*Executing Hadoop Commands*

1(a) Create a directory in HDFS with the name of Assignment 1 and sub folder - Amazon

[cloudera@quickstart ~]$ hadoop fs -mkdir Assignment1

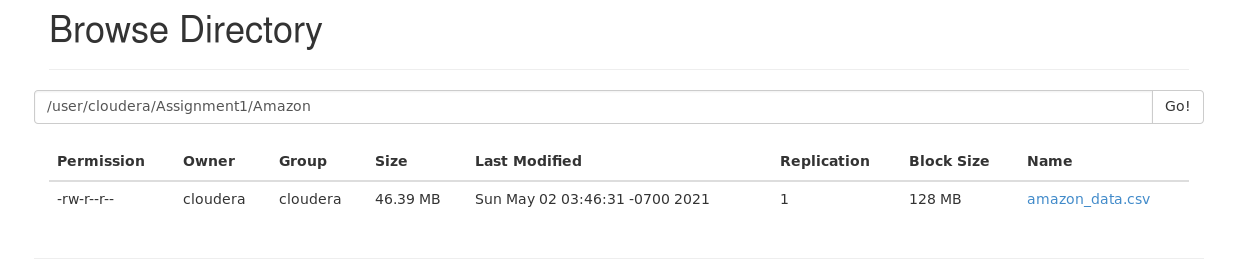
1(b) Put the CSV file into the Amazon folder

[cloudera@quickstart ~]$ hadoop fs –put

/home/cloudera/Desktop/Amazon/user/cloudera/Assignment1

1(c) View the data using Cat Command

[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/Assignment1/Amazon/amazon\_data.csv



*Execution of Hive Queries*

Question- 2.

Show the name of top 10 customers whose reviews are most recommended by others along with number of recommendation.

hive> create table Amazon(id string,name string,asins string,brand string,categories string,keys string,manufacturer string,reviews\_date string,reviews\_dateAdded string,reviews\_dateSeen string,reviews\_didPurchase string,reviews\_doRecommend string,reviews\_id string,reviews\_numHelpful int,reviews\_rating int,reviews\_sourceURLs string,reviews\_text string,reviews\_title string,reviews\_userCity string,reviews\_userProvince string,reviews\_username string)

> row format delimited

> fields terminated by ','

> stored as textfile;

OK

Time taken: 4.021 seconds

hive> load data inpath '/user/cloudera/Assignment1/Amazon/amazon\_data.csv' into table Amazon

> ;

Loading data to table default.amazon

Table default.amazon stats: [numFiles=1, totalSize=48647008]

OK

Time taken: 1.762 seconds



Question- 3.

Compare the number of product rated in various years.

hive> select year(reviews\_date),sum(reviews\_numHelpful) as tc from Amazon group by year(reviews\_date) order by tc desc limit 15;

Query ID = cloudera\_20210502042929\_9414b467-50de-4c30-b354-75dec55f6b1d

Total jobs = 2

Launching Job 1 out of 2

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1619950792006\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1619950792006\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1619950792006\_0003

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2021-05-02 04:29:59,516 Stage-1 map = 0%, reduce = 0%

2021-05-02 04:30:28,807 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 10.63 sec

2021-05-02 04:30:51,329 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 14.25 sec

MapReduce Total cumulative CPU time: 14 seconds 250 msec

Ended Job = job\_1619950792006\_0003

Launching Job 2 out of 2

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1619950792006\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1619950792006\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1619950792006\_0004

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2021-05-02 04:31:20,374 Stage-2 map = 0%, reduce = 0%

2021-05-02 04:31:38,113 Stage-2 map = 100%, reduce = 0%

2021-05-02 04:32:00,275 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 7.02 sec

MapReduce Total cumulative CPU time: 7 seconds 20 msec

Ended Job = job\_1619950792006\_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 14.25 sec HDFS Read: 48657057 HDFS Write: 316 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.02 sec HDFS Read: 5409 HDFS Write: 88 SUCCESS

Total MapReduce CPU Time Spent: 21 seconds 270 msec

OK

NULL 36964

2015 6542

2016 5278

2011 774

2010 455

2014 404

2017 373

2012 11

2018 1

2013 NULL

Time taken: 151.163 seconds, Fetched: 10 row(s)

Question- 4

Which product is purchased by maximum number of customers?

hive> select name , count(reviews\_numHelpful) as cd from Amazon group by name order by cd desc limit 10;

Query ID = cloudera\_20210502043434\_4535327f-b6af-460f-9618-25ba8133d055

Total jobs = 2

Launching Job 1 out of 2

Number of reduce tasks not specified. Estimated from input data size: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1619950792006\_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1619950792006\_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1619950792006\_0005

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2021-05-02 04:34:27,485 Stage-1 map = 0%, reduce = 0%

2021-05-02 04:34:51,632 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.52 sec

2021-05-02 04:35:18,361 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.68 sec

MapReduce Total cumulative CPU time: 11 seconds 680 msec

Ended Job = job\_1619950792006\_0005

Launching Job 2 out of 2

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1619950792006\_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1619950792006\_0006/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1619950792006\_0006

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2021-05-02 04:35:46,863 Stage-2 map = 0%, reduce = 0%

2021-05-02 04:36:05,660 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.83 sec

2021-05-02 04:36:28,095 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 6.73 sec

MapReduce Total cumulative CPU time: 6 seconds 730 msec

Ended Job = job\_1619950792006\_0006

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 11.68 sec HDFS Read: 48656658 HDFS Write: 4236 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 6.73 sec HDFS Read: 9297 HDFS Write: 704 SUCCESS

Total MapReduce CPU Time Spent: 18 seconds 410 msec

OK

Fire Tablet 7 Display Wi-Fi 8 GB - Includes Special Offers Magenta 10962

6689

B00L9EPT8O B01E6AO69U 5939

Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black 3176

All-New Fire HD 8 Tablet 8 HD Display Wi-Fi 16 GB - Includes Special Offers Magenta 2814

Fire Kids Edition Tablet 7 Display Wi-Fi 16 GB Green Kid-Proof Case 1685

Brand New Amazon Kindle Fire 16gb 7 Ips Display Tablet Wifi 16 Gb Blue 1030

Kindle Voyage E-reader 6 High-Resolution Display (300 ppi) with Adaptive Built-in Light PagePress Sensors Wi-Fi - Includes Special Offers 580

B00TSUGXKE 372

Fire Tablet 7 Display Wi-Fi 8 GB - Includes Special Offers Black 346

Time taken: 146.181 seconds, Fetched: 10 row(s)