

# EEL 6761 - Cloud Computing and Storage

## 1. WordCount

a. Mapper reads the input values and emits the (key, 1) pairs to be read by the reducer class. The reducer class sums the values of same key and prints the final (key, value) pair to the output file. The results can be found in the following S3 links

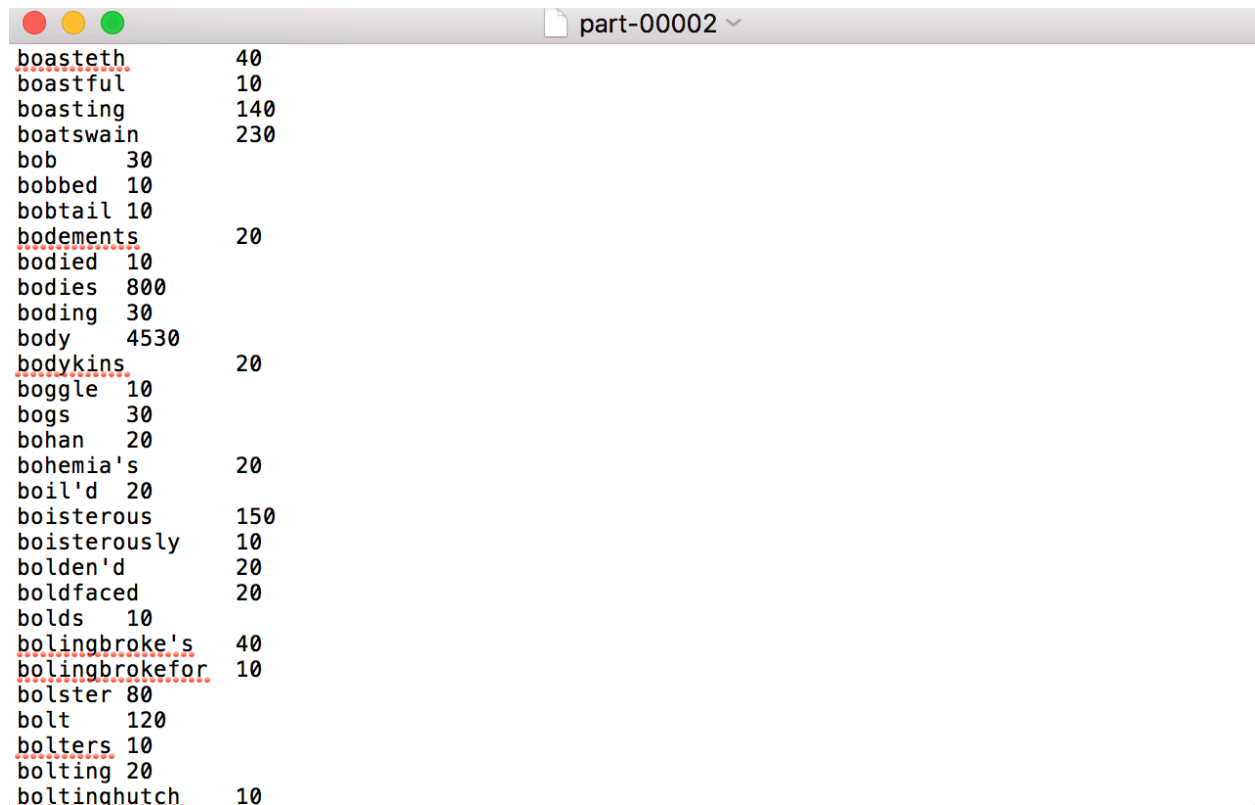
<https://s3.us-east-2.amazonaws.com/mantha/results/single/part-00000>

<https://s3.us-east-2.amazonaws.com/mantha/results/single/part-00001>

<https://s3.us-east-2.amazonaws.com/mantha/results/single/part-00002>

## b. Results

Below is a screenshot of one of the result file from S3. It has the count of each word in the input file against it.



boasteth	40
boastful	10
boasting	140
boatswain	230
bob	30
bobbed	10
bobtail	10
bodements	20
bodied	10
bodies	800
boding	30
body	4530
bodykins	20
boggle	10
bogs	30
bohan	20
bohemia's	20
boil'd	20
boisterous	150
boisterously	10
bolden'd	20
boldfaced	20
bolds	10
bolingbroke's	40
bolingbrokefor	10
bolster	80
bolt	120
bolters	10
bolting	20
boltinghutch	10

## 2. Double WordCount

a. Mapper class emits keys that are two words separated by one space. Reducer sums the values of same key and prints the final pair to the output file. The results can be found in the following S3 links

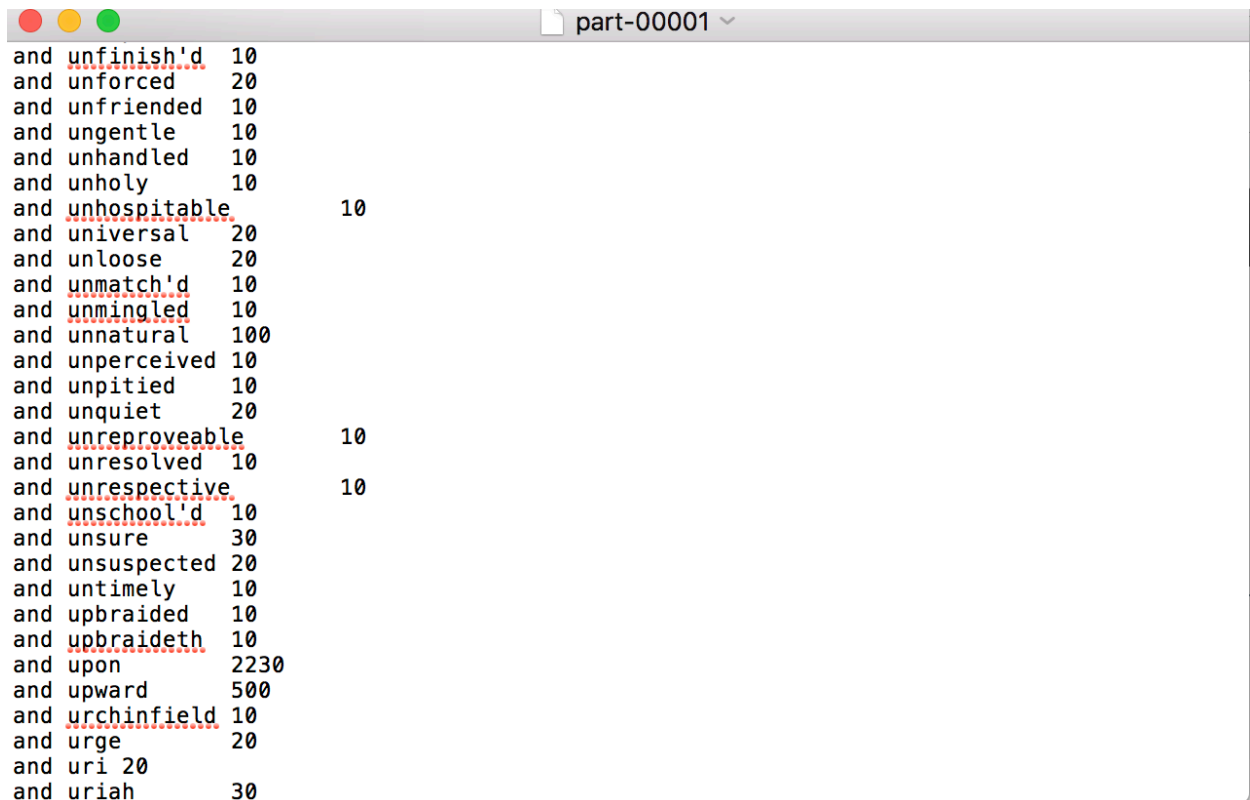
<https://s3.us-east-2.amazonaws.com/mantha/results/double/part-00000>

<https://s3.us-east-2.amazonaws.com/mantha/results/double/part-00001>

<https://s3.us-east-2.amazonaws.com/mantha/results/double/part-00002>

### b. Results

Below is a screenshot of one of the result file from S3. It has the count of each pair of words in the input file against it.



```
part-00001
and unfinish'd 10
and unforced 20
and unfriended 10
and ungentle 10
and unhandled 10
and unholy 10
and unhospitable 10
and universal 20
and unloose 20
and unmatch'd 10
and unmingled 10
and unnatural 100
and unperceived 10
and unpitied 10
and unquiet 20
and unreprieveable 10
and unresolved 10
and unrespectful 10
and unschool'd 10
and unsure 30
and unsuspected 20
and untimely 10
and upbraided 10
and upbraideth 10
and upon 2230
and upward 500
and urchinfield 10
and urge 20
and uri 20
and uriah 30
```

### 3. Distributed Cache

a. The mapper class reads the word-patterns file and stores its words in a buffer and compares them to that of input (bible) and emits only those key, value pairs in which the key is also found in the buffer. The results can be found in the following S3 links

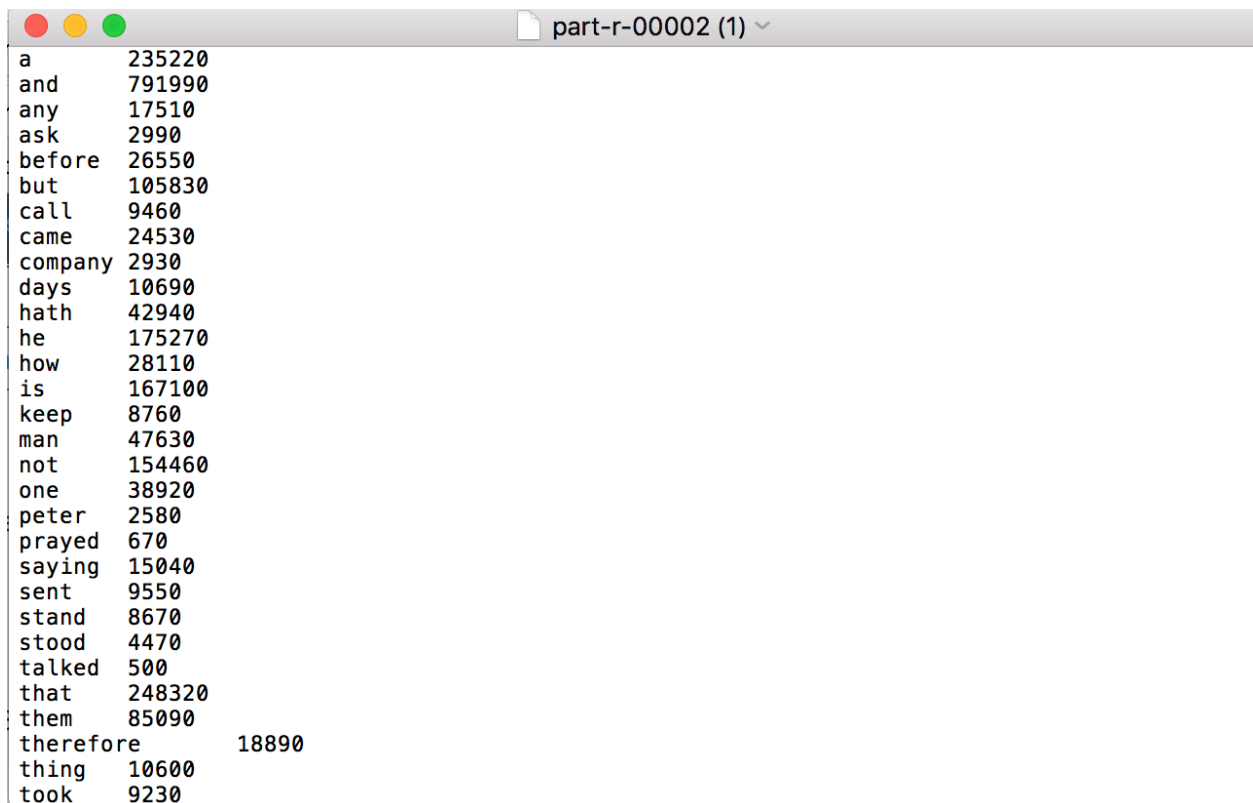
<https://s3.us-east-2.amazonaws.com/mantha/results/distcache/part-r-00000>

<https://s3.us-east-2.amazonaws.com/mantha/results/distcache/part-r-00001>

<https://s3.us-east-2.amazonaws.com/mantha/results/distcache/part-r-00002>

#### b. Results

Below is a screenshot of one of the result file from S3. It has the count of each word from the word-pattern file in the input file against it.



a	235220	
and	791990	
any	17510	
ask	2990	
before	26550	
but	105830	
call	9460	
came	24530	
company	2930	
days	10690	
hath	42940	
he	175270	
how	28110	
is	167100	
keep	8760	
man	47630	
not	154460	
one	38920	
peter	2580	
prayed	670	
saying	15040	
sent	9550	
stand	8670	
stood	4470	
talked	500	
that	248320	
them	85090	
therefore	18890	
thing	10600	
took	9230	

## Screenshots( AWS Cluster setup, steps and Jar files)

Services

Resource Groups

★

VenkateshOhioSupport

Amazon EMR

Cluster listSecurity configurationsVPC subnetsEventsHelp

Cluster: My clusterWaitingCluster ready after last step completed.

Connections:Enable Web Connection – Resource Manager ... (View All)  
Master public DNS:ec2-18-221-87-243.us-east-2.compute.amazonaws.com SSH  
Tags:-- View All / Edit

Summary

ID: j-17QLTH4D2EF2U  
Creation date: 2017-09-10 13:40 (UTC-4)  
Elapsed time: 2 days, 12 hours  
Auto-terminate: No  
Termination protection: On Change

Configuration Details

Release label: emr-5.8.0  
Hadoop distribution: Amazon 2.7.3  
Applications: --  
Log URI: s3://aws-logs-027807388749-us-east-2/elasticmapreduce/  
EMRFS consistent view: Disabled  
Custom AMI ID: --

Network and Hardware

Availability zone: us-east-2a  
Subnet ID: subnet-0780b36e  
Master: Running 1 m4.large  
Core: Running 2 m4.large  
Task: --

Security and Access

Key name: wordcount  
EC2 instance profile: EMR\_EC2\_DefaultRole  
EMR role: EMR\_DefaultRole  
Auto Scaling role: EMR\_AutoScaling\_DefaultRole  
Visible to all users: All Change  
Security groups for sg-3f2f8957 (ElasticMapReduce-Master: master)  
Security groups for sg-732d8b1b (ElasticMapReduce-Core & Task: slave)

FeedbackEnglish (US)© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved.Privacy PolicyTerms of Use

### Hardware

Add task instance group

#### Instance Groups

Filter: Filter instance groups ... 2 instance groups (all loaded)

ID	Status	Node type & name	Instance Type	Instance Count
ig-1UZ925M5PCE7X	Running	MASTER Master - 1	m4.large 4 vCPU, 8 GiB memory, EBS only storage EBS Storage: 32 GiB	1 Instances
ig-FREYS2PSOCJW	Running	CORE Core - 2	m4.large 4 vCPU, 8 GiB memory, EBS only storage EBS Storage: 32 GiB	2 Instances <a href="#">Resize</a>

### Steps

Add stepClone stepCancel step

#### Steps

[View all interactive jobs](#) | [View all jobs](#)

Filter: All steps Filter steps ... 16 steps (all loaded)

ID	Name	Status	Start time (UTC-4)	Elapsed time	Log files
s-3VR8K7T2LC9DZ	distcache	Pending			<a href="#">controller</a>   <a href="#">syslog</a>   <a href="#">stderr*</a>   <a href="#">stdout*</a>
s-3IWCJNLUINU4A	doublewordcount	Completed	2017-09-13 01:22 (UTC-4)	1 minute	<a href="#">View logs</a>
s-358OQT70VAK70	wordcount	Completed	2017-09-13 01:14 (UTC-4)	1 minute	<a href="#">View logs</a>
s-18YALE7A01440	distcache	Completed	2017-09-12 20:47 (UTC-4)	50 seconds	<a href="#">View logs</a>

Overview

Properties

Permissions

Management

🔍 Type a prefix and press Enter to search. Press ESC to clear.

📁 Upload

+ Create folder

More ▾

US East (Ohio) ↻

Viewing 1 to 6				
<input type="checkbox"/>	Name	Last modified	Size	Storage class
<input type="checkbox"/>	results	--	--	--
<input type="checkbox"/>	DistCache.jar	Sep 13, 2017 1:13:19 AM	4.6 KB	Standard
<input type="checkbox"/>	DoubleWordCount.jar	Sep 13, 2017 1:13:20 AM	3.5 KB	Standard
<input type="checkbox"/>	WordCount.jar	Sep 13, 2017 1:13:20 AM	3.3 KB	Standard
<input type="checkbox"/>	bible_big	Sep 10, 2017 2:26:04 PM	86.5 MB	Standard
<input type="checkbox"/>	word-patterns.txt	Sep 12, 2017 8:46:16 PM	628.0 B	Standard