1. Hive 설치

1-1. Hive 다운로드 및 압축해체

```
hadoop@hadoop01:~$ wget http://apache.mirror.cdnetworks.com/hive/hive-1.2.2/apac
he-hive-1.2.2-bin.tar.gz
--2017-06-06 08:02:50--  http://apache.mirror.cdnetworks.com/hive/hive-1.2.2/apa
che-hive-1.2.2-bin.tar.gz
Resolving apache.mirror.cdnetworks.com (apache.mirror.cdnetworks.com)... 14.0.10
1.165
Connecting to apache.mirror.cdnetworks.com (apache.mirror.cdnetworks.com)|14.0.1
01.165|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 90859180 (87M) [application/x-gzip]
Saving to: 'apache-hive-1.2.2-bin.tar.gz'

apache-hive-1.2.2-b 100%[===================>]  86.65M  3.87MB/s    in 22s

2017-06-06 08:03:13 (3.89 MB/s) - 'apache-hive-1.2.2-bin.tar.gz' saved [90859180
/90859180]

hadoop@hadoop01:~$ tar -zxvf apache-hive-1.2.2-bin.tar.gz
```

```
wget http://apache.mirror.cdnetworks.com/hive/hive-1.2.2/apache-hive-1.2.2-bin.tar.gz
tar -zxvf apache-hive-1.2.2-bin.tar.gz
```

1-2. ~/.bashrc 파일에 환경변수 설정

```
119 export JAVA_HOME=/usr/lib/jvm/java-7-oracle
120 export HADOOP_HOME=/home/hadoop/hadoop-2.7.3
121 export HIVE_HOME=/home/hadoop/apache-hive-1.2.2-bin
122 export PATH=$HIVE_HOME/bin:$HADOOP_HOME/sbin:$HADOOP_HOME/bin:$JAVA_HOME/bin
    :$PATH
```

```
export HIVE_HOME=/home/hadoop/apache-hive-1.2.2-bin
export PATH=$HIVE_HOME/bin:$HADOOP_HOME/sbin:$HADOOP_HOME/bin:
$JAVA_HOME/bin:$PATH
```

1-3. 하둡 디렉토리 생성

```
hadoop@hadoop01:~/hadoop-2.7.3$ bin/hadoop fs -mkdir /tmp
hadoop@hadoop01:~/hadoop-2.7.3$ bin/hadoop fs -mkdir /user/hive
hadoop@hadoop01:~/hadoop-2.7.3$ bin/hadoop fs -mkdir /user/hive/warehouse
hadoop@hadoop01:~/hadoop-2.7.3$ bin/hadoop fs -chmod g+w /tmp
hadoop@hadoop01:~/hadoop-2.7.3$ bin/hadoop fs -chmod g+w /user/hive/warehouse
```

```
bin/hadoop fs -mkdir /tmp
bin/hadoop fs -mkdir /user/hive
bin/hadoop fs -mkdir /user/hive/warehouse
bin/hadoop fs -chmod g+w /tmp
bin/hadoop fs -chmod g+w /user/hive/warehouse
```

2. 타슈데이터 Hive 에 업로드하기

2-1. tashu table 생성

```
hadoop@hadoop01:~/apache-hive-1.2.2-bin$ hive

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-1.2
.2-bin/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> create table tashu (RENT_STATION int, RENT_DATE string, RETURN_STATION int
, RETURN_DATE string) row format delimited fields terminated by ',';
OK
Time taken: 1.519 seconds
hive> show tables;
OK
tashu
Time taken: 0.273 seconds, Fetched: 1 row(s)
hive> describe tashu;
OK
rent_station            int
rent_date               string
return_station          int
return_date             string
Time taken: 0.502 seconds, Fetched: 4 row(s)
hive>
```

```
create table tashu (RENT_STATION int, RENT_DATE string, RETURN_STATION int,
RETURN_DATE string) row format delimited fields terminated by ',';
show tables;
describe tashu;
```

2-2. tashi table 에 데이터 업로드

```
hive> load data local inpath 'tashu.csv' overwrite into table tashu;
Loading data to table default.tashu
Table default.tashu stats: [numFiles=1, numRows=0, totalSize=126400696, rawDataS
ize=0]
OK
Time taken: 8.522 seconds
hive> select * from tashu limit 10;
OK
43      20130101055603  34      20130101060217
97      20130101060400  NULL    20130101102037
2       20130101060406  10      20130101061859
106     20130101105305  105     20130101105743
4       20130101112223  4       20130101121753
21      20130101113953  105     20130101114943
90      20130101120833  91      20130101125136
13      20130101131429  30      20130101133039
1       20130101133742  1       20130101133815
1       20130101133813  2       20130101150958
Time taken: 0.431 seconds, Fetched: 10 row(s)
hive>
```

```
load data local inpath 'tashu.csv' overwrite into table tashu;
select * from tashu limit 10;
```

3. 타슈데이터에 쿼리를 날려 통계구하기 (RENT_DATE 기준)

3-1. 연도별 대여량

```
hive> select year(from_unixtime(unix_timestamp(RENT_DATE, 'yyyyMMddHHmmss'))), c
ount(RENT_DATE) from tashu group by year(from_unixtime(unix_timestamp(RENT_DATE,
 'yyyyMMddHHmmss')));
```

```
select year(from_unixtime(unix_timestamp(RENT_DATE, 'yyyyMMddHHmmss'))),
count(RENT_DATE) from tashu group by year(from_unixtime(unix_timestamp(RENT_DATE,
'yyyyMMddHHmmss')));
```

결과

```
Total MapReduce CPU Time Spent: 24 seconds 30 msec
OK
2013    1036614
2014    1200187
2015    1167862
Time taken: 59.502 seconds, Fetched: 3 row(s)
hive>
```

3-2. 일별 대여량

```
hive> select day(from_unixtime(unix_timestamp(RENT_DATE, 'yyyyMMddHHmmss'))), co
unt(RENT_DATE) from tashu group by day(from_unixtime(unix_timestamp(RENT_DATE, '
yyyyMMddHHmmss')));
```

```
select day(from_unixtime(unix_timestamp(RENT_DATE, 'yyyyMMddHHmmss'))),
count(RENT_DATE) from tashu group by day(from_unixtime(unix_timestamp(RENT_DATE,
'yyyyMMddHHmmss')));
```

결과

```
Total MapReduce CPU Time Spent: 24 seconds 140 msec
OK
1       116298
2       104499
3       106474
4       115649
5       115198
6       110271
7       105940
8       107095
9       117166
10      110197
11      105053
12      104176
13      107960
14      110600
15      120347
16      120099
17      114330
18      105077
19      112020
20      106850
21      113472
22      120796
23      111161
24      110613
25      104908
26      120407
27      106867
28      119053
29      102871
30      114542
31      64674
Time taken: 58.556 seconds, Fetched: 31 row(s)
```

3-3. 시간대별 대여량

```
hive> select hour(from_unixtime(unix_timestamp(RENT_DATE, 'yyyyMMddHHmmss'))), c
ount(RENT_DATE) from tashu group by hour(from_unixtime(unix_timestamp(RENT_DATE,
 'yyyyMMddHHmmss')));
```

```
select hour(from_unixtime(unix_timestamp(RENT_DATE, 'yyyyMMddHHmmss'))),
count(RENT_DATE) from tashu group by hour(from_unixtime(unix_timestamp(RENT_DATE,
'yyyyMMddHHmmss')));
```

결과

```
Total MapReduce CPU Time Spent: 24 seconds 470 msec
OK
0       63022
1       15205
2       199
3       3
4       11
5       16591
6       22152
7       129110
8       186421
9       142126
10      111652
11      96250
12      118768
13      157488
14      173437
15      167018
16      194413
17      250842
18      292905
19      241023
20      259816
21      273561
22      251385
23      241265
Time taken: 53.045 seconds, Fetched: 24 row(s)
```