

Assignment 1

Jingyi Ding 69364001

Codes and outputs in https://github.com/dingjy94/EEL6761_Cloud_Computing

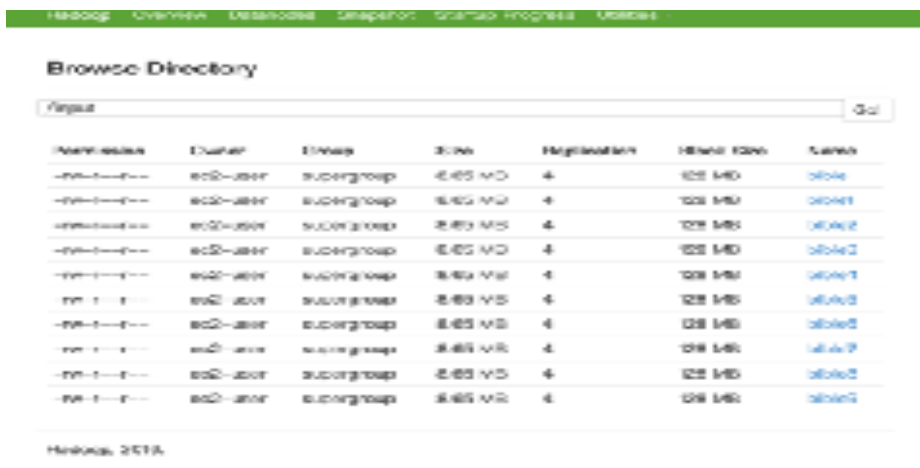
1. Some aws steps:



3 aws sc2 instances, one for namenode, two for datanodes;



2 datanode in operation when hdfs started;



10 bibles (input files).

2. part1



Show 20 entries		Search:		
Attempt ID	Started	Node	Logs	Blacklisted Nodes
appattempt_1504755544562_0092_000002	Thu, 07 Sep 2017 03:49:42 GMT	http://p-172-31-27-19.us-east-2.compute.internal:8042	Logs	0
appattempt_1504755544562_0092_000001	Thu, 07 Sep 2017 03:48:25 GMT	http://p-172-31-30-56.us-east-2.compute.internal:8042	Logs	0
Showing 1 to 2 of 2 entries				First
				Previous
				Next Last

The code is just similar to Hadoop's sample WordCount.

3. part2

My solution is sample, just use two StringTokenizer to read the input file, one StringTokenizer is one word faster than another. Thus, we can easily get double words key by combine these two StringTokenizer.

job_1504891155723_0004

[Job Overview](#)

Job Name: word count

State: RUNNING

Uberized: false

Started: Fri Sep 08 18:02:07 UTC 2017

Elapsed: 1 mins 49 sec

ApplicationMaster

Attempt Number	Start Time	Node	Logs
1	Fri Sep 08 18:02:07 UTC 2017	ip-172-31-27-19.us-east-2.compute.internal:8042	logs

Task Type	Progress	Total	Pending	Running	Complete
Map	<div></div>	0	0	0	0
Reduce	<div></div>	1	0	1	0

Attempt Type	New	Running	Failed	Killed	Successful
Maps	0	0	0	0	0
Reducers	0	1	0	0	0

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Det
4	0	0	4	0	0 B	8 GB	0 B	0	8	0	1	0
Show 20 entries												
ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus				
application_1504891155723_0004	ec2-user	word count	MAPREDUCE	default	Fri, 08 Sep 2017 18:02:00 GMT	Fri, 08 Sep 2017 18:16:40 GMT	FINISHED	SUCCEEDED				



Application application_1504891155723_0004

Kill Application		Application Overview	
		User:	ec2-user
		Name:	word count
		Application Type:	MAPREDUCE
		Application Tags:	
		YarnApplicationState:	FINISHED
		FinalStatus Reported by AM:	SUCCEEDED
		Started:	8-Sep-2017 18:02:00
		Elapsed:	10mins, 48sec
		Tracking URL:	History
		Diagnostics:	

4. part3

The program get three arguments, input, output and word-pattern file's paths instead of only two arguments. Then, add word-pattern into local cache. In mapper class, when setup the map, read the local file and turns it to HashSet. Then, map method is still similiar to WordCount, but only write the word contained in HashSet

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	vCores Used	vCores Total	vCores Reserved	Active Nodes	Decommission Nodes
1	0	0	1	0	0 B	24 GB	0 B	0	8	0	1	0

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress
application_1505420087431_0001	ec2-user	word count	MAPREDUCE	default	Thu, 14 Sep 2017 21:55:11 GMT	Thu, 14 Sep 2017 22:00:49 GMT	FINISHED	SUCCEEDED	

Showing 1 to 1 of 1 entries

Kill Application

Kill Application		Application Overview	
		User:	ec2-user
		Name:	word count
		Application Type:	MAPREDUCE
		Application Tags:	
		YarnApplicationState:	FINISHED
		FinalStatus Reported by AM:	SUCCEEDED
		Started:	14-Sep-2017 21:55:11
		Elapsed:	13mins, 37sec
		Tracking URL:	History
		Diagnostics:	

Kill Application		Application Metrics	
		Total Resource Preempted:	<memory:0, vCores:0>