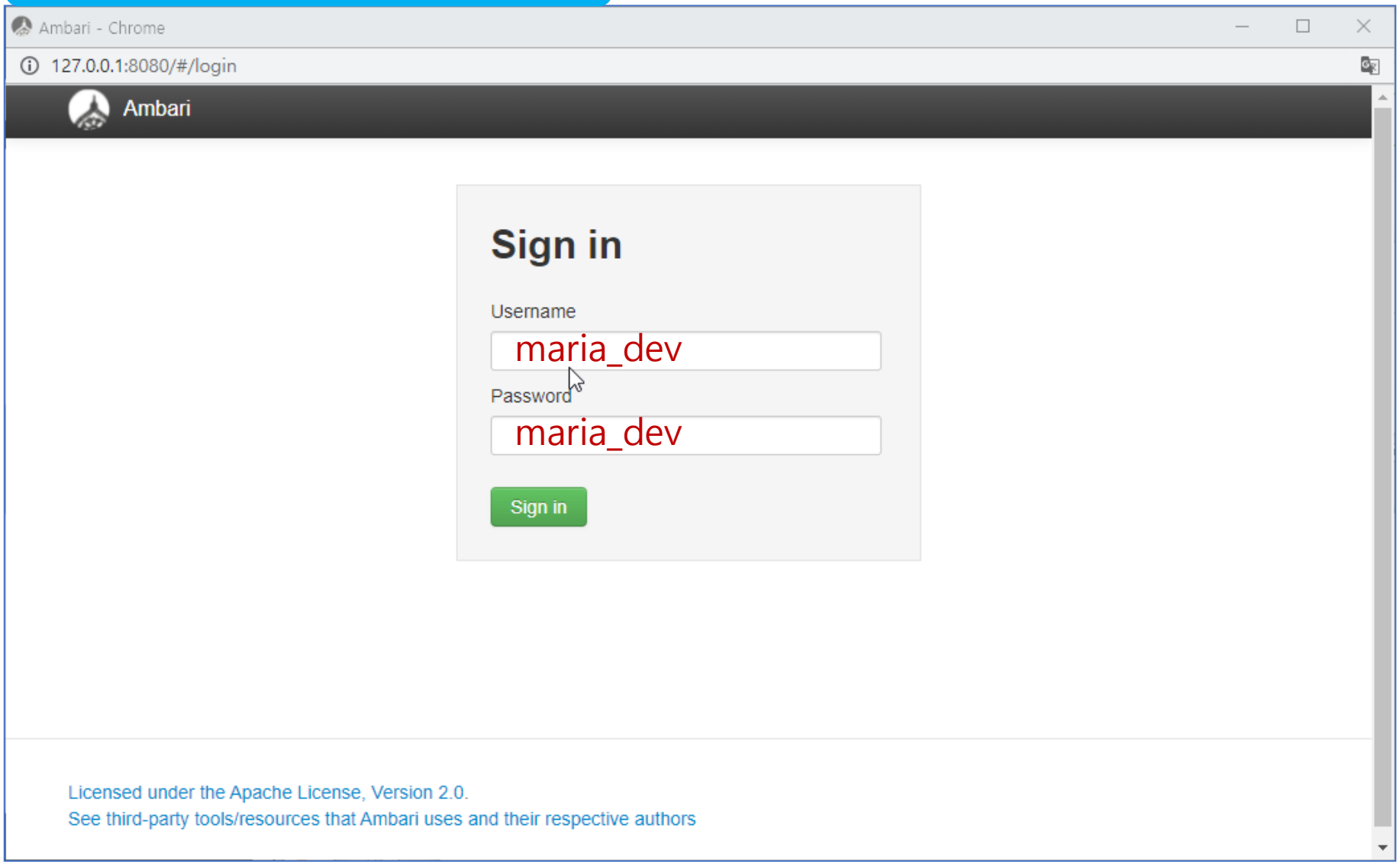
Hadoop 에코시스템

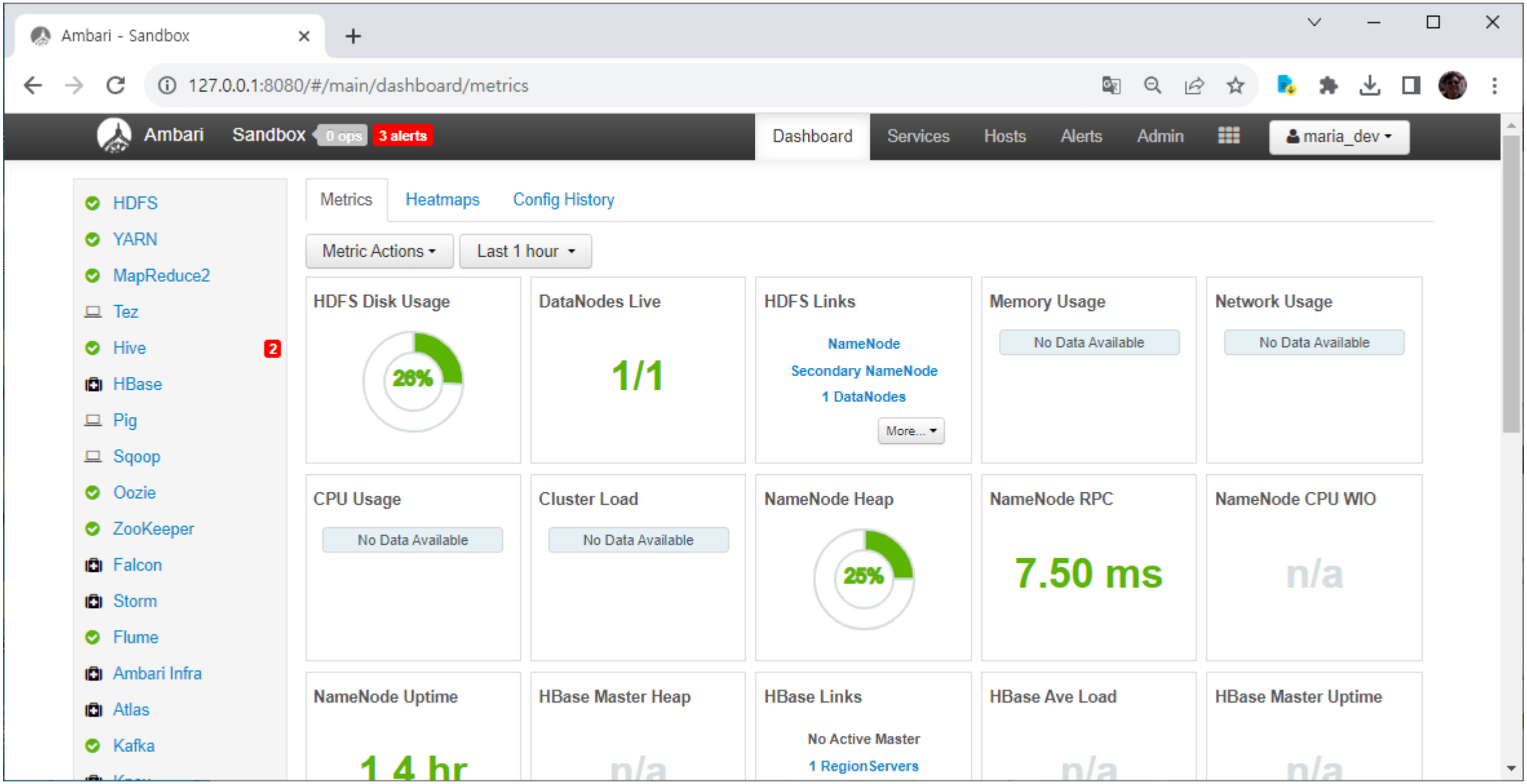
Ambari

Ambari 접속

http://127.0.0.1:8080/



Ambari Dashboard



Ambari Service

Ambari - Sandbox

localhost:8080/#/main/services/HDFS/summary

Ambari Sandbox 0 ops 0 alerts

Dashboard | Services | Hosts | Alerts | Admin | maria_dev

- HDFS
- YARN
- MapReduce2
- Tez
- Hive
- HBase
- Pig
- Sqoop
- Oozie
- ZooKeeper
- Falcon
- Storm
- Flume
- Ambari Infra
- Atlas
- Kafka
- Knox

Summary | Heatmaps | Configs | Quick Links | Service Actions

Summary

No alerts

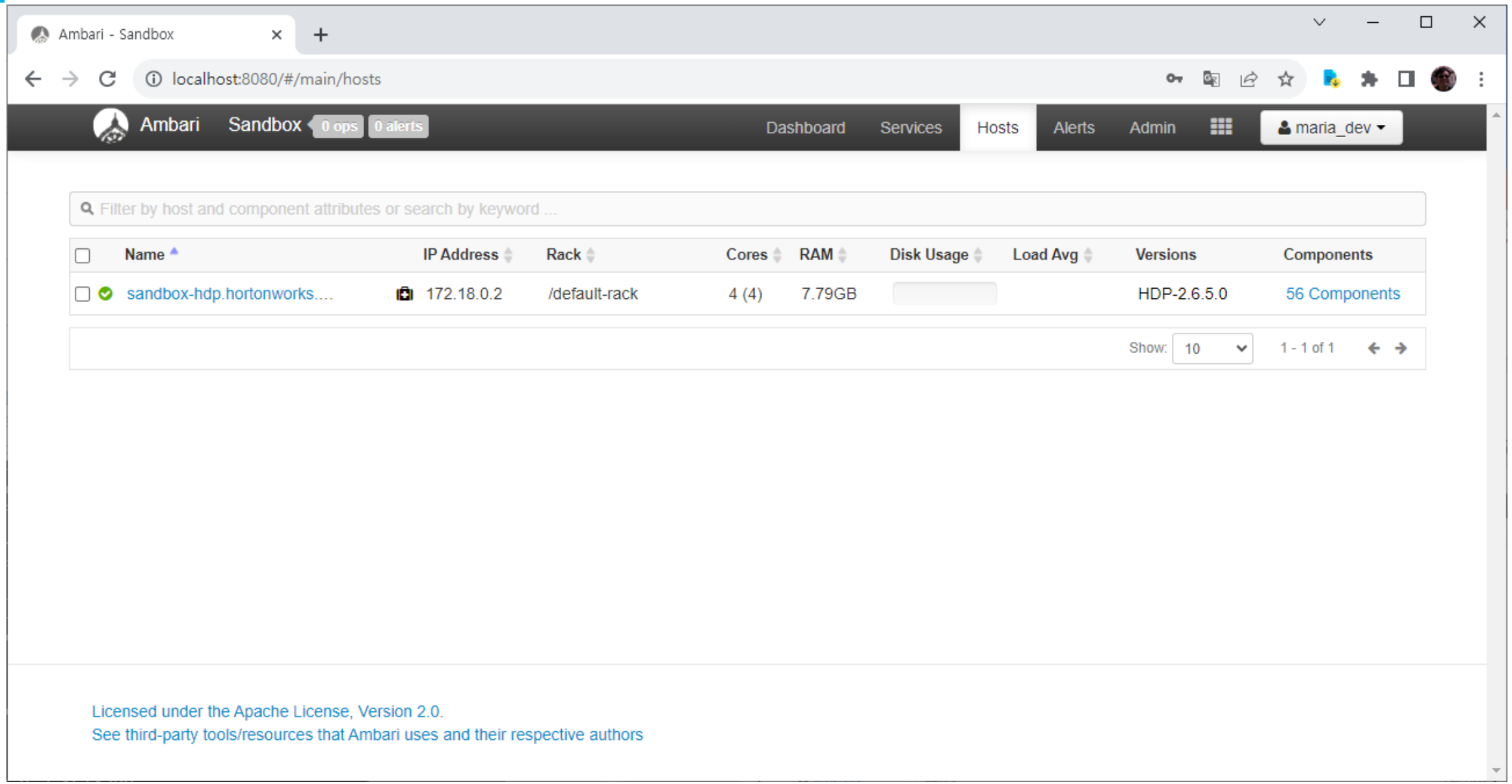
NameNode	Started No alerts
SNameNode	Started No alerts
DataNodes	1/1 Started
DataNodes Status	1 live / 0 dead / 0 decommissioning
JournalNodes	1/1 JournalNodes Live
NFSGateways	0/0 Started
NameNode Uptime	24.76 mins
NameNode Heap	27.7 MB / 240.0 MB (11.5% used)
Disk Usage (DFS Used)	2.1 GB / 106.0 GB (1.95%)
Disk Usage (Non DFS Used)	25.5 GB / 106.0 GB (24.08%)
Disk Remaining	78.4 GB / 106.0 GB (73.97%)
Blocks (total)	1121
Block Errors	0 corrupt replica / 0 missing / 0 under replicated
Total Files + Directories	1352
Upgrade Status	No pending upgrade
Safe Mode Status	Not in safe mode

Metrics

Last 1 hour

NameNode GC count	NameNode GC time	NN Connection Load	NameNode Heap	NameNode Host Load
No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.	No Data No available data for the time period.

Ambari Host


Ambari - Sandbox
localhost:8080/#/main/hosts
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Filter by host and component attributes or search by keyword ...
Name
IP Address
Rack
Cores
RAM
Disk Usage
Load Avg
Versions
Components
sandbox-hdp.hortonworks....
172.18.0.2
/default-rack
4 (4)
7.79GB
HDP-2.6.5.0
56 Components
Show:
10
1 - 1 of 1
Licensed under the Apache License, Version 2.0.
See third-party tools/resources that Ambari uses and their respective authors

Ambari Alerts

Ambari - Sandbox

127.0.0.1:8080/#/main/alerts

Ambari Sandbox 0 ops 3 alerts

Dashboard Services Hosts Alerts Admin maria_dev

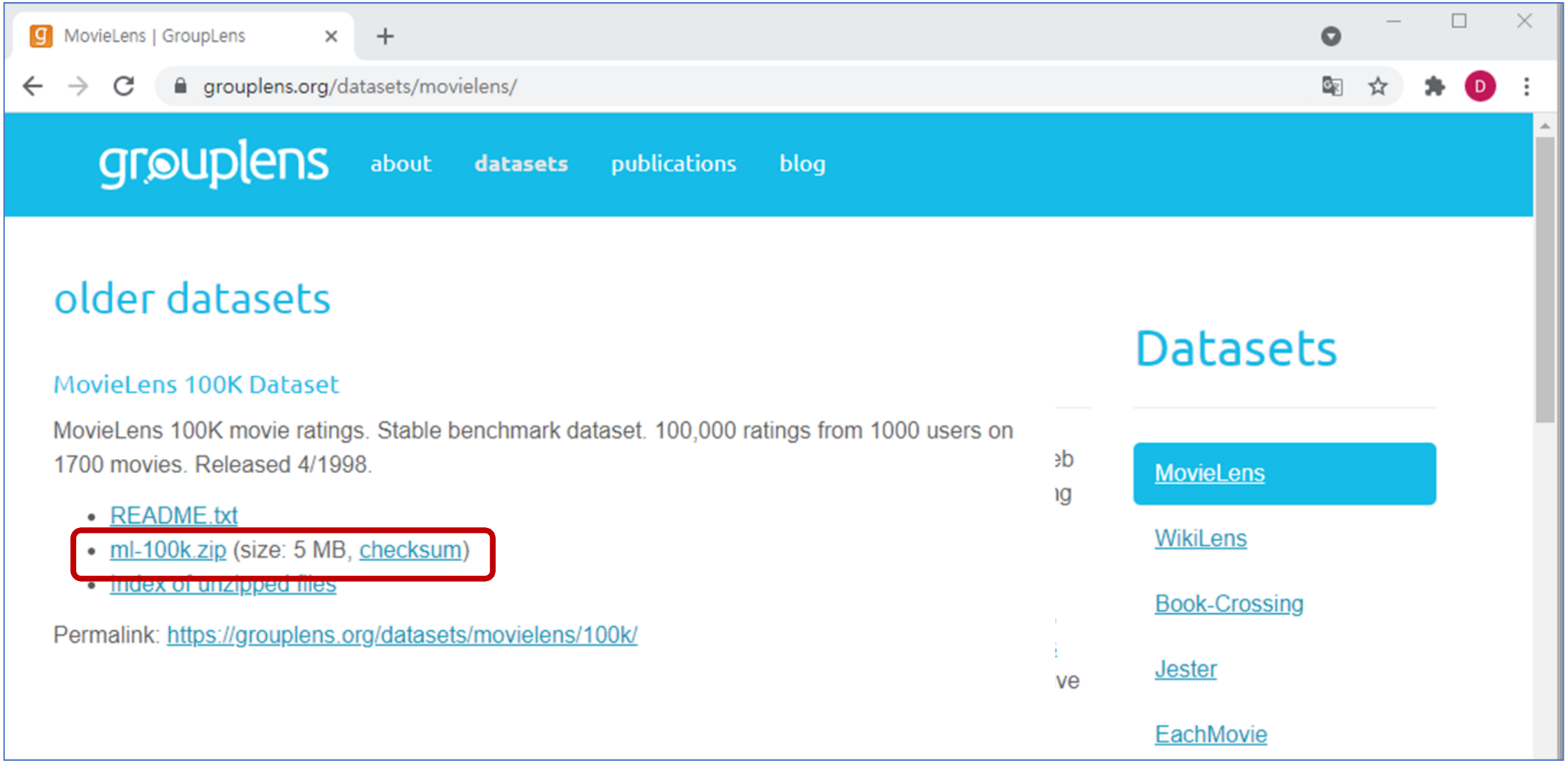
Groups: All (85)

Alert Definition Name	Status	Service	Last Status Changed	State
Any	All	All	Any	All
Falcon Server Web UI	CRIT	Falcon	5 years ago	Enabled
Falcon Server Process	CRIT	Falcon	5 years ago	Enabled
Metadata Server Web UI	CRIT	Atlas	5 years ago	Enabled
HBase Master Process	CRIT	HBase	5 years ago	Enabled
HBase RegionServer Process	CRIT	HBase	5 years ago	Enabled
Knox Gateway Process	CRIT	Knox	5 years ago	Enabled
Hive Metastore Process	CRIT	Hive	2 hours ago	Enabled
HiveServer2 Process	CRIT	Hive	2 hours ago	Enabled
Infra Solr Web UI	CRIT	Ambari Infra	5 years ago	Enabled

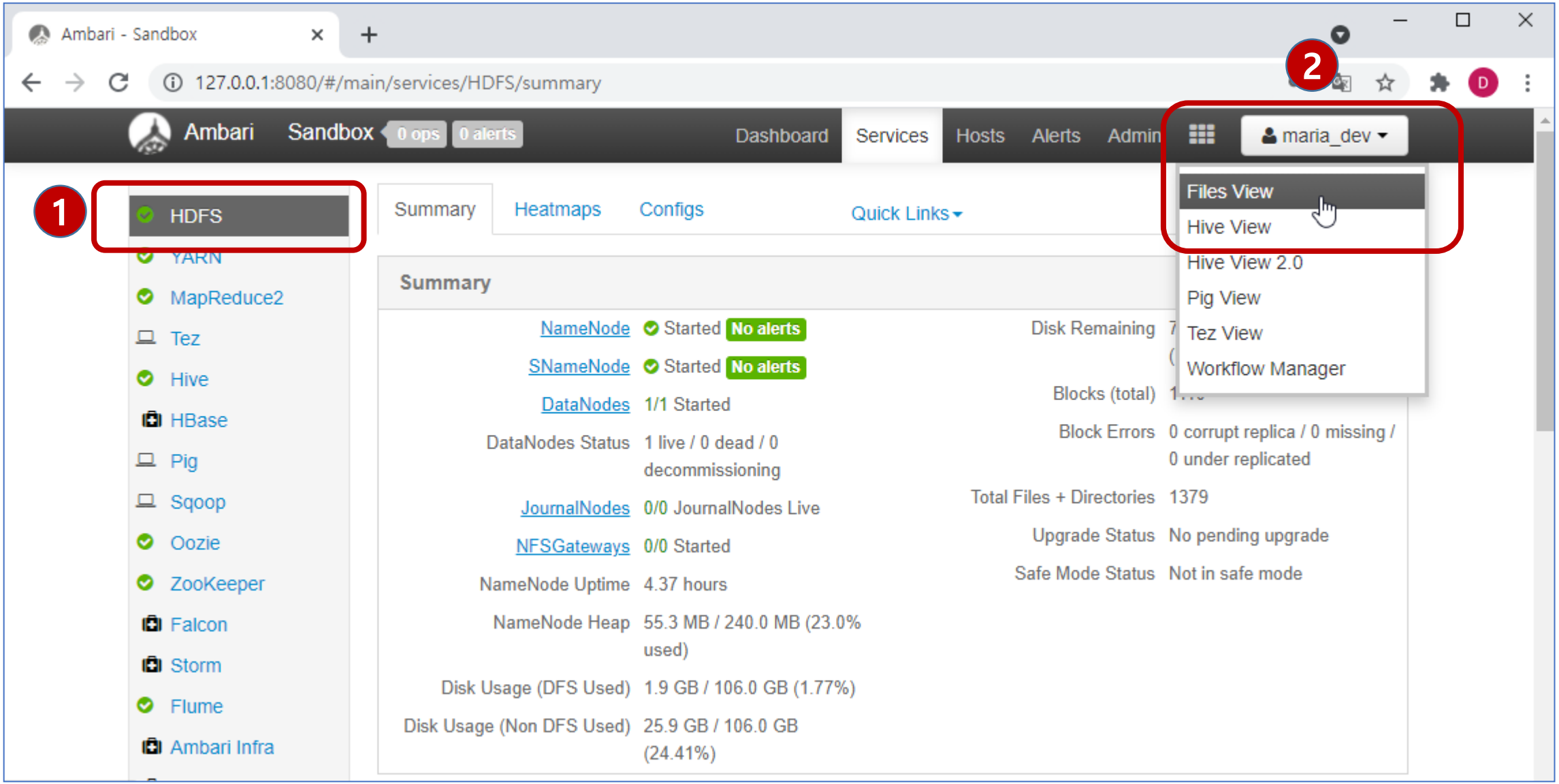
HDFS 실습

MovieLens 데이터 다운로드

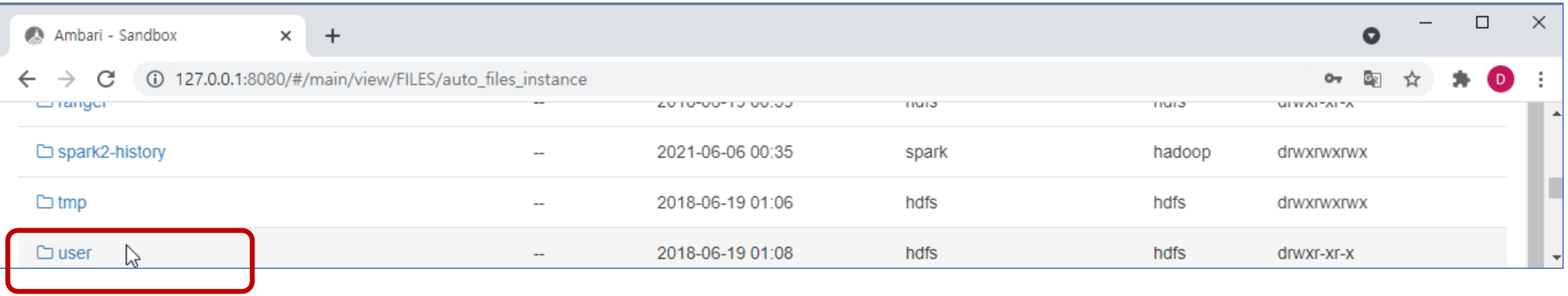
https://grouplens.org/datasets/movielens/

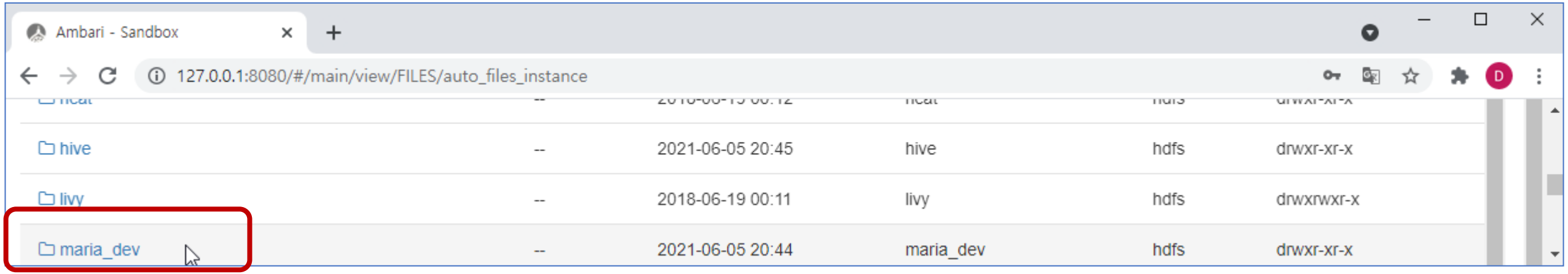


HDFS

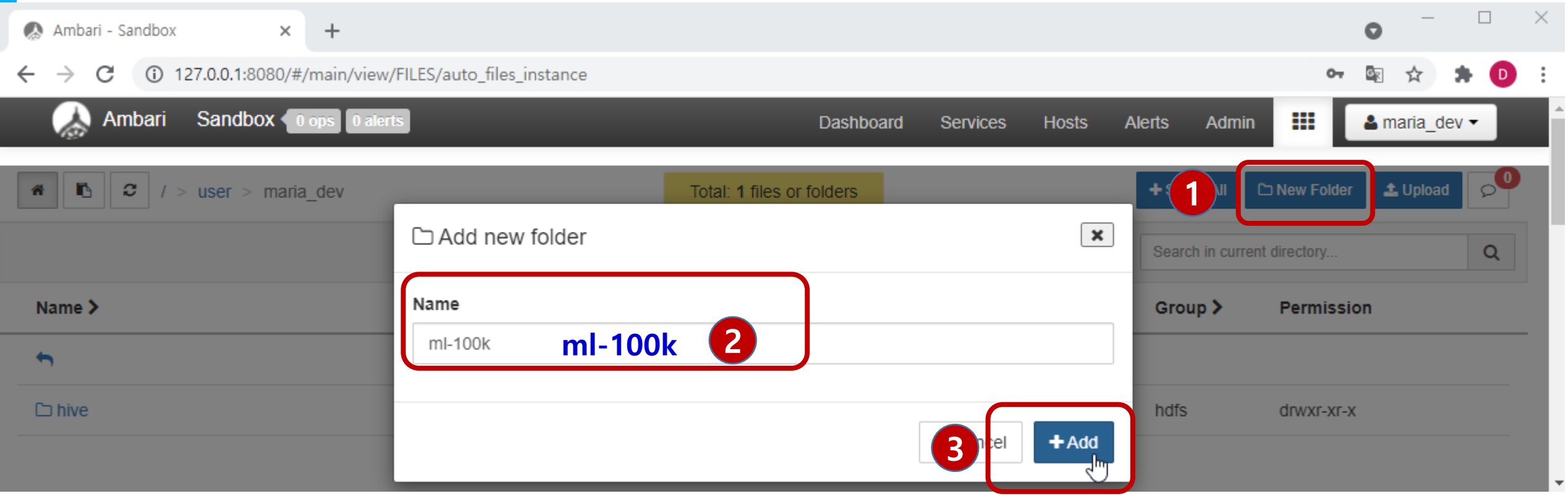

Ambari - Sandbox
127.0.0.1:8080/#/main/services/HDFS/summary
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
1
2
HDFS
YARN
MapReduce2
Tez
Hive
HBase
Pig
Sqoop
Oozie
ZooKeeper
Falcon
Storm
Flume
Ambari Infra
Summary
Heatmaps
Configs
Quick Links
Files View
Hive View
Hive View 2.0
Pig View
Tez View
Workflow Manager
Summary
NameNode Started No alerts
SNameNode Started No alerts
DataNodes 1/1 Started
DataNodes Status 1 live / 0 dead / 0 decommissioning
JournalNodes 0/0 JournalNodes Live
NFSGateways 0/0 Started
NameNode Uptime 4.37 hours
NameNode Heap 55.3 MB / 240.0 MB (23.0% used)
Disk Usage (DFS Used) 1.9 GB / 106.0 GB (1.77%)
Disk Usage (Non DFS Used) 25.9 GB / 106.0 GB (24.41%)
Disk Remaining
Blocks (total)
Block Errors 0 corrupt replica / 0 missing / 0 under replicated
Total Files + Directories 1379
Upgrade Status No pending upgrade
Safe Mode Status Not in safe mode

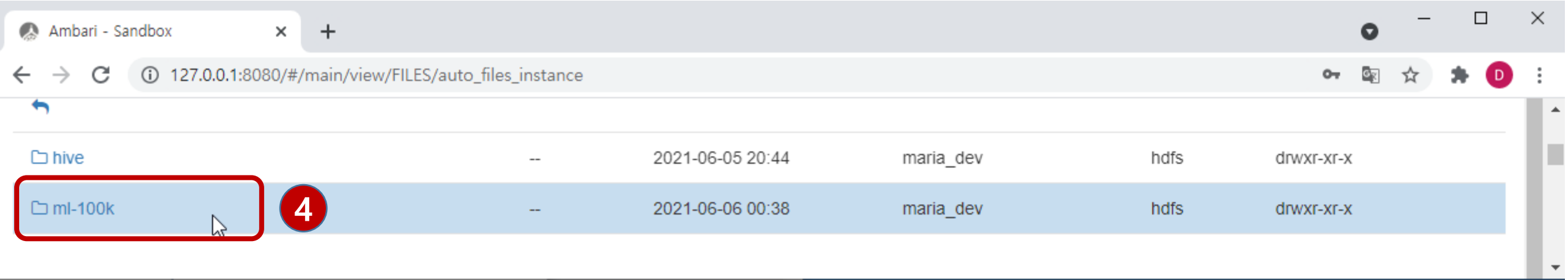
HDFS


Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
spark2-history -- 2021-06-06 00:35 spark hadoop drwxrwxrwx
tmp -- 2018-06-19 01:06 hdfs hdfs drwxrwxrwx
user -- 2018-06-19 01:08 hdfs hdfs drwxr-xr-x

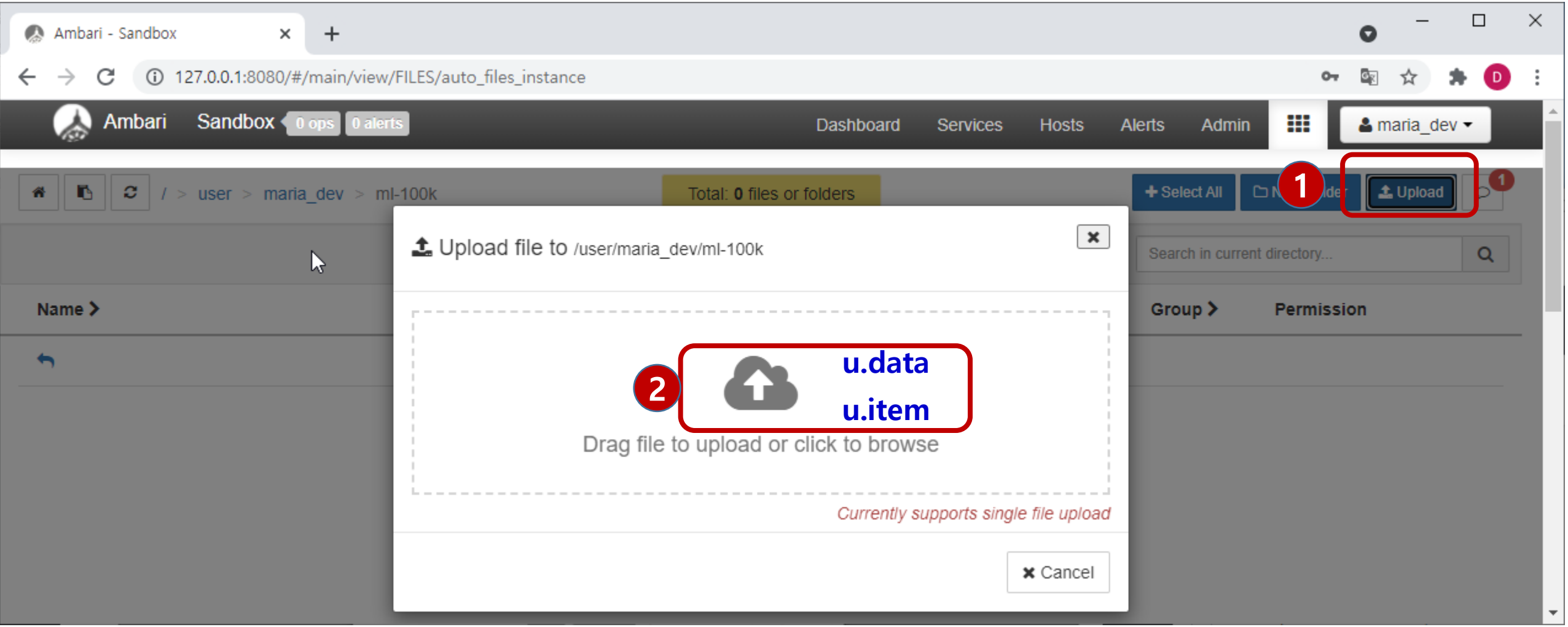

Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
hive -- 2021-06-05 20:45 hive hdfs drwxr-xr-x
livy -- 2018-06-19 00:11 livy hdfs drwxrwxr-x
maria_dev -- 2021-06-05 20:44 maria_dev hdfs drwxr-xr-x

HDFS

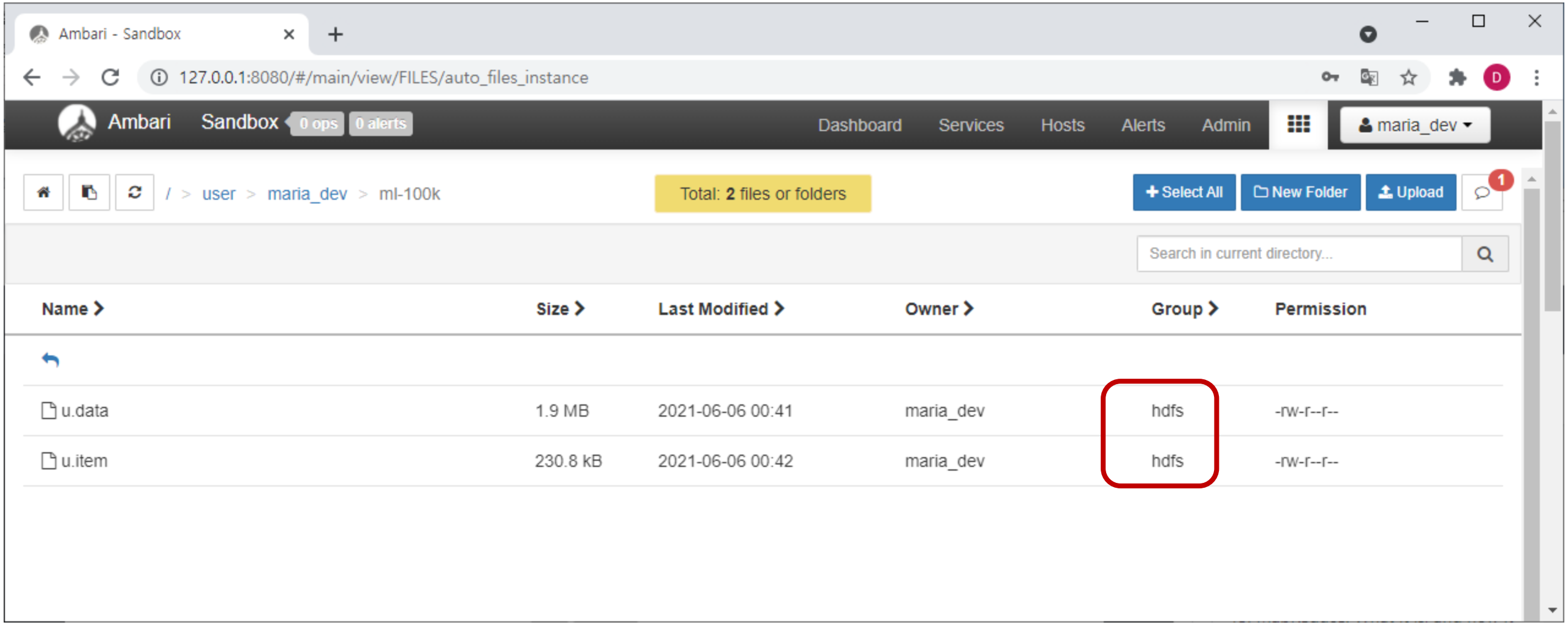

Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
/ > user > maria_dev
Total: 1 files or folders
1
New Folder
Upload
Search in current directory...
Add new folder
Name
ml-100k
ml-100k
2
3
Add
Name
Group
Permission
hive
hdfs
drwxr-xr-x


Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
hive -- 2021-06-05 20:44 maria_dev hdfs drwxr-xr-x
ml-100k -- 2021-06-06 00:38 maria_dev hdfs drwxr-xr-x
4

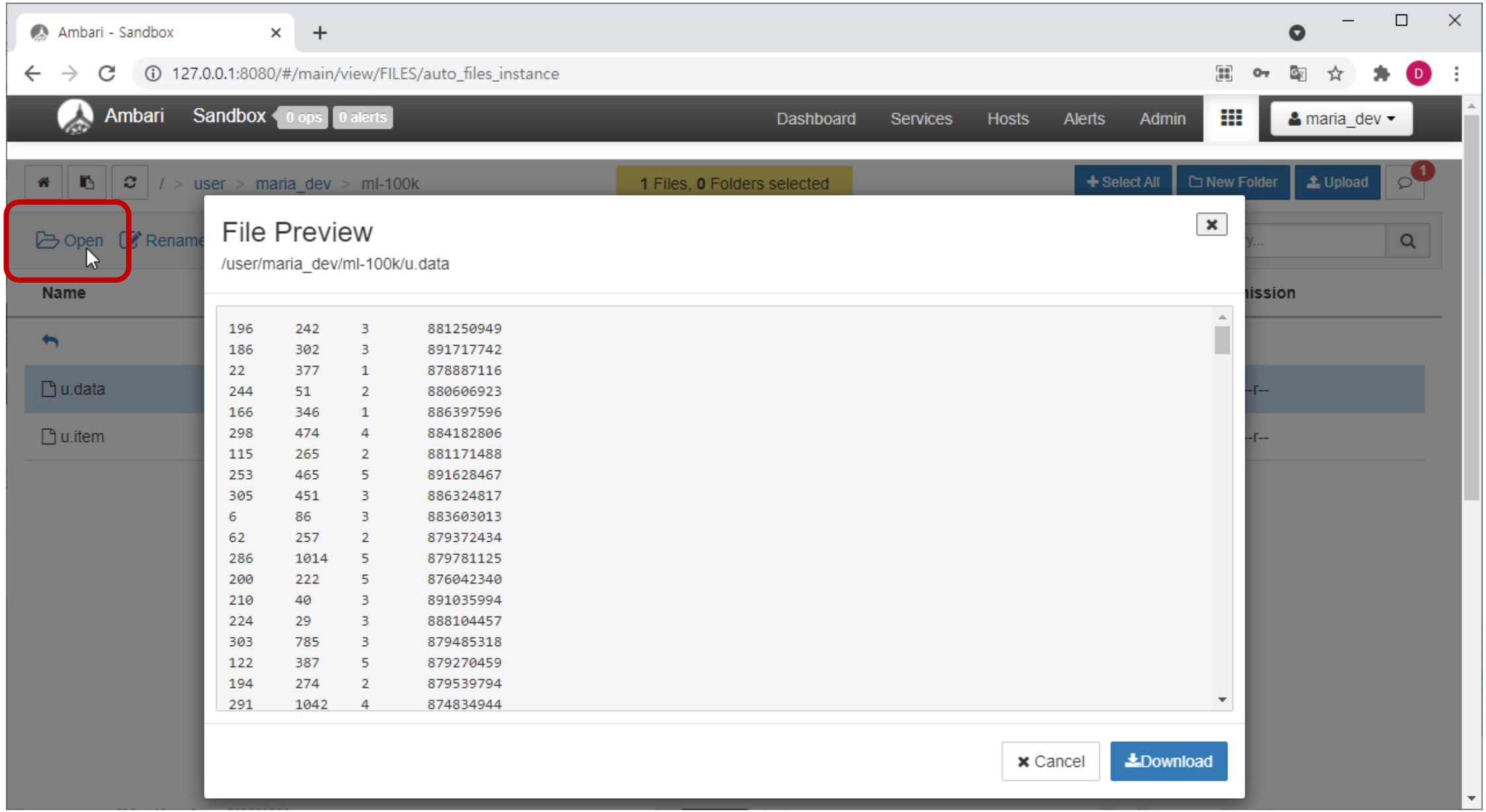
HDFS


Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
/ > user > maria_dev > ml-100k
Total: 0 files or folders
Select All
Upload
1
Search in current directory...
Name
Group
Permission
Upload file to /user/maria_dev/ml-100k
2
u.data
u.item
Drag file to upload or click to browse
Currently supports single file upload
Cancel

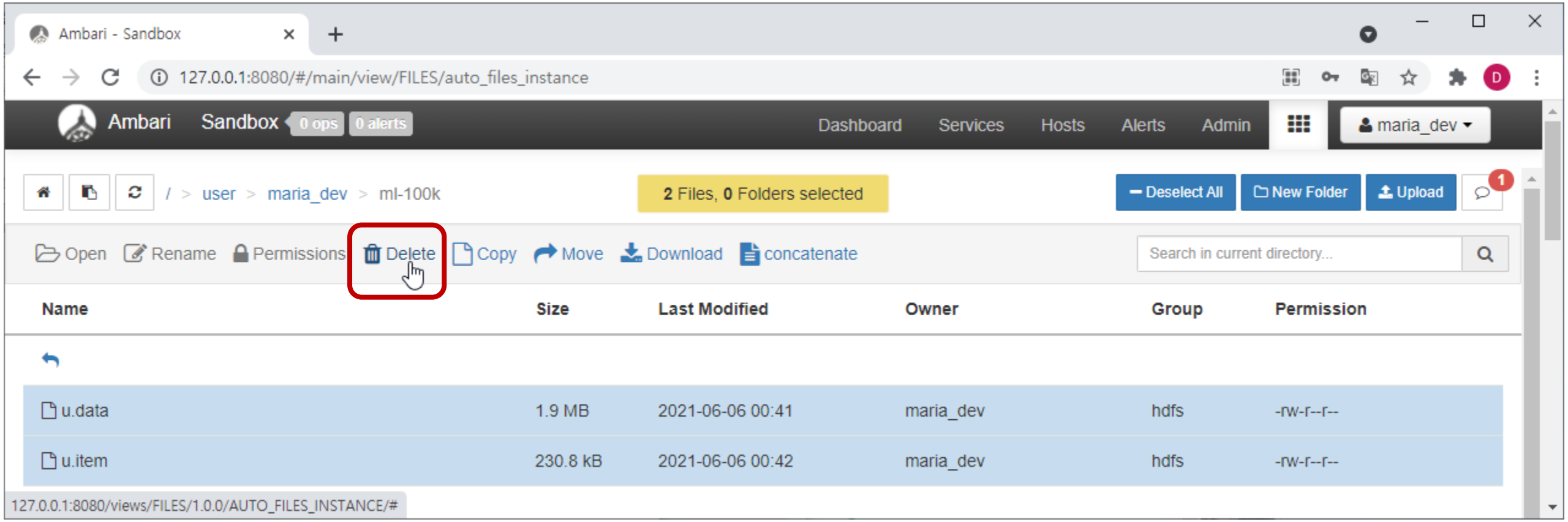
HDFS


Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
/ > user > maria_dev > ml-100k
Total: 2 files or folders
Select All
New Folder
Upload
Search in current directory...
Name
Size
Last Modified
Owner
Group
Permission
u.data
1.9 MB
2021-06-06 00:41
maria_dev
hdfs
-rw-r--r--
u.item
230.8 kB
2021-06-06 00:42
maria_dev
hdfs
-rw-r--r--

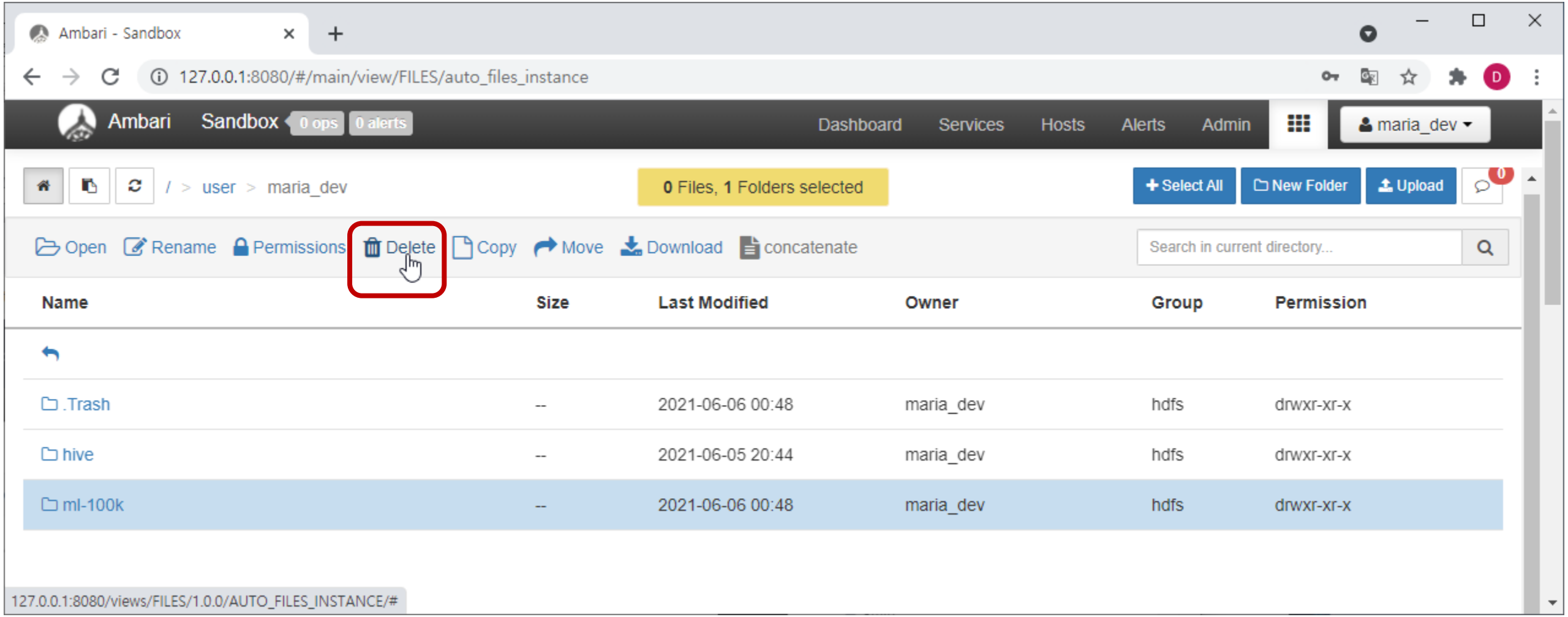
HDFS


Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
/ > user > maria_dev > ml-100k
1 Files, 0 Folders selected
Select All
New Folder
Upload
Open
Rename
Name
u.data
u.item
File Preview
/user/maria_dev/ml-100k/u.data
196 242 3 881250949
186 302 3 891717742
22 377 1 878887116
244 51 2 880606923
166 346 1 886397596
298 474 4 884182806
115 265 2 881171488
253 465 5 891628467
305 451 3 886324817
6 86 3 883603013
62 257 2 879372434
286 1014 5 879781125
200 222 5 876042340
210 40 3 891035994
224 29 3 888104457
303 785 3 879485318
122 387 5 879270459
194 274 2 879539794
291 1042 4 874834944
Cancel
Download

HDFS

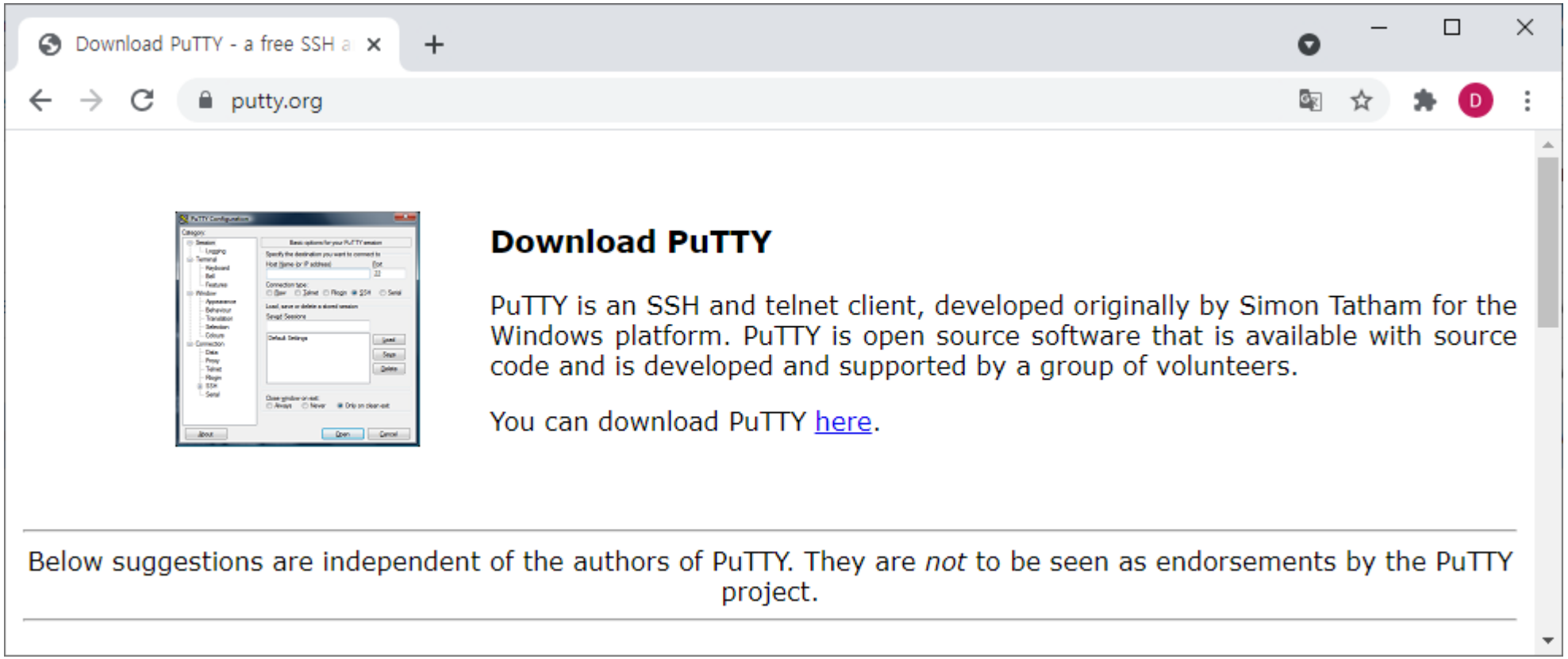

Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
/ > user > maria_dev > ml-100k
2 Files, 0 Folders selected
Deselect All
New Folder
Upload
Open
Rename
Permissions
Delete
Copy
Move
Download
concatenate
Search in current directory...
Name
Size
Last Modified
Owner
Group
Permission
u.data
1.9 MB
2021-06-06 00:41
maria_dev
hdfs
-rw-r--r--
u.item
230.8 kB
2021-06-06 00:42
maria_dev
hdfs
-rw-r--r--
127.0.0.1:8080/views/FILES/1.0.0/AUTO_FILES_INSTANCE/#

HDFS


Ambari - Sandbox
127.0.0.1:8080/#/main/view/FILES/auto_files_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
/ > user > maria_dev
0 Files, 1 Folders selected
Select All
New Folder
Upload
Open
Rename
Permissions
Delete
Copy
Move
Download
concatenate
Search in current directory...
Name
Size
Last Modified
Owner
Group
Permission
.Trash
--
2021-06-06 00:48
maria_dev
hdfs
drwxr-xr-x
hive
--
2021-06-05 20:44
maria_dev
hdfs
drwxr-xr-x
ml-100k
--
2021-06-06 00:48
maria_dev
hdfs
drwxr-xr-x
127.0.0.1:8080/views/FILES/1.0.0/AUTO_FILES_INSTANCE/#

HDFS (터미널 환경)

https://www.putty.org/

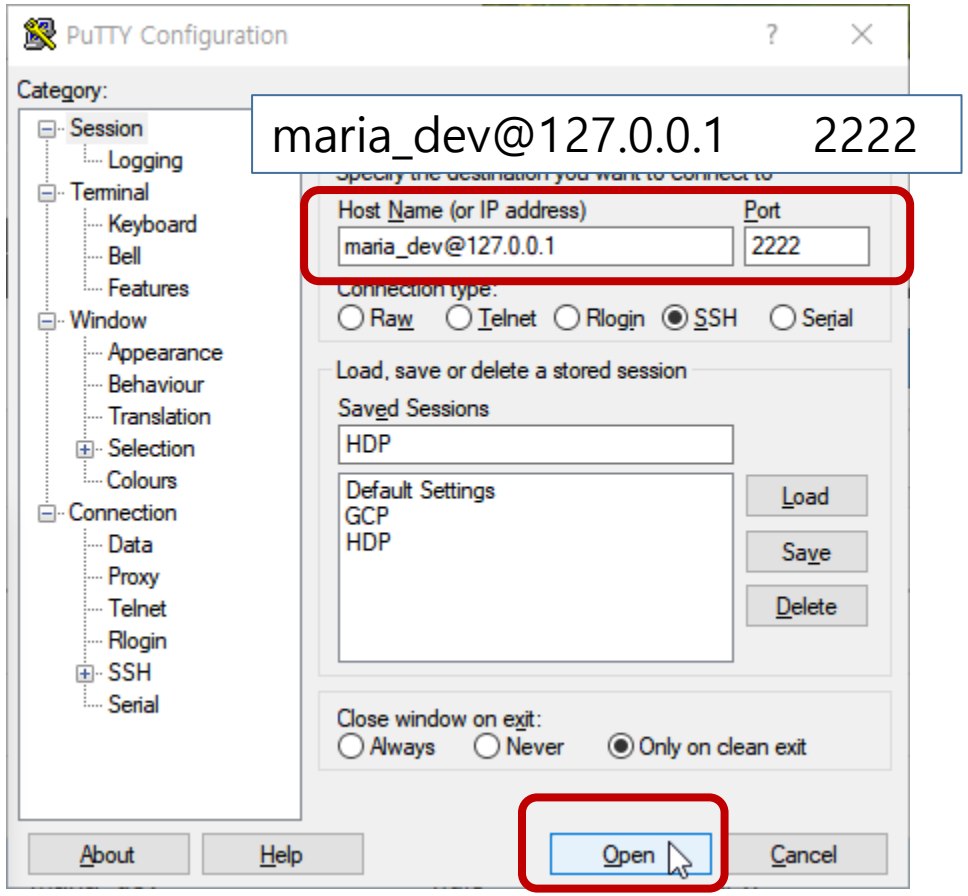


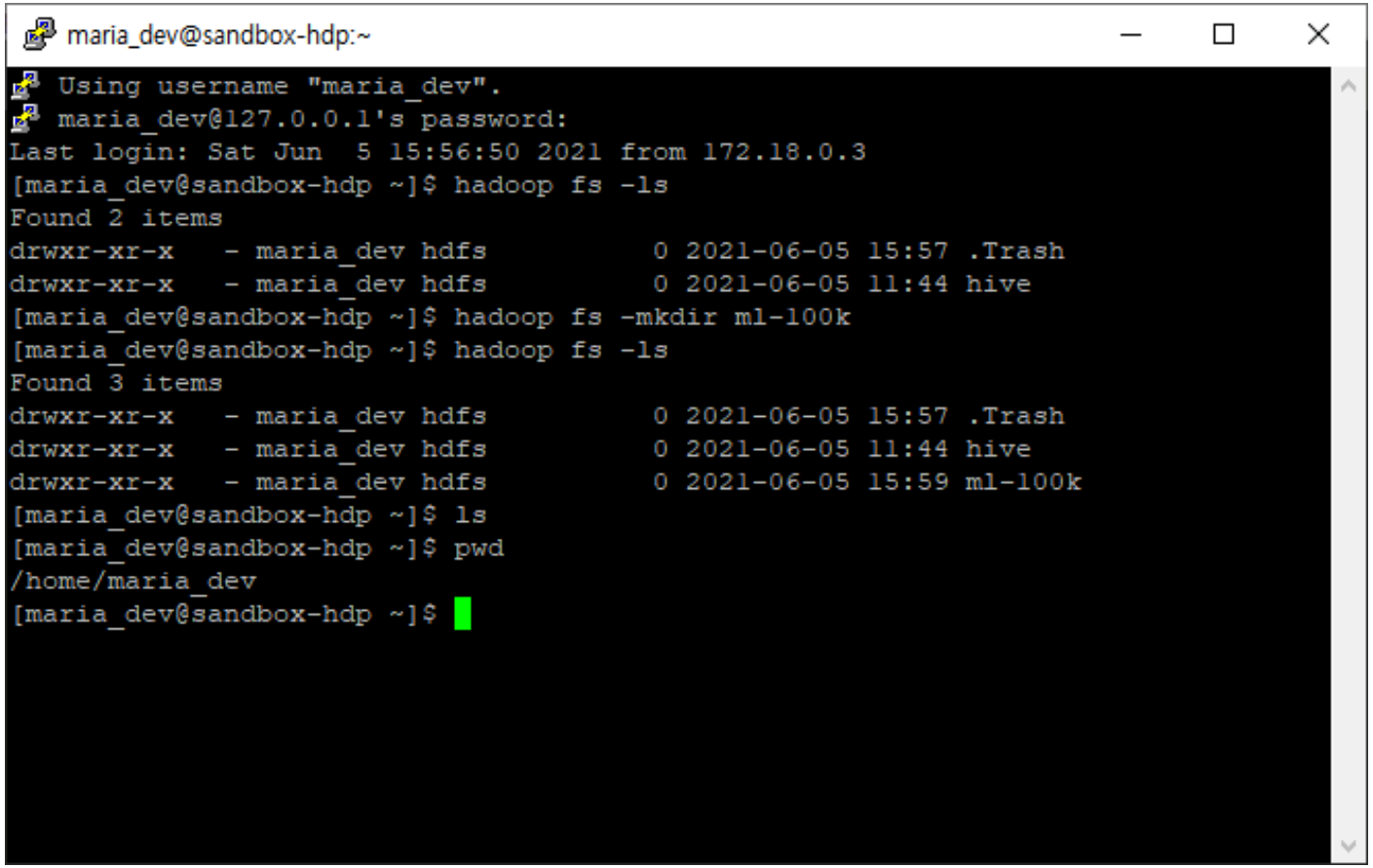
HDFS (터미널 환경)

putty 실행

hadoop fs -ls

hadoop fs -mkdir ml-100k





HDFS (터미널 환경)

wget https://github.com/kgpark88/bigdata/raw/main/ml-100k/u.data

ls

maria_dev@sandbox-hdp:~

```
[maria_dev@sandbox-hdp ~]$ wget https://github.com/kgpark88/bigdata/raw/main/ml-100k/u.data
--2021-06-05 16:12:19--  https://github.com/kgpark88/bigdata/raw/main/ml-100k/u.data
Resolving github.com (github.com)... 15.164.81.167
Connecting to github.com (github.com)|15.164.81.167|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://raw.githubusercontent.com/kgpark88/bigdata/main/ml-100k/u.data [following]
--2021-06-05 16:12:19--  https://raw.githubusercontent.com/kgpark88/bigdata/main/ml-100k/u.data
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.109.13
3, 185.199.111.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.108.133|:443... connec
ted.
HTTP request sent, awaiting response... 200 OK
Length: 1979173 (1.9M) [text/plain]
Saving to: 'u.data'

100%[===================================================>] 1,979,173   8.33MB/s   in 0.2s

2021-06-05 16:12:19 (8.33 MB/s) - 'u.data' saved [1979173/1979173]

[maria_dev@sandbox-hdp ~]$ ls
u.data
[maria_dev@sandbox-hdp ~]$
```

HDFS (터미널 환경)

hadoop fs -copyFromLocal u.data ml-100k/u.data

hadoop fs -ls ml-100k

hadoop fs -rm ml-100k/u.data

hadoop fs -rmdir ml-100k

hadoop fs -ls

maria_dev@sandbox-hdp:~

```
[maria_dev@sandbox-hdp ~]$ hadoop fs -copyFromLocal u.data ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls ml-100k
Found 1 items
-rw-r--r--   1 maria_dev hdfs    1979173 2021-06-05 16:13 ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -rm ml-100k/u.data
21/06/05 16:15:19 INFO fs.TrashPolicyDefault: Moved: 'hdfs://sandbox-hdp.hortonworks.com:8020/user
/maria_dev/ml-100k/u.data' to trash at: hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/.Tr
ash/Current/user/maria_dev/ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -rmdir ml-100k
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x   - maria_dev hdfs          0 2021-06-05 16:15 .Trash
drwxr-xr-x   - maria_dev hdfs          0 2021-06-05 11:44 hive
[maria_dev@sandbox-hdp ~]$
```

HDFS (터미널 환경)

■ HDFS Commands Guide

https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html

hadoop fs

maria_dev@sandbox-hdp:~

```
[maria_dev@sandbox-hdp ~]$ hadoop fs
Usage: hadoop fs [generic options]
        [-appendToFile <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
        [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-count [-q] [-h] [-v] [-t [<storage type>]] [-u] <path> ...]
        [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
        [-createSnapshot <snapshotDir> [<snapshotName>]]
        [-deleteSnapshot <snapshotDir> <snapshotName>]
```

MapReduce

MapReduce

Mapper는 데이터를 변환(Transform)하고, Reducer는 데이터를 집계(Aggregate) 하는 것입니다.

MapReduce

Map each input line to (rating, 1)
Redude each rating with the sum of all the 1's

USER ID	MOVIE ID	RATING	TIMESTAMP
196	242	3	881250949
186	302	3	891717742
196	377	1	878887116
244	51	2	880606923
166	346	1	886397596
186	474	4	884182806
186	265	2	881171488


```
def mapper_get_ratings(self, _, line):
    (userID, movieID, rating, timestamp) = line.split('\t')
    yield rating, 1
```

```
def reducer_count_ratings(self, key, values):
    yield key, sum(values)
```

https://youtu.be/EmHc9hV5Xi8

MapReduce

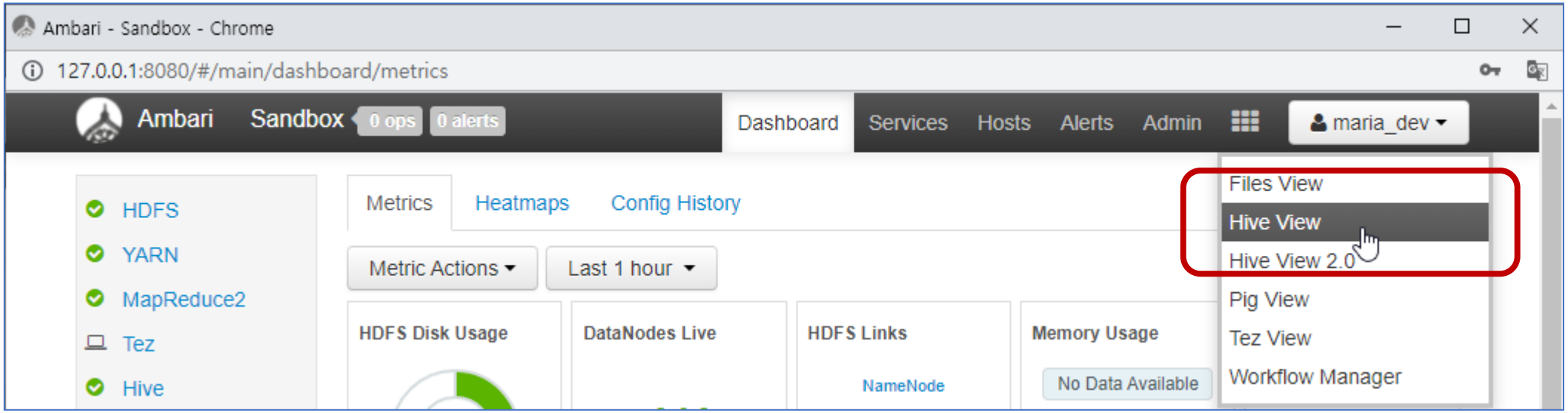
RatingsBreakdown.py

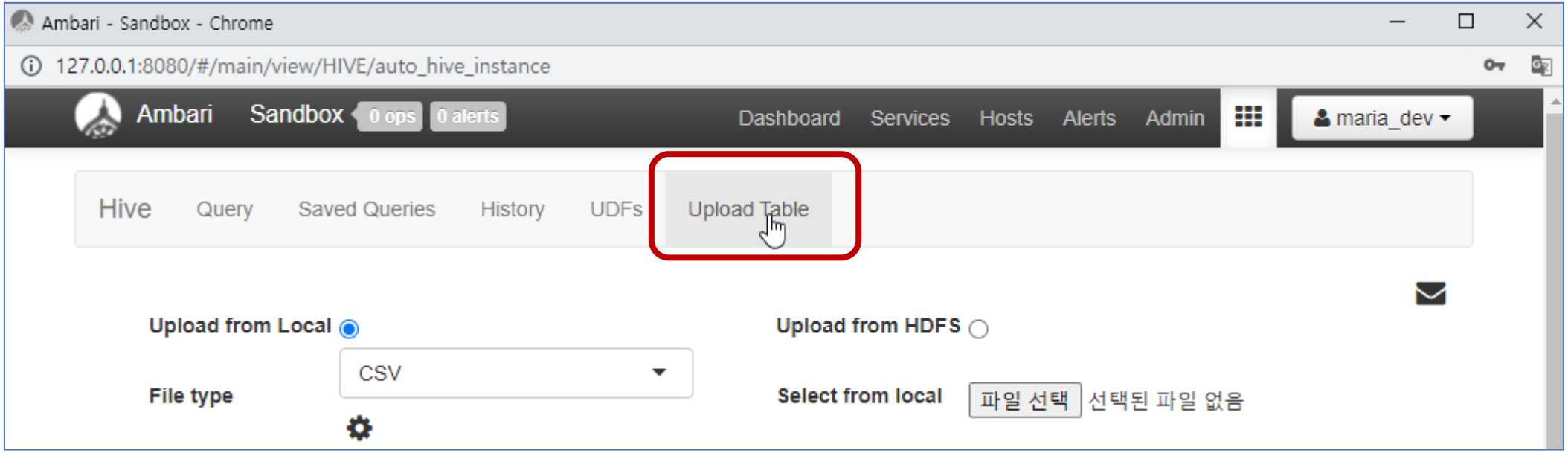
```python
from mrjob.job import MRJob
from mrjob.step import MRStep

class RatingsBreakdown(MRJob):
    def steps(self):
        return [
            MRStep(mapper=self.mapper_get_ratings,
                   reducer=self.reducer_count_ratings)
        ]

    def mapper_get_ratings(self, _, line):
        (userID, movieID, rating, timestamp) = line.split('\t')
        yield rating, 1

    def reducer_count_ratings(self, key, values):
        yield key, sum(values)

if __name__ == '__main__':
    RatingsBreakdown.run()
```

https://youtu.be/EmHc9hV5Xi8

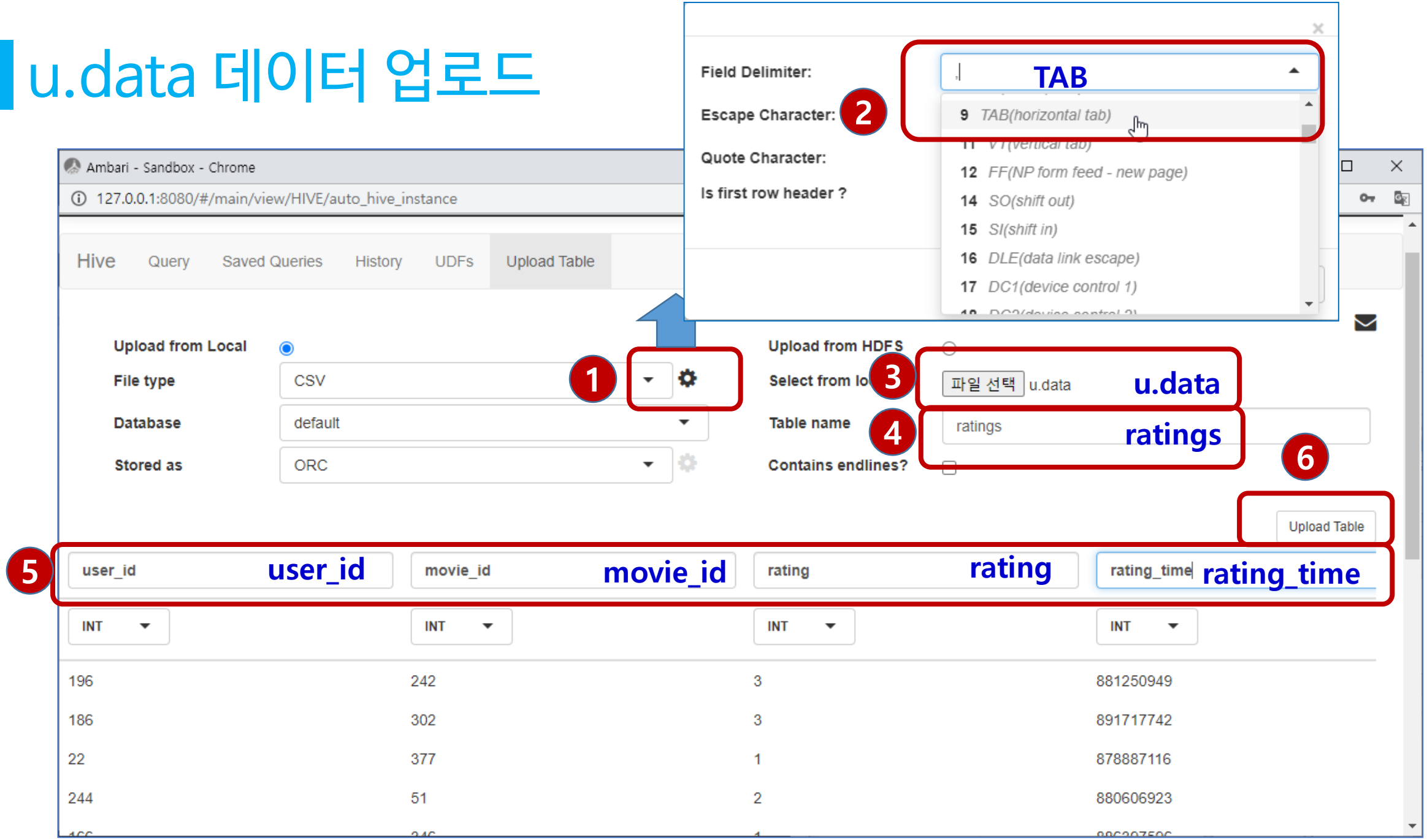
HIVE 실습

데이터 업로드

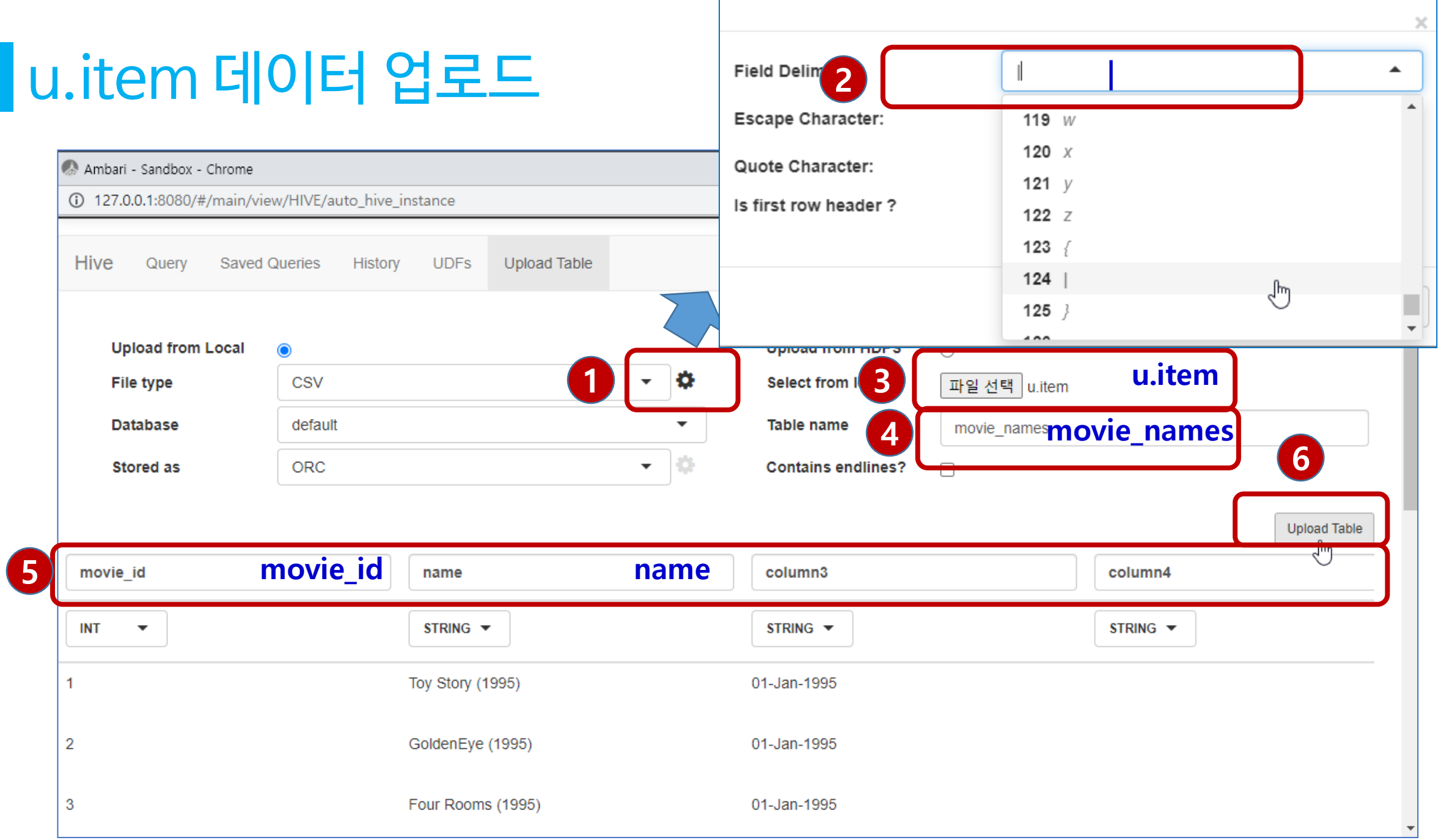

Ambari - Sandbox - Chrome
127.0.0.1:8080/#/main/dashboard/metrics
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
HDFS
YARN
MapReduce2
Tez
Hive
Metrics
Heatmaps
Config History
Metric Actions
Last 1 hour
HDFS Disk Usage
DataNodes Live
HDFS Links
NameNode
Memory Usage
No Data Available
Files View
Hive View
Hive View 2.0
Pig View
Tez View
Workflow Manager


Ambari - Sandbox - Chrome
127.0.0.1:8080/#/main/view/HIVE/auto_hive_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Hive
Query
Saved Queries
History
UDFs
Upload Table
Upload from Local
Upload from HDFS
File type
CSV
Select from local
파일 선택
선택된 파일 없음

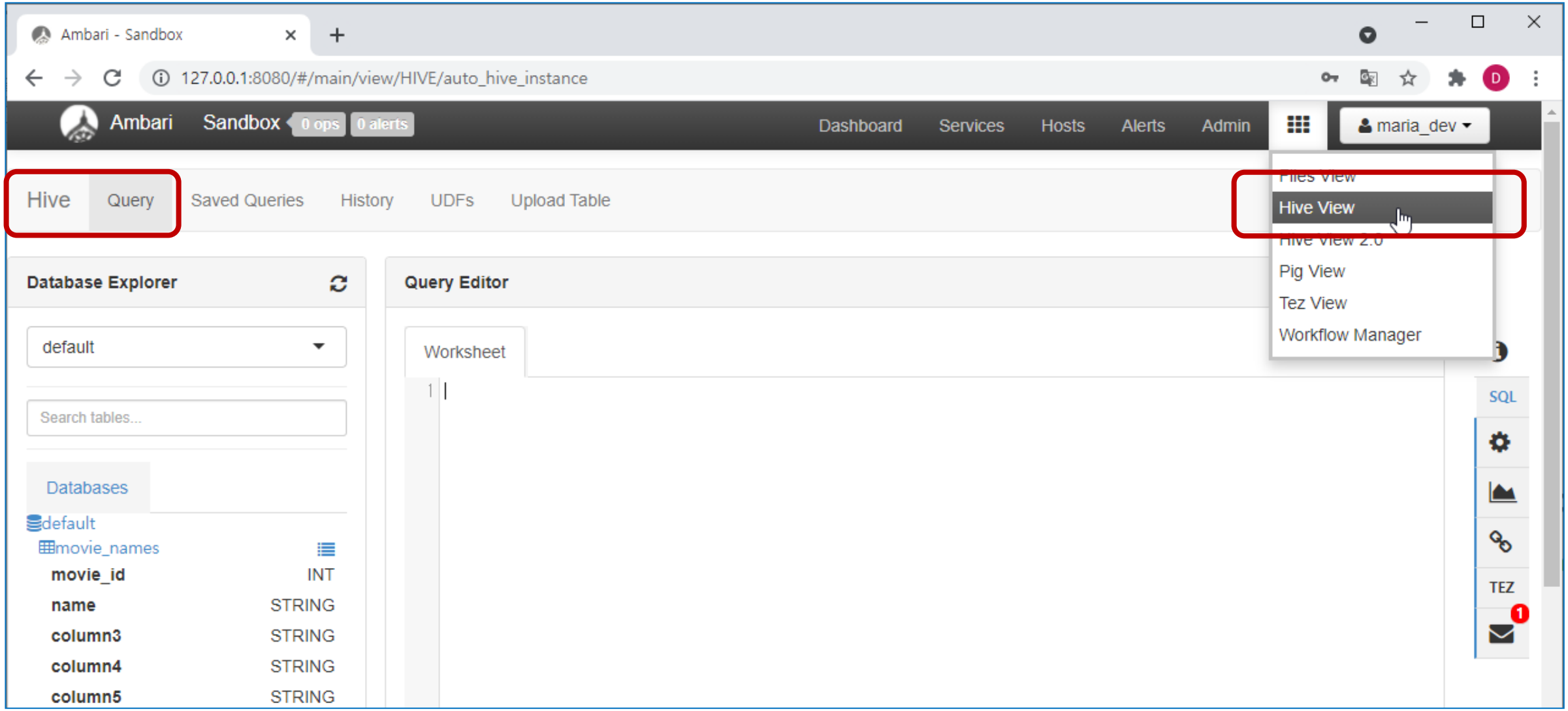
u.data 데이터 업로드



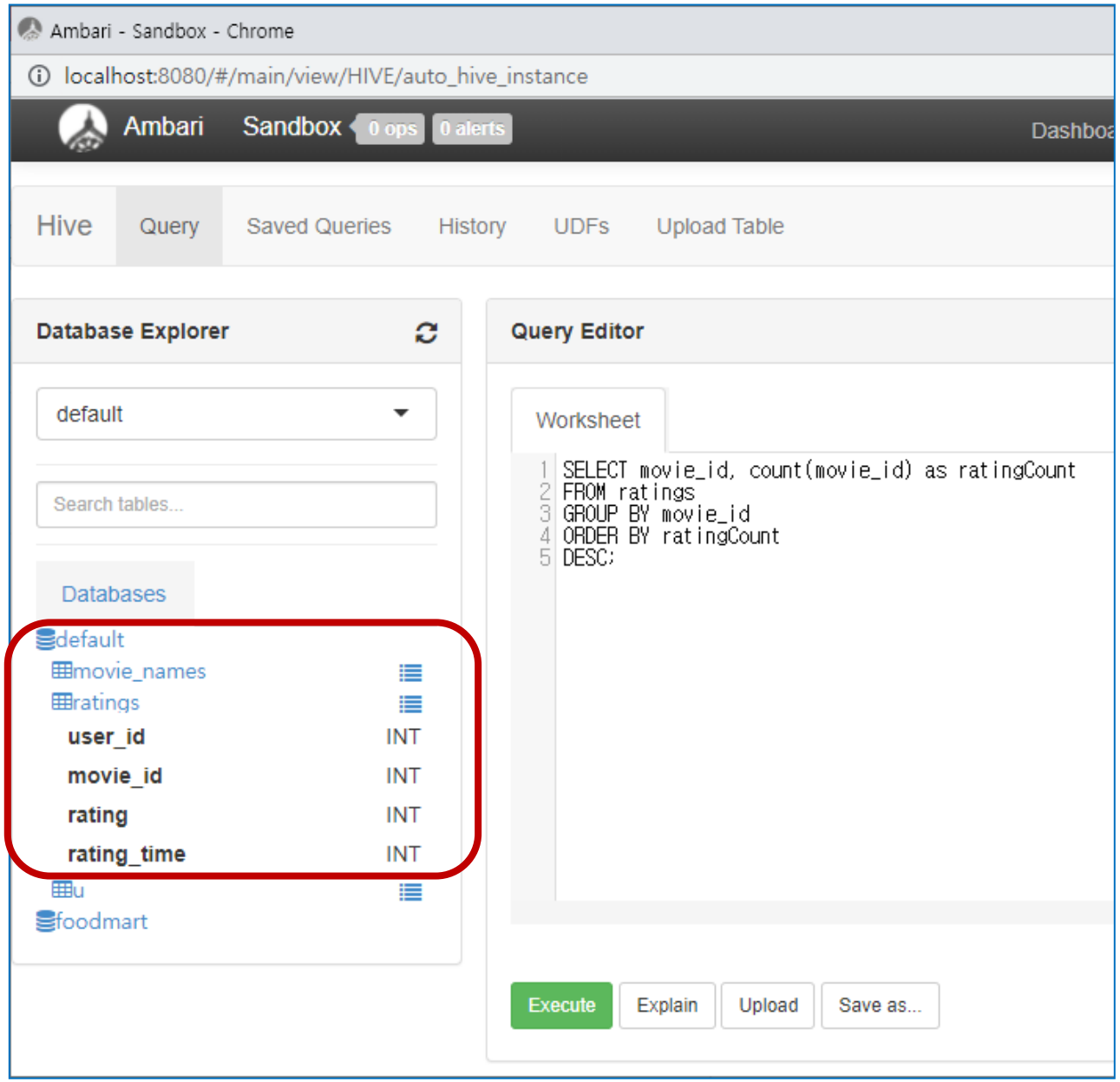
u.item 데이터 업로드



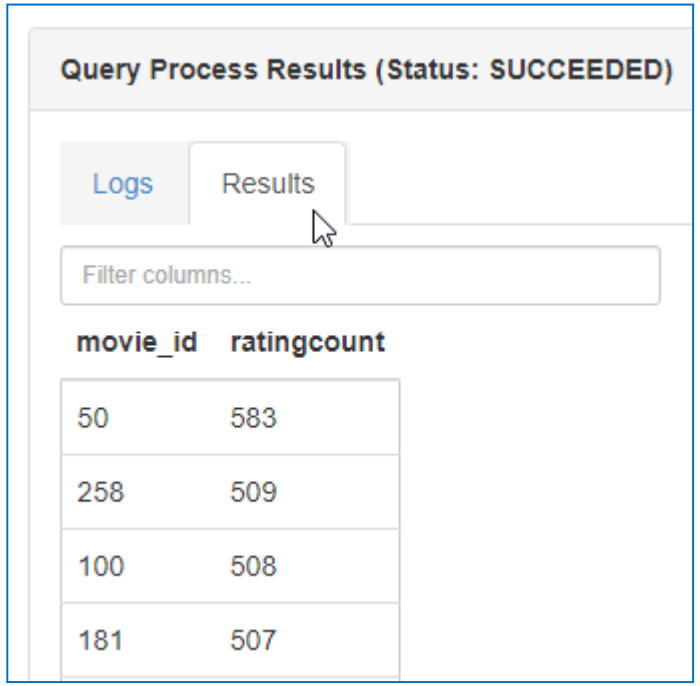
Hive


Ambari - Sandbox
127.0.0.1:8080/#/main/view/HIVE/auto_hive_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Hive
Query
Saved Queries
History
UDFs
Upload Table
Hive View
Pig View
Tez View
Workflow Manager
Database Explorer
default
Search tables...
Databases
default
movie_names
movie_id INT
name STRING
column3 STRING
column4 STRING
column5 STRING
Query Editor
Worksheet
SQL
TEZ

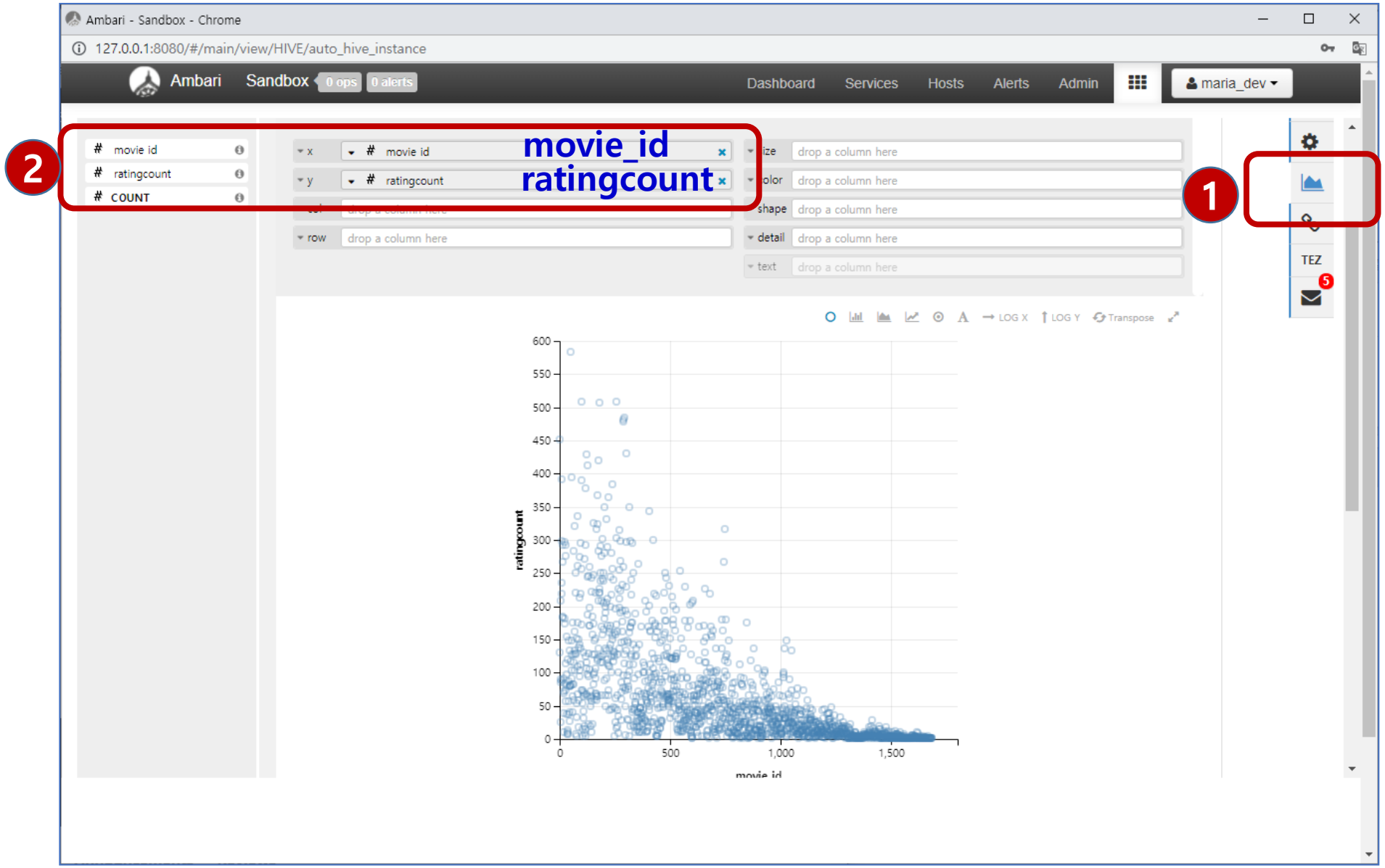
Hive




```
SELECT movie_id, count(movie_id) as ratingCount
FROM ratings
GROUP BY movie_id
ORDER BY ratingCount
DESC;
```



Hive


Ambari - Sandbox - Chrome
127.0.0.1:8080/#/main/view/HIVE/auto_hive_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
2
movie id
ratingcount
COUNT
x
movie id
y
ratingcount
row
drop a column here
movie_id
ratingcount
size
color
shape
detail
text
1
TEZ
LOG X
LOG Y
Transpose
ratingcount
movie id
600
550
500
450
400
350
300
250
200
150
100
50
0
0
500
1,000
1,500

Hive

Ambari - Sandbox - Chrome

127.0.0.1:8080/#/main/view/HIVE/auto_hive_instance

Search tables...

Databases

default

movie_names

movie_id	INT
name	STRING
column3	STRING
column4	STRING
column5	STRING
column6	INT
column7	INT
column8	INT
column9	INT
column10	INT

Load more...

ratings

foodmart

```
SELECT name
FROM movie_names
WHERE movie_id = 50;
```

SELECT name
FROM movie_names
WHERE movie_id = 50;

SQL

TEZ

6

Execute | Explain | Upload | Save as... | New Worksheet

Query Process Results (Status: SUCCEEDED)

Save results...

Logs | Results

Filter columns...

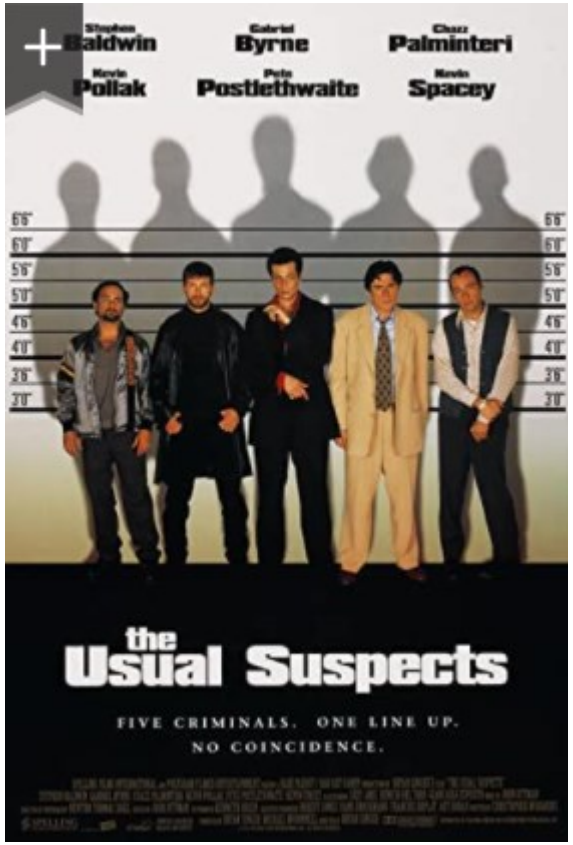
previous | next

name
Star Wars (1977)

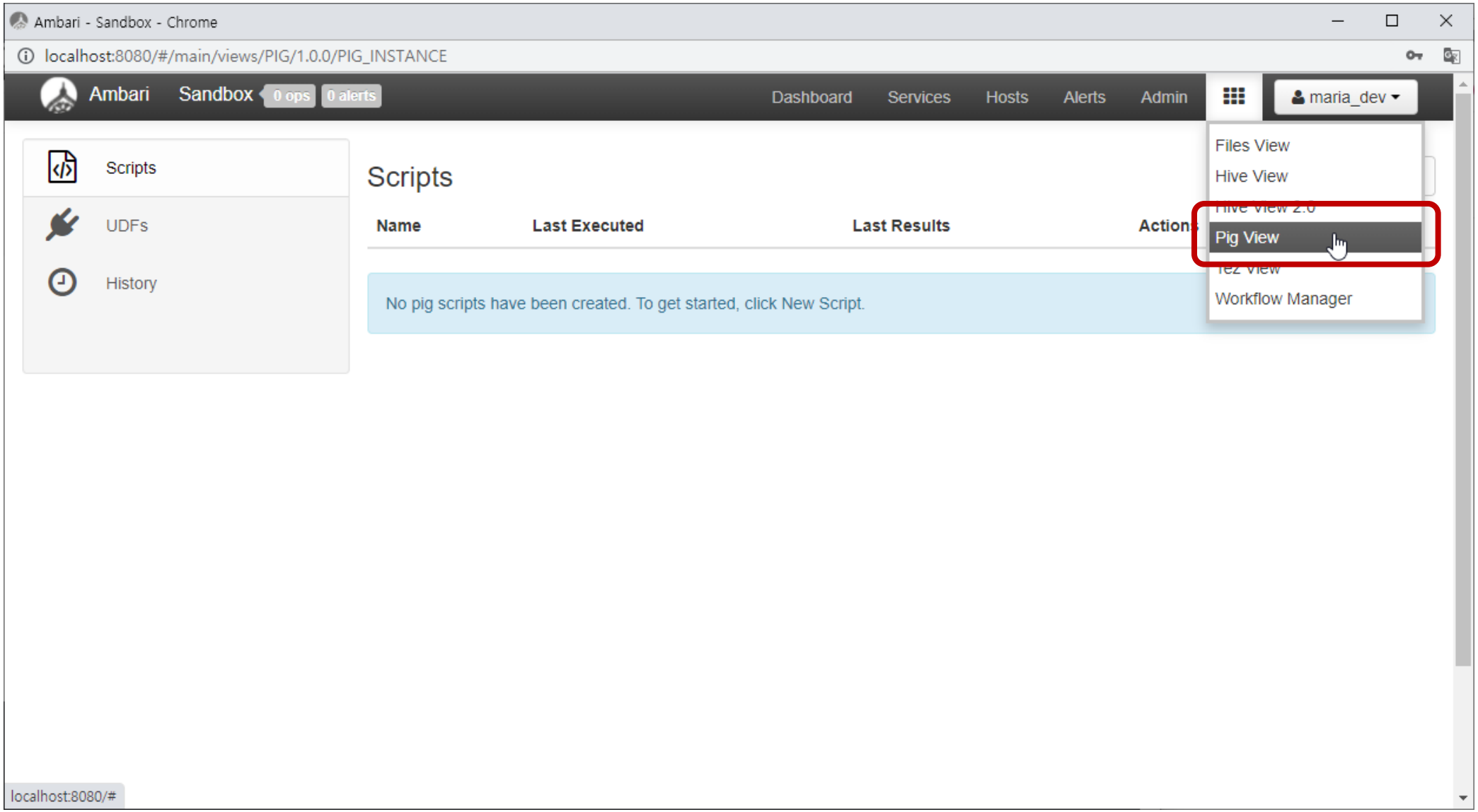
Pig 실습

Pig

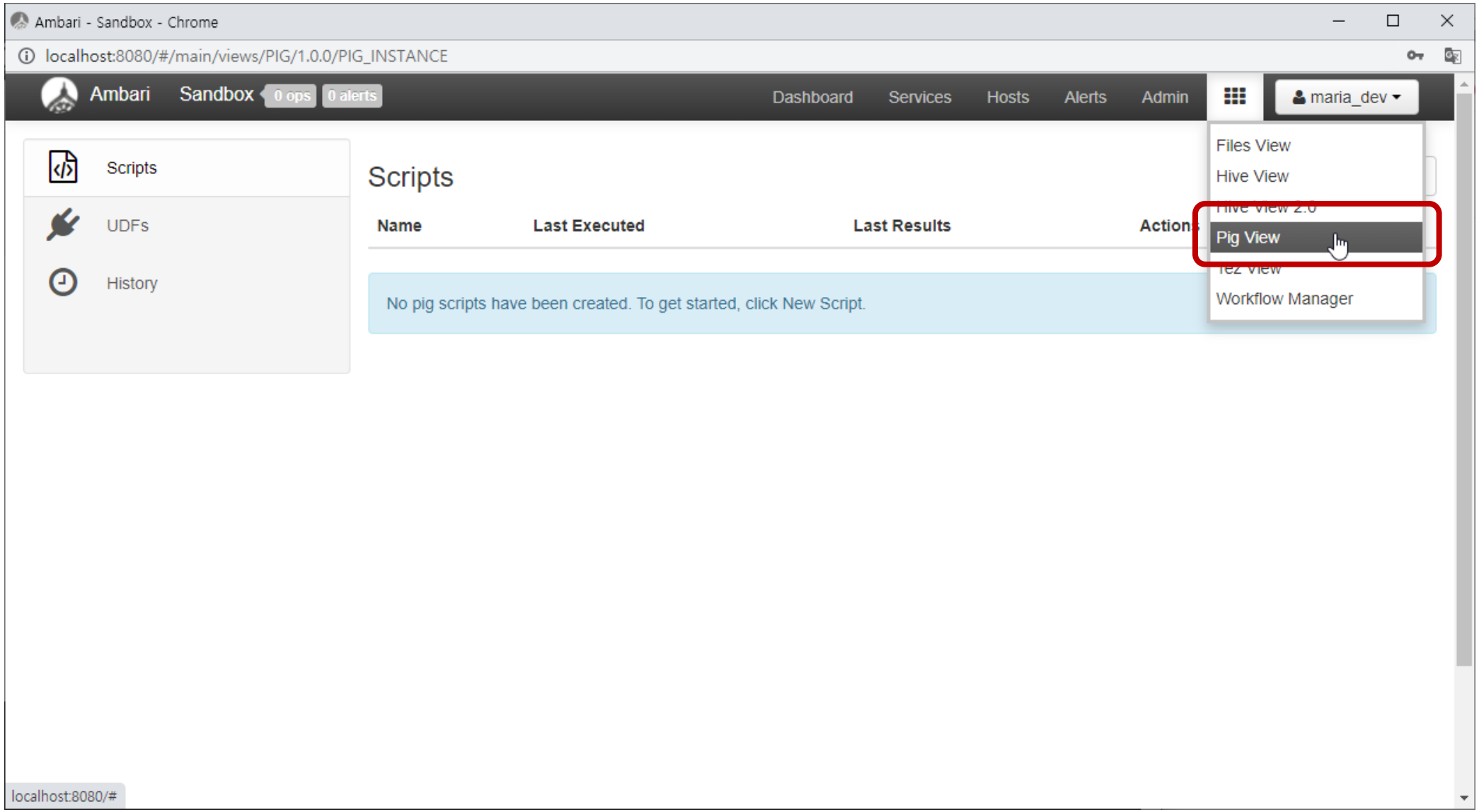
Find the oldest 5-star movies



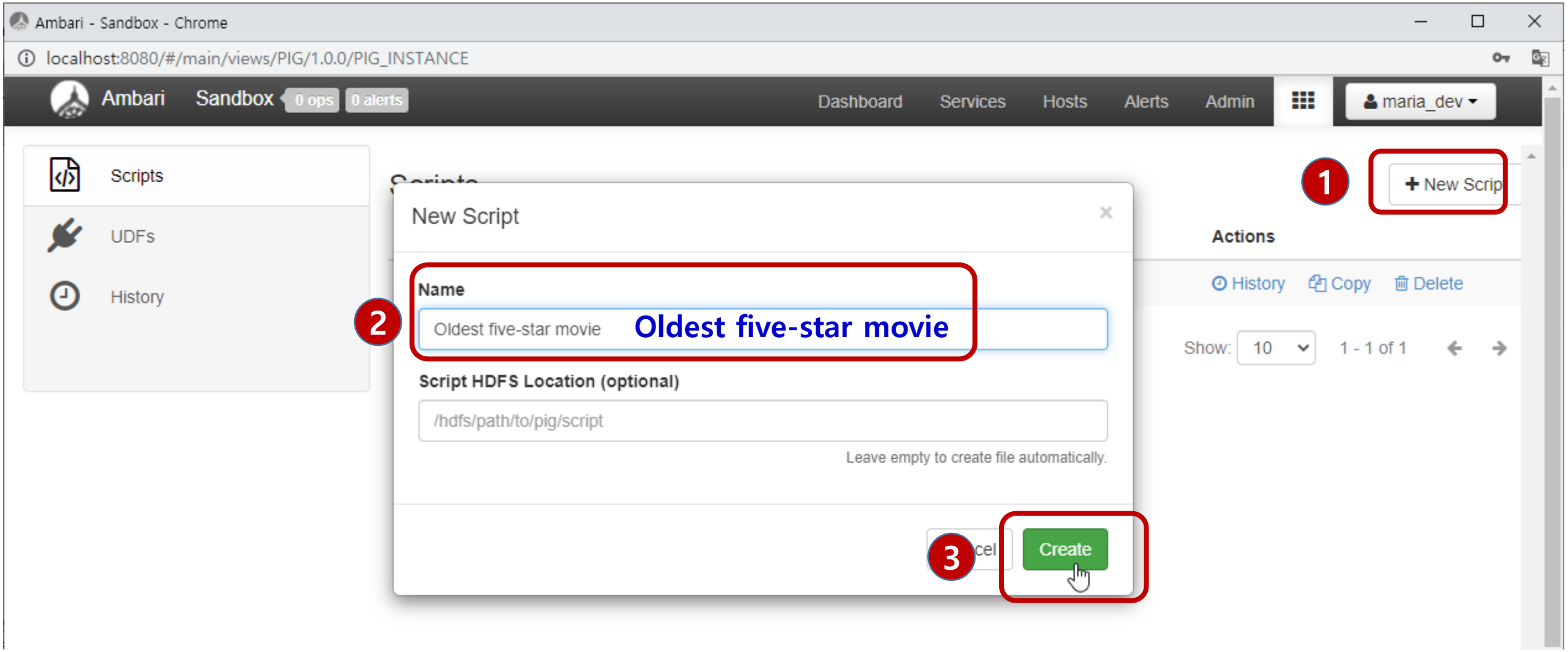
Pig


Ambari - Sandbox - Chrome
localhost:8080/#/main/views/PIG/1.0.0/PIG_INSTANCE
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Files View
Hive View
Pig View
Workflow Manager
Scripts
UDFs
History
Scripts
Name
Last Executed
Last Results
Actions
No pig scripts have been created. To get started, click New Script.
localhost:8080/#

Pig


Ambari - Sandbox - Chrome
localhost:8080/#/main/views/PIG/1.0.0/PIG_INSTANCE
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Files View
Hive View
Pig View
Workflow Manager
Scripts
UDFs
History
Scripts
Name
Last Executed
Last Results
Actions
No pig scripts have been created. To get started, click New Script.
localhost:8080/#

Pig


Ambari - Sandbox - Chrome
localhost:8080/#/main/views/PIG/1.0.0/PIG_INSTANCE
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Scripts
UDFs
History
1
+ New Script
Actions
History
Copy
Delete
Show:
10
1 - 1 of 1
New Script
2
Name
Oldest five-star movie
Oldest five-star movie
Script HDFS Location (optional)
/hdfs/path/to/pig/script
Leave empty to create file automatically.
3
Create

Pig

Ambari - Sandbox - Chrome

localhost:8080/#/main/views/PIG/1.0.0/PIG_INSTANCE

Ambari Sandbox 0 ops 0 alerts Dashboard Services Hosts Alerts Admin maria_dev

Oldest five-star movie

Save

Copy

Delete

Script History

Oldest five-star movie

Execute 2 Execute

PIG helper UDF helper /user/maria_dev/pig/scripts/oldest_fivestar_movie-2021-06-06_04-27.pig

```
ratings = LOAD '/user/maria_dev/ml-100k/u.data' AS (userID:int, movieID:int, rating:int, ratingTime:int);
metadata = LOAD '/user/maria_dev/ml-100k/u.item' USING PigStorage('|')
    AS (movieID:int, movieTitle:chararray, releaseDate:chararray, videoRealese:chararray, imdblink:chararray);

nameLookup = FOREACH metadata GENERATE movieID, movieTitle,
    ToUnixTime(ToDate(releaseDate, 'dd-MMM-yyyy')) AS releaseTime;

ratingsByMovie = GROUP ratings BY movieID;
avgRatings = FOREACH ratingsByMovie GENERATE group as movieID, AVG(ratings.rating) as avgRating;
fiveStarMovies = FILTER avgRatings BY avgRating > 4.0;
fiveStarsWithData = JOIN fiveStarMovies BY movieID, nameLookup BY movieID;
oldestFiveStarMovies = ORDER fiveStarsWithData BY nameLookup::releaseTime;
DUMP oldestFiveStarMovies;
```

1

pig.txt 내용 복사

Pig

Ambari - Sandbox - Chrome

localhost:8080/#/main/views/PIG/1.0.0/PIG_INSTANCE

Save

Copy

Delete

Oldest five-star movie - COMPLETED

Job ID job_1622947887856_0007

Started 2021-06-06 13:29

Results

Download

```
(493,4.15,493,Thin Man, The (1934),-1136073600)
(604,4.012345679012346,604,It Happened One Night (1934),-1136073600)
(615,4.0508474576271185,615,39 Steps, The (1935),-1104537600)
(1203,4.0476190476190474,1203,Top Hat (1935),-1104537600)
(613,4.037037037037037,613,My Man Godfrey (1936),-1073001600)
(633,4.057971014492754,633,Christmas Carol, A (1938),-1009843200)
(132,4.0772357723577235,132,Wizard of Oz, The (1939),-978307200)
(1122,5.0,1122,They Made Me a Criminal (1939),-978307200)
(136,4.123809523809523,136,Mr. Smith Goes to Washington (1939),-978307200)
(478,4.115384615384615,478,Philadelphia Story, The (1940),-946771200)
(524,4.021739130434782,524,Great Dictator, The (1940),-946771200)
(484,4.2101449275362315,484,Maltese Falcon, The (1941),-915148800)
(134,4.292929292929293,134,Citizen Kane (1941),-915148800)
(483,4.45679012345679,483,Casablanca (1942),-883612800)
(659,4.078260869565217,659,Arsenic and Old Lace (1944),-820540800)
```

HBase 실습

HBase 실습

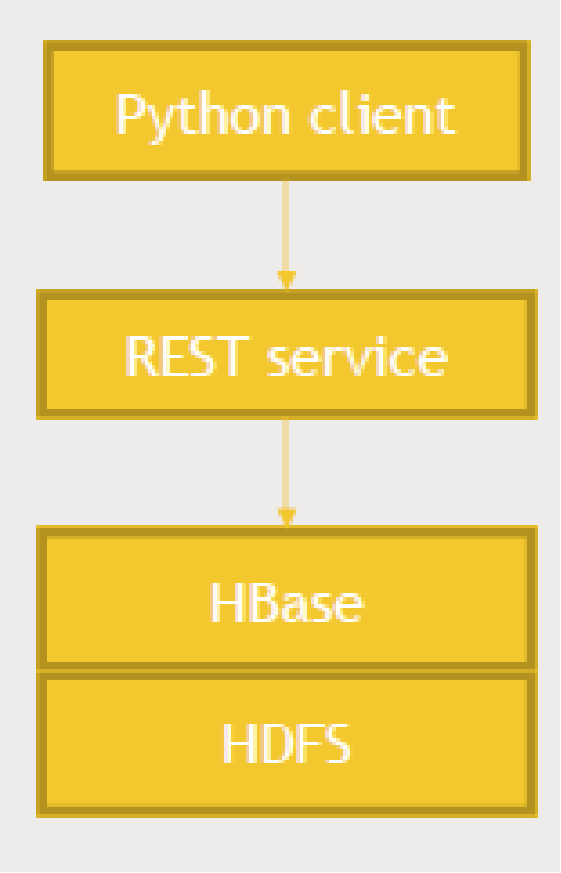
■ 실습내용

사용자별 영화 등급에 대한 Hbase 테이블 생성

사용자별 영화 등급 개수 집계

	Ratings : 50	Ratings : 33	Ratings 223
UserID	1	5	5

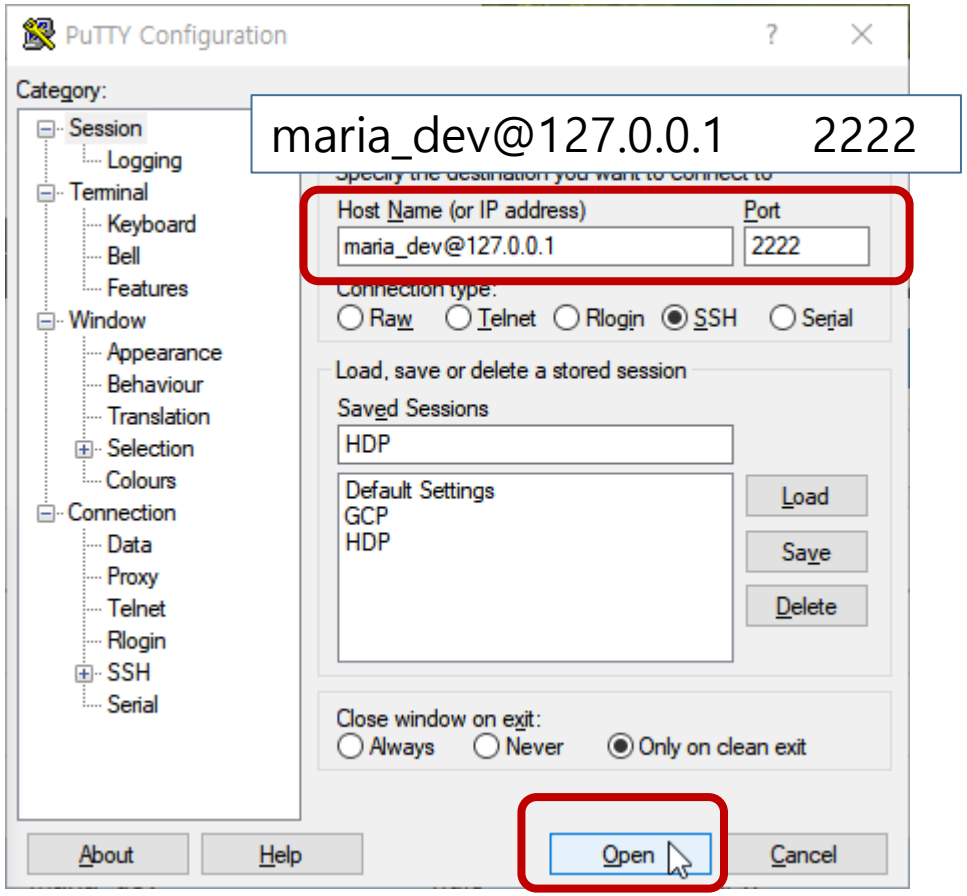
■ 구조



HBase 실습

■ HDP 접속

putty 실행



■ HBase 서버 실행

- su root #root 초기 비밀번호는 hadoop

- /usr/hdp/current/hbase-master/bin/hbase-daemon.sh start rest -p 8000

HBase 실습

■ starbase 패키지를 설치

pip install starbase

■ HBaseExamples.py

```python
from starbase import Connection

c = Connection("127.0.0.1", "8000")

ratings = c.table('ratings')

if (ratings.exists()):
    print("Dropping existing ratings table\n")
    ratings.drop()

ratings.create('rating')

print("Parsing the ml-100k ratings data...\n")
ratingFile = open("u.data", "r")

batch = ratings.batch()
print(batch)

for line in ratingFile:
    (userID, movieID, rating, timestamp) = line.split()
    batch.update(userID, {'rating': {movieID: rating}})

ratingFile.close()

print ("Committing ratings data to HBase via REST service\n")
batch.commit(finalize=True)

print ("Get back ratings for some users...\n")
print ("Ratings for user ID 1:\n")
print (ratings.fetch("1"))
print ("Ratings for user ID 33:\n")
print (ratings.fetch("33"))

ratings.drop()
```

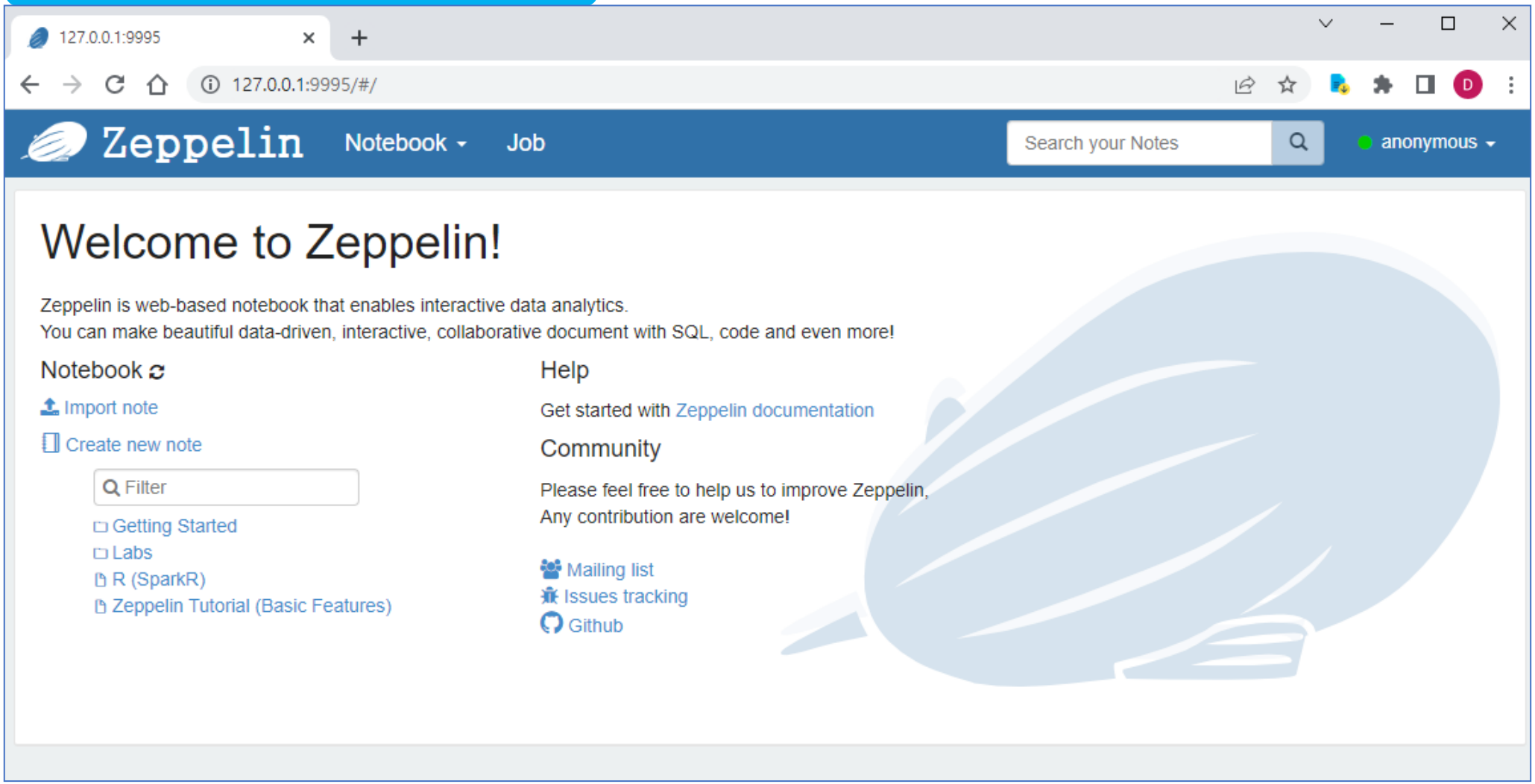
■ HBaseExamples 실행

Python HBaseExamples.py

Zeppelin 실습

Zeppelin

http://127.0.0.1:9995/



Zeppelin

```
sc.version

%sh

wget http://media.sundog-soft.com/hadoop/ml-100k/u.data -O /tmp/u.data
wget http://media.sundog-soft.com/hadoop/ml-100k/u.item -O /tmp/u.item

%sh

hadoop fs –rm –r –f /tmp/m1-100k

hadoop fs mkdir /tmp/m1-100k

hadoop fs –put /tmp/u.data /tmp/m1-100k

hadoop fs –put /tmp/u.item item /tmp/m1-100k
```

Zeppelin

127.0.0.1:9995/#/notebook/2J8MM35T5

Zeppelin Notebook Job Search your Notes anonymous

zeppelin_demo Head default

```
sc.version
```

READY

```
%sh

wget https://raw.githubusercontent.com/kgpark88/bigdata/main/ml-100k/u.data -O /tmp/u.data
wget https://raw.githubusercontent.com/kgpark88/bigdata/main/ml-100k/u.item -O /tmp/u.item
```

READY

```
%sh

hadoop fs –rm –r –f /tmp/m1-100k
hadoop fs mkdir /tmp/m1-100k
hadoop fs –put /tmp/u.data /tmp/m1-100k
hadoop fs –put /tmp/u.item item /tmp/m1-100k
```

READY

Zeppelin

127.0.0.1:9995/#/notebook/2A94M5J1Z

Zeppelin Notebook Job

Search your Notes

anonymous

Zeppelin Tutorial (Basic Features)

Head

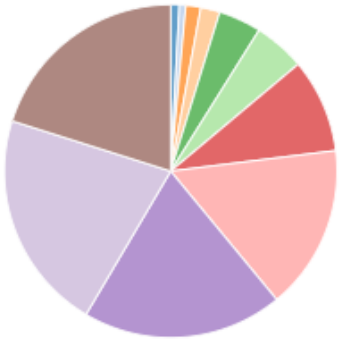
default

```
%spark.sql
select age, count(1) value
from bank
where age < 30
group by age
order by age
```

FINISHED

settings

19 20 21 22 23 24 25 26 27 28 29



Took 51 sec. Last updated by anonymous at December 18 2016, 12:31:04 AM. (outdated)

```
%spark.sql
select age, count(1) value
from bank
where age < ${maxAge=30}
group by age
order by age
```

FINISHED

maxAge

35

age	value
19	4
20	3
21	7
22	9
23	20
24	24
25	44
26	77
27	94

Took 9 sec. Last updated by anonymous at December 18 2016, 12:31:07 AM. (outdated)

```
%spark.sql
select age, count(1) value
from bank
where marital="${marital=single,single|divorced|married}"
group by age
order by age
```

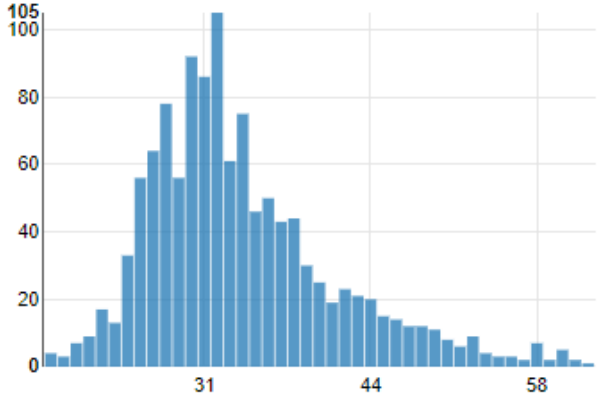
FINISHED

marital

single

settings

Grouped Stacked value



Took 4 sec. Last updated by anonymous at December 18 2016, 12:31:09 AM. (outdated)

Thank you