□ 수행평가 – 빅데이터 분석시스템 구축 과정

1. Hadoop을 설치 한다.

하둡과 jdk의 압축을 푼다

```
jdk1.8.0_161/jre/Welcome.html
jdk1.8.0_161/jre/README
jdk1.8.0_161/README.html
[root@master 다운로드]# tar xvfz hadoop-1.2.1.tar.gz ▉
```

```
dk1.8.0_161/jre/README
dk1.8.0_161/README.html
root@master 다운로드]# tar xvfz jdk-8u161-linux-x64.tar.gz ▉
```

/Usr/local/ 에 폴더를 복사한다

```
jdk-8u161-linux-x64.tar.gz  jdk1.8.0_161/
[root@master 다운로드]# cp -r jdk1.8.0_161/ /usr/local
[root@master 다운로드]# cp -r hadoop-1.2.1 /usr/local
root@master 다운로드]#
```

2. 설정 파일을 설정 한다.

환경설정 파일 초기화(jdk버전 확인 필수)

```
52 JAVA_HOME=/usr/local/jdk-9.0.4
53 HADOOP_HOME=/usr/local/hadoop-1.2.1
54 CLASSPATH=/usr/local/jdk-9.0.4/lib
55 export JAVA_HOME
56 export HADOOP_HOME
57 export CLASSPATH
58 PATH=$HADOOP_HOME/bin:$JAVA_HOME/bin:$PATH
59
```

```
[root@master bin]# ln -s /usr/local/jdk1.8.0_161/bin/java java
[root@master bin]# cd
[root@master ~]# java -version
java version "1.8.0_161"
Java(TM) SE Runtime Environment (build 1.8.0_161-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.161-b12, mixed mode)
[root@master ~]# ls -la java
ls: cannot access java: 그런 파일이나 디렉터리가 없습니다
[root@master ~]# cd /usr/bin
[root@master bin]# ls -la java
lrwxrwxrwx. 1 root root 32  3월  22 11:42 java -> /usr/local/jdk1.8.
0_161/bin/java
```

Ssh 설정

```
[root@master ~]# ssh-keygen -t dsa -P '' -f ~/.ssh/id_dsa
Generating public/private dsa key pair.
Created directory '/root/.ssh'.
Your identification has been saved in /root/.ssh/id_dsa.
Your public key has been saved in /root/.ssh/id_dsa.pub.
The key fingerprint is:
c7:d5:f1:e6:4a:72:1c:d2:2b:b6:f6:61:df:1d:fb:13 root@master
The key's randomart image is:
+--[ DSA 1024]----+
|             .   |
|            o o  |
|           o + o|
|        . . o = |
|       S o + = .|
|        . . * E |
|           o +..|
|          . o +=|
|            ..B|
+----------------+
[root@master ~]# cd .ssh
[root@master .ssh]# ls
id_dsa  id_dsa.pub
[root@master .ssh]# 
```

```
[root@master .ssh]# cat id_dsa.pub >> authorized_keys
[root@master .ssh]# ls
authorized_keys  id_dsa  id_dsa.pub
[root@master .ssh]# 
```

**vi core-site.xml, vi hdfs-site.xml, vi mapred-site.xml 를 차례로 아래와 같이 기
입.**

```xml
<configuration>

<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>
<property>
<name>dfs.tmp.dir</name>
<value>/usr/local/hadoop-1.2.1/tmp</value>
</property>

</configuration>
~
~
~
```

```xml
<configuration>

<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.name.dir</name>
<value>/usr/local/hadoop-1.2.1/name</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/usr/local/hadoop-1.2.1/date</value>
</property>

<property>
<name>dfs.webhdfs.enabled</name>
<value>true</value>
</property>
```

```xml
<configuration>

<property>
<name>mapred.job.tracker</name>
<value>localhost:9001</value>
</property>

</configuration>
~
~
```

방화벽 해제

```
root@master conf]# systemctl stop firewalld
root@master conf]# systemctl disable firewalld
root@master conf]#
```

Cd /usr/local/hadoop-1.2.1/conf 에서 hadoop-env.sh 열어서 아래와 같이 추가

```
 8 # The java implementation to use.  Required.
 9 export JAVA_HOME=/usr/local/jdk1.8.0_161
10 export HADOOP_HOME_WARN_SUPPRESS="TRUE"
```

하고 재부팅 한후 하둡 실행시켜본다

3.Hadoop을 가동 한다.

```
[root@master ~]# hadoop
Usage: hadoop [--config confdir] COMMAND
where COMMAND is one of:
  namenode -format     format the DFS filesystem
  secondarynamenode    run the DFS secondary namenode
  namenode             run the DFS namenode
  datanode             run a DFS datanode
  dfsadmin             run a DFS admin client
  mradmin              run a Map-Reduce admin client
  fsck                 run a DFS filesystem checking utility
  fs                   run a generic filesystem user client
  balancer             run a cluster balancing utility
  oiv                  apply the offline fsimage viewer to an fsimage
  fetchdt              fetch a delegation token from the NameNode
  jobtracker           run the MapReduce job Tracker node
  pipes                run a Pipes job
  tasktracker          run a MapReduce task Tracker node
  historyserver        run job history servers as a standalone daemon
  job                  manipulate MapReduce jobs
  queue                get information regarding JobQueues
  version              print the version
  jar <jar>            run a jar file
  distcp <srcurl> <desturl> copy file or directories recursively
  distcp2 <srcurl> <desturl> DistCp version 2
  archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
  classpath            prints the class path needed to get the
                       Hadoop jar and the required libraries
  daemonlog            get/set the log level for each daemon
 or
  CLASSNAME            run the class named CLASSNAME
Most commands print help when invoked w/o parameters.
```

4.Hadoop 관리화면 가동

http://localhost:50070 입력

NameNode 'localhost:9000'

Started: Thu Mar 22 14:09:58 KST 2018
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

6 files and directories, 2 blocks = 8 total. Heap Size is 59.94 MB / 966.69 MB (6%)

Configured Capacity	:	38.03 GB
DFS Used	:	24 KB
Non DFS Used	:	7.05 GB
DFS Remaining	:	30.98 GB
DFS Used%	:	0 %
DFS Remaining%	:	81.46 %
Live Nodes	:	1
Dead Nodes	:	0
Decommissioning Nodes	:	0
Number of Under-Replicated Blocks	:	0

NameNode Storage:

Storage Directory	Type	State
/usr/local/hadoop-1.2.1/name	IMAGE_AND_EDITS	Active

5. Hive 설치

리눅스 책 555~57p 따라서 마리아디비 깔기, profile설정까지 한다.

```
Nothing to do
[root@master ~]# cd 다운로드
[root@master 다운로드]# yum -y localinstall Maria*
[root@master 다운로드]# cp -r apache-hive-1.0.1-bin /usr/local/hive
[root@master 다운로드]# cd
[root@master ~]# vi /etc/profile
[root@master ~]# tar xvfz apache-hive-1.0.1-bin.tar.gz
```

```
root@master 다운로드]# systemctl restart mysql
root@master 다운로드]# systemctl status mysql
mysql.service - LSB: start and stop MySQL
   Loaded: loaded (/etc/rc.d/init.d/mysql)
   Active: active (running) since 목 2018-03-22 14:31:51 KST; 5s ag
)
  Process: 5063 ExecStart=/etc/rc.d/init.d/mysql start (code=exited
 status=0/SUCCESS)
   CGroup: /system.slice/mysql.service
           ├─5068 /bin/sh /usr/bin/mysqld_safe --datadir=/var/li...
           └─5142 /usr/sbin/mysqld --basedir=/usr --datadir=/var...

3월 22 14:31:51 master mysql[5063]: Starting MySQL. SUCCESS!
3월 22 14:31:51 master systemd[1]: Started LSB: start and stop...
Hint: Some lines were ellipsized, use -l to show in full.
root@master 다운로드]# chkconfig mysql on
root@master 다운로드]# mysqladmin -u root password '111111'
root@master 다운로드]# mysql -u root -p mysql
Enter password: █
```

로컬호스트에서 접속하는 hive에게 권한을 부여

hive_db를 만들고 그 권한을 hive에게 부여

```
MariaDB [mysql]> grant all privileges on *.* to 'hive'@localhost'
identified by '111111';
Query OK, 0 rows affected (0.00 sec)

MariaDB [mysql]> create database hive_db;
Query OK, 1 row affected (0.00 sec)

MariaDB [mysql]> grant all privileges on hive_db.* to 'hive'@% id
entified by '111111' with grant option;
Query OK, 0 rows affected (0.00 sec)

MariaDB [mysql]> grant all privileges on hive_db.* to 'hive'@local
host' identified by '111111' with grant option;
Query OK, 0 rows affected (0.00 sec)

MariaDB [mysql]> commit
    -> ;
Query OK, 0 rows affected (0.00 sec)

MariaDB [mysql]> use hive_db
Database changed
MariaDB [hive_db]> show database;
```

mysql 작업 끝

HIVE_HOME 추가 57,58라인 추가하고 61라인에 패쓰에 하이브홈을 넣어줌

```
[root@hadoopserver ~]# vi /etc/profile

52 JAVA_HOME=/usr/local/jdk1.8.0_161
53 HADOOP_HOME=/usr/local/hadoop-1.2.1
54 CLASSPATH=/usr/local/jdk1.8.0_161/lib
55 HIVE_HOME=/usr/local/hive
56 export HIVE_HOME
57 export JAVA_HOME
58 export HADOOP_HOME
59 export CLASSPATH
60 PATH=$HADOOP_HOME/bin:$JAVA_HOME/bin:$PATH
61
```

```
[root@master conf]# touch hive-site.xml
[root@master conf]# vi hive-site.xml
[root@master conf]# cd ..
[root@master hive]# ls
LICENSE   README.txt         bin    examples   lib
NOTICE    RELEASE_NOTES.txt  conf   hcatalog   scripts
[root@master hive]# cd
[root@master ~]# cd 다운로드
[root@master 다운로드]# cp mariadb-java-client-1.3.5.jar /usr/local
/hive/lib
[root@master 다운로드]# cd
[root@master ~]# cd /usr/local/hive/lib/
[root@master lib]# ls
```

```
[root@master ~]# hive
18/03/22 15:30:52 WARN conf.HiveConf: DEPRECATED: Configuration
perty hive.metastore.local no longer has any effect. Make sure
rovide a valid value for hive.metastore.uris if you are connect
to a remote metastore.
18/03/22 15:30:52 WARN conf.HiveConf: HiveConf of name hive.met
re.local does not exist

Logging initialized using configuration in jar:file:/usr/local/
/lib/hive-common-1.0.1.jar!/hive-log4j.properties
hive> ▉
```

6. Java Application 연동 테스트

태이블 생성

```
Logging initialized using configuration in
/lib/hive-common-1.0.1.jar!/hive-log4j.pro
hive> CREATE TABLE airline_delay(
    > Year INT,
    > MONTH INT,
    > DayofMonth INT,
    > DayofWeek INT,
    > DepTime INT,
    > CRSDepTime INT,
    > ArrTime INT,
    > CRSArrTime INT,
    > UniqueCarrier STRING,
    > FlightNum INT,
    > TailNum STRING,
    > ActualElapsedTime INT,
    > CRSElapsedTime INT.
```

DATA load 를 해준다

```
Loading data to table default.airline_delay partition (delayyear=20
08)
Partition default.airline_delay{delayyear=2008} stats: [numFiles=1,
 numRows=0, totalSize=689413344, rawDataSize=0]
OK
Time taken: 13.471 seconds
hive> LOAD DATA LOCAL INPATH '/root/airline/2007.csv'
    > OVERWRITE INTO TABLE airline_delay
    > PARTITION (delayYear='2007');
Loading data to table default.airline_delay partition (delayyear=20
07)
Partition default.airline_delay{delayyear=2007} stats: [numFiles=1,
 numRows=0, totalSize=702878193, rawDataSize=0]
OK
Time taken: 14.966 seconds
hive>
```

Hive --service hiveserver2 를 실행시켜 연동을 시킨다

```
e CPU 13.04 sec
2018-03-22 16:44:34,675 Stage-1 map = 100%   reduce = 67% Cumulati
ve CPU 16.07 sec
2018-03-22 16:44:42,753 Stage-1 map = 100%   reduce = 78% Cumulati
ve CPU 16.07 sec
2018-03-22 16:44:43,765 Stage-1 map = 100%   reduce = 100% Cumulat
ive CPU 17.51 sec
MapReduce Total cumulative CPU time: 17 seconds 510 msec
Ended Job = job_201803221529_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 3 Reduce: 3  Cumulative CPU: 17.51 sec   HDFS
 Read: 672110523 HDFS Write: 56 SUCCESS
Total MapReduce CPU Time Spent: 17 seconds 510 msec
OK
```

연동확인

```
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactor
18/03/22 16:43:51 INFO jdbc.HiveConnection: Will try to open clier
[["1월",233144],["4월",250085],["2월",233511],["3월",272779]]
Success....
```