

🌸 52. HIVE 설치

빅데이터 분석

=> R, Python 등을 이용해서 데이터 분석

빅데이터 저장

=> MONGODB, Hbase, Hcatlog, Mahout

분산 처리지원

=> Hive, Pig, sqoop, zookeeper

분산 배치 처리

=> 하둡(hadoop) -> MapReduce

분산 파일 관리

=> 하둡(hadoop) -> HDFS (Hadoop 분산 파일 시스템)

하이브(hive)

=> 페이스북에서 만든 오픈 소스

=> java 를 몰라도 rdbms 에 SQL 에 익숙한 사용자들을 위해서 하둡 분산 파일 시스템에 올린 데이터를 조회하는 언어

- emp.csv 의 데이터에서 이름과 월급을 조회하려면 java 를 알아야 했는데 hive 가 나오면서 java를 몰라도 SQL과 같은 언어로 조회가 가능합니다.

(HIVE)

```
select ename, sal
from emp
where job='SALESMAN';
```

하이브의 특징

1. HDFS에 있는 데이터를 SELECT 할 수 있는데 update와 delete 가 안된다.

=> 수백 TB 카톡 메시지중에 어떤 특정 메시지를 찾는 데는 하둡이 빠르다. 다른 사람이 카톡한 메시지를 update 할 일은 없다.

2. from 절의 서브쿼리 사용가능 (in line view)

3. select 문 사용시 having 절 사용 불가능

4. PL/SQL 은 지원안함 (자바로 해야합니다. Hive 2.0 가능)

오라클: 샵

하둡과 하이브: 포크레인

작은 데이터를 조회할 때는 오라클이 훨씬 빠르지만 큰 데이터를 조회할 때는 하둡이 빠르다.

HIVE 설치

1. oracle 유저로 접속해서 hive 설치 파일을 다운로드 받는다.

```
wget https://downloads.apache.org/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz
```

```
[oracle@centos ~]$ wget https://downloads.apache.org/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz
--2024-01-24 13:47:54-- https://downloads.apache.org/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 135.181.214.104, 2a01:4f9:3a:2c57::2,
...
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 90859180 (87M) [application/x-gzip]
Saving to: 'apache-hive-1.2.2-bin.tar.gz'

60% [=====] 55,009,280 3.65MB/s eta 9s w
100%[=====] 90,859,180 2.23MB/s in 37s

2024-01-24 13:48:32 (2.34 MB/s) - 'apache-hive-1.2.2-bin.tar.gz' saved [90859180/90859180]
```

2. 압축 파일의 권한을 777로 올린다.

chmod 777 apache-hive-1.2.2-bin.tar.gz

```
[oracle@centos ~]$ chmod 777 apache-hive-1.2.2-bin.tar.gz
[oracle@centos ~]$ ls -l apache-hive-1.2.2-bin.tar.gz
-rwxrwxrwx. 1 oracle oracle 90859180 4월 7 2017 apache-hive-1.2.2-bin.tar.gz
```

3. hive 설치 파일의 압축을 푼다.

tar xzvf apache-hive-1.2.2-bin.tar.gz

ls -ld apach*

```
[oracle@centos ~]$ ls -ld apach*
drwxrwxr-x. 8 oracle oracle 159 1월 24 13:50 apache-hive-1.2.2-bin
-rwxrwxrwx. 1 oracle oracle 90859180 4월 7 2017 apache-hive-1.2.2-bin.tar.gz
```

4. 압축을 푼 apache-hive-1.2.2-bin 디렉토리로 이동해서 ls 해본다.

cd apache-hive-1.2.2-bin

ls

5. bin 폴더로 이동한다.

cd bin

ls

⇒ hive 가 보이면 잘 되고 있다는 거다.

```
[oracle@centos apache-hive-1.2.2-bin]$ cd bin
[oracle@centos bin]$ ls
beeline ext hive hive-config.sh hiveserver2 metatool schematool
```

6. jps 를 수행해서 하둡이 정상인지 확인하기

jps

```
[oracle@centos bin]$ jps
3183 JobTracker
3081 SecondaryNameNode
3317 TaskTracker
2818 NameNode
6744 Jps
5492 FsShell
2938 DataNode
```

7. hive 를 실행한다.

./hive

=> 위와 같이 접속하려면 매번 하이브 홈 디렉토리로 와서 bin 디렉토리로 간 다음에 hive를 ./hive 로

실행해야 하므로 너무 불편하다. 그래서 이렇게 접속 안하려고 다음과 같이 .bash_profile에 HIVE_HOME 디렉토리가 어디라고 지정을 한다. 그러면 위와 같이 안하고 그냥 아무데서나 hive라고 하면 접속된다.

```
cd
vi .bash_profile
-----
export HIVE_HOME=/home/oracle/apache-hive-1.2.2-bin
export PATH=$HIVE_HOME/bin:$PATH
-----
source .bash_profile
```

```
[oracle@centos bin]$ cd
[oracle@centos ~]$ vi .bash_profile
[oracle@centos ~]$ source .bash_profile
[oracle@centos ~]$ hive

Logging initialized using configuration in jar:file:/home/oracle/apache-hive-1.2.2-bin/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> █
```

에러 발생 시

이슈 1. Hive or Hiveserver2 실행 후 아래의 에러 발생 시,
Exception in thread "main" java.lang.RuntimeException: java.io.IOException: Filesystem closed

아래의 코드를 oracle 계정에서 실행하여 조치한다.

```
hadoop dfs -mkdir /tmp/hive
hadoop dfs -chmod 777 /tmp
hadoop dfs -chmod 777 /tmp/hive
```

```
./hive
```

```
[oracle@centos bin]$ cd
[oracle@centos bin]$ hadoop dfs -mkdir /tmp/hive
mkdir: cannot create directory /tmp/hive: File exists
[oracle@centos bin]$ hadoop dfs -chmod 777 /tmp
[oracle@centos bin]$ hadoop dfs -chmod 777 /tmp/hive
[oracle@centos bin]$ ./hive

Logging initialized using configuration in jar:file:/home/oracle/apache-hive-1.2.2-bin/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> █
```

이슈 2. hiveserver2 명령어를 실행 한 후 hive 명령어가 정상적으로 작동하지 않는 에러

root 계정으로 jps 실행 후 RunJar 라고 되어있는 프로세스 kill 한다.

예시)

```
kill -9 (RunJar_Process_Number)
```

이슈 3. hive 실행 후 hiveserver2 명령어가 정상적으로 작동하지 않는 에러. hiveserver2 실행 이후 아래 명령어를 실행 했을 때 아무것도 안나오면 정상적으로 작동하지 않는 상태다.

```
netstat -ant |grep 10000
```

oracle 계정에서 해당 로그를 찾아서 확인한다.

```
find / -name hive.log
tail -f hive.log
```

만약 아래 내용의 에러가 발생했다면 ps -ef |grep hive 프로세스 전부 kill 하고 다시 실행
ERROR XSDB6: Another instance of Derby may have already booted the database /home/oracle/metastore/derby
아래 명령어 실행시 10000번 포트가 바인딩되면 정상작동 완료
hiverserver2 &