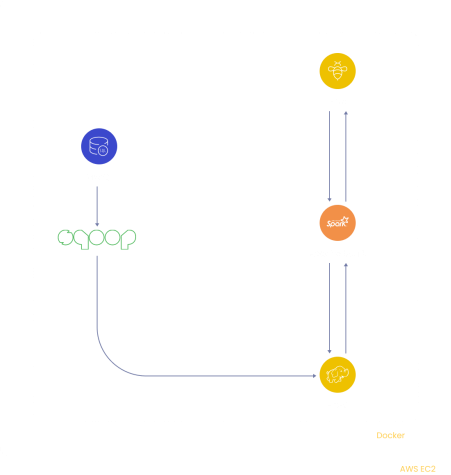
**Ecommerce Sales Data Analysis:**

**Hive Mini Project to Build a Data Warehouse for e-Commerce Sales Analysis:**

In this hive project, you will design a data warehouse for e-commerce application to perform Hive analytics on Sales and Customer Demographics data using big data tools such as Sqoop, Hadoop, and HDFS ,HIVE

**ARCHITECTURE: HADOOP BIGDATA SOLUTION DESIGN**

**Ecommerce Sales Data Analysis: **

Ecommerce public dataset of orders made at Olist Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status,

**Ecommerce Sales Data Analysis:**

price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates.

This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

**Join it With the Marketing Funnel by Olist**

We have also released a Marketing Funnel Dataset. You may join both datasets and see an order from Marketing perspective now!

**Instructions on joining are available on this Kernel.**

**Context**

This dataset was generously provided by Olist, the largest department store in Brazilian marketplaces. Olist connects small businesses from all over Brazil to channels without hassle and with a single contract. Those merchants are able to sell their products through the Olist Store and ship them directly to the customers using Olist logistics partners. See more on our website: www.olist.com

After a customer purchases the product from Olist Store a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he can give a note for the purchase experience and write down some comments.

**Ecommerce Sales Data Analysis:**

**Attention**

1.An order might have multiple items.

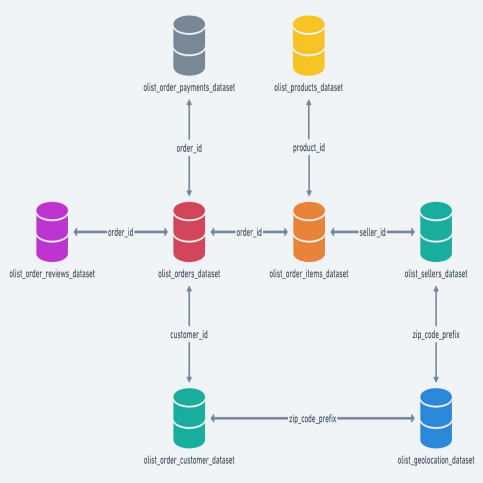
2.Each item might be fulfilled by a distinct seller.

3.All text identifying stores and partners where replaced by the names of Game of Thrones great houses.

**Data Schema**

**The data is divided in multiple datasets for better understanding and organization. Please refer to the following data schema when working with it:**

**Ecommerce Sales Data Analysis:**

****

**About the Data:**

Data consists of Ecommerce data from 04-09-2016 to 03-09-2018, which is about 2 years of data. The dataset we have used is a combination of 9 sub-datasets which originally is 120.3 MB sized dataset. But we have pre-processed and removed many unwanted feature columns and used the modified dataset for our project analysis. Dataset rows : 1,16,573

**Ecommerce Sales Data Analysis:**

Dataset columns : 21

Dataset size : 27.4 MB

**Data Description**

**S.No Name Description** 1 order\_id unique id for each order (32 fixed-size number)

2 customer\_id unique id for each customer (32 fixed-size number)

3 quantity 1-21

4 price\_MRP cost price, 0.85-6735

5 payment selling price, 0-13664.8

6 timestamp order purchase time (local, day-month-year hour:min:sec AM/PM)

7 rating 1-5

8 product\_category category under which product belongs

9 product\_id unique id for each product (32 fixed-size number)

10 payment\_typeType of payment - credit card/debit card/boleto/voucher

11 order\_status delivered/shipped/invoiced

**Ecommerce Sales Data Analysis:**

**S.No Name Description** 12 product\_weight\_g weight of product (in grams), 0-40425 13 product\_length\_cm length of product (in centimeter), 7-105 14 product\_height\_cm height of product (in centimeter), 2-105 15 product\_width\_cm width of product (in centimeter), 6-118 16 customer\_city city where order is placed

17 customer\_state state where order is placed

18 seller\_id unique id for each seller (32 fixed-size number) 19 seller\_city city where order is picked up 20 seller\_state state where order is picked up

21 payment\_installments no. of installments taken by customer to pay bill, 0-24

**Analysis using Hadoop Hive Batch process Analysis.**

∙ **ETL pipeline using Sqoop.**

∙ **Internal table**

∙ **Partition table**

∙ **Indexes**

∙ **External table to the client db**

∙ **DB with Tableau or Power bi**

**Ecommerce Sales Data Analysis:**

**HIVE JOBS:**

1. **Customer Segmentation**

Categorizing customers based on their spendings

2. **Monthly Trend Forecasting**

the monthly trend of sales

3. **Hourly Sales Analysis**

Which hour has more no. of sales?

4. **Product Based Analysis**

Which category product has more rating?

Which product has sold more?

Top 10 highest & least product rating?

Order Count for each rating

5. **Payment Preference**

What are the most commonly used payment types?

Count of Orders With each No. of Payment Installments

6. **Potential Customer's Location**

Where do most customers come from?

7. **Seller Rating**

Which seller sold more?

Which seller got more rating?

8. **Logistics based Optimization Insights**

Which city buys heavy weight products and low weight products? How much products sold within seller state?

**Ecommerce Sales Data Analysis: Visualization with Tableau**