A1Mart Online Retail analysis

Introduction:

A1Mart is a one-stop supermarket chain that aims to offer customers a wide range of basic home and personal products under one roof. A1Mart, is an Indian retail corporation that operates a chain of hypermarkets in India. It has multiple stores in various cities across states

Project Introduction:

Data is stored in MySQL. With market demands and the success of A1mart they decided to use big data to increase the business and show the current progress. We are considering this as online store and small scenario for project with limited dataset due to timing constraint.

Input Files: Orders and customers.

Note:

1. Project can be done using Scala or python.
2. All requirements will be done in two versions
   a. Using spark SQL
   b. Dataframe /Dataset

Requirements:

Step 1: Import data from MySQL to HDFS. (Tables: Orders and customers)

Step 2: Analysis can be done using Hive or Spark. **Here we will be using spark.** Following analysis needs to be done.

a. Retrieve all records for particular customer. (Better to accept customer name from user)
b. List count of orders based on status and month.
c. List count of orders based on status and month for particular customer.
d. List count of orders based on customer and status.
e. Find the customers who have placed orders.
f. Find the customers who have not placed orders yet.
g. Find top 5 customer
    a. Highest number of orders
    b. Highest sum of total orders
h. Find the customer who did not order in last 1 month or for long time
i. Find the last order date for all customers.
j. Find open and close number of orders for a customer.
k. Find number of customers in every state.

l. Create jar and run using spark-submit. Also seek customer id or name as input wherever required.

Step 3: Add 3 requirements from your side. If required you can add columns to dataset with dummy values.

Enhancements:

1. You can use Hive integration with spark for some requirements.
2. As you can see customers data, email and mobile number is masked in MySQL. Implement masking and unmasking in spark.

Submission:

1. Input files
2. Submit code on Github using Jira
3. Documentation: Include the code reference as filename for very analysis and screenshot of the output

Final Presentation will be done on the day as communicated.

!!!!!!!!!!!!!!!! All the Best!!!!!!!!!!!!!!!!!!!!!

TRENDY TECH
UPLIFT YOUR CAREER!