

# Demo

## Hbase

1. 输入 `start-hbase.sh` 命令并输入 `yes` 启动Hbase

```
root@usert-cluster-master:~# start-hbase.sh
The authenticity of host 'usert-cluster-master (10.244.5.150)' can't be established.
ECDSA key fingerprint is 2b:88:aa:6c:bb:e2:4f:fc:40:5f:4c:b8:a2:e5:56:73.
Are you sure you want to continue connecting (yes/no)? yes
usert-cluster-master: Warning: Permanently added 'usert-cluster-master,10.244.5.150' (ECDSA) to the list of known hosts.
usert-cluster-master: running zookeeper, logging to /usr/local/hbase-2.2.2/bin/../logs/hbase-root-zookeeper-usert-cluster-master.out
running master, logging to /usr/local/hbase-2.2.2/logs/hbase--master-usert-cluster-master.out
: running regionserver, logging to /usr/local/hbase-2.2.2/logs/hbase--regionserver-usert-cluster-master.out
root@usert-cluster-master:~#
```

2. 输入 `hbase shell` 命令进入Hbase命令行界面

```
root@usert-cluster-master:~# hbase shell
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.2.2, re6513a76c91cceda95dad7af246ac81d46fa2589, Sat Oct 19 10:10:12 UTC 2019
Took 0.0027 seconds
hbase(main):001:0>
```

3. 输入如下命令创建表

```
1 | create 'students','ID','Description','Course','Home'
```

```
hbase(main):001:0> create 'students','ID','Description','Course','Home'
Created table students
Took 2.9393 seconds
=> Hbase::Table - students
hbase(main):002:0>
```

4. 输入如下命令插入数据

```
1 | put 'students','001','Description:Name','Li Lei'
2 | put 'students','001','Description:Height','176'
3 | put 'students','001','Course:Chinese','80'
```

```
hbase(main):002:0> put 'students','001','Description:Name','Li Lei'
Took 0.2325 seconds
hbase(main):003:0> put 'students','001','Description:Height','176'
Took 0.0069 seconds
hbase(main):005:0> put 'students','001','Course:Chinese','80'
Took 0.0103 seconds
```

5. 输入如下命令查看描述表信息

```
1 | describe 'students'
```

```

hbase(main):007:0> describe 'students'
Table students is ENABLED
students
COLUMN FAMILIES DESCRIPTION
{NAME => 'Course', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}

{NAME => 'Description', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}

{NAME => 'Home', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}

{NAME => 'ID', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CACHE_DATA_ON_WRITE => 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0', REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', CACHE_INDEX_ON_WRITE => 'false', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPEN => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}

4 row(s)

QUOTAS
0 row(s)
Took 0.1460 seconds
hbase(main):008:0>

```

## hive

1. 选择带有hive的镜像创建集群后进入terminal，输入 `hive` 进入Hive的命令行界面，输入 `show tables;` 显示OK表明Hive正常启动

```

root@usert-cluster-master:~# hive
Hive Session ID = 4c700ba9-8dee-488c-90ca-28be6aa46d24

Logging initialized using configuration in jar:file:/usr/local/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
2020-04-09 12:23:02,249 INFO [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.Persistences: Property datanucleus.cache.level2 unknown - will be ignored
2020-04-09 12:23:08,915 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:08,916 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:08,917 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:08,917 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:08,918 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:08,918 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:09,989 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:09,989 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:09,989 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:09,990 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:09,990 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
2020-04-09 12:23:09,990 WARN [4c700ba9-8dee-488c-90ca-28be6aa46d24 main] DataNucleus.MetaData: Metadata has jdbc-type of null yet this is not valid. Ignored
Hive Session ID = 341c8a90-1b62-4cd9-b19f-f4e459c64f02
Hive on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show tables;
OK
Time taken: 0.6 seconds
hive>

```

2. 输入如下命令创建表

```

1 create table logs(
2     user_id int,item_id int,cat_id int,mer_id int,brand_id int,month
int,day int,action int,age_range int,gender int,province string)
3     row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
4     WITH SERDEPROPERTIES( "separatorChar"="," ) STORED AS TEXTFILE;

```

```

hive> create table logs(
ay int,action int,age_range int,gender int,province string)nd_id int,month int,d
>     row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
>     WITH SERDEPROPERTIES( "separatorChar"="," ) STORED AS TEXTFILE;
OK

```

3. 输入如下命令从csv文件导入到创建的表中

```

1 load data local inpath '/workspace/dataset1/million_user_log.csv' overwrite
into table logs;

```

```

e into table logs; al inpath '/workspace/dataset1/million_user_log.csv' overwrit
Loading data to table default.logs
OK
Time taken: 0.837 seconds
hive>

```

4. 输入如下命令测试导入成功:

```

1 select count(*) from logs;

```

```

hive> select count(*) from logs;
Query ID = root_20200409130653_003b9574-dda6-4569-b240-d80acc392fcd
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
2020-04-09 13:06:53,758 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] client.RMProxy: Connecting to ResourceManager at usert-cluster-master/10.244.5.150:8032
2020-04-09 13:06:53,811 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] client.RMProxy: Connecting to ResourceManager at usert-cluster-master/10.244.5.150:8032
2020-04-09 13:06:53,824 WARN [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execu
te your application with ToolRunner to remedy this.
2020-04-09 13:06:54,269 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] input.FileInputFormat: Total input paths to process : 1
2020-04-09 13:06:54,270 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] input.CombineFileInputFormat: DEBUG: Terminated node allocation with : CompletedNodes: 2, size left: 0
2020-04-09 13:06:54,370 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.JobSubmitter: number of splits:1
2020-04-09 13:06:54,428 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.JobSubmitter: Submitting tokens for job: job_1586436266729_0002
2020-04-09 13:06:54,450 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] impl.YarnClientImpl: Submitted application application_1586436266729_0002
2020-04-09 13:06:54,454 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.Job: The url to track the job: http://usert-cluster-master:8088/proxy/application_1586436266729_0002/
Starting Job = job_1586436266729_0002, Tracking URL = http://usert-cluster-master:8088/proxy/application_1586436266729_0002/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1586436266729_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-04-09 13:07:04,793 WARN [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.Counters: Group org.apache.hadoop.mapred.Task$Counter is deprecated. Use org.apache.hadoop.mapreduce.TaskCoun
ter instead
2020-04-09 13:07:04,793 Stage-1 map = 0%, reduce = 0%
2020-04-09 13:07:27,906 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 32.8 sec
2020-04-09 13:07:36,295 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 36.97 sec
MapReduce Total cumulative CPU time: 36 seconds 970 msec
Ended Job = job_1586436266729_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 36.97 sec HDFS Read: 47319613 HDFS Write: 107 SUCCESS
Total MapReduce CPU Time Spent: 36 seconds 970 msec
OK
2020-04-09 13:07:37,472 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] Configuration.deprecation: mapred.input.dir is deprecated. Instead, use mapreduce.input.fileinputformat.inputdir
2020-04-09 13:07:37,477 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapred.FileInputFormat: Total input paths to process : 1
1000000
Time taken: 44.157 seconds, Fetched: 1 row(s)

```

执行结果为1000000条

5. 查询记录: action为2的user数量, 输入命令如下:

```

1 select count(distinct user_id) from logs where action=2;

```

```
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 41.9 sec HDFS Read: 47311600 HDFS Write: 0 FAIL
Total MapReduce CPU Time Spent: 41 seconds 900 msec
hive> select count(distinct user_id) from logs where action=2;
Query ID = root_20200409131159_b7ea7bf5-a9db-4afl-9299-3c425b1783eb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
2020-04-09 13:11:59,467 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] client.RMProxy: Connecting to ResourceManager at usert-cluster-master/10.244.5.150:8032
2020-04-09 13:11:59,498 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] client.RMProxy: Connecting to ResourceManager at usert-cluster-master/10.244.5.150:8032
2020-04-09 13:11:59,504 WARN [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2020-04-09 13:12:00,008 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] input.FileInputFormat: Total input paths to process : 1
2020-04-09 13:12:00,008 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] input.CombineFileInputFormat: DEBUG: Terminated node allocation with : CompletedNodes: 2, size left: 0
2020-04-09 13:12:00,113 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.JobSubmitter: number of splits:1
2020-04-09 13:12:00,172 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.JobSubmitter: Submitting tokens for job: job_1586436266729_0005
2020-04-09 13:12:00,189 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] impl.YarnClientImpl: Submitted application application_1586436266729_0005
2020-04-09 13:12:00,192 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.Job: The url to track the job: http://usert-cluster-master:8088/proxy/application_1586436266729_0005/
Starting Job = job_1586436266729_0005, Tracking URL = http://usert-cluster-master:8088/proxy/application_1586436266729_0005/
Kill Command = /usr/local/hadoop/bin/kill job -kill job_1586436266729_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2020-04-09 13:12:08,481 WARN [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapreduce.Counters: Group org.apache.hadoop.mapred.Task$Counter is deprecated. Use org.apache.hadoop.mapreduce.TaskCounter instead
2020-04-09 13:12:08,480 Stage-1 map = 0%, reduce = 0%
2020-04-09 13:12:27,321 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 30.61 sec
2020-04-09 13:12:32,570 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 35.35 sec
MapReduce Total cumulative CPU time: 35 seconds 350 msec
Ended Job = job_1586436266729_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 35.35 sec HDFS Read: 47316456 HDFS Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 35 seconds 350 msec
OK
2020-04-09 13:12:34,698 INFO [60e35147-1c76-4d8e-83ce-9be99800b4a8 main] mapped.FileInputFormat: Total input paths to process : 1
37202
Time taken: 35.559 seconds, Fetched: 1 row(s)
```