Architecting Enterprise Cloud Solutions Assignment

* **DEVELOPER IQ**

Name : Thirunayan Dinesh Jeeva

Student No: MS20908188

GitHub: <https://github.com/Thirunayan22/DeveloperIQ>

Contents

1. Introduction
   1. What is DeveloperIQ
   2. Definition of developer productivity
2. Calculating developer productivity
3. Micro service architecture
   1. Architecture diagram of microservices
   2. Usage of CQRS Pattern
   3. Architecture diagrams of microservices and functionality of each service.
      1. Productivity Calculation Services
      2. Database upload service
      3. Database update check service
4. Database
   1. Reason for choosing DynamoDB (A no-sql database)
   2. DB Structure
5. AWS deployment structure diagram. Along with control EC2 instance
6. Containerization
   1. Reason to containerize.
   2. Cloud container registry and why choose AWS ECR
7. Cluster Orchestration
   1. Kubernetes deployments
   2. Kubernetes services
   3. EKS Cluster Configuration
   4. Node group configuration
   5. EKS Cluster Architecture
8. Using AWS CloudFormation
   1. Purpose of using cloud formation
   2. Cloud formation script
9. Fault tolerance scenarios
10. Conclusion

Introduction

Developer IQ is a developer productivity calculation application which can quantitatively calculate the productivity of a developer on a particular project by tracking the developer’s activity on GitHub such as commits, pull requests, issues created and issue comment interactions. These activity metrices are then used to calculate an aggregated DeveloperIQ productivity score for the developer, for a “weekly”, “monthly” and “yearly” timeframe.

Calculating Developer Productivity

Calculating an aggregated developer productivity score quantitatively is a complex task since, productivity itself is an abstract concept with multiple definitions, in the case of DeveloperIQ we consider developer productivity as the total contribution of the developer to the project repository. Thus, our mathematical formulation for calculating productivity also considers the direct and indirect contributions of a developer like there activeness for responding to issues, issues created along with direct contribution factors like commit frequency ,commit addition and deletion.

The DeveloperIQ productivity metric

Developer Productivity :=

Commit addition : source code addition volume in commits

Commit deletions : source code deletion volume in commits

Number of commits : total number of commits the developer has made for a particular time frame.

Issues created : Issues created by the developer this includes the pull requests and common bug issues as well

Issues comments: Number of times the developer has interacted with open issues, this includes code reviews and answers to open bug issues.

The total commit addition and commit deletion is added to get an aggregate score of the total code contribution which is then divided by the number of commits. The developers are penalized for a high number of commits with low contributions, because the standard best practice is to push meaningful commits with sufficient contributions and functionalities instead smaller commits without any significant additions or deletions. The calculated score is called the **commit score.** The commit score is then added to the sum of the number of issues created by the developer and issue interactions by the developer, this is called the **interaction score**. The **commit score** and **interaction score** are added and are passed into the *log* function to normalize the outputs within a much shorter range.

The output will be the productivity score of the developer, and higher DeveloperIQ scores denote higher developer productivity.

Microservice Architecture

A completely cloud based microservices architecture is used for the implementation of this project.

We use 3 core services:

1. Productivity Calculation Service
2. DB Update Check Service
3. DB Update Service

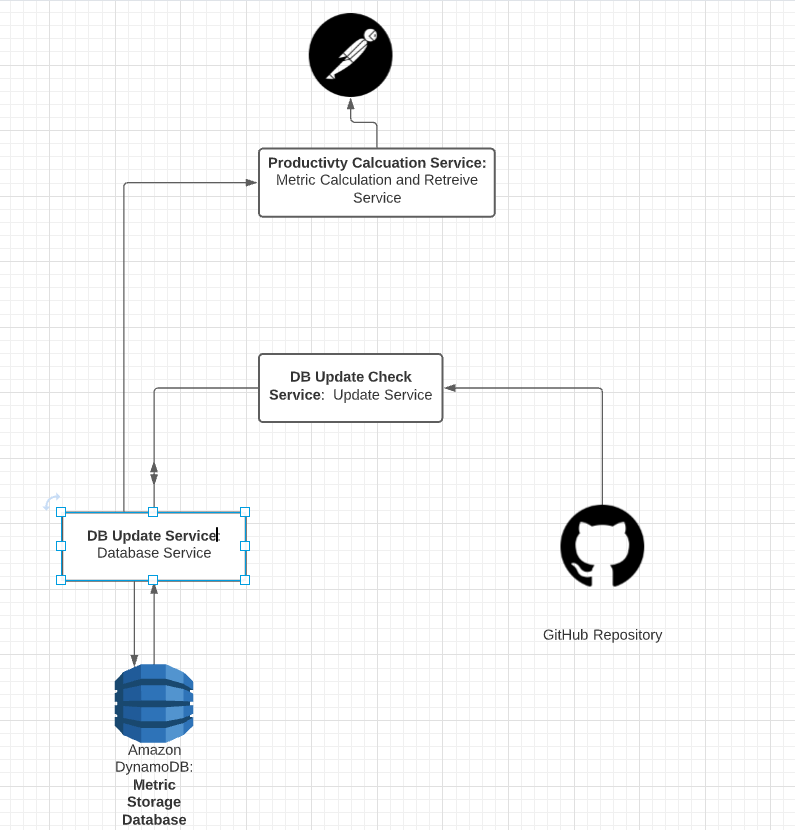
A diagram depicting the inter-service communication is given below:

Figure - Service Communciation Model

**Productivity Calculation Service** : This service fetches data from the AWS DynamoDB cache database consisting of the developer productivity scores for each contributor of a repository and then calculates the DeveloperIQ productivity score for each contributor using the equation above. This is the service that the frontend will directly interact with to get the developer productivity metrics.

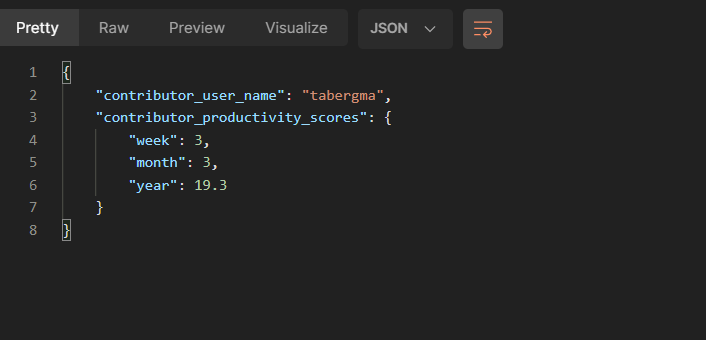


Figure - Productivity calculation service output

The service provides the contributor metrics in 3 time frames,”week”, “month” and “year”.

**DB Update Check Service :** This service calls the GitHub API and gets the real-time productivity metrics for a particular contributor by making multiple requests to the different endpoints which contain the metric values.

**DB Update Service :** This service is responsible for updating the caching database. The DB update service calls the DB update check service in a regular time frame to look for changes in contributor activity in the repository and update the changes to the cache db. The service takes a live snapshot of the cache database and another live snapshot from GitHub to get the contributor’s real-time metrics, and then compares both the snapshots and if it detects any difference it updates cache database with the latest snapshot from GitHub.

This routine check is run every 1 hour within and additional “delay” time which can be controlled as a parameter. This type of a batch-based cache update mechanism is used because GitHub imposes a limit of 5000 API requests per hour for an authenticated user, if this limit is exceeded the requests are bounced back and rejected.

**Cloud Architecture Pattern**

We used CQRS (Command Query Segregation Pattern) as the base pattern to extract and update the database. The CQRS pattern proposes the idea of separating read and update functions into two different domains or models.

The core reason for the choice of this cloud design pattern is because this application contains a lot of frequent writes and the volume of reads can vary depending on the number of requests. The CQRS pattern enables us to independently scale the read and write services depending on the traffic, in addition it also provides a better separation of concerns thus enabling simpler queries to the database.

Database

We utilize a database for caching so that for every request the api does not have to directly call the GitHub API. The database was implemented using AWS DynamoDB which is a no-sql database. A no-sql database was chosen because it enabled high speed read-write operations along with un-structured schemas which allowed flexibility in the structure of the contributor data storage format.

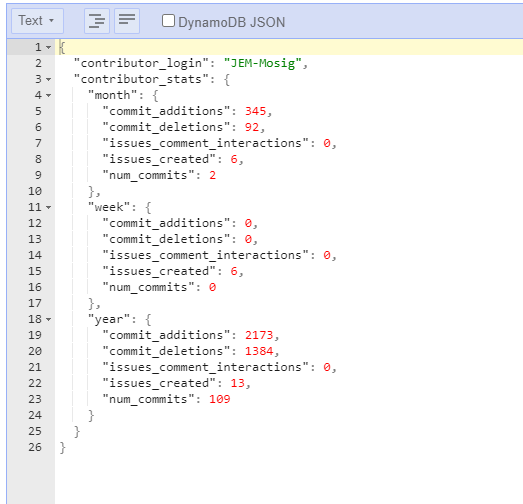


Figure 3 - A sample user's metrics stored in the dynamo db database

The database update is executed by the DB Update Service which routinely updates the database if there are any changes in contributor activity.

AWS

In this project we use amazon web services (AWS) for completely to develop the backend infrastructure. The services in AWS utilized for this in this project are:

1. AWS EC2 - For cluster deployment and administration
2. AWS ECR - Used as a registry for the storing and versioning of containers
3. AWS EKS - Used to deploy and manage the Kubernetes cluster and master control plane.
4. AWS DynamoDB – For cache storage and retrieval
5. AWS IAM – For access management for node-groups within the cluster
6. AWS VPC – For secure private cloud networking within the cluster
7. AWS CloudFormation – For the automation of configuration and deployment of the entire cluster infrastructure

Containerization

All the 3 services use the Dockerfile configuration since the packages and environment needed to run the 3 services are identical. After a code change or environment change is made to the codebase, the container would be built again and then tagged with an incremental version number and then pushed to the repository. The complete container workflow is depicted below.

After a container is pushed to ECR registry, a Kubernetes deployment is rolled out and the existing services and deployments are restarted to work with the updated code base and containers, this is done without a down time since all the Kubernetes services use a load balancer to expose the deployments. The complete container deployment pipeline is presented in the diagram below.

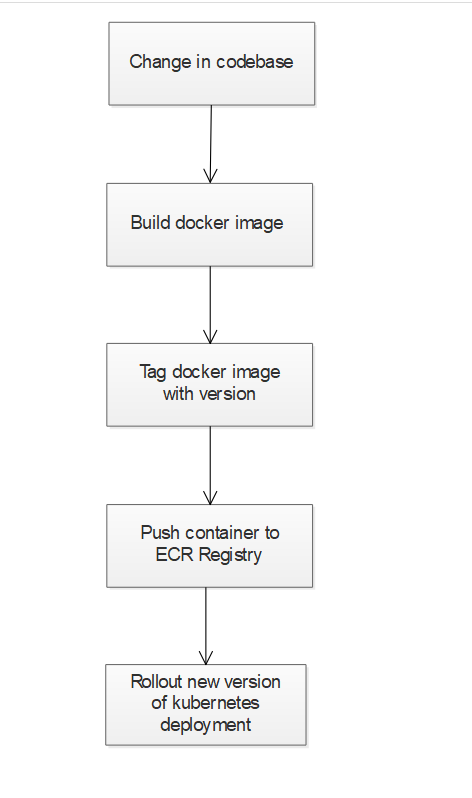


Figure 4 - Container deployment pipeline

EKS Cluster and Kubernetes Orchestration

Kubernetes was used for the orchestration of the deployed containers in this project and AWS EKS was used to manage the Kubernetes master control plane.

There were 2 deployments created for the 3 services. The productivity calculation service was deployed in a separate deployment whereas the GitHub repository monitoring service and database update service were contained in a single deployment because both the services need to communicate frequently and aggregately serve the same purpose of caching.

In the EKS cluster, a node-group consisting of two t3 medium instances were used, t3 medium was chosen as the node compute category since running Kubernetes requires a moderate level of memory. All the nodes were made SSH accessible so that each of the nodes can be accessed by the user and each deployment were configured to have 2 replicas running on each of the nodes, so that even in the case that a single node fails the services can continue to run.

The number of nodes in the cluster were configured to be 2 because of the cost factors associated with creating nodes.

The complete cluster infrastructure is wrapped with a AWS VPC (Virtual Private Cloud) network to enable inter-communication within the services as communication with external services.

The deployment.yaml file for the productivity calculation service is given below

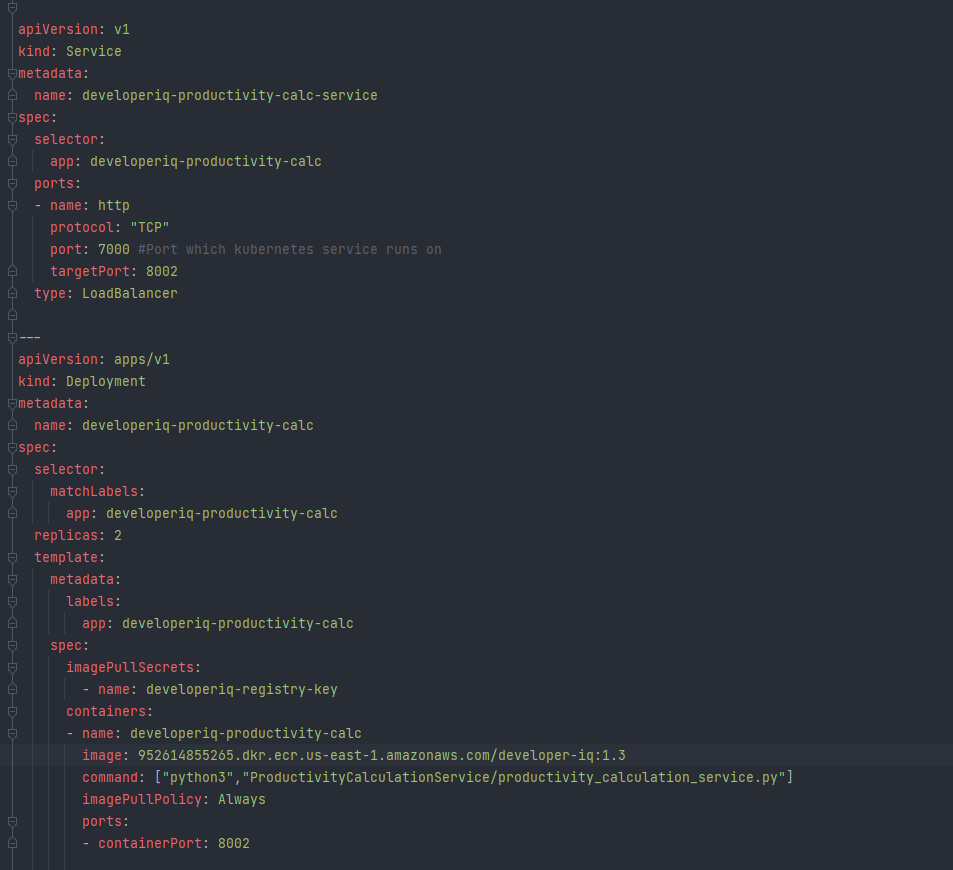


Figure - Deployment file of productivity calculation service

The EKS cluster deployment diagram with VPC configuration is visualized in the cluster deployment architecture diagram below.

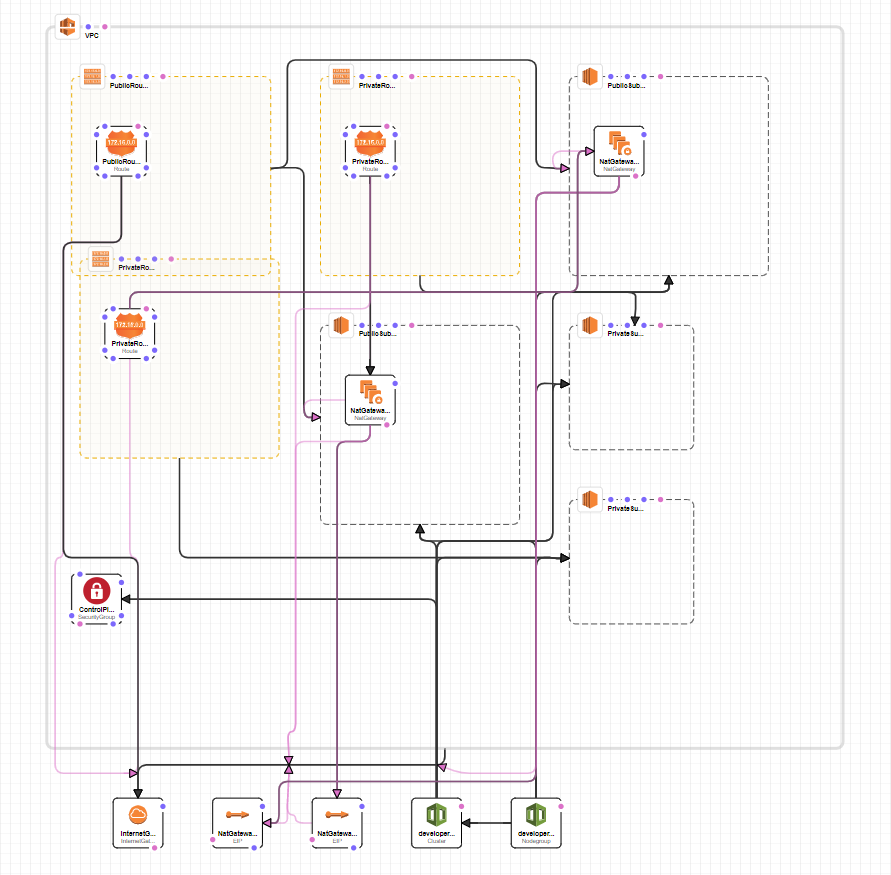
**AWS EKS Cluster Configuration With VPC Configured**

Figure – AWS EKS Cluster with VPC Configuration

Cloud formation

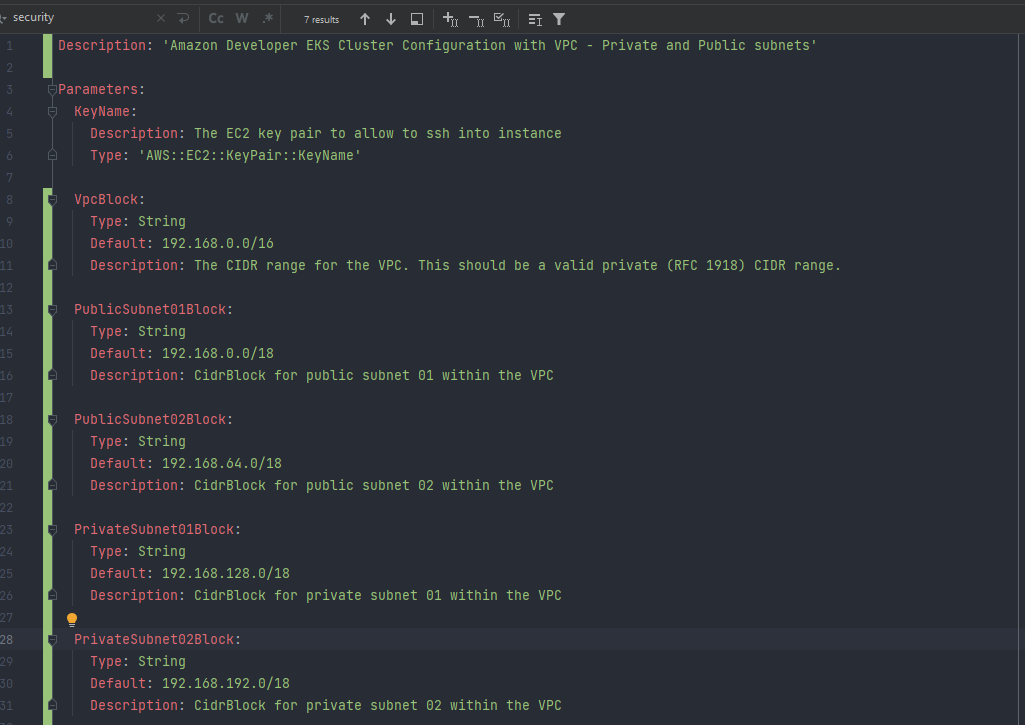
The deployment of a complete EKS cluster with node-group and VPC configurations can take a lot of time, to avoid this delay and to make infrastructure replication and versioning more seamless, AWS cloud formation is used to script the complete infrastructure configuration and then create all the components. We configure both the VPC and EKS Cluster infrastructure in a single CloudFormation yaml file.

Figure 7 - CloudFormation Script Yaml File – Parameters Section

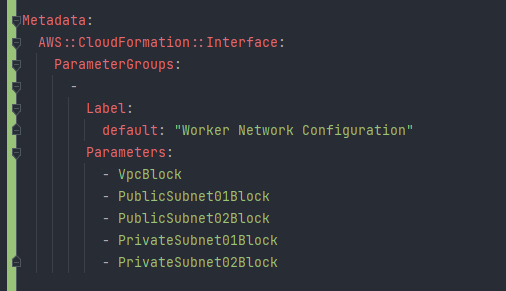


Figure 8 - CloudFormation Script -Metadata Section

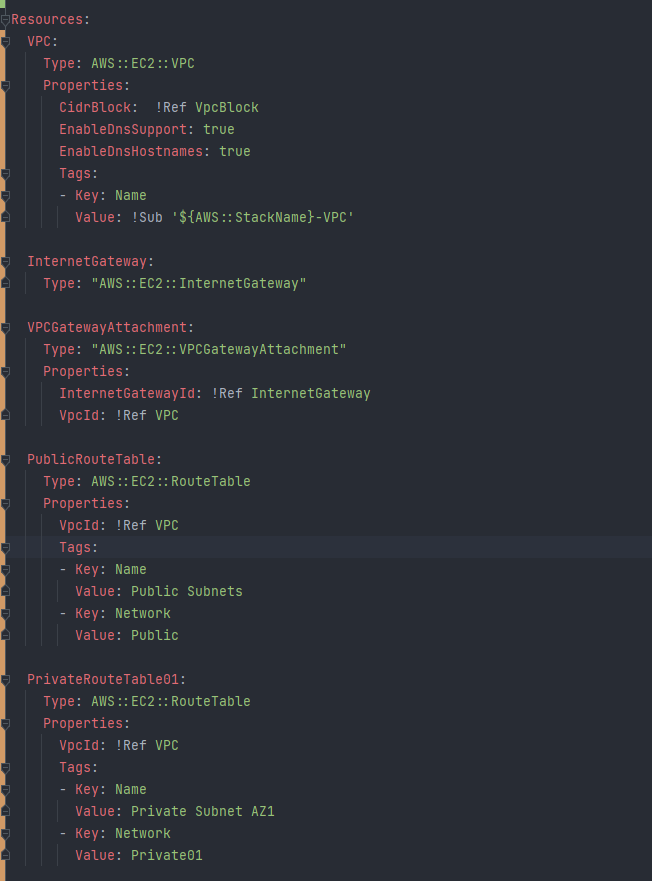


Figure 9 - CloudFormationScript - Resource Section 1

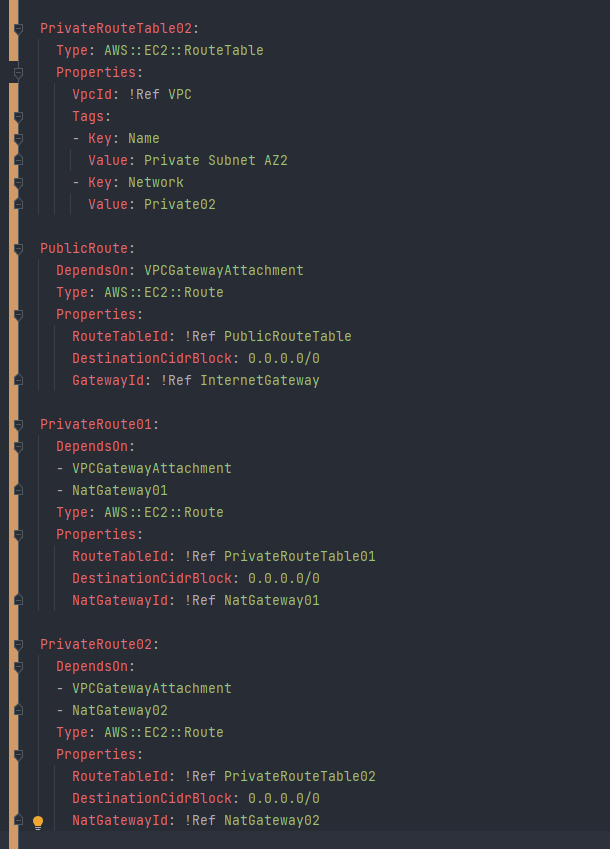


Figure 10 - CloudFormation – Resources Section - 2

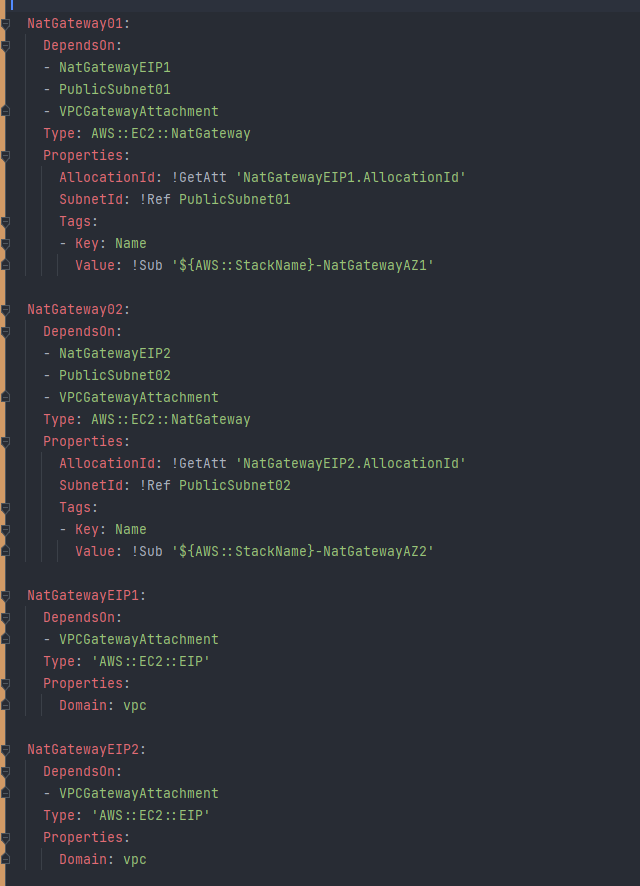


Figure 11 - CloudFormation - Resource Section - 3

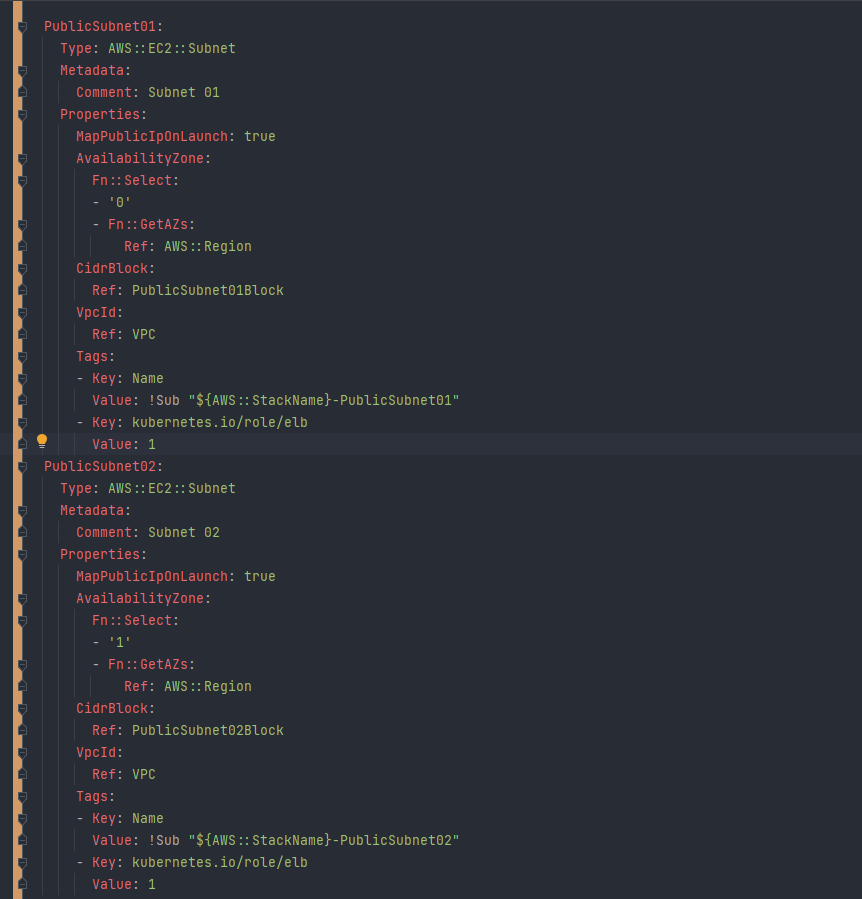


Figure 12 – CloudFormation – Resource Section 4

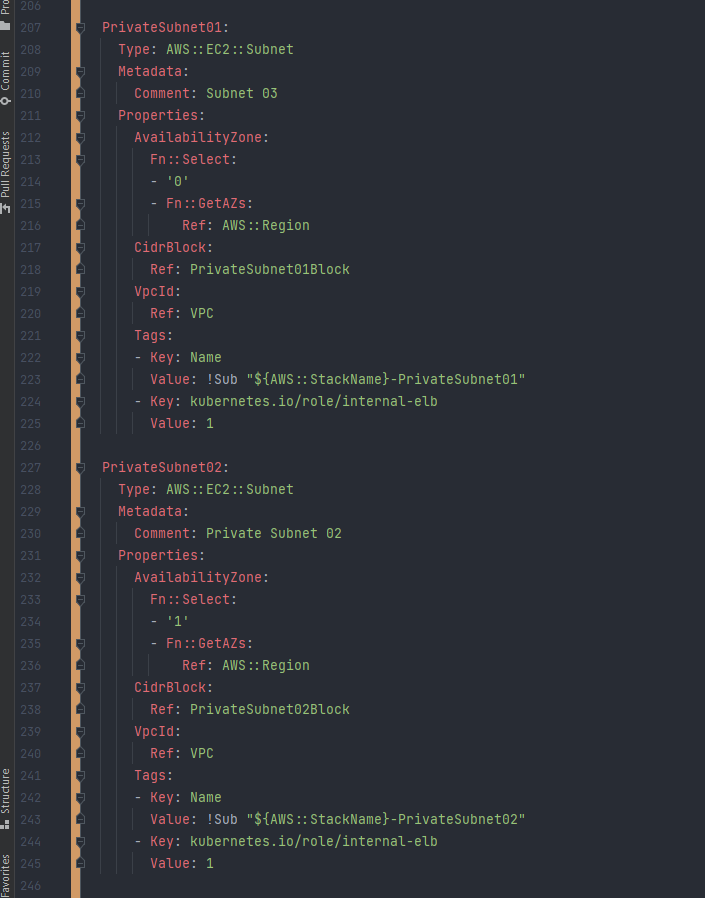


Figure 13 - CloudFormation - Resource Section - 5

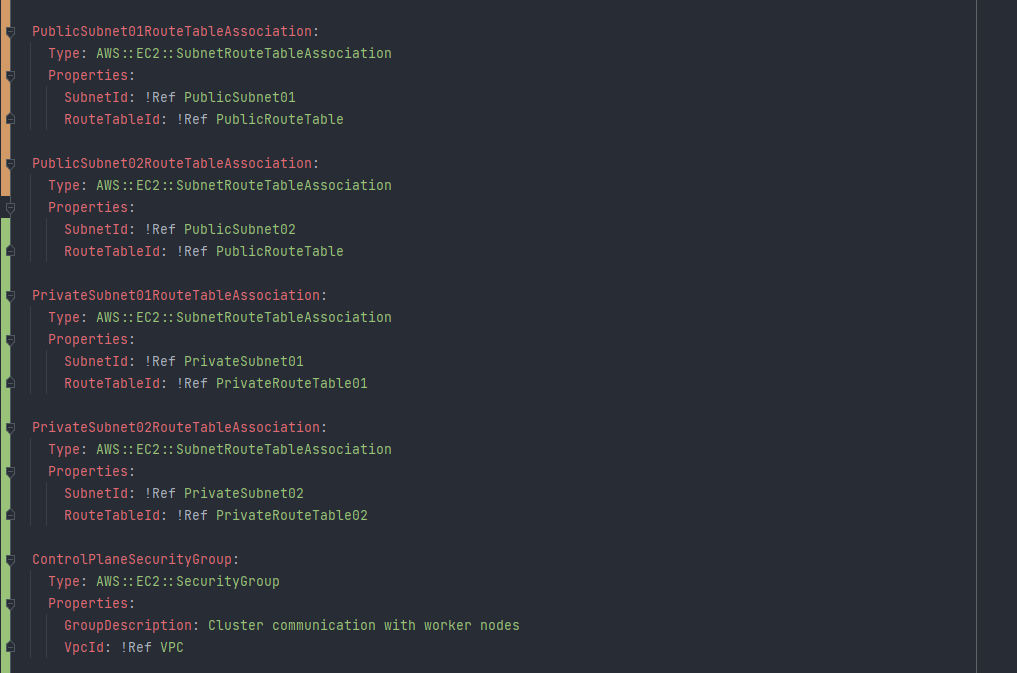


Figure 14 - CloudFormation - Resource Section – 6

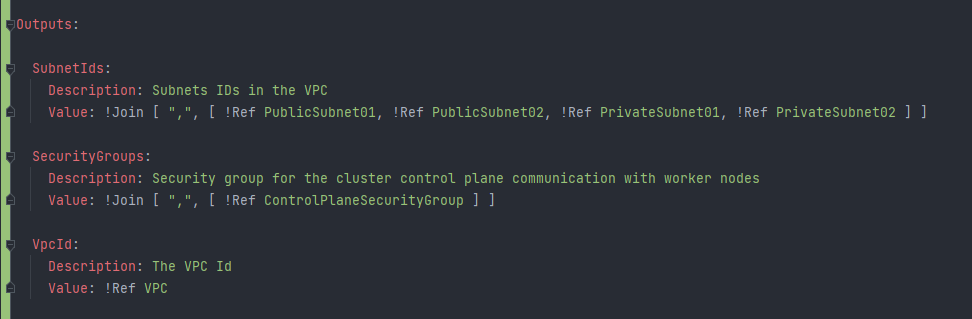


Figure 15 - CloudFormation Script - Outputs Section

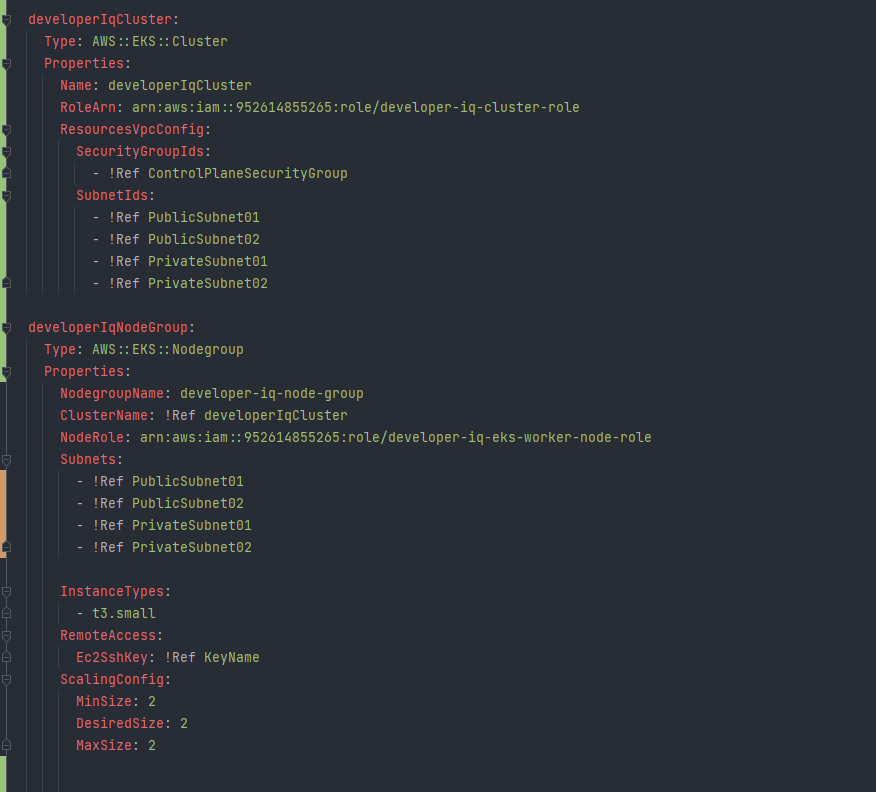


Figure 16 - CloudFormation Script - Cluster Section

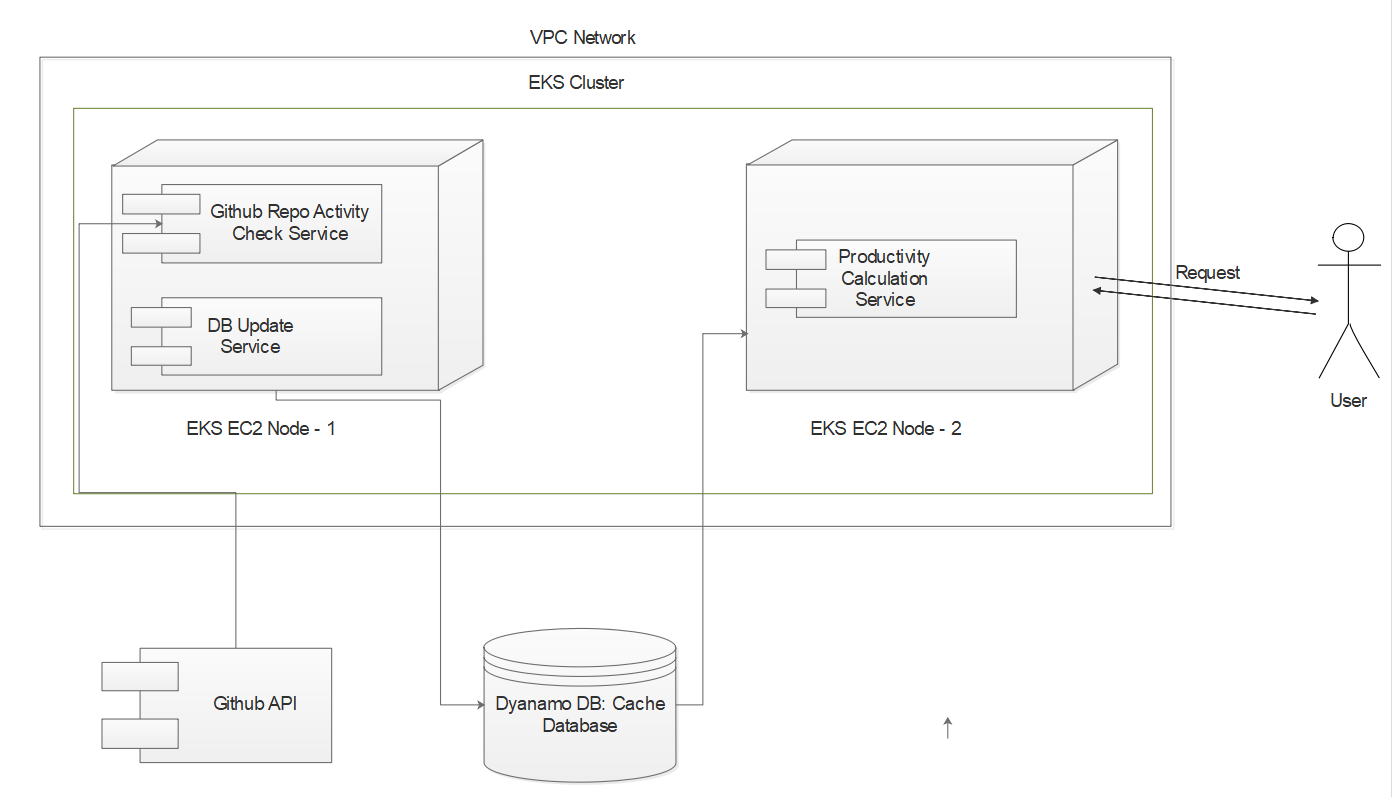
Deployment Diagram

Figure 17 - Deployment Diagram

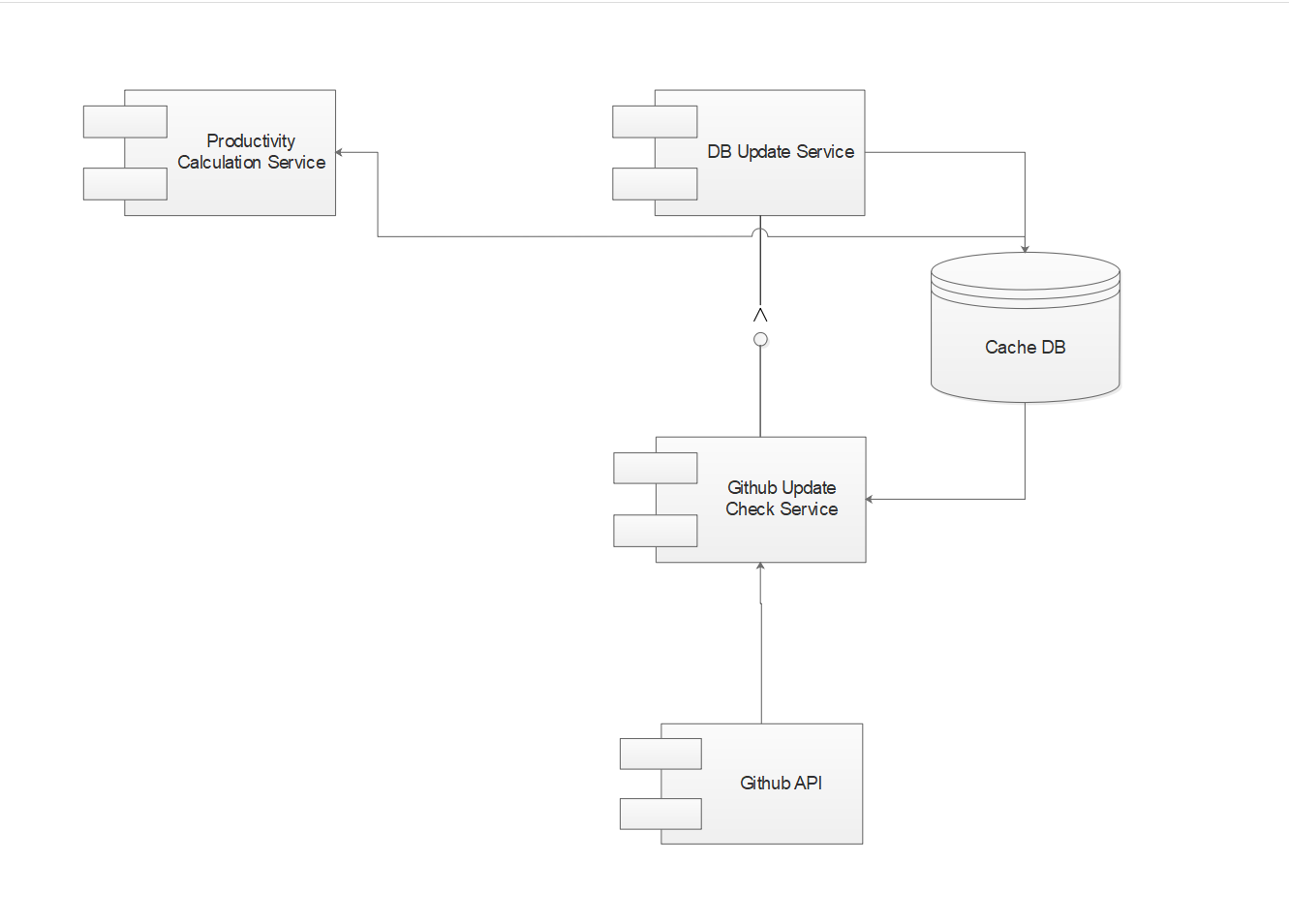
Component Diagram

Figure 18 - Component Diagram

Fault Scenarios and Fault Tolerance Mechanisms

***Scenario 1: Node Failure***

**Fault tolerance approach:** The Kubernetes master control plane is configured to be responsible to solve the issue of node failure by spawning new nodes to replace failed nodes.

***Scenario 2 : Service Failure***

**Fault tolerance approach :**  The concept of Kubernetes replicas is utilized in the deployments so that in the case of a service failure, the Kubernetes control plane realizes that the number of active pods are lower than the number of replicas, thus restarts the pods containing the services

***Scenario 3: Unauthorized Access***

**Fault tolerance approach:**  IAM roles are created for the node-groups to make sure that the nodes in the cluster only have limited and required control over the AWS services.

***Scenario 4 : Interservice communication intercepting***

**Fault tolerance approach :** To make interservice communication efficient and secure we use a hybrid VPC network for the EKS Cluster where the services can communicate with each other using a private network but services can be accessed externally through the public network as well. This VPC configuration prevents against attacks to intercept data in transit between the services.

Conclusion

In conclusion DeveloperIQ presents a new logical and mathematical method for calculating developer productivity by considering multiple dimensions of their GitHub activity along with a highly scalable and fault tolerant infrastructure. One area of improvement in this application is the automation of the manual workflow of container deployment and ECR key validation by using Jenkins or GitHub actions.