7. Spark 설치 및 환경설정

목 차

1. Spark 소개
2. Spark 다운로드
3. Spark 설치
4. Spark 환경설정
5. Spark 실행
6. Spark SQL CLI 실행

빅데이터 탐색에 활용하는 기술 – Spark 등장배경

1. 기존 RDBMS를 대신할 빅데이터 저장 매체 Hadoop 등장

2. Hadoop에서도 SQL을 사용하고자 만든 것이 바로 Hive
   Hive를 통해 Hadoop에서도 SQL을 이용하여 DW 생성(편의성 제공)
   Hive는 Hadoop의 MapReduce 방법을 이용하여 연산 수행

   매 연산마다 다음과 같은 작업 반복
   1. Disk에서 Memory로 연산에 필요한 Data 로딩
   2. Memory에서 연산을 진행하고, 다시 Disk에 변경사항 저장
   위 과정으로 불필요한 I/O 연산 많아지고, 처리 속도 떨어짐

3. hive 한계를 극복하기 위한 대안으로 Spark 등장
   Spark는 한번에 연산을 수행할 Data를 모두 Memory에 불러온 후,
   Memory에서 연산을 수행하기 때문에 Hive보다 훨씬 빠른 연산 가능

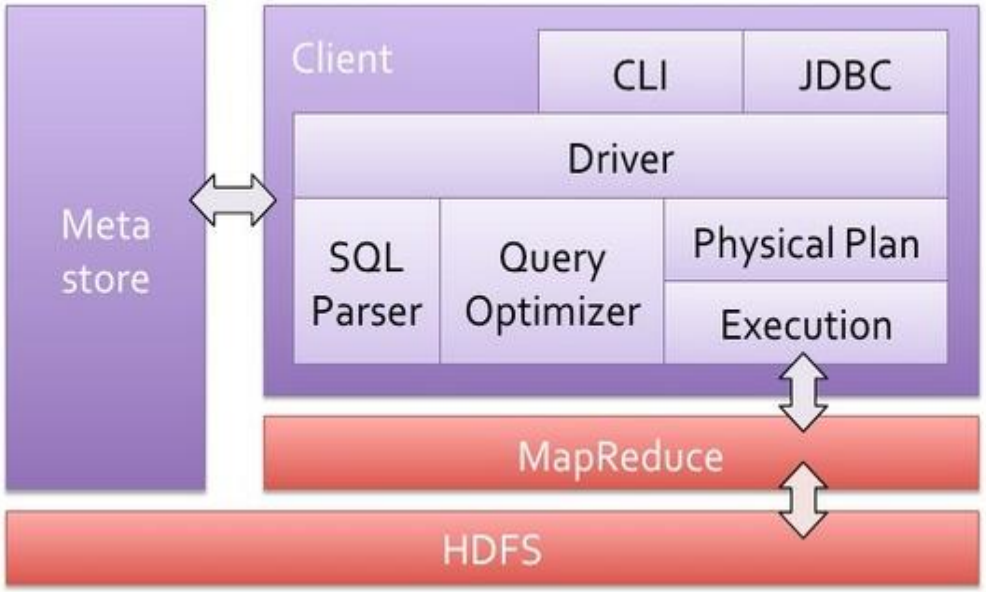
빅데이터 탐색에 활용하는 기술 – Spark

➢ Spark 소개

- 맵리듀스 코어를 그대로 사용하는 하이브는 성능면에서 만족스럽지 못함.
- 그로 인해 반복적인 대화형 연산 작업에서는 하이브가 적합하지 못함.
- 이 단점을 극복한 고성능 인메모리 분석.
- UC 버클리의 AMPLab에서 2009년 개발, 2010년 오픈 소스로 공개.
- 2013년 6월 아파치 재단으로 이관되어 최상위 프로젝트.
- 최근 빅데이터 분야에서 가장 핫한 기술 중 하나.
- 데이터 가공 처리를 인메모리에서 수행함으로써 대용량 데이터 작업에도 빠른 성능을 보장.

Hive vs Spark

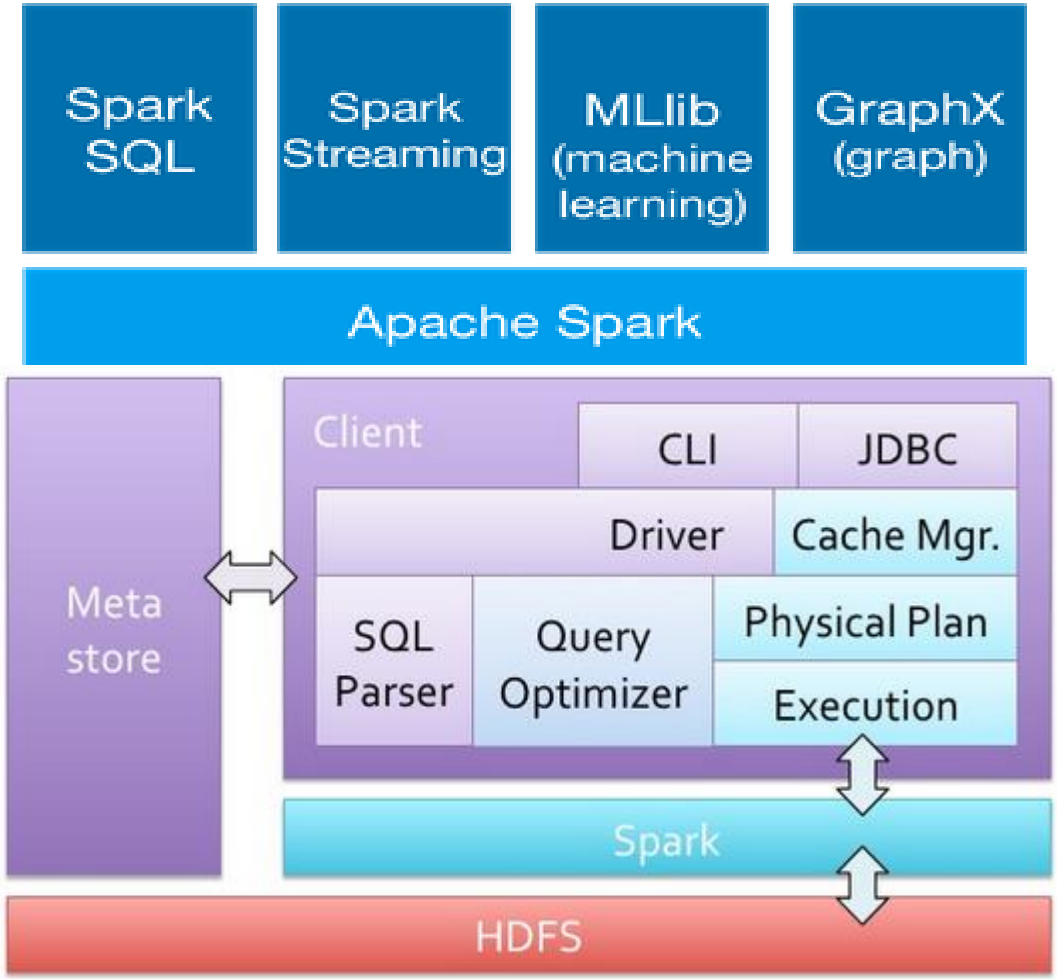
➢ Hive 아키텍처



- HiveQL은 MapReduce으로 변환하여 HDFS 데이터를 대상으로 DW를 생성하기 때문에 처리 속도가 느린 단점

빅데이터 탐색에 활용하는 기술 – Spark

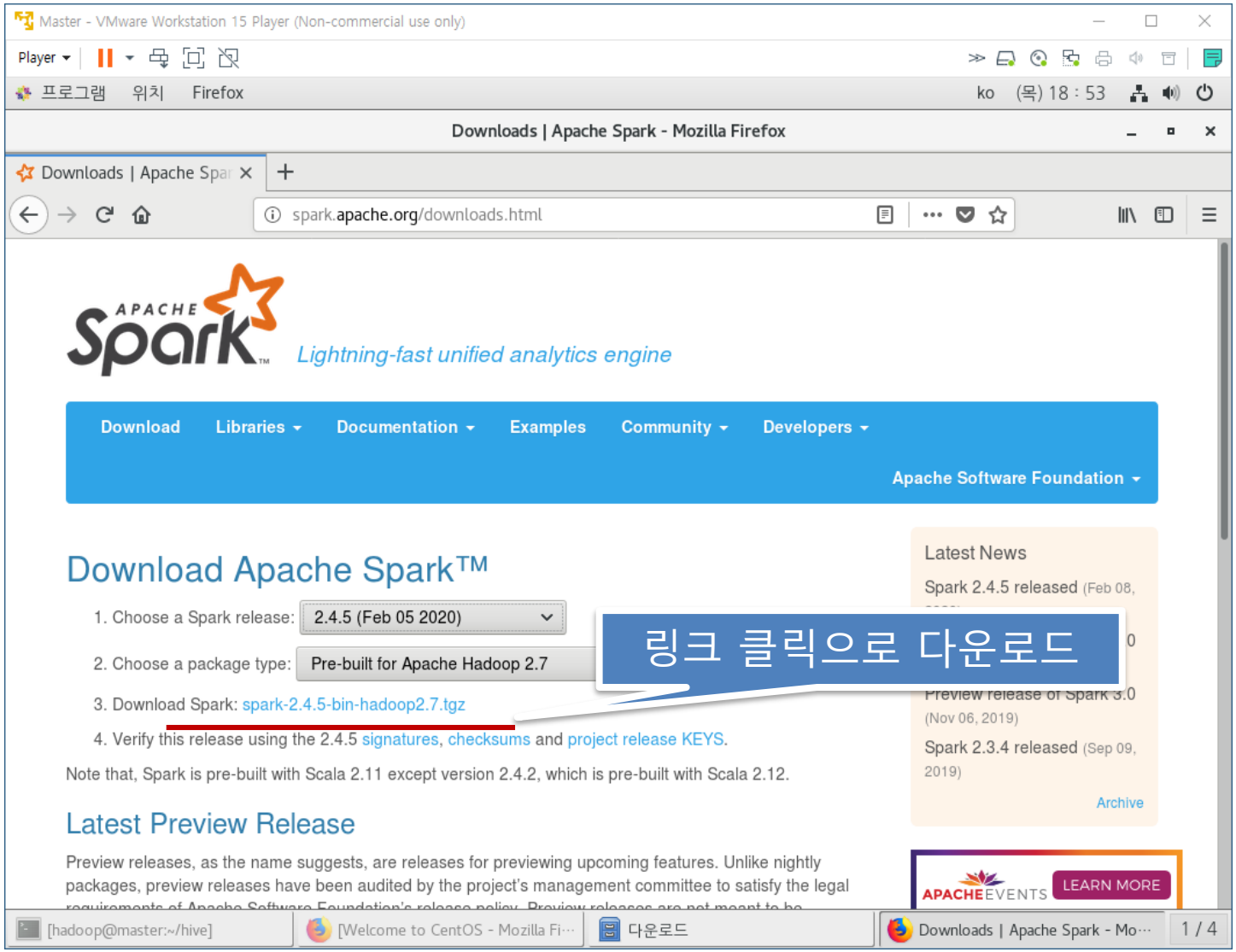
➢ Spark 아키텍처

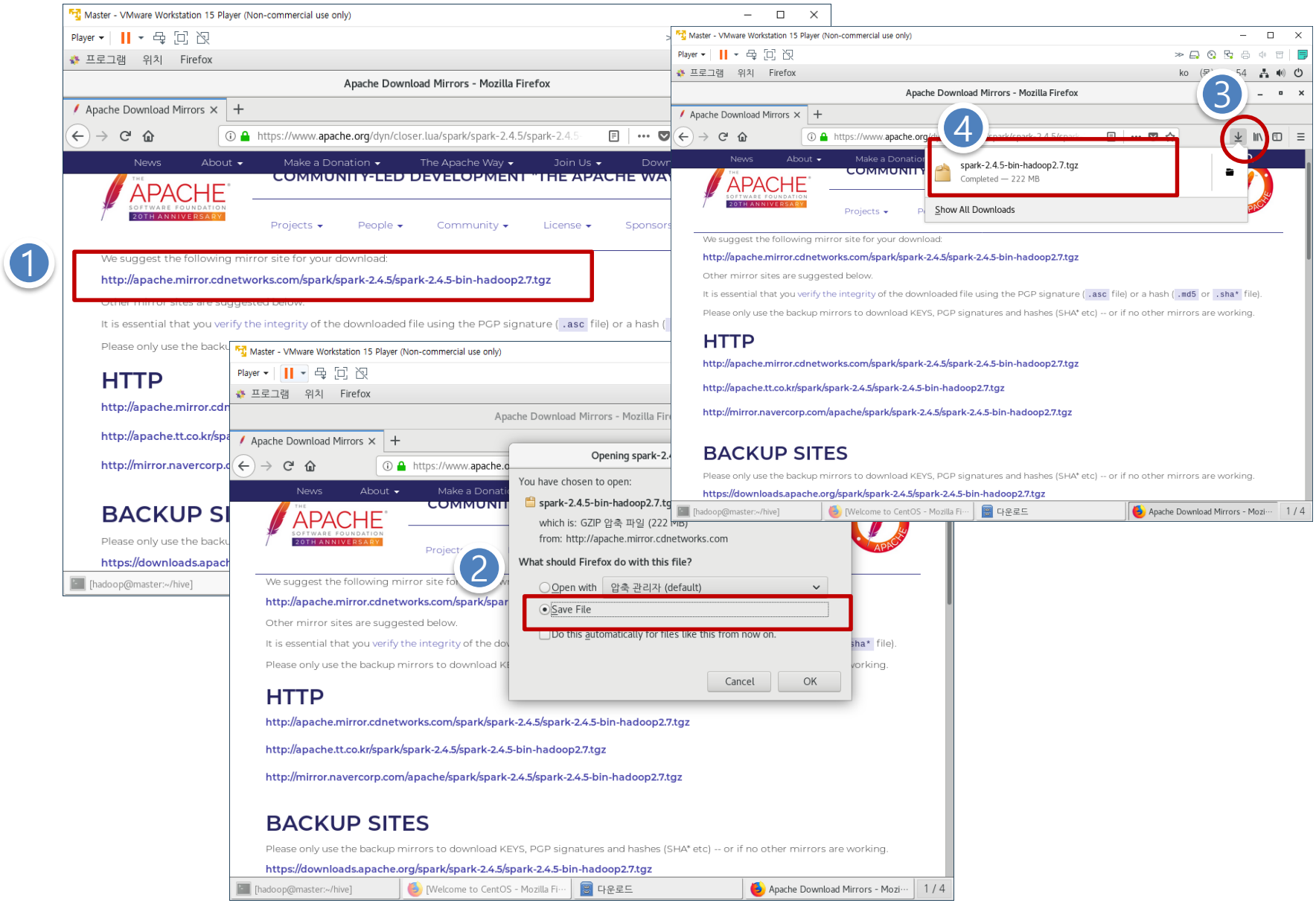


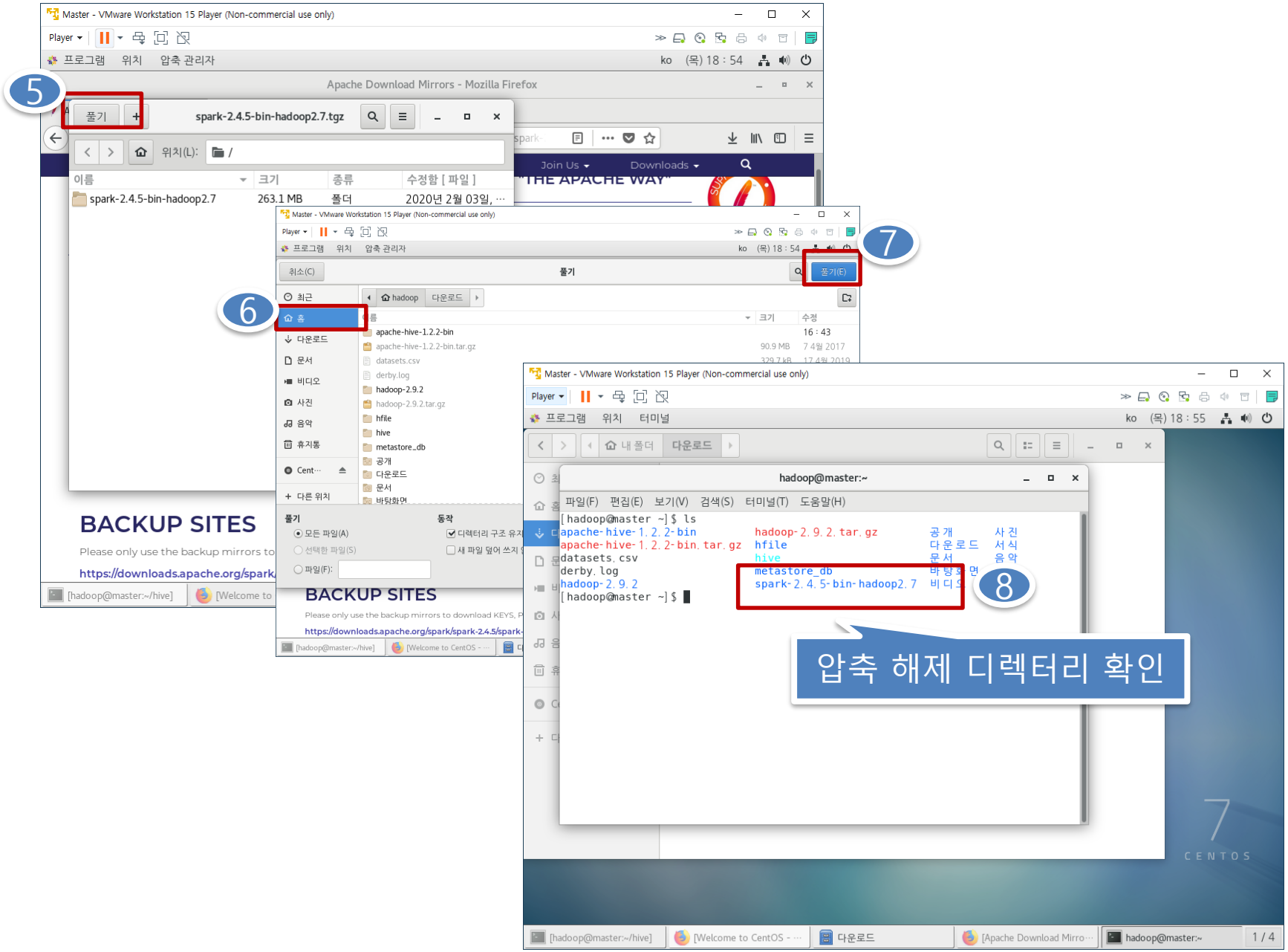
- Spark SQL
  Hive 대신 Spark SQL를 통해 MapReduce없이 빠르게 처리
- Spark SQL CLI
  HiveQL을 이용하여 테이블을 작성하거나 테이블에 데이터를 로드하고, 테이블에 대화식으로 쿼리를 발행하여 분산처리 구현
- Spark Steaming
  스트림 데이터를 짧은 간격으로 읽어서 처리하는 처리하는 준 실시간 데이터 처리 방식
- MLib for machine learning
  Classification, Regression, Clustering 등의 다양한 ML 알고리즘 지원
- GraphX
  그래픽스 처리용 라이브러리 지원

2. Spark 다운로드

http://spark.apache.org/downloads.html 에서 다운로드 가능한 spark 버전 확인




1
2
3
4
Master - VMware Workstation 15 Player (Non-commercial use only)
Apache Download Mirrors - Mozilla Firefox
Apache Download Mirrors
https://www.apache.org/dyn/closer.lua/spark/spark-2.4.5/spark-2.4.5
We suggest the following mirror site for your download:
http://apache.mirror.cdnetworks.com/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz
Other mirror sites are suggested below.
It is essential that you verify the integrity of the downloaded file using the PGP signature (.asc file) or a hash (.md5 or .sha* file).
Please only use the backup mirrors to download KEYS, PGP signatures and hashes (SHA* etc) -- or if no other mirrors are working.
HTTP
http://apache.mirror.cdnetworks.com/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz
http://apache.tt.co.kr/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz
http://mirror.navercorp.com/apache/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz
BACKUP SITES
https://downloads.apache.org/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz
Opening spark-2.4.5-bin-hadoop2.7.tgz
You have chosen to open:
which is: GZIP 압축 파일 (222 MB)
from: http://apache.mirror.cdnetworks.com
What should Firefox do with this file?
Open with 압축 관리자 (default)
Save File
Do this automatically for files like this from now on.
Cancel
OK
spark-2.4.5-bin-hadoop2.7.tgz
Completed — 222 MB
Show All Downloads


5
6
7
8
spark-2.4.5-bin-hadoop2.7.tgz
spark-2.4.5-bin-hadoop2.7
hadoop@master:~
[hadoop@master ~]$ ls
apache-hive-1.2.2-bin
apache-hive-1.2.2-bin.tar.gz
datasets.csv
derby.log
hadoop-2.9.2
hadoop-2.9.2.tar.gz
hfile
hive
metastore_db
spark-2.4.5-bin-hadoop2.7
압축 해제 디렉터리 확인

3. Spark 설치

- **Soft link**

[hadoop@master ~]$ ln -s spark-2.2.0-bin-hadoop2.7 spark

```
drwxrwxr-x.  2 hadoop hadoop      4096  7월 11 18:42 hfile
lrwxrwxrwx.  1 hadoop hadoop        21  7월 10 11:50 hive -> apache-hive-1.2.2-bin
drwxrwxr-x.  3 hadoop hadoop      4096  7월 11 12:46 hive_result
drwxrwxr-x.  3 hadoop hadoop      4096  7월 11 17:03 hive_result2
drwxrwxr-x.  3 hadoop hadoop      4096  7월 11 16:48 home
drwxrwxr-x.  5 hadoop hadoop      4096  7월 11 18:54 metastore_db
lrwxrwxrwx.  1 hadoop hadoop        25  7월 11 19:08 spark -> spark-2.2.0-bin-hadoop2.7
drwxr-xr-x. 12 hadoop hadoop      4096  7월  1  2017 spark-2.2.0-bin-hadoop2.7
-rw-rw-r--.  1 hadoop hadoop 203728858  7월  9 17:58 spark-2.2.0-bin-hadoop2.7.tgz
drwxrwxr-x.  2 hadoop hadoop      4096  7월  9 10:09 test
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 공개
drwxr-xr-x.  2 hadoop hadoop      4096  7월 11 16:07 다운로드
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 문서
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 바탕화면
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 비디오
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 사진
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 서식
drwxr-xr-x.  2 hadoop hadoop      4096  7월  6 10:45 음악
[hadoop@master ~]$
```

4. Spark 환경설정

1) .bash_profile 수정

[hadoop@master ~]$ vi .bash_profile

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

export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.171-8.b10.el7_5.x86_64
export HADOOP_HOME=/home/hadoop/hadoop-2.7.6
export HIVE_HOME=/home/hadoop/hive

PATH=$PATH:$HOME/.local/bin:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$H
IVE_HOME/bin

export SPARK_HOME=/home/hadoop/spark
export PATH=$PATH:$SPARK_HOME/bin


export PATH

~
".bash_profile" 21L, 485C                                    21,0-1        모두
```

Spark 홈 디렉터리 및 Path 설정

2) .bash_profile 적용/테스트

```
hadoop@master:~
파일(F)  편집(E)  보기(V)  검색(S)  터미널(T)  도움말(H)
[hadoop@master ~]$ vi .bash_profile
[hadoop@master ~]$
[hadoop@master ~]$
[hadoop@master ~]$ source .bash_profile
[hadoop@master ~]$
[hadoop@master ~]$
[hadoop@master ~]$ $SPARK_HOME
bash: /home/hadoop/spark: 디렉터리입니다
[hadoop@master ~]$
[hadoop@master ~]$
```

3) Spark 홈 디렉터리 이동

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
[hadoop@master ~]$ vi .bash_profile
[hadoop@master ~]$
[hadoop@master ~]$
[hadoop@master ~]$ source .bash_profile
[hadoop@master ~]$
[hadoop@master ~]$
[hadoop@master ~]$ $SPARK_HOME
bash: /home/hadoop/spark: 디렉터리입니다
[hadoop@master ~]$
[hadoop@master ~]$ cd $SPARK_HOME
[hadoop@master spark]$
```

4) Spark-env.sh 파일 생성/수정

```
hadoop@master:~/spark/conf
파일(F)  편집(E)  보기(V)  검색(S)  터미널(T)  도움말(H)
[hadoop@master ~]$
[hadoop@master ~]$ cd $SPARK_HOME
[hadoop@master spark]$ ls
LICENSE   README.md   conf      jars      sbin
NOTICE    RELEASE     data      licenses  spark-warehouse
R         bin         examples  python    yarn
[hadoop@master spark]$ cd conf/
[hadoop@master conf]$ ls
docker.properties.template    slaves.template
fairscheduler.xml.template    spark-defaults.conf.template
log4j.properties.template     spark-env.sh.template
metrics.properties.template
[hadoop@master conf]$
[hadoop@master conf]$
[hadoop@master conf]$
[hadoop@master conf]$ pwd
/home/hadoop/spark/conf
[hadoop@master conf]$ cp spark-env.sh.template spark-env.sh
[hadoop@master conf]$
[hadoop@master conf]$ vi spark-env.sh
[hadoop@master conf]$
```

4) Spark-env.sh 파일 생성/수정

hadoop@master:~/spark/conf

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
# - SPARK_HISTORY_OPTS, to set config properties only for the history server (e.g. "-D
x=y")
# - SPARK_SHUFFLE_OPTS, to set config properties only for the external shuffle service
 (e.g. "-Dx=y")
# - SPARK_DAEMON_JAVA_OPTS, to set config properties for all daemons (e.g. "-Dx=y")
# - SPARK_PUBLIC_DNS, to set the public dns name of the master or workers

# Generic options for the daemons used in the standalone deploy mode
# - SPARK_CONF_DIR      Alternate conf dir. (Default: ${SPARK_HOME}/conf)
# - SPARK_LOG_DIR       Where log files are stored.  (Default: ${SPARK_HOME}/logs)
# - SPARK_PID_DIR       Where the pid file i
# - SPARK_IDENT_STRING  A string representin
# - SPARK_NICENESS      The scheduling prior
# - SPARK_NO_DAEMONIZE  Run the proposed commar                    reground. It will not output
 a PID file.

export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop

~
                                                                    66,0-1        바닥
```

Hadoop을 이용할 수 있도록
환경변수 추가

5. Spark 실행

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
[hadoop@master spark]$
[hadoop@master spark]$ ls
LICENSE  R          RELEASE  conf  examples  licenses  sbin
NOTICE   README.md  bin      data  jars      python    yarn
[hadoop@master spark]$
[hadoop@master spark]$
[hadoop@master spark]$ spark-submit --version
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 2.2.0
      /_/

Using Scala version 2.11.8, OpenJDK 64-Bit Server VM, 1.8.0_171
Branch
Compiled by user jenkins on 2017-06-30T22:58:04Z
Revision
Url
Type --help for more information.
[hadoop@master spark]$
```

Spark 관련 디렉터리 보기

Spark 설치 버전 확인

- **원주율 근사치 구하기**

spark-submit --class 애플리케이션class 애플리케이션의클래스가포함된JAR파일 파라미터

```
[hadoop@master spark]$ spark-submit --class org.apache.spark.examples.SparkPi examples/jars/spark-examples_2.11-2.2.0.jar 10
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
18/07/10 17:03:14 INFO SparkContext: Running Spark version 2.2.0
18/07/10 17:03:15 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin
```

원주율 근사치 결과

```
18/07/10 17:03:19 INFO Executor: Finished task 8.0 in stage 0.0 (TID 8). 824 bytes result sent to driver
18/07/10 17:03:19 INFO TaskSetManager: Finished task 8.0 in stage 0.0 (TID 8) in 39 ms on localhost (executor driver) (9/10)
18/07/10 17:03:19 INFO Executor: Running task 9.0 in stage 0.0 (TID 9)
18/07/10 17:03:19 INFO Executor: Finished task 9.0 in stage 0.0 (TID 9). 824 bytes result sent to driver
18/07/10 17:03:19 INFO TaskSetManager: Finished task 9.0 in stage 0.0 (TID 9) in 28 ms on localhost (executor driver) (10/10)
18/07/10 17:03:19 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/07/10 17:03:19 INFO DAGScheduler: ResultStage 0 (reduce at SparkPi.scala:38) finished in 0.590 s
18/07/10 17:03:19 INFO DAGScheduler: Job 0 finished: reduce at SparkPi.scala:38, took 0.972947 s
Pi is roughly 3.1415871415871415
18/07/10 17:03:19 INFO SparkUI: Stopped Spark web UI at http://192.168.220.5:4040
18/07/10 17:03:19 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/07/10 17:03:19 INFO MemoryStore: MemoryStore cleared
18/07/10 17:03:19 INFO BlockManager: BlockManager stopped
18/07/10 17:03:19 INFO BlockManagerMaster: BlockManagerMaster stopped
18/07/10 17:03:19 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/07/10 17:03:19 INFO SparkContext: Successfully stopped SparkContext
18/07/10 17:03:19 INFO ShutdownHookManager: Shutdown hook called
18/07/10 17:03:19 INFO ShutdownHookManager: Deleting directory /tmp/spark-c6519cb3-d9fd-4f45-9f73-a66110b029ee
[hadoop@master spark]$
```

6. Spark SQL CLI 실행

- **Hive 서버 연동으로 Spark SQL 실행**

1. **Hive 서버 연동을 위해서 $SPARK_HOME/conf 디렉터리 Hive와 Hadoop 설정 파일 복사**
   대상 파일 : hive-site.xml, core-site.xml, hdfs-site.xml 복사

```
hadoop@master:~/spark/conf
파일(F)  편집(E)  보기(V)  검색(S)  터미널(T)  도움말(H)
[hadoop@master conf]$
[hadoop@master conf]$
[hadoop@master conf]$ pwd
/home/hadoop/spark/conf
[hadoop@master conf]$
[hadoop@master conf]$ cp $HIVE_HOME/conf/hive-site.xml .
[hadoop@master conf]$
[hadoop@master conf]$ cp $HADOOP_HOME/etc/hadoop/core-site.xml .
[hadoop@master conf]$
[hadoop@master conf]$ cp $HADOOP_HOME/etc/hadoop/hdfs-site.xml .
[hadoop@master conf]$
[hadoop@master conf]$ ls
core-site.xml                hive-site.xml                spark-defaults.conf.template
docker.properties.template   log4j.properties.template     spark-env.sh
fairscheduler.xml.template   metrics.properties.template   spark-env.sh.template
hdfs-site.xml                slaves.template
[hadoop@master conf]$
```

2. Hadoop 실행

hadoop@master:~/spark/conf

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
[hadoop@master conf]$
[hadoop@master conf]$
[hadoop@master conf]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [master]
master: starting namenode, logging to /home/hadoop/hadoop-2.7.6/logs/hadoop-hadoop-namenode-ma
ster.out
slave1: starting datanode, logging to /home/hadoop/hadoop-2.7.6/logs/hadoop-hadoop-datanode-sl
ave1.out
slave2: starting datanode, logging to /home/hadoop/hadoop-2.7.6/logs/hadoop-hadoop-datanode-sl
ave2.out
Starting secondary namenodes [slave1]
slave1: starting secondarynamenode, logging to /home/hadoop/hadoop-2.7.6/logs/hadoop-hadoop-se
condarynamenode-slave1.out
starting yarn daemons
starting resourcemanager, logging to /home/hadoop/hadoop-2.7.6/logs/yarn-hadoop-resourcemanage
r-master.out
slave1: starting nodemanager, logging to /home/hadoop/hadoop-2.7.6/logs/yarn-hadoop-nodemanage
r-slave1.out
slave2: starting nodemanager, logging to /home/hadoop/hadoop-2.7.6/logs/yarn-hadoop-nodemanage
r-slave2.out
[hadoop@master conf]$
```

3. Spark-sql 실행

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
[hadoop@master spark]$
[hadoop@master spark]$
[hadoop@master spark]$ pwd
/home/hadoop/spark
[hadoop@master spark]$
[hadoop@master spark]$ spark-sql
18/07/11 19:35:34 WARN conf.HiveConf: HiveConf of name hive.conf.hidden.list does not ex
ist
18/0
our
18/0
clas
18/0
18/0
hdow
```

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
18/07/11 19:35:47 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator
endpoint
18/07/11 19:35:47 WARN conf.HiveConf: HiveConf of name hive.conf.hidden.list does not ex
ist
18/07/11 19:35:47 INFO session.SessionState: Created local directory: /tmp/hive/java/7d1
584ba-9eb9-45a6-8946-b9bd7b1cc9b6_resources
18/07/11 19:35:47 INFO session.SessionState: Created HDFS directory: /tmp/hive/hadoop/7d
1584ba-9eb9-45a6-8946-b9bd7b1cc9b6
18/07/11 19:35:47 INFO session.SessionSta                    ve/java/had
oop/7d1584ba-9eb9-45a6-8946-b9bd7b1cc9b6
18/07/11 19:35:47 INFO session.SessionSta                    e/hadoop/7d
1584ba-9eb9-45a6-8946-b9bd7b1cc9b6/_tmp_s
18/07/11 19:35:47 INFO client.HiveClientImpl:                ouse location for Hive client (versio
n 1.2.1) is /user/hive/warehouse
spark-sql>
```

mataStore와
Spark-sql 프롬프트

4. Spark SQL 실습 : iris 테이블 생성

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
2bb0/_tmp_space.db
18/07/11 19:41:04 INFO client.HiveClientImpl: Warehouse location for Hive client (version 1.2.1)
se
spark-sql> create table iris_tab(
         > col1 float, col2 float, col3 float, col4 float, col5 string)
         > row format delimited fields terminated by ',' stored as textfile;
18/07/11 19:43:41 INFO execution.SparkSqlParser: Parsing command: create table iris_tab(
col1 float, col2 float, col3 float, col4 float, col5 string)
row format delimited fields terminated by ',' stored as textfile
18/07/1
18/07/1
18/07/1
18/07/1
18/07/
18/07/1
```

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
Time taken: 2.34 seconds
18/07/11 19:43:43 INFO CliDriver: Time taken: 2.34 seconds
spark-sql> show tables;
18/07/11 19:44:49 INFO execution.SparkSqlParser: Parsing command: show tables
18/07/11 19:44:49 INFO metastore.HiveMetaStore: 0: get_database: default
18/07/11 19:44:49 INFO HiveMetaStore.audit: ugi=hadoop ip=unknown-ip-addr cmd=get_datab
18/07/11 19:44:49 INFO metastore.HiveMetaStore: 0: get_database: default
18/07/11 19:44:49 INFO HiveMetaStore.audit: ugi=hadoop ip=unknown-ip-addr cmd=get_datab
18/07/11 19:44:49 INFO metastore.HiveMetaStore: 0: get_tables: db=default pat=*
18/07/11 19:44:49 INFO HiveMetaStore.audit: ugi=hadoop ip=unknown-ip-addr cmd=get_table
18/07/11 19:44:50 INFO codegen.CodeGenerator: Code generated in 197.609696 ms
default iris_tab false
Time taken: 0.599 seconds, Fetched 1 row(s)
18/07/11 19:44:50 INFO CliDriver: Time taken: 0.599 seconds, Fetched 1 row(s)
spark-sql>
```

4. Spark SQL 실습 : iris 테이블에 데이터 삽입

hadoop@master:~/spark

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
18/07/11 19:44:50 INFO CliDriver: Time taken: 0.599 seconds, Fetched 1 row(s)
spark-sql>
         >
         > load data local inpath '/home/hadoop/hfile/iris.csv' into table iris_tab;
18/07/11 19:49:11 INFO execution.SparkSqlParser: Parsing command: load data inpath '/
home/hadoop/hfile/iris.csv' into table iris_tab
18/07/11 19:49:11 INFO metastore.HiveMetaStore: 0: get_database: default
18/07/11 19:49:11 INFO HiveMetaStore.audit: ugi=hadoop  ip=unknown-ip-addr      cmd=g
et_database: default
18/07/11 19:49:11 INFO metastore.HiveMetaStore: 0: get_table : db=default tbl=iris_ta
b
18/07/11 19:49:11 INFO HiveMetaStore.audit: ugi=hadoop  ip=unknown-ip-addr      cmd=g
et_table : db=default tbl=iris_tab
18/07/11 19:49:11 INFO metastore.HiveMetaStore: 0: get_table : db=default tbl=iris_ta
b
```

4. Spark SQL 실습 : iris 테이블 조회

```
18/07/11 19:49:11 INFO HiveMetaStore.audit: ugi=hadoop  ip=unknown-ip-addr      cmd=a
lter_table: db=default tbl=iris_tab newtbl=iris_tab
18/07/11 19:49:11 INFO hive.log: Updating table stats fast for iris_tab
18/07/11 19:49:11 INFO hive.log: Updated size of table iris_tab to 4177
Time taken: 0.524 seconds
18/07/11 19:49:11 INFO CliDriver: Time taken: 0.524 seconds
spark-sql> select * from iris_tab;
18/07/11 19:49:32 INFO execution.SparkSqlParser: Parsing command: select * from iris_
tab
18/07/11 19:49:32 INFO metastore.HiveMetaStore: 0: get_table : db=default tbl=iris_ta
b
18/07/11 19:49:32 INFO H.java:376) finished in 0.466 s
et_table : db=default tb18/07/11 19:49:34 INFO scheduler.DAGScheduler: Job 0 finished:
18/07/11 19:49:32 INFO pr.java:376, took 0.618930 s
NULL    NULL    NULL    NULL    "Species"
5.1     3.5     1.4     0.2     "setosa"
4.9     3.0     1.4     0.2     "setosa"
4.7     3.2     1.3     0.2     "setosa"
4.6     3.1     1.5     0.2     "setosa"
5.0     3.6     1.4     0.2     "setosa"
5.4     3.9     1.7     0.4     "setosa"
4.6     3.4     1.4     0.3     "setosa"
5.0     3.4     1.5     0.2     "setosa"
4.4     2.9     1.4     0.2     "setosa"
4.9     3.1     1.5     0.1     "setosa"
5.4     3.7     1.5     0.2     "setosa"
4.8     3.4     1.6     0.2     "setosa"
4.8     3.0     1.4     0.1     "setosa"
```

5. Hive meta Sore에서 Spark 테이블 확인

hadoop@master:~

파일(F) 편집(E) 보기(V) 검색(S) 터미널(T) 도움말(H)

```
[hadoop@master ~]$ hdfs dfs -ls /user/hive/warehouse
Found 10 items
-rwxr-xr-x   3 hadoop supergroup      28124 2018-07-11 17:05 /user/hive/warehouse/000000_0
-rwxr-xr-x   3 hadoop supergroup      28286 2018-07-11 17:05 /user/hive/warehouse/000001_0
-rwxr-xr-x   3 hadoop supergroup      28212 2018-07-11 17:05 /user/hive/warehouse/000002_0
drwxr-xr-x   - hadoop supergroup          0 2018-07-11 16:11 /user/hive/warehouse/airline_delay
drwxr-xr-x   - hadoop supergroup          0 2018-07-11 10:26 /user/hive/warehouse/init_table
-rwxr-xr-x   3 hadoop supergroup       4177 2018-07-11 15:36 /user/hive/warehouse/iris_csv
drwxr-xr-x   - hadoop supergroup          0 2018-07-11 18:50 /user/hive/warehouse/iris_tab
drwxr-xr-x   - hadoop supergroup          0 2018-07-11 18:30 /user/hive/warehouse/spark_table
drwxr-xr-x   - hadoop supergroup          0 2018-07-11 11:13 /user/hive/warehouse/stocks
drwxr-xr-x   - hadoop supergroup          0 2018-07-11 10:16 /user/hive/warehouse/test_tab
[hadoop@master ~]$
```

5. 테이블 삭제 및 종료

```
ble.
18/07/11 19:57:30 INFO metastore.hivemetastoressimpl: deleting  hdfs://master:9000/us
er/hive/warehouse/iris_tab
18/07/11 19:57:30 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion
interval = 0 minutes, Emptier interval = 0 minutes.
18/07/11 19:57:31 INFO metastore.hivemetastoressimpl: Deleted the diretory hdfs://mas
ter:9000/user/hive/warehouse/iris_tab
Time taken: 1.357 seconds
18/07/11 19:57:31 INFO CliDriver: Time taken: 1.357 seconds
spark-sql> drop table iris_tab;
```

```
spark-sql> quit;
18/07/11 20:00:18 INFO server.AbstractConnector: Stopped Spark@1e6bd263{HTTP/1.1,[htt
p/1.1]}{0.0.0.0:4040}
18/07/11 20:00:18 INFO ui.SparkUI: Stopped Spark web UI at http://192.168.220.5:4040
18/07/11 20:00:18 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEn
dpoint stopped!
18/07/11 20:00:18 INFO memory.MemoryStore: MemoryStore cleared
18/07/11 20:00:18 INFO storage.BlockManager: BlockManager stopped
18/07/11 20:00:18 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/07/11 20:00:18 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpo
int: OutputCommitCoordinator stopped!
18/07/11 20:00:18 INFO spark.SparkContext: Successfully stopped SparkContext
18/07/11 20:00:18 INFO util.ShutdownHookManager: Shutdown hook called
18/07/11 20:00:18 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-6b9a3e
55-7e27-43ac-8319-5b52175c2c6d
18/07/11 20:00:18 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-f024a6
f4-e14a-49e0-b458-a2a86961c80b
[hadoop@master spark]$
```