Hadoop with KV-store using Cygwin

Ms. Namrata Rawal[1], Ms. Vatika Sharma[2]

[1]*Research Scholar, Network Security, GTU PG School, Ahmedabad, Gujarat, India*

[2]*Developer, Hadoop Technology, I-verve Infoweb Company Ahmedabad, Gujarat, India*

ABSTRACT

Handling Big Data means to handle huge databases. So to handle bigdata, there is Hadoop technology to maintain and manage bigdata with KVstore that is one type of NoSQL database on the Cygwin which provides Linux based platform on windows.

Keyword: - *KVLite, Hadoop cluster, Cygwin on Windows, Oracle NoSQL*

1. Introduction to KV store

KV store is one of the types of NoSQL datastores. KV store uses the concept of associative array as the basic data model where each key is correspond with only single value in collection. Collection is storage of data, like database. This relation is known as a key value pair. Key will be unique identity in the collection.

It provides advantages like scalability, increase performance, flexibility in data retrieval. As it provides performance of data retrieving, so it stores much data in main memory to avoid expensive I/O operations [3].

2. Introduction to Oracle NoSQL Database

Oracle NoSQL database aim is to provide high reliability, flexibility and availability of data and manages across a configurable set of nodes. It provides large throughput and performance of multi-terabyte distributed key-value pair storage [1]. It also offers the services to store and retrieve data of key value pair form the database. Oracle NoSQL database provides perfect CRUD (Create, Read, Update, Delete) operations with high performance, consistency and durability [1].

Oracle NoSQL database is traditionally three tier based architecture:web server, application server and backend database. To make use of Oracle NoSQL Database, code must be written (using Java or C) that runs on the application server [1].

An application makes use of Oracle NoSQL Database by performing network requests against Oracle NoSQL Database's key-value store, which is referred to as the KVStore. The requests are made using the Oracle NoSQL Database Driver, which is linked into your application as a Java library (.jar file), and then accessed using a series of Java APIs.

3. Introduction to KVLite

KVLite is simple version of the Oracle NoSQL Database. It provides a single-node store that is not replicated. It runs in a single process without requiring any administrative interface. Able to configure, start, and stop KVLite using a command line interface. KVLite is installed when KVstore is installed [4]

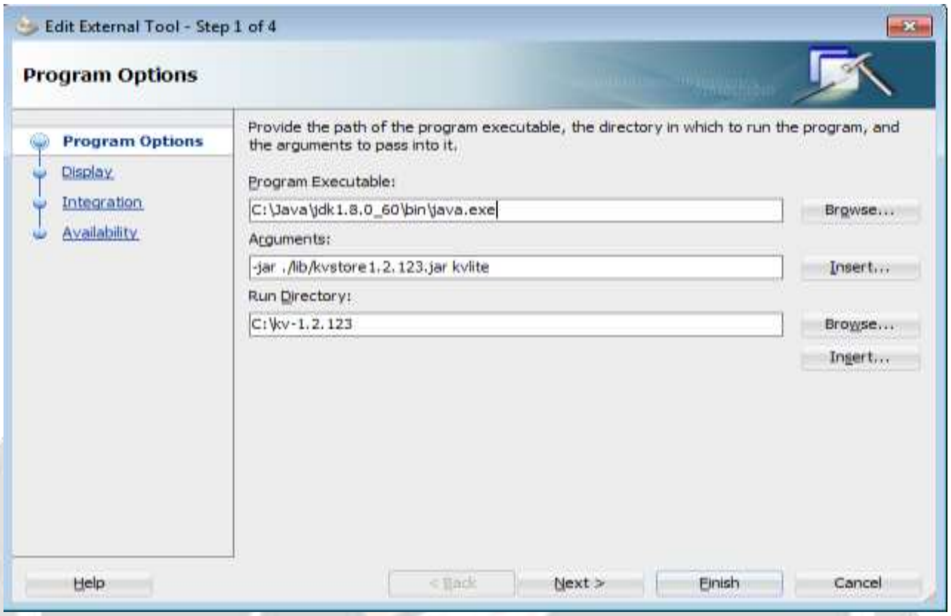
Configuring KVLite in Oracle JDeveloper

KVLite is started by using the kvlite utility which is available in the KVHOME/lib/kvstore.jar. If use this utility without any command line options then KVLite will run with the default values:

- The store name is kvstore.
- The hostname is the local machine.

- The registry port is 5000.
- The directory where Oracle NoSQL Database data is placed (known as KVROOT) is./kvroot.
- The administration process is turned on using port 5001.

For example:
>java –jar ./lib/kvstore.jar kvlite



Edit External Tool - Step 1 of 4

Program Options

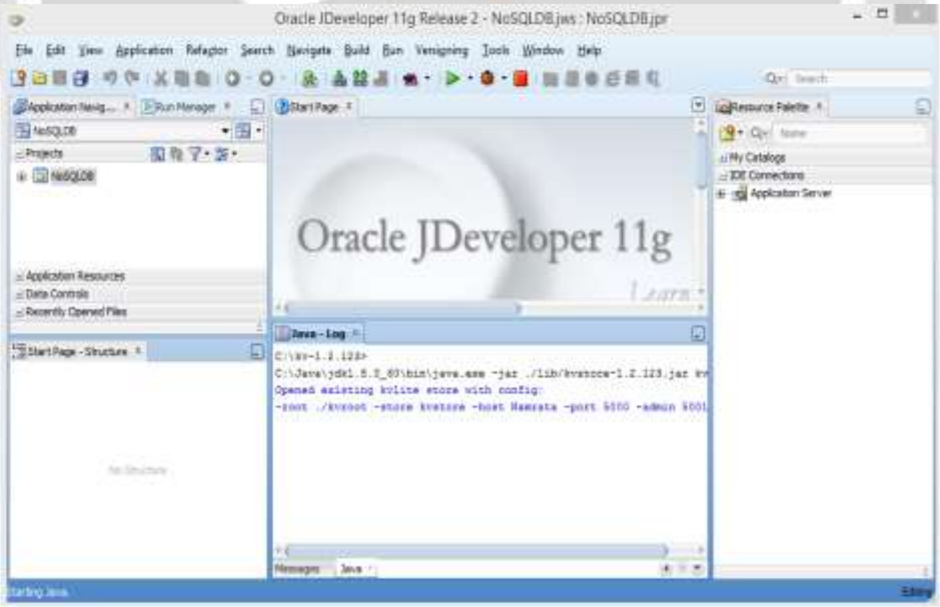
Provide the path of the program executable, the directory in which to run the program, and the arguments to pass into it.

Program Executable: C:\Java\jdk1.8.0_60\bin\java.exe

Arguments: -jar ./lib/kvstore1.2.123.jar kvlite

Run Directory: C:\kv-1.2.123

Fig-1 Setup of oracle NoSQL



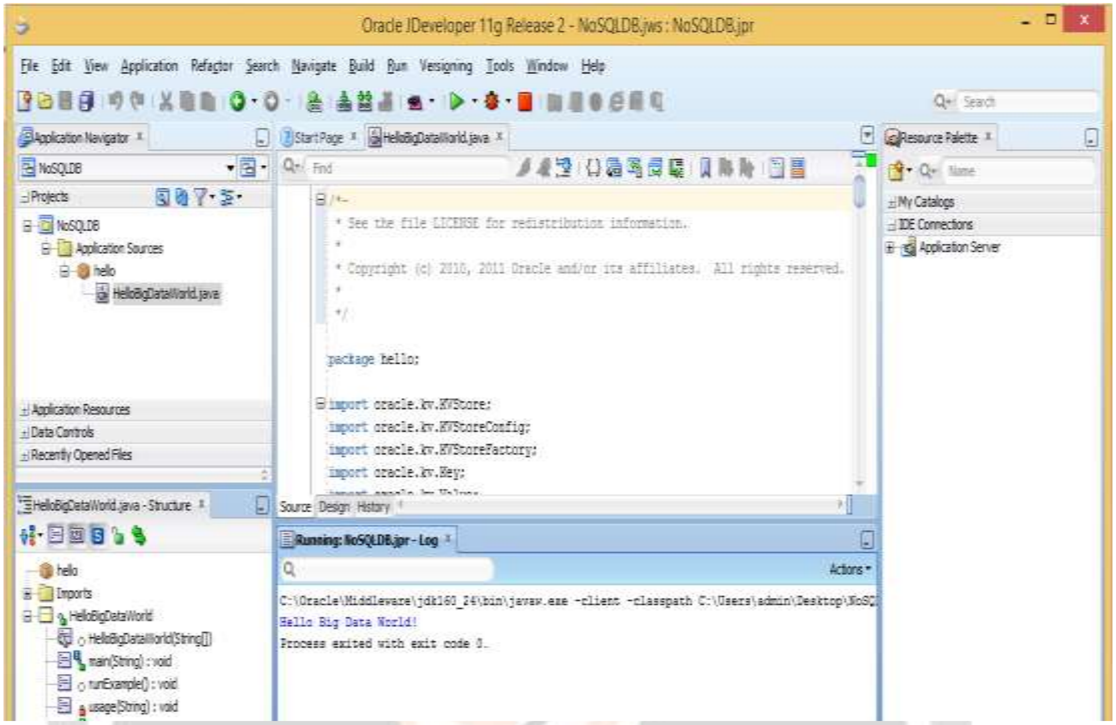
Oracle JDeveloper 11g Release 2 - NoSQLDB.jws : NoSQLDB.jpr

```
C:\kv-1.2.123>
C:\Java\jdk1.8.0_60\bin\java.exe -jar ./lib/kvstore-1.2.123.jar kv
Opened existing kvlite store with config:
-root ./kvroot -store kvstore -host Namrata -port 5000 -admin 5001
```

Fig-2 start up Oracle NoSQL Database server

Opened existing kvlite store with config:
-root ./kvroot –store <kvstore name> -host <localhost> -port 5000

-admin 5001



```
/*-
 * See the file LICENSE for redistribution information.
 *
 * Copyright (c) 2010, 2011 Oracle and/or its affiliates.  All rights reserved.
 *
 */

package hello;

import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;
import oracle.kv.Key;
```

```
C:\Oracle\Middleware\jdk160_24\bin\javaw.exe -client -classpath C:\Users\admin\Desktop\NoSQ
Hello Big Data World!
Process exited with exit code 0.
```

Fig-3 Running NoSQL database as java application

Oracle.kv.KVStore is created using the KVStoreFactory class:

store = KVStoreFactory.getStore(new KVStoreConfig(storeName, hostname + ":" +hostPort));

Key/value pairs are created and stored in the KV store:

final String keyString = "Hello";
final String valueString = "Big Data World!";
store.put(Key.createKey(keyString), Value.createValue(valueString.getBytes()));

The key-value are retrieved from the store and output.Subsequently the KVstore is closed:

final ValueVersion valueVersion = store.get(Key.createKey(keyString));
System.ut.println(keyString + "" + new String(valueVersion.getValue().getValue())+ "\n");
Store.close();

4. Cygwin

Cygwin is the best tool for the users that face the problem in switching operating system (windows to Unix and Unix to windows). Cygwin provides the UNIX environment on the windows operating system. It provides Cygwin command shell prompt in windows system without the need to understand or learn the new commands. Cygwin is a POSIX (Portable Operating System Interface for UNIX) [5]. Cygwin is able to work with any of 32 bit or 64 bit system. Cygwin packages are selected at the installation as per the requirement.

```
admin@Namrata ~
$ |
```

Fig-4 Cygwin Terminal

5. Hadoop

Hadoop is an open source framework it is java based programming which maintain the storage and process of huge amount of data in a distributed computing environment.

6. Hadoop Cluster

Hadoop cluster defined as a type of computational cluster used for storage of huge amount of unstructured data in distributed environment. Hadoop cluster runs on a low cost.

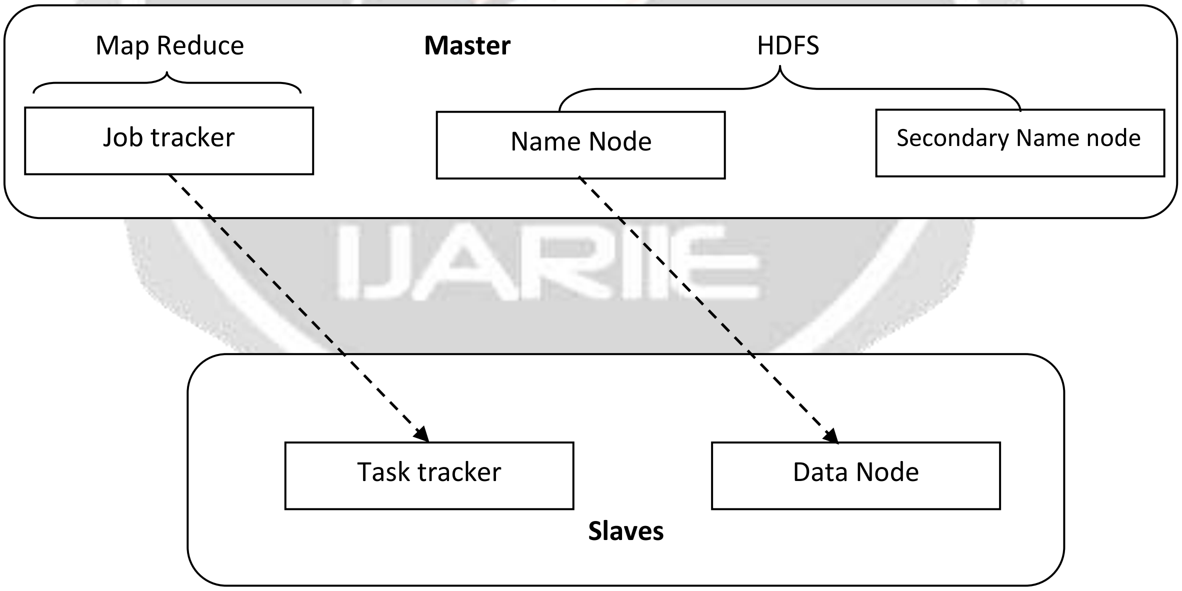


Fig-5 Hadoop components

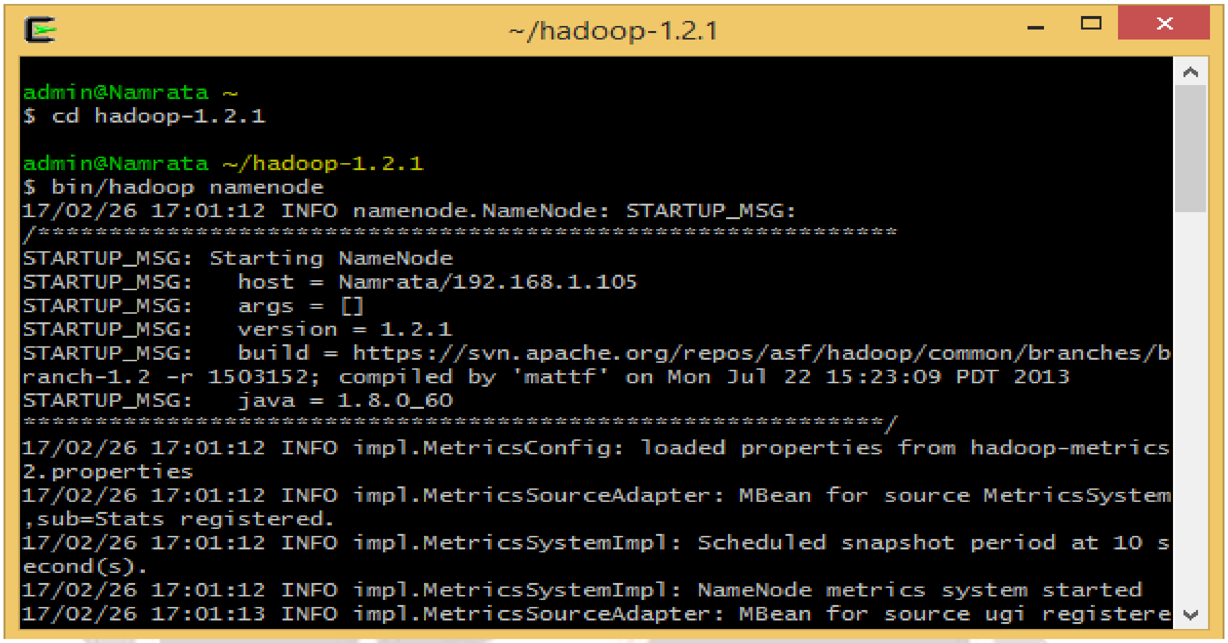
Steps to run Hadoop cluster

Step: 1 Install Hadoop and configure it with Cygwin
Step: 2 creating single node cluster with Namenode, Secondarynamenode, Datanode, Jobtracker and Tasktracker

Below screenshot shows creation of Hadoop cluster:

>$ bin/start-all.sh(to start all nodes on cluster)
Run the following nodes one by one to run the cluster
>$cd hadoop-1.2.1   >$bin/hadoop **namenode**

.

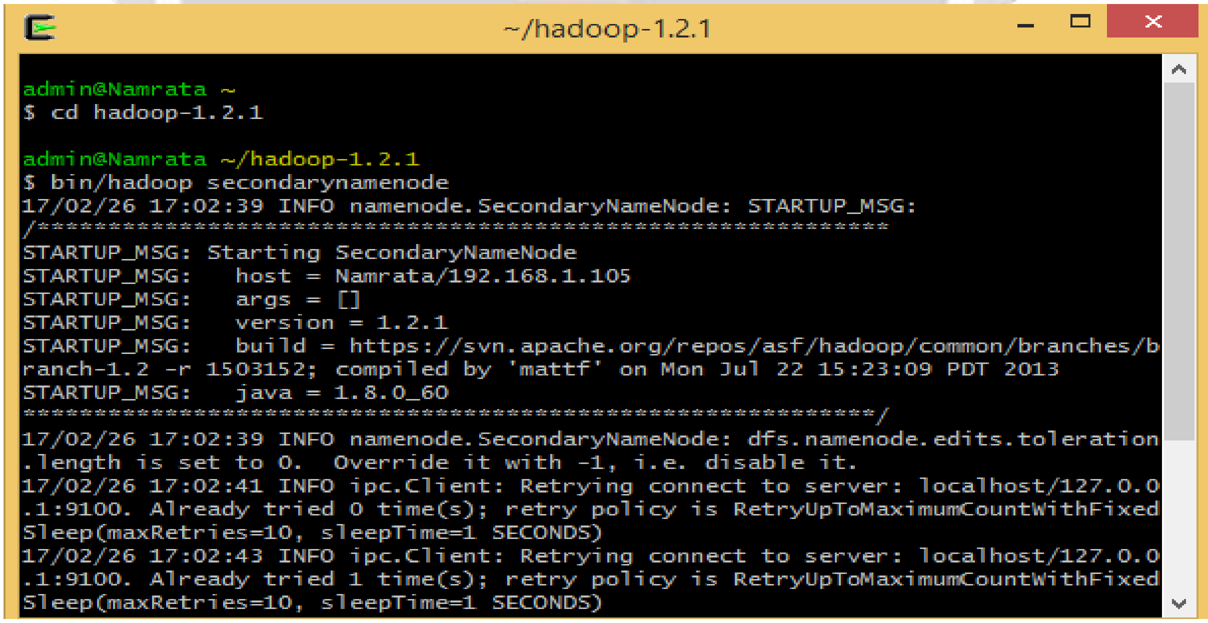


```
admin@Namrata ~
$ cd hadoop-1.2.1

admin@Namrata ~/hadoop-1.2.1
$ bin/hadoop namenode
17/02/26 17:01:12 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = Namrata/192.168.1.105
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-1.2 -r 1503152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.8.0_60
************************************************************/
17/02/26 17:01:12 INFO impl.MetricsConfig: loaded properties from hadoop-metrics
2.properties
17/02/26 17:01:12 INFO impl.MetricsSourceAdapter: MBean for source MetricsSystem
,sub=Stats registered.
17/02/26 17:01:12 INFO impl.MetricsSystemImpl: Scheduled snapshot period at 10 s
econd(s).
17/02/26 17:01:12 INFO impl.MetricsSystemImpl: NameNode metrics system started
17/02/26 17:01:13 INFO impl.MetricsSourceAdapter: MBean for source ugi registere
```

Fig-6 Namenode

>bin/hadoop **secondarynamenode**

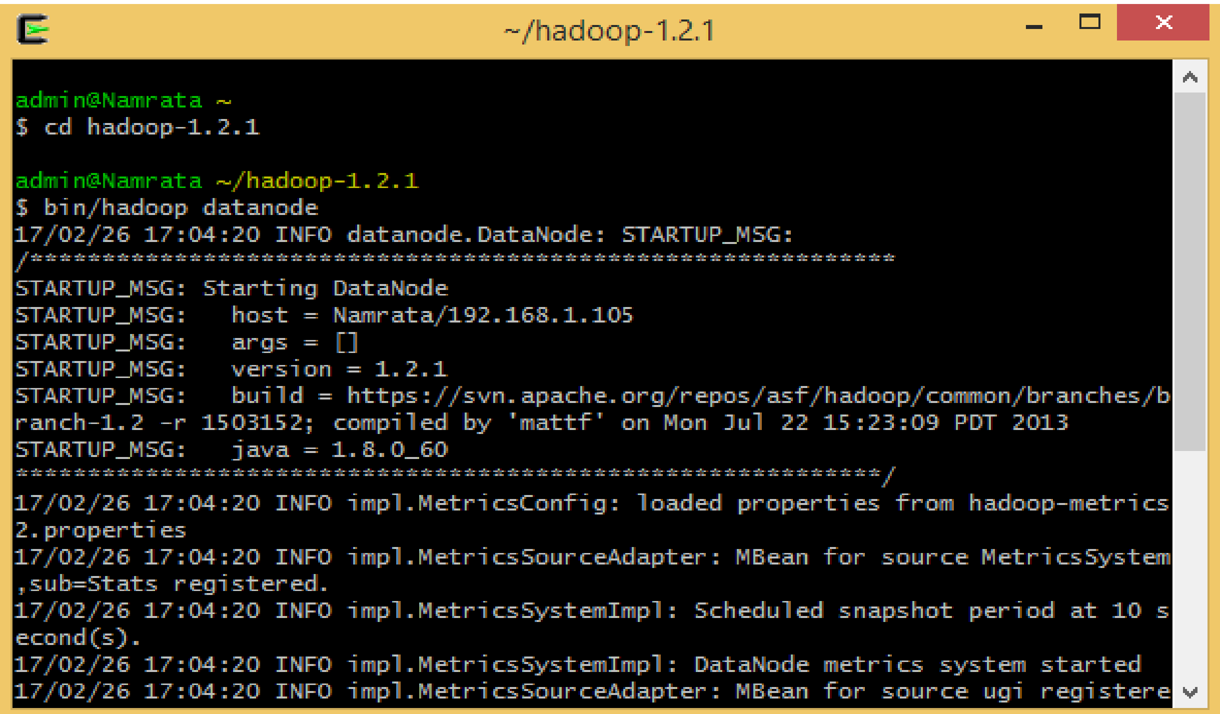


```
admin@Namrata ~
$ cd hadoop-1.2.1

admin@Namrata ~/hadoop-1.2.1
$ bin/hadoop secondarynamenode
17/02/26 17:02:39 INFO namenode.SecondaryNameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting SecondaryNameNode
STARTUP_MSG:   host = Namrata/192.168.1.105
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-1.2 -r 1503152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.8.0_60
************************************************************/
17/02/26 17:02:39 INFO namenode.SecondaryNameNode: dfs.namenode.edits.toleration
.length is set to 0.  Override it with -1, i.e. disable it.
17/02/26 17:02:41 INFO ipc.Client: Retrying connect to server: localhost/127.0.0
.1:9100. Already tried 0 time(s); retry policy is RetryUpToMaximumCountWithFixed
Sleep(maxRetries=10, sleepTime=1 SECONDS)
17/02/26 17:02:43 INFO ipc.Client: Retrying connect to server: localhost/127.0.0
.1:9100. Already tried 1 time(s); retry policy is RetryUpToMaximumCountWithFixed
Sleep(maxRetries=10, sleepTime=1 SECONDS)
```

Fig-7 Secondary Namenode

>bin/hadoop **datanode**

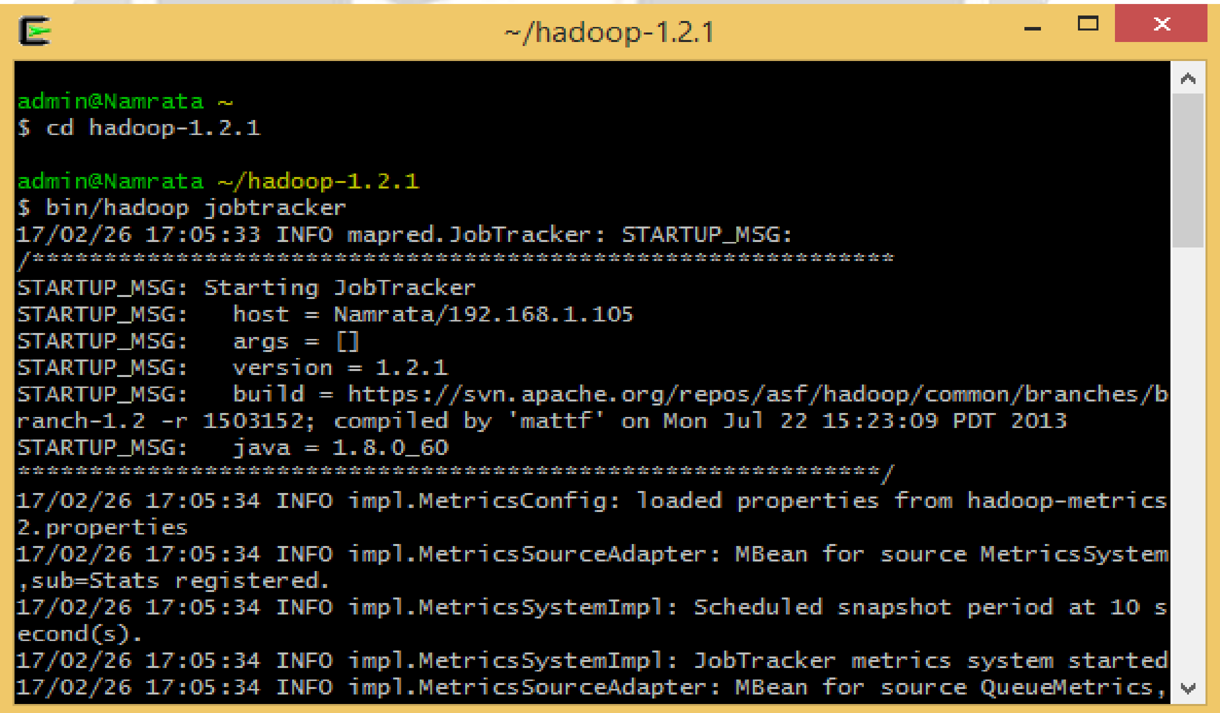


```
admin@Namrata ~
$ cd hadoop-1.2.1

admin@Namrata ~/hadoop-1.2.1
$ bin/hadoop datanode
17/02/26 17:04:20 INFO datanode.DataNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting DataNode
STARTUP_MSG:   host = Namrata/192.168.1.105
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-1.2 -r 1503152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.8.0_60
************************************************************/
17/02/26 17:04:20 INFO impl.MetricsConfig: loaded properties from hadoop-metrics
2.properties
17/02/26 17:04:20 INFO impl.MetricsSourceAdapter: MBean for source MetricsSystem
,sub=Stats registered.
17/02/26 17:04:20 INFO impl.MetricsSystemImpl: Scheduled snapshot period at 10 s
econd(s).
17/02/26 17:04:20 INFO impl.MetricsSystemImpl: DataNode metrics system started
17/02/26 17:04:20 INFO impl.MetricsSourceAdapter: MBean for source ugi registere
```

Fig-8 Datanode

>bin/hadoop **jobtracker**

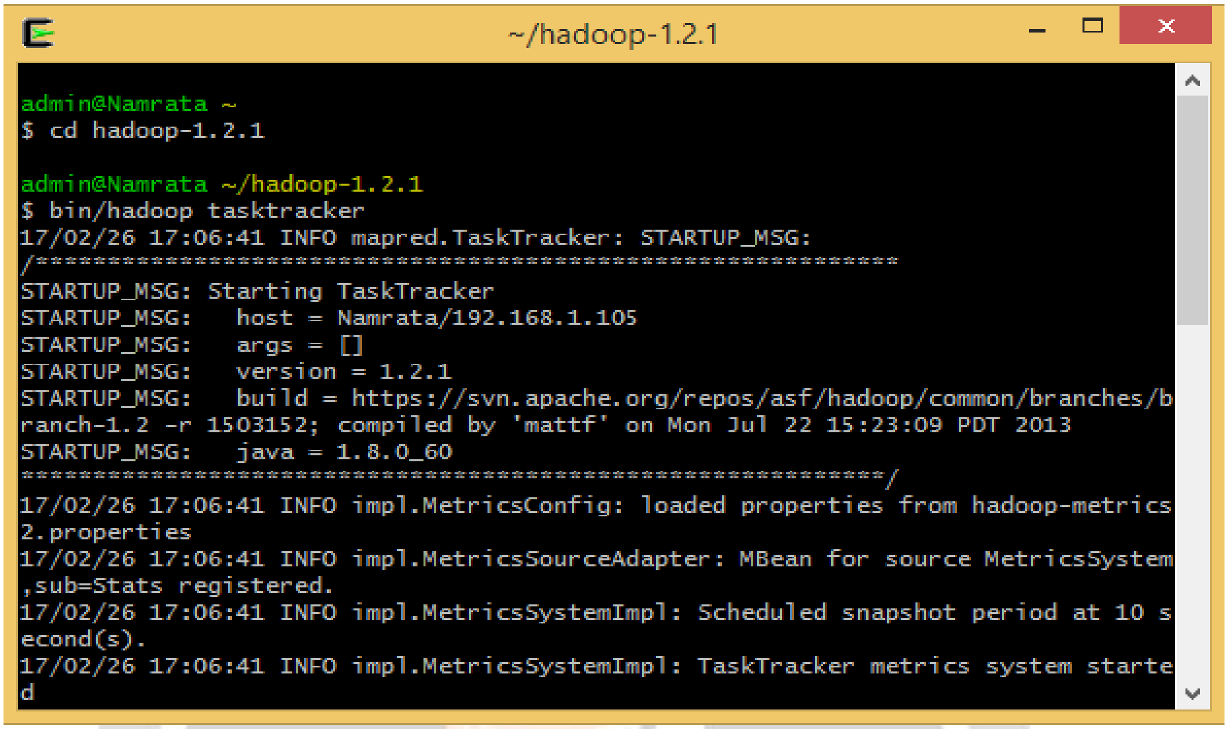


```
admin@Namrata ~
$ cd hadoop-1.2.1

admin@Namrata ~/hadoop-1.2.1
$ bin/hadoop jobtracker
17/02/26 17:05:33 INFO mapred.JobTracker: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting JobTracker
STARTUP_MSG:   host = Namrata/192.168.1.105
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-1.2 -r 1503152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.8.0_60
************************************************************/
17/02/26 17:05:34 INFO impl.MetricsConfig: loaded properties from hadoop-metrics
2.properties
17/02/26 17:05:34 INFO impl.MetricsSourceAdapter: MBean for source MetricsSystem
,sub=Stats registered.
17/02/26 17:05:34 INFO impl.MetricsSystemImpl: Scheduled snapshot period at 10 s
econd(s).
17/02/26 17:05:34 INFO impl.MetricsSystemImpl: JobTracker metrics system started
17/02/26 17:05:34 INFO impl.MetricsSourceAdapter: MBean for source QueueMetrics,
```

Fig-9 Jobtracker

>bin/Hadoop **tasktracker**

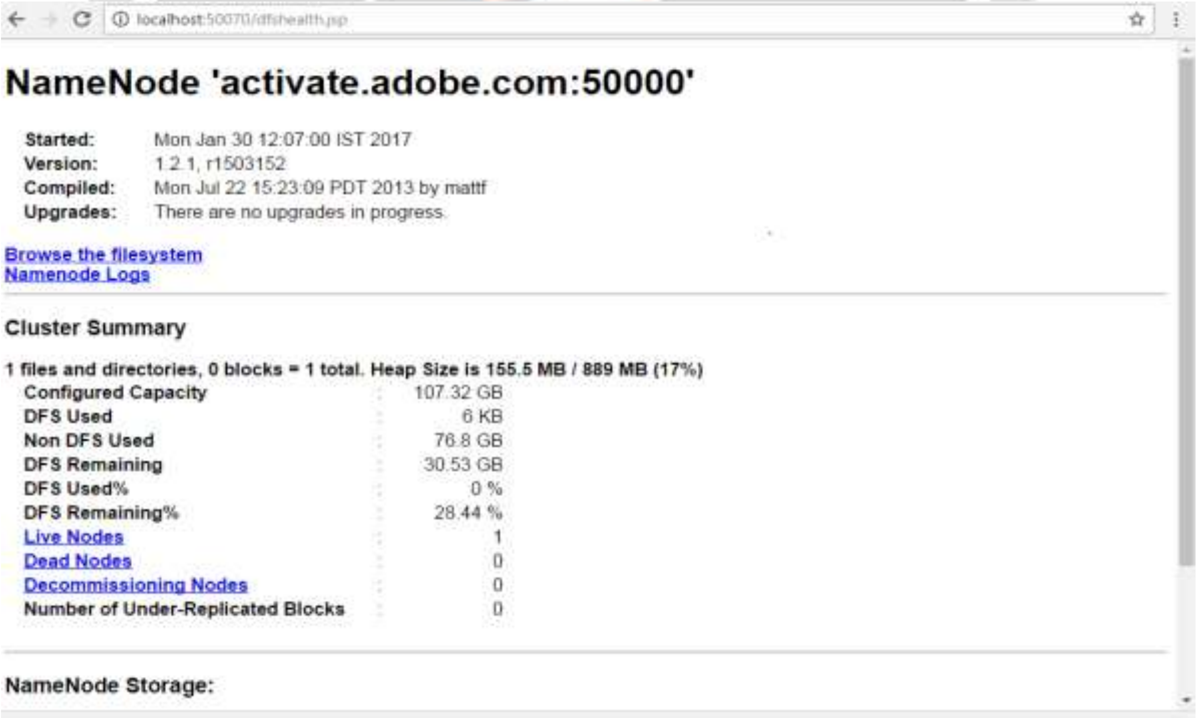


```
admin@Namrata ~
$ cd hadoop-1.2.1

admin@Namrata ~/hadoop-1.2.1
$ bin/hadoop tasktracker
17/02/26 17:06:41 INFO mapred.TaskTracker: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting TaskTracker
STARTUP_MSG:   host = Namrata/192.168.1.105
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/b
ranch-1.2 -r 1503152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.8.0_60
************************************************************/
17/02/26 17:06:41 INFO impl.MetricsConfig: loaded properties from hadoop-metrics
2.properties
17/02/26 17:06:41 INFO impl.MetricsSourceAdapter: MBean for source MetricsSystem
,sub=Stats registered.
17/02/26 17:06:41 INFO impl.MetricsSystemImpl: Scheduled snapshot period at 10 s
econd(s).
17/02/26 17:06:41 INFO impl.MetricsSystemImpl: TaskTracker metrics system starte
d
```

Fig-10 Task Tracker

Hadoop cluster summary



NameNode 'activate.adobe.com:50000'

Started:	Mon Jan 30 12:07:00 IST 2017
Version:	1.2.1, r1503152
Compiled:	Mon Jul 22 15:23:09 PDT 2013 by mattf
Upgrades:	There are no upgrades in progress.

Browse the filesystem
Namenode Logs

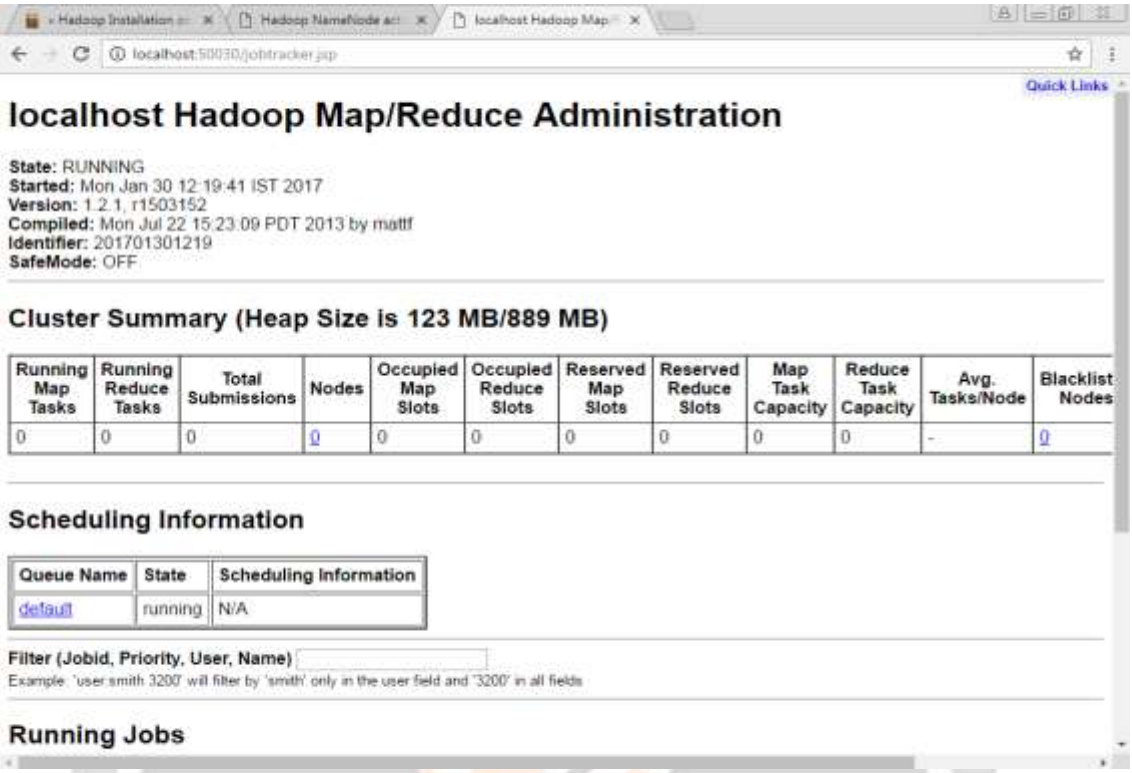
Cluster Summary

1 files and directories, 0 blocks = 1 total. Heap Size is 155.5 MB / 889 MB (17%)

Configured Capacity	:	107.32 GB
DFS Used	:	6 KB
Non DFS Used	:	76.8 GB
DFS Remaining	:	30.53 GB
DFS Used%	:	0 %
DFS Remaining%	:	28.44 %
Live Nodes	:	1
Dead Nodes	:	0
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	0

NameNode Storage:

Fig-11 Cluster summary



localhost Hadoop Map/Reduce Administration

State: RUNNING
Started: Mon Jan 30 12:19:41 IST 2017
Version: 1.2.1, r1503152
Compiled: Mon Jul 22 15:23:09 PDT 2013 by mattf
Identifier: 201701301219
SafeMode: OFF

Cluster Summary (Heap Size is 123 MB/889 MB)

Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklist Nodes
0	0	0	0	0	0	0	0	0	0	-	0

Scheduling Information

Queue Name	State	Scheduling Information
default	running	N/A

Filter (Jobid, Priority, User, Name)
Example: 'user:smith 3200' will filter by 'smith' only in the user field and '3200' in all fields

Running Jobs

Fig-12 Hadoop cluster summary

7. Conclusion and future work

This paper concludes with the basics of KVStore, NoSQL databases, accessing KVLite with the Oracle Jdeveloper for key-value data storage then installation process of Cygwin and creating Hadoop cluster of Hadoop components/nodes with Cygwin. Future work states to implement the performance parameters that achieve polyglot persistence

Referrences

[1] Oracle, "NoSQL Database Administrator's Guide" published in 4/23/2013
[2] Doug Merritt, Snehal Antani, and Richard Campione, web frameworks, 2005. http://dev.splunk.com/view/webframework-developapps/SP-CAAAEZK
[3] Mateusz Berezecki, Eitan Frachtenberg, Mike Paleczny and Kenneth Steele, "Many-Core Key-Value Store"
[4] , in 2011 IEEE
[5] Oracle, "http://docs.oracle.com/cd/NOSQL/html/GettingStartedGuide/index.html" in 2011
[6] Harold L Hunt, II Jon Turney, "Cygwin/X User's Guide", http://x.cygwin.com/docs/ug/cygwin-x-ug.pdf