SOFE 4790U
Distributed Systems

Lab

Fall 2022

Group # 1

Due Date: September 18th, 2022

Walid Ayub - 100695612

Discussion:

Part 1

The problem in this article is that a user has to work with various services and is required to be updated upon any changes. The problems that occurred could be multiple disparate services, multiple instances of the same services, or multiple versions of the same services. The best possible solution could be that have a gateway between the services and the clients so that the most applicable service can be routed to the consumer. The usage of this pattern almost always has to do with any problems occurring with the client or trying to make it easier and more efficient for them.

Part 2

In this phase, we followed the directions to deploy the request splitting ambassador splits the weight of each deployment between 90% and 10%. The YAML file will show the percentage weight for both the web-development and the experimental deployments and divide it accordingly. The configuration gives the weight an integer of 9 while the experimental will automatically become a 1 which is the default weight. The ambassador deployment is a little different where it creates replicas and each pod will have its own local path.

Part 3

We use load balancing services for this stage that will deploy 3 replicas of servers. Similar to previous steps in the lab but this will deploy 3 pods of the image.

Design:

Autoscaling allows a user the opportunity to improve the memory or CPU and its resources with the use of patterns. It is really important for us to learn this as it is a big game changer for deployments in google cloud computing. I had issues with the CPU and memory in part 3 of the lab and autoscaling is a concept that increases the memory and CPU when it's too much as a default. How it works is that there

are clients who communicate with servers along with any other replicas. It requires the use of the internet and a load balancer. Autoscaling allows the load balancer to scale up or down during its operations for incoming traffic from whether the deployments or pods or nodes. Autoscaling is just another tool in this technological world to make a performance of an operation run smoother and more efficient, and make it easier for the clients. Without it, there are more limits to what a user can do.