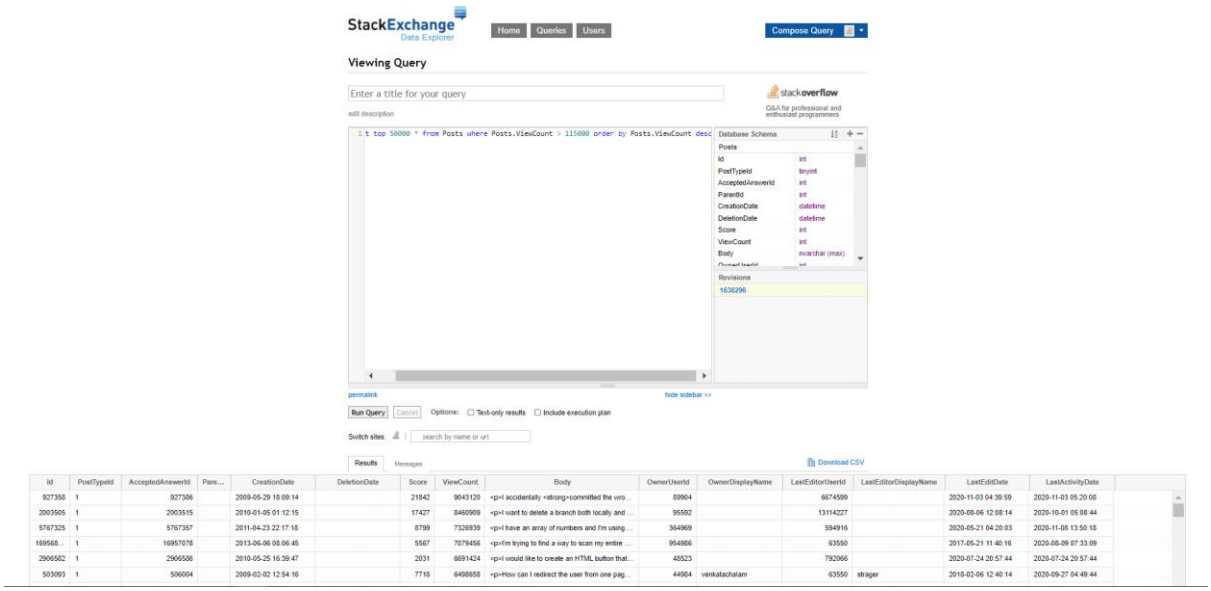
# Assignment 1- Data Analysis

Srijith Unni
DCU Student ID: 20211114
MSc in Computing – Data Analytics
Cloud Technologies CA675

**Complete chronological screenshots:**




Results Messages Download CSV

| Id | PostTypeId | AcceptedAnswerId | Pare... | CreationDate | DeletionDate | Score | ViewCount | Body | OwnerUserId | OwnerDisplayName |
|---|---|---|---|---|---|---|---|---|---|---|
| 38107085 | 1 | 38112235 | | 2016-06-29 18:06:57 | | 17 | 49999 | I'm trying to implement a login system wh... | 1326269 | |
| 24624980 | 1 | | | 2014-07-08 06:24:25 | | 26 | 49999 | According to the documentation, <a href=... | 136476 | |
| 12067929 | 1 | | | 2012-08-22 07:07:52 | | 20 | 49999 | I'm using codeigniter as a framework.</p... | 1616141 | |
| 9090072 | 1 | 9090375 | | 2012-02-01 02:50:26 | | 37 | 49999 | I have a SQL script that will insert a long ... | 385387 | |
| 31198835 | 1 | | | 2015-07-03 05:01:55 | | 77 | 49998 | I am working on a task that involves buildi... | 3044967 | |
| 397754 | 1 | 397812 | | 2008-12-29 14:02:24 | | 21 | 49998 | Is there a way to record the screen, either... | 23294 | CJCraft.com |
| 36191182 | 1 | 36191199 | | 2016-03-24 00:25:42 | | 22 | 49996 | Writing this html: <pre><code>lorem ... | 3044737 | |
| 25223958 | 1 | 25230593 | | 2014-08-09 23:08:13 | | 12 | 49996 | When I try to modify the format of the def... | 2047843 | |
| 7407935 | 1 | 7408036 | | 2011-09-13 20:19:23 | | 19 | 49996 | According to jQuery document on <code>... | 707381 | |
| 16383237 | 1 | | | 2013-05-05 09:41:32 | | 67 | 49996 | Is there a program or workflow to convert ... | 901946 | |
| 48012662 | 1 | 48013403 | | 2017-12-28 18:38:21 | | 16 | 49996 | In an Angular 2+ component, how do I pa... | 79677 | |
| 42618872 | 1 | 42619368 | | 2017-03-06 06:27:02 | | 14 | 49996 | I know this is so common question &... | 928715 | |
| 13115784 | 1 | 13116162 | | 2012-10-29 04:15:33 | | 16 | 49996 | So now, I am making a client server app ... | 1318743 | |
| 3549890 | 1 | 3549932 | | 2010-08-23 17:04:59 | | 18 | 49995 | There is a web page from which I want to ... | 310626 | |
| 7762771 | 1 | 7762882 | | 2011-10-14 03:33:42 | | 7 | 49995 | I'm making a simple program that takes t... | 994684 | |
| 7810326 | 1 | | | 2011-10-18 16:07:55 | | 15 | 49994 | URL: <a href="http://cyberbat.co.uk/test" r... | 995913 | |

44177 rows returned in 897 ms (cached)

Run Query | Cancel | Options: Text-only results | Include execution plan

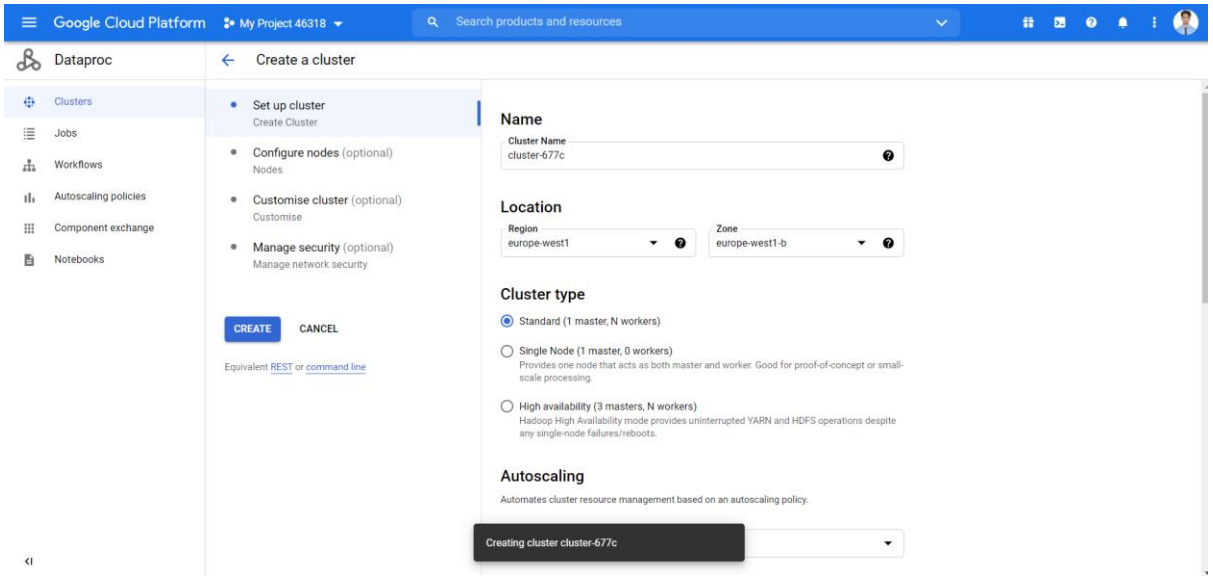
Switch sites: search by name or url

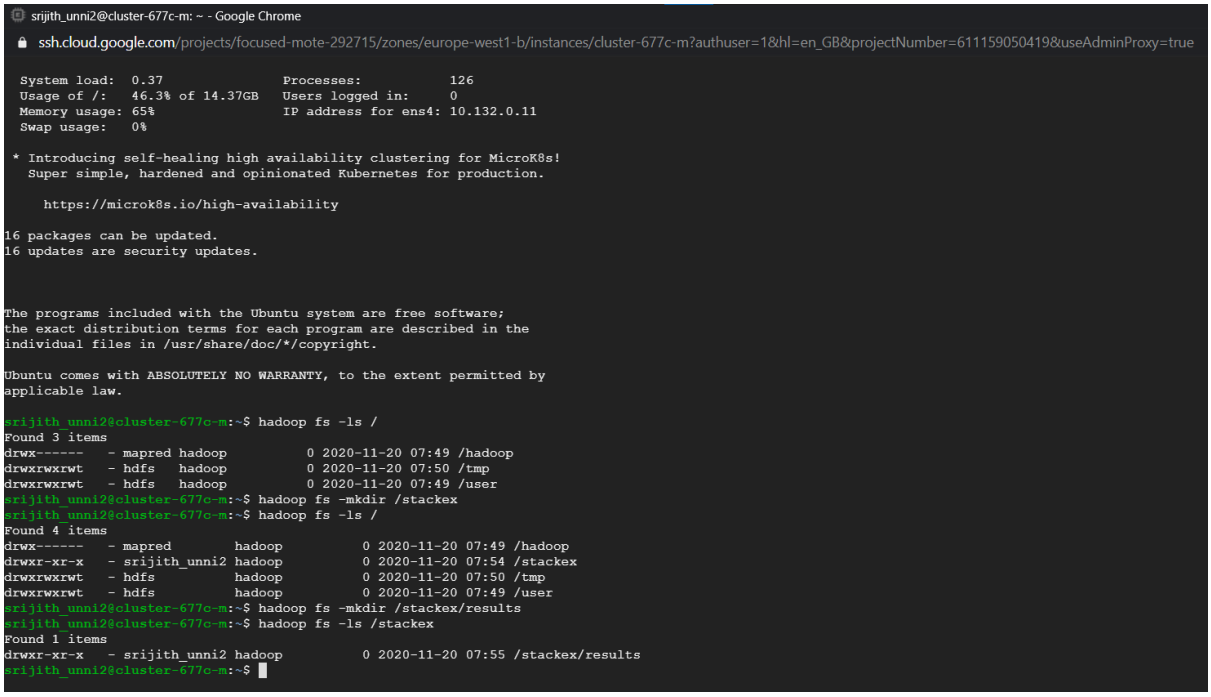
Hold tight while we fetch your results

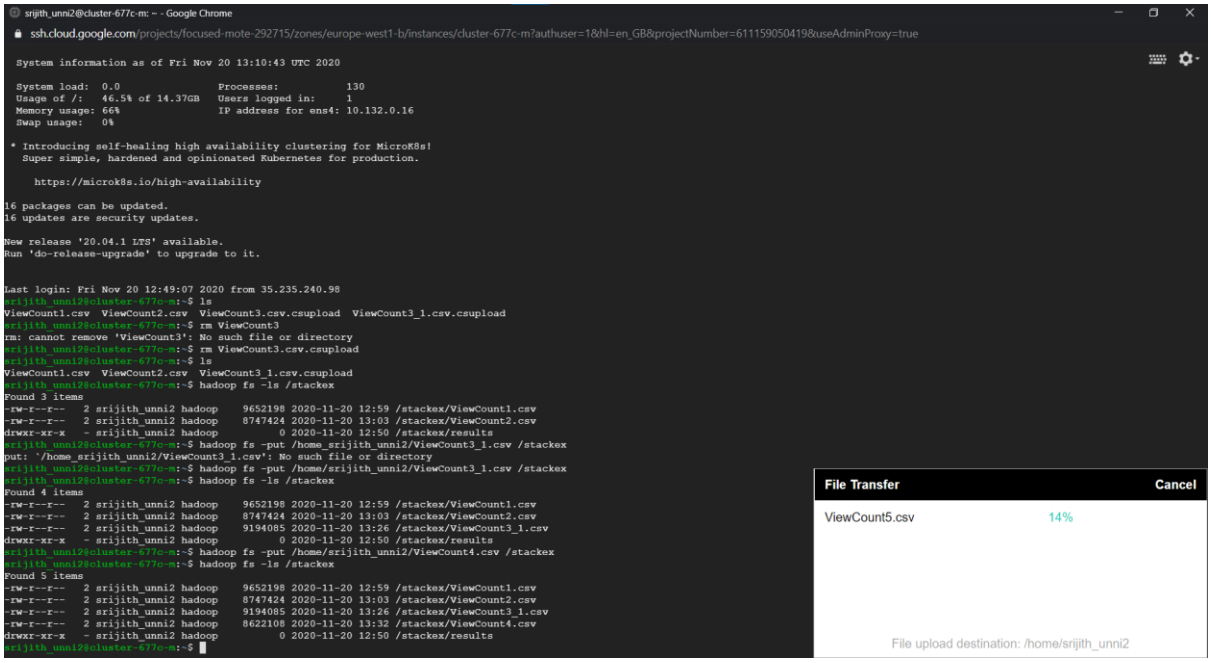
Results | Messages | Download CSV

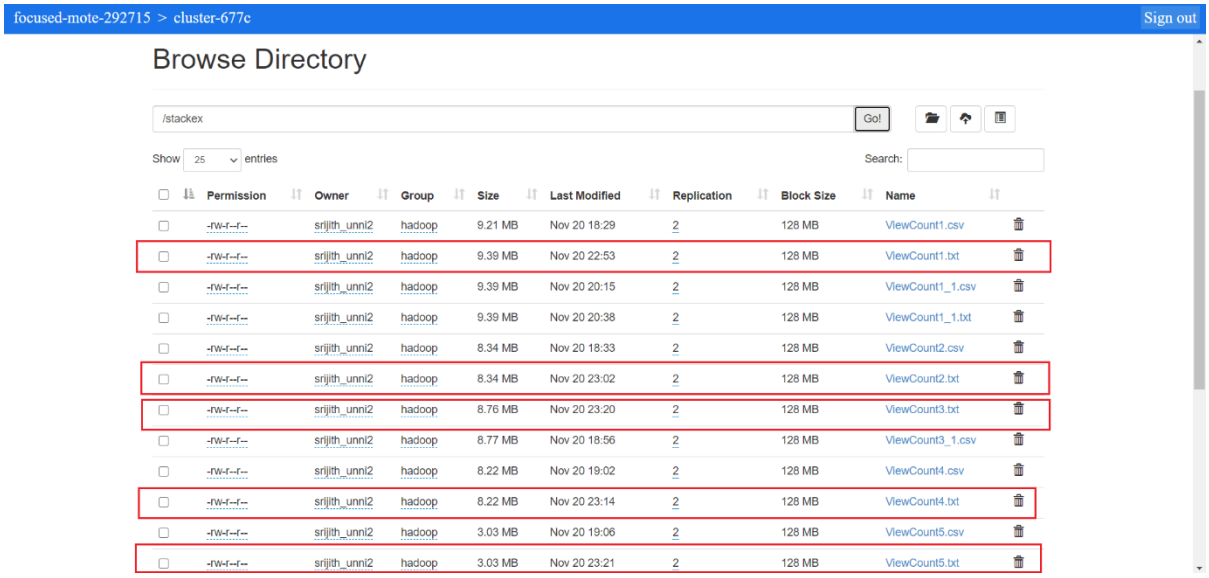
| Id | PostTypeId | AcceptedAnswerId | Pare... | CreationDate | DeletionDate | Score | ViewCount | Body | OwnerUserId | OwnerDisplayName |
|---|---|---|---|---|---|---|---|---|---|---|
| 21376645 | 1 | 21376885 | | 2014-01-27 09:01:51 | | 7 | 69999 | As i know, i can create an array with item ... | 3122881 | |
| 20829348 | 1 | 20829349 | | 2013-12-29 19:43:17 | | 162 | 69999 | How do I join two strings in a list with a sp... | 1139353 | |
| 5323349 | 1 | 5323350 | | 2011-03-16 05:59:12 | | 42 | 69999 | What is the <code>#error</code> directiv... | 390452 | Shine |
| 29574732 | 1 | | | 2015-04-11 06:09:43 | | 44 | 69997 | As of 2015, I see that Android studio is an... | 1800583 | |
| 22842691 | 1 | 22844164 | | 2014-04-03 15:46:59 | | 159 | 69996 | Since I first saw a <code>dist/</code> dir... | 2391795 | |
| 14870596 | 1 | 14871317 | | 2013-02-14 08:27:44 | | 36 | 69996 | I'm trying to run a sample android code in... | 893411 | |
| 26307920 | 1 | 26308081 | | 2014-10-10 20:56:46 | | 22 | 69995 | I'm learning node Js and i'm trying to do a... | 3595813 | |
| 26033239 | 1 | 26033300 | | 2014-09-25 07:46:20 | | 31 | 69995 | I have a problem converting <code>Obje... | 4046638 | |
| 5193973 | 1 | 5193991 | | 2011-03-04 12:55:02 | | 17 | 69995 | How would I loop through all the td eleme... | 517406 | |
| 8707753 | 1 | | | 2012-01-03 04:00:10 | | 9 | 69995 | I was thinking of developing an iPhone an... | 1080731 | |
| 6993132 | 1 | 6993177 | | 2011-08-09 08:05:02 | | 56 | 69994 | If I have an integer variable I can use <co... | 184759 | |
| 6825834 | 1 | 6826076 | | 2011-07-26 05:52:19 | | 32 | 69993 | I have a string as shown below. <pre... | 683898 | |
| 3192095 | 1 | | | 2010-07-07 05:17:09 | | 50 | 69993 | I was just curious where exactly the singl... | 311498 | |
| 11260659 | 1 | 11262192 | | 2012-06-29 11:05:59 | | 14 | 69992 | I have a question about if statement in tcl ... | | user707549 |
| 12336234 | 1 | 12340302 | | 2012-09-09 03:11:19 | | 28 | 69990 | I have data stored in a CSV where the firs... | 907714 | |
| 12999899 | 1 | 13000013 | | 2012-10-21 16:44:03 | | 7 | 69990 | I am trying to have a scanner take input i... | 605328 | |

46925 rows returned in 84053 ms

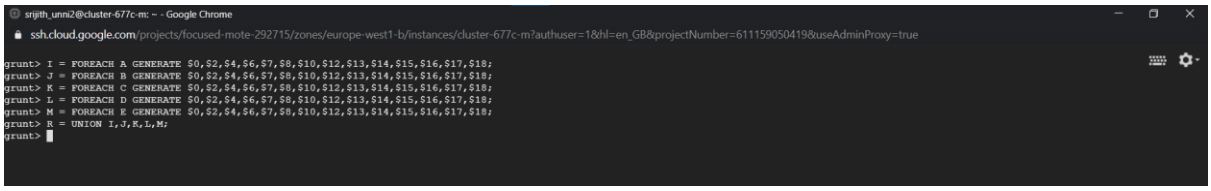



srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
System load: 0.37 Processes: 126
Usage of /: 46.3% of 14.37GB Users logged in: 0
Memory usage: 65% IP address for ens4: 10.132.0.11
Swap usage: 0%
* Introducing self-healing high availability clustering for MicroK8s!
Super simple, hardened and opinionated Kubernetes for production.
https://microk8s.io/high-availability
16 packages can be updated.
16 updates are security updates.
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /
Found 3 items
drwx------ - mapred hadoop 0 2020-11-20 07:49 /hadoop
drwxrwxrwt - hdfs hadoop 0 2020-11-20 07:50 /tmp
drwxrwxrwt - hdfs hadoop 0 2020-11-20 07:49 /user
srijith_unni2@cluster-677c-m:~$ hadoop fs -mkdir /stackex
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /
Found 4 items
drwx------ - mapred hadoop 0 2020-11-20 07:49 /hadoop
drwxr-xr-x - srijith_unni2 hadoop 0 2020-11-20 07:54 /stackex
drwxrwxrwt - hdfs hadoop 0 2020-11-20 07:50 /tmp
drwxrwxrwt - hdfs hadoop 0 2020-11-20 07:49 /user
srijith_unni2@cluster-677c-m:~$ hadoop fs -mkdir /stackex/results
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /stackex
Found 1 items
drwxr-xr-x - srijith_unni2 hadoop 0 2020-11-20 07:55 /stackex/results
srijith_unni2@cluster-677c-m:~$

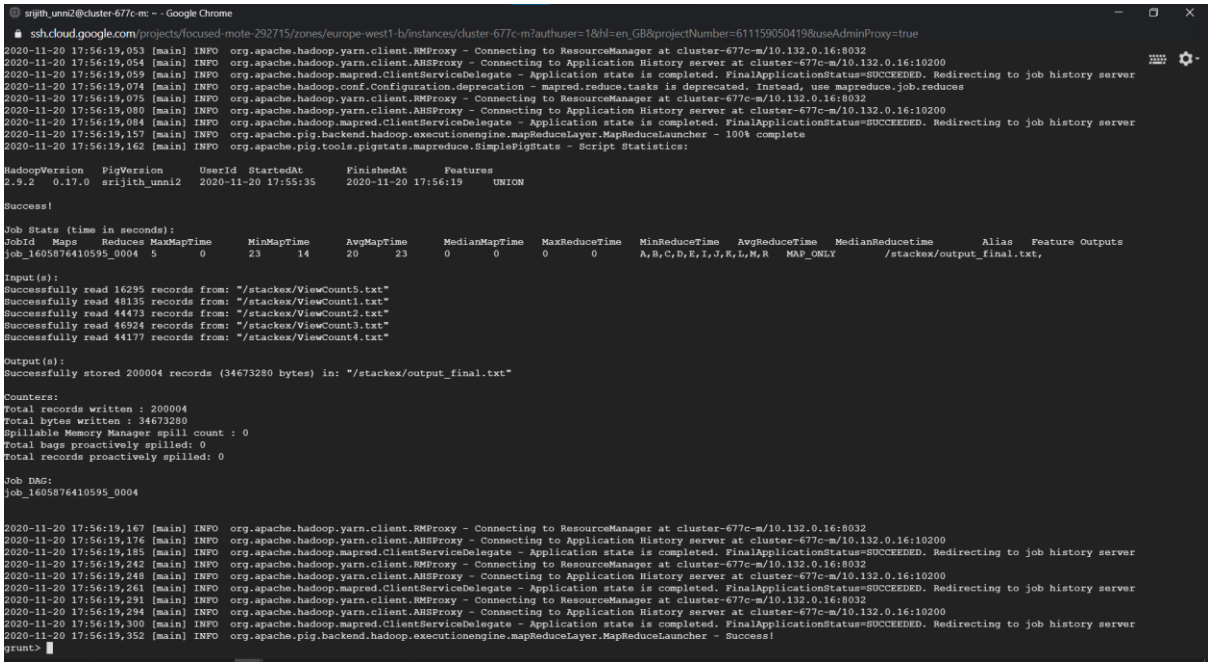

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
System information as of Fri Nov 20 13:10:43 UTC 2020
System load: 0.0 Processes: 130
Usage of /: 46.5% of 14.37GB Users logged in: 1
Memory usage: 66% IP address for ens4: 10.132.0.16
Swap usage: 0%
* Introducing self-healing high availability clustering for MicroK8s!
Super simple, hardened and opinionated Kubernetes for production.
https://microk8s.io/high-availability
16 packages can be updated.
16 updates are security updates.
New release '20.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.
Last login: Fri Nov 20 12:49:07 2020 from 35.235.240.98
srijith_unni2@cluster-677c-m:~$ ls
ViewCount1.csv ViewCount2.csv ViewCount3.csv.csupload ViewCount3_1.csv.csupload
srijith_unni2@cluster-677c-m:~$ rm ViewCount3
rm: cannot remove 'ViewCount3': No such file or directory
srijith_unni2@cluster-677c-m:~$ rm ViewCount3.csv.csupload
srijith_unni2@cluster-677c-m:~$ ls
ViewCount1.csv ViewCount2.csv ViewCount3_1.csv.csupload
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /stackex
Found 3 items
-rw-r--r-- 2 srijith_unni2 hadoop 9652198 2020-11-20 12:59 /stackex/ViewCount1.csv
-rw-r--r-- 2 srijith_unni2 hadoop 8747424 2020-11-20 13:03 /stackex/ViewCount2.csv
drwxr-xr-x - srijith_unni2 hadoop 0 2020-11-20 12:50 /stackex/results
srijith_unni2@cluster-677c-m:~$ hadoop fs -put /home_srijith_unni2/ViewCount3_1.csv /stackex
put: `/home_srijith_unni2/ViewCount3_1.csv': No such file or directory
srijith_unni2@cluster-677c-m:~$ hadoop fs -put /home/srijith_unni2/ViewCount3_1.csv /stackex
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /stackex
Found 4 items
-rw-r--r-- 2 srijith_unni2 hadoop 9652198 2020-11-20 12:59 /stackex/ViewCount1.csv
-rw-r--r-- 2 srijith_unni2 hadoop 8747424 2020-11-20 13:03 /stackex/ViewCount2.csv
-rw-r--r-- 2 srijith_unni2 hadoop 9194085 2020-11-20 13:26 /stackex/ViewCount3_1.csv
drwxr-xr-x - srijith_unni2 hadoop 0 2020-11-20 12:50 /stackex/results
srijith_unni2@cluster-677c-m:~$ hadoop fs -put /home/srijith_unni2/ViewCount4.csv /stackex
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /stackex
Found 5 items
-rw-r--r-- 2 srijith_unni2 hadoop 9652198 2020-11-20 12:59 /stackex/ViewCount1.csv
-rw-r--r-- 2 srijith_unni2 hadoop 8747424 2020-11-20 13:03 /stackex/ViewCount2.csv
-rw-r--r-- 2 srijith_unni2 hadoop 9194085 2020-11-20 13:26 /stackex/ViewCount3_1.csv
-rw-r--r-- 2 srijith_unni2 hadoop 8622108 2020-11-20 13:32 /stackex/ViewCount4.csv
drwxr-xr-x - srijith_unni2 hadoop 0 2020-11-20 12:50 /stackex/results
srijith_unni2@cluster-677c-m:~$
File Transfer
Cancel
ViewCount5.csv
14%
File upload destination: /home/srijith_unni2

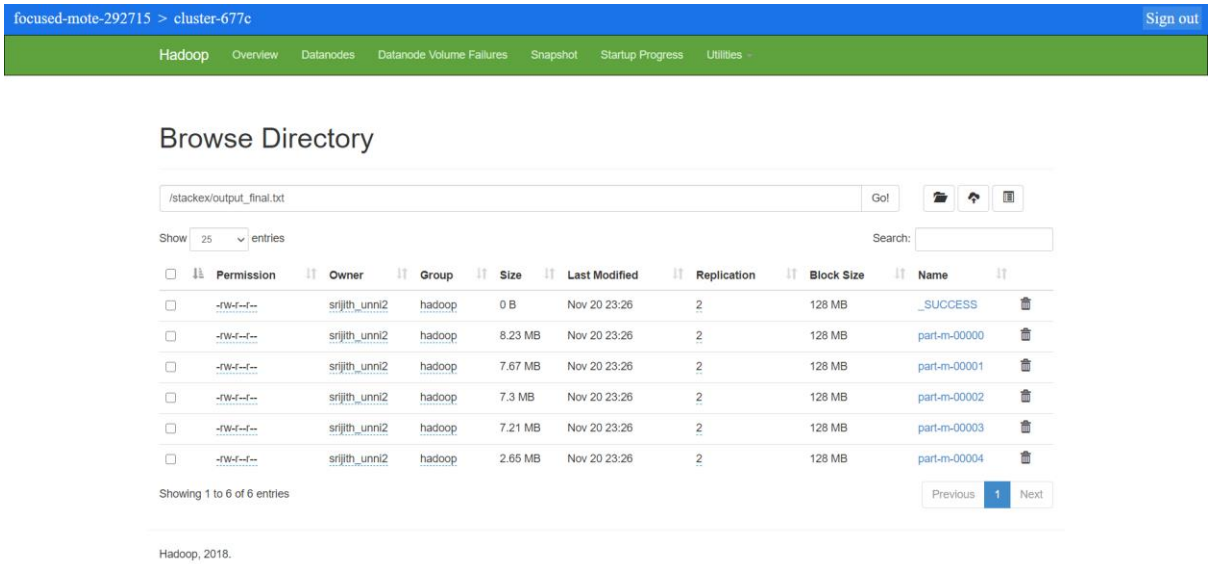

focused-mote-292715 > cluster-677c
Sign out
Browse Directory
/stackex
Go!
Show 25 entries
Search:
Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name
-rw-r--r-- | srijith_unni2 | hadoop | 9.21 MB | Nov 20 18:29 | 2 | 128 MB | ViewCount1.csv
-rw-r--r-- | srijith_unni2 | hadoop | 9.39 MB | Nov 20 22:53 | 2 | 128 MB | ViewCount1.txt
-rw-r--r-- | srijith_unni2 | hadoop | 9.39 MB | Nov 20 20:15 | 2 | 128 MB | ViewCount1_1.csv
-rw-r--r-- | srijith_unni2 | hadoop | 9.39 MB | Nov 20 20:38 | 2 | 128 MB | ViewCount1_1.txt
-rw-r--r-- | srijith_unni2 | hadoop | 8.34 MB | Nov 20 18:33 | 2 | 128 MB | ViewCount2.csv
-rw-r--r-- | srijith_unni2 | hadoop | 8.34 MB | Nov 20 23:02 | 2 | 128 MB | ViewCount2.txt
-rw-r--r-- | srijith_unni2 | hadoop | 8.76 MB | Nov 20 23:20 | 2 | 128 MB | ViewCount3.txt
-rw-r--r-- | srijith_unni2 | hadoop | 8.77 MB | Nov 20 18:56 | 2 | 128 MB | ViewCount3_1.csv
-rw-r--r-- | srijith_unni2 | hadoop | 8.22 MB | Nov 20 19:02 | 2 | 128 MB | ViewCount4.csv
-rw-r--r-- | srijith_unni2 | hadoop | 8.22 MB | Nov 20 23:14 | 2 | 128 MB | ViewCount4.txt
-rw-r--r-- | srijith_unni2 | hadoop | 3.03 MB | Nov 20 19:06 | 2 | 128 MB | ViewCount5.csv
-rw-r--r-- | srijith_unni2 | hadoop | 3.03 MB | Nov 20 23:21 | 2 | 128 MB | ViewCount5.txt

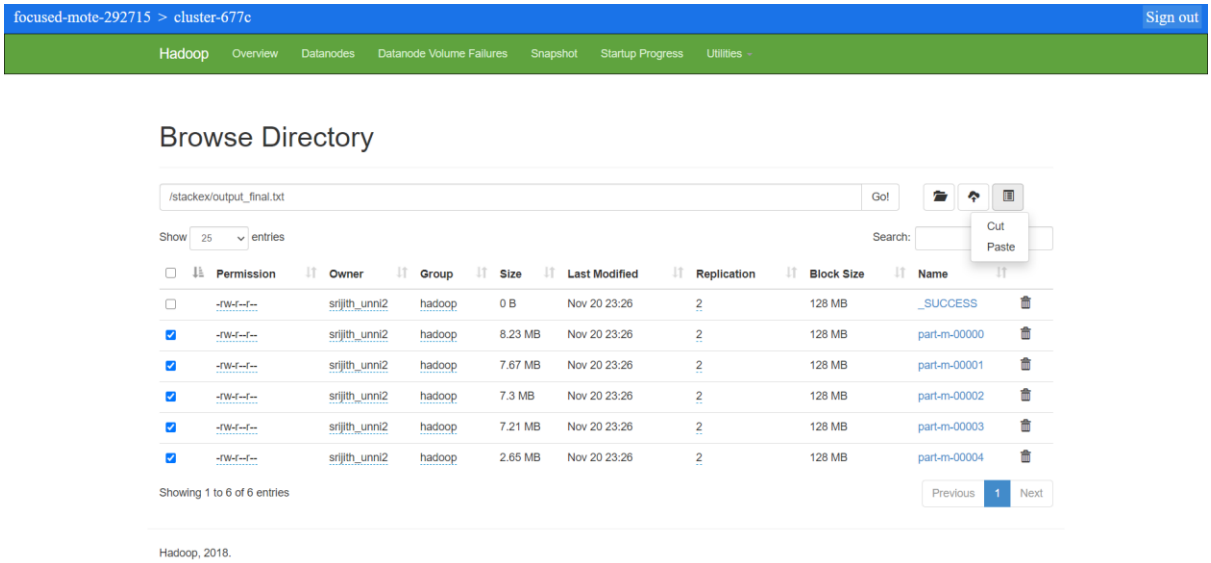

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
srijith_unni2@cluster-677c-m:~$ pig
20/11/20 17:53:10 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
20/11/20 17:53:10 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDUCE
20/11/20 17:53:10 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the ExecType
2020-11-20 17:53:10,503 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r: unknown) compiled Oct 18 2020, 10:13:54
2020-11-20 17:53:10,503 [main] INFO org.apache.pig.Main - Logging error messages to: /home/srijith_unni2/pig_1605894790495.log
2020-11-20 17:53:10,542 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/srijith_unni2/.pigbootup not found
2020-11-20 17:53:11,165 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2020-11-20 17:53:11,165 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://cluster-677c-m
2020-11-20 17:53:12,404 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-d926933d-8a37-4dd6-9d31-d3d6b512e528
2020-11-20 17:53:12,457 [main] INFO org.apache.hadoop.yarn.client.api.impl.TimelineClientImpl - Timeline service address: null
2020-11-20 17:53:13,030 [main] INFO org.apache.pig.backend.hadoop.PigATSClient - Created ATS Hook
2020-11-20 17:53:13,083 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt> A = LOAD '/stackex/ViewCount1.txt' USING PigStorage('\t') AS(Id:chararray, PostTypeId:chararray, AcceptedAnswerId:chararray, ParentId:chararray, CreationDate:chararray, DeletionDate:chararray, Score:chararray, ViewCount:chararray, OwnerUserId:chararray, OwnerDisplayName:chararray, LastEditorUserId:chararray, LastEditorDisplayName:chararray, LastEditDate:chararray, LastActivtityTime:chararray, Title:chararray, Tags:chararray, AnswerCount:chararray, CommentCount:chararray, FavouriteCount:chararray, ClosedDate:chararray, CommunityOwnedDate:chararray, ContentLicense:chararray);
2020-11-20 17:53:34,437 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt> B = LOAD '/stackex/ViewCount2.txt' USING PigStorage('\t') AS(Id:chararray, PostTypeId:chararray, AcceptedAnswerId:chararray, ParentId:chararray, CreationDate:chararray, DeletionDate:chararray, Score:chararray, ViewCount:chararray, OwnerUserId:chararray, OwnerDisplayName:chararray, LastEditorUserId:chararray, LastEditorDisplayName:chararray, LastEditDate:chararray, LastActivtityTime:chararray, Title:chararray, Tags:chararray, AnswerCount:chararray, CommentCount:chararray, FavouriteCount:chararray, ClosedDate:chararray, CommunityOwnedDate:chararray, ContentLicense:chararray);
2020-11-20 17:53:41,531 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt> C = LOAD '/stackex/ViewCount3.txt' USING PigStorage('\t') AS(Id:chararray, PostTypeId:chararray, AcceptedAnswerId:chararray, ParentId:chararray, CreationDate:chararray, DeletionDate:chararray, Score:chararray, ViewCount:chararray, OwnerUserId:chararray, OwnerDisplayName:chararray, LastEditorUserId:chararray, LastEditorDisplayName:chararray, LastEditDate:chararray, LastActivtityTime:chararray, Title:chararray, Tags:chararray, AnswerCount:chararray, CommentCount:chararray, FavouriteCount:chararray, ClosedDate:chararray, CommunityOwnedDate:chararray, ContentLicense:chararray);
2020-11-20 17:53:49,143 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt> D = LOAD '/stackex/ViewCount4.txt' USING PigStorage('\t') AS(Id:chararray, PostTypeId:chararray, AcceptedAnswerId:chararray, ParentId:chararray, CreationDate:chararray, DeletionDate:chararray, Score:chararray, ViewCount:chararray, OwnerUserId:chararray, OwnerDisplayName:chararray, LastEditorUserId:chararray, LastEditorDisplayName:chararray, LastEditDate:chararray, LastActivtityTime:chararray, Title:chararray, Tags:chararray, AnswerCount:chararray, CommentCount:chararray, FavouriteCount:chararray, ClosedDate:chararray, CommunityOwnedDate:chararray, ContentLicense:chararray);
2020-11-20 17:53:56,091 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt> E = LOAD '/stackex/ViewCount5.txt' USING PigStorage('\t') AS(Id:chararray, PostTypeId:chararray, AcceptedAnswerId:chararray, ParentId:chararray, CreationDate:chararray, DeletionDate:chararray, Score:chararray, ViewCount:chararray, OwnerUserId:chararray, OwnerDisplayName:chararray, LastEditorUserId:chararray, LastEditorDisplayName:chararray, LastEditDate:chararray, LastActivtityTime:chararray, Title:chararray, Tags:chararray, AnswerCount:chararray, CommentCount:chararray, FavouriteCount:chararray, ClosedDate:chararray, CommunityOwnedDate:chararray, ContentLicense:chararray);
2020-11-20 17:54:03,063 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
grunt>

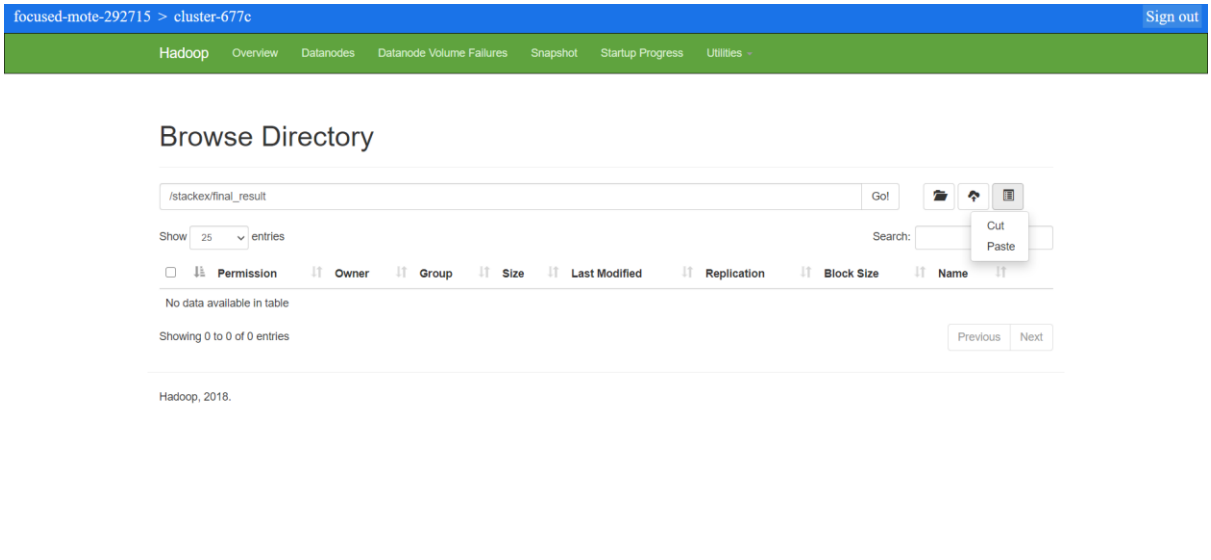

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
grunt> I = FOREACH A GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> J = FOREACH B GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> K = FOREACH C GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> L = FOREACH D GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> M = FOREACH E GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> R = UNION I,J,K,L,M;
grunt>

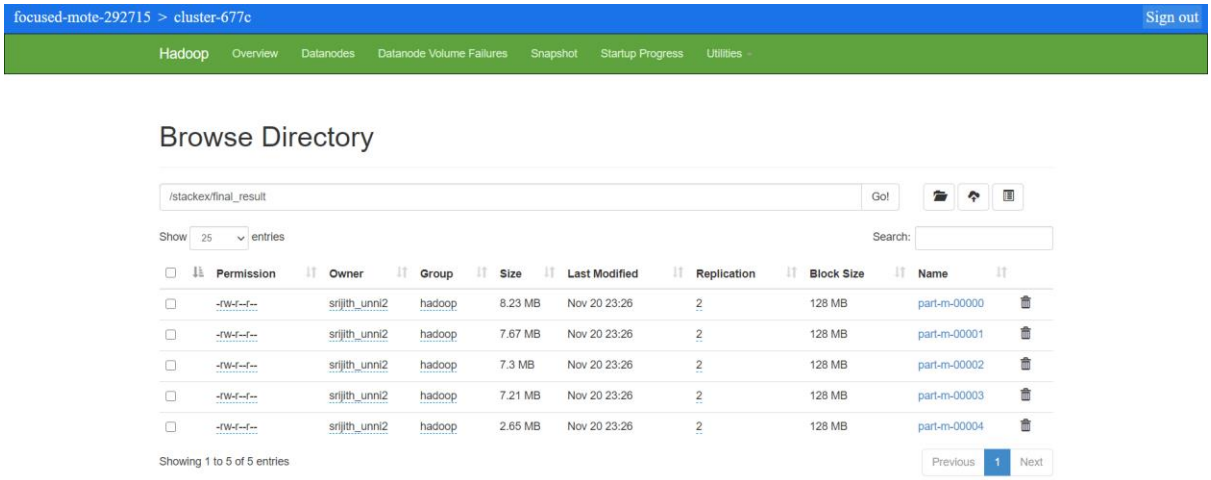

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
grunt> I = FOREACH A GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> J = FOREACH B GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> K = FOREACH C GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> L = FOREACH D GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> M = FOREACH E GENERATE $0,$2,$4,$6,$7,$8,$10,$12,$13,$14,$15,$16,$17,$18;
grunt> R = UNION I,J,K,L,M;
grunt> STORE R INTO '/stackex/output_final.txt' USING PigStorage('\t');
2020-11-20 17:55:34,425 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2020-11-20 17:55:34,542 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutputformat.separator
2020-11-20 17:55:34,662 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: UNION
2020-11-20 17:55:34,696 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2020-11-20 17:55:34,707 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2020-11-20 17:55:34,762 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2020-11-20 17:55:34,864 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for A: $1, $3, $5, $9, $11, $19, $20, $21
2020-11-20 17:55:34,868 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for B: $1, $3, $5, $9, $11, $19, $20, $21
2020-11-20 17:55:34,868 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for C: $1, $3, $5, $9, $11, $19, $20, $21
2020-11-20 17:55:34,869 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for D: $1, $3, $5, $9, $11, $19, $20, $21
2020-11-20 17:55:34,869 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for E: $1, $3, $5, $9, $11, $19, $20, $21
2020-11-20 17:55:34,955 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (Tenured Gen) of size 699072512 to monitor. collectionUsageThreshold = 489350752, usageThreshold = 489350752
2020-11-20 17:55:35,107 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2020-11-20 17:55:35,172 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2020-11-20 17:55:35,173 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2020-11-20 17:55:35,212 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2020-11-20 17:55:35,301 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at cluster-677c-m/10.132.0.16:8032
2020-11-20 17:55:35,571 [main] INFO org.apache.hadoop.yarn.client.AHSProxy - Connecting to Application History server at cluster-677c-m/10.132.0.16:10200
2020-11-20 17:55:35,653 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2020-11-20 17:55:35,665 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecated. Instead, use mapreduce.reduce.markreset.buffer.percent
2020-11-20 17:55:35,669 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2020-11-20 17:55:35,673 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use mapreduce.output.fileoutputformat.compress
2020-11-20 17:55:35,679 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - This job cannot be converted run in-process
2020-11-20 17:55:35,709 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.submit.replication is deprecated. Instead, use mapreduce.client.submit.file.replication
2020-11-20 17:55:36,043 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/pig/pig-0.17.0-core-h2.jar to DistributedCache through /tmp/temp-381137324/tmp-693856074/pig-0.17.0-core-h2.jar
2020-11-20 17:55:36,085 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/pig/lib/automaton-1.11-8.jar to DistributedCache through /tmp/temp-381137324/tmp392829647/automaton-1.11-8.jar
2020-11-20 17:55:36,123 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/pig/lib/antlr-runtime-3.4.jar to DistributedCache through /tmp/temp-381137324/tmp-472282150/antlr-runtime-3.4.jar
2020-11-20 17:55:36,491 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Added jar file:/usr/lib/hive/lib/hive-exec-2.3.7.jar to DistributedCache through /tmp/temp-381137324/tmp-1252980932/hive-exec-2.3.7.jar
2020-11-20 17:55:36,520 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2020-11-20 17:55:36,543 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2020-11-20 17:55:36,549 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2020-11-20 17:55:36,549 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []

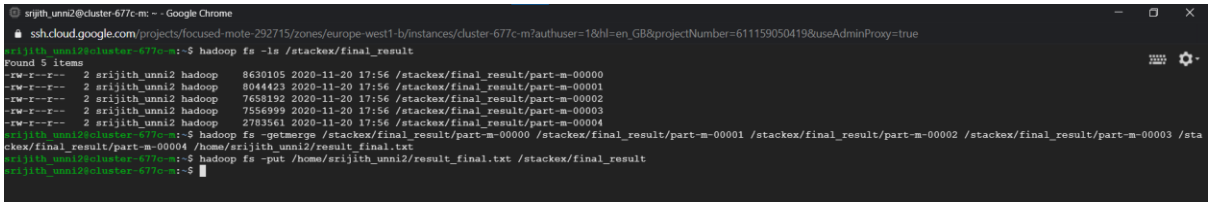

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
2020-11-20 17:56:19,053 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at cluster-677c-m/10.132.0.16:8032
2020-11-20 17:56:19,054 [main] INFO org.apache.hadoop.yarn.client.AHSProxy - Connecting to Application History server at cluster-677c-m/10.132.0.16:10200
2020-11-20 17:56:19,059 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2020-11-20 17:56:19,074 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.reduce.tasks is deprecated. Instead, use mapreduce.job.reduces
2020-11-20 17:56:19,075 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at cluster-677c-m/10.132.0.16:8032
2020-11-20 17:56:19,080 [main] INFO org.apache.hadoop.yarn.client.AHSProxy - Connecting to Application History server at cluster-677c-m/10.132.0.16:10200
2020-11-20 17:56:19,084 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2020-11-20 17:56:19,157 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2020-11-20 17:56:19,162 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.9.2 0.17.0 srijith_unni2 2020-11-20 17:55:35 2020-11-20 17:56:19 UNION
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_1605876410595_0004 5 0 23 14 20 23 0 0 0 0 A,B,C,D,E,I,J,K,L,M,R MAP_ONLY /stackex/output_final.txt,
Input(s):
Successfully read 16295 records from: "/stackex/ViewCount5.txt"
Successfully read 48135 records from: "/stackex/ViewCount1.txt"
Successfully read 44473 records from: "/stackex/ViewCount2.txt"
Successfully read 46924 records from: "/stackex/ViewCount3.txt"
Successfully read 44177 records from: "/stackex/ViewCount4.txt"
Output(s):
Successfully stored 200004 records (34673280 bytes) in: "/stackex/output_final.txt"
Counters:
Total records written : 200004
Total bytes written : 34673280
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1605876410595_0004
2020-11-20 17:56:19,167 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at cluster-677c-m/10.132.0.16:8032
2020-11-20 17:56:19,176 [main] INFO org.apache.hadoop.yarn.client.AHSProxy - Connecting to Application History server at cluster-677c-m/10.132.0.16:10200
2020-11-20 17:56:19,185 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2020-11-20 17:56:19,242 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at cluster-677c-m/10.132.0.16:8032
2020-11-20 17:56:19,248 [main] INFO org.apache.hadoop.yarn.client.AHSProxy - Connecting to Application History server at cluster-677c-m/10.132.0.16:10200
2020-11-20 17:56:19,261 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2020-11-20 17:56:19,291 [main] INFO org.apache.hadoop.yarn.client.RMProxy - Connecting to ResourceManager at cluster-677c-m/10.132.0.16:8032
2020-11-20 17:56:19,294 [main] INFO org.apache.hadoop.yarn.client.AHSProxy - Connecting to Application History server at cluster-677c-m/10.132.0.16:10200
2020-11-20 17:56:19,300 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2020-11-20 17:56:19,352 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
grunt>

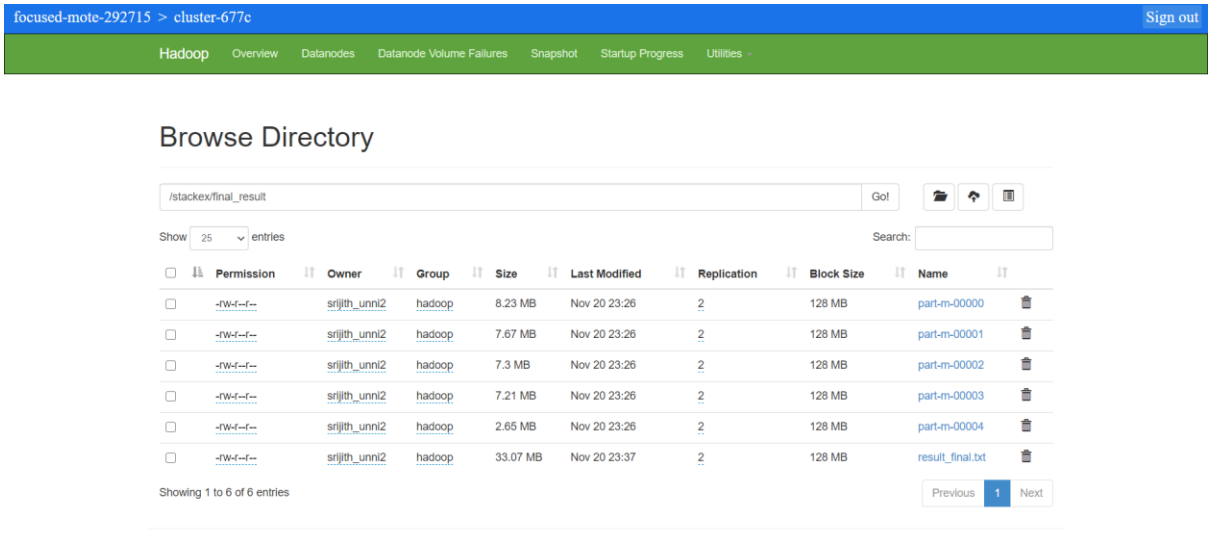

focused-mote-292715 > cluster-677c
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/stackex/output_final.txt
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
srijith_unni2
hadoop
0 B
Nov 20 23:26
2
128 MB
_SUCCESS
-rw-r--r--
srijith_unni2
hadoop
8.23 MB
Nov 20 23:26
2
128 MB
part-m-00000
-rw-r--r--
srijith_unni2
hadoop
7.67 MB
Nov 20 23:26
2
128 MB
part-m-00001
-rw-r--r--
srijith_unni2
hadoop
7.3 MB
Nov 20 23:26
2
128 MB
part-m-00002
-rw-r--r--
srijith_unni2
hadoop
7.21 MB
Nov 20 23:26
2
128 MB
part-m-00003
-rw-r--r--
srijith_unni2
hadoop
2.65 MB
Nov 20 23:26
2
128 MB
part-m-00004
Showing 1 to 6 of 6 entries
Previous
1
Next
Hadoop, 2018.

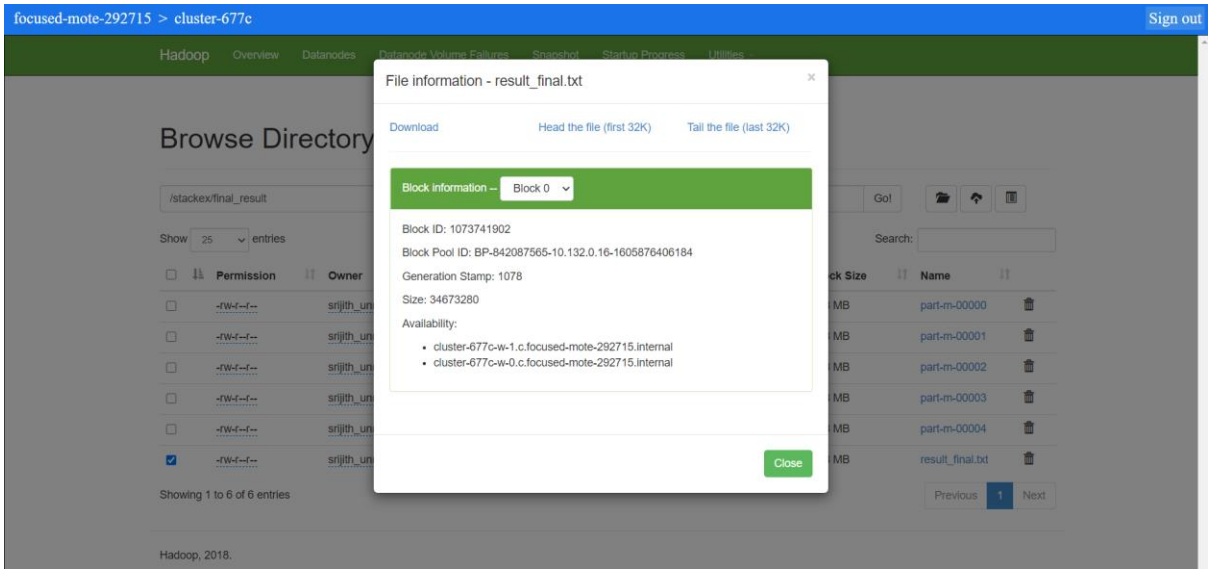

focused-mote-292715 > cluster-677c
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/stackex/output_final.txt
Go!
Cut
Paste
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
srijith_unni2
hadoop
0 B
Nov 20 23:26
2
128 MB
_SUCCESS
-rw-r--r--
srijith_unni2
hadoop
8.23 MB
Nov 20 23:26
2
128 MB
part-m-00000
-rw-r--r--
srijith_unni2
hadoop
7.67 MB
Nov 20 23:26
2
128 MB
part-m-00001
-rw-r--r--
srijith_unni2
hadoop
7.3 MB
Nov 20 23:26
2
128 MB
part-m-00002
-rw-r--r--
srijith_unni2
hadoop
7.21 MB
Nov 20 23:26
2
128 MB
part-m-00003
-rw-r--r--
srijith_unni2
hadoop
2.65 MB
Nov 20 23:26
2
128 MB
part-m-00004
Showing 1 to 6 of 6 entries
Previous
1
Next
Hadoop, 2018.

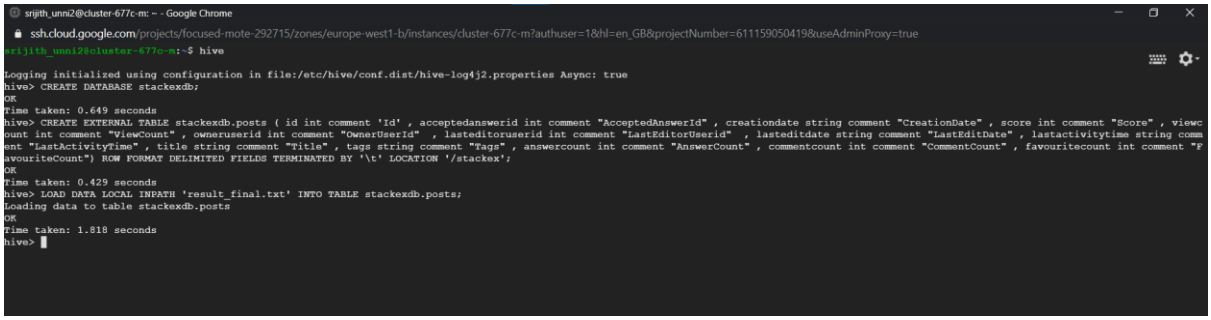

focused-mote-292715 > cluster-677c
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/stackex/final_result
Go!
Cut
Paste
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
No data available in table
Showing 0 to 0 of 0 entries
Previous
Next
Hadoop, 2018.


focused-mote-292715 > cluster-677c
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/stackex/final_result
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
srijith_unni2
hadoop
8.23 MB
Nov 20 23:26
2
128 MB
part-m-00000
-rw-r--r--
srijith_unni2
hadoop
7.67 MB
Nov 20 23:26
2
128 MB
part-m-00001
-rw-r--r--
srijith_unni2
hadoop
7.3 MB
Nov 20 23:26
2
128 MB
part-m-00002
-rw-r--r--
srijith_unni2
hadoop
7.21 MB
Nov 20 23:26
2
128 MB
part-m-00003
-rw-r--r--
srijith_unni2
hadoop
2.65 MB
Nov 20 23:26
2
128 MB
part-m-00004
Showing 1 to 5 of 5 entries
Previous
1
Next
Hadoop, 2018.

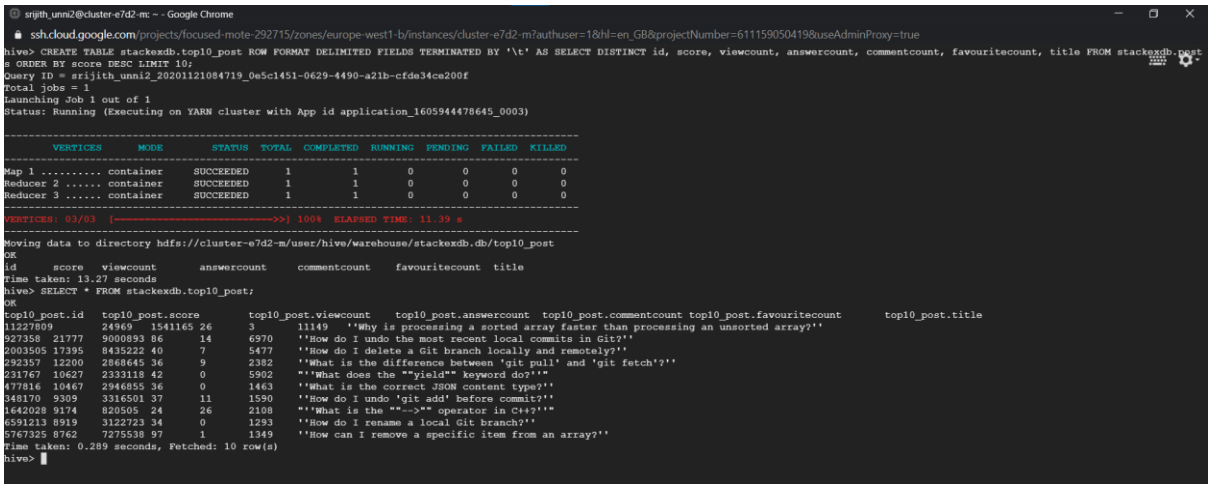

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
srijith_unni2@cluster-677c-m:~$ hadoop fs -ls /stackex/final_result
Found 5 items
-rw-r--r--   2 srijith_unni2 hadoop    8630105 2020-11-20 17:56 /stackex/final_result/part-m-00000
-rw-r--r--   2 srijith_unni2 hadoop    8044423 2020-11-20 17:56 /stackex/final_result/part-m-00001
-rw-r--r--   2 srijith_unni2 hadoop    7658192 2020-11-20 17:56 /stackex/final_result/part-m-00002
-rw-r--r--   2 srijith_unni2 hadoop    7556999 2020-11-20 17:56 /stackex/final_result/part-m-00003
-rw-r--r--   2 srijith_unni2 hadoop    2783561 2020-11-20 17:56 /stackex/final_result/part-m-00004
srijith_unni2@cluster-677c-m:~$ hadoop fs -getmerge /stackex/final_result/part-m-00000 /stackex/final_result/part-m-00001 /stackex/final_result/part-m-00002 /stackex/final_result/part-m-00003 /stackex/final_result/part-m-00004 /home/srijith_unni2/result_final.txt
srijith_unni2@cluster-677c-m:~$ hadoop fs -put /home/srijith_unni2/result_final.txt /stackex/final_result
srijith_unni2@cluster-677c-m:~$

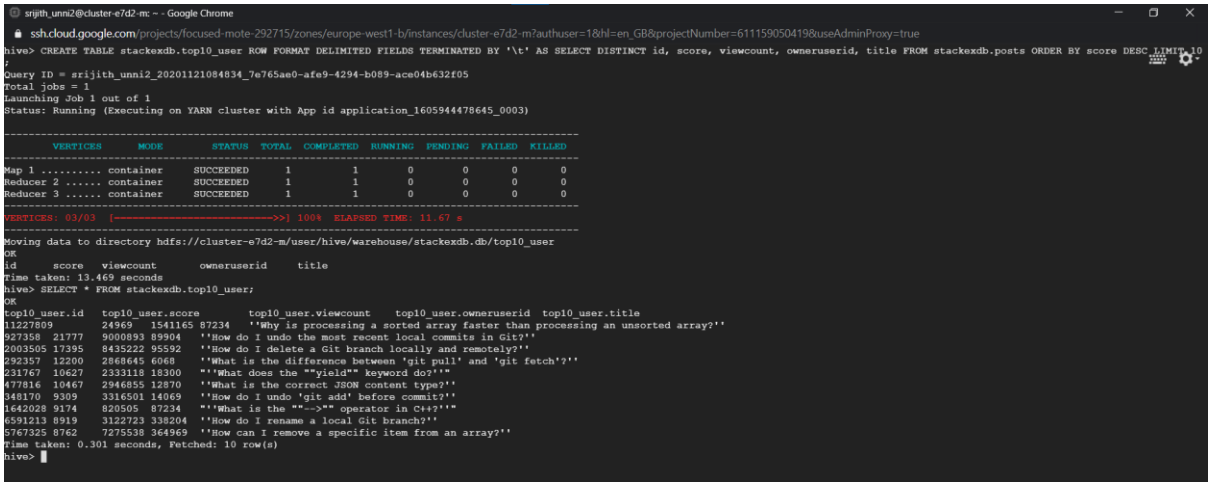

focused-mote-292715 > cluster-677c
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/stackex/final_result
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
-rw-r--r--
srijith_unni2
hadoop
8.23 MB
Nov 20 23:26
2
128 MB
part-m-00000
-rw-r--r--
srijith_unni2
hadoop
7.67 MB
Nov 20 23:26
2
128 MB
part-m-00001
-rw-r--r--
srijith_unni2
hadoop
7.3 MB
Nov 20 23:26
2
128 MB
part-m-00002
-rw-r--r--
srijith_unni2
hadoop
7.21 MB
Nov 20 23:26
2
128 MB
part-m-00003
-rw-r--r--
srijith_unni2
hadoop
2.65 MB
Nov 20 23:26
2
128 MB
part-m-00004
-rw-r--r--
srijith_unni2
hadoop
33.07 MB
Nov 20 23:37
2
128 MB
result_final.txt
Showing 1 to 6 of 6 entries
Previous
1
Next
Hadoop, 2018.

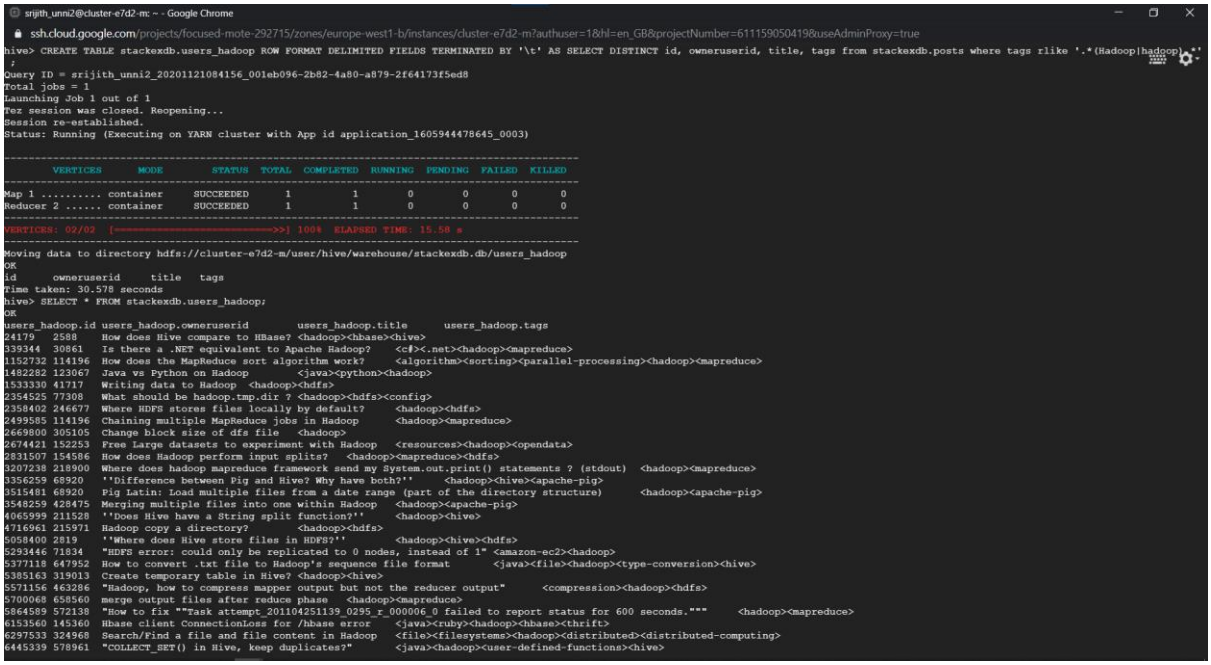

focused-mote-292715 > cluster-677c
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/stackex/final_result
Go!
Show 25 entries
Search:
Permission
Owner
Name
part-m-00000
part-m-00001
part-m-00002
part-m-00003
part-m-00004
result_final.txt
Showing 1 to 6 of 6 entries
Previous
1
Next
Hadoop, 2018.
File information - result_final.txt
Download
Head the file (first 32K)
Tail the file (last 32K)
Block information --
Block 0
Block ID: 1073741902
Block Pool ID: BP-842087565-10.132.0.16-1605876406184
Generation Stamp: 1078
Size: 34673280
Availability:
cluster-677c-w-1.c.focused-mote-292715.internal
cluster-677c-w-0.c.focused-mote-292715.internal
Close

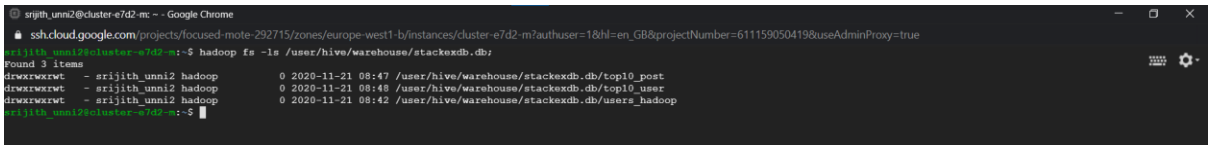

srijith_unni2@cluster-677c-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-677c-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
srijith_unni2@cluster-677c-m:~$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true
hive> CREATE DATABASE stackexdb;
OK
Time taken: 0.649 seconds
hive> CREATE EXTERNAL TABLE stackexdb.posts ( id int comment 'Id' , acceptedanswerid int comment "AcceptedAnswerId" , creationdate string comment "CreationDate" , score int comment "Score" , viewcount int comment "ViewCount" , owneruserid int comment "OwnerUserId" , lasteditoruserid int comment "LastEditorUserid" , lasteditdate string comment "LastEditDate" , lastactivitytime string comment "LastActivityTime" , title string comment "Title" , tags string comment "Tags" , answercount int comment "AnswerCount" , commentcount int comment "CommentCount" , favouritecount int comment "FavouriteCount") ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LOCATION '/stackex';
OK
Time taken: 0.429 seconds
hive> LOAD DATA LOCAL INPATH 'result_final.txt' INTO TABLE stackexdb.posts;
Loading data to table stackexdb.posts
OK
Time taken: 1.818 seconds
hive>

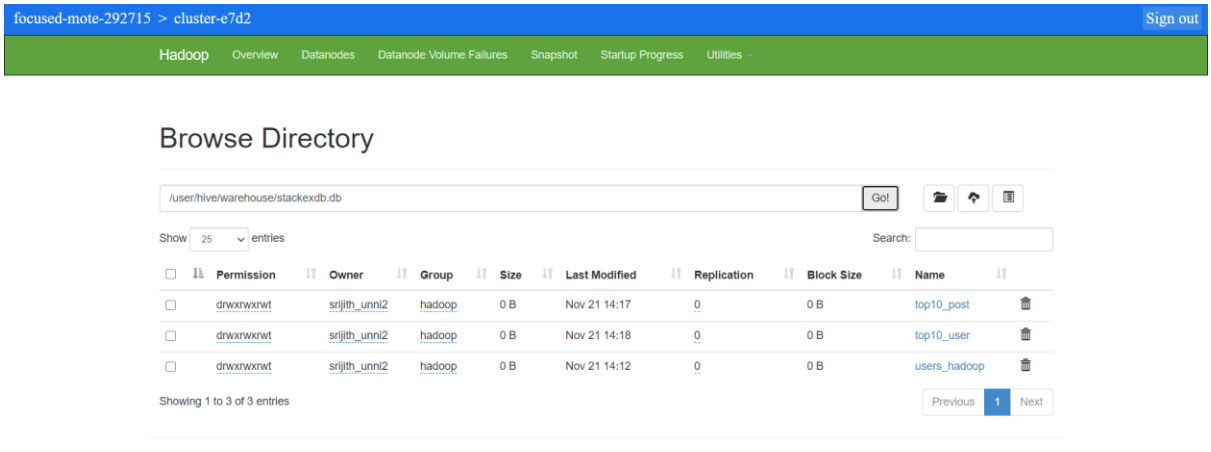

srijith_unni2@cluster-e7d2-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-e7d2-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
hive> set hive.cli.print.header=true;
hive>


srijith_unni2@cluster-e7d2-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-e7d2-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
hive> CREATE TABLE stackexdb.top10_post ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' AS SELECT DISTINCT id, score, viewcount, answercount, commentcount, favouritecount, title FROM stackexdb.posts ORDER BY score DESC LIMIT 10;
Query ID = srijith_unni2_20201121084719_0e5c1451-0629-4490-a21b-cfde34ce200f
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1605944478645_0003)
VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 2 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 3 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 03/03 [==========>>] 100% ELAPSED TIME: 11.39 s
Moving data to directory hdfs://cluster-e7d2-m/user/hive/warehouse/stackexdb.db/top10_post
OK
id score viewcount answercount commentcount favouritecount title
Time taken: 13.27 seconds
hive> SELECT * FROM stackexdb.top10_post;
OK
top10_post.id top10_post.score top10_post.viewcount top10_post.answercount top10_post.commentcount top10_post.favouritecount top10_post.title
11227809 24969 1541165 26 3 11149 ''Why is processing a sorted array faster than processing an unsorted array?''
927358 21777 9000893 86 14 6970 ''How do I undo the most recent local commits in Git?''
2003505 17395 8435222 40 7 5477 ''How do I delete a Git branch locally and remotely?''
292357 12200 2868645 36 9 2382 ''What is the difference between 'git pull' and 'git fetch'?''
231767 10627 2333118 42 0 5902 "''What does the ""yield"" keyword do?'"
477816 10467 2946855 36 0 1463 ''What is the correct JSON content type?''
348170 9309 3316501 37 11 1590 ''How do I undo 'git add' before commit?''
1642028 9174 820505 24 26 2108 "''What is the ""-->"" operator in C++?'"
6591213 8919 3122723 34 0 1293 ''How do I rename a local Git branch?''
5767325 8762 7275538 97 1 1349 ''How can I remove a specific item from an array?''
Time taken: 0.289 seconds, Fetched: 10 row(s)
hive>


srijith_unni2@cluster-e7d2-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-e7d2-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
hive> CREATE TABLE stackexdb.top10_user ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' AS SELECT DISTINCT id, score, viewcount, owneruserid, title FROM stackexdb.posts ORDER BY score DESC LIMIT 10
;
Query ID = srijith_unni2_20201121084834_7e765ae0-afe9-4294-b089-ace04b632f05
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1605944478645_0003)
VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 2 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 3 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 03/03 [==========>>] 100% ELAPSED TIME: 11.67 s
Moving data to directory hdfs://cluster-e7d2-m/user/hive/warehouse/stackexdb.db/top10_user
OK
id score viewcount owneruserid title
Time taken: 13.469 seconds
hive> SELECT * FROM stackexdb.top10_user;
OK
top10_user.id top10_user.score top10_user.viewcount top10_user.owneruserid top10_user.title
11227809 24969 1541165 87234 ''Why is processing a sorted array faster than processing an unsorted array?''
927358 21777 9000893 89904 ''How do I undo the most recent local commits in Git?''
2003505 17395 8435222 95592 ''How do I delete a Git branch locally and remotely?''
292357 12200 2868645 6068 ''What is the difference between 'git pull' and 'git fetch'?''
231767 10627 2333118 18300 "'What does the ""yield"" keyword do?'"
477816 10467 2946855 12870 ''What is the correct JSON content type?''
348170 9309 3316501 14069 ''How do I undo 'git add' before commit?''
1642028 9174 820505 87234 "'What is the ""-->"" operator in C++?'"
6591213 8919 3122723 338204 ''How do I rename a local Git branch?''
5767325 8762 7275538 364969 ''How can I remove a specific item from an array?''
Time taken: 0.301 seconds, Fetched: 10 row(s)
hive>


srijith_unni2@cluster-e7d2-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-e7d2-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
hive> CREATE TABLE stackexdb.users_hadoop ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' AS SELECT DISTINCT id, owneruserid, title, tags from stackexdb.posts where tags rlike '.*(Hadoop|hadoop).*'
;
Query ID = srijith_unni2_20201121084156_001eb096-2b82-4a80-a879-2f64173f5ed8
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1605944478645_0003)
VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 2 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 02/02 [==========>>] 100% ELAPSED TIME: 15.58 s
Moving data to directory hdfs://cluster-e7d2-m/user/hive/warehouse/stackexdb.db/users_hadoop
OK
id owneruserid title tags
Time taken: 30.578 seconds
hive> SELECT * FROM stackexdb.users_hadoop;
OK
users_hadoop.id users_hadoop.owneruserid users_hadoop.title users_hadoop.tags
24179 2588 How does Hive compare to HBase? <hadoop><hbase><hive>
339344 30861 Is there a .NET equivalent to Apache Hadoop? <c#><.net><hadoop><mapreduce>
1152732 114196 How does the MapReduce sort algorithm work? <algorithm><sorting><parallel-processing><hadoop><mapreduce>
1482282 123067 Java vs Python on Hadoop <java><python><hadoop>
1533330 41717 Writing data to Hadoop <hadoop><hdfs>
2354525 77308 What should be hadoop.tmp.dir ? <hadoop><hdfs><config>
2358402 246677 Where HDFS stores files locally by default? <hadoop><hdfs>
2499585 114196 Chaining multiple MapReduce jobs in Hadoop <hadoop><mapreduce>
2669800 305105 Change block size of dfs file <hadoop>
2674421 152253 Free Large datasets to experiment with Hadoop <resources><hadoop><opendata>
2831507 154586 How does Hadoop perform input splits? <hadoop><mapreduce><hdfs>
3207238 218900 Where does hadoop mapreduce framework send my System.out.print() statements ? (stdout) <hadoop><mapreduce>
3356259 68920 ''Difference between Pig and Hive? Why have both?'' <hadoop><hive><apache-pig>
3515481 68920 Pig Latin: Load multiple files from a date range (part of the directory structure) <hadoop><apache-pig>
3548259 428475 Merging multiple files into one within Hadoop <hadoop><apache-pig>
4065999 211528 ''Does Hive have a String split function?'' <hadoop><hive>
4716961 215971 Hadoop copy a directory? <hadoop><hdfs>
5058400 2819 ''Where does Hive store files in HDFS?'' <hadoop><hive><hdfs>
5293446 71834 "HDFS error: could only be replicated to 0 nodes, instead of 1" <amazon-ec2><hadoop>
5377118 647952 How to convert .txt file to Hadoop's sequence file format <java><file><hadoop><type-conversion><hive>
5385163 319013 Create temporary table in Hive? <hadoop><hive>
5571156 463286 "Hadoop, how to compress mapper output but not the reducer output" <compression><hadoop><hdfs>
5700068 658560 merge output files after reduce phase <hadoop><mapreduce>
5864589 572138 "How to fix ""task attempt_201104251139_0295_r_000006_0 failed to report status for 600 seconds.""" <hadoop><mapreduce>
6153560 145360 Hbase client ConnectionLoss for /hbase error <java><ruby><hadoop><hbase><thrift>
6297533 324968 Search/Find a file and file content in Hadoop <file><filesystems><hadoop><distributed><distributed-computing>
6445339 578961 "COLLECT_SET() in Hive, keep duplicates?" <java><hadoop><user-defined-functions><hive>


srijith_unni2@cluster-e7d2-m: ~ - Google Chrome
ssh.cloud.google.com/projects/focused-mote-292715/zones/europe-west1-b/instances/cluster-e7d2-m?authuser=1&hl=en_GB&projectNumber=611159050419&useAdminProxy=true
srijith_unni2@cluster-e7d2-m:~$ hadoop fs -ls /user/hive/warehouse/stackexdb.db;
Found 3 items
drwxrwxrwt - srijith_unni2 hadoop 0 2020-11-21 08:47 /user/hive/warehouse/stackexdb.db/top10_post
drwxrwxrwt - srijith_unni2 hadoop 0 2020-11-21 08:48 /user/hive/warehouse/stackexdb.db/top10_user
drwxrwxrwt - srijith_unni2 hadoop 0 2020-11-21 08:42 /user/hive/warehouse/stackexdb.db/users_hadoop
srijith_unni2@cluster-e7d2-m:~$


focused-mote-292715 > cluster-e7d2
Sign out
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/user/hive/warehouse/stackexdb.db
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
drwxrwxrwt srijith_unni2 hadoop 0 B Nov 21 14:17 0 0 B top10_post
drwxrwxrwt srijith_unni2 hadoop 0 B Nov 21 14:18 0 0 B top10_user
drwxrwxrwt srijith_unni2 hadoop 0 B Nov 21 14:12 0 0 B users_hadoop
Showing 1 to 3 of 3 entries
Previous
1
Next
Hadoop, 2018.