Apache Hive

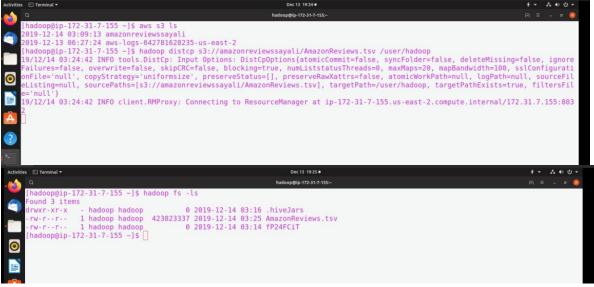
Apache Hive provides a SQL-like interface to data stored in HDP(Hortonworks Data Platform) whose component is Hive.

Steps:

1] I have also created a Amazon EMR Hadoop cluster and I am going to use HIVE on it



2] Uploaded my dataset on Amazon S3 bucket and then into hdfs:



Amazon Review Data Analysis HIVE:

Create a table into Hive:

```
CREATE TABLE IF NOT EXISTS AmazonReviews (
marketplace String,
customer id int,
review id String,
product id String,
product parent String,
product title String,
product category String,
star rating float,
helpful votes float,
total votes int,
Vine String,
verified purchase String,
review headline String,
review body String,
review date String)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY
'\n' tblproperties("skip.header.line.count"="1");
```

```
hive> show tables;

Ok

Time taken: 0.503 seconds
hive> CREATE TABLE IF NOT EXISTS AmazonReviews (

marketplace String,

customer id int,

review id String,

product jastring,

product title String,

product title String,

product category String,

star rating float,

helpful votes float,

votal votes int,

vine String,

review body Body Body Body Body Body Body B
```

Load the dataset into hive table:

```
hive> show tables;

OK

amazonreviews
amazonreviews
amazonreviews1 limit 5;

OK

US 1797882 R3I2DHQBR577SS B001AN000E 2102612 The Naked Bee Vitmin C Moisturizing Sunscreen SPF 30, 5.5 oz (163 ml.) Bea uty 5.0 0.0 0 N Y Five Stars Love this, excellent sun block! 2015-08-31

US 18381298 R1QNESNOFJC2Y4 B00160J22EQ 106393691 Alba Botanica Sunless Tanning Lotion, 4 Ounce Beauty 5.0 0.0 0 N Y Thank you Alba Bontanica! The great thing about this cream is that it doesn't smell weird like all those chemical laden ones. I get a nice healthy un-fake looking tan that isn't orange and it makes my skin soft too. 2015-08-31

US 19242472 R3LIDG204LJBAO B00HUGUGAG 375449471 Elysee Infusion Skin Therapy Elixir, 20z. Beauty 5.0 0.0 0 N Y Five Stars Great Product, I'm 65 years old and this is all it claims to be! 2015-08-31

US 19551372 R3KSZHPAEVPEAL B002HWS7RM 255651889 Diane D722 Color, Perm And Conditioner Processing Caps - 10 0-Pack - Clear Beauty 5.0 0.0 0 N Y G000 DEAL! I use them as shower caps & conditioning caps. I like that they're in bulk. It saves a lot of money. 2015-08-31

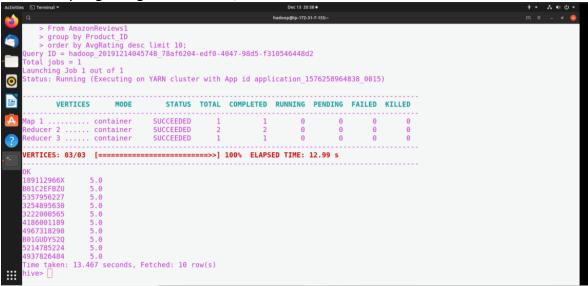
US 14802407 RAI20IG50KZ43 B005M99KWU 116158747 Biore UV Aqua Rich Watery Essence SPF50+/PA++++ (pack of 2) Beauty 5.0 0.0 0 N Y this soaks in quick and provides a nice base for makeup. I've been using this brand for over a year. With daily use, this tube will last you a couple months. 2015-08-31

Time taken: 1.345 seconds, Fetched: 5 row(s) hive>
```

Top 10 Products based on average rating

I have calculated average rating for each product. I have demonstrated similar script on map reduce however hive gave the fastest performance.

- >Select product id, avg(star rating) as AvgRating
- >From AmazonReviews1
- >group by Product ID
- >order by AvgRating DESC limit 10;

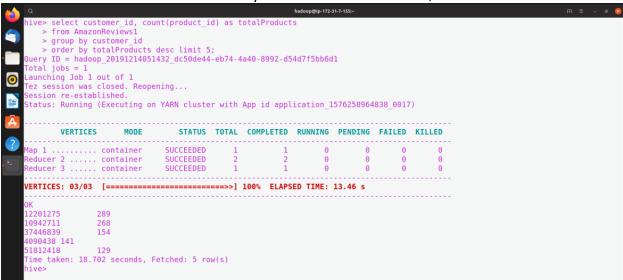


Most valuable Customer

I have counted no of products ordered by each customer and Customer who ordered highest number of products is the most valuable customer.

- >select customer_id, count(product_id) as totalProducts
- > from AmazonReviews1
- > group by customer_id

> order by totalProducts desc limit 5;



Most Popular Product

I have counted how many times each product was ordered.

- >select product id, count(product_id) as mostordered
- > from AmazonReviews1
- > group by product id
- > order by mostordered desc limit 5;

```
select product_id, count(product_id) as mostordered
        > from AmazonReviews1
        > group by product id
  > order by mostordered desc limit 5;
uery ID = hadoop_20191214052439_5deb1f73-850b-4091-acb1-c27eb23041a8
 otal jobs = 1
aunching Job 1 out of 1
  tatus: Running (Executing on YARN cluster with App id application_1576258964838_0018)
                  VERTICES
                                                                                   STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED

        Iap 1 ......
        container
        SUCCEEDED
        1
        1
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
        0
1631
1351
  9049LUT90
  90I3VHKVK
 000JN0ST0
                                        1245
  ime taken: 13.968 seconds, Fetched: 5 row(s)
```

Number of Products Per day

- > Select review date, count(product id) as noOfProducts
- > From AmazonReviews1
- > group by review_date
- > order by noOfProducts desc limit 5;



Number of Products Per Rating

- >select star rating, count(product id) as noOfProducts
- > From AmazonReviews1
- > group by star rating
- > order by noOfProducts desc limit 5;

