

Part 1: Simply running Wordcount1.java

2017.02.18 17:47:32 UTC	2017.02.18 17:47:38 UTC	2017.02.18 17:48:23 UTC	job_1487438491589_0002	word count	vagrant	default	SUCCEEDED	7	7	1	1
-------------------------------	-------------------------------	-------------------------------	------------------------	---------------	---------	---------	-----------	---	---	---	---

Screenshot of the top 10 words



	A	B
1	the	1867
2	to	1433
3	and	1217
4	of	1142
5	a	757
6	our	657
7	in	640
8	that	571
9	we	560
10	for	445

Part 2: Simply running Wordcount2.java

2017.02.18 18:01:47 UTC	2017.02.18 18:01:55 UTC	2017.02.18 18:02:39 UTC	job_1487440869693_0001	word count	vagrant	default	SUCCEEDED	7	7	1	1
-------------------------------	-------------------------------	-------------------------------	------------------------	---------------	---------	---------	-----------	---	---	---	---

Screenshot of the top 10 words



	A	B
1	the	1867
2	to	1433
3	and	1217
4	of	1142
5	a	757
6	our	657
7	in	640
8	that	571
9	we	560
10	for	445

Part 3: Running Wordcount1.java to produce words which occur more than 4 times

2017.02.18 18:08:00 UTC	2017.02.18 18:08:05 UTC	2017.02.18 18:08:49 UTC	job_1487440869693_0002	word count	vagrant	default	SUCCEEDED	7	7	1	1
-------------------------------	-------------------------------	-------------------------------	------------------------	---------------	---------	---------	-----------	---	---	---	---

Screenshot of the top 10 words

	A	B
1	the	1867
2	to	1433
3	and	1217
4	of	1142
5	a	757
6	our	657
7	in	640
8	that	571
9	we	560
10	for	445

A SMALL ANALYSIS

The interesting thing to note is that the top 10 words of all the 3 outputs IS THE SAME. This is because in step 1 and 2, we are simply running the wordcount.java codes. In the step 3, we skipped all words who occurred less than 4 times. Even in that case, our output list was only trimmed, it did not change.

Part 4: Running Wordcount2.java and skipping prepositions and punctuations

2017.02.19 01:54:01 UTC	2017.02.19 01:54:07 UTC	2017.02.19 01:54:53 UTC	job_1487440869693_0006	word count	vagrant	default	SUCCEEDED	7	7	1	1
2017.02.19 01:54:01 UTC	2017.02.19 01:54:07 UTC	2017.02.19 01:54:53 UTC	job_1487440869693_0006	word count	vagrant	default	FAILED	0	0	0	0

Screenshot of the top 10 words

	A	B
1	the	1868
2	and	1222
3	a	759
4	our	657
5	th	582
6	we	562
7	is	399
8	will	398
9	I	269
10	this	265

In this case, the output was a little different as compared to the other 3 steps.