

하둡 설치

Hadoop Installation.

2020. 09.25 홍태희



하둡 & 설치

Enjoy your stylish business and campus life with BIZCAM

JAVA

VERSION

<https://www.oracle.com/kr/java/technologies/javase/javase-jdk8-downloads.html>

JDK 1.8

이상 버전으로 설치해야 됨

HADOOP

VERSION

<http://apache.mirror.cdnetworks.com/hadoop/common/hadoop-3.3.0/>

하둡 3.3.0 설치

HIVE

VERSION

하이브 3.1.2 설치

<http://apache.mirror.cdnetworks.com/hive/hive-3.1.2/>

설치 참고주소 : https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-common/SingleCluster.html#Standalone_Operation

1. 압축풀기

```
tar xvfz hadoop-3.3.0.tar.gz
```

2. 하둡 설정 디렉토리로 이동

```
cd hadoop-3.3.0/etc/hadoop/
```

3. 환경파일(hadoop-env.sh) 수정

```
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_65.jdk/Contents/Home
```

```
export HADOOP_HOME=/Users/hongtebari/Platform/hadoop-3.3.0
```

```
export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop
```

설치 참고주소 : https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-common/SingleCluster.html#Standalone_Operation

4. 설정파일 수정 : [core-site.xml](#)

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

5. 설정파일 수정 : [hdfs-site.xml](#)

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```



5. SSH 로그인 가능 여부 테스트

```
ssh localhost
```

→ 로그인 되지 않는 경우,

```
ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
```

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
chmod 0600 ~/.ssh/authorized_keys
```

6. 하둡 네임노드 포맷

```
bin/hdfs namenode -format
```

...

```
2020-09-20 14:00:23,053 INFO util.GSet: 0.25% max memory 3.6 GB = 9.1 MB2020-09-20 14:00:23,053 INFO util.GSet: capacity = 2^20 = 1048576 entries
```

```
2020-09-20 14:00:23,060 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
```

```
2020-09-20 14:00:23,060 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
```

...

```
2020-09-20 14:00:23,064 INFO util.GSet: VM type = 64-bit
```

```
2020-09-20 14:00:23,065 INFO util.GSet: 0.029999999329447746% max memory 3.6 GB = 1.1 MB
```

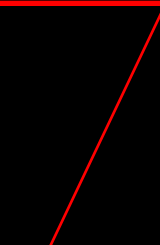
```
2020-09-20 14:00:23,065 INFO util.GSet: capacity = 2^17 = 131072 entries
```

```
Format filesystem in Storage Directory root= /tmp/hadoop-hongtebari/dfs/name; location= null ? (Y or N) Y
```

7. DFS 데몬 실행

sbin/start-dfs.sh

```
→ hadoop-3.3.0 sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [hongtebariui-MacBook-Pro.local]
2020-09-20 14:18:06,408 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
→ hadoop-3.3.0 jps
16099 NameNode
16200 DataNode
73512 RunJar
16524 Jps
16335 SecondaryNameNode
```



Optional 설정 : hadoop-env.sh

```
export HADOOP_COMMON_LIB_NATIVE_DIR=${HADOOP_PREFIX}/lib/native
export HADOOP_OPTS="${HADOOP_OPTS} -Djava.library.path=${HADOOP_PREFIX}/lib"
```

8. Name Node Web Interface 확인

<http://localhost:9870/>

--> 웹인터페이스 접속이 잘되면 성공

localhost:9870/dfshealth.html#tab-overview

ie3 Wind Ms... Quickstart Center... AWS 서울 e학습터 :: 홈 Classes Popit www.google.com Dia

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Overview 'localhost:9000' (active)

Started:	Sun Sep 20 14:17:59 +0900 2020
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af
Compiled:	Tue Jul 07 03:44:00 +0900 2020 by brahma from branch-3.3.0
Cluster ID:	CID-309675fc-d2e8-43ce-afad-0519c53a0d2f
Block Pool ID:	BP-597156304-127.0.0.1-1600443775516

Summary

Security is off.

Safemode is off.

13 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 13 total filesystem object(s).

Heap Memory used 132.39 MB of 353 MB Heap Memory. Max Heap Memory is 3.56 GB.

Non Heap Memory used 49.44 MB of 51.56 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	931.64 GB
----------------------	-----------

설치 참고주소 : https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-common/SingleCluster.html#Standalone_Operation

9. 설정파일 수정 : [mapred-site.xml](#)

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.application.classpath</name>
    <value>${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/*:${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/lib/* </value>
  </property>
</configuration>
```

설치 참고주소 : https://hadoop.apache.org/docs/r3.3.0/hadoop-project-dist/hadoop-common/SingleCluster.html#Standalone_Operation

10. 설정파일 수정 : [yarn-site.xml](#)

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,
HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME
    </value>
  </property>
</configuration>
```



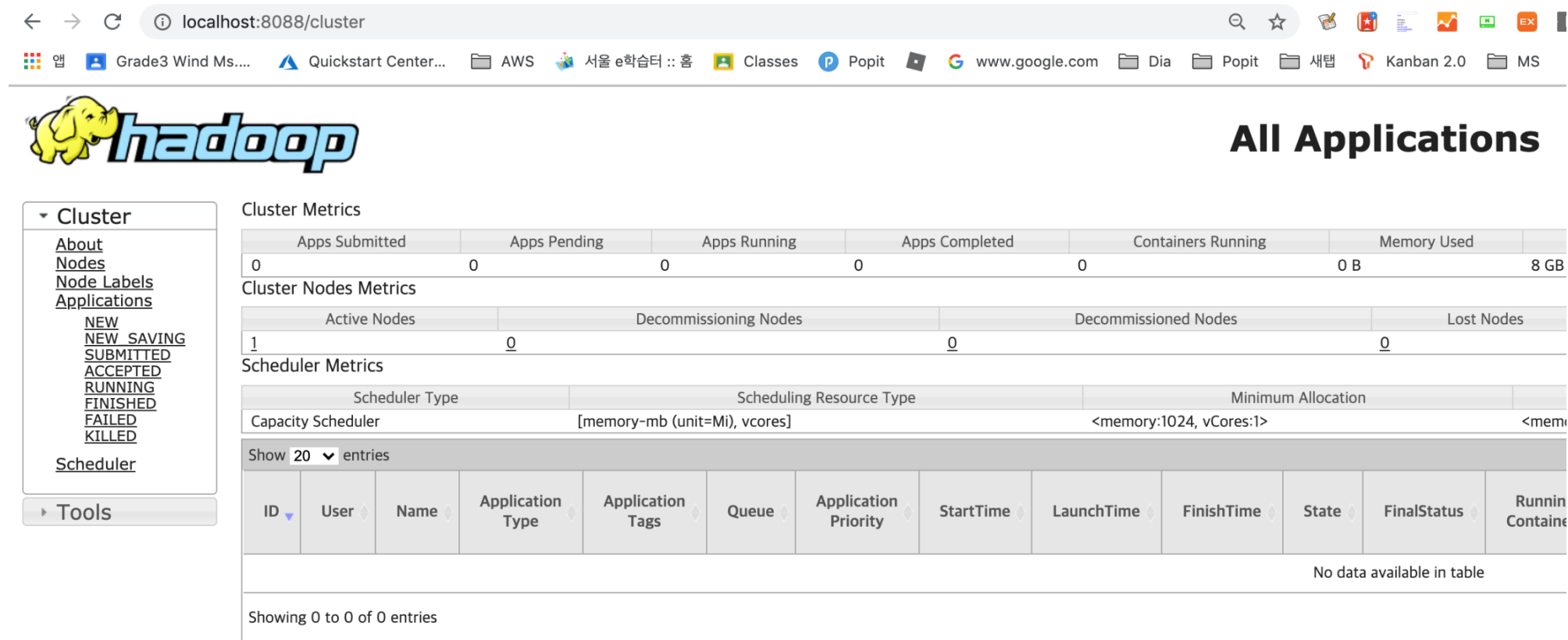
11. YARN 데몬 실행

sbin/start-yarn.sh

```
→ hadoop-3.3.0 sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
→ hadoop-3.3.0 jps
16099 NameNode
16200 DataNode
73512 RunJar
23581 ResourceManager
23678 NodeManager
23854 Jps
16335 SecondaryNameNode
```

12. Resource Manager 웹 인터페이스 확인

<http://localhost:8088/>



The screenshot shows the Hadoop Resource Manager web interface. The browser address bar displays `localhost:8088/cluster`. The page features the Hadoop logo on the left and the title "All Applications" on the right. A sidebar on the left contains navigation links under "Cluster" and "Tools". The main content area displays several metrics tables:

- Cluster Metrics:** A table showing application status counts.

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used
0	0	0	0	0 B	8 GB
- Cluster Nodes Metrics:** A table showing node status counts.

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0
- Scheduler Metrics:** A table showing scheduler configuration.

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mb), vcores]	<memory:1024, vCores:1>
- Applications Table:** A table with columns for application details. The table is currently empty, showing "No data available in table".

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers
No data available in table												

At the bottom, it indicates "Showing 0 to 0 of 0 entries".



감사합니다.