ABSTRACT

Video conferencing has become a popular activity today and the development of technology. It is not uncommon for the service to experience downtime with user traffic. To avoid this, an infrastructure that can replicate itself is built so that downtime can be avoided.

The infrastructure is built using a container orchestration tool called Kubernetes. Kubernetes will run the service inside a virtual machine named *node*. Inside the *node*, the minor Kubernetes component is the *pod*. These components can run services concurrently. The Kubernetes Cluster is connected to DigitalOcena and Li*node*, which functions as a Cloud Service.

This final project discusses the design of a Kubernetes cluster in which two Data Centers originate from two different countries. The Data Center can accommodate several Kubernetes *node*s where each *node* accommodates one *pod* that can run one application through a service intermediary. *Node*s in one Data Center can automatically replicate *pod*s and resize them (Auto scalable). This can improvethe High Availability infrastructure and avoid downtime.

There is one service that will be implemented, namely WebRTC. This service will be observed using several parameters testing. There are four parameters to be observed: service quality (Jitter, Throughput, *Delay*, Packet loss) and error rate.

Keywords: *Kubernetes, Cluster, High Availability, Data Center*