ABSTRACT

This final project aims to add server resources and improve performance on the Central AI server with horizontal scaling techniques. Because under certain conditions with the current system, the application server becomes very slow in responding to requests given by the user. Therefore, it is necessary to upgrade resources on the server side, one of which is by using the horizontal scaling technique. This technique is done by assigning several additional servers to the cluster and making them work concurrently using Load Balancer technology with the help of Nginx software. The workings of the load balancer technology are to divide the load of requests from users to servers in the cluster evenly. In addition, this final project also discusses a feature of the load balancer called Failover. The purpose of the whole process is to improve performance and provide higher capabilities on the server, or High Availability (HA).

Keywords : Load Balancer, Failover, Horizontal Scaling, Server, High Availability.