start joining the linux server to the domain.

1. yum install adcli
2. adcli info <domain>

[root@ad-testmachine1 ~]# adcli info testad.com

[domain]

domain-name = testad.com

domain-short = TESTAD

domain-forest = testad.com

domain-controller = rtp-testad01.testad.com

domain-controller-site = Default-First-Site-Name

domain-controller-flags = pdc gc ldap ds kdc timeserv closest writable full-secret ads-web

domain-controller-usable = yes

domain-controllers = rtp-testad01.testad.com

[computer]

computer-site = Default-First-Site-Name

1. Join the domain using a user who has privilege to join computer on the OU

adcli join -v --domain-controller=<ip> --domain-ou="Millicom-Test-OU,OU=servers" --login-user="<user>"

\* Sending netlogon pings to domain controller: cldap://10.11.12.7

\* Received NetLogon info from: rtp-testad01.testad.com

\* Discovered domain name: testad.com

\* Calculated computer account name from fqdn: AD-TESTMACHINE1

\* Calculated domain realm from name: TESTAD.COM

\* Wrote out krb5.conf snippet to /tmp/adcli-krb5-rH6Uxd/krb5.d/adcli-krb5-conf-aOklHf

Password for millicom-user@TESTAD.COM:

\* Authenticated as user: millicom-user@TESTAD.COM

\* Looked up short domain name: TESTAD

\* Using fully qualified name: ad-testmachine1

\* Using domain name: testad.com

\* Using computer account name: AD-TESTMACHINE1

\* Using domain realm: testad.com

[logging]

\* Calculated computer account name from fqdn: AD-TESTMACHINE1

\* Generated 120 character computer password

\* Using keytab: FILE:/etc/krb5.keytab

\* Computer account for AD-TESTMACHINE1$ does not exist

! Couldn't find a computer container in the ou, creating computer account directly in: OU=Millicom-Test-OU,DC=testad,DC=com

\* Calculated computer account: CN=AD-TESTMACHINE1,OU=Millicom-Test-OU,DC=testad,DC=com

\* Created computer account: CN=AD-TESTMACHINE1,OU=Millicom-Test-OU,DC=testad,DC=com

\* Set computer password

\* Retrieved kvno '2' for computer account in directory: CN=AD-TESTMACHINE1,OU=Millicom-Test-OU,DC=testad,DC=com

\* Modifying computer account: dNSHostName

\* Modifying computer account: userAccountControl

\* Modifying computer account: operatingSystem, operatingSystemVersion, operatingSystemServicePack

\* Modifying computer account: userPrincipalName

! Couldn't set service principals on computer account CN=AD-TESTMACHINE1,OU=Millicom-Test-OU,DC=testad,DC=com: 00002083: AtrErr: DSID-03151785, #1:

0: 00002083: DSID-03151785, problem 1006 (ATT\_OR\_VALUE\_EXISTS), data 0, Att 90303 (servicePrincipalName)

\* Discovered which keytab salt to use

\* Added the entries to the keytab: AD-TESTMACHINE1$@TESTAD.COM: FILE:/etc/krb5.keytab

\* Added the entries to the keytab: host/AD-TESTMACHINE1@TESTAD.COM: FILE:/etc/krb5.keytab

\* Added the entries to the keytab: host/ad-testmachine1@TESTAD.COM: FILE:/etc/krb5.keytab

\* Added the entries to the keytab: RestrictedKrbHost/AD-TESTMACHINE1@TESTAD.COM: FILE:/etc/krb5.keytab

\* Added the entries to the keytab: RestrictedKrbHost/ad-testmachine1@TESTAD.COM: FILE:/etc/krb5.keytab

1. Modify the /etc/krb5.conf to point to the right REALM

[logging]

default = FILE:/var/log/krb5libs.log

kdc = FILE:/var/log/krb5kdc.log

admin\_server = FILE:/var/log/kadmind.log

[libdefaults]

default\_realm = TESTAD.COM

dns\_lookup\_realm = true

dns\_lookup\_kdc = true

ticket\_lifetime = 24h

renew\_lifetime = 7d

forwardable = true

[realms]

TESTAD.COM = {

 kdc = rtp-testad01.testad.com

 admin\_server = rtp-testad01.testad.com

}

[domain\_realm]

.example.com = TESTAD.COM

example.com = TESTAD.COM

1. Setting up SSSD to use keytab to resolve users . Make sure authconfig & sssd are installed

yum install authconfig sssd

1. authconfig --enablesssd --enablesssdauth --enablemkhomedir --update
2. Final step define a single domain in SSSD

[sssd]

services = nss, pam, ssh, autofs

config\_file\_version = 2

domains = TESTAD.COM

[domain/TESTAD.COM]

id\_provider = ad

# Uncomment if service discovery is not working

#ad\_server = rtp-testad01.testad.com

[nss]

       override\_shell = /bin/bash

       override\_homedir = /home/%u

# Active Directory structure for Big data OU

Similar to any technology platform, having an OU under AD helps manage the Hadoop platform better in context of users & groups having necessary access to the platform .

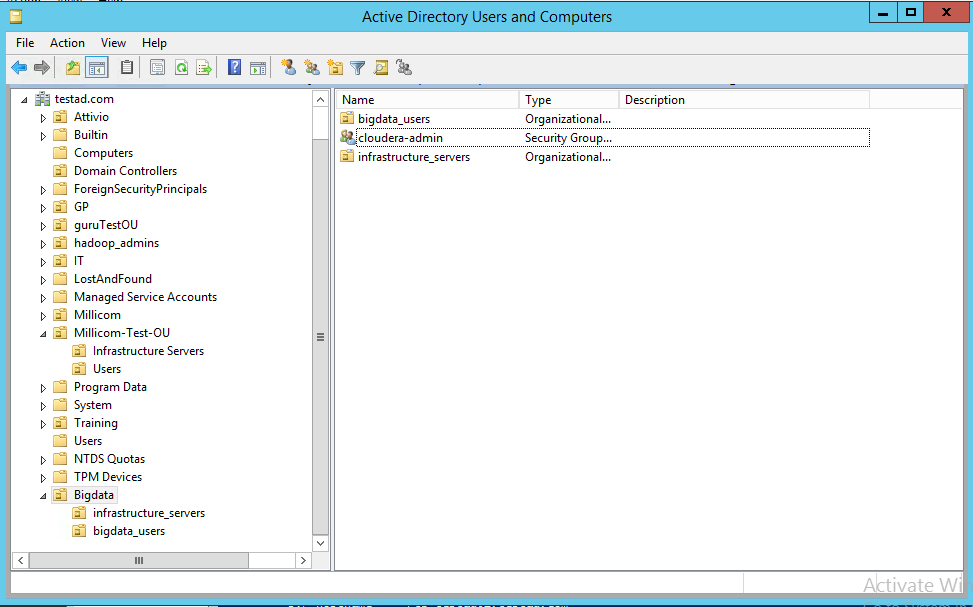
Below is a sample structure that can be followed : Please make sure there are no space between the OU names . (There are some cloudera code issues when Cloudera is AD kerberized and cloudera manager tries to regenerate tickets for service accounts .)

Parent OU=Bigdata : Top level OU that includes all resources related to Big Data Project

Child OU: infrastructure\_servers : All the servers belonging to bigdata initiative can be under this OU .

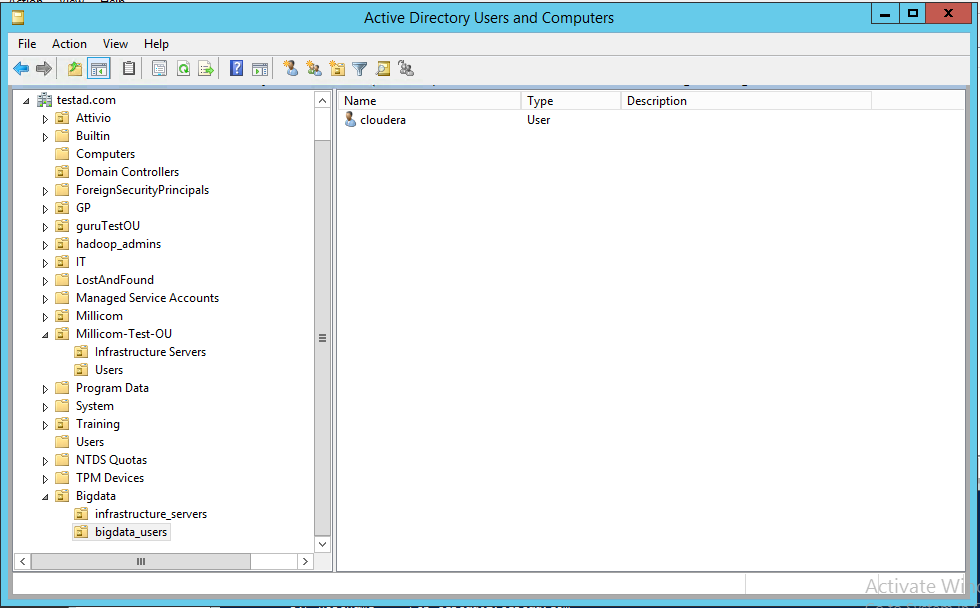
Child OU: bigdata\_users : All the users involved in bigdata related projects can be placed under this OU . This way we can narrow down access to the BigData Linux servers.

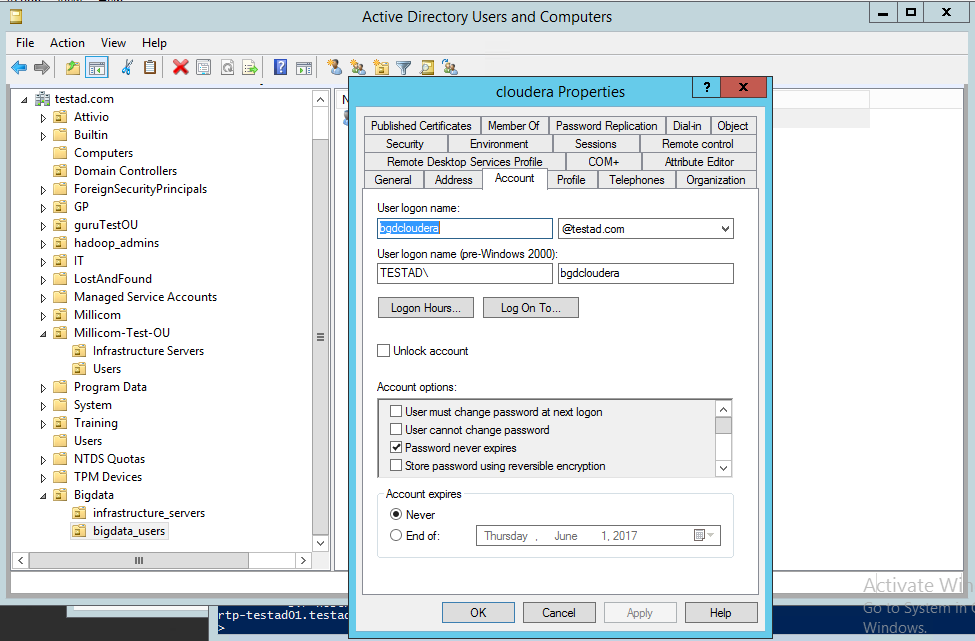
Security Groups can also be defined under the BigData OU .



## Service Account for Kerberos Integration to Active Directory

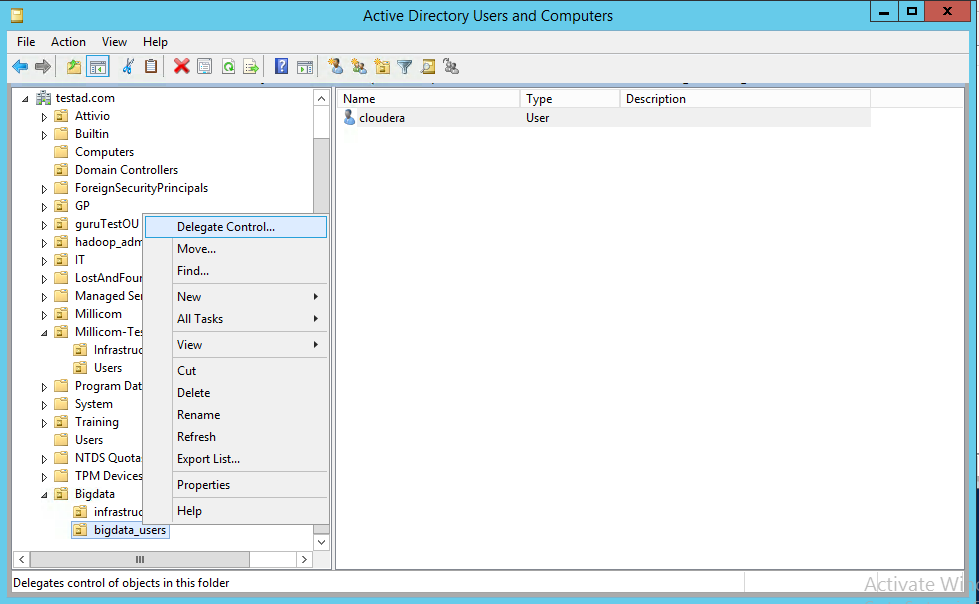
For Kerberos security to integrate with Active Directory, infrastructure team would need to provide a privileged service account with non-expiring password . Below is how a sample account might look like :

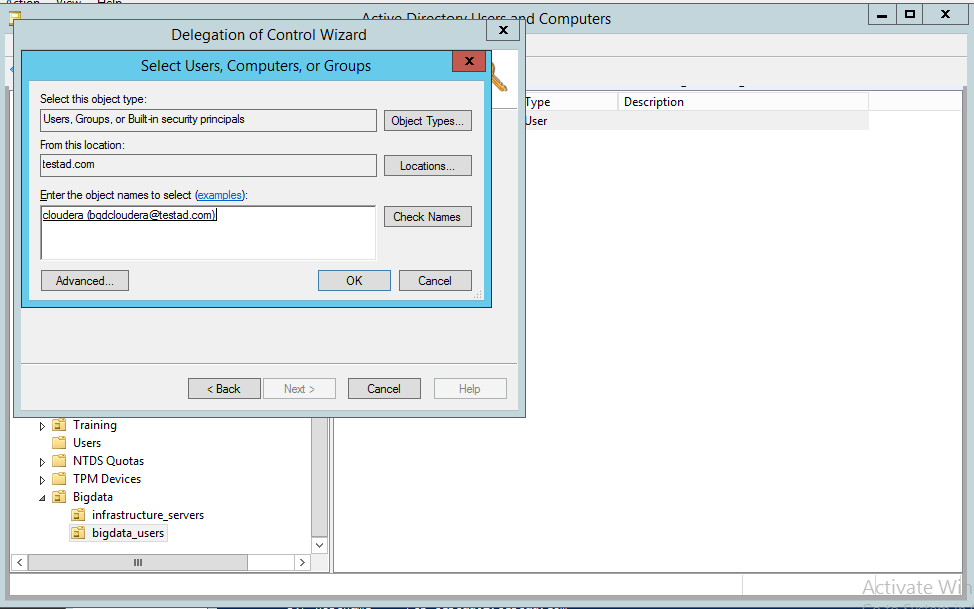


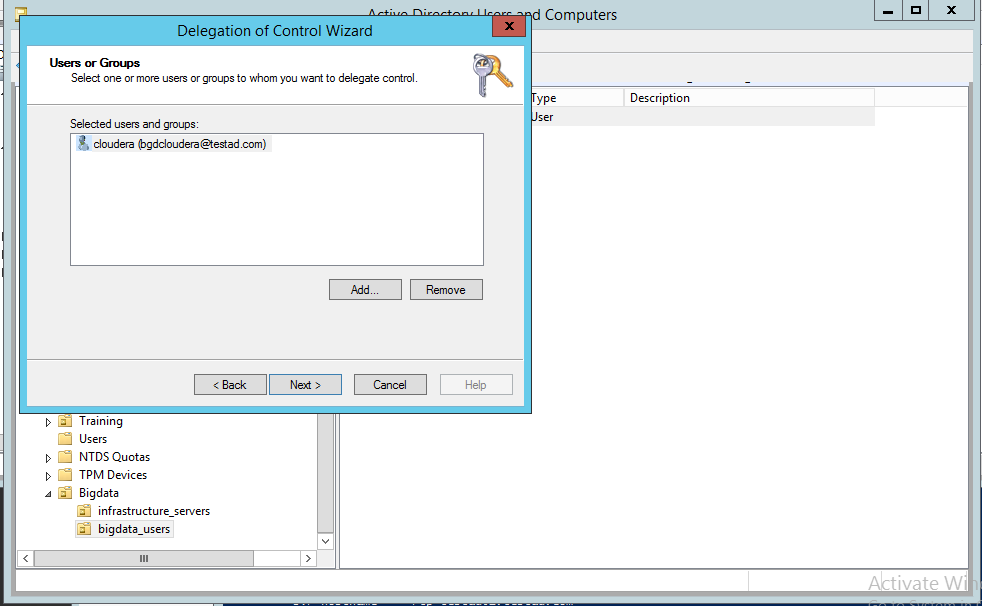


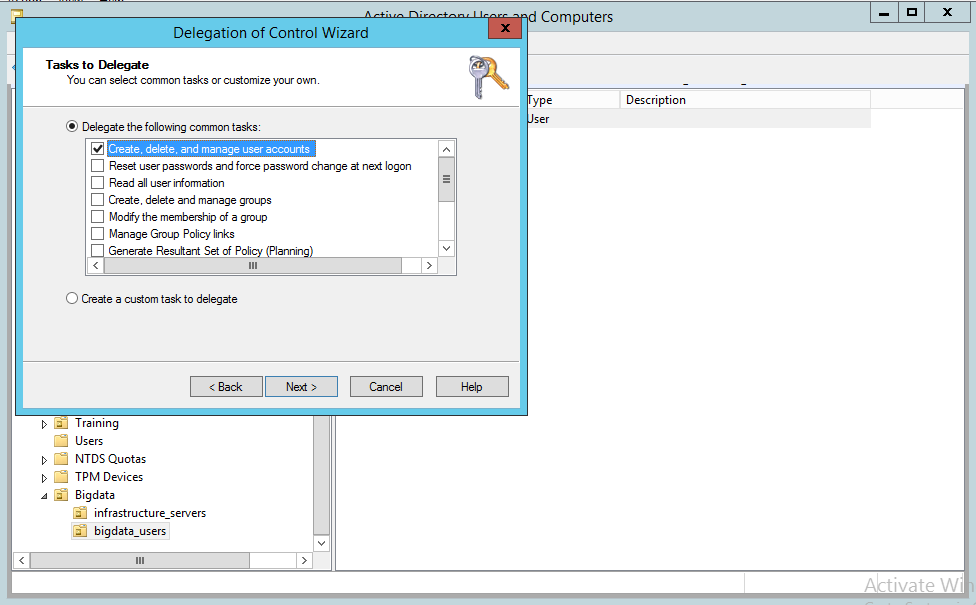
privileges the service account should have : The service account should have “create,delete and manager user accounts” on the OU (in this case bigdata\_users)

Below are the steps to provide those privileges on Active Directory:









# Web Addresess from where Packages Needs to be Downloaded

Below are the web addresses that needs access from the server to download the binaries.

* Cloudera :
  + <https://archive.cloudera.com/cm5/repo-as-tarball/5.8.4/>
  + http://archive.cloudera.com/cdh5/parcels/5.8.4/
* Miscelaneous : Access to CentOS base repo .

# Hosting an Internal Repo using Apache Web Server

Internal Repository needs to be hosted to access the binaries for below two scenarios:

1. Corporate policy does not allow access to external web url’s
2. Faster access to binaries in case of installations/upgrades.

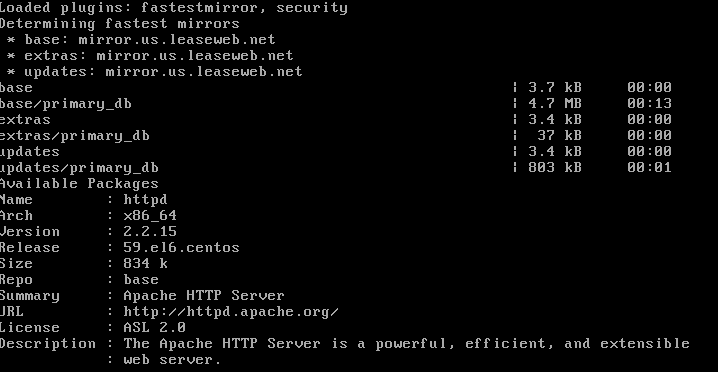
One of the most common webserver that can be used is Apache Http server . Verify if the Linux distribution has “httpd” package already installed .

There are multiple ways to verify, one of the ways is as below :

> yum info httpd

This will show an output similar to below :

\* If httpd is not installed, it will show the package only under “Available packages”

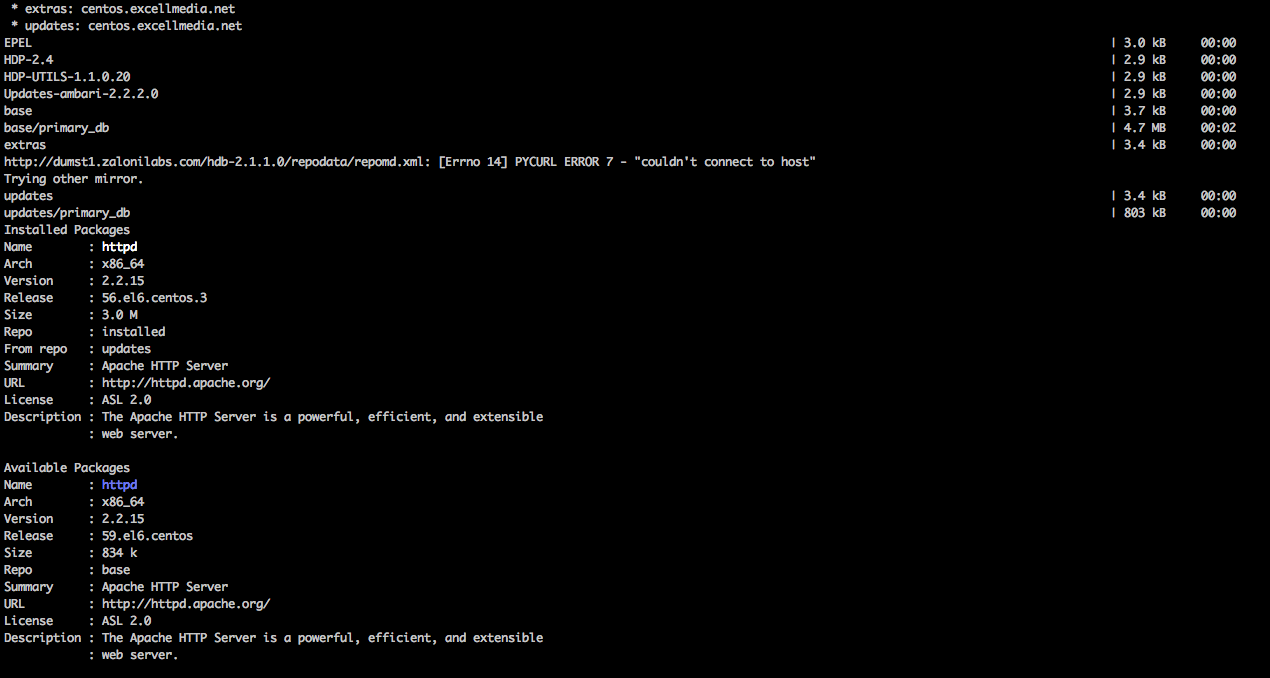


\* If httpd is installed and there are higher versions available , you can wither upgrade or start using the available version.

To Install httpd :

> yum install httpd –y

The above command will fetch the necessary binaries and dependencies from the external url and install the package. Once installation is successful , you will be able to start the web server as described in section [starting apache web-server](#_Starting_a_Apache).



If the webserver is already installed , chose to update using below command:

> yum update httpd

After that you can start the webserver as mentioned in Step A under Starting webserver section.

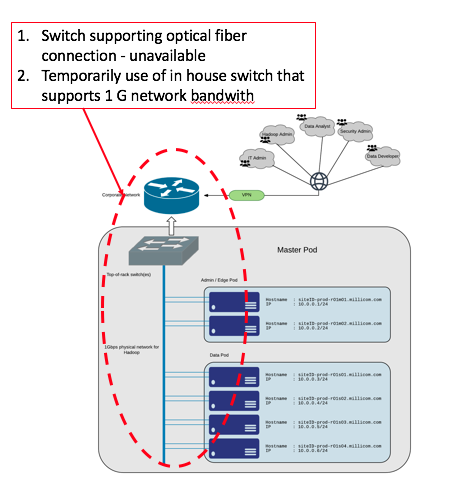
## Starting a Apache http web-server:

1. Log on to server , & start web-server

> service httpd start

# Temporary Network re-architecture around 1G compatible TOR

Since there is a delay in procuring the compatible TOR for 10G network interface . For now hadoop and datalake implementation will proceed. Once the switch is procured , network will be upgraded to fiber . Below noted points need to be taken care .



Implication of the network upgrade :

1. Change from copper to fiber would mean change in use of available NICs, which could imply change in IP address .
2. Hadoop downtime would be required and reconfiguring the OS cluster IPs will be required .

Recommendation to avoid any adverse impact to cluster:

1. Network team to make sure , same IP’s are assigned to the servers .
2. DNS entries should made sure are accurate.