Accelerating Analytics with Data bricks and AWS S3
Solution Design Document

Created by: Anisha Kale
Date: 25-04-2023

Project Overview:

Our project aims to build a data warehouse in Databricks by importing data from AWS S3 and extracting insights. This will enable you to make data-driven business decisions and gain important insight into sales data. The Purpose of the Document is to outlines the proposed solution for the project based on the user stories identified.

User stories

S.no	User Stories
6.	Finding products with minimum sales.
7.	Finding Stores which has generated maximum and minimum revenue
8.	Generating BI dashboards

Solution:

The solution of above user stories will be achieved using following tools.

1. Data bricks: Query out the required data set using data frames.
2. Power BI: To create visualization of output data and help in business insights.

Design:

1. Data Bricks: Following things are done in data bricks
   a. There are two demands from the client. First one is client want to balance its inventory and remove the products which are not in demand. Therefore as a data analyst we need to query out products which have minimum sales.
   b. Another thing is client also want to open new store in the city where revenue is good and close the store in the city with least revenue so as a data analyst we need to query out stores which has generated maximum and minimum revenue.
   c. Best coding practices: While querying best coding practices should be kept in mind. For reference
   Repository
   https://github.com/Aniishak/AAwDS3

2. Power BI: Data visualization and creating dashboard will be on Power BI following are the things that need to be performed

   a. The output of queries needs to be input in power bi.
   b. Using the output, dashboard is created and organized
   c. Dashboards can be created using multiple designs.

Work Flow:

The workflow for above user stories will look like,

1. Query out product with minimum sales.
2. Query out stores which has minimum and maximum revenue
3. Take the output of both queries and put it in power bi to generate visual data.
4. Organize the visuals and create dashboard.