

## 实验四 Hbase, Hive 安装与使用

141070010, 141180016, 141290008

### 1、在自己电脑正确安装和运行 HBase 和 Hive

下载安装包后, 依照网上教程修改配置文件即可

```
p_installs/hbase-1.1.10/bin/htrb.rb)
  at org.jruby.Ruby.runScript(Ruby.java:697)
  at org.jruby.Ruby.runScript(Ruby.java:690)
  at org.jruby.Ruby.runNormally(Ruby.java:597)
  at org.jruby.Ruby.runFromMain(Ruby.java:446)
  at org.jruby.Main.doRunFromMain(Main.java:369)
  at org.jruby.Main.internalRun(Main.java:258)
  at org.jruby.Main.run(Main.java:224)
  at org.jruby.Main.run(Main.java:208)
  at org.jruby.Main.main(Main.java:188)
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.1.10, raf9719e45a1ee9f4509607aff7e554119d29e930, Tue Apr 18 20:57:55
hbase(main):001:0>
```

```
must be established by default if explicit option isn't set. For compliance with existing
Archive Manager SSL the verifyServerCertificate property is set to 'false'. You need either
licitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for
certificate verification.
Fri May 12 07:46:06 PDT 2017 WARN: Establishing SSL connection without server's identity ve
ion is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL con
must be established by default if explicit option isn't set. For compliance with existing
tions not using SSL the verifyServerCertificate property is set to 'false'. You need either
licitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for
certificate verification.
Fri May 12 07:46:07 PDT 2017 WARN: Establishing SSL connection without server's identity ve
ion is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL con
must be established by default if explicit option isn't set. For compliance with existing
tions not using SSL the verifyServerCertificate property is set to 'false'. You need either
licitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for
certificate verification.
hive>
```

### 2、在 Hbase 中创建表 Wuxia:

Create 'Wuxia','datas'

### 3、将 mapreduce 结果写入 Hbase 的表 Wuxia 中

```
if(!preword.equals(word) && !preword.equals(new Text(" "))) {
```

```
    StringBuilder out = new StringBuilder();
    long size = postingList.size();
    double average = wordsum / (double) size;
```

```
    HTable table = new HTable(HBASE_CONFIG, tableName);
    //以当前单词作为 Row_keys
    Put put = new Put(preword.getBytes());
    //参数分别对应列族, 修饰符, 列值
    put.add(Bytes.toBytes("datas"), Bytes.toBytes("average"),
        Bytes.toBytes(String.format("%.2f", average)));
    table.put(put);
    table.close();
```

### 4、遍历 Hbase 中的表

```
String tableName = "Wuxia";
Configuration HBASE_CONFIG = HBaseConfiguration.create();

Scan scan = new Scan();
HTable table = new HTable(HBASE_CONFIG, tableName);
ResultScanner resultScanner = table.getScanner(scan);

for(Result r:resultScanner) {
    //由上知每行的行键就是单词
    String word = Bytes.toString(r.getRow());
    //获取出现频率
    byte[] valueBytes = r.getValue(Bytes.toBytes("datas"),
        Bytes.toBytes("average"));
    String value = Bytes.toString(valueBytes);
    String fileName = "Wuxia.txt";
    FileWriter writer = new FileWriter(fileName, true);
    writer.write(word + "\t" + value + "\n");
    writer.close();
}
}
```

0	1.69	
007	1.0	
01	1.0	
01章	1.0	
02章	1.0	
03章	1.0	
04	1.0	
04章	1.0	
05	1.0	
05章	1.0	
06	1.0	
06章	1.0	
07章	1.0	
08张	1.0	
08章	1.0	
09章	1.0	
0一	1.5	
0年	2.5	
1	3.31	
1.	1.0	
10	2.0	
1000万钱	1.0	1.0
100间	1.0	
102个	1.0	
107	1.0	
10zz	9.0	
10岁	1.0	
10月	3.0	
10章	1.0	
11	1.0	
1100	1.0	
11岁	1.0	
11月	4.0	
11章	1.0	
12	1.0	
120卷	1.0	
1216	1.0	
1237年	2.0	
1238年	1.0	
1240年	1.0	
125个	1.0	
1269年	2.0	
12层	2.0	
12岁	1.0	
12月	1.0	
12月份	3.0	
12盘	1.0	
13	1.0	
1300页	1.0	
1301年	1.0	
1329年	1.0	
1360年	1.0	

5、

(1) 创建表并导入数据:

```
hive> create table Wuxia(word string,average double)
> row format delimited
> fields terminated by '\t'
> lines terminated by '\n';
OK
Time taken: 2.272 seconds
hive> load data local inpath '/home/hadoop/Wuxia.txt' into table Wuxia;
Loading data to table default.wuxia
Table default.wuxia stats: [numFiles=1, totalSize=1897164]
OK
Time taken: 1.335 seconds
```

(2) 查询出现次数大于 300 的词语:

```
Select * from Wuxia where average > 300
```

```
hive> select * from wuxia where average > 300
> ;
OK
_ 753.2
_ 448.27
一个 327.49
一声 364.0
丁典 586.5
丁玲 962.5
万成 333.0
万震山 494.76
不 544.0
东方龙 1471.89
两利 541.02
中之 391.53
老 302.0
大 889.5
乐圣 1345.75
也 1574.48
了 340.51
人 332.74
什么 2614.89
他 568.61
他们 1905.0
令狐冲 729.0
仪琳 934.0
伍元 597.12
但 378.0
余沧海 304.0
余鱼同 2517.53
你
```

(3) 前 100 个出现次数最多的词:

Select \* from Wuxia order by average desc limit 100

```
OK
的 7168.19
的 4230.0
英豪 3277.0
小宝 3272.61
道 2614.89
他 2517.53
你 2438.67
玮 2373.63
我 2338.0
无忌 1905.0
令狐冲 1898.5
程小灵 1756.25
小鱼儿 1744.0
齐金麟 1688.0
段誉 1686.0
虚竹 1574.48
了 1528.4
两利 1471.89
自维 1416.0
云 1408.0
也 1345.75
也 1336.1
尊 1262.5
尊 1229.0
行 1184.0
不群 1110.5
天舒 1073.0
前靖 1021.6
杨过 1013.33
志 997.33
胡斐 995.14
寒秋 983.27
耶 975.0
开道
```