1、 HBase 创建一张表 Wuxia

```
hbase(main):007:0*
hbase(main):008:0* create 'Wuxia','count'
0 row(s) in 2.3940 seconds

=> Hbase::Table - Wuxia
hbase(main):009:0>
```

2、 scan ‘Wuxia’

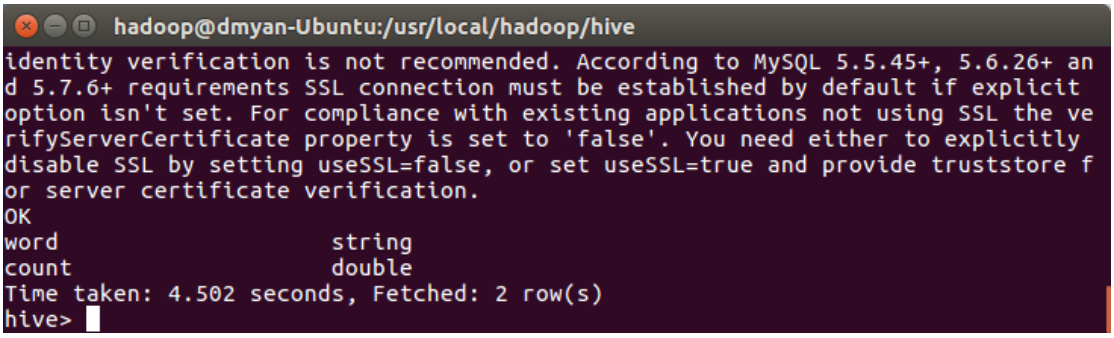
```
hadoop@dmyan-Ubuntu: /usr/local/hbase

hbase(main):002:0> scan 'Wuxia'
ROW                   COLUMN+CELL
 0                    column=count:freqs, timestamp=1510128897908, value=1.00
 1                    column=count:freqs, timestamp=1510128897924, value=2.33
 1238\xE5\xB9\xB4     column=count:freqs, timestamp=1510128897927, value=1.00
 1240\xE5\xB9\xB4     column=count:freqs, timestamp=1510128897931, value=1.00
 1368                 column=count:freqs, timestamp=1510128897935, value=1.00
 1480\xE5\xB9\xB4     column=count:freqs, timestamp=1510128897939, value=1.00
 1911\xE4\xB8\x80     column=count:freqs, timestamp=1510128897943, value=1.00
 1975\xE5\xB9\xB4     column=count:freqs, timestamp=1510128897946, value=1.00
 2                    column=count:freqs, timestamp=1510128897951, value=1.67
 240                  column=count:freqs, timestamp=1510128897955, value=1.00
 3                    column=count:freqs, timestamp=1510128897958, value=1.00
 37                   column=count:freqs, timestamp=1510128897962, value=1.00
 4                    column=count:freqs, timestamp=1510128897965, value=1.00
 5                    column=count:freqs, timestamp=1510128897970, value=1.00
 5\xE6\x9C\x88        column=count:freqs, timestamp=1510128897974, value=1.00
 6                    column=count:freqs, timestamp=1510128897979, value=1.00
 7                    column=count:freqs, timestamp=1510128897983, value=1.00
 BertrandRussell      column=count:freqs, timestamp=1510128897986, value=1.00
 Boleslaw             column=count:freqs, timestamp=1510128897991, value=1.00
 Brunei               column=count:freqs, timestamp=1510128897994, value=1.00
 C                    column=count:freqs, timestamp=1510128897998, value=4.27
```

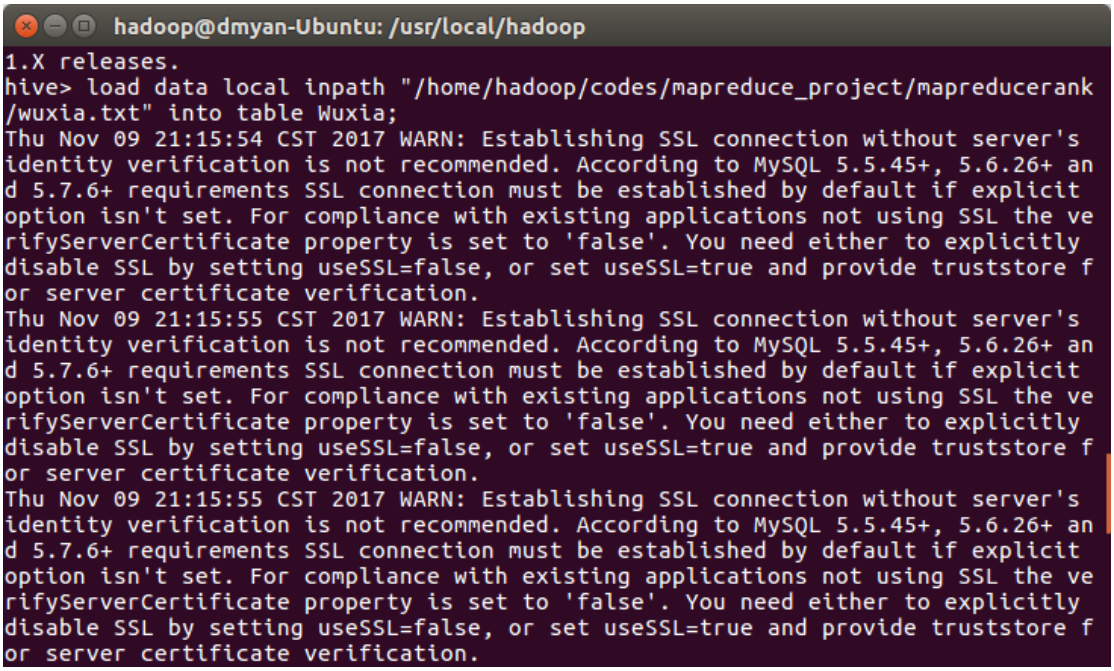
```
hadoop@dmyan-Ubuntu: /usr/local/hbase
 \x8B
 \xE9\xBE\x9F\xE7\x94 column=count:freqs, timestamp=1510232233134, value=2.33
 \xB2
 \xE9\xBE\x9F\xE7\xBA column=count:freqs, timestamp=1510232233135, value=3.00
 \xB9
 \xE9\xBE\x9F\xE7\xBC column=count:freqs, timestamp=1510232233136, value=1.44
 \xA9
 \xE9\xBE\x9F\xE8\x82 column=count:freqs, timestamp=1510232233137, value=1.33
 \x89
 \xE9\xBE\x9F\xE8\x83 column=count:freqs, timestamp=1510232233138, value=2.50
 \x8C
 \xE9\xBE\x9F\xE8\xA3 column=count:freqs, timestamp=1510232233139, value=1.29
 \x82
 \xE9\xBE\x9F\xE9\xB3 column=count:freqs, timestamp=1510232233140, value=2.00
 \x96
 \xE9\xBE\x9F\xE9\xB9 column=count:freqs, timestamp=1510232233141, value=1.00
 \xA4\xE9\x81\x90\xE9
 \xBE\x84
 \xEF\xA8\x8C         column=count:freqs, timestamp=1510232233142, value=5.00
 \xEF\xBF\xA1         column=count:freqs, timestamp=1510232233143, value=1.50
 \xEF\xBF\xA5         column=count:freqs, timestamp=1510232233144, value=3.33
134882 row(s) in 74.8350 seconds

hbase(main):011:0>
```

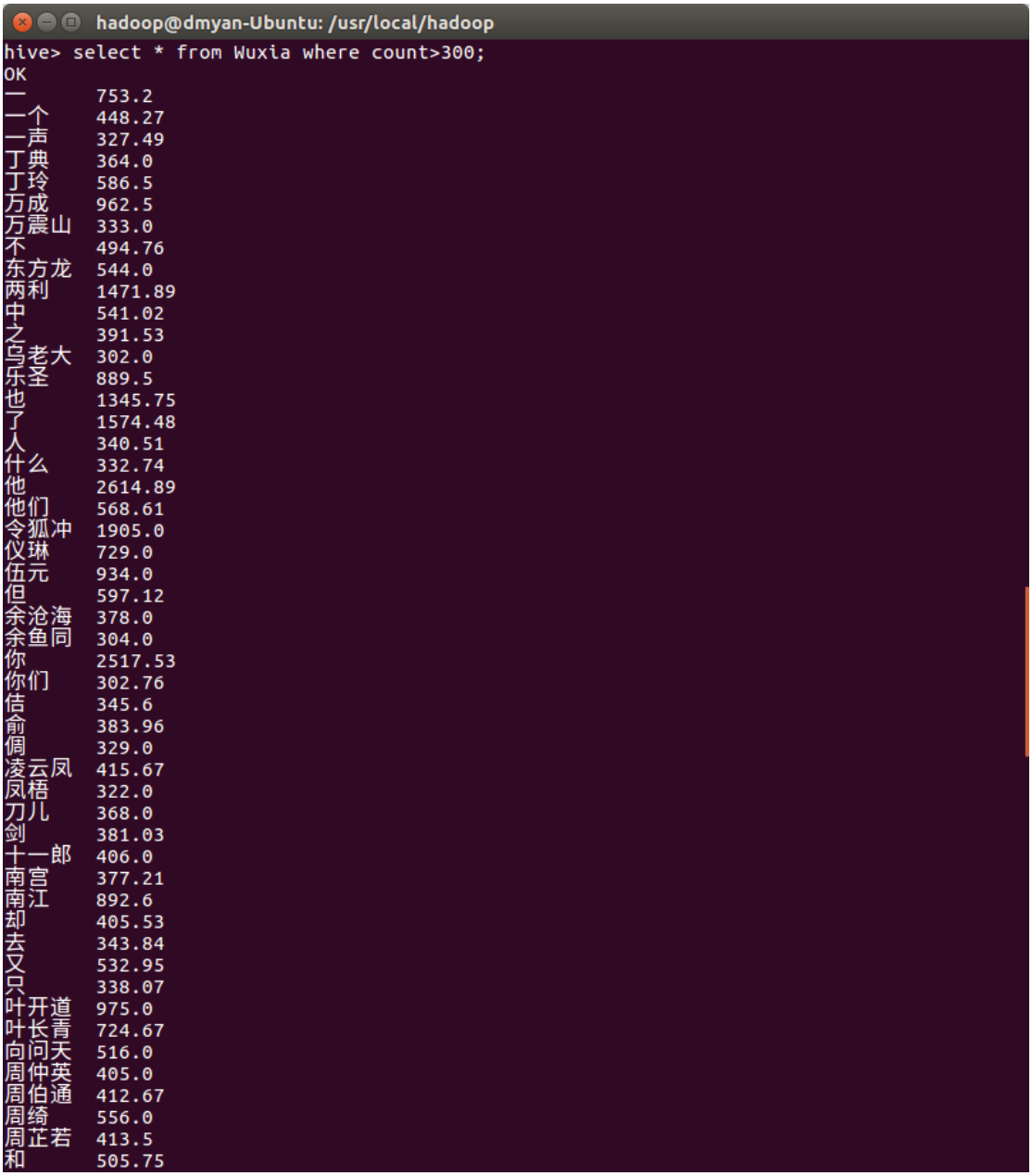
3、 Hive 创建 Wuxia 数据表



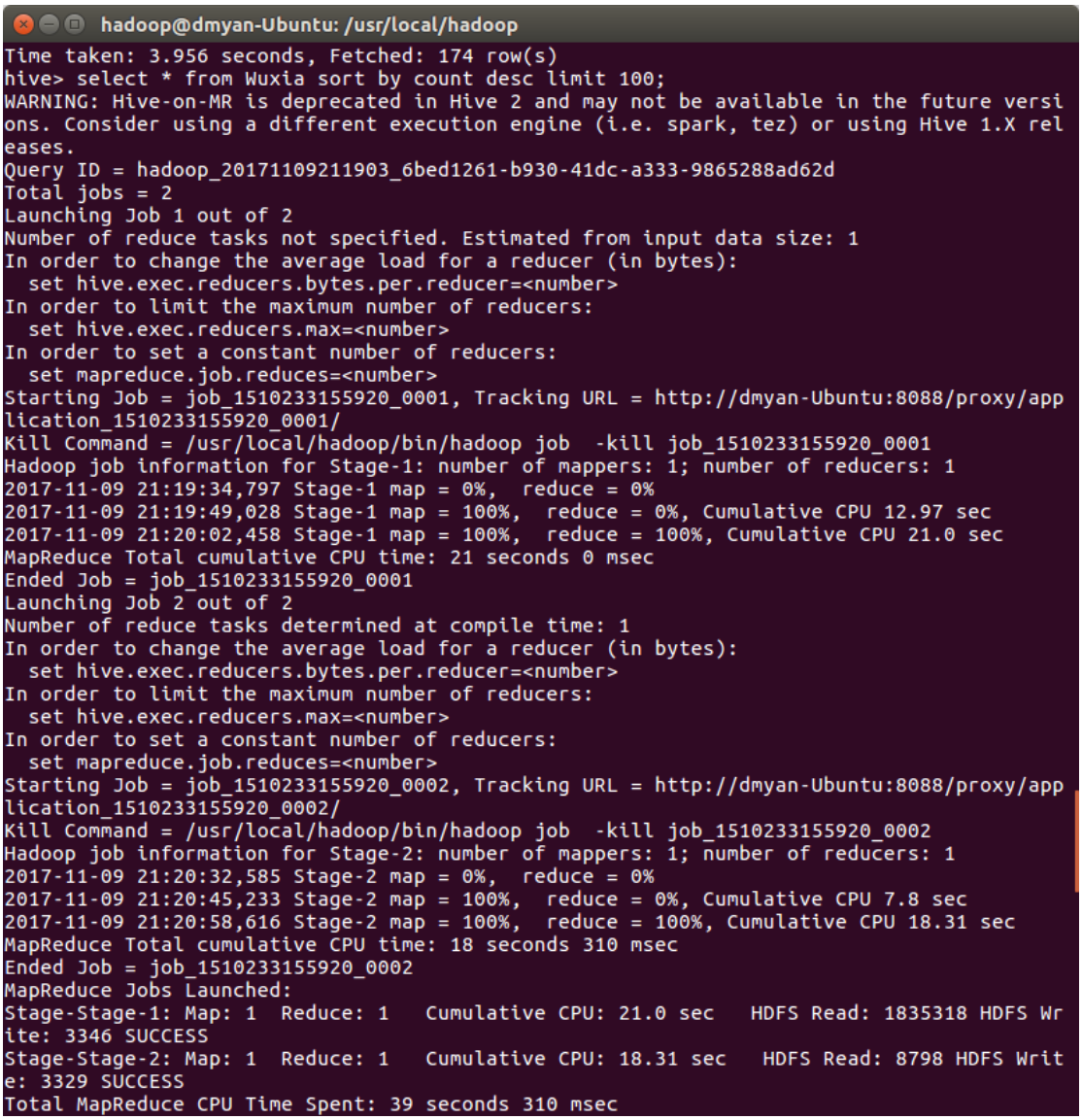
4、 导入数据到 Hive

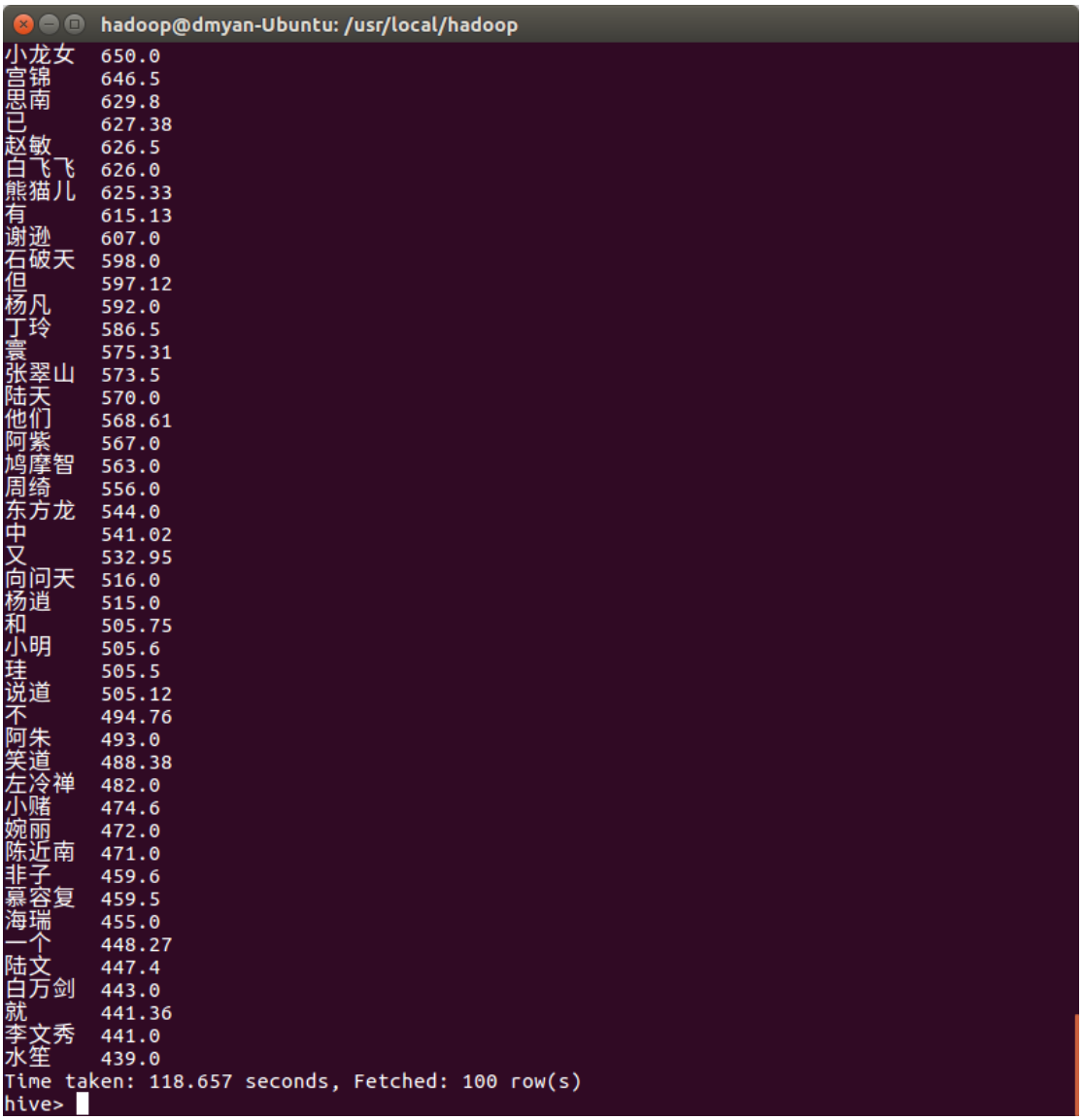


5、 查询出现次数大于 300 的词语



6、 查询 100 个出现次数最多的词


hadoop@dmyan-Ubuntu: /usr/local/hadoop
Time taken: 3.956 seconds, Fetched: 174 row(s)
hive> select * from Wuxia sort by count desc limit 100;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versi
ons. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X rel
eases.
Query ID = hadoop_20171109211903_6bed1261-b930-41dc-a333-9865288ad62d
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 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>
Starting Job = job_1510233155920_0001, Tracking URL = http://dmyan-Ubuntu:8088/proxy/app
lication_1510233155920_0001/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1510233155920_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2017-11-09 21:19:34,797 Stage-1 map = 0%, reduce = 0%
2017-11-09 21:19:49,028 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.97 sec
2017-11-09 21:20:02,458 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 21.0 sec
MapReduce Total cumulative CPU time: 21 seconds 0 msec
Ended Job = job_1510233155920_0001
Launching Job 2 out of 2
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>
Starting Job = job_1510233155920_0002, Tracking URL = http://dmyan-Ubuntu:8088/proxy/app
lication_1510233155920_0002/
Kill Command = /usr/local/hadoop/bin/hadoop job -kill job_1510233155920_0002
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2017-11-09 21:20:32,585 Stage-2 map = 0%, reduce = 0%
2017-11-09 21:20:45,233 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 7.8 sec
2017-11-09 21:20:58,616 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 18.31 sec
MapReduce Total cumulative CPU time: 18 seconds 310 msec
Ended Job = job_1510233155920_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 21.0 sec HDFS Read: 1835318 HDFS Wr
ite: 3346 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 18.31 sec HDFS Read: 8798 HDFS Writ
e: 3329 SUCCESS
Total MapReduce CPU Time Spent: 39 seconds 310 msec



7、 输出到本文文件的内容附在最后

8、 实验体会

通过本次实验学会了安装和使用 Hbase,Hive，在 MapReduce 过程中使用 Hbase 保存数据，对 MapReduce 的过程更加清晰了，对以后的学习和使用 MapReduce 大有益处。

部分输出内容

诏命 1.33

诏安 2.00

诏旨 3.11

诏曰 1.17

诏示 1.00

诏谕 1.33

译　 2.83

译为 2.33

译传 1.00

译作 1.50

译出 1.71

译名 1.50

译员 3.00

译官 1.00

译意 1.00

译成 2.05

译文 1.75

译本 1.17

译注 1.00

译着 1.00

译笔 1.00

译经 1.00

译者 1.00

译述 1.00

译错 1.00

译音 1.00

诒　1.00

诓　2.33

诓骗 1.33

诔　1.33

试　12.44

试一下 1.43

试一试 2.96

试了 3.79

试了试 1.24

试以 1.00

试作 1.00

试做 1.00

试出 2.62

试出来 1.15

试制 1.00

试加 2.00

试卷 1.25

试吃 1.29

试听 1.00

试图 1.40

试场 1.00

试射 1.25

试得 2.00

试想 3.10

试成 1.00

试戴 1.00

试手 1.27

试探 4.96

试探性 1.42

试来试去 1.00