



Analysis of Pricing pattern based on ratings from Ecommerce website to strategize product portfolio for user.



Group No: 25

Group Members:

- Amol Chilkulwar
- Diksha Wanjari
- Komal Satpute
- Piyush Bhomale (TL)
- Snehal Bhajikhaye
- Vaishnavi Chandak

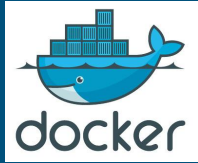
Mentors:

- Nandkishor Rajput (Neiron)
- Amar Jadhav (Morning Star)

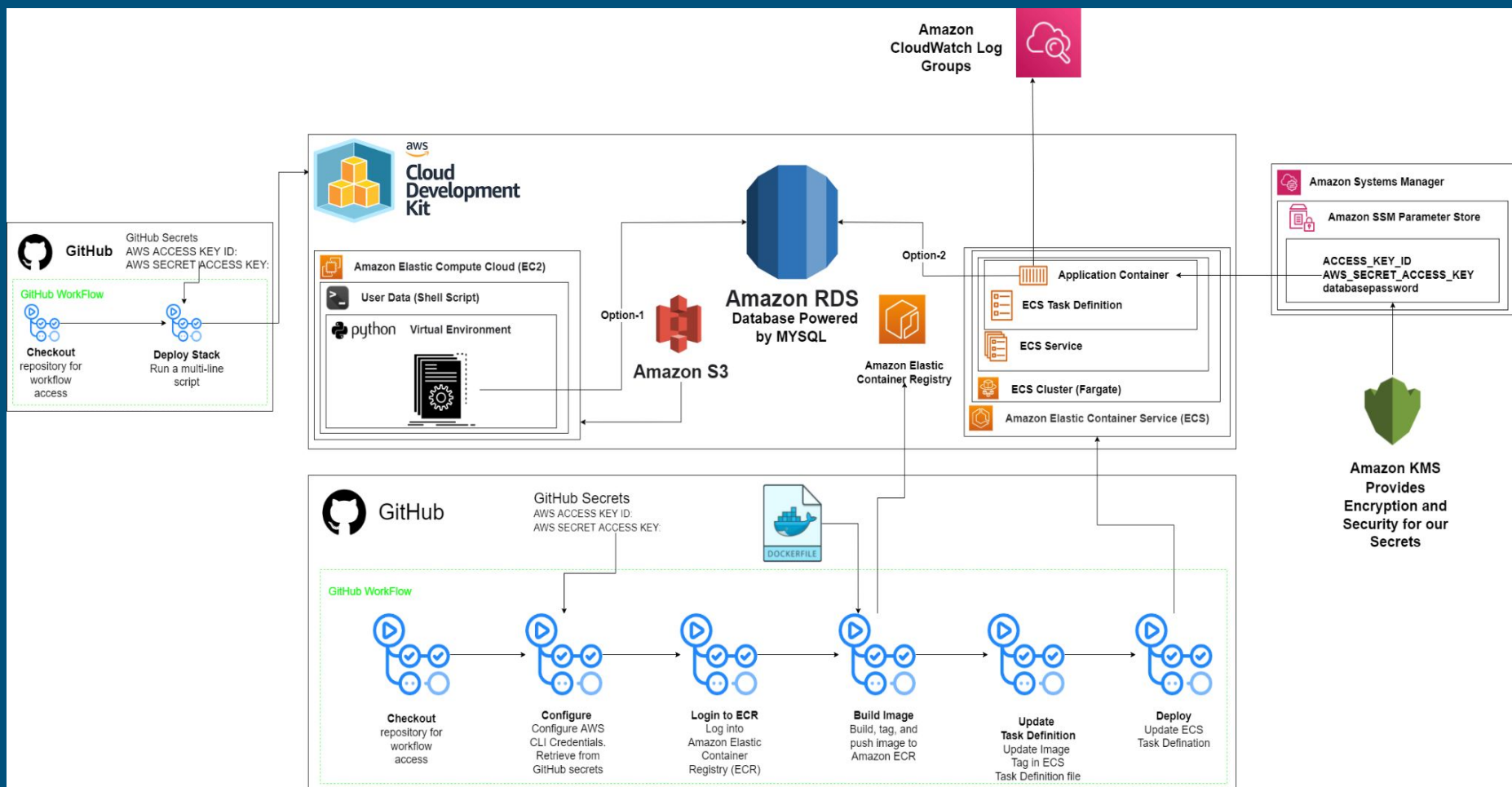
Problem Statement

- We are working on a project for a client where you want to setup product lineups for their e-commerce website. They want to understand the market value and outlook of products so that the portfolio is more customer-oriented and satisfying to customers.
- Our team decided that the first step in analyzing the market is to collect data and for this purpose, we scraped data from various e-commerce websites so that we could collect all the details of products and store them on Amazon Web Services in a structured format where client can easily query the data.
- This entire process is automated in such way that the client will be able to get the desired output within a single click.

Technologies Used:



Architecture



SENTIMENT ANALYSIS

Scrap the product review from E-commerce website



HUGGING FACE TRANSFORMER



Collect sentiments whether positive or negative sentiments along with confidence

Web Scrapping in Python

- Fetching the data (Using requests library : import requests)
- Used BeautifulSoup to parsing Data
 - Provides simple methods to
 - search
 - navigate
 - select
 - Deals with broken web-pages really well
- Collected Various Product Details such as
 - Product Name
 - Rating
 - Total Rating
 - Discounted Price
 - Original Price
 - Product Url
 - Product Sentiments

Why we need Docker?

An application works in developer's laptop but not in testing or production. This is due to difference in computing environment between Dev, Test and Prod.



Dev

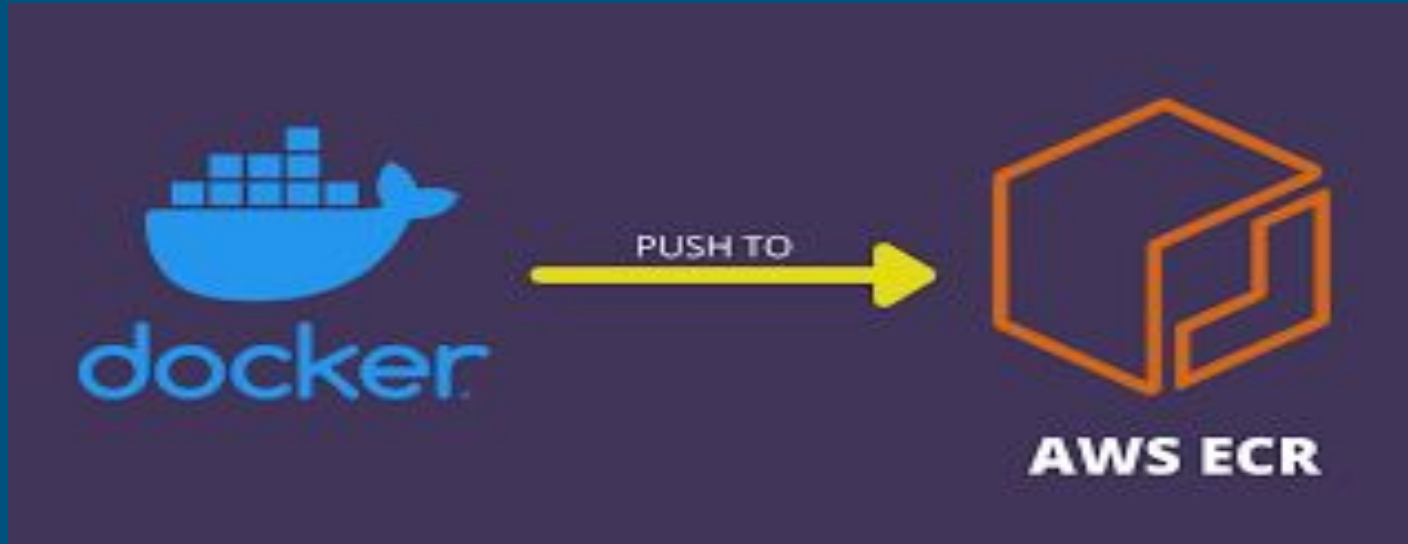


Prod

In Dev there can be a software that is upgraded and in Prod the old version of software might be present

DOCKER

- Docker is an open source containerization platform. It enables developers to package applications into containers
- Dockerfile - Build image - Tag -Push Image-ECR



AWS EC2

- Elastic compute cloud provides resizable,secure,compute capacity in the cloud via a virtual machine.
- Represent a physical server for you to deploy your applications.
- Creation of EC2 instance, set up required configurations for EC2 instance.
- Passing USER DATA for performing common automated tasks by system.
- For automating EC2 required bootstrapping(shell script).
- Shells script contains all the required tools and configurations for virtual environment setup.
- Data Execution after fetching data(object) from S3 Bucket.

AWS S3

- Amazon Simple Storage Service is scalable, high-speed, web based cloud storage service.
- This is designed for online-backup and archiving of data and application on AWS.
- S3 provides object storage for storing and recovering of data from anywhere over the web.
- Tasks: Bucket Creation and Uploading Dependencies.
- Data loading from scraper file(zip) and extracted it into S3 Bucket.

AWS RDS

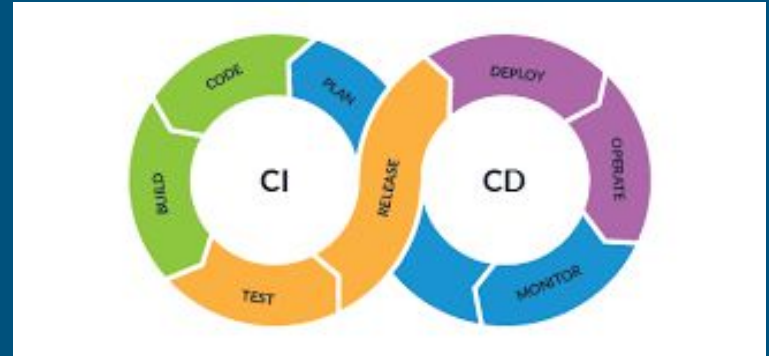
- Boto3 client connected to AWS RDS
- Checked “Endpoint” of database (RDS)
- Connection to database (mydb) using mysql.connector
- Schema created
- Features:
 - Multi-engine support(Aurora, MySQL, MariaDB, PostgreSQL, Oracle, SQL Server)
 - Easy to manage
 - Security
 - High Availability
 - Designed for use with other Amazon Web Services

CI/CD Pipeline:

- CI and CD stand for continuous integration and continuous delivery/continuous deployment
- Continuous Integration is the process that allows developers to integrate new code into a shared repository (such as GitHub)
- Continuous Delivery (CD) occurs at the end of the CI cycle and is responsible for the automated delivery of the integrated code from the development to the production stage

The benefits of GitHub Actions:

- Automate everything within the GitHub flow
- Hosted virtual machines on multiple operating systems
- Pre-written CI templates that are ready to use
- Simple container and operating system testing
- Use it on your public repository for free



Environment Management



CDK Advantages

- Can maintain Production and Development environment
- Cross Account Deployment
- Infrastructure as a Code
- Dependencies between Application is possible
- Uses Object Oriented Programming Concepts
- Multiple Programming Language support

Running Status

100%

Extras

- PyTest
- Python Docs
- AWS Batch
- Kinesis



Special Thanks !

Thanks to Pradeep Tripathi Sir for continuously monitoring, suggesting changes, introducing to new Technologies as well as for challenging every day to push our limits.

Thanks to our Mentors Amar Jadhav and Nandkishor Rajput to keep us charged as well as helping us whenever needed.

Thanks to Nikhil Ramgiri (AJACKUS) for providing resources to understand tools we used.

THANK YOU !!