ABSTRACT

The rapid development of cloud technology brings developers to keep innovate in order to achieve better performance and resource utilization in building servers. Containerization technique is one of the efforts made by developers as an alternative solution to traditional Virtual Machine (VM). Docker container allows applications to be run on the same Operating System (OS) kernel, as a result server can work more efficient, flexible and lightweight. Kubernetes is a container orchestration system that can automate management, deployment and scaling of these containers.

Microservice architecture offered by Kubernetes allows developers to run multiple applications on single sever independently. With such architecture obviously will cause problem in order of sharing resource usage among applications. Therefore limitation and auto-scaling system needed to be apply in case if there's a spike traffic on the application. Horizontal Pod Auto-scaling (HPA) is one of the scaling system that can be used. The HPA response depend on target average CPU utilization. Furthermore cloud storage Nextcloud will be built on top of Kubernetes cluster with distribution system storage Ceph to provide a dynamic storage system.

The final results of this Final Project shows that service Nextcloud with auto-scaling system can provide better performance compared with system without auto-scaling. The difference values obtained by the parameters are 768 milicore on minimum CPU allocation, 0,61 Mbps on throughput, 6008,16 on successful transactions, 20,03 req/sec on transactions rate and 27,85% on availability.

Keywords *: Ceph, cluster, container, Docker, horizontal pod auto-scaling, microservice.*