ABSTRACT

Currently, the world has entered Industry 4.0, where many technologies can be found in everyday life. Even technology cannot just go away, because technology has been able to simplify and assist humans in doing everything, be it in searching for information or communicating remotely. Technology can be fully integrated with the internet, so that systems and devices can communicate and exchange information via the internet. Moreover, the services provided today are already massive and can be found in many online networks. With these developments, of course, the scalability of these services is very large and increasingly complex. Even servers can accommodate these services with large amounts of data and very quickly. The traffic received by the server must have qualified hardware specifications to overcome overload. If at any time the server cannot handle a very large amount of traffic, the server will experience downtime and cannot serve the various applications provided by the server.

In the Final Project, load balancing is designed and implemented using two algorithms, namely Least Connection and IP Hash on the Kubernetes platform provided by Oracle Cloud Infrastructure. The application that will be carried out for the process of implementing load balancing is a wordpress application with database integration using MySQL to store all wordpress component data. Then, an analysis is carried out for a comparison of the use of the two algorithms. The design of this final project uses the Kubernetes platform, which is a platform for creating server clusters with 3 integrated virtual server configurations. Furthermore, it is carried out to measure the process performance of running applications based on parameters of response time, throughput, request loss, as well as the performance of computing resource usage from servers such as CPU Utilization.

The test results show that the Least Connection algorithm has good performance for distributing requests from clients to worker nodes based on parameters of response time, throughput, request loss on application performance and CPU Utilization on worker node performance. The use of CPU Utilization is below the average usage standard of 80%, the response time varies for each user variant, the throughput increases every time more users send requests to the server, and the request loss is better compared to the IP Hash algorithm. The load balancing implementation can be accessed publicly, so that users can access the application via the internet.

Keywords: Load Balancing, Least Connection, IP Hash, Kubernetes, Cloud Computing