ABSTRACT

Internet users in the world according to the Internet World Stat in 2020 reached 58.7% of internet users worldwide, this percentage increases from year to year. The growth of internet users worldwide, namely 1,167% from 2000 to 2020. A large number of internet users makes web application service providers have to think about the availability of resources for the managed web server. Based on these problems, we need a system that can help server performance and availability. Server clustering allows a number of computers to work together to perform the computation process. Server clustering technology has the advantage of being able to produce a system with a high level of reliability and availability. Kubernetes is a cluster platform for container or container orchestrators. Kubernetes is expected to be a solution so that the computation process is more efficient and the creation of a system with a high level of availability. The system is simulated using a public cloud platform, namely the google cloud platform with google service kubernetes engine. The system is simulated using the google kubernetes engine on the google cloud platform. The average value of Availability was 99.96% of the total 27526 minutes of uptime. With the server computation load on google kubernetes engine, the average CPU usage for each test scenario is quite stable with the highest CPU being 27.178%. From the results of throughput testing, the average value of throughput for each connection is 7,126 Mbit / s with the highest value for 1000 connections, namely 8,764 Mbit / s. From the results of the delay test, the average value for each connection is 23.402 ms with the fastest value is 1000 connections with the results of 16.7854 ms indicating the category "Very Good" based on TIPHON standardization which is less than 150 ms. The highest percentage of packet loss was obtained at 5000 connections at 16.27%..

Keywords: *Web server, high availability, cluster, cloud computing, docker, kubernetes, orchestration*