

实验三倒排索引——实验报告

一. 小组信息:

集群账号: 2017st15

141220138 臧凤云 2714359840@qq.com

141220132 银琦 141220132@smail.nju.edu.cn

二. 实验环境

Ubuntu 14、Java 1.7.0_25、IntelliJ 17

三. 实验流程

1.在 IntelliJ 上生成 InvertedIndex.jar, 拷贝到本机的 hadoop 用户下

2.在本机 Hadoop 用户下将 jar 包拷贝到集群下

```
scp -r InvertedIndex.jar 2017st15@114.212.190.91:~/
```

3.执行 jar 包

基础:hadoop jar InvertedIndex.jar InvertedIndex /data/wuxia_novels/. /user/2017st15/output

选作 1: hadoop jar InvertedIndex.jar FrequencySort /output/part-r-00000 /user/2017st15/sort-out

选作 2: hadoop jar InvertedIndex.jar TFIDF /output/part-r-00000 /user/2017st15/tf_out

4. 将执行后的文件保存到集群账户上

```
hadoop fs -get /user/2017st15/output
```

```
hadoop fs -get /user/2017st15/sort-out
```

```
hadoop fs -get /user/2017st15/tf_out
```

5. 将文件复制到本地查看

```
scp -r 2017st15@114.212.190.91:~/output ~/
```

```
scp -r 2017st15@114.212.190.91:~/sort-out ~/
```

```
scp -r 2017st15@114.212.190.91:~/tf_out ~/
```

四. Map 和 Reduce 的设计思路

(1) 倒排索引

Map: 将输入文本 value 信息按照空格进行拆分, 获取单词, 输入的 context 获取 mapper 分配的 split, 得到作家小说名, 将单词和作家小说名作为输出的 key, 输出的 value 为“1”。

Combiner: 将 map 的输出作为输入, 将一个 key 的所有 value 进行合并, 得到一个单词在一个文件的总次数。对输入的 key 进行划分, 将单词作为输出的 key, 将作家小说名+':'+count 作为输出 value。

Reducer: 对 Combiner 的输出进行处理, 输入的 key 为单词, 输入的 values 为所有的作家小说名+':'+count, 计算 key 的总 count, 除以该 key 出现的文件总数得到频率。对 values 进行合并作为输出的 value。

(2) 选作 1——频率排序

Map: 输入 key 无意义, value 为倒排索引输出文件的一行信息即一个单词的信息, 因为 MapReduce 会按照 key 自动排序, 对 value 进行拆分, 将频率作为输出的 key (DoubleWritable 类型), 将 value 的其他信息作为输出的 value。

Reduce: 将 value 中的单词作为输出的 key, value 中的其他信息和频率(即输入的 key)作

为输出的 value。

(3) 选作 2——TF-IDF

Map: 输入的 key 无意义, 输入的 value 为倒排索引的输出文件的一行信息。建立 `ArrayList<String> Authors` 存储作者信息, 建立 `ArrayList<Integer> times` 存储词的数目, 根据‘.’号来提取小说名, 以及提取小说名中的作者名, 并统计该词出现的数目。输出 key 为单词加作者, value 为 TF 和包含该词的文档数。

Reduce: 对 Map 输出进行处理, 使用公式 $\text{Math.log}((\text{double})218 / (\text{double})(\text{count} + 1)) / \text{Math.log}(2)$; 计算 IDF。输出的 key 即为输入的 key, 输出的 value 为 TF-IDF。(这里对文档总数的获取没有实现, 所以直接使用了 218)。

五. 倒排索引代码分析

//mapper

```
class InvertedIndexMapper extends Mapper<Object,Text,Text,Text>{
    @Override
    //key 意义, value 为一行文本信息, context 包含文件名信息
    protected void map(Object key, Text value, Context context) throws IOException,
        InterruptedException
    {
        //获取书名信息, 根据文件名中的‘.’进行拆分, 去掉.txt.segmented
        FileSplit fileSplit = (FileSplit)context.getInputSplit();
        String[] split = String.valueOf(fileSplit.getPath().getName()).split("\\.");
        String filename;
        if(split.length==4)
            filename=split[0]+split[1];
        else
            filename=split[0];
        //对 value 进行划分获得单词组
        StringTokenizer itr = new StringTokenizer(value.toString());
        while(itr.hasMoreTokens()) {
            //获得 keyinfo ,格式为 “单词+: +filename”
            Text keyinfo = new Text(itr.nextToken()+":"+filename);
            //对每个单词依次输出, value 为‘1’
            context.write(keyinfo, new Text("1"));
        }
    }
}
```

//combiner

```
class InvertedIndexCombiner extends Reducer<Text,Text,Text,Text>{
    //key 格式为 “单词+: +filename”, values 为含有 key 的所有键值对中的 value
    protected void reduce(Text key, Iterable<Text>values ,Context context)throws IOException,
        InterruptedException
    {
        Text valueinfo =new Text();//输出 value
        Text keyinfo = new Text();//输出 key
    }
}
```

```

//计算一个单词在一个文件出现的次数
int sum=0;
for(Text x:values){
    sum+= Integer.parseInt(x.toString());
}
//按照 ‘:’ 位置对 key 划分
int index = key.toString().indexOf(":");
//单词作为输出 key
keyinfo=new Text(key.toString().substring(0,index));
//输出的 value 格式为“filename+: +count”
valueinfo=new
Text(key.toString().substring(index+1,key.toString().length())+' '+Integer.toString(sum));
//输出
context.write(keyinfo,valueinfo);
}
}
//reducer
class InvertedIndexReducer extends Reducer<Text,Text,Text,Text> {
    @Override
    //key 为单词， values 为含有 key 所有文件中数量的集合， 格式为 “filename+: +count”
    protected void reduce(Text key, Iterable<Text> values, Context context) throws IOException,
    InterruptedException
    {
        ArrayList<String>novel = new ArrayList<String>();//用于记录 filename
        ArrayList<Integer>times = new ArrayList<Integer>();//记录 key 在每个文件出现次数
        for(Text x : values) {
            //根据‘:’拆分
            int index = x.toString().indexOf(':');
            String novelName = x.toString().substring(0,index);
            //count 计算单词出现作品数量
            int count = Integer.parseInt(x.toString().substring(index+1,x.toString().length()));
            if(novel.indexOf(novelName)==-1){
                novel.add(novelName);
                times.add(count);
            }
            else{
                int temp = times.get(novel.indexOf(novelName));
                temp+=count;
                times.set(novel.indexOf(novelName),temp);
            }
        }
    }

    //sum 计算单词在每个小说中出现次数之和
    double sum = 0;

```

```

for(int i : times){
    sum+=(double)i;
}
//计算频率,保留两位小数
double avg = sum / times.size();
String avgf = String.format("%.2f",avg);
//value,格式为 “频率+, +filename1+: +count1+: +filename2+: +count2+: +....”
StringBuilder all = new StringBuilder();
all.append(avgf+',');
for(int i=0; i < times.size(); i++){
    all.append(novel.get(i)+'.'+ Integer.toString(times.get(i))+';');
}

//输出
context.write(key, new Text(all.toString().substring(0,all.length()-1)));
}
}

```

六. 输出结果截图

(1) 倒排索引

```

[2017st15@master01 ~]$ hadoop fs -ls /user/2017st15/output
17/04/30 17:22:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  3 2017st15 hadoop_user      0 2017-04-30 16:58 /user/2017st15/output/_SUCCESS
-rw-r--r--  3 2017st15 hadoop_user 105642849 2017-04-30 16:58 /user/2017st15/output/part-r-00000

```

```

yq-mapreduce@ubuntu: ~
1359 reality 1.00,梁羽生32随笔集:笔花六照:1;梁羽生33随笔集:三剑楼随笔:1;梁羽生31随笔集:笔不花:1
1360 rebelon 1.00,梁羽生23龙虎斗京华:1
1361 ribbon 1.00,梁羽生32随笔集:笔花六照:1
1362 rtiso 1.00,梁羽生32随笔集:笔花六照:1
1363 rtiver 1.00,梁羽生32随笔集:笔花六照:1
1364 ronance 3.00,梁羽生32随笔集:笔花六照:3
1365 rpo 1.00,卧龙生45燕子传奇:1
1366 ry 3.00,梁羽生32随笔集:笔花六照:3
1367 s 1.11,卧龙生45燕子传奇:1;金庸13碧血剑:1;梁羽生38云海玉弓缘:1;李凉04本尊分身:1;古龙31剑气书香:1;李凉21六宝江湖行:1;卧龙生11黑白双娇:1;李凉11会醉才会赢:2;李凉16江湖一品郎:1
1368 salatt 1.00,梁羽生32随笔集:笔花六照:1
1369 satya 1.00,梁羽生32随笔集:笔花六照:1
1370 sea 1.00,梁羽生32随笔集:笔花六照:1;梁羽生31随笔集:笔不花:1
1371 seven 1.00,梁羽生32随笔集:笔花六照:1
1372 shlp 2.00,梁羽生32随笔集:笔花六照:2;梁羽生31随笔集:笔不花:2
1373 slnon 1999 1.00,古龙35绝代双骄:1
1374 skadarlija 1.00,梁羽生32随笔集:笔花六照:1
1375 slqlzf 33.00,古龙37流星·蝴蝶·剑:33
1376 smith 1.00,梁羽生32随笔集:笔花六照:1
1377 somelinesmissinghere 1.00,李凉26奇神扬小邪续集:1
1378 south 1.00,梁羽生32随笔集:笔花六照:1;梁羽生31随笔集:笔不花:1
1379 sp 1.00,金庸02雪山飞狐:1
1380 stephen 1.00,梁羽生32随笔集:笔花六照:1
1381 style 1.00,古龙63武林外史:1
1382 sunny 1.00,梁羽生32随笔集:笔花六照:1;梁羽生31随笔集:笔不花:1
1383 swordmen 1.00,梁羽生32随笔集:笔花六照:1
1384 swordmenfrom 1.00,梁羽生32随笔集:笔花六照:1
1385 sydney 1.00,梁羽生32随笔集:笔花六照:1
1386 t 1.17,梁羽生38云海玉弓缘:1;梁羽生37游剑江湖:1;卧龙生15绛雪玄霜:1;卧龙生30素手劫:1;李凉16江湖一品郎:1;李凉09红顶记:2;李凉34天下第一当:1;卧龙生25七绝剑:1;还情剑:1;卧龙生23飘花令:1;卧龙生04地獄门:1;李凉20狂侠南宮鹰:2;李凉41杨小邪发威:1;梁羽生24鸣镝风云录:1;李凉30淘气世家:1;卧龙生42新仙鹤神针:1;卧龙生29双凤旗:1;李凉17惊神关小刀:1;李凉37小鱼吃大鱼:2
1387 t. 1.00,李凉02霸枪艳血:1
1388 tabb 1.00,卧龙生29双凤旗:1
1389 tbs 1.00,李凉26奇神扬小邪续集:1
1390 temple 1.00,梁羽生32随笔集:笔花六照:1
1391 the 4.50,梁羽生32随笔集:笔花六照:5;梁羽生31随笔集:笔不花:4
1392 thxde 1.00,卧龙生30素手劫:1
1393 tl 1.00,李凉20狂侠南宮鹰:1
1394 tian 1.00,梁羽生32随笔集:笔花六照:1
1395 tj 1.00,李凉04本尊分身:1

```

(2) 选作 1——频数排序

yq-mapredude@ubuntu:~\$ cat 129992 129993 129994 129995 129996 129997 129998 129999 130000 130001 130002 130003 130004 130005 130006 130007 130008 130009 130010 130011 130012 130013 130014 130015 130016 130017 130018 130019 130020 130021 130022 130023 130024 130025 130026 130027 130028 130029 130030 130031 130032 130033 130034 130035 130036 130037 130038 130039 130040 130041 130042 130043 130044 130045 130046 130047 130048 130049 130050 130051 130052 130053 130054 130055 130056 130057 130058 130059 130060 130061 130062 130063 130064 130065 130066 130067 130068 130069 130070 130071 130072 130073 130074 130075 130076 130077 130078 130079 130080 130081 130082 130083 130084 130085 130086 130087 130088 130089 130090 130091 130092 130093 130094 130095 130096 130097 130098 130099 130100 130101 130102 130103 130104 130105 130106 130107 130108 130109 130110 130111 130112 130113 130114 130115 130116 130117 130118 130119 130120 130121 130122 130123 130124 130125 130126 130127 130128 130129 130130 130131 130132 130133 130134 130135 130136 130137 130138 130139 130140 130141 130142 130143 130144 130145 130146 130147 130148 130149 130150 130151 130152 130153 130154 130155 130156 130157 130158 130159 130160 130161 130162 130163 130164 130165 130166 130167 130168 130169 130170 130171 130172 130173 130174 130175 130176 130177 130178 130179 130180 130181 130182 130183 130184 130185 130186 130187 130188 130189 130190 130191 130192 130193 130194 130195 130196 130197 130198 130199 130200 130201 130202 130203 130204 130205 130206 130207 130208 130209 130210 130211 130212 130213 130214 130215 130216 130217 130218 130219 130220 130221 130222 130223 130224 130225 130226 130227 130228 130229 130230 130231 130232 130233 130234 130235 130236 130237 130238 130239 130240 130241 130242 130243 130244 130245 130246 130247 130248 130249 130250 130251 130252 130253 130254 130255 130256 130257 130258 130259 130260 130261 130262 130263 130264 130265 130266 130267 130268 130269 130270 130271 130272 130273 130274 130275 130276 130277 130278 130279 130280 130281 130282 130283 130284 130285 130286 130287 130288 130289 130290 130291 130292 130293 130294 130295 130296 130297 130298 130299 130300 130301 130302 130303 130304 130305 130306 130307 130308 130309 130310 130311 130312 130313 130314 130315 130316 130317 130318 130319 130320 130321 130322 130323 130324 130325 130326 130327 130328 130329 130330 130331 130332 130333 130334 130335 130336 130337 130338 130339 130340 130341 130342 130343 130344 130345 130346 130347 130348 130349 130350 130351 130352 130353 130354 130355 130356 130357 130358 130359 130360 130361 130362 130363 130364 130365 130366 130367 130368 130369 130370 130371 130372 130373 130374 130375 130376 130377 130378 130379 130380 130381 130382 130383 130384 130385 130386 130387 130388 130389 130390 130391 130392 130393 130394 130395 130396 130397 130398 130399 130400 130401 130402 130403 130404 130405 130406 130407 130408 130409 130410 130411 130412 130413 130414 130415 130416 130417 130418 130419 130420 130421 130422 130423 130424 130425 130426 130427 130428 130429 130430 130431 130432 130433 130434 130435 130436 130437 130438 130439 130440 130441 130442 130443 130444 130445 130446 130447 130448 130449 130450 130451 130452 130453 130454 130455 130456 130457 130458 130459 130460 130461 130462 130463 130464 130465 130466 130467 130468 130469 130470 130471 130472 130473 130474 130475 130476 130477 130478 130479 130480 130481 130482 130483 130484 130485 130486 130487 130488 130489 130490 130491 130492 130493 130494 130495 130496 130497 130498 130499 130500 130501 130502 130503 130504 130505 130506 130507 130508 130509 130510 130511 130512 130513 130514 130515 130516 130517 130518 130519 130520 130521 130522 130523 130524 130525 130526 130527 130528 130529 130530 130531 130532 130533 130534 130535 130536 130537 130538 130539 130540 130541 130542 130543 130544 130545 130546 130547 130548 130549 130550 130551 130552 130553 130554 130555 130556 130557 130558 130559 130560 130561 130562 130563 130564 130565 130566 130567 130568 130569 130570

```
[2017st15@master01 ~]$ hadoop fs -ls /user/2017st15/tf_out
17/04/30 17:26:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 3 2017st15 hadoop_user 0 2017-04-30 17:05 /user/2017st15/tf_out/_SUCCESS
-rw-r--r-- 3 2017st15 hadoop_user 16013861 2017-04-30 17:05 /user/2017st15/tf_out/part-r-000000
```

七. “江湖”，“风雪”输出结果

江湖 116.06, 卧龙生31桃花血令:131; 李凉07暗器小狂侠:31; 梁羽生36快骨丹心:103; 古龙44陆小凤07剑神一笑:47; 卧龙生40无名箫:136; 梁羽生02冰川天女传:65; 卧龙生02春秋笔:329; 古龙43陆小凤06凤舞九天:42; 古龙37流星·蝴蝶·剑:34; 梁羽生09风雷震九州:135; 古龙03碧血银枪:81; 卧龙生26情剑无刀:7; 卧龙生32天鹤谱:106; 李凉39新蜀山剑侠传:20; 卧龙生42新仙鹤神针:188; 古龙18大旗英雄传:151; 李凉14江湖双响炮:291; 古龙69圆月弯刀:155; 古龙46那一剑的风情:32; 卧龙生19惊鸿一剑震江湖:346; 古龙50七星龙王:55; 李凉29神偷小千:17; 古龙62天涯·明月·刀:69; 古龙59三少爷的剑:86; 李凉12圣宝小淘气:121; 金庸06白马啸西风:5; 卧龙生12剑气洞彻九重天:299; 卧龙生18金剑雕龙:397; 李凉02霸枪艳血:34; 古龙12楚留香04借尸还魂(鬼恋传奇):41; 古龙38陆小凤01陆小凤前传(金鹏王朝):15; 古龙45名剑风流:201; 李凉01暗器高手:53; 李凉03百败小霸王:79; 金庸12倚天屠龙记:145; 梁羽生26女帝奇英传:49; 李凉18酒醉小浪子:17; 梁羽生18绝塞传烽录:40; 梁羽生14幻剑侠情:27; 李凉10拳头小子:43; 古龙34绝不低头:3; 金庸14鸳鸯刀:35; 古龙16楚留香08午夜兰花:68; 梁羽生19狂侠天娇女:180; 李凉06超级邪侠:209; 卧龙生11黑白双骄:117; 卧龙生34天龙甲:293; 梁羽生30散花女侠:48; 梁羽生37游剑江湖:275; 梁羽生25牧野流星:35; 李凉04本尊分身:97; 古龙32剑玄录:160; 卧龙生04地狱门:105; 李凉27奇神杨小邪:102; 古龙07残金缺玉:82; 古龙39陆小凤02绣花大盗:23; 金庸09书剑恩仇录:64; 卧龙生24七绝剑:130; 梁羽生20联剑风云录:126; 古龙05边城浪子:62; 卧龙生41无形剑:256; 古龙33九月鹰飞:50; 李凉22矛盾天师:59; 李凉23妙贼小勾:16; 古龙31剑气书香:27; 梁羽生01白发魔女传:63; 卧龙生35天马追衣:278; 卧龙生23飘花令:263; 梁羽生17江湖三女侠:252; 梁羽生15慧剑心魔:102; 古龙06彩环曲:76; 古龙47怒剑狂花:102; 梁羽生28七剑下天山:77; 李凉08公孙小刀:69; 梁羽生11广陵剑:149; 古龙28剑·花·烟雨·江南:18; 卧龙生28神州豪侠:261; 古龙24护花铃:193; 卧龙生50王权盟:320; 古龙64湘妃剑:169; 金庸10神雕侠侣:81; 李凉15江湖一担皮:197; 古龙35绝代双骄:249; 卧龙生13刺龙:1261; 李凉28忍辱角:77; 古龙21飞刀·又见飞刀:30; 梁羽生27萍踪侠影录:40; 古龙40陆小凤03决战前后:26; 古龙55七种武器05霸王枪:54; 梁羽生29塞外奇侠传:2; 梁羽生22龙凤宝钗缘:109; 古龙23孤星传:175; 李凉11会醉才金:189; 卧龙生16金笔点龙记:318; 古龙66火并萧十一郎:89; 金庸04天龙八部:126; 梁羽生24鸣镝风云录:134; 梁羽生07弹指惊雷:77; 古龙65萧十一郎:56; 李凉25魔手邪怪:6; 古龙09楚留香01血海飘香:47; 古龙10楚留香02大沙漠:23; 李凉21六宝江湖行:72; 古龙57七种武器07拳·头:16; 卧龙生51王手点将录:150; 梁羽生03冰河洗剑录:65; 卧龙生22女捕头:317; 李凉20狂侠南侠:15; 卧龙生20梦幻之刀:106; 卧龙生25七绝剑II还情剑:144; 李凉33江湖急救站:118; 梁羽生21梁羽生传奇:74; 梁羽生32随笔集·笔花六照:15; 卧龙生07飞燕惊龙:269; 卧龙生01镖旗:275; 古龙49七杀手:20; 古龙68游侠录:103; 卧龙生33天剑绝刀:321; 卧龙生47一代天骄:353; 古龙17大地飞鹰:56; 古龙22风铃中的刀声:76; 李凉38笑笑江湖:69; 李凉32天才混世外传:127; 梁羽生35武林天骄:45; 古龙20多情剑客无情剑:150; 古龙52七种武器02碧玉刀:22; 古龙04边城刀声:50; 梁羽生31随笔集·笔不花:4; 古龙13楚留香05蝙蝠传奇:80; 卧龙生37天涯情偶:113; 李凉35武林漫游记:180; 古龙03武林外史:179; 李凉37小鱼吃大鱼:42; 古龙01白王老虎:111; 李凉36小霸王:29; 金庸05射雕英雄传:57; 梁羽生06大唐游侠传:97; 卧龙生54指剑为魔:57; 梁羽生38云海玉弓缘:81; 金庸11快客行:61; 卧龙生52紫紫烟:30; 古龙15楚留香07新月传奇:20; 古龙60神君别传:12; 古龙56七种武器06离别钩:39; 梁羽生16剑网尘丝:127; 卧龙生08风尘侠侣:228; 李凉09红顶记:48; 古龙48飘香剑雨:67; 卧龙生10黑白剑:223; 卧龙生44烟锁江湖:238; 金庸13碧血令:65; 梁羽生23龙虎斗京华:182; 梁羽生12瀚海雄风:169; 梁羽生34武当一剑:94; 卧龙生46烟花放屠城:446; 梁羽生08飞鹰雄风:4; 金庸02雪山飞狐:129; 卧龙生06飞铃:244; 李凉24妙贼丁小勾魂集:113; 古龙53七种武器03孔雀翎:10; 卧龙生39铁苗神剑:358; 李凉31天才混世:89; 卧龙生53岳小钗:308; 卧龙生05飞花逐月:298; 李凉26奇神杨小邪续集:98; 卧龙生45豪杰传:132; 李凉41杨小邪续集:34; 金庸03连城诀:35; 卧龙生43血刺丹心:350; 梁羽生13还剑·奇情录:8; 梁羽生04冰魄寒光剑:4; 古龙27血鸢:11; 古龙41陆小凤04银钩顺带:10; 古龙08苍苔神剑:131; 卧龙生14剑仙列传:25; 李凉05超霸的男人:7; 卧龙生03翠袖玉环:402; 古龙29剑毒梅香:110; 卧龙生29双凤旗:261; 古龙30剑客行:187; 古龙61失魂引:96; 李凉16江湖一品郎:186; 卧龙生49幽灵四艳:221; 卧龙生15锦雪玄霜:274; 卧龙生27琼楼十二曲:181; 古龙70月异星那:102; 古龙26浣花洗剑录:285; 李凉40新蜀山剑侠传:1; 古龙51七种武器01长生剑:23; 卧龙生36天香录:246; 金庸08笑傲江湖:199; 卧龙生21妙绝天香:28; 李凉34天下第一当:52; 卧龙生09风雨燕归来:198; 古龙14楚留香06桃花传奇:18; 卧龙生17金凤剪:326; 李凉30海气世家:81; 金庸01飞狐外传:67; 卧龙生38铁剑玉佩:71; 金庸07鹿鼎记:75; 古龙58情人箭:197; 李凉33天齐大帝:62; 梁羽生10风云雷电:139; 古龙36猎鹰·暗局:107; 古龙54七种武器04多情环:8; 古龙19大人物:37; 古龙02白玉屠龙:24; 卧龙生30素手劫:374; 古龙42陆小凤03幽灵山庄:40; 古龙67英雄无泪:54; 梁羽生05草莽龙蛇传:143; 李凉19酒狂任小腰:167; 卧龙生48银月飞:132; 古龙25欢乐英雄:64; 李凉17惊神大小刀:38; 古龙11楚留香03圆月弯刀:96; 李凉33金庸12倚天屠龙记:145; 卧龙生23飘花令:263; 李凉03百败小霸王:79; 金庸12倚天屠龙记:145

风雪 4.53, 梁羽生38云海玉弓缘:1; 卧龙生47一代天骄:353; 卧龙生19惊鸿一剑震江湖:346; 梁羽生03冰河洗剑录:65; 金庸07鹿鼎记:75; 金庸03连城诀:1; 李凉02枪艳血:5; 李凉36小霸王:29; 梁羽生20联剑风云录:126; 古龙25欢乐英雄:3; 古龙42陆小凤05幽灵山庄:1; 卧龙生07飞燕惊龙:16; 李凉11会醉才金:189; 卧龙生8铁剑玉佩:7; 梁羽生09风雷震九州:135; 梁羽生26女帝奇英传:49; 古龙31剑气书香:33; 梁羽生12瀚海雄风:1; 古龙46那一剑的风情:32; 卧龙生09风雨燕归来:1; 卧龙生08风尘侠侣:228; 古龙63武林外史:179; 古龙56七种武器06离别钩:39; 李凉01暗器高手:53; 卧龙生31桃花血令:131; 卧龙生25七绝剑II还情剑:144; 李凉03百败小霸王:79; 李凉34天下第一当:52; 梁羽生27萍踪侠影录:40; 李凉04本尊分身:97; 梁羽生01白发魔女传:63; 古龙13楚留香05蝙蝠传奇:80; 李凉39新蜀山剑侠传:20; 古龙54指剑为魔:57; 梁羽生02冰川天女传:63; 梁羽生21梁羽生传奇:74; 古龙61失魂引:96; 卧龙生15锦雪玄霜:274; 古龙14楚留香06桃花传奇:18; 金庸05射雕英雄传:57; 金庸12倚天屠龙记:145; 卧龙生42新仙鹤神针:188; 卧龙生36天香录:246; 古龙27血鸢:11; 梁羽生25牧野流星:35; 金庸10神雕侠侣:81; 梁羽生23龙虎斗京华:182; 梁羽生12瀚海雄风:169; 梁羽生34武当一剑:94; 金庸02雪山飞狐:129; 卧龙生06飞铃:244; 李凉24妙贼丁小勾魂集:113; 古龙53七种武器03孔雀翎:10; 卧龙生39铁苗神剑:358; 李凉31天才混世:89; 卧龙生53岳小钗:308; 卧龙生05飞花逐月:298; 李凉26奇神杨小邪续集:98; 卧龙生45豪杰传:132; 李凉41杨小邪续集:34; 金庸03连城诀:35; 卧龙生43血刺丹心:350; 梁羽生13还剑·奇情录:8; 梁羽生04冰魄寒光剑:4; 古龙27血鸢:11; 古龙41陆小凤04银钩顺带:10; 古龙08苍苔神剑:131; 卧龙生14剑仙列传:25; 李凉05超霸的男人:7; 卧龙生03翠袖玉环:402; 古龙29剑毒梅香:110; 卧龙生29双凤旗:261; 古龙30剑客行:187; 古龙61失魂引:96; 李凉16江湖一品郎:186; 卧龙生49幽灵四艳:221; 卧龙生15锦雪玄霜:274; 卧龙生27琼楼十二曲:181; 古龙70月异星那:102; 古龙26浣花洗剑录:285; 李凉40新蜀山剑侠传:1; 古龙51七种武器01长生剑:23; 卧龙生36天香录:246; 金庸08笑傲江湖:199; 卧龙生21妙绝天香:28; 李凉34天下第一当:52; 卧龙生09风雨燕归来:198; 古龙14楚留香06桃花传奇:18; 卧龙生17金凤剪:326; 李凉30海气世家:81; 金庸01飞狐外传:67; 卧龙生38铁剑玉佩:71; 金庸07鹿鼎记:75; 古龙58情人箭:197; 李凉33天齐大帝:62; 梁羽生10风云雷电:139; 古龙36猎鹰·暗局:107; 古龙54七种武器04多情环:8; 古龙19大人物:37; 古龙02白玉屠龙:24; 卧龙生30素手劫:374; 古龙42陆小凤03幽灵山庄:40; 古龙67英雄无泪:54; 梁羽生05草莽龙蛇传:143; 李凉19酒狂任小腰:167; 卧龙生48银月飞:132; 古龙25欢乐英雄:64; 李凉17惊神大小刀:38; 古龙11楚留香03圆月弯刀:96; 李凉33金庸12倚天屠龙记:145; 卧龙生23飘花令:263; 李凉03百败小霸王:79; 金庸12倚天屠龙记:145

八. MapReduce Job 截图

The screenshot shows the Hadoop All Applications page. The main table lists the following jobs:

| ID | User | Name | Application Type | Queue | StartTime | FinishTime | State | FinalStatus |
|--------------------------------|----------|---------------|------------------|---------|---------------------|---------------------|----------|-------------|
| application_1492438200769_0224 | 2017st15 | TFIDF | MAPREDUCE | default | Sun Apr 30 02:05:30 | Sun Apr 30 02:05:57 | FINISHED | SUCCEEDED |
| application_1492438200769_0223 | 2017st03 | Invert index | MAPREDUCE | default | Sun Apr 30 02:04:20 | Sun Apr 30 02:05:21 | FINISHED | SUCCEEDED |
| application_1492438200769_0222 | 2017st15 | FrequencySort | MAPREDUCE | default | Sun Apr 30 02:02:42 | Sun Apr 30 02:03:36 | FINISHED | SUCCEEDED |
| application_1492438200769_0221 | 2017st15 | InvertedIndex | MAPREDUCE | default | Sun Apr 30 01:57:19 | Sun Apr 30 01:58:44 | FINISHED | SUCCEEDED |
| application_1492438200769_0220 | 2017st15 | InvertedIndex | MAPREDUCE | default | Sun Apr 30 01:40:17 | Sun Apr 30 01:40:44 | FINISHED | FAILED |
| application_1492438200769_0219 | 2017st03 | Invert index | MAPREDUCE | default | Sun Apr 30 01:40:17 | Sun Apr 30 01:40:44 | FINISHED | SUCCEEDED |

倒排索引:

Counters for job_1492438200769_0221 - Mozilla Firefox

114.212.190.91:19888/jobhistory/jobcounters/job_1492438200769_0221

hadoop

Counters for job_1492438200769_0221

Logged in as: drwho

Application

Job

Overview

Counters

Configuration

Map tasks

Reduce tasks

Tools

| Counter Group | Name | Map | Reduce | Total |
|----------------------|--|-------------|-------------|----------------|
| File System Counters | FILE: Number of bytes read | 0 | 21,961,629 | 21,961,629 |
| | FILE: Number of bytes written | 48,216,124 | 22,077,857 | 70,293,981 |
| | FILE: Number of large read operations | 0 | 0 | 0 |
| | FILE: Number of read operations | 0 | 0 | 0 |
| | FILE: Number of write operations | 0 | 0 | 0 |
| | HDFS: Number of bytes read | 268,310,944 | 0 | 268,310,944 |
| | HDFS: Number of bytes written | 0 | 105,642,849 | 105,642,849 |
| | HDFS: Number of large read operations | 0 | 0 | 0 |
| | HDFS: Number of read operations | 654 | 3 | 657 |
| | HDFS: Number of write operations | 0 | 2 | 2 |
| Job Counters | Data-local map tasks | 0 | 0 | 214 |
| | Killed map tasks | 0 | 0 | 1 |
| | Launched map tasks | 0 | 0 | 219 |
| | Launched reduce tasks | 0 | 0 | 1 |
| | Rack-local map tasks | 0 | 0 | 5 |
| | Total megabyte-seconds taken by all map tasks | 0 | 0 | 11,360,534,528 |
| | Total megabyte-seconds taken by all reduce tasks | 0 | 0 | 567,091,200 |
| | Total time spent by all map tasks (ms) | 0 | 0 | 1,386,784 |
| | Total time spent by all maps in occupied slots (ms) | 0 | 0 | 2,773,568 |
| | Total time spent by all reduce tasks (ms) | 0 | 0 | 69,225 |
| | Total time spent by all reduces in occupied slots (ms) | 0 | 0 | 138,450 |
| | Total vcore-seconds taken by all map tasks | 0 | 0 | 1,386,784 |
| | Total vcore-seconds taken by all reduce tasks | 0 | 0 | 69,225 |
| | | 0 | 0 | |
| | | 0 | 0 | |

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Choose What I Share

频率排序:

Counters for job_1492438200769_0222 - Mozilla Firefox

114.212.190.91:19888/jobhistory/jobcounters/job_1492438200769_0222

hadoop

Counters for job_1492438200769_0222

Logged in as: drwho

Application

Job

Overview

Counters

Configuration

Map tasks

Reduce tasks

Tools

| Counter Group | Name | Map | Reduce | Total |
|----------------------|--|-------------|-------------|-------------|
| File System Counters | FILE: Number of bytes read | 14,415,305 | 14,391,899 | 28,807,204 |
| | FILE: Number of bytes written | 28,923,155 | 14,507,803 | 43,430,958 |
| | FILE: Number of large read operations | 0 | 0 | 0 |
| | FILE: Number of read operations | 0 | 0 | 0 |
| | FILE: Number of write operations | 0 | 0 | 0 |
| | HDFS: Number of bytes read | 105,642,968 | 0 | 105,642,968 |
| | HDFS: Number of bytes written | 0 | 105,575,612 | 105,575,612 |
| | HDFS: Number of large read operations | 0 | 0 | 0 |
| | HDFS: Number of read operations | 3 | 3 | 6 |
| | HDFS: Number of write operations | 0 | 2 | 2 |
| Job Counters | Launched map tasks | 0 | 0 | 1 |
| | Launched reduce tasks | 0 | 0 | 1 |
| | Rack-local map tasks | 0 | 0 | 1 |
| | Total megabyte-seconds taken by all map tasks | 0 | 0 | 199,499,776 |
| | Total megabyte-seconds taken by all reduce tasks | 0 | 0 | 150,052,864 |
| | Total time spent by all map tasks (ms) | 0 | 0 | 24,353 |
| | Total time spent by all maps in occupied slots (ms) | 0 | 0 | 48,706 |
| | Total time spent by all reduce tasks (ms) | 0 | 0 | 18,317 |
| | Total time spent by all reduces in occupied slots (ms) | 0 | 0 | 36,634 |
| | Total vcore-seconds taken by all map tasks | 0 | 0 | 24,353 |
| | Total vcore-seconds taken by all reduce tasks | 0 | 0 | 18,317 |
| | | 0 | 0 | |
| | | 0 | 0 | |
| | | 0 | 0 | |

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Choose What I Share

TF-IDF:

Counters for job_1492438200769_0224 - Mozilla Firefox

114.212.190.91:19888/jobhistory/jobcounters/job_1492438200769_0224

hadoop

Counters for job_1492438200769_0224

Logged in as: drwho

Application

- Job
 - Overview
 - Counters
 - Configuration
 - Map tasks
 - Reduce tasks
- Tools

| Counter Group | Name | Map | Reduce | Total |
|----------------------|--|-------------|------------|-------------|
| File System Counters | FILE: Number of bytes read | 0 | 3,274,380 | 3,274,380 |
| | FILE: Number of bytes written | 3,390,295 | 3,390,248 | 6,780,543 |
| | FILE: Number of large read operations | 0 | 0 | 0 |
| | FILE: Number of read operations | 0 | 0 | 0 |
| | FILE: Number of write operations | 0 | 0 | 0 |
| | HDFS: Number of bytes read | 105,642,968 | 0 | 105,642,968 |
| | HDFS: Number of bytes written | 0 | 16,013,861 | 16,013,861 |
| | HDFS: Number of large read operations | 0 | 0 | 0 |
| | HDFS: Number of read operations | 3 | 3 | 6 |
| | HDFS: Number of write operations | 0 | 2 | 2 |
| Job Counters | Data-local map tasks | 0 | 0 | 1 |
| | Launched map tasks | 0 | 0 | 1 |
| | Launched reduce tasks | 0 | 0 | 1 |
| | Total megabyte-seconds taken by all map tasks | 0 | 0 | 75,554,816 |
| | Total megabyte-seconds taken by all reduce tasks | 0 | 0 | 64,454,656 |
| | Total time spent by all map tasks (ms) | 0 | 0 | 9,223 |
| | Total time spent by all maps in occupied slots (ms) | 0 | 0 | 18,446 |
| | Total time spent by all reduce tasks (ms) | 0 | 0 | 7,868 |
| | Total time spent by all reduces in occupied slots (ms) | 0 | 0 | 15,736 |
| | Total vcore-seconds taken by all map tasks | 0 | 0 | 9,223 |
| | Total vcore-seconds taken by all reduce tasks | 0 | 0 | 7,868 |
| | Combine input records | 0 | 0 | 0 |
| | Combine output records | 0 | 0 | 0 |
| | Combine input records | 0 | 0 | 0 |
| | Combine output records | 0 | 0 | 0 |

Firefox automatically sends some data to Mozilla so that we can improve your experience. Choose What I Share