

5. Hive 설치 및 환경설정

목 차

1. Hive 소개
2. Hive 설치
3. Hive 환경설정
4. Hadoop DFS 작업
5. Hive 실행

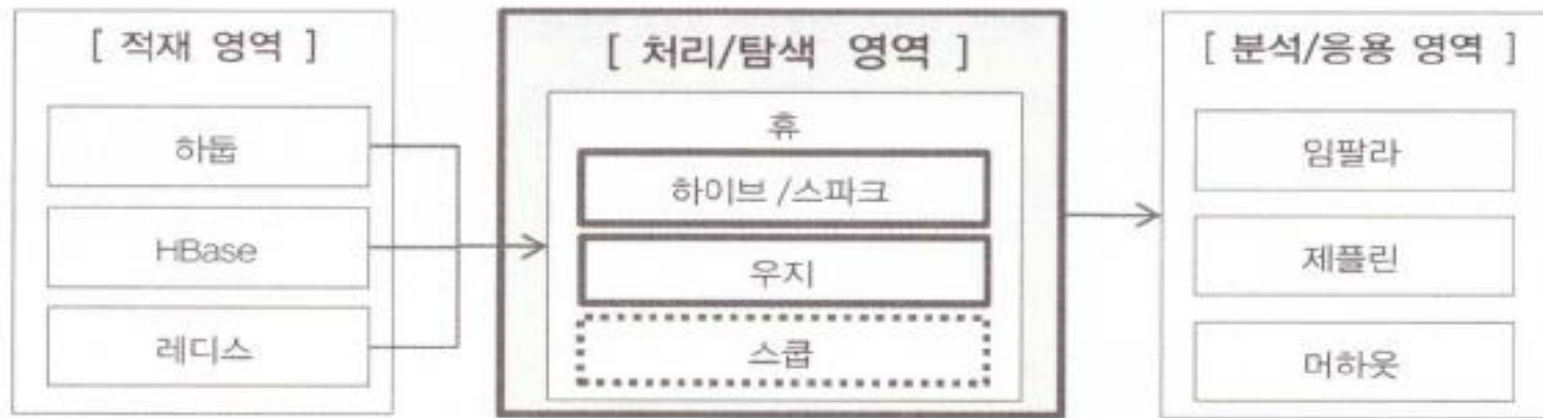
1. Hive 소개

- 하둡 초기에 적재된 데이터 탐색/분석 도구로 MapReduce 이용
- MapReduce는 복잡도가 높은 프로그래밍 기법이 요구되고, 이는 업무 분석가 및 관리자들에게 빅데이터 접근을 어렵게 만듦
- 이를 해결하기 위해 페이스북에서 SQL과 매우 유사한 방식으로 하둡 데이터에 접근성을 높인 Hive 개발
- 오픈 소스로 공개되면서 2016년 2월 하이버 2.0이 릴리스
- 빅데이터의 가장 대표적인 SQL on Hadoop 제품으로 자리 잡음

➤ Hive 주요 특징

- HiveQL이라고 하는 SQL과 유사한 쿼리를 사용
- HDFS 적재 데이터 이용 DW(data warehouse) 구축
- HiveQL로 정의한 내용을 Hive가 MapReduce 변환/ 실행
- Hive 구성요소에서 Metastore 존재
 - 이 부분에 DB 스키마 저장
 - 실제 데이터는 하둡의 HDFS에 분산되어 저장

처리/탐색 Layer



➤ 하이브/스파크

- 하둡에 적재된 데이터를 정제/변형/통합/분리/탐색 등의 작업 수행
- 데이터를 정형화된 구조로 정규화해 데이터 마트 생성

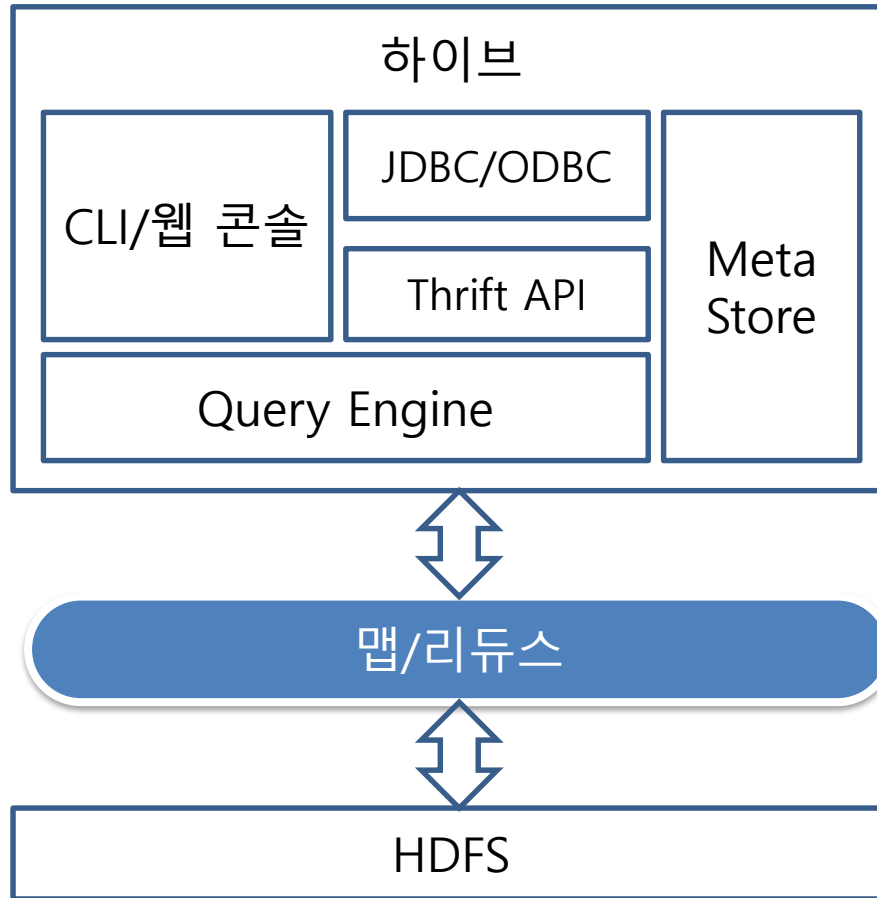
➤ 스쿱

- 가공/분석된 데이터를 외부로 제공하며, 필요 시 분석/응용 단계에서도 사용

➤ 우지

- 데이터의 품질을 높이는 단계, 과정이 길고 복잡해지기 마련 .
- 프로세스를 구성해 복잡도를 낮추고, 자동화.

➤ Hive 아키텍처



CLI : 사용자가 Hive 쿼리를 입력 및 실행 인터페이스

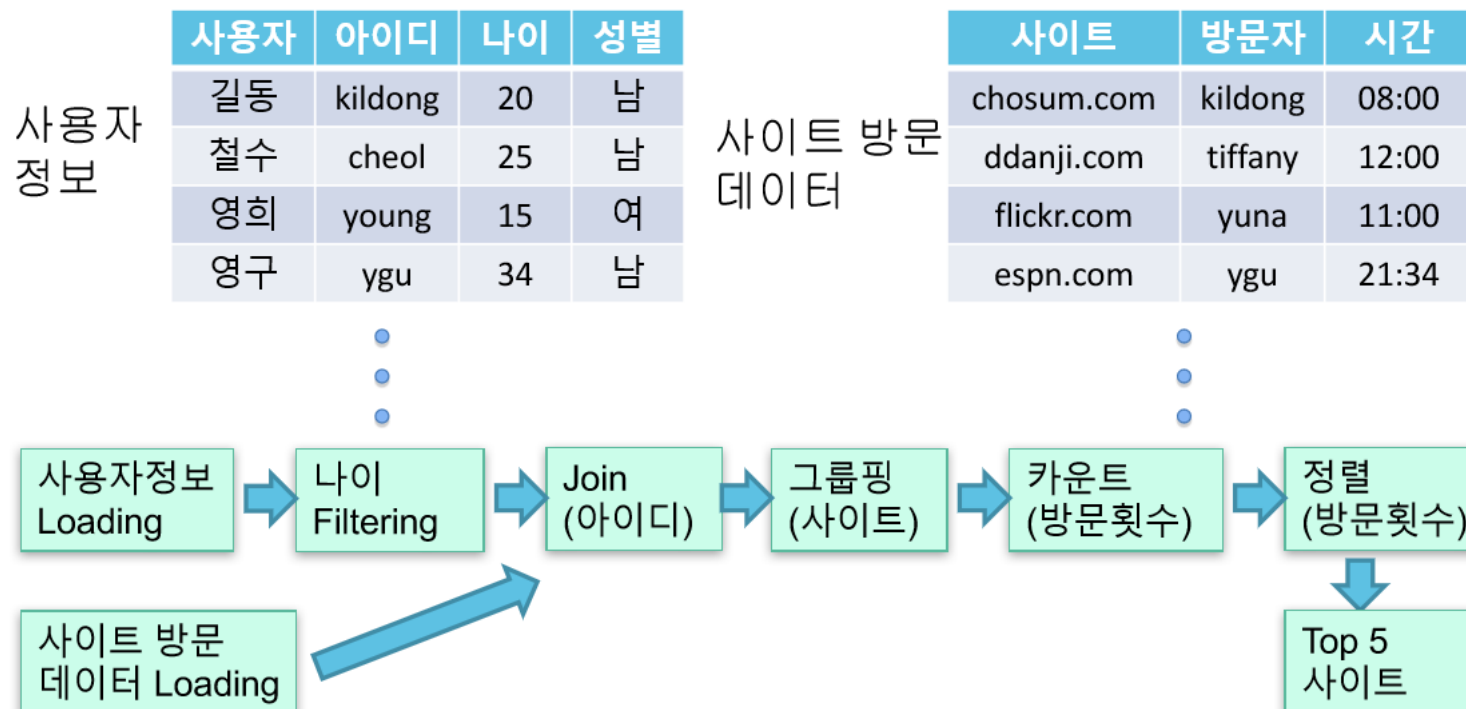
JDBC/ODBC Driver
- Hive 쿼리를 다양한 DB와 연결하기 위한 드라이버

Query Engine : Hive QL을 맵리듀스 코드로 변환/실행

MetaStore : Hive에서 사용하는 테이블의 스키마 정보를 저장 및 관리 기본적으로 Derby DB 사용(DBMS 변경 가능)

MapReduce vs Hive

- 18 ~ 25세 연령대의 사용자가 가장 많이 방문하는 사이트 5개를 찾아라



사용자 보다는 개발자 중심, 중복된 코딩 및 노력의 반복, 실행을 위한 버전 관리, 환경 설정의 복잡성

- 직접 MapReduce 프로그램을 코딩 할 경우

[illegible]

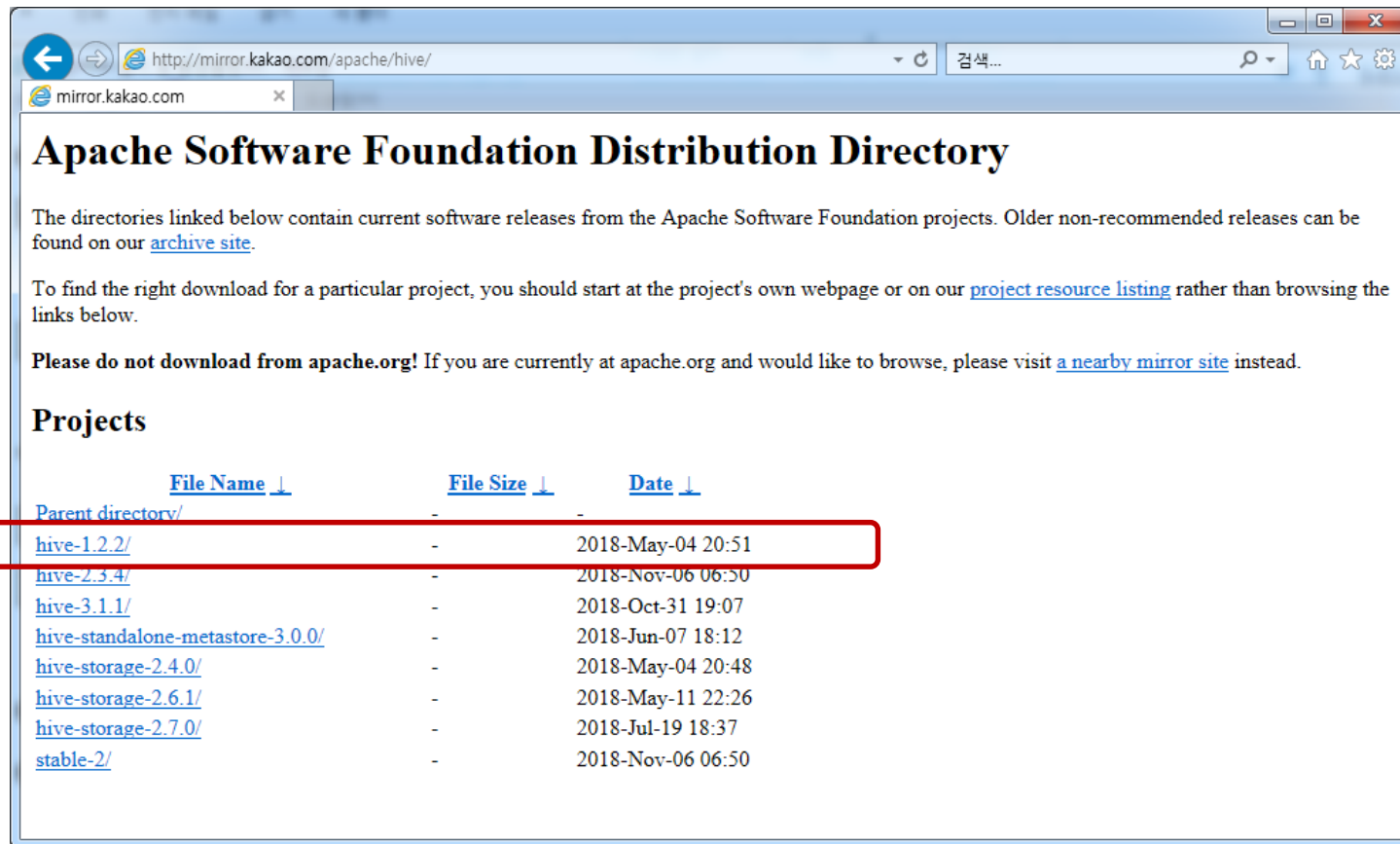
Hive SQL

사용자 중심 편의성 제공, 기본 SQL문 이용 빅데이터 검색, 시간과 노력 절감, 간단한 쿼리문

```
select  t2.url,      count(1) as      visits      from      userinfo t1
        join      visitinfo t2      on      (t1.id=t2.id)      where
        t1.age    >      17      and      t1.age    <      26
        group    by      t2.url      sort      by      visits      DESC
        limit    5;
```


2. Hive 설치

<http://ftp.daumkakao.com/apache/hive/> 에서 다운로드 가능한 hive 버전 확인



Apache Software Foundation Distribution Directory

The directories linked below contain current software releases from the Apache Software Foundation projects. Older non-recommended releases can be found on our [archive site](#).

To find the right download for a particular project, you should start at the project's own webpage or on our [project resource listing](#) rather than browsing the links below.

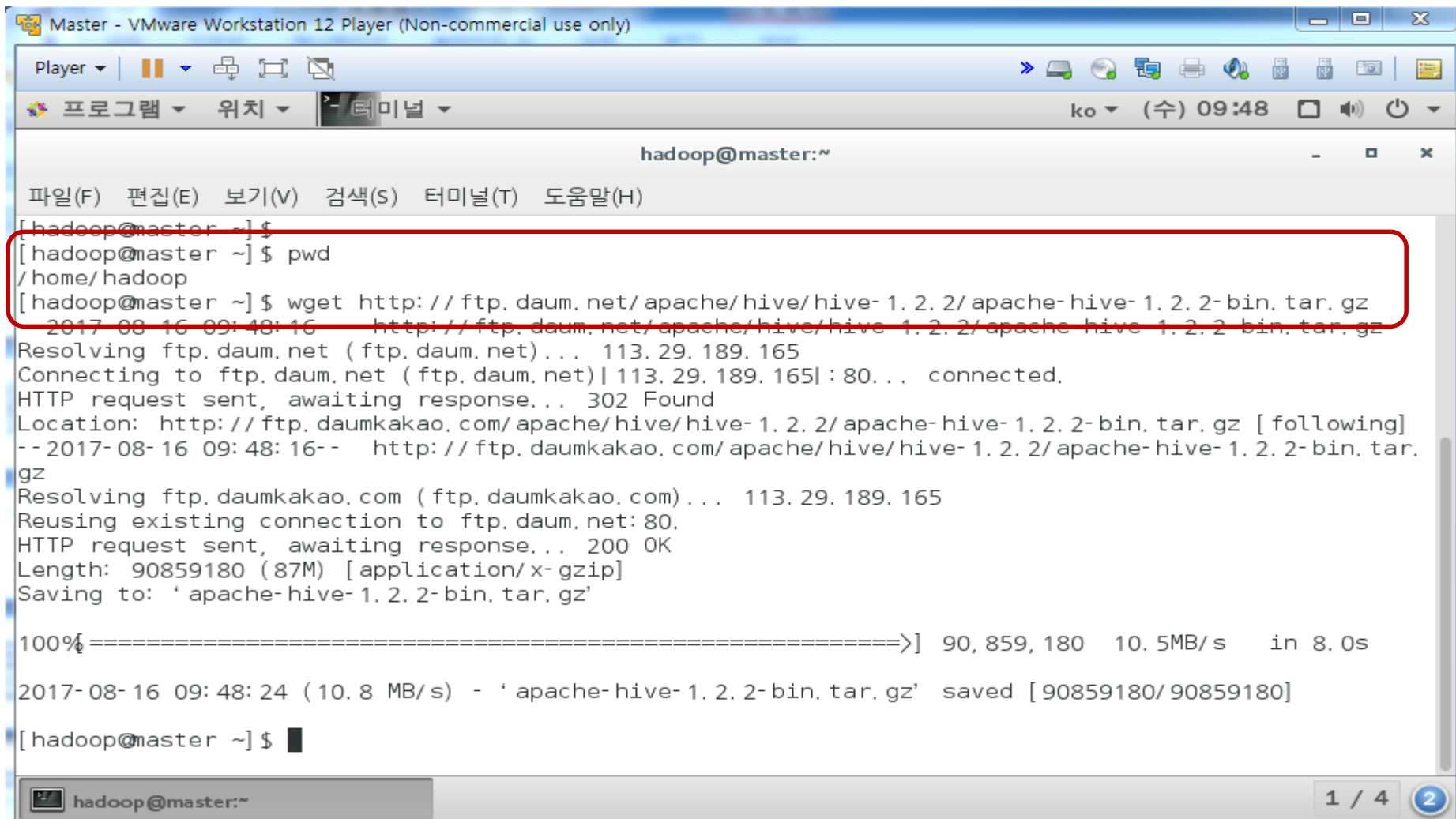
Please do not download from apache.org! If you are currently at apache.org and would like to browse, please visit [a nearby mirror site](#) instead.

Projects

File Name ↓	File Size ↓	Date ↓
Parent directory/	-	-
hive-1.2.2/	-	2018-May-04 20:51
hive-2.3.4/	-	2018-Nov-06 06:50
hive-3.1.1/	-	2018-Oct-31 19:07
hive-standalone-metastore-3.0.0/	-	2018-Jun-07 18:12
hive-storage-2.4.0/	-	2018-May-04 20:48
hive-storage-2.6.1/	-	2018-May-11 22:26
hive-storage-2.7.0/	-	2018-Jul-19 18:37
stable-2/	-	2018-Nov-06 06:50

2. Hive 설치

<http://ftp.daumkakao.com/apache/hive/> 에서 다운로드 가능한 hive 버전 확인



The screenshot shows a terminal window titled "hadoop@master:~" within a "Master - VMware Workstation 12 Player" environment. The terminal displays the following commands and output:

```
[hadoop@master ~]$  
[hadoop@master ~]$ pwd  
/home/hadoop  
[hadoop@master ~]$ wget http://ftp.daum.net/apache/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz  
2017-08-16 09:48:16 http://ftp.daum.net/apache/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz  
Resolving ftp.daum.net (ftp.daum.net)... 113.29.189.165  
Connecting to ftp.daum.net (ftp.daum.net)[113.29.189.165]:80... connected.  
HTTP request sent, awaiting response... 302 Found  
Location: http://ftp.daumkakao.com/apache/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz [following]  
--2017-08-16 09:48:16-- http://ftp.daumkakao.com/apache/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz  
Resolving ftp.daumkakao.com (ftp.daumkakao.com)... 113.29.189.165  
Reusing existing connection to ftp.daum.net:80.  
HTTP request sent, awaiting response... 200 OK  
Length: 90859180 (87M) [application/x-gzip]  
Saving to: 'apache-hive-1.2.2-bin.tar.gz'  
  
100%=====>] 90,859,180 10.5MB/s in 8.0s  
2017-08-16 09:48:24 (10.8 MB/s) - 'apache-hive-1.2.2-bin.tar.gz' saved [90859180/90859180]  
[hadoop@master ~]$
```

A red rectangle highlights the initial commands: `pwd` and the `wget` command. The terminal interface includes a menu bar with options like "파일(F)", "편집(E)", "보기(V)", "검색(S)", "터미널(T)", and "도움말(H)". The status bar at the bottom shows "hadoop@master:~" and a page indicator "1 / 4".

hadoop@master:~

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

Hive 압축 풀기

```
[hadoop@master ~]$  
[hadoop@master ~]$  
[hadoop@master ~]$ tar xvfz apache-hive-1.2.2-bin.tar.gz  
apache-hive-1.2.2-bin/LICENSE  
apache-hive-1.2.2-bin/NOTICE  
apache-hive-1.2.2-bin/README.txt  
apache-hive-1.2.2-bin/RELEASE_NOTES.txt  
apache-hive-1.2.2-bin/examples/files/2000_cols_data.csv  
apache-hive-1.2.2-bin/examples/files/agg_01-p1.txt  
apache-hive-1.2.2-bin/examples/files/agg_01-p2.txt  
apache-hive-1.2.2-bin/examples/files/agg_01-p3.txt  
apache-hive-1.2.2-bin/examples/files/alltypes.txt  
apache-hive-1.2.2-bin/examples/files/alltypes2.txt  
apache-hive-1.2.2-bin/examples/files/apache.access.2.log  
apache-hive-1.2.2-bin/examples/files/apache.access.log  
apache-hive-1.2.2-bin/examples/files/archive_corrupt.rc  
apache-hive-1.2.2-bin/examples/files/array_table.txt  
apache-hive-1.2.2-bin/examples/files/avro_charvarchar.txt  
apache-hive-1.2.2-bin/examples/files/avro_date.txt  
apache-hive-1.2.2-bin/examples/files/avro_timestamp.txt  
apache-hive-1.2.2-bin/examples/files/AvroPrimitiveInList.parquet  
apache-hive-1.2.2-bin/examples/files/AvroSingleFieldGroupInList.parquet
```

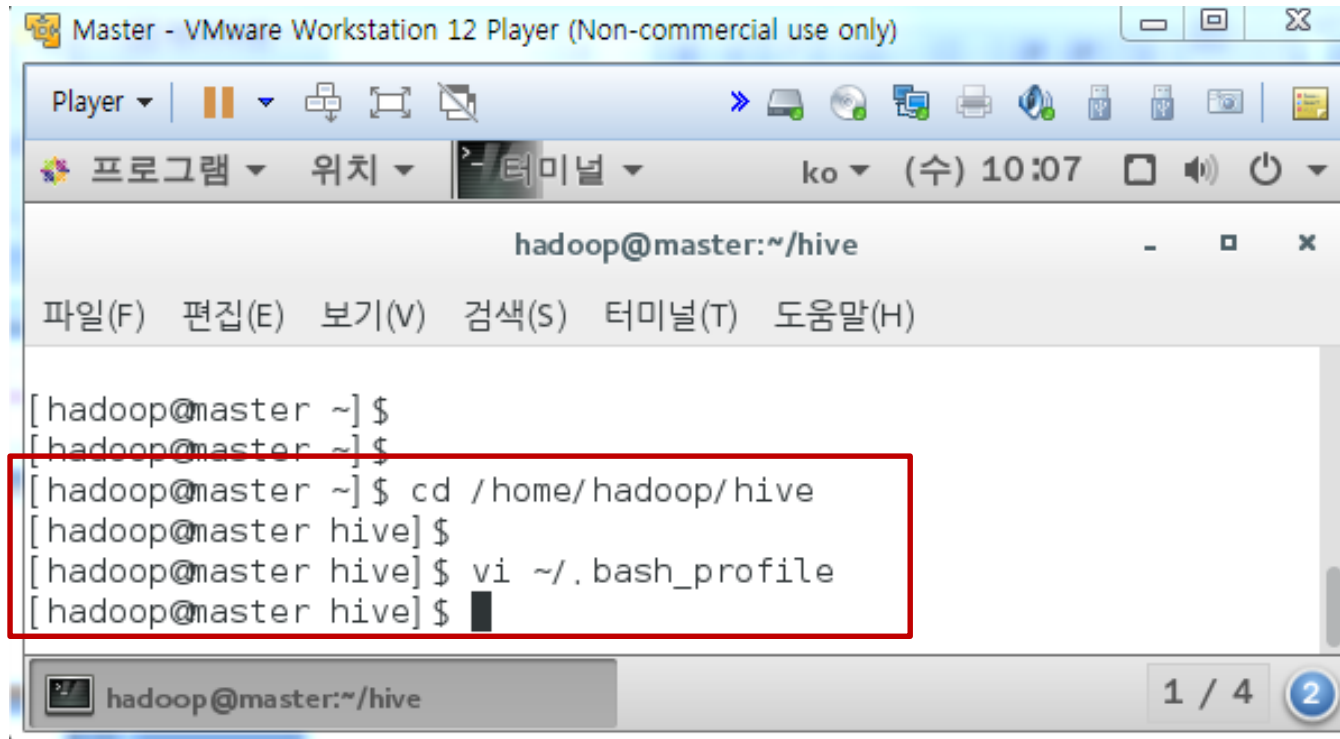
hadoop@master:~

1 / 4

2

3. Hive 환경설정

1) 디렉터리 이동

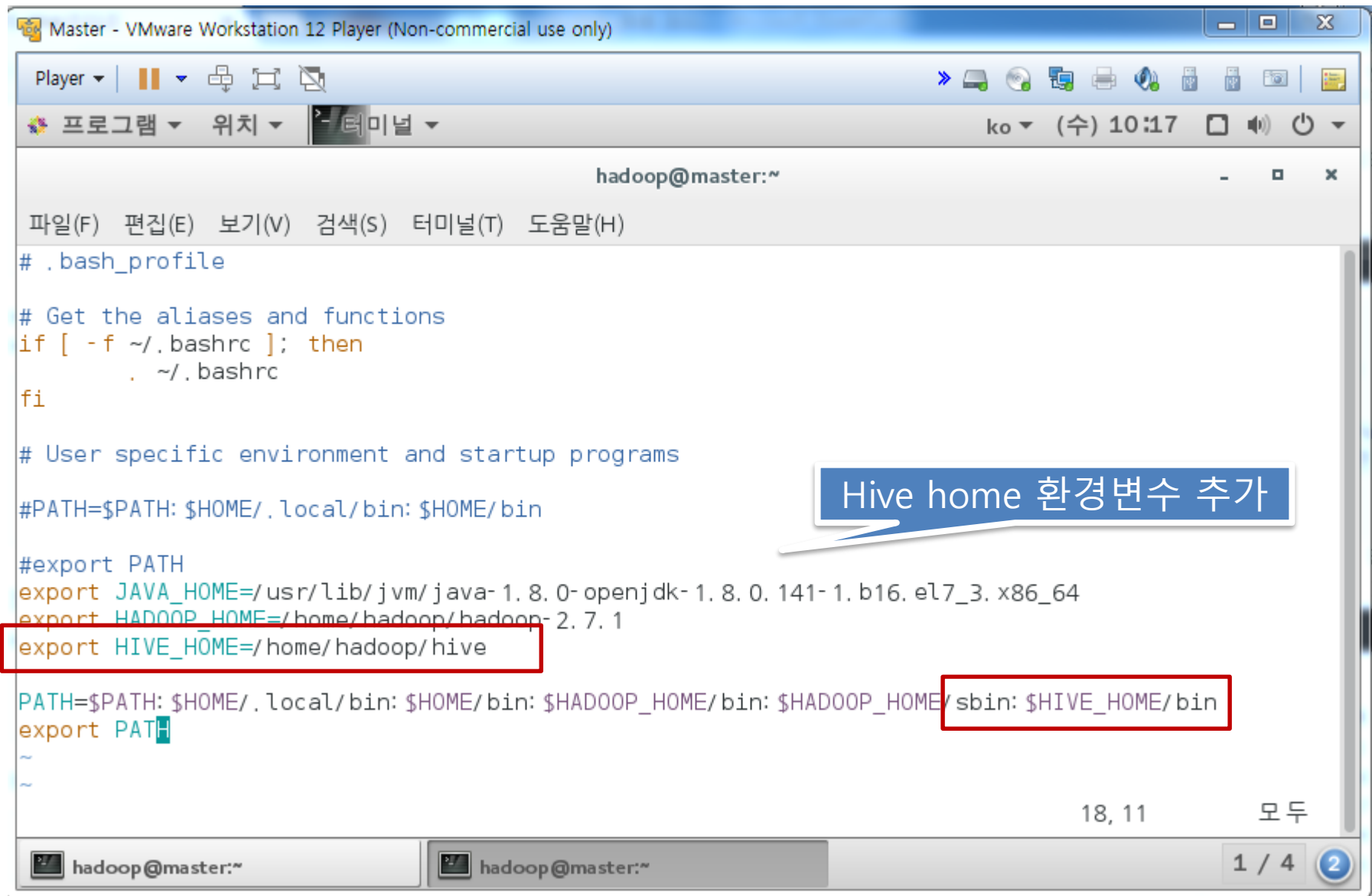


The screenshot shows a VMware Workstation 12 Player window titled "Master - VMware Workstation 12 Player (Non-commercial use only)". Inside, a terminal window titled "hadoop@master:~/hive" is open. The terminal displays a menu bar with "파일(F)", "편집(E)", "보기(V)", "검색(S)", "터미널(T)", and "도움말(H)". The command history shows the user navigating from the home directory to the Hive directory and then editing the bash profile. The last four lines of the terminal output are highlighted with a red rectangle.

```
[hadoop@master ~]$  
[hadoop@master ~]$  
[hadoop@master ~]$ cd /home/hadoop/hive  
[hadoop@master hive]$  
[hadoop@master hive]$ vi ~/.bash_profile  
[hadoop@master hive]$
```

At the bottom of the terminal window, there is a status bar showing "hadoop@master:~/hive" on the left, "1 / 4" in the middle, and a blue circle with the number "2" on the right.

2) .bash_profile 설정



```
hadoop@master:~  
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)  
# .bash_profile  
  
# Get the aliases and functions  
if [ -f ~/.bashrc ]; then  
    . ~/.bashrc  
fi  
  
# User specific environment and startup programs  
  
#PATH=$PATH:$HOME/.local/bin:$HOME/bin  
  
#export PATH  
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.141-1.b16.el7_3.x86_64  
export HADOOP_HOME=/home/hadoop/hadoop-2.7.1  
export HIVE_HOME=/home/hadoop/hive  
  
PATH=$PATH:$HOME/.local/bin:$HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$HIVE_HOME/bin  
export PATH  
~  
~
```

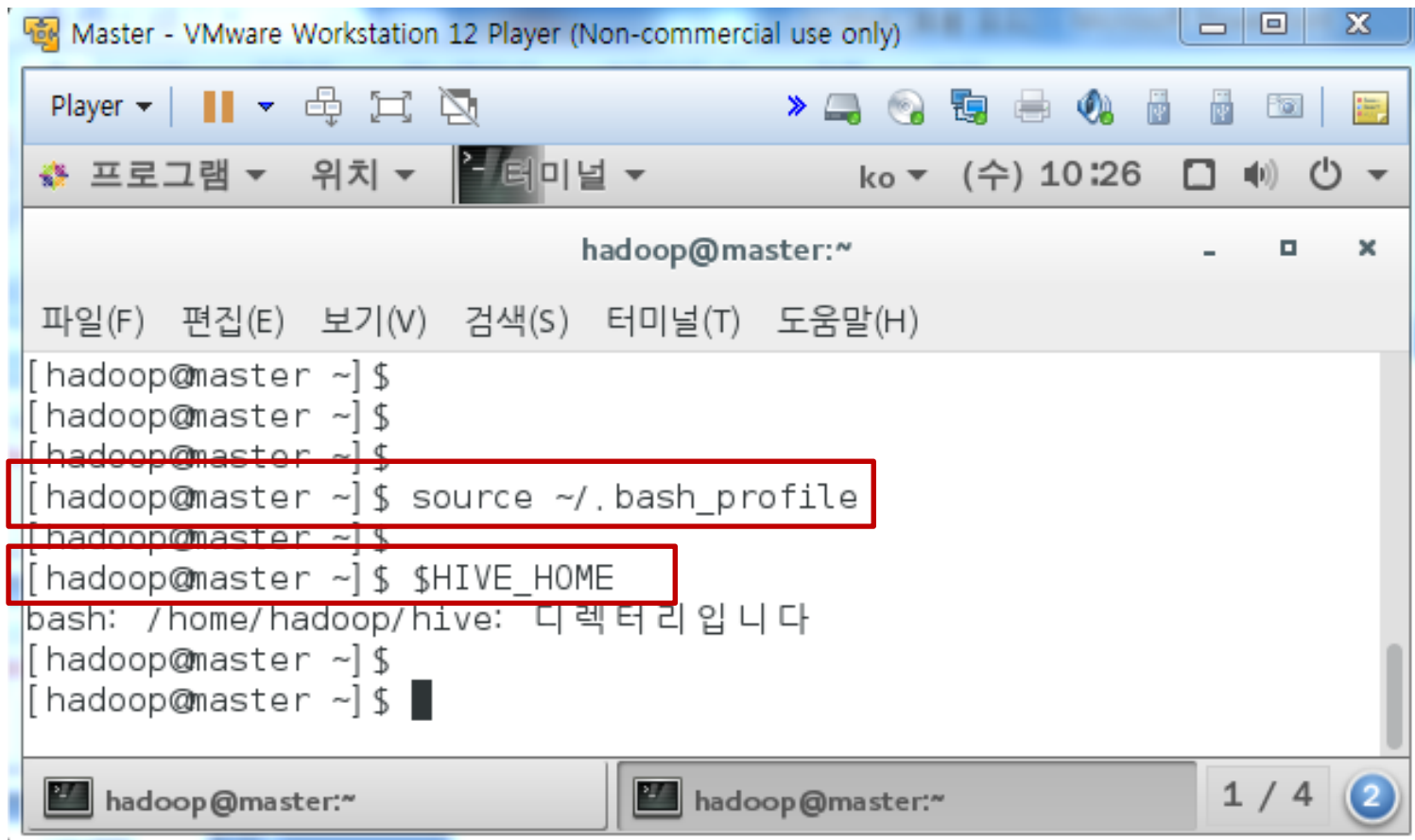
Hive home 환경변수 추가

18, 11 모두

hadoop@master:~

1 / 4 2

3) .bash_profile 적용/테스트

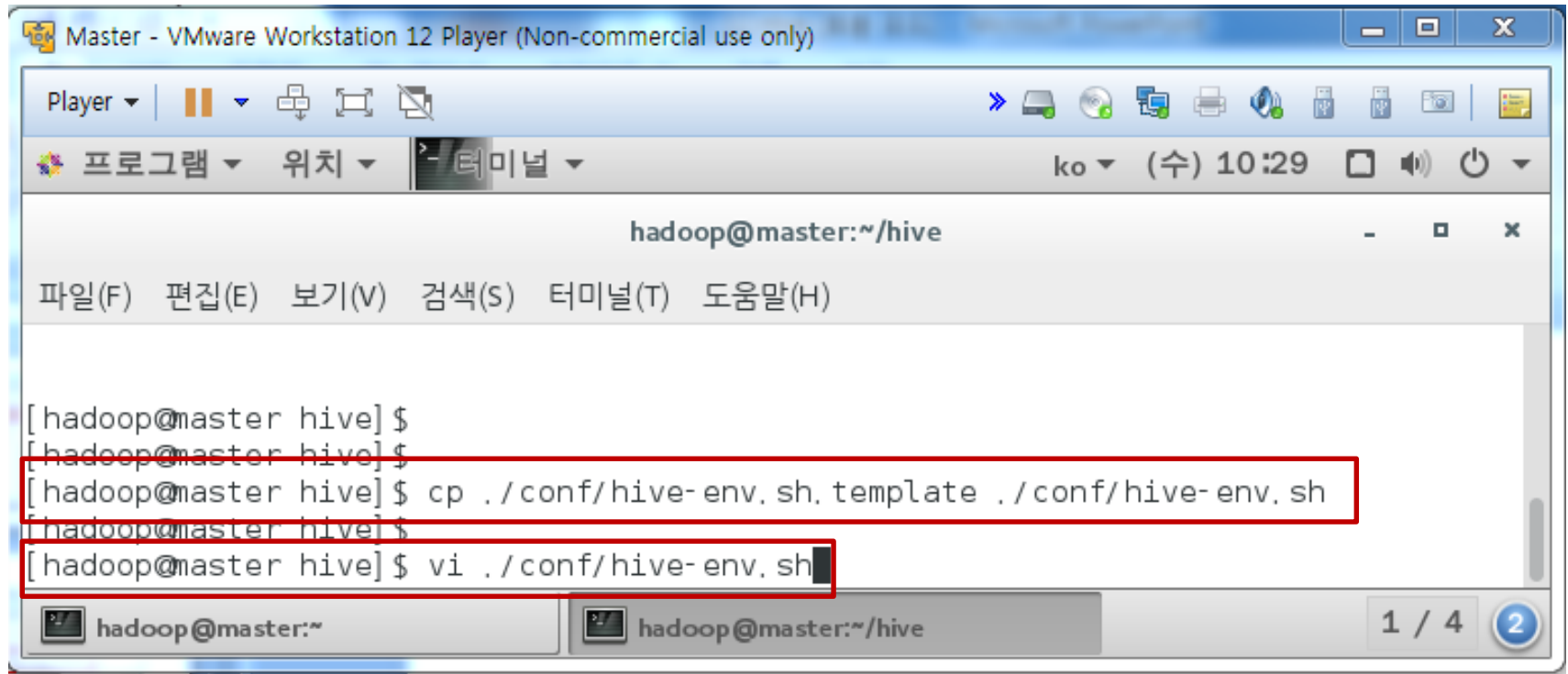


The screenshot shows a terminal window titled "hadoop@master:~" within a VMware Workstation 12 Player. The terminal displays the following commands and output:

```
hadoop@master:~$  
hadoop@master:~$  
hadoop@master:~$  
hadoop@master:~$ source ~/.bash_profile  
hadoop@master:~$  
hadoop@master:~$ $HIVE_HOME  
bash: /home/hadoop/hive: 디렉터리입니다  
hadoop@master:~$  
hadoop@master:~$
```

Two red boxes highlight the commands `source ~/.bash_profile` and `$HIVE_HOME`. The terminal also shows a menu bar with options like "파일(F)", "편집(E)", "보기(V)", "검색(S)", "터미널(T)", and "도움말(H)". The bottom status bar indicates the current page is 1 of 4.

4) hive-env.sh 환경 설정

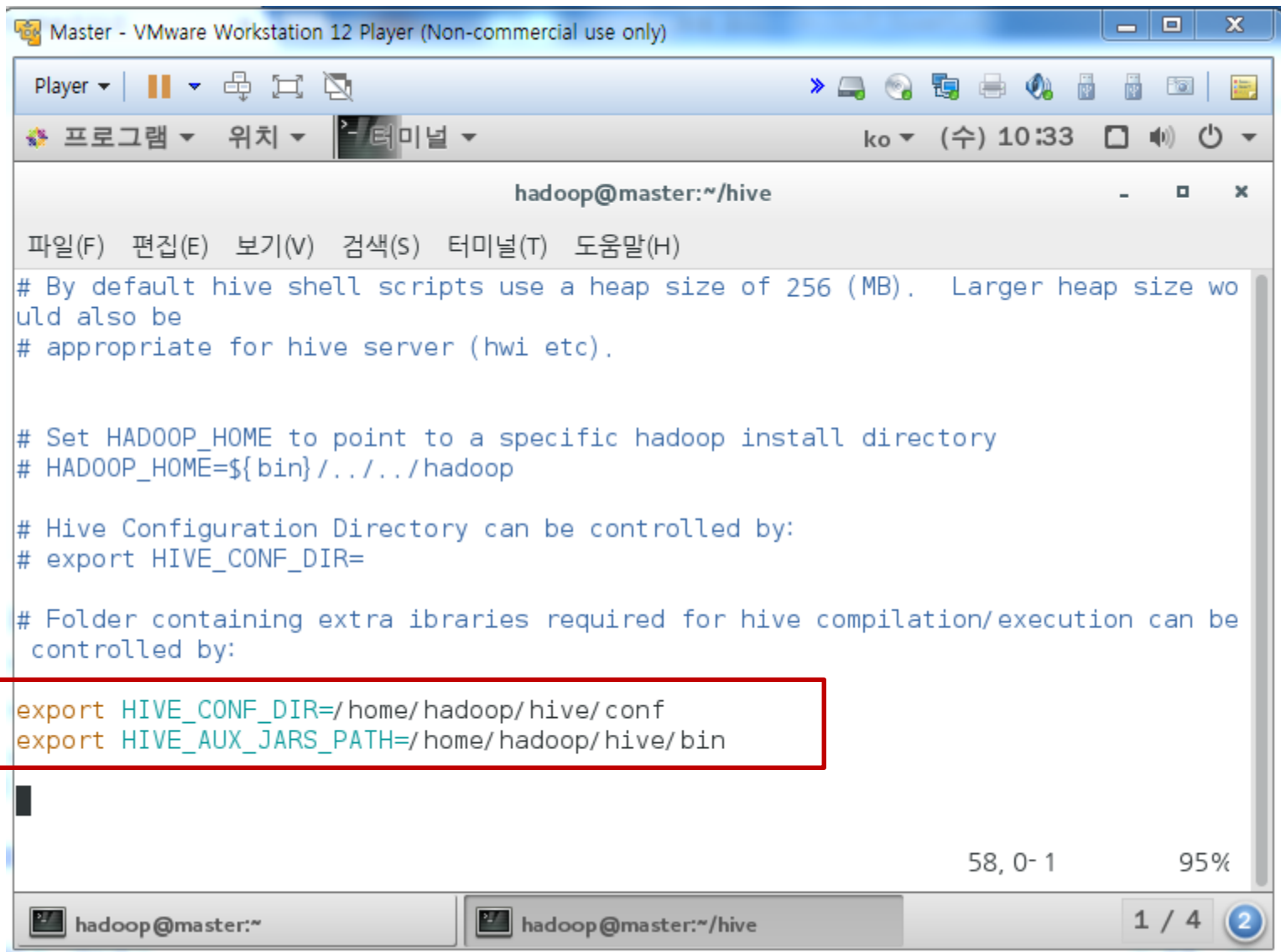


The screenshot shows a VMware Workstation 12 Player window titled "Master - VMware Workstation 12 Player (Non-commercial use only)". Inside the player, there is a terminal window titled "hadoop@master:~/hive". The terminal has a menu bar with "파일(F)", "편집(E)", "보기(V)", "검색(S)", "터미널(T)", and "도움말(H)". The terminal shows the following commands and prompts:

```
[hadoop@master hive]$  
[hadoop@master hive]$  
[hadoop@master hive]$ cp ./conf/hive-env.sh.template ./conf/hive-env.sh  
[hadoop@master hive]$  
[hadoop@master hive]$ vi ./conf/hive-env.sh
```

The last two lines of the terminal output are highlighted with red boxes. The bottom of the terminal window shows a tab bar with two tabs: "hadoop@master:~" and "hadoop@master:~/hive". The "hadoop@master:~/hive" tab is active, and the page number "1 / 4" and a blue circle with the number "2" are visible in the bottom right corner.

4) hive-env.sh 환경 설정



```
hadoop@master:~/hive
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
# By default hive shell scripts use a heap size of 256 (MB). Larger heap size would also be
# appropriate for hive server (hwi etc).

# Set HADOOP_HOME to point to a specific hadoop install directory
# HADOOP_HOME=${ bin} ../../hadoop

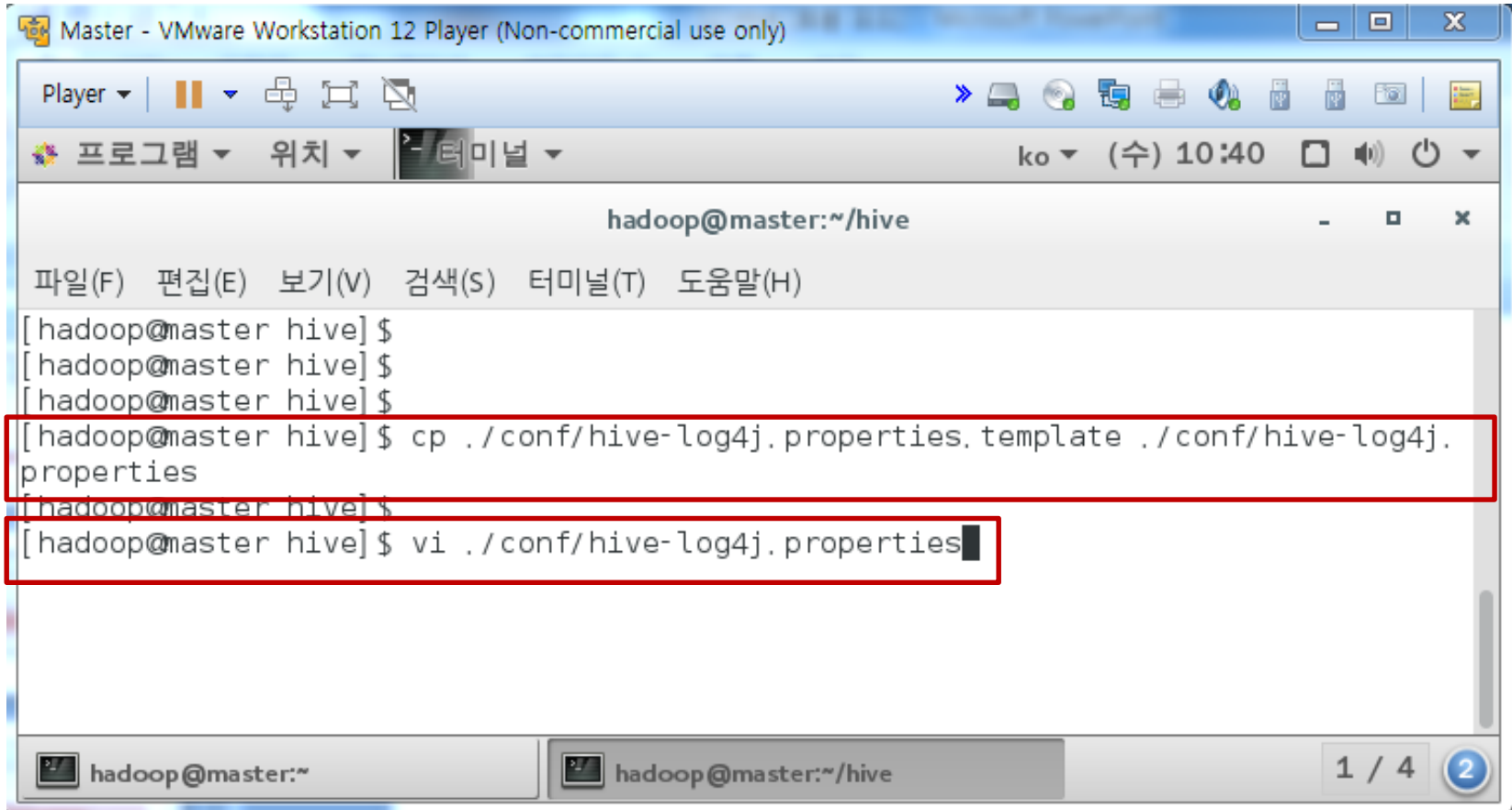
# Hive Configuration Directory can be controlled by:
# export HIVE_CONF_DIR=

# Folder containing extra ibraries required for hive compilation/execution can be
# controlled by:

export HIVE_CONF_DIR=/home/hadoop/hive/conf
export HIVE_AUX_JARS_PATH=/home/hadoop/hive/bin

58, 0-1 95%
hadoop@master:~ hadoop@master:~/hive 1 / 4 2
```

5) hive-log4j.properties 환경 설정



The screenshot shows a VMware Workstation 12 Player window titled "Master - VMware Workstation 12 Player (Non-commercial use only)". Inside the player, there is a terminal window titled "hadoop@master:~/hive". The terminal displays the following commands and prompts:

```
hadoop@master:~/hive
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[hadoop@master hive]$
[hadoop@master hive]$
[hadoop@master hive]$
[hadoop@master hive]$ cp ./conf/hive-log4j.properties.template ./conf/hive-log4j.
properties
[hadoop@master hive]$
[hadoop@master hive]$ vi ./conf/hive-log4j.properties
```

The last two commands are highlighted with red boxes. The terminal window has a menu bar with "파일(F)", "편집(E)", "보기(V)", "검색(S)", "터미널(T)", and "도움말(H)". The status bar at the bottom shows "hadoop@master:~" and "hadoop@master:~/hive" with a page indicator "1 / 4" and a blue circle with the number "2".

Master - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | [Icons] | ko ▾ (수) 11:08 [Icons]

프로그램 ▾ 위치 ▾ 터미널 ▾

hadoop@master:~/hive

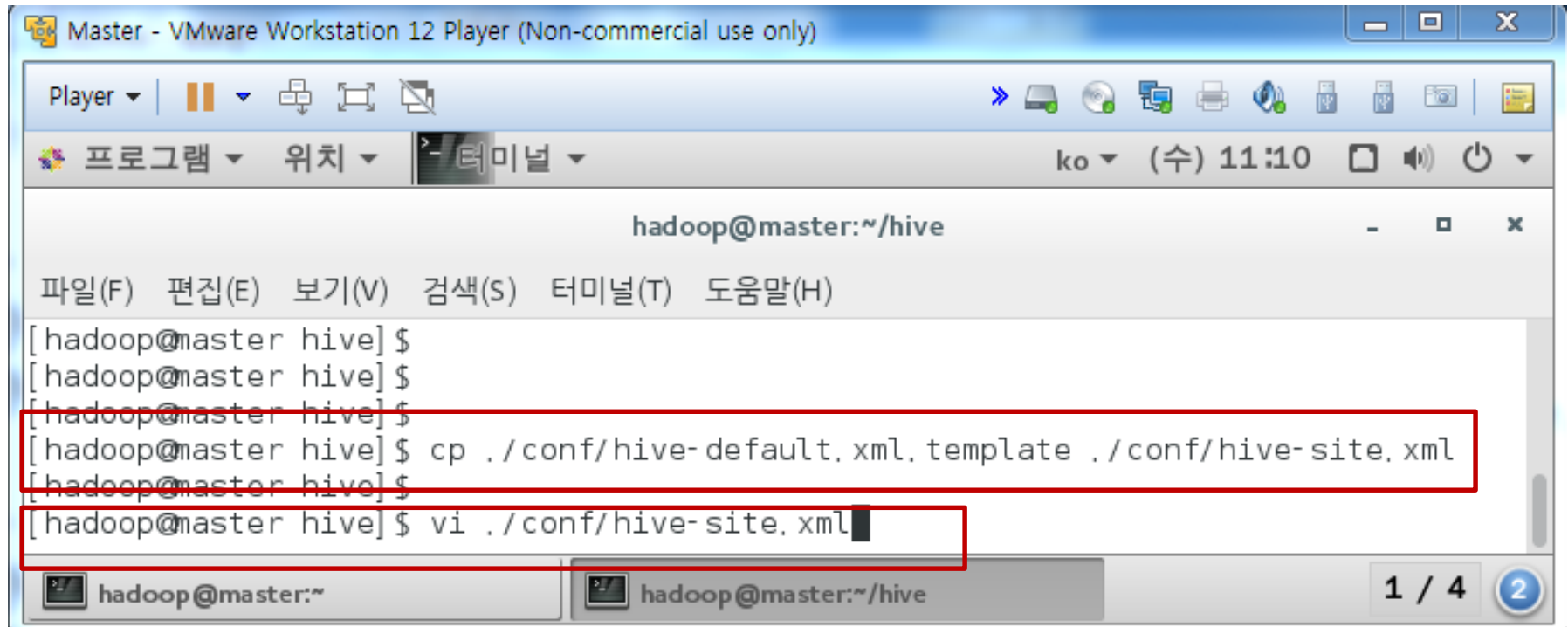
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements.  See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership.  The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License.  You may obtain a copy of the License at
#
#     http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# Define some default values that can be overridden by system properties
hive.log.threshold=ALL
hive.root.logger=INFO,DRFA
#hive.log.dir=${java.io.tmpdir}/${user.name}
hive.log.dir=/home/hadoop/hive/${user.name}

hive.log.file=hive.log
```

6) hive-site.xml 환경 설정



```
Master - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | [Icons] | ko ▾ (수) 11:10 [Icons]

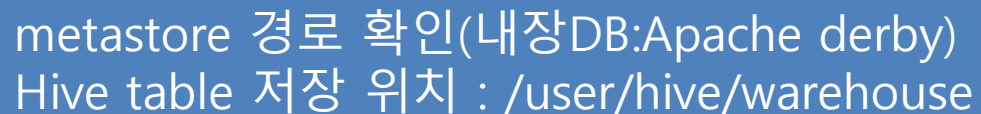
프로그램 ▾ 위치 ▾ 터미널 ▾

hadoop@master:~/hive

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

[hadoop@master hive]$
[hadoop@master hive]$
[hadoop@master hive]$
[hadoop@master hive]$ cp ./conf/hive-default.xml.template ./conf/hive-site.xml
[hadoop@master hive]$
[hadoop@master hive]$ vi ./conf/hive-site.xml

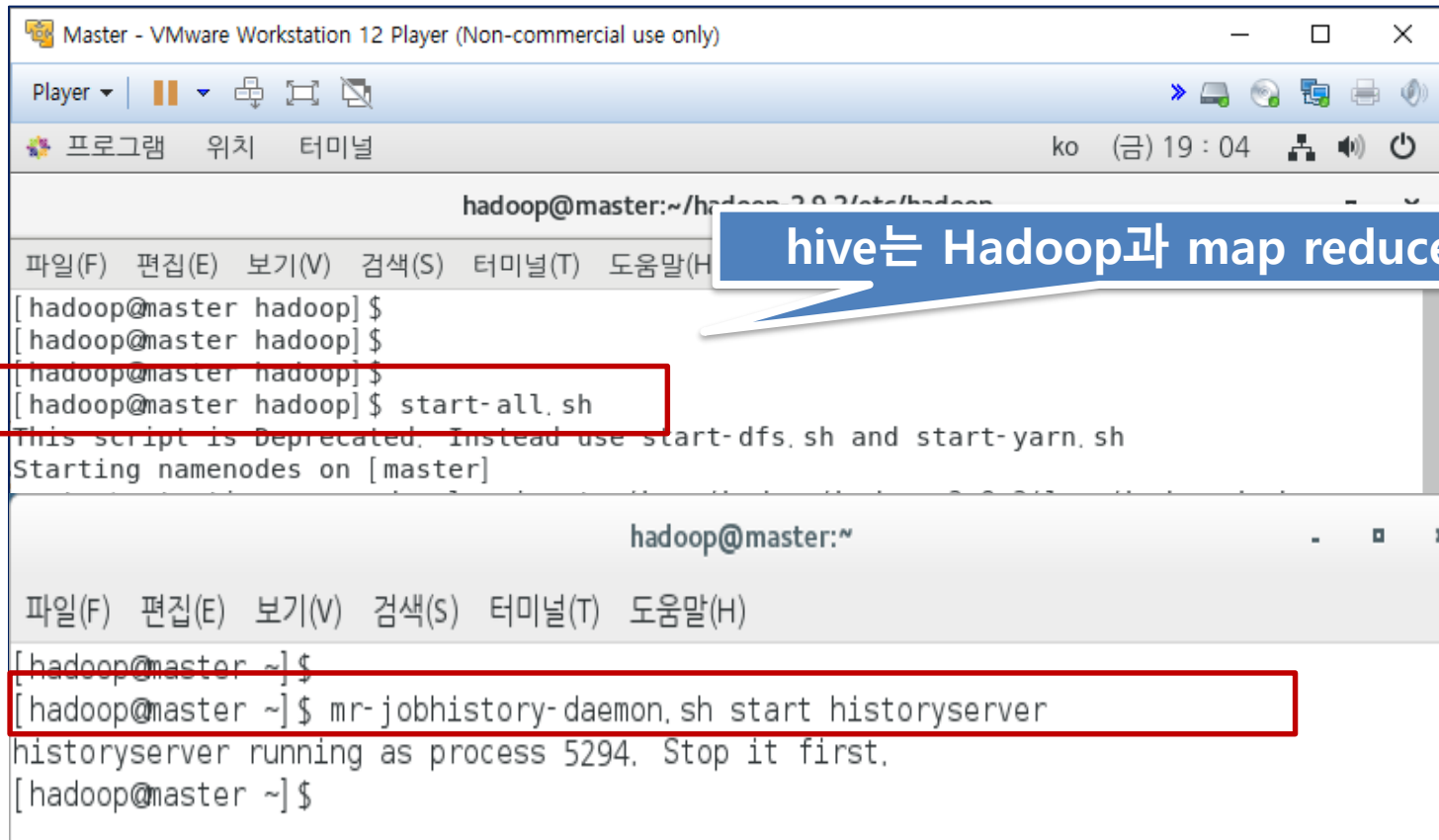
hadoop@master:~ hadoop@master:~/hive 1 / 4 2
```



4. Hadoop DFS 작업

준비 : Master/Slave1/Slave2 가상머신 ON

- 1) **start-all.sh** # 하둡/얀 실행
- 2) **mr-jobhistory-daemon.sh start historyserver** # 데몬실행



```
Master - VMware Workstation 12 Player (Non-commercial use only)
Player | [Icons]
프로그램 위치 터미널 ko (금) 19:04 [Icons]

hadoop@master:~/hadoop-2.8.2/etc/hadoop

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[hadoop@master hadoop]$
[hadoop@master hadoop]$
[hadoop@master hadoop]$
[hadoop@master hadoop]$ start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [master]

hadoop@master:~
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[hadoop@master ~]$
[hadoop@master ~]$ mr-jobhistory-daemon.sh start historyserver
historyserver running as process 5294. Stop it first.
[hadoop@master ~]$
```

3) 작업 디렉터리 생성/접근권한 설정

Master - VMware Workstation 12 Player (Non-commercial use only)

Player | [Icons] | ko (수) 11:37 [Icons]

hadoop@master:~/hive

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

[hadoop@master hive]\$ \$HADOOP_HOME

[hadoop@master hive]\$ \$HADOOP_HOME/bin/hdfs dfs -mkdir /tmp

mkdir: /tmp: File exists

[hadoop@master hive]\$ \$HADOOP_HOME/bin/hdfs dfs -mkdir -p /user/hive/warehouse

[hadoop@master hive]\$

[hadoop@master hive]\$ \$HADOOP_HOME/bin/hdfs dfs -chmod g+w /tmp

[hadoop@master hive]\$ \$HADOOP_HOME/bin/hdfs dfs -chmod g+w /user/hive/warehouse

[hadoop@master hive]\$

[hadoop@master hive]\$

hadoop@master:~ | hadoop@master:~/hive | 1 / 4 | 2

Hive에서 작업한 table 저장 디렉터리 생성(hive-site.xml 파일 참고)

```
[hadoop@master hive]$ ./bin/hive
```



[hadoop@master hive]\$ vi ./conf/hive-site.xml

Hive 재실행/테이블보기

Master - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | [Icons] | ko ▾ (수) 12:33 [Icons]

프로그램 ▾ 위치 ▾ 터미널 ▾

hadoop

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T)

```
[hadoop@master hive]$  
[hadoop@master hive]$ ./bin/hive  
Logging initialized using configuration in file:/home/hadoop/apache-hive-1.2.2  
bin/conf/hive-log4j.properties  
hive> show tables;  
OK  
Time taken: 1.852 seconds
```

Hive 실행, table 목록 보기

Hive 실행, table 목록 보기

Show tables; 실행 시 오류 해결

```
hadoop@master:~/hive/conf
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
[hadoop@master bin]$
[hadoop@master bin]$
[hadoop@master bin]$ hive
Logging initialized using configuration file:/usr/hadoop-2.2.2-bin/conf/hive-log4j.properties
hive> show tables;
FAILED: IllegalArgumentExcpetion java.net.URISyntaxException: Relative path in absolute URI
: file: /%20tmp/hive/java/hadoop/c535e3ed-73f6-489b-a089-1aaab580b078/hive_2018-07-10_15-13-28_007_3340507514741482861-1
hive> [hadoop@master bin]$ cd $HISTCMD
$HISTCMD $HISTCONTROL $HISTFILE
[hadoop@master bin]$ cd $HIVE_HOME
```

hive-site.xml 설정파일에서
경로 앞에 공백이 있는 경우 오류 발생

```
hadoop@master:~/hive/conf
파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)
```

```
</property>
<property>
  <name>system: java.io.tmpdir</name>
  <value>/tmp/hive/java</value>
</property>
<property>
  <name>system: user.name</name>
  <value>${user.name}</value>
</property>
</configuration>
"hive-site.xml" 3924L, 168913C
```

<value> 태그의 경로 앞뒤 공백 제거