

## 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

```

hadoop@dmyan-Ubuntu:/usr/local/hadoop/hive
identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
OK
word                string
count               double
Time taken: 4.502 seconds, Fetched: 2 row(s)
hive>

```

#### 4、 导入数据到 Hive

```

hadoop@dmyan-Ubuntu: /usr/local/hadoop
1.X releases.
hive> load data local inpath "/home/hadoop/codes/mapreduce_project/mapreducerank/wuxia.txt" into table Wuxia;
Thu Nov 09 21:15:54 CST 2017 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu Nov 09 21:15:55 CST 2017 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
Thu Nov 09 21:15:55 CST 2017 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.

```

#### 5、 查询出现次数大于 300 的词语

```
hadoop@dmyan-Ubuntu: /usr/local/hadoop
hive> select * from Wuxia where count>300;
OK
一个 753.2
一声 448.27
丁典 327.49
丁玲 364.0
万成 586.5
万震山 962.5
不 333.0
东方龙 494.76
两利 544.0
中之 1471.89
乌老大 541.02
乐圣 391.53
也 302.0
了 889.5
人 1345.75
什么 1574.48
他们 340.51
令狐冲 332.74
仪琳 2614.89
伍元 568.61
但 1905.0
余沧海 729.0
余鱼同 934.0
你们 597.12
你 378.0
估 304.0
俞 2517.53
个 302.76
凌 345.6
云凤 383.96
凤梧 329.0
刀儿 415.67
剑 322.0
十一郎 368.0
南宫 381.03
南江 406.0
却 377.21
去 892.6
又 405.53
只 343.84
叶开 532.95
长青 338.07
问天 975.0
周仲英 724.67
周伯通 516.0
周绮 405.0
周芷若 412.67
和 556.0
505.75
```

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 versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
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/application_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/application_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 Write: 3346 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 18.31 sec HDFS Read: 8798 HDFS Write: 3329 SUCCESS
Total MapReduce CPU Time Spent: 39 seconds 310 msec

```

```
hadoop@dmyan-Ubuntu: /usr/local/hadoop
小龙女 650.0
宫锦 646.5
思南 629.8
已敏 627.38
赵敏 626.5
白飞飞 626.0
熊猫儿 625.33
有 615.13
谢逊 607.0
石破天 598.0
但 597.12
杨凡 592.0
丁玲 586.5
袁 575.31
张翠山 573.5
陆天 570.0
他们 568.61
阿紫 567.0
鸠摩智 563.0
周绮 556.0
东方龙 544.0
中 541.02
又 532.95
向问天 516.0
杨逍 515.0
和 505.75
小明 505.6
珪 505.5
说道 505.12
不 494.76
阿朱 493.0
笑道 488.38
左冷禅 482.0
小赌 474.6
婉丽 472.0
陈近南 471.0
非子 459.6
慕容复 459.5
海瑞 455.0
一个 448.27
陆文 447.4
白万剑 443.0
就 441.36
李文秀 441.0
水笙 439.0
Time taken: 118.657 seconds, Fetched: 100 row(s)
hive>
```

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

诤 2.33

诤骗 1.33

诤 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