Accelerating Analytics with Data bricks and AWS S3
Solution Design Document

Project Overview:

Our project aims to build a data warehouse in Data bricks 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
3.	Uploading data into S3 buckets.
4.	Finding products with maximum sales.

Solution:

The solution of above two will be achieved by following AWS services and workforce.

1. AWS S3: For uploading the tables created after normalization and data cleaning.
2. Data bricks: For analyzing the data and querying out products with maximum sales.

Design:

1. AWS S3: Configuration in s3 includes following things
   a. Importing the data into S3. The data should be readily accessible.
   b. The versioning option must enabled to avoid overwriting existing data

2. Data Bricks: Once data ingestion is done following things will be done in data bricks,
   a. Using read commands input the entire tables into Data bricks.
   b. Based on clients demand data analyst need to query out products with maximum sales. The query should find in demand product for each month in a year. For code reference.
   Repository
   https://github.com/Aniishak/AAwDS3

3. Testing: For testing things need to be performed are as following
   a. Unit testing: Testing the individual functions or queries that we have written to perform analysis. Also check if the functions are working as expected and handle edge cases like null values or invalid inputs
   b. Integration testing: Here we need to check if the different components like S3 bucket and data bricks are working together correctly and handle any errors that may arise due to interactions between them.
   c. Acceptance Testing: - Test if all the acceptance criteria are met , and ensure that it meets their business requirements and helps client make better decisions

Work Flow:

The workflow for above user stories will look like,

1. Once the data is ready, it needs to be uploaded in S3.
2. Data analyst need to query out products with maximum sales.
3. Testing needs to be performed which include unit testing and integration testing.
4. Once done with above steps testing team need to give acceptance to the solution and further scrum master will get the business acceptance.