

Государственное автономное образовательное
учреждение высшего образования «Московский городской педагогический
университет»

Институт цифрового образования
Департамент информатики, управления и технологий

Отчет по лабораторной работе № 3.2

Выполнил: студент БД-231
Кряжева К.С.

2024 г

```
cloudera@quickstart:~/ex_3_2
File Edit View Search Terminal Help
Documents      geolocation.zip      Music           trucks.csv
Downloads      kerberos             parcels         Videos
eclipse        Kryazheva.gz         part-r-000000  workspace
[cloudera@quickstart ~]$ mkdir ex_3_2
[cloudera@quickstart ~]$ ls
cloudera-manager  ex_3_2              lib             Templates
cm_api.py         express-deployment.json  __MACOSX       trucks.csv
Desktop           geolocation.csv       Music           Videos
Documents         geolocation.zip       parcels         workspace
Downloads         kerberos              part-r-000000
eclipse           Kryazheva.gz          Pictures
enterprise-deployment.json Kryazheva.txt        Public
[cloudera@quickstart ~]$ cd ex_3_2/
[cloudera@quickstart ex_3_2]$ ls
[cloudera@quickstart ex_3_2]$ wget https://github.com/BosenkoTM/cloudera-quickstart/blob/main/data/geolocation.zip
--2024-04-05 09:22:54-- https://github.com/BosenkoTM/cloudera-quickstart/blob/main/data/geolocation.zip
Resolving github.com... 140.82.121.4
Connecting to github.com|140.82.121.4|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: "geolocation.zip"
```

```
cloudera@quickstart:~/ex_3_2
File Edit View Search Terminal Help
Length: unspecified [text/html]
Saving to: "geolocation.csv"

[ <=> ] 842,986 1.95M/s in 0.4s

2024-04-05 09:26:03 (1.95 MB/s) - "geolocation.csv" saved [842986]

[cloudera@quickstart ex_3_2]$ wget https://github.com/BosenkoTM/cloudera-quickst
art/blob/main/data/trucks.csv
--2024-04-05 09:26:48-- https://github.com/BosenkoTM/cloudera-quickstart/blob/m
ain/data/trucks.csv
Resolving github.com... 140.82.121.3
Connecting to github.com|140.82.121.3|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/html]
Saving to: "trucks.csv"

[ <=> ] 232,131 1.18M/s in 0.2s

2024-04-05 09:26:48 (1.18 MB/s) - "trucks.csv" saved [232131]

[cloudera@quickstart ex_3_2]$ ls
geolocation.csv geolocation.zip trucks.csv
[cloudera@quickstart ex_3_2]$
```

cloudera-quickstart-vm-5.13.0-0-virtualbox [Pa60raet] - Oracle VM VirtualBox

Applications Places System Fri Apr 5, 9:46 AM cloudera

cloud

File Edit View Search Termin Length: unspecified [text/html] Saving to: "geolocation.csv"

[<=>]

2024-04-05 09:26:03 (1.95 MB/s)

[cloudera@quickstart ex_3_2]\$ art/blob/main/data/trucks.csv --2024-04-05 09:26:48-- https://github.com/BosenkoTM/cloudera-quickstart/blob/main/data/trucks.csv Resolving github.com... 140.82.121.3 Connecting to github.com|140.82.121.3|:443... connected. HTTP request sent, awaiting response... 200 OK Length: unspecified [text/html] Saving to: "trucks.csv"

[<=>]

2024-04-05 09:26:48 (1.18 MB/s)

[cloudera@quickstart ex_3_2]\$ geolocation.csv geolocation.zip trucks.csv [cloudera@quickstart ex_3_2]\$

Hue - Editor - Mozilla Firefox

about:sessionrestore x Oozie Web Console x Hue - Editor x

quickstart.cloudera:8888/hue/editor?editor=33

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Search data and saved documents...

Tables (2) 201402_trip_data athlete

1 geoloc = LOAD 'geoloc/geolocation.csv' USING PigStorage(',') AS (truckid:chararray, driverid:chararray, event:chararray, latitude:double, longitude:double, city:chararray, state:chararray, velocity:double, event_ind:long, idling_ind:long); 2 geoloc_limit = LIMIT geoloc 10; 3 STORE geoloc_limit INTO 'results-geoloc'; 4 DUMP geoloc_limit;

Query History 4 minutes ago geoloc = LOAD 'geoloc/geolocation.csv' USING PigStorage(',') AS (truckid:chararray, driverid:chararray, event:chararray, latitude:double, longitude:double, city:chararray, state:chararray, velocity:double, event_ind:long, idling_ind:long); geoloc_limit = LIMIT geoloc 10; STORE geoloc_limit INTO 'results-geoloc'; DUMP geoloc_limit;

cloudera-quickstart-vm-5.13.0-0-virtualbox [Pa60raer] - Oracle VM VirtualBox

Applications Places System

Fri Apr 5, 9:46 AM cloudera

cloud

File Edit View Search Termin

length: unspecified [text/html]
aving to: "geolocation.csv"

[< >]

124-04-05 09:26:03 (1.95 MB/s)

:cloudera@quickstart ex 3 2]\$
t/blob/main/data/trucks.csv
2024-04-05 09:26:48-- https://
in/data/trucks.csv
solving github.com... 140.82
nnecting to github.com[140.82]
TP request sent, awaiting re
length: unspecified [text/html]
aving to: "trucks.csv"

[< >]

124-04-05 09:26:48 (1.18 MB/s)

:cloudera@quickstart ex 3 2]\$
oolocation.csv geolocation.z
:cloudera@quickstart ex 3 2]\$

Oozie Web Console - Mozilla Firefox

about:sessionrestore x Oozie Web Console x Hue - Editor x

quickstart.cloudera:11000/oozie/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Oozie Web Console

Workflow Jobs Coordinator Jobs Bundle Jobs SLA System Info Instrumentation Metrics Settings

All Jobs Active Jobs Done Jobs Custom Filter

Job Id	Name	Status	Run	User	Group	Created	Started	Last Modified	Ended
1 0000000-240405090405368-oozie-oozi-W	Batch job for qu...	RUNNING	0	cloudera		Fri, 05 Apr 2024 16:43:19 GMT	Fri, 05 Apr 2024 16:43:20 GMT	Fri, 05 Apr 2024 16:43:25 GMT	

File Edit View

length: unspecified
aving to: "geol

[< >]

2024-04-05 09:26:48 (1.18 MB/s)

[cloudera@quicks
rt/blob/main/d
--2024-04-05 09:
in/data/trucks.
esolving github
nnecting to ga
HTTP request sen
length: unspecif
aving to: "trud

[< >]

2024-04-05 09:26:48 (1.18 MB/s)

[cloudera@quicks
geolocation.csv
[cloudera@quicks

Oozie Web Console - Mozilla Firefox

about:sessionrestore x Oozie Web Console x Hue - Editor x

quickstart.cloudera:11000/oozie/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Job (Name: Batch job for query-pig|JobId: 0000000-240405090405368-oozie-oozi-W)

Job (Name: Batch job for query-pig|JobId: 0000000-240405090405368-oozie-oozi-W)

Job Info Job Definition Job Configuration Job Log Job DAG

start

End

End

End

All Applications - Mozilla Firefox

about:sessionrestore x Oozie Web Console x Hue - Editor x All Applications x

quickstart.cloudera:8088/cluster

Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores
3	0	1	2	2	3 GB	8 GB	0 B	2	8

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
1	0	0	0	0

User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCoers	Reserved Memory MB
application_1712332974359_0003	cloudera	PigLatin:pig-1cc3.pig	MAPREDUCE	root.cloudera	Fri Apr 5 09:48:23 -0700 2024	Fri Apr 5 09:49:16 -0700 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A
application_1712332974359_0002	cloudera	PigLatin:pig-1cc3.pig	MAPREDUCE	root.cloudera	Fri Apr 5 09:45:30 -0700 2024	Fri Apr 5 09:48:12 -0700 2024	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A
application_1712332974359_0001	cloudera	oozie:launcher:T=pig;W=Batch job for query-pig:A=pig-1cc3;ID=0000000-240405090405368-oozie-oozi-W	MAPREDUCE	root.cloudera	Fri Apr 5 09:43:24 -0700 2024	N/A	RUNNING	UNDEFINED	2	2	3072	0	0

Showing 1 to 3 of 3 entries

cloudera-quickstart-vm-5.13.0-0-virtualbox [Pa60taer] - Oracle VM VirtualBox

Applications Places System Fri Apr 5, 9:55 AM cloudera

Click to view your appointments and tasks - Mozilla Firefox

about:sessionrestore x Oozie Web Console x Hue - File Browser x All Applications x

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/results-geoloc/part-r-00000

Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE

Query

Search data and saved documents...

File Browser

View as binary

Edit file

Download

View file location

Refresh

Last modified 04/05/2024 4:49 PM

User cloudera

Group cloudera

Size 630 B

Mode 100644

Home / user / cloudera / results-geoloc / part-r-00000

A19	A19	normal	37.962146	-122.345526	San Pablo	California	0.0	0	1
A20	A20	normal	36.977173	-121.899402	Aptos	California	27.0	0	0
A31	A31	normal	39.409608	-123.355566	Willits	California	22.0	0	0
A40	A40	overspeed	37.957702	-121.29078	Stockton	California	77.0	1	0
A50	A50	normal	38.40765	-122.947713	Occidental	California	0.0	0	1
A51	A51	normal	37.639097	-120.996078	Modesto	California	0.0	0	1
A54	A54	normal	38.440467	-122.714431	Santa Rosa	California	17.0	0	0
A71	A71	normal	33.683947	-117.794694	Irvine	California	43.0	0	0
A77	A77	normal	37.962146	-122.345526	San Pablo	California	25.0	0	0
truckid	driverid	event		city	state				

HUE

Search data and saved documents...





Job Browser Jobs Workflows Schedules Bundles SLAs

default user:cloudera [Succeeded] [Running] [Failed in the last 7 days]

ID	Name	User	Type	Status	Progress	Group	Started	Duration
application_1712332974359_0006	oozie.launcher:T=pig W=Batch job for query-pig-A:pig-2621-ID=0000001-240405090405368-oozie-ooz-W	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	April 5, 2024 10:02 AM	21.19s
application_1712332974359_0005	PigLatin.pig-1cc3.pig	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	April 5, 2024 9:53 AM	46.56s
application_1712332974359_0004	PigLatin.pig-1cc3.pig	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	April 5, 2024 9:51 AM	1m, 10s
application_1712332974359_0003	PigLatin.pig-1cc3.pig	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	April 5, 2024 9:48 AM	52.95s
application_1712332974359_0002	PigLatin.pig-1cc3.pig	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	April 5, 2024 9:45 AM	2m, 41s
application_1712332974359_0001	oozie.launcher:T=pig W=Batch job for query-pig-A:pig-1cc3-ID=0000000-240405090405368-oozie-ooz-W	cloudera	MAPREDUCE	SUCCEEDED	100%	root.cloudera	April 5, 2024 9:43 AM	12m, 38s

6 jobs

[Home](#) / [user](#) / [cloudera](#) / [geoloc](#)

<input type="checkbox"/>	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	 f		cloudera	cloudera	drwxr-xr-x	April 14, 2024 10:55 AM
<input type="checkbox"/>	 .		cloudera	cloudera	drwxr-xr-x	April 14, 2024 11:02 AM
<input type="checkbox"/>	 geolocation.csv	514.3 KB	cloudera	cloudera	-rw-r--r--	April 14, 2024 11:02 AM
<input type="checkbox"/>	 trucks.csv	59.9 KB	cloudera	cloudera	-rw-r--r--	April 14, 2024 11:02 AM

Show of 2 items

Page of 1 ⏪ ⏩ ⏴ ⏵

Restore Session x Hue - Editor x Oozie Web Console x

quickstart.cloudera:8888/hue/editor?editor=45

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Search data and saved documents...

Pig Add a name... Add a description...

58.52s

```
1 geoloc = LOAD 'geoloc/geolocation.csv' USING PigStorage(',') AS (truckid:chararray,
2 driverid:chararray, event:chararray, latitude:double, longitude:double, city:chararray,
3 state:chararray, velocity:double, event_ind:long, idling_ind:long);
4 geoloc_limit = LIMIT geoloc 10;
5 STORE geoloc_limit INTO 'results-geoloc';
6 DUMP geoloc_limit;
```

2024-04-14 11:06:54,175 INFO WorkflowNotificationXCommand:520 - SERVER[quickstart.cloudera] USER[-] GROUP[-] TOKEN[-] APP[-] JOB[0000000-240414103048965-oozi-e-oozi-W] ACTION[] No Notification URL is defined. Therefore nothing to notify for job 0000000-240414103048965-oozi-e-oozi-W job_1713115779898_0001

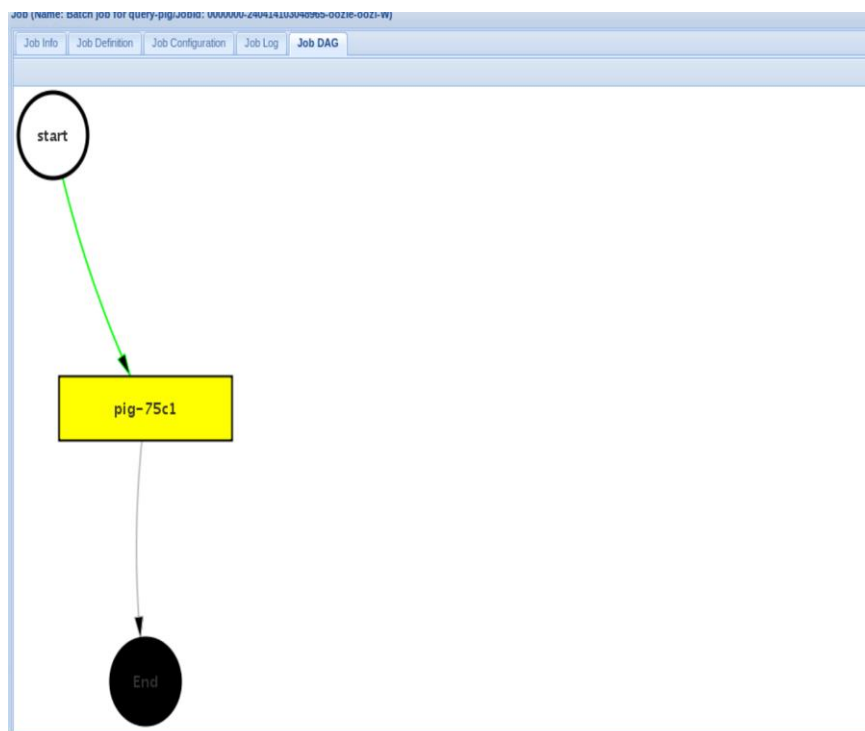
2024-04-14 11:06:54,219 INFO ActionStartXCommand:520 - SERVER[quickstart.cloudera] USER[cloudera] GROUP[-] TOKEN[] APP[Batch job for query-pig] JOB[0000000-240414103048965-oozi-e-oozi-W] ACTION[0000000-240414103048965-oozi-e-oozi-W@pig-75c1] Start action [0000000-240414103048965-oozi-e-oozi-W@pig-75c1] with user-retry state : userRetryCount [0], userRetryMax [0], userRetryInterval [10]

2024-04-14 11:06:57,164 INFO PigActionExecutor:520 - SERVER[quickstart.cloudera] USER[cloudera] GROUP[-] TOKEN[] APP[Batch job for query-pig] JOB[0000000-240414103048965-oozi-e-oozi-W] ACTION[0000000-240414103048965-oozi-e-oozi-W@pig-75c1] checking action, hadoop job ID [job_1713115779898_0001] status [RUNNING]

2024-04-14 11:06:57,167 INFO ActionStartXCommand:520 - SERVER[quickstart.cloudera] USER[cloudera] GROUP[-] TOKEN[] APP[Batch job for query-pig] JOB[0000000-240414103048965-oozi-e-oozi-W] ACTION[0000000-240414103048965-oozi-e-oozi-W@pig-75c1] [***0000000-240414103048965-oozi-e-oozi-W@pig-75c1***]Action status=RUNNING

2024-04-14 11:06:57,168 INFO ActionStartXCommand:520 - SERVER[quickstart.cloudera] USER[cloudera] GROUP[-] TOKEN[] APP[Batch job for query-pig] JOB[0000000-240414103048965-oozi-e-oozi-W] ACTION[0000000-240414103048965-oozi-e-oozi-W@pig-75c1] [***0000000-240414103048965-oozi-e-oozi-W@pig-75c1***]Action updated in DB!

2024-04-14 11:06:57,181 INFO WorkflowNotificationXCommand:520 - SERVER[quickstart.cloudera] USER[-] GROUP[-] TOKEN[-] APP[-] JOB[0000000-240414103048965-oozi-e-oozi-W] ACTION[0000000-240414103048965-oozi-e-oozi-W@pig-75c1] No Notification URL is defined. Therefore nothing to notify for job 0000000-240414103048965-oozi-e-oozi-W@pig-75c1



Hue - Editor - Mozilla Firefox

quickstart.cloudera:8888/hue/editor/editor=45

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Search data and saved documents...

Header

1 >>> Invoking Pig command line now >>>

2

3

4 Run pig script using PigRunner.run() for Pig version 0.8+

5 Apache Pig version 0.12.0-cdh5.13.0 (exported)

6 compiled Oct 04 2017, 11:09:03

7

8 Run pig script using PigRunner.run() for Pig version 0.8+

9 2024-04-14 11:08:28,452 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.13.0 (exported) compiled Oct 04 2017, 11:09:03

10 2024-04-14 11:08:28,453 [main] INFO org.apache.pig.Main - Logging error messages to: /var/lib/hadoop-yarn/cache/yarn-local-dir/usercache/cloudera/appcache/application_17

11 2024-04-14 11:08:28,715 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /var/lib/hadoop-yarn/pigbootstrap not found

12 2024-04-14 11:08:29,005 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

13 2024-04-14 11:08:29,006 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

14 2024-04-14 11:08:29,006 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020

15 2024-04-14 11:08:29,028 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8032

16 2024-04-14 11:08:29,031 [main] WARN org.apache.pig.PigServer - Empty string specified for jar path

17 2024-04-14 11:08:31,173 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT

18 2024-04-14 11:08:31,349 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColur

19 2024-04-14 11:08:34,205 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.textoutputformat.separator is deprecated. Instead, use mapreduce.output.textoutput

Hue - Editor - Mozilla Firefox

quickstart.cloudera:8888/hue/editor/editor=45

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Query Search data and saved documents...

Header

304 2024-04-14 11:14:40,417 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

305 2024-04-14 11:14:40,417 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

306 2024-04-14 11:14:40,417 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.

307 2024-04-14 11:14:40,423 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

308 2024-04-14 11:14:40,423 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

309 (A19,A19,nomal,37.962146,-122.345526,San Pablo,California,0.0,0,1)

310 (A20,A20,nomal,36.977173,-121.899402,Aptos,California,27.0,0,0)

311 (A31,A31,nomal,39.409608,-123.355566,Willits,California,22.0,0,0)

312 (A40,A40,overspeed,37.957702,-121.29078,Stockton,California,77.0,1,0)

313 (A50,A50,nomal,38.40765,-122.947713,Occidental,California,0.0,0,1)

314 (A51,A51,nomal,37.639097,-120.996678,Modesto,California,0.0,0,1)

315 (A54,A54,nomal,38.440467,-122.714431,Santa Rosa,California,17.0,0,0)

316 (A71,A71,nomal,33.683947,-117.794694,Irvine,California,43.0,0,0)

317 (A77,A77,nomal,37.962146,-122.345526,San Pablo,California,25.0,0,0)

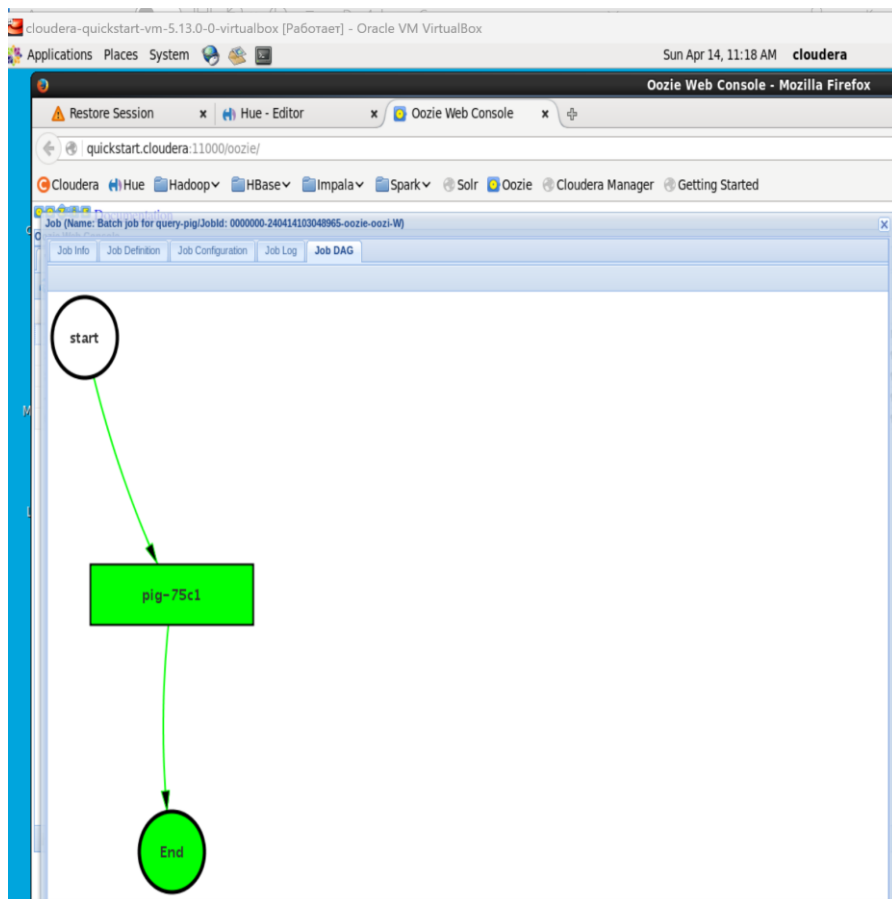
318 (truckId,driverId,event_city,state,...)

319 Hadoop Job IDs executed by Pig job_1713115779898_0004.job_1713115779898_0005

320

321 <<< Invocation of Pig command completed <<<

All Jobs Active Jobs Done Jobs Custom Filter										
Job Id	Name	Status	Run	User	Group	Created	Started	Last Modified	Ended	
1 0000000-240414103048965-oozie-oozi-W	Batch job for qu...	SUCCEE...	0	cloudera		Sun, 14 Apr 2024 18:06:53 GMT	Sun, 14 Apr 2024 18:06:53 GMT	Sun, 14 Apr 2024 18:14:59 GMT	Sun, 14 Apr 2024 18:14:59 GMT	
2 0000003-240405090405368-oozie-oozi-W	Batch job for qu...	SUSPEN...	0	cloudera		Fri, 05 Apr 2024 17:14:05 GMT	Fri, 05 Apr 2024 17:14:05 GMT	Fri, 05 Apr 2024 17:17:05 GMT	Fri, 05 Apr 2024 17:17:05 GMT	
3 0000002-240405090405368-oozie-oozi-W	Batch job for qu...	KILLED	0	cloudera		Fri, 05 Apr 2024 17:10:50 GMT	Fri, 05 Apr 2024 17:10:50 GMT	Fri, 05 Apr 2024 17:12:22 GMT	Fri, 05 Apr 2024 17:12:14 GMT	
4 0000001-240405090405368-oozie-oozi-W	Batch job for qu...	SUSPEN...	0	cloudera		Fri, 05 Apr 2024 17:02:20 GMT	Fri, 05 Apr 2024 17:02:20 GMT	Fri, 05 Apr 2024 17:05:18 GMT	Fri, 05 Apr 2024 17:05:18 GMT	
5 0000000-240405090405368-oozie-oozi-W	Batch job for qu...	KILLED	0	cloudera		Fri, 05 Apr 2024 16:43:19 GMT	Fri, 05 Apr 2024 16:43:20 GMT	Fri, 05 Apr 2024 17:02:03 GMT	Fri, 05 Apr 2024 17:02:03 GMT	



otaet] - Oracle VM VirtualBox

Sun Apr 14, 11:20 AM cloudera

Hue - File Browser - Mozilla Firefox

Browser x Oozie Web Console x

view=/user/cloudera/results-geoloc/part-r-00000

Impala Spark Solr Oozie Cloudera Manager Getting Started

Search data and saved documents...

File Browser

View as binary Home / user / cloudera / results-geoloc / part-r-00000 Page

Edit file

Download

View file location

Refresh

Last modified: 4/14/2024 6:12 PM

A19	A19	normal	37.962146	-122.345526	San Pablo	California	0.0	0	1
A20	A20	normal	36.977173	-121.899402	Aptos	California	27.0	0	0
A31	A31	normal	39.489688	-123.355566	Willits	California	22.0	0	0
A40	A40	overspeed	37.957782	-121.29078	Stockton	California	77.0	1	0
A50	A50	normal	38.40765	-122.947713	Occidental	California	0.0	0	1
A51	A51	normal	37.639097	-120.996878	Modesto	California	0.0	0	1
A54	A54	normal	38.440467	-122.714431	Santa Rosa	California	17.0	0	0
A71	A71	normal	33.683947	-117.794694	Irvine	California	43.0	0	0
A77	A77	normal	37.962146	-122.345526	San Pablo	California	25.0	0	0
truckid	driverid	event		city	state				

3-2.4 Посчитать статистику по файлу.

Add a name...

Add a description...

2m, 37s ?

```
1 geoloc = LOAD 'geoloc/geolocation.csv' USING PigStorage(',') AS (truckid:chararray, driverid:chararray, event:chararray, latitude:double, l
2 truck_ids = GROUP geoloc BY truckid;
3 result = FOREACH truck_ids GENERATE group AS truckid, COUNT(geoloc) as count;
4 STORE result INTO 'results2';
5 DUMP result;
```

< >

>>> Invoking Main class now >>>

job_1713115779898_0007

INFO: loading log4j config file log4j.properties.
INFO: log4j config file log4j.properties loaded successfully.

Oozie Pig action configuration
=====

Setting [tez.application.tags] tag: oozie-7baa7cc809584ccb69137b261c86c2c2

Setting env property for mapreduce.job.credentials.binary to:/var/lib/hadoop-yarn/cache/yarn/nm-local-dir/usercache/cloudera/appcache/application_1713115779898_0007/container_1713115779898_0007_01_000002/container_tokens

pig.properties:

Query History Saved Queries Results (277)

Header

Soraet] - Oracle VM VirtualBox

Sun Apr 14, 11:29 AM cloudera

Hue - Editor - Mozilla Firefox

Oozie Web Console

or=51

Search

Impala Spark Solr Oozie Cloudera Manager Getting Started

Search data and saved documents...

Header

1 >>> Invoking Pig command line now >>>

2

3

4 Run pig script using PigRunner.run() for Pig version 0.8+

5 Apache Pig version 0.12.0-cdh5.13.0 (reexported)

6 compiled Oct 04 2017, 11:09:03

7

8 Run pig script using PigRunner.run() for Pig version 0.8+

9 2024-04-14 11:26:16,169 [main] INFO org.apache.pig.Main - Apache Pig version 0.12.0-cdh5.13.0 (reexported) compiled Oct 04 2017, 11:09:03

10 2024-04-14 11:26:16,170 [main] INFO org.apache.pig.Main - Logging error messages to: /var/lib/hadoop-yarn/cache/yarn/nm-local-dir/usercache/cloudera/appcache/application_17

11 2024-04-14 11:26:16,214 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file /var/lib/hadoop-yarn/pigbootstrap not found

12 2024-04-14 11:26:16,297 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address

13 2024-04-14 11:26:16,297 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

14 2024-04-14 11:26:16,297 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020

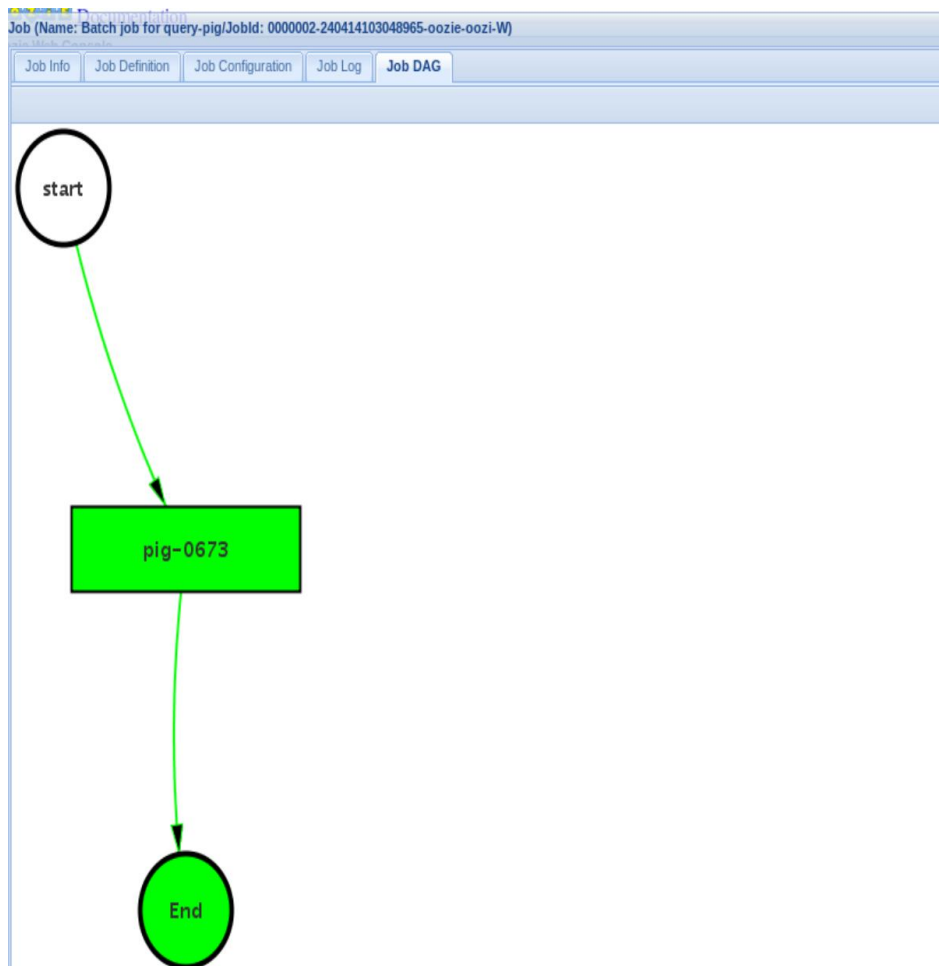
15 2024-04-14 11:26:16,301 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to map-reduce job tracker at: localhost:8032

16 2024-04-14 11:26:16,303 [main] WARN org.apache.pig.PigServer - Empty string specified for jar path

17 2024-04-14 11:26:16,716 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY

18 2024-04-14 11:26:16,745 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - (RULES_ENABLED={AddForEach, ColumnMapKeyPrune, DuplicateForEachColour

1 0000002-240414103048965-oozie-oozi-W Batch job for qu... SUCCEE... 0 cloudera Sun, 14 Apr 2024 18:26:02 GMT Sun, 14 Apr 2024 18:26:02 GMT Sun, 14 Apr 2024 18:28:40 GMT Sun, 14 Apr 2024 18:28:40 GMT



File Browser

View as binary

Home

/ user / cloudera / results2 / part-r-00000

Edit file

Download

View file location

Refresh

Last modified
04/14/2024 6:27 PM

User
cloudera

Group
cloudera

Size
702 B

Mode
100644

A1	80
A2	80
A3	80
A4	80
A5	80
A6	80
A7	80
A8	80
A9	80
A10	80
A11	80
A12	80
A13	80
A14	80
A15	80
A16	80
A17	80
A18	80
A19	80
A20	80
A21	80
A22	80
A23	80
A24	80
A25	80

