

하둡 멀티 노드 클러스터 구성

하둡 멀티 노드 클러스터

요구사항

두 개 이상의 하둡 멀티 노드 클러스터를 구성하고 체계에 대한 활용 목적과 시스템 디자인 이유 등에 대한 설명을 포함하여 실제 설치 과정을 각종 캡처 화면과 함께 설명하세요.

01

하둡 멀티 노드
클러스터 구성

02

시스템 디자인 이유

03

설치 과정

01

하둡 멀티노드 클러스터 구성

하둡 멀티노드 클러스터 구성

4개의 서버로 운영

myserver (1개) : namenode

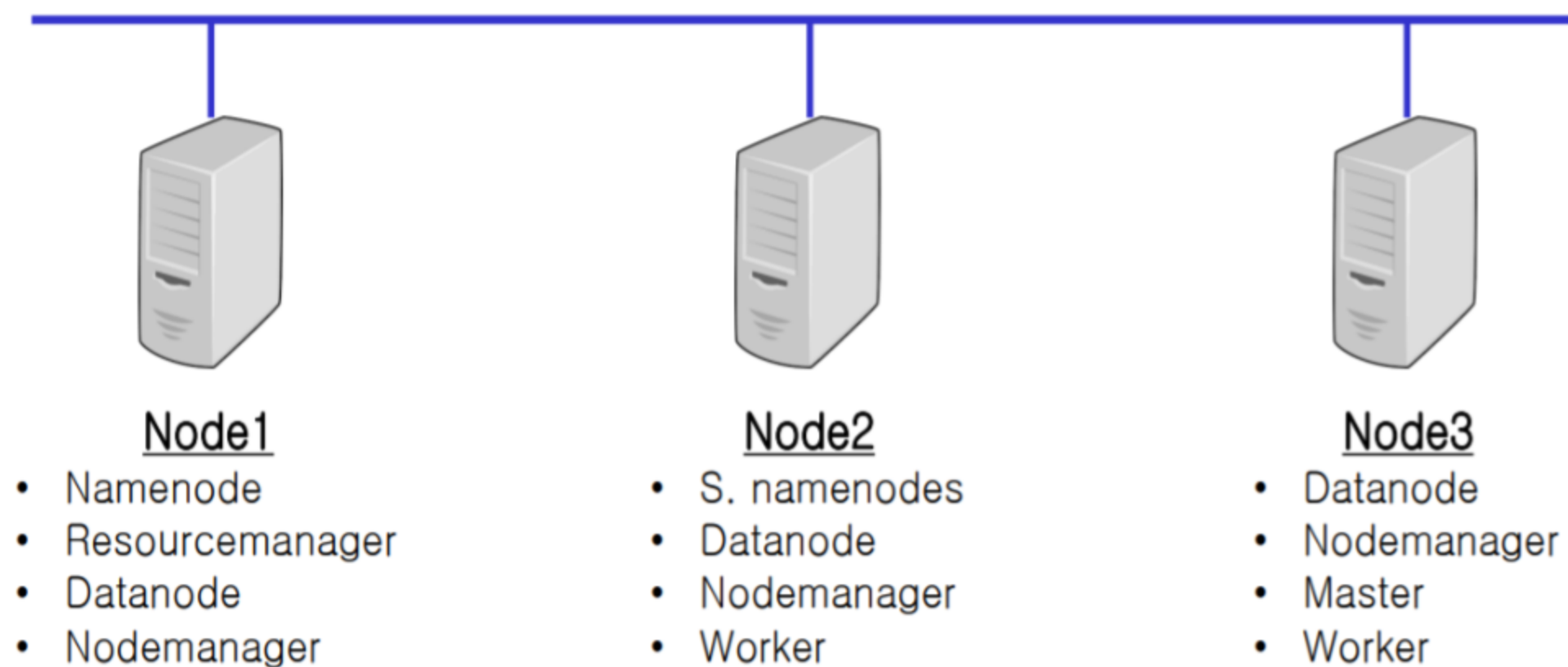
node1 (4개) : resource manager, datanode, nodemanager, worker

node2 (4개) : secondary namenode, datanode, nodemanager, worker

node3 (3개) : datanode, nodemanager, master

시스템 디자인 이유

Cluster Design



- Namenode와 보조 Namenode 분산
- 하둡 Namenode와 스파크 Master의 분산
- Namenode와 ResourceManager의 분산
- Datanodes와 파일 복사 규모 3/2, Workers 2 운용 규모 고려

시스템 디자인 이유

myserver (1개) : namenode

node1 (4개) : resource manager, datanode, nodemanager, worker

node2 (4개) : secondary namenode, datanode, nodemanager, worker

node3 (3개) : datanode, nodemanager, master

02

설치 과정

Path 설정

자바, 하둡, 스파크

우분투 20.04, JAVA 1.8 이상, 하둡 ver. 3.3.1,
스파크 ver. 3.2.0 설치

```
#java home setup
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:$JAVA_HOME/bin

#Hadoop Related Options
export HADOOP_HOME=/home/kmk/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

export SPARK_HOME=/home/kmk/spark
export PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin
```


호스트서버 및 RSA 인증 설정

호스트서버 설정

- 서로 통신 호스트로 인식될 수 있도록

RSA 인증

- 패스워드 없이 로그인 가능하도록

`sudo nano /etc/hosts`

```
GNU nano 4.8
127.0.0.1 localhost
192.168.219.101 myserver
192.168.219.146 node1
192.168.219.133 node2
192.168.219.105 node3
```

RSA 인증

```
connection to node1 closed.
kmk@node2:~$ scp ~/.ssh/id_rsa.pub kmk@node3:/home/kmk
The authenticity of host 'node3 (192.168.219.105)' can't be established.
ECDSA key fingerprint is SHA256:pWysR53ubSgpcdZq+Gsle33t6gLaPP4XKBac4pl3ARg.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'node3,192.168.219.105' (ECDSA) to the list of known
hosts.
kmk@node3's password:
id_rsa.pub                                100% 563   367.3KB/s   00:00
kmk@node2:~$ cat ~/id_rsa.pub >> ~/.ssh/authorized_keys
kmk@node2:~$ ssh node3
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.4.0-91-generic x86_64)
```

파일 수정

〈core-site.xml 파일 수정〉

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

〈hdfs-site.xml 파일 수정〉

```
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>/home/kmk/hdata/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/home/kmk/hdata/datanode</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>node2:50090</value>
  </property>
</configuration>
```

myserver (1개) : namenode

node1 (4개) : resource manager, datanode, nodemanager, worker

node2 (4개) : secondary namenode, datanode, nodemanager, worker

node3 (3개) : datanode, nodemanager, master

파일 수정

myserver (1개) : namenode

node1 (4개) : resource manager, datanode, nodemanager, worker

node2 (4개) : secondary namenode, datanode, nodemanager, worker

node3 (3개) : datanode, nodemanager, master

〈yarn-site.xml 파일 수정〉

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>node1</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

〈mapred-site.xml 파일 수정〉

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```


파일 수정

〈workers 파일 수정〉

```
node1
node2
node3
```

〈spark-env.sh 파일 수정〉

```
# - SPARK_PUBLIC_DNS, to set the public DNS name of the driver program
# - SPARK_LOCAL_DIRS, storage directories to use on this node for shuffle a
# - MESOS_NATIVE_JAVA_LIBRARY, to point to your libmesos.so if you use Mesos

# Options read in YARN client/cluster mode
# - SPARK_CONF_DIR, Alternate conf dir. (Default: ${SPARK_HOME}/conf)
# - HADOOP_CONF_DIR, to point Spark towards Hadoop configuration files
# - YARN_CONF_DIR, to point Spark towards YARN configuration files when you
# - SPARK_EXECUTOR_CORES, Number of cores for the executors (Default: 1).
# - SPARK_EXECUTOR_MEMORY, Memory per Executor (e.g. 1000M, 2G) (Default: 1
# - SPARK_DRIVER_MEMORY, Memory for Driver (e.g. 1000M, 2G) (Default: 1G)

# Options for the daemons used in the standalone deploy mode
# - SPARK_MASTER_HOST, to bind the master to a different IP address or host
SPARK_MASTER_HOST='192.168.219.105'
# - SPARK_MASTER_PORT / SPARK_MASTER_WEBUI_PORT, to use non-default ports f
# - SPARK_MASTER_OPTS, to set config properties only for the master (e.g. "
# - SPARK_WORKER_CORES, to set the number of cores to use on this machine
# - SPARK_WORKER_MEMORY, to set how much total memory workers have to give
# - SPARK_WORKER_PORT / SPARK_WORKER_WEBUI_PORT, to use non-default ports f
```

myserver (1개) : namenode

node1 (4개) : resource manager, datanode, nodemanager, worker

node2 (4개) : secondary namenode, datanode, nodemanager, worker

node3 (3개) : datanode, nodemanager, master

파일 수정

<workers 파일 수정>

```
# A Spark Worker will be started on each of the machines
node1
node2
```

<파일 전송>

```
workers 100% 868 554.3KB/s 00:00
kmk@node3:~$ scp spark/conf/spark-env.sh node1:/home/kmk/spark/conf/
spark-env.sh 100% 4464 3.8MB/s 00:00
kmk@node3:~$ scp spark/conf/workers node1:/home/kmk/spark/conf/
workers 100% 868 554.3KB/s 00:00
```

myserver (1개) : namenode

node1 (4개) : resource manager, datanode, nodemanager, worker

node2 (4개) : secondary namenode, datanode, nodemanager, worker

node3 (3개) : datanode, nodemanager, master

myserver (17H) : namenode

```
kmk@myserver:~$ jps  
26741 NameNode  
27631 Jps  
kmk@myserver:~$
```

node1 (47H) : resource manager, datanode, nodemanager, worker

```
kmk@node1:~$ jps  
26946 ResourceManager  
27541 Jps  
26108 NodeManager  
25773 Worker  
25935 DataNode  
kmk@node1:~$
```

node2 (4개) : secondary namenode, datanode, nodemanager, worker

```
kmk@node2:~$ jps
25393 Worker
26882 Jps
25556 DataNode
25870 NodeManager
25727 SecondaryNameNode
kmk@node2:~$
```

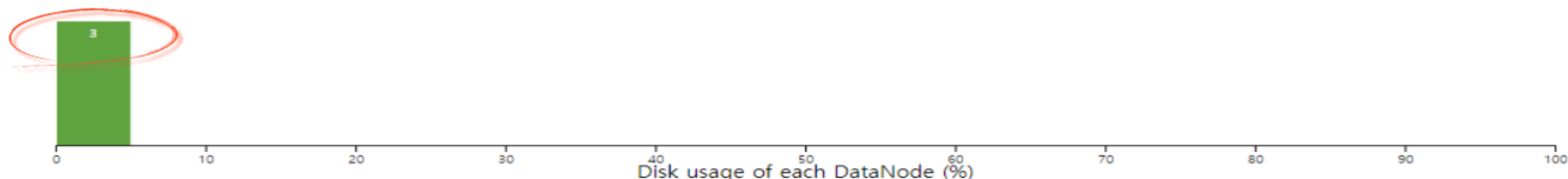
node3 (3개) : datanode, nodemanager, master

```
kmk@node3:~$ jps
25975 NodeManager
26121 Jps
25565 DataNode
25023 Master
kmk@node3:~$
```

Datanode Information

✓ In service ⚠ Down ⏸ Decommissioning ⚡ Decommissioned ⛔ Decommissioned & dead
🔧 Entering Maintenance 🛠 In Maintenance 🛑 In Maintenance & dead

Datanode usage histogram



In operation

DataNode State AllShow 25 entriesSearch:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
✓ /default-rack/node3:9866 (192.168.219.105:9866)	http://node3:9864	0s	16m	32 KB	7.15 GB	18.57 GB	0	32 KB (0%)	3.3.1
✓ /default-rack/node1:9866 (192.168.219.146:9866)	http://node1:9864	2s	16m	32 KB	7.16 GB	18.57 GB	0	32 KB (0%)	3.3.1
✓ /default-rack/node2:9866 (192.168.219.133:9866)	http://node2:9864	2s	16m	32 KB	7.17 GB	18.57 GB	0	32 KB (0%)	3.3.1

Showing 1 to 3 of 3 entries

Previous 1 Next

Spark 3.2.0

Jobs

Stages

Storage

Environment

Executors

Spark shell application UI

Spark Jobs (?)

User: kmk

Total Uptime: 13 min

Scheduling Mode: FIFO

Completed Jobs: 9

Event Timeline

Enable zooming

Executors

Added

Removed

Jobs

Succeeded

Failed

Running

5 December 16:33

5 December 16:34

Completed Jobs (9)

Page: 1

1 Pages. Jump to 1. Show 100 items in a page. Go

Job Id *	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
8	top at <console>:24 top at <console>:24	2021/12/05 16:34:34	96 ms	1/1	1/1
7	first at <console>:24 first at <console>:24	2021/12/05 16:34:34	40 ms	1/1	1/1
6	foreach at <console>:24 foreach at <console>:24	2021/12/05 16:34:33	53 ms	1/1	1/1
5	foreach at <console>:24 foreach at <console>:24	2021/12/05 16:34:28	38 ms	1/1	1/1
4	foreach at <console>:24 foreach at <console>:24	2021/12/05 16:34:26	94 ms	1/1	1/1
3	take at <console>:24 take at <console>:24	2021/12/05 16:34:24	0.1 s	1/1	1/1
2	count at <console>:24 count at <console>:24	2021/12/05 16:34:16	57 ms	1/1	1/1
1	count at <console>:24 count at <console>:24	2021/12/05 16:34:10	94 ms	1/1	1/1
0	count at <console>:23 count at <console>:23	2021/12/05 16:34:01	2 s	1/1	1/1

02



Spark Master at spark://192.168.219.105:7077

URL: spark://192.168.219.105:7077

Alive Workers: 2

Cores in use: 2 Total, 0 Used

Memory in use: 2.0 GiB Total, 0.0 B Used

Resources in use:

Applications: 0 Running, 0 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

▼ Workers (2)

Worker Id		Address	State	Cores	Memory	Resources
worker-20211205163248-192.168.219.133-46397	node2	192.168.219.133:46397	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	
worker-20211205163248-192.168.219.146-37885	node1	192.168.219.146:37885	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	

Thank you
발표를 들어주셔서
감사합니다 :)