

1、 系统安装运行情况

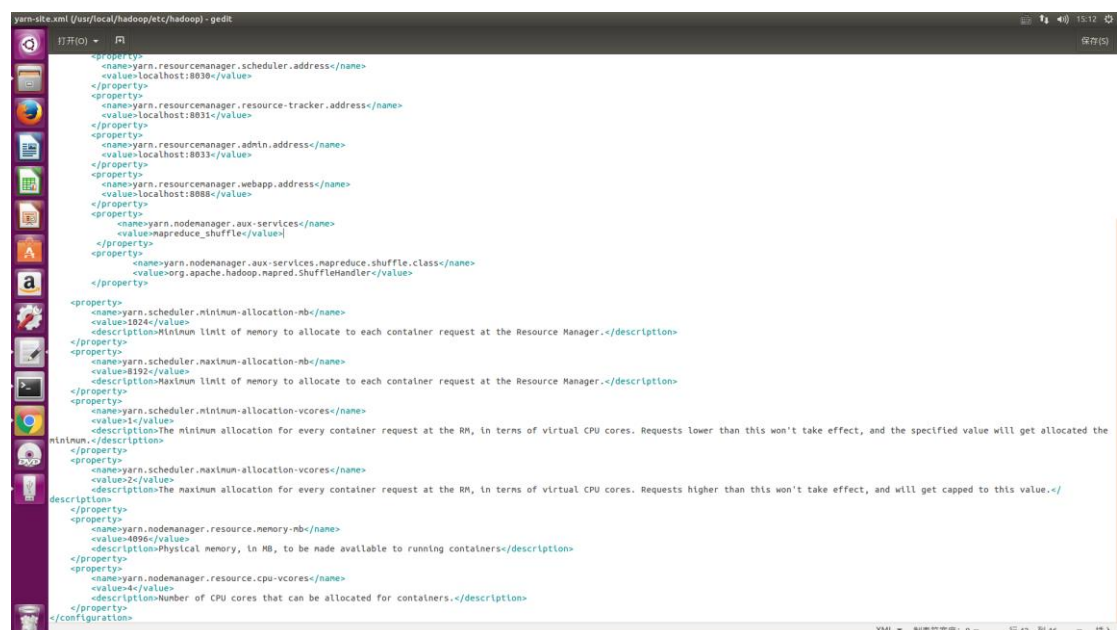
etc/hadoop/core-site.xml:

```
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/usr/local/hadoop/tmp</value>
    <description>Abase for other temporary directories.</description>
  </property>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
  <property>
    <name>hadoop.native.lib</name>
    <value>true</value>
    <description>Should native hadoop libraries, if present, be used.</description>
  </property>
</configuration>
```

etc/hadoop/hdfs-site.xml:

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/usr/local/hadoop/tmp/dfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/usr/local/hadoop/tmp/dfs/data</value>
  </property>
</configuration>
```

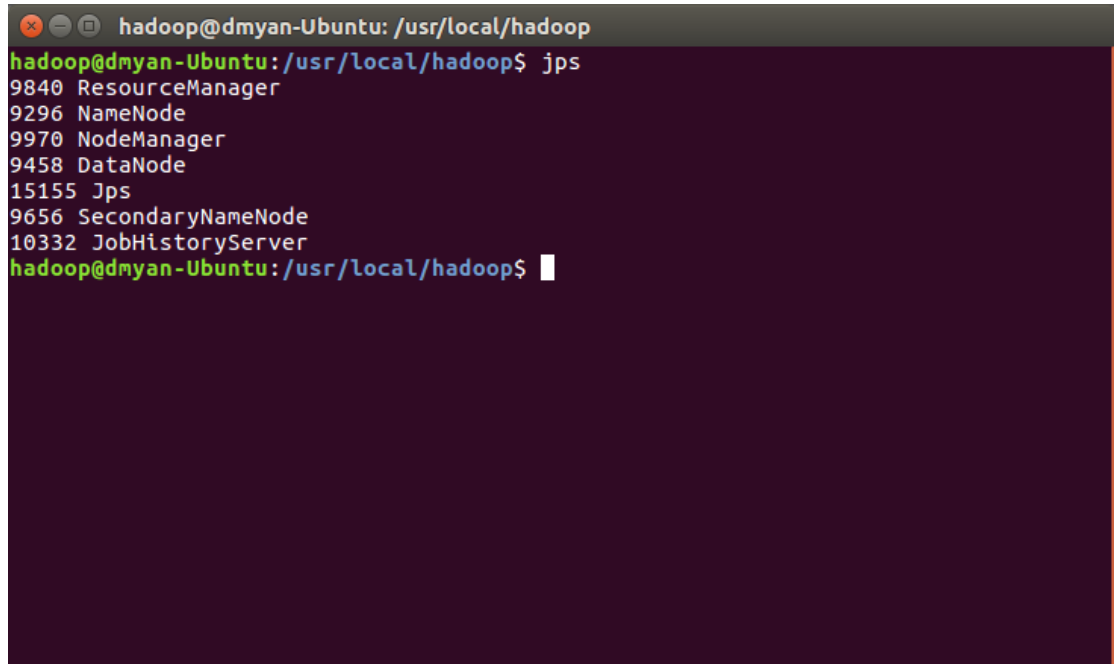
etc/hadoop/yarn-site.xml:



etc/hadoop/mapred-site.xml:

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

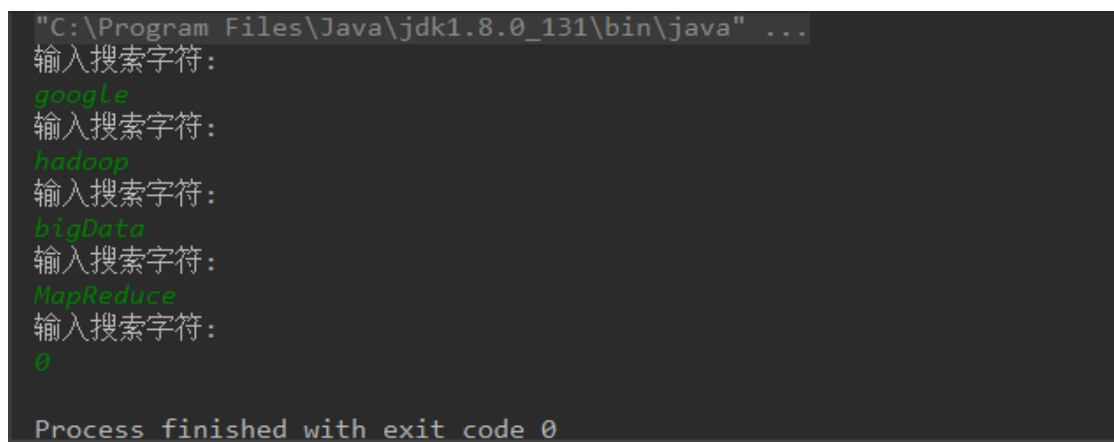
运行情况：

A terminal window titled 'hadoop@dmyan-Ubuntu: /usr/local/hadoop' shows the output of the 'jps' command. The output lists several Hadoop processes and their corresponding PIDs: 9840 ResourceManager, 9296 NameNode, 9970 NodeManager, 9458 DataNode, 15155 Jps, 9656 SecondaryNameNode, and 10332 JobHistoryServer. The prompt 'hadoop@dmyan-Ubuntu: /usr/local/hadoop\$' is visible at the bottom.





```
hadoop@dmyan-Ubuntu: /usr/local/hadoop$ jps
9840 ResourceManager
9296 NameNode
9970 NodeManager
9458 DataNode
15155 Jps
9656 SecondaryNameNode
10332 JobHistoryServer
hadoop@dmyan-Ubuntu: /usr/local/hadoop$
```

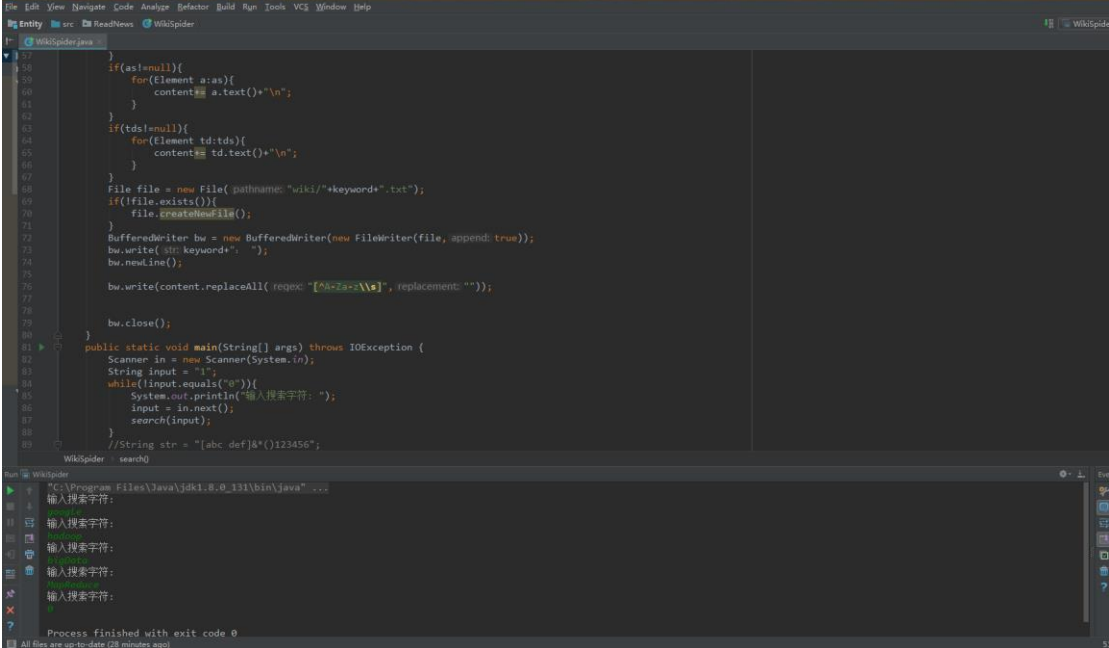
2、 实验数据说明：

从 wikipedia 网页中抓取了关于 bigData、google、hadoop、MapReduce 等关键词的英文网页中的单词数据

A terminal window shows a series of prompts '输入搜索字符:' followed by the words 'google', 'hadoop', 'bigData', 'MapReduce', and '0'. The process concludes with the message 'Process finished with exit code 0'.

```
"C:\Program Files\Java\jdk1.8.0_131\bin\java" ...
输入搜索字符:
google
输入搜索字符:
hadoop
输入搜索字符:
bigData
输入搜索字符:
MapReduce
输入搜索字符:
0
Process finished with exit code 0
```

 bigData.txt	2017/10/18 14:55	文本文档	53 KB
 google.txt	2017/10/18 14:57	文本文档	109 KB
 hadoop.txt	2017/10/18 14:55	文本文档	41 KB
 MapReduce.txt	2017/10/18 14:55	文本文档	31 KB



```

1  Wikispider.java
2
3  // 从文件中读取关键词
4
5  // 从文件中读取关键词
6
7  // 从文件中读取关键词
8
9  // 从文件中读取关键词
10
11 // 从文件中读取关键词
12
13 // 从文件中读取关键词
14
15 // 从文件中读取关键词
16
17 // 从文件中读取关键词
18
19 // 从文件中读取关键词
20
21 // 从文件中读取关键词
22
23 // 从文件中读取关键词
24
25 // 从文件中读取关键词
26
27 // 从文件中读取关键词
28
29 // 从文件中读取关键词
30
31 // 从文件中读取关键词
32
33 // 从文件中读取关键词
34
35 // 从文件中读取关键词
36
37 // 从文件中读取关键词
38
39 // 从文件中读取关键词
40
41 // 从文件中读取关键词
42
43 // 从文件中读取关键词
44
45 // 从文件中读取关键词
46
47 // 从文件中读取关键词
48
49 // 从文件中读取关键词
50
51 // 从文件中读取关键词
52
53 // 从文件中读取关键词
54
55 // 从文件中读取关键词
56
57 // 从文件中读取关键词
58
59 // 从文件中读取关键词
60
61 // 从文件中读取关键词
62
63 // 从文件中读取关键词
64
65 // 从文件中读取关键词
66
67 // 从文件中读取关键词
68
69 // 从文件中读取关键词
70
71 // 从文件中读取关键词
72
73 // 从文件中读取关键词
74
75 // 从文件中读取关键词
76
77 // 从文件中读取关键词
78
79 // 从文件中读取关键词
80
81 // 从文件中读取关键词
82
83 // 从文件中读取关键词
84
85 // 从文件中读取关键词
86
87 // 从文件中读取关键词
88
89 // 从文件中读取关键词
90
91 // 从文件中读取关键词
92
93 // 从文件中读取关键词
94
95 // 从文件中读取关键词
96
97 // 从文件中读取关键词
98
99 // 从文件中读取关键词
100
101 // 从文件中读取关键词
102
103 // 从文件中读取关键词
104
105 // 从文件中读取关键词
106
107 // 从文件中读取关键词
108
109 // 从文件中读取关键词
110
111 // 从文件中读取关键词
112
113 // 从文件中读取关键词
114
115 // 从文件中读取关键词
116
117 // 从文件中读取关键词
118
119 // 从文件中读取关键词
120
121 // 从文件中读取关键词
122
123 // 从文件中读取关键词
124
125 // 从文件中读取关键词
126
127 // 从文件中读取关键词
128
129 // 从文件中读取关键词
130
131 // 从文件中读取关键词
132
133 // 从文件中读取关键词
134
135 // 从文件中读取关键词
136
137 // 从文件中读取关键词
138
139 // 从文件中读取关键词
140
141 // 从文件中读取关键词
142
143 // 从文件中读取关键词
144
145 // 从文件中读取关键词
146
147 // 从文件中读取关键词
148
149 // 从文件中读取关键词
150
151 // 从文件中读取关键词
152
153 // 从文件中读取关键词
154
155 // 从文件中读取关键词
156
157 // 从文件中读取关键词
158
159 // 从文件中读取关键词
160
161 // 从文件中读取关键词
162
163 // 从文件中读取关键词
164
165 // 从文件中读取关键词
166
167 // 从文件中读取关键词
168
169 // 从文件中读取关键词
170
171 // 从文件中读取关键词
172
173 // 从文件中读取关键词
174
175 // 从文件中读取关键词
176
177 // 从文件中读取关键词
178
179 // 从文件中读取关键词
180
181 // 从文件中读取关键词
182
183 // 从文件中读取关键词
184
185 // 从文件中读取关键词
186
187 // 从文件中读取关键词
188
189 // 从文件中读取关键词
190
191 // 从文件中读取关键词
192
193 // 从文件中读取关键词
194
195 // 从文件中读取关键词
196
197 // 从文件中读取关键词
198
199 // 从文件中读取关键词
200
201 // 从文件中读取关键词
202
203 // 从文件中读取关键词
204
205 // 从文件中读取关键词
206
207 // 从文件中读取关键词
208
209 // 从文件中读取关键词
210
211 // 从文件中读取关键词
212
213 // 从文件中读取关键词
214
215 // 从文件中读取关键词
216
217 // 从文件中读取关键词
218
219 // 从文件中读取关键词
220
221 // 从文件中读取关键词
222
223 // 从文件中读取关键词
224
225 // 从文件中读取关键词
226
227 // 从文件中读取关键词
228
229 // 从文件中读取关键词
230
231 // 从文件中读取关键词
232
233 // 从文件中读取关键词
234
235 // 从文件中读取关键词
236
237 // 从文件中读取关键词
238
239 // 从文件中读取关键词
240
241 // 从文件中读取关键词
242
243 // 从文件中读取关键词
244
245 // 从文件中读取关键词
246
247 // 从文件中读取关键词
248
249 // 从文件中读取关键词
250
251 // 从文件中读取关键词
252
253 // 从文件中读取关键词
254
255 // 从文件中读取关键词
256
257 // 从文件中读取关键词
258
259 // 从文件中读取关键词
260
261 // 从文件中读取关键词
262
263 // 从文件中读取关键词
264
265 // 从文件中读取关键词
266
267 // 从文件中读取关键词
268
269 // 从文件中读取关键词
270
271 // 从文件中读取关键词
272
273 // 从文件中读取关键词
274
275 // 从文件中读取关键词
276
277 // 从文件中读取关键词
278
279 // 从文件中读取关键词
280
281 // 从文件中读取关键词
282
283 // 从文件中读取关键词
284
285 // 从文件中读取关键词
286
287 // 从文件中读取关键词
288
289 // 从文件中读取关键词
290
291 // 从文件中读取关键词
292
293 // 从文件中读取关键词
294
295 // 从文件中读取关键词
296
297 // 从文件中读取关键词
298
299 // 从文件中读取关键词
300
301 // 从文件中读取关键词
302
303 // 从文件中读取关键词
304
305 // 从文件中读取关键词
306
307 // 从文件中读取关键词
308
309 // 从文件中读取关键词
310
311 // 从文件中读取关键词
312
313 // 从文件中读取关键词
314
315 // 从文件中读取关键词
316
317 // 从文件中读取关键词
318
319 // 从文件中读取关键词
320
321 // 从文件中读取关键词
322
323 // 从文件中读取关键词
324
325 // 从文件中读取关键词
326
327 // 从文件中读取关键词
328
329 // 从文件中读取关键词
330
331 // 从文件中读取关键词
332
333 // 从文件中读取关键词
334
335 // 从文件中读取关键词
336
337 // 从文件中读取关键词
338
339 // 从文件中读取关键词
340
341 // 从文件中读取关键词
342
343 // 从文件中读取关键词
344
345 // 从文件中读取关键词
346
347 // 从文件中读取关键词
348
349 // 从文件中读取关键词
350
351 // 从文件中读取关键词
352
353 // 从文件中读取关键词
354
355 // 从文件中读取关键词
356
357 // 从文件中读取关键词
358
359 // 从文件中读取关键词
360
361 // 从文件中读取关键词
362
363 // 从文件中读取关键词
364
365 // 从文件中读取关键词
366
367 // 从文件中读取关键词
368
369 // 从文件中读取关键词
370
371 // 从文件中读取关键词
372
373 // 从文件中读取关键词
374
375 // 从文件中读取关键词
376
377 // 从文件中读取关键词
378
379 // 从文件中读取关键词
380
381 // 从文件中读取关键词
382
383 // 从文件中读取关键词
384
385 // 从文件中读取关键词
386
387 // 从文件中读取关键词
388
389 // 从文件中读取关键词
390
391 // 从文件中读取关键词
392
393 // 从文件中读取关键词
394
395 // 从文件中读取关键词
396
397 // 从文件中读取关键词
398
399 // 从文件中读取关键词
400
401 // 从文件中读取关键词
402
403 // 从文件中读取关键词
404
405 // 从文件中读取关键词
406
407 // 从文件中读取关键词
408
409 // 从文件中读取关键词
410
411 // 从文件中读取关键词
412
413 // 从文件中读取关键词
414
415 // 从文件中读取关键词
416
417 // 从文件中读取关键词
418
419 // 从文件中读取关键词
420
421 // 从文件中读取关键词
422
423 // 从文件中读取关键词
424
425 // 从文件中读取关键词
426
427 // 从文件中读取关键词
428
429 // 从文件中读取关键词
430
431 // 从文件中读取关键词
432
433 // 从文件中读取关键词
434
435 // 从文件中读取关键词
436
437 // 从文件中读取关键词
438
439 // 从文件中读取关键词
440
441 // 从文件中读取关键词
442
443 // 从文件中读取关键词
444
445 // 从文件中读取关键词
446
447 // 从文件中读取关键词
448
449 // 从文件中读取关键词
450
451 // 从文件中读取关键词
452
453 // 从文件中读取关键词
454
455 // 从文件中读取关键词
456
457 // 从文件中读取关键词
458
459 // 从文件中读取关键词
460
461 // 从文件中读取关键词
462
463 // 从文件中读取关键词
464
465 // 从文件中读取关键词
466
467 // 从文件中读取关键词
468
469 // 从文件中读取关键词
470
471 // 从文件中读取关键词
472
473 // 从文件中读取关键词
474
475 // 从文件中读取关键词
476
477 // 从文件中读取关键词
478
479 // 从文件中读取关键词
480
481 // 从文件中读取关键词
482
483 // 从文件中读取关键词
484
485 // 从文件中读取关键词
486
487 // 从文件中读取关键词
488
489 // 从文件中读取关键词
490
491 // 从文件中读取关键词
492
493 // 从文件中读取关键词
494
495 // 从文件中读取关键词
496
497 // 从文件中读取关键词
498
499 // 从文件中读取关键词
500
501 // 从文件中读取关键词
502
503 // 从文件中读取关键词
504
505 // 从文件中读取关键词
506
507 // 从文件中读取关键词
508
509 // 从文件中读取关键词
510
511 // 从文件中读取关键词
512
513 // 从文件中读取关键词
514
515 // 从文件中读取关键词
516
517 // 从文件中读取关键词
518
519 // 从文件中读取关键词
520
521 // 从文件中读取关键词
522
523 // 从文件中读取关键词
524
525 // 从文件中读取关键词
526
527 // 从文件中读取关键词
528
529 // 从文件中读取关键词
530
531 // 从文件中读取关键词
532
533 // 从文件中读取关键词
534
535 // 从文件中读取关键词
536
537 // 从文件中读取关键词
538
539 // 从文件中读取关键词
540
541 // 从文件中读取关键词
542
543 // 从文件中读取关键词
544
545 // 从文件中读取关键词
546
547 // 从文件中读取关键词
548
549 // 从文件中读取关键词
550
551 // 从文件中读取关键词
552
553 // 从文件中读取关键词
554
555 // 从文件中读取关键词
556
557 // 从文件中读取关键词
558
559 // 从文件中读取关键词
560
561 // 从文件中读取关键词
562
563 // 从文件中读取关键词
564
565 // 从文件中读取关键词
566
567 // 从文件中读取关键词
568
569 // 从文件中读取关键词
570
571 // 从文件中读取关键词
572
573 // 从文件中读取关键词
574
575 // 从文件中读取关键词
576
577 // 从文件中读取关键词
578
579 // 从文件中读取关键词
580
581 // 从文件中读取关键词
582
583 // 从文件中读取关键词
584
585 // 从文件中读取关键词
586
587 // 从文件中读取关键词
588
589 // 从文件中读取关键词
590
591 // 从文件中读取关键词
592
593 // 从文件中读取关键词
594
595 // 从文件中读取关键词
596
597 // 从文件中读取关键词
598
599 // 从文件中读取关键词
600
601 // 从文件中读取关键词
602
603 // 从文件中读取关键词
604
605 // 从文件中读取关键词
606
607 // 从文件中读取关键词
608
609 // 从文件中读取关键词
610
611 // 从文件中读取关键词
612
613 // 从文件中读取关键词
614
615 // 从文件中读取关键词
616
617 // 从文件中读取关键词
618
619 // 从文件中读取关键词
620
621 // 从文件中读取关键词
622
623 // 从文件中读取关键词
624
625 // 从文件中读取关键词
626
627 // 从文件中读取关键词
628
629 // 从文件中读取关键词
630
631 // 从文件中读取关键词
632
633 // 从文件中读取关键词
634
635 // 从文件中读取关键词
636
637 // 从文件中读取关键词
638
639 // 从文件中读取关键词
640
641 // 从文件中读取关键词
642
643 // 从文件中读取关键词
644
645 // 从文件中读取关键词
646
647 // 从文件中读取关键词
648
649 // 从文件中读取关键词
650
651 // 从文件中读取关键词
652
653 // 从文件中读取关键词
654
655 // 从文件中读取关键词
656
657 // 从文件中读取关键词
658
659 // 从文件中读取关键词
660
661 // 从文件中读取关键词
662
663 // 从文件中读取关键词
664
665 // 从文件中读取关键词
666
667 // 从文件中读取关键词
668
669 // 从文件中读取关键词
670
671 // 从文件中读取关键词
672
673 // 从文件中读取关键词
674
675 // 从文件中读取关键词
676
677 // 从文件中读取关键词
678
679 // 从文件中读取关键词
680
681 // 从文件中读取关键词
682
683 // 从文件中读取关键词
684
685 // 从文件中读取关键词
686
687 // 从文件中读取关键词
688
689 // 从文件中读取关键词
690
691 // 从文件中读取关键词
692
693 // 从文件中读取关键词
694
695 // 从文件中读取关键词
696
697 // 从文件中读取关键词
698
699 // 从文件中读取关键词
700
701 // 从文件中读取关键词
702
703 // 从文件中读取关键词
704
705 // 从文件中读取关键词
706
707 // 从文件中读取关键词
708
709 // 从文件中读取关键词
710
711 // 从文件中读取关键词
712
713 // 从文件中读取关键词
714
715 // 从文件中读取关键词
716
717 // 从文件中读取关键词
718
719 // 从文件中读取关键词
720
721 // 从文件中读取关键词
722
723 // 从文件中读取关键词
724
725 // 从文件中读取关键词
726
727 // 从文件中读取关键词
728
729 // 从文件中读取关键词
730
731 // 从文件中读取关键词
732
733 // 从文件中读取关键词
734
735 // 从文件中读取关键词
736
737 // 从文件中读取关键词
738
739 // 从文件中读取关键词
740
741 // 从文件中读取关键词
742
743 // 从文件中读取关键词
744
745 // 从文件中读取关键词
746
747 // 从文件中读取关键词
748
749 // 从文件中读取关键词
750
751 // 从文件中读取关键词
752
753 // 从文件中读取关键词
754
755 // 从文件中读取关键词
756
757 // 从文件中读取关键词
758
759 // 从文件中读取关键词
760
761 // 从文件中读取关键词
762
763 // 从文件中读取关键词
764
765 // 从文件中读取关键词
766
767 // 从文件中读取关键词
768
769 // 从文件中读取关键词
770
771 // 从文件中读取关键词
772
773 // 从文件中读取关键词
774
775 // 从文件中读取关键词
776
777 // 从文件中读取关键词
778
779 // 从文件中读取关键词
780
781 // 从文件中读取关键词
782
783 // 从文件中读取关键词
784
785 // 从文件中读取关键词
786
787 // 从文件中读取关键词
788
789 // 从文件中读取关键词
790
791 // 从文件中读取关键词
792
793 // 从文件中读取关键词
794
795 // 从文件中读取关键词
796
797 // 从文件中读取关键词
798
799 // 从文件中读取关键词
800
801 // 从文件中读取关键词
802
803 // 从文件中读取关键词
804
805 // 从文件中读取关键词
806
807 // 从文件中读取关键词
808
809 // 从文件中读取关键词
810
811 // 从文件中读取关键词
812
813 // 从文件中读取关键词
814
815 // 从文件中读取关键词
816
817 // 从文件中读取关键词
818
819 // 从文件中读取关键词
820
821 // 从文件中读取关键词
822
823 // 从文件中读取关键词
824
825 // 从文件中读取关键词
826
827 // 从文件中读取关键词
828
829 // 从文件中读取关键词
830
831 // 从文件中读取关键词
832
833 // 从文件中读取关键词
834
835 // 从文件中读取关键词
836
837 // 从文件中读取关键词
838
839 // 从文件中读取关键词
840
841 // 从文件中读取关键词
842
843 // 从文件中读取关键词
844
845 // 从文件中读取关键词
846
847 // 从文件中读取关键词
848
849 // 从文件中读取关键词
850
851 // 从文件中读取关键词
852
853 // 从文件中读取关键词
854
855 // 从文件中读取关键词
856
857 // 从文件中读取关键词
858
859 // 从文件中读取关键词
860
861 // 从文件中读取关键词
862
863 // 从文件中读取关键词
864
865 // 从文件中读取关键词
866
867 // 从文件中读取关键词
868
869 // 从文件中读取关键词
870
871 // 从文件中读取关键词
872
873 // 从文件中读取关键词
874
875 // 从文件中读取关键词
876
877 // 从文件中读取关键词
878
879 // 从文件中读取关键词
880
881 // 从文件中读取关键词
882
883 // 从文件中读取关键词
884
885 // 从文件中读取关键词
886
887 // 从文件中读取关键词
888
889 // 从文件中读取关键词
890
891 // 从文件中读取关键词
892
893 // 从文件中读取关键词
894
895 // 从文件中读取关键词
896
897 // 从文件中读取关键词
898
899 // 从文件中读取关键词
900
901 // 从文件中读取关键词
902
903 // 从文件中读取关键词
904
905 // 从文件中读取关键词
906
907 // 从文件中读取关键词
908
909 // 从文件中读取关键词
910
911 // 从文件中读取关键词
912
913 // 从文件中读取关键词
914
915 // 从文件中读取关键词
916
917 // 从文件中读取关键词
918
919 // 从文件中读取关键词
920
921 // 从文件中读取关键词
922
923 // 从文件中读取关键词
924
925 // 从文件中读取关键词
926
927 // 从文件中读取关键词
928
929 // 从文件中读取关键词
930
931 // 从文件中读取关键词
932
933 // 从文件中读取关键词
934
935 // 从文件中读取关键词
936
937 // 从文件中读取关键词
938
939 // 从文件中读取关键词
940
941 // 从文件中读取关键词
942
943 // 从文件中读取关键词
944
945 // 从文件中读取关键词
946
947 // 从文件中读取关键词
948
949 // 从文件中读取关键词
950
951 // 从文件中读取关键词
952
953 // 从文件中读取关键词
954
955 // 从文件中读取关键词
956
957 // 从文件中读取关键词
958
959 // 从文件中读取关键词
960
961 // 从文件中读取关键词
962
963 // 从文件中读取关键词
964
965 // 从文件中读取关键词
966
967 // 从文件中读取关键词
968
969 // 从文件中读取关键词
970
971 // 从文件中读取关键词
972
973 // 从文件中读取关键词
974
975 // 从文件中读取关键词
976
977 // 从文件中读取关键词
978
979 // 从文件中读取关键词
980
981 // 从文件中读取关键词
982
983 // 从文件中读取关键词
984
985 // 从文件中读取关键词
986
987 // 从文件中读取关键词
988
989 // 从文件中读取关键词
990
991 // 从文件中读取关键词
992
993 // 从文件中读取关键词
994
995 // 从文件中读取关键词
996
997 // 从文件中读取关键词
998
999 // 从文件中读取关键词
1000

```

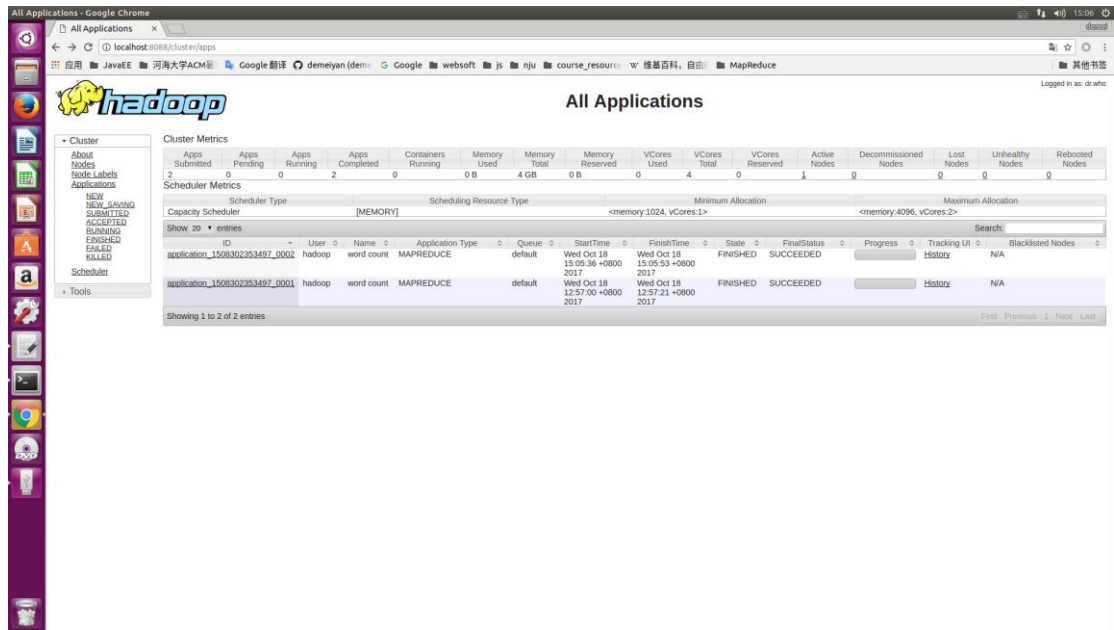
下面是关键词为 google 的 wikipedia 上词汇抓取结果（关键词 google 的结果一共 3425 行单词数据）：

```

1 google:
2 Google Inc is an American multinational technology company that specializes in Internet-related services and products These include online advertising technologies search cloud computing software and hardware Google was foun
3 The company's rapid growth since incorporation has triggered a chain of products acquisitions and partnerships beyond Google's core search engine Google Search It offers services designed for work and productivity Google Docs
4 Alexa a company that monitors commercial web traffic lists Google.com as the most visited website in the world Several other Google services also figure in the top most visited websites including YouTube and Blogger Google
5
6
7 Google began in January as a research project by Larry Page and Sergey Brin when they were both PhD students at Stanford University in Stanford California
8 While conventional search engines ranked results by counting how many times the search term appeared on the page the two theorized about a better system that analyzed the relationships among websites They called this new t
9 Page and Brin originally nicknamed their new search engine BackRub because the system checked backlinks to estimate the importance of a site Eventually they changed the name to Google the name of the search engine originate
10 The domain name for Google was registered on September 7 and the company was incorporated on September 7 It was based in the garage of a friend Susan Wojcicki in Menlo Park California Craig Silverstein a fellow PhD student
11 Google was initially funded by an August contribution of from Andy Bechtolsheim cofounder of Sun Microsystems the money was given before Google was incorporated Google received money from three other angel investors in A
12 After some additional small investments through the end of to early a new million round of funding was announced on June with major investors including the venture capital firms Kleiner Perkins Caufield Byers and Sequ
13 Early in Brin and Page decided they wanted to sell Google to Excite They went to Excite CEO George Bell and offered to sell it to him for million He rejected the offer Vinod Khosla one of Excite's venture capitalists take
14 Google's initial public offering IPO took place five years later on August At that time Larry Page Sergey Brin and Eric Schmidt agreed to work together at Google for years until the year
15 At IPO the company offered shares at a price of per share Shares were sold in an online auction format using a system built by Morgan Stanley and Credit Suisse underwriters for the deal The sale of an billion gave Google
16 There were concerns that Google's IPO would lead to changes in company culture Reasons ranged from shareholder pressure for employee benefit reductions to the fact that many company executives would become instant paper mill
17 The stock performed well after the first time on October primarily because of strong sales and earnings in the online advertising market The surge in stock price was fueled mainly by indiv
18 In March the company moved its office to Palo Alto California which is home to several prominent Silicon Valley technology startups The next year Google began selling advertisements associated with search keywords against
19 This model of selling keyword advertising was first pioneered by Overture an Idealab spinoff created by Bill Gross when the company changed names to Overture Services It used Google over alleged infringements of the company
20 In Google received a patent for its PageRank mechanism The patent was officially assigned to Stanford University and lists Lawrence Page as the inventor In after outgrowing the other locations the company leased an office
21 In The Washington Post reported on a percent increase in third-quarter profit for Google largely thanks to large companies shifting their advertising strategies from newspapers magazines and television to the Internet In 3
22 The year was the first time that Google generated billion in annual revenue which topped the billion that was generated the previous year In January then-CEO Larry Page commented He ended with a strong quarter Revenues
23 Google announced the launch of a new company called Calico on September to be led by Apple Inc chairman Arthur Levinson In the official public statement Page explained that the health and wellbeing company would focus on
24 Google celebrated its year anniversary on September and in it celebrated its birthday with an animated Google shown on web browsers around the world although it has used other dates for its official birthday The reason
25 The Alliance for Affordable Internet AAI was launched in October Google is part of the coalition of public and private organizations that also includes Facebook Intel and Microsoft Led by Sir Tim Berners-Lee the AAI seeks t
26 The corporations consolidated revenue for the third quarter of was reported in mid-October as billion a percent increase compared to the previous quarter Google's Internet business was responsible for billion of this tot
27 According to Interbrand's annual Best Global Brands report Google has been the second most valuable brand in the world behind Apple Inc in and with a valuation of billion
28 In September Google engineering manager Rachel Potvin revealed details about Google's software code at an engineering conference She revealed that the entire Google codebase which spans every single service it develops cons
29 As of October Google operates offices in more than countries Alexa a company that monitors commercial web traffic lists Google.com as the most visited website in the world Several other Google services also figure in the
30 In Google acquired Beta News the operators of a large archive of materials from Usenet Google rebranded the archive as Google Groups and by the end of the year it had expanded the history back to
31 In April Google acquired Applied Semantics a company specializing in making software applications for the online advertising space The AdSense contextual advertising technology developed by Applied Semantics was adopted in
32 In Google acquired Keyhole Inc Keyhole's eponymous product was later renamed Google Earth
33 In Google acquired iStock Software in April using their iStock on Demand product along with ideas from Adaptive Paths Measure Map to create Google Analytics in
34 In October Google announced that it had acquired the video-sharing site YouTube for billion in Google stock and the deal was finalized on November
35 On April Google reached an agreement to acquire DoubleClick for billion transferring to Google valuable relationships that DoubleClick had with web publishers and advertising agencies The deal was approved despite antitru
36 In addition to the many companies Google has purchased the firm has partnered with other organizations for research advertising and other activities In Google partnered with NASA Ames Research Center to build square feet
37 In Google partnered with AOL to enhance each others video search services In Google and Fox Interactive Media of News Corporation entered into a million agreement to provide search and advertising on the then-popular socia
38 In Google began sponsoring NASCAR Tracks Santa dislodging the former sponsor AOL NASCAR Tracks Santa purports to follow Santa Claus progress on Christmas Eve using Google Earth to track Santa in 3D for the first time
39 In Google developed a partnership with Google to launch a satellite providing Google with high-resolution monochrome color imagery for Google Earth The satellite was launched from Vandenberg Air Force Base on September
40 In Google Energy made its first investment in a renewable energy project putting million into two wind farms in North Dakota The company announced the two locations will generate megawatts of power enough to supply homes
41 Also in Google purchased Global IP Solutions a Norway-based company that provides web-based teleconferencing and other related services This acquisition enabled Google to add telephony-style services to its list of products 0
42 On April The Globe and Mail reported that Google bid million for Nortel Networks patents
43 On August Google made its largest-ever acquisition today when it announced that it would acquire Motorola Mobility for billion subject to approval from regulators in the United States and Europe In a post on Google's blog
44 This purchase was made in part to help Google gain Motorola's considerable patent portfolio on mobile phones and wireless technologies to help protect Google in its ongoing patent disputes with other companies mainly Apple a
45 In June Google acquired Waze a million deal While Waze would remain an independent entity its social features such as its crowdsourced location platform were reportedly valuable integrations between Waze and Google Maps 0
46 On January Google announced it had agreed to acquire DeepMind Technologies a privately held artificial intelligence company from London DeepMind describes itself as having the ability to combine the best techniques from m
47 On January Google announced that it would divest Motorola Mobility to Lenovo for billion a fraction of the original billion price paid by Google to acquire the company Google retained all but of Motorola patents and e
48 On September HTC announced a cooperation agreement in which it would sell nonexclusive rights to certain intellectual property as well as smartphone talent to Google for billion
49 As of Google owned and operated nine data centers across North and South America two in Asia and four in Europe
50 In the company had announced plans to build three data centers at a cost of more than million in Asia Singapore Hong Kong and Taiwan and said they would be operational within two years In December Google announced that it
51 In October The Washington Post reported that the US National Security Agency intercepted communications between Google's data centers as part of a program named MUSCULAR This wiretapping was made possible because Google did
52 Google's most efficient data center runs at F C using only fresh air cooling requiring no electrically powered air conditioning the servers run so hot that humans cannot go near them for extended periods
53 An August report estimated that Google had about servers in their data centers based on energy usage The report does state that Google never says how many servers are running in its data centers
54 In December Google announced that starting in it will power all of its data centers as well as all of its offices from renewable energy The commitment will make Google the world's largest corporate buyer of renewable powe
55 On August Google announced plans to reorganize its various interests as a conglomerate called Alphabet Google became Alphabet's leading subsidiary and will continue to be the umbrella company for Alphabet's Internet interes
56 On September Google Inc announced its plans of restructuring as a limited liability company Google LLC as a wholly owned subsidiary of XXVI Holdings Inc which is formed as a subsidiary of Alphabet Inc to hold the equity o
57 For the fiscal year the company reported billion in total advertising revenues and only million in licensing and other revenues In of Google's revenue was derived from its advertising programs In addition to its own algor
58 In Google launched AdSense for mobile taking advantage of the emerging mobile advertising market
59 Google Analytics allows website owners to track where and how people use their website for example by examining click rates for all the links on a page Google advertisements can be placed on third-party websites in a twopart
60 One of the criticisms of this program is the possibility of click fraud which occurs when a person or automated script clicks on advertisements without being interested in the product causing the advertiser to pay money to
61 In February Google stopped showing the advertisements of Oceana a nonprofit organization protesting a major cruise ship's sewage treatment practices Google cited its editorial policy at the time stating Google does not acc
62 According to market research published by comScore in November Google Search is the dominant search engine in the United States market with a market share of Google indexes billions of web pages so that users can search f
63 In The New York Times complained about Google's indexing claiming that Google's caching of content on its site infringed its copyright for the content In this case the United States District Court of Nevada ruled in favor of
64 Google's search engine retained Google's ownership of the data distributed under the same name and terms as the original content

```

3、程序运行后在 Hadoop Web 作业状态查看界面上的作业运行状态



4、实验输出结果开头部分

```
hadoop@dmyan-Ubuntu: /usr/local/hadoop
17/10/18 15:05:33 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
17/10/18 15:05:36 INFO InputFileInputFormat: Total input paths to process : 4
17/10/18 15:05:36 INFO mapreduce.JobSubmitter: number of splits:4
17/10/18 15:05:36 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1508302353497_0002
17/10/18 15:05:36 INFO Impl.YarnClientImpl: Submitted application application_1508302353497_0002
17/10/18 15:05:36 INFO mapreduce.Job: The url to track the job: http://dmyan-Ubuntu:8088/proxy/application_1508302353497_0002/
17/10/18 15:05:36 INFO mapreduce.Job: Running job: job_1508302353497_0002
17/10/18 15:05:42 INFO mapreduce.Job: map 0% reduce 0%
17/10/18 15:05:42 INFO mapreduce.Job: map 50% reduce 0%
17/10/18 15:05:51 INFO mapreduce.Job: map 100% reduce 0%
17/10/18 15:05:55 INFO mapreduce.Job: map 100% reduce 100%
17/10/18 15:05:55 INFO mapreduce.Job: Job_1508302353497_0002 completed successfully
17/10/18 15:05:55 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=142344
  FILE: Number of bytes written=993691
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=218192
  HDFS: Number of bytes written=72037
  HDFS: Number of read operations=15
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2

Job Counters
  Launched map tasks=4
  Launched reduce tasks=1
  Data-local map tasks=4
  Total time spent by all maps in occupied slots (ms)=12848
  Total time spent by all reduces in occupied slots (ms)=2538
  Total time spent by all map tasks (ms)=12848
  Total time spent by all reduce tasks (ms)=2538
  Total vcore-milliseconds taken by all map tasks=12848
  Total vcore-milliseconds taken by all reduce tasks=2538
  Total megabyte-milliseconds taken by all map tasks=13116332
  Total megabyte-milliseconds taken by all reduce tasks=2598912

Map-Reduce Framework
  Map input records=7861
  Map output records=35703
  Map output bytes=377723
  Map output materialized bytes=142362
  Input split bytes=480
  Combine input records=35703
  Combine output records=10216
  Reduce input groups=7020
  Reduce shuffle bytes=142362
  Reduce input records=10216
  Reduce output records=7020
  Spilled Records=26432
  Shuffled Maps=4
  Merged Shuffles=0
  Merged Map outputs=4
  GC time elapsed (ms)=357
  CPU time spent (ms)=4908
  Physical memory (bytes) snapshot=1297881216
  Virtual memory (bytes) snapshot=987468688
  Total committed heap usage (bytes)=95069856

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
```

```
hadoop@dmyan-Ubuntu: /usr/local/hadoop
hadoop@dmyan-Ubuntu: /usr/local/hadoop$ bin/hadoop fs -cat hdfsoutput/part-r-00000
A 57
AAI 2
AA 1
ACCESS 1
ACM 2
ADATA 3
ADPS 3
AFP 4
AFS 2
AG 2
AI 6
Aibased 1
AJAX 4
AKN 3
ALEC 1
ALECS 1
AMPLab 3
ANDSI 1
AOL 30
APFS 6
API 6
APIS 14
APISincluding 1
APR 2
ARCA 3
ARM 3
Archive 2
ASEE 1
ASF 4
ASML 4
ATT 5
AWS 1
AXFS 3
Aardvark 4
Adera 2
About 14
Abroad 1
Absolutely 1
Abstraction 1
Academy 2
Accelerator 5
Accenture 7
Access 11
Accord 1
According 5
Account 4
Accounts 2
Accumulo 4
Accurately 1
Acer 9
Acquire 4
Acquired 1
Acquires 2
Acquisition 2
Acquisitions 3
Across 1
Active 2
ActivemQ 2
Activision 3
Ad 2
AdMob 8
```

5、实验体会

本次实验经过一步步的实践学会了安装 hadoop 平台，简单的使用 Hadoop, Yarn 进行 Hadoop 集群的资源管理，运行 WordCount 程序并进行词频统计。