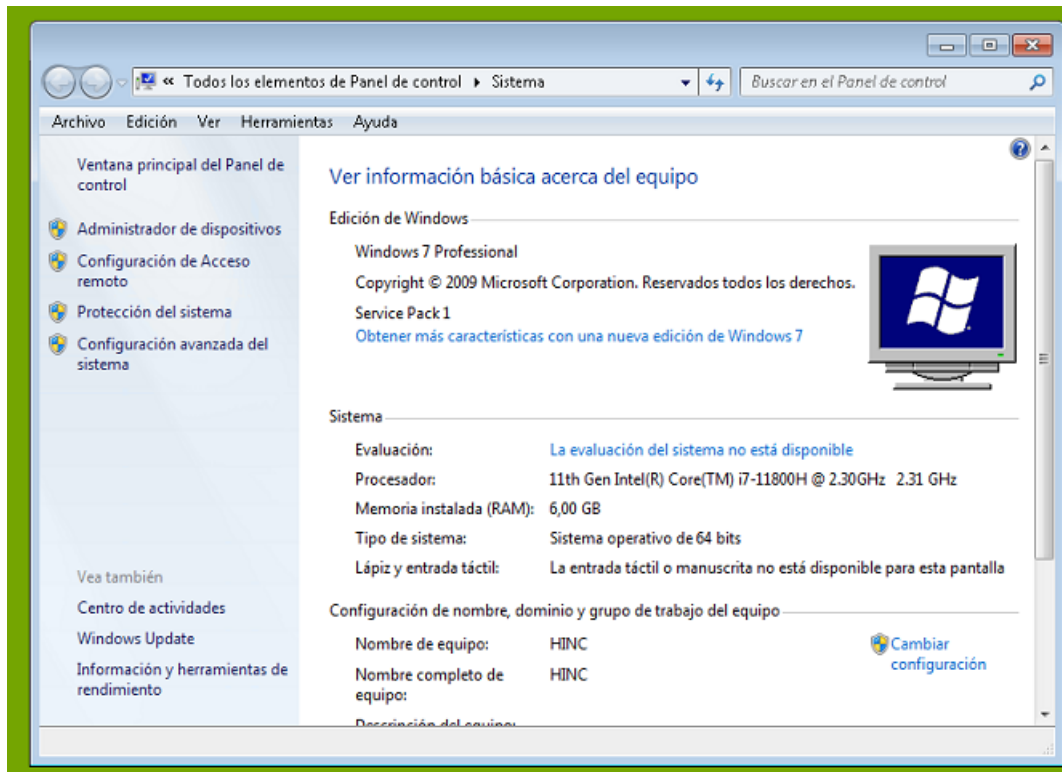




Sistema base: Entorno Virtualizado

Repositorio: <https://github.com/cpalacios01/cluster-hadoop>

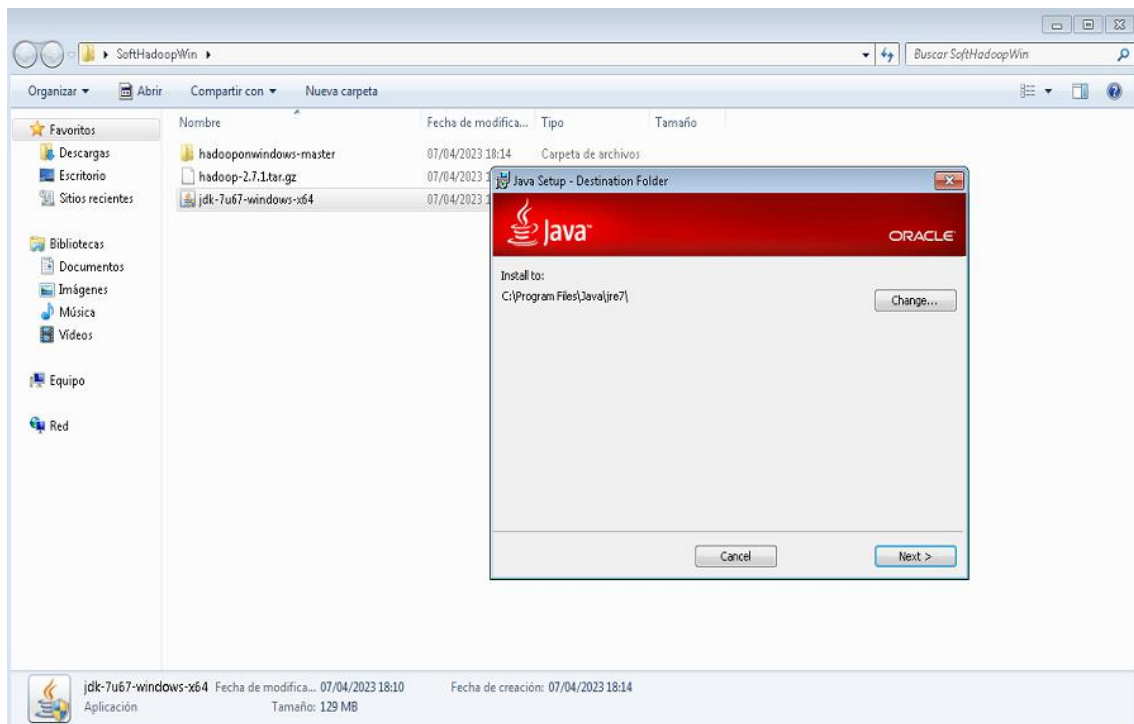
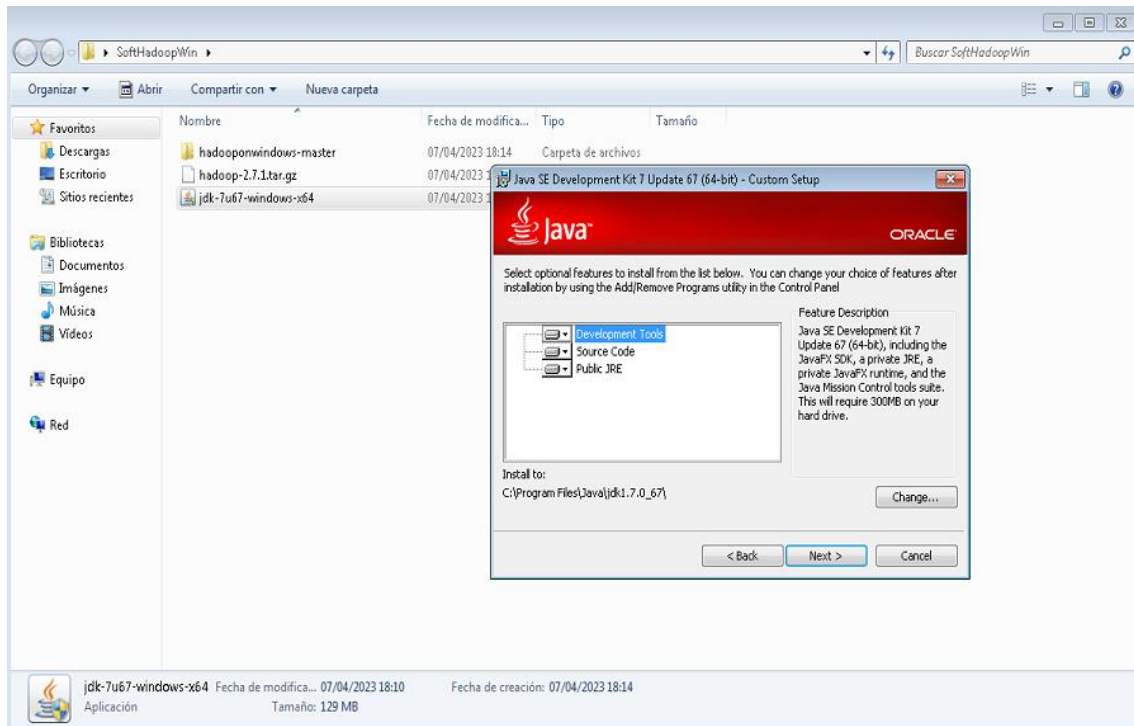


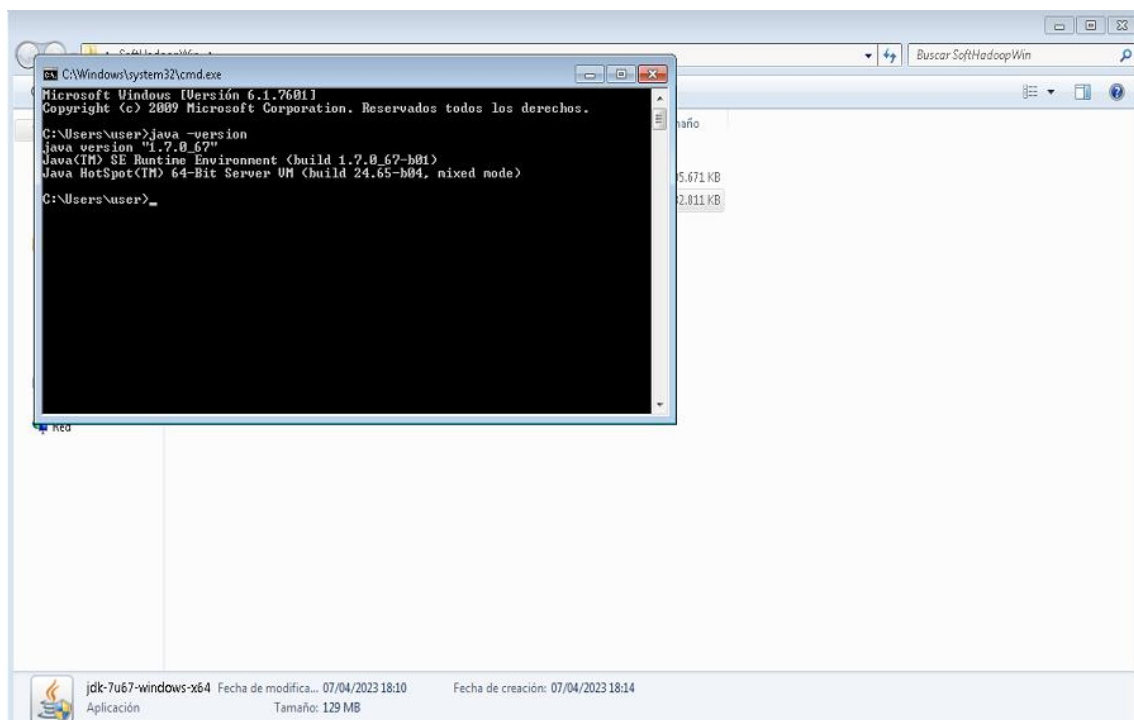
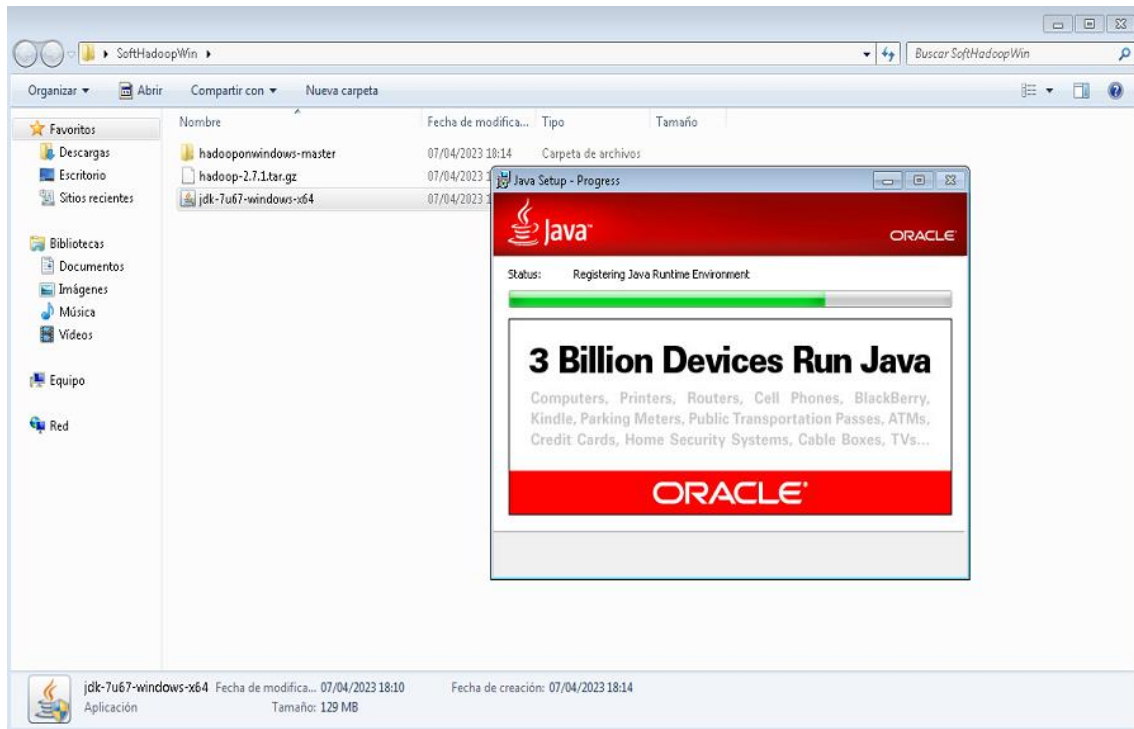
Nota: La versión 2.7 y posterior de Apache Hadoop requiere Java 7.

Descarga:

<https://www.oracle.com/java/technologies/javase/javase7-archive-downloads.html>

Versión: jdk-7u67-windows-x64.exe





C:\Program Files\Java
C:\Program Files\Java\jdk1.7.0_67\bin

0- Descarga =====

<https://archive.apache.org/dist/hadoop/core/hadoop-2.7.1/>

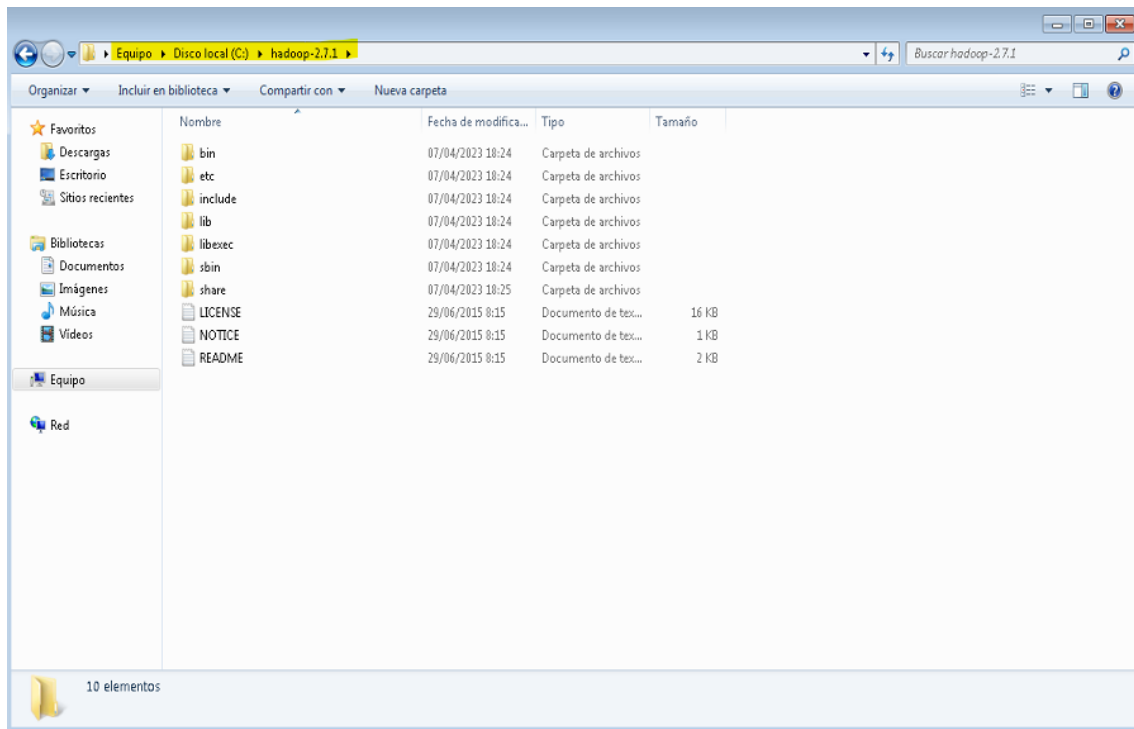
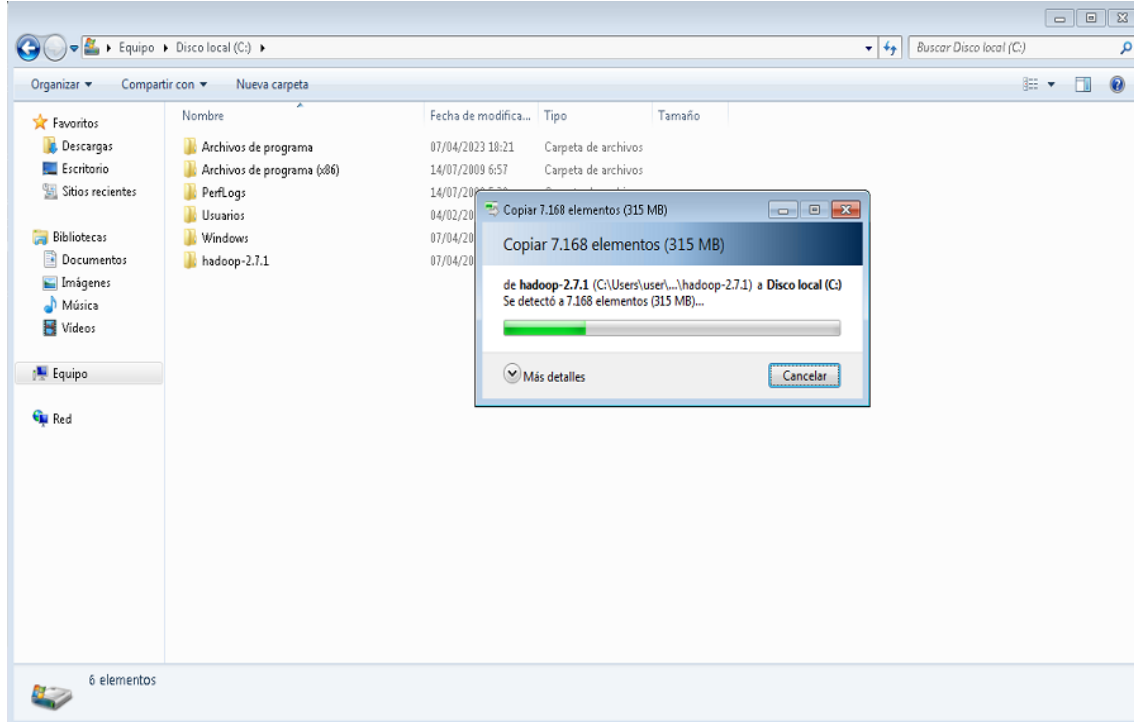


Core: <https://archive.apache.org/dist/hadoop/core/>

1- Directorio de trabajo =====

-Extraer hadoop-2.7.1.tar.gz

-Copiar en C: (en mi caso, puede ser el lugar que consideren)

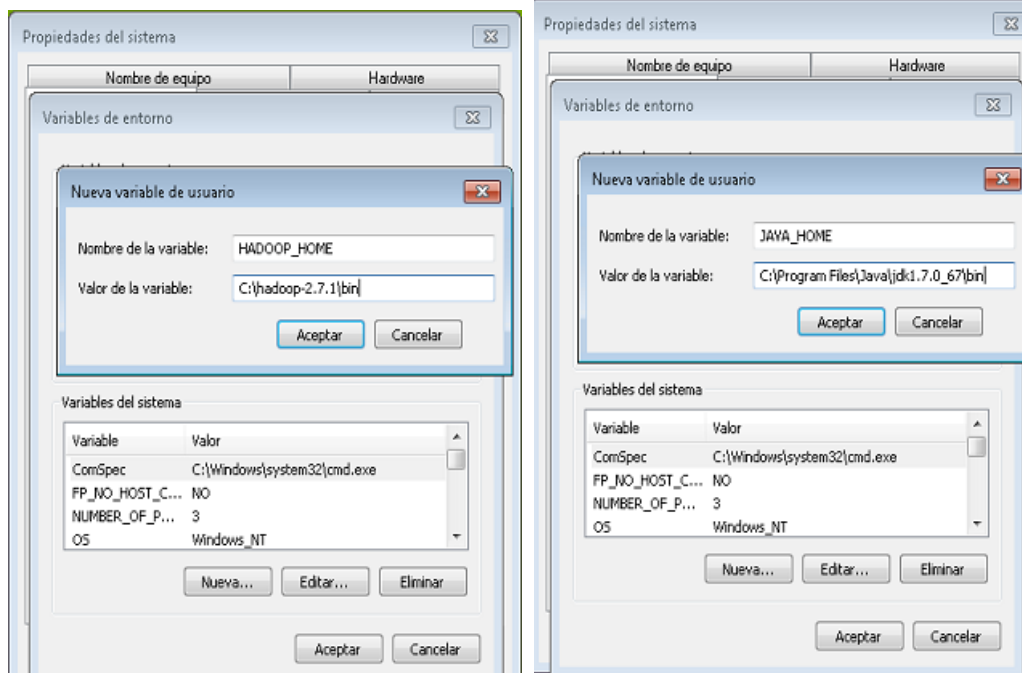


2- VARIABLES DE ENTORNO =====

-Crear Variables de entorno de Usuario

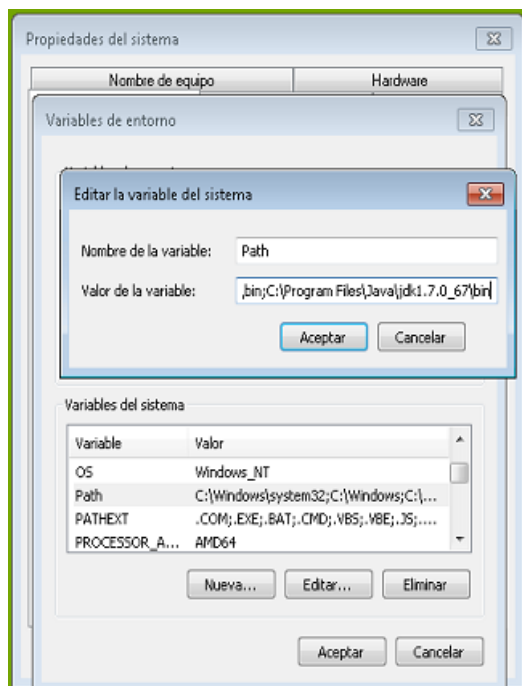
HADOOP_HOME = C:\hadoop-2.7.1\bin

JAVA_HOME = C:\Program Files\Java\jdk1.7.0_67\bin



Agregar al Path:

; C:\hadoop-2.7.1\bin;C:\Program Files\Java\jdk1.7.0_67\bin





Algunas verificaciones:

```
C:\Windows\system32\cmd.exe

C:\Users\user>java -version
java version "1.7.0_67"
Java(TM) SE Runtime Environment (build 1.7.0_67-b01)
Java HotSpot(TM) 64-Bit Server VM (build 24.65-b04, mixed mode)

C:\Users\user>set JAVA_HOME
JAVA_HOME=C:\Program Files\Java\jdk1.7.0_67\bin

C:\Users\user>set HADOOP_HOME
HADOOP_HOME=C:\hadoop-2.7.1\bin

C:\Users\user>_
```

3- CONFIGURACIÓN HADOOP =====

[3.1] C:\hadoop-2.7.1\etc\hadoop\core-site.xml

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

```
C:\hadoop-2.7.1\etc\hadoop\core-site.xml - Notepad++

Archivo  Editar  Buscar  Vista  Codificación  Lenguaje  Configuración  Herramientas  Macro  Ejecutar  Complementos  Pestañas  ?

core-site.xml x
1  <?xml version="1.0" encoding="UTF-8"?>
2  <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3  <!--
4  Licensed under the Apache License, Version 2.0 (the "License");
5  you may not use this file except in compliance with the License.
6  You may obtain a copy of the License at
7
8  http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20   <property>
21     <name>fs.defaultFS</name>
22     <value>hdfs://localhost:9000</value>
23   </property>
24 </configuration>
```



[3.2] C:\hadoop-2.7.1\etc\hadoop\mapred-site.xml

<configuration>

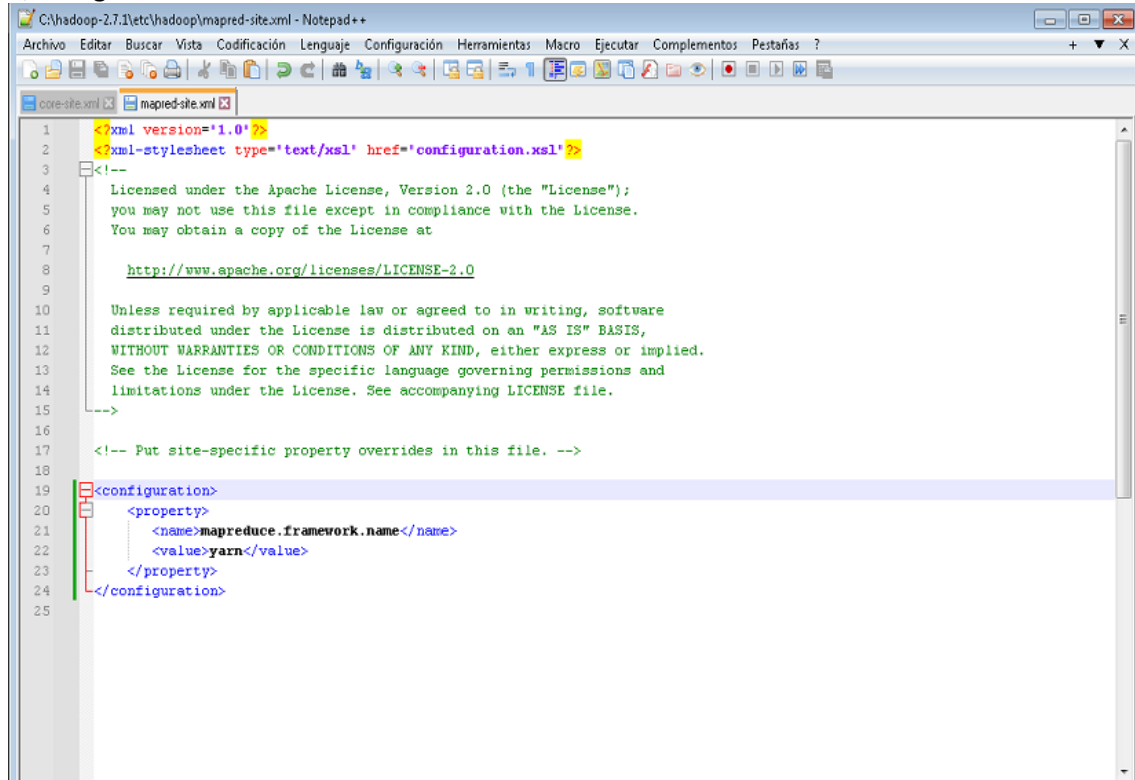
<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

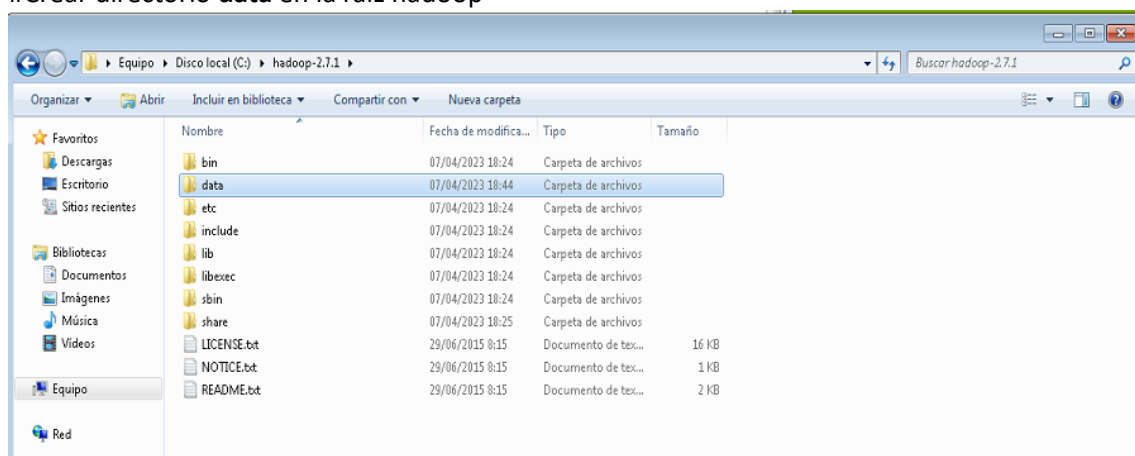
</property>

</configuration>



[3.3]

#Crear directorio **data** en la raíz hadoop



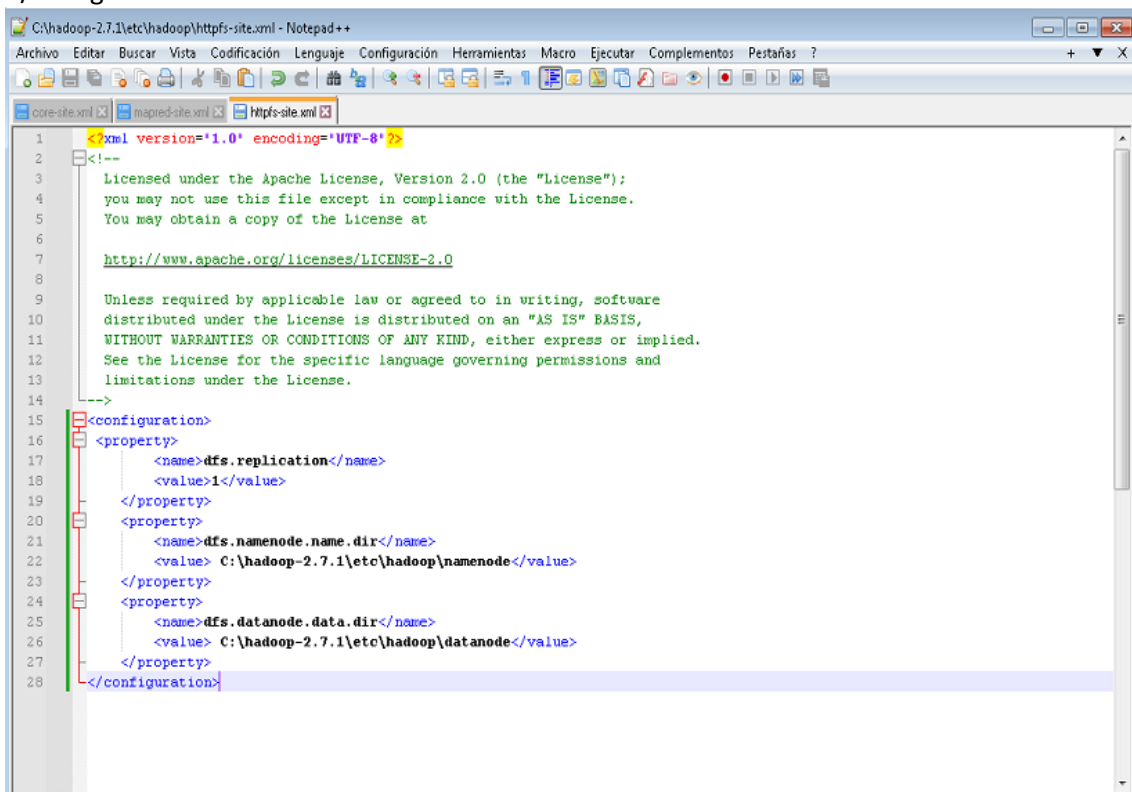
En mi caso, ruta: C:\hadoop-2.7.1\data



[3.4]

- C:\hadoop-2.7.1\etc\hadoop\httpfs-site.xml

```
<configuration>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>C:\hadoop-2.7.1\etc\hadoop\namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>C:\hadoop-2.7.1\etc\hadoop\datanode</value>
</property>
</configuration>
```



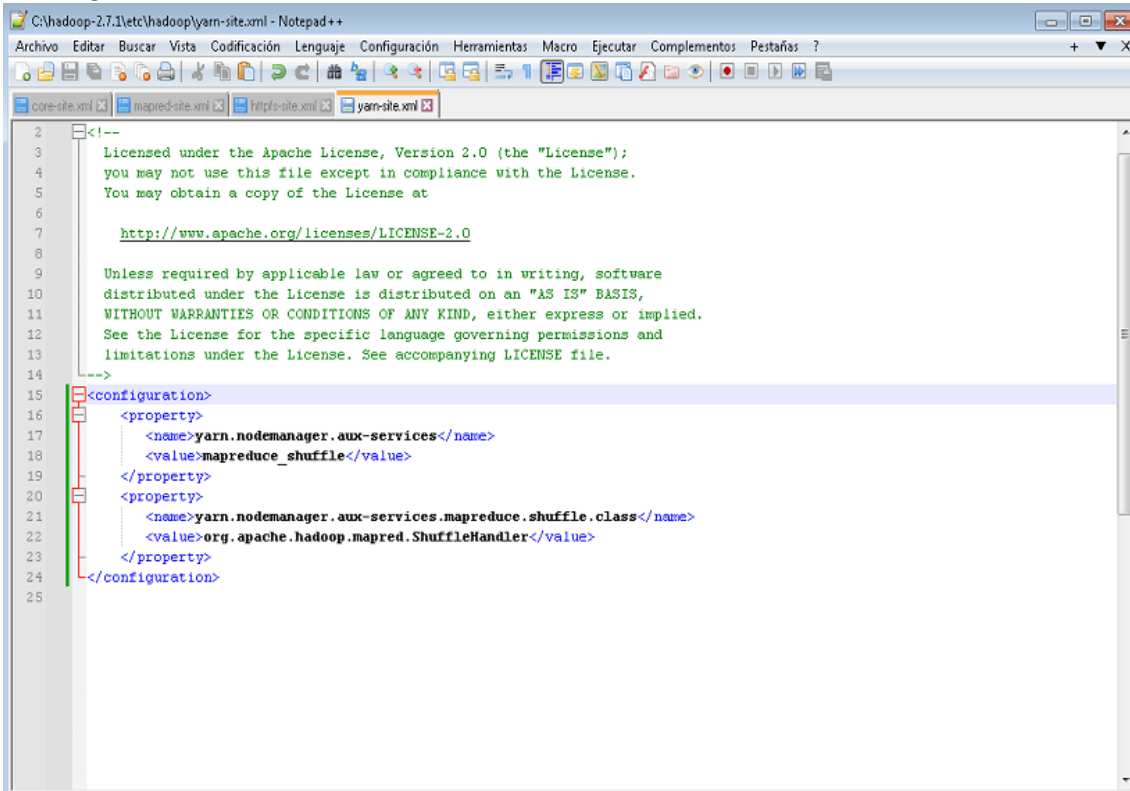
[3.5]

- C:\hadoop-2.7.1\etc\hadoop\yarn-site.xml

```
<configuration>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
```




```
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>
```



4- CONFIGURACIÓN ENTORNO =====

- C:\hadoop-2.7.1\etc\hadoop\hadoop-env.cmd

Línea 25 (mi caso)

set JAVA_HOME=JAVA_HOME=C:\PROGRA~1\Java\jdk1.7.0_67



```
C:\hadoop-2.7.1\etc\hadoop\hadoop-env.cmd - Notepad++
Archivo  Editar  Buscar  Vista  Codificación  Lenguaje  Configuración  Herramientas  Macro  Ejecutar  Complementos  Pestañas  ?

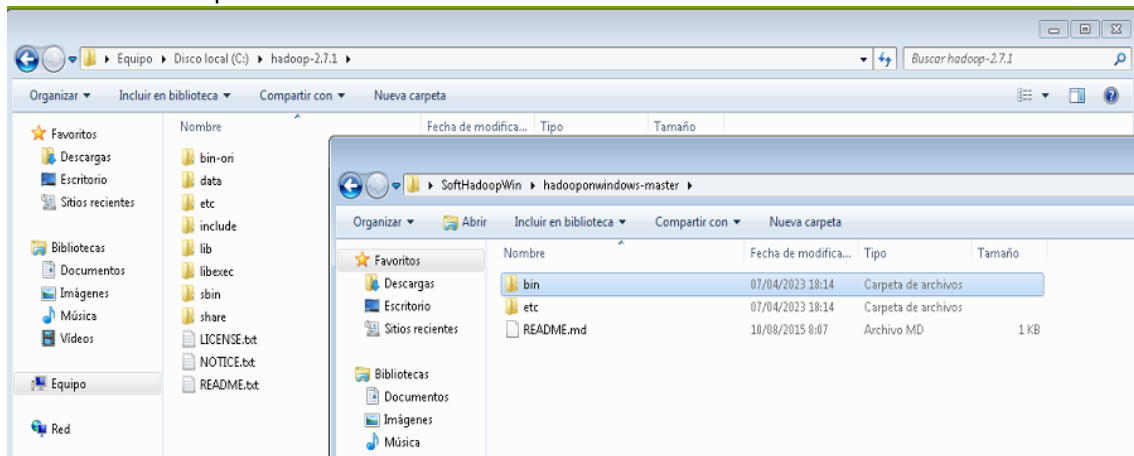
21  rem set JAVA_HOME in this file, so that it is correctly defined on
22  rem remote nodes.
23
24  rem The java implementation to use. Required.
25  set JAVA_HOME=JAVA_HOME=C:\PROGRA~1\Java\jdk1.7.0_67
26
27  rem The jsvc implementation to use. Jsvc is required to run secure datanodes.
28  rem set JSVC_HOME=%JSVC_HOME%
29
30  rem set HADOOP_CONF_DIR=
31
32  rem Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.
33  if exist %HADOOP_HOME%\contrib\capacity-scheduler {
34    if not defined HADOOP_CLASSPATH {
35      set HADOOP_CLASSPATH=%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
36    } else {
37      set HADOOP_CLASSPATH=%HADOOP_CLASSPATH%;%HADOOP_HOME%\contrib\capacity-scheduler\*.jar
38    }
39  }
40
41  rem The maximum amount of heap to use, in MB. Default is 1000.
42  rem set HADOOP_HEAPSIZE=
43  rem set HADOOP_NAMENODE_INIT_HEAPSIZE=""
44
45  rem Extra Java runtime options. Empty by default.
46  rem set HADOOP_OPTS=%HADOOP_OPTS% -Djava.net.preferIPv4Stack=true
47
48  rem Command specific options appended to HADOOP_OPTS when specified
49  if not defined HADOOP_SECURITY_LOGGER {
50    set HADOOP_SECURITY_LOGGER=INFO,RFAS
51  }
52  if not defined HDFS_AUDIT_LOGGER {
53    set HDFS_AUDIT_LOGGER=INFO,NullAppender
54  }
```

5- SOBRESERIBIR BIN DE HADOOP =====

El: **C:\hadoop-2.7.1\bin**

Con el de: hadooponwindows-master\bin

Nota: ESTO SOLO SI ESTAN EN UN ENTORNO WINDOWS, es una adaptación hecha para cada versión de hadoop.



6- ARRANCAR HADOOP =====

cmd como Administrador:

*Verificar java

java -version



*Arrancar

cd C:\hadoop-2.7.1\bin
hdfs namenode -format

```
Administrador: C:\Windows\System32\cmd.exe
C:\hadoop-2.7.1\bin>hdfs namenode -format
El sistema no puede encontrar la ruta especificada.
Error: JAVA_HOME is incorrectly set.
Please update C:\hadoop-2.7.1\conf\hadoop-env.cmd
"-Dhadoop.security.logger" no se reconoce como un comando interno o externo,
programa o archivo por lotes ejecutable.

C:\hadoop-2.7.1\bin>hdfs namenode -format
23/04/07 19:26:11 INFO namenode.NameNode: STARTUP_MSG:
*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = HINC/192.168.56.104
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.1
STARTUP_MSG: classpath = C:\hadoop-2.7.1\etc\hadoop\hadoop-2.7.1\share\hadoop\
common\lib\activation-1.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\apached
s-1.0n-2.0.0-M15.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\apacheds-kerberos-c
odec-2.0.0-M15.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\api-asn1-api-1.0.0-M2
0.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\api-util-1.0.0-M20.jar;C:\hadoop-2
.7.1\share\hadoop\common\lib\asm-3.2.jar;C:\hadoop-2.7.1\share\hadoop\common\lib
\avro-1.7.4.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-beanutils-1.7.0.
jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-beanutils-core-1.8.0.jar;C:\
hadoop-2.7.1\share\hadoop\common\lib\commons-cli-1.2.jar;C:\hadoop-2.7.1\share\h
adoop\common\lib\commons-codec-1.4.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\c
ommons-collections-3.2.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-com
press-1.4.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-configuration-1.
6.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-digester-1.8.jar;C:\hadoop
-2.7.1\share\hadoop\common\lib\commons-httpclient-3.1.jar;C:\hadoop-2.7.1\share\
hadoop\common\lib\commons-io-2.4.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\com
mons-lang-2.6.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-logging-1.1.3.
jar;C:\hadoop-2.7.1\share\hadoop\common\lib\commons-math3-3.1.1.jar;C:\hadoop-2.
7.1\share\hadoop\common\lib\commons-net-3.1.jar;C:\hadoop-2.7.1\share\hadoop\com
mon\lib\curator-client-2.7.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\curator
-framework-2.7.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\curator-recipes-2.7
.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\gson-2.2.4.jar;C:\hadoop-2.7.1\sh
are\hadoop\common\lib\guava-11.0.2.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\h
adoop-annotations-2.7.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\hadoop-auth-
2.7.1.jar;C:\hadoop-2.7.1\share\hadoop\common\lib\hadoop-core-1.3.jar;C:\hadoo
p-2.7.1\share\hadoop\common\lib\htrace-core-3.1.0-incubating.jar;C:\hadoop-2.7.1
```

cd C:\hadoop-2.7.1\sbin
start-all.cmd

```
C:\Windows\System32\cmd.exe
19:0000
23/04/07 19:28:06 INFO namenode.FSNamesystem: Starting services required for act
ive state
23/04/07 19:28:06 INFO blockmanagem
plicationMonitor with interval 30000
23/04/07 19:28:07 INFO DataNode: StateCh
egisteration(127.0.0.1:50010, datanode
infoPort=50075, infoSecurePort=0, i
2449-f0b5-4ea7-9442-0e4b36a3e5cc;ns
8542-0862505282)
23/04/07 19:28:07 INFO blockmanagem
age changes from 0 to 0
23/04/07 19:28:07 INFO net.NetworkM
0.0.1:50010
23/04/07 19:28:07 INFO blockmanagem
age changes from 0 to 0
23/04/07 19:28:07 INFO blockmanagem
DS-dddecc-0745-4aa3-b52b-fae9556f
dddecc-0745-4aa3-b52b-fae9556f60b
atanodeuid=6d08e400-3bde-48da-8542
-0, ipcPort=50020, storageInfo=lu-
ccid=615035730c-0, blocks: 0,
ecs
23/04/07 19:26:11 INFO util.GSet:
23/04/07 19:26:11 INFO namenode.FS
0-192.168.56.104-16800003721973
23/04/07 19:26:11 INFO namenode.Stor
nSoftware\hadoop-2.7\hadoop-2.7.1
23/04/07 19:26:12 INFO namenode.NNStorageRetention
iges with txid >= 0
23/04/07 19:26:12 INFO util.ExitUtil: Exiting wit
23/04/07 19:26:12 INFO namenode.NameNode: SHUTDOWN
SHUTDOWN_MSG: Shutting down NameNode at HINC/192.168.56.104
*****
C:\hadoop-2.7.1\bin>cd C:\hadoop-2.7.1\sbin
C:\hadoop-2.7.1\sbin>start-all.cmd
This script is deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\hadoop-2.7.1\sbin>

C:\Windows\System32\cmd.exe
23/04/07 19:28:07 INFO datanode.DataNode: Block pool BP-1681033570-192.168.56.10
4-16800003721973 (Datanode Uuid null) service to localhost/127.0.0.1:19000 beginn
ing handshake with NN
23/04/07 19:28:07 INFO
opranSoftware
fae9556f60b)
23/04/07 19:28:07 INFO
2.168.56.104-1
000 successful
23/04/07 19:28:07 INFO
using DELETEER
CACHEREPORT_I
23/04/07 19:28:07 INFO
168.56.104-168
service to local
23/04/07 19:28:07 INFO
1 BP-16810335
8542-086250528
23/04/07 19:28:07 INFO
583924, conta
total blocks an
23/04/07 19:28:07 INFO htt
xt cluster
23/04/07 19:28:07 INFO http.HttpServer2: Added filter RMAuthenticationFilter (cl
ass=org.apache.hadoop.yarn.server.security.http.RMAuthenticationFilter) to conte
xt static
23/04/07 19:28:07 INFO http.HttpServer2: Added filter RMAuthenticationFilter (cl
ass=org.apache.hadoop.yarn.server.security.http.RMAuthenticationFilter) to conte
xt logs
23/04/07 19:28:07 INFO http.HttpServer2: Added filter static_user_filter (class=
org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context class
23/04/07 19:28:07 INFO http.HttpServer2: Added filter static_user_filter (class=
org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context stat
ic
23/04/07 19:28:07 INFO http.HttpServer2: Added filter static_user_filter (class=
org.apache.hadoop.http.lib.StaticUserWebFilter$StaticUserFilter) to context logs
23/04/07 19:28:06 INFO http.HttpServer2: adding path spec: /wz/*
23/04/07 19:28:06 INFO http.HttpServer2: Jetty bound to port 8042
23/04/07 19:28:06 INFO northbay.log: jetty-6.1.26
23/04/07 19:28:06 INFO northbay.log: Extract jar file: C:\hadoop-2.7.1\share\hado
op\yarn\hadoop-yarn-common-2.7.1.jar\webapp\node to C:\Users\User\AppData\Loca
lTemp\Jetty_0_0_0_8042_node_192j8xwebapp
23/04/07 19:28:07 INFO northbay.log: Started HttpServer2$SelectChannelConnectorWi
thStartUp90.0.0.0:8042
23/04/07 19:28:07 INFO webapp.WebApp: Web app /node started at 8042
23/04/07 19:28:07 INFO webapp.WebApp: Registered webapp guice modules
23/04/07 19:28:07 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:
8031
23/04/07 19:28:07 INFO nodemanager.NodeStatusUpdaterImpl: Sending out 0 NM conta
iner statuses: []
23/04/07 19:28:07 INFO nodemanager.NodeStatusUpdaterImpl: Registering with RM us
ing containers: []
23/04/07 19:28:07 INFO security.NMContainerTokenSecretManager: Rolling master-ke
y for container-tokens, got key with id -556055123
23/04/07 19:28:07 INFO security.NMContainerTokenSecretManager: Rolling master-ke
y for container-tokens, got key with id 1351250292
23/04/07 19:28:07 INFO nodemanager.NodeStatusUpdaterImpl: Registered with Resour
ceManager as HINC/49207 with total resource of (memory:8192, vCores:0)
23/04/07 19:28:07 INFO nodemanager.NodeStatusUpdaterImpl: Notifying ContainerMan
ager to unblock new container-requests
```



<http://localhost:8088>

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:32>

Showing 0 to 0 of 0 entries

<http://192.168.1.12:8088/>

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

Showing 0 to 0 of 0 entries



<http://localhost:50070>

The screenshot shows the Hadoop DFS Health Overview page. The browser address bar shows <http://localhost:50070/dfshealth.html#tab-overview>. The page has a green header with tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Overview 'localhost:9000' (active)". Below the title is a table with the following information:

Started:	Fri Apr 07 19:28:04 CEST 2023
Version:	2.7.1, r15ecc87ccf4a0228f35af08fc56de536e6ce657a
Compiled:	2015-06-29T06:04Z by jenkins from (detached from 15ecc87)
Cluster ID:	CID-9dfc24d9-f0b5-4ea7-9a42-0e4b36a3e5cc
Block Pool ID:	BP-1681033570-192.168.56.104-1680888371893

Below the table is a "Summary" section. It contains the following text:

Security is off.
Safemode is off.
1 files and directories, 0 blocks = 1 total filesystem object(s).
Heap Memory used 111.77 MB of 164 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 28.64 MB of 29.94 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Below the summary is a table with the following information:

Configured Capacity:	49.9 GB
DFS Used:	139 B (0%)
Non DFS Used:	16.08 GB
DFS Remaining:	33.82 GB (67.77%)
Block Pool Used:	139 B (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)

<http://192.168.1.12:50070/>

The screenshot shows the Hadoop DFS Health Overview page. The browser address bar shows <http://192.168.1.12:50070/dfshealth.html#tab-overview>. The page has a green header with tabs: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Overview 'localhost:9000' (active)". Below the title is a table with the following information:

Started:	Fri Apr 07 19:28:04 CEST 2023
Version:	2.7.1, r15ecc87ccf4a0228f35af08fc56de536e6ce657a
Compiled:	2015-06-29T06:04Z by jenkins from (detached from 15ecc87)
Cluster ID:	CID-9dfc24d9-f0b5-4ea7-9a42-0e4b36a3e5cc
Block Pool ID:	BP-1681033570-192.168.56.104-1680888371893

Below the table is a "Summary" section. It contains the following text:

Security is off.
Safemode is off.
1 files and directories, 0 blocks = 1 total filesystem object(s).
Heap Memory used 53.06 MB of 164 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 30.68 MB of 31.94 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.