Team 7

[Price Comparisons and Product Analysis of groceries](#)
[Title](#)
[Objective](#)
[Block Diagram](#)
[Workflow](#)
[Tools](#)
[Milestones](#)

Online Retail and Product Review Comparison

Team Member
1. Abel Philip
2. Aishwarya Parab
3. Vidhi Patel

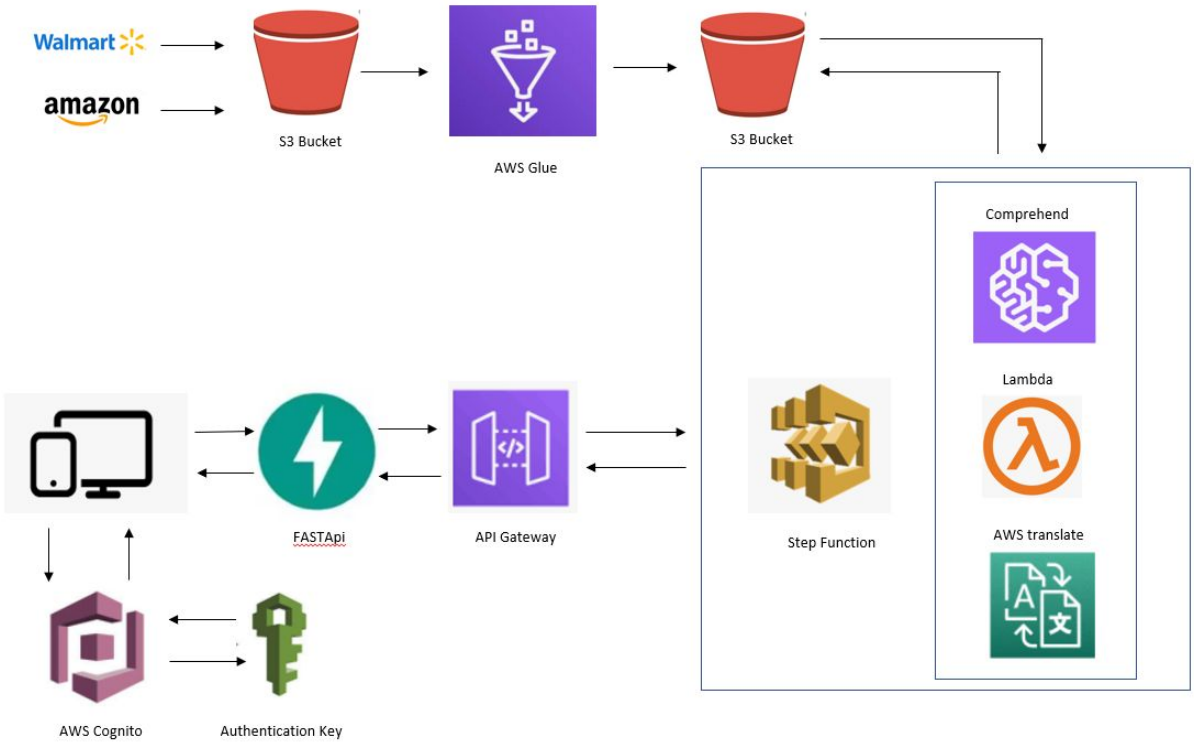
Super helpful
amazon reviews
that I totally made up

Walmart
Help Center
Share your opinions. Write a review.

Objective

Currently multiple platforms are available to shop online. Over time innumerous reviews have been left by users over the pool of products. These reviews play a significant role for all other users in determining whether to buy the product or not. However, it is not possible for one to go through all the reviews and determine the product's quality and compare it with the other products available on the same or different online platforms. A platform that summarizes the reviews, analyses the sentiment behind it and also compares the product prices from different sites is extremely paramount. When we talk about retail sites we consider the two major ones.

Architecture



Workflow

1. This Project has 6 stages to run the workflow.

Stage 1- Scraping and Input

a) The links of the products to be scraped are first uploaded to S3.
b) Once the product is hit, the link is fetched from S3 and all the products and its reviews are scraped and stored into another S3 bucket.
c) Cloudwatch is used to monitor PUT operations on the S3 bucket

Stage 2 - Data Processing

a) When data is loaded into S3 the Amazon Glue will be triggered.
b) ETL processing will be carried out by using AWS Glue Service and the processed data is stored in S3 bucket

Stage 3 - Streamlit

a) The user will need to be authenticated and only then can the APIs
b) The user will select the product for which they need comparison or reviews

Stage 4 - Sentiment Analysis Pipeline

a) After the user hits an API for a selected product, Comprehend will summarize the reviews stored in S3 bucket
b) Depending on the user requirement and the language chosen the step function will process the output.

Timeline

Authentication, Web Scraping & Storing data – 12/9/20

ETL & APIs – 12/12/20

Implementing Step functions, Text Summarization & - 12/15/20

Final Changes – 12/17/20