

ABSTRAK

Cloud computing merupakan teknologi yang memberikan sumber daya yang dapat diakses oleh siapapun yang memiliki akses tanpa memiliki komputer secara fisik. Pada *cloud computing* ditemui permasalahan dimana beberapa *virtual machine* menjadi kelebihan beban pekerjaan dan yang lain memiliki lebih sedikit. Permasalahan tersebut terjadi pada server di Jurusan Informatika Universitas Pembangunan Nasional "Veteran" Yogyakarta yang memiliki *web service* didalamnya. *Quality of Service* dari server tersebut akan menurun jika *request* yang diberikan dilakukan secara masif dan dalam rentang waktu yang singkat. Keadaan ini terjadi karena *request* yang tidak dapat ditangani dan terjadinya *timeout* yang disebabkan oleh *resource* yang sudah tidak mampu menangani *request* yang masuk.

Pada penelitian ini mengimplementasikan *load balancing* dalam arsitektur *cloud computing* yang digunakan pada server di Jurusan Informatika Universitas Pembangunan Nasional "Veteran" Yogyakarta. Dilakukan pula perbandingan dari algoritma *round-robin* dan *least-connection* dalam *load balancing* untuk mengetahui performa kedua algoritma dalam menangani *traffic request* yang tinggi pada server.

Pengujian dilakukan pada tiga skenario yaitu, tanpa dan menggunakan *load balancing* pada dua dan tiga *virtual machine*. Didapati penggunaan *load balancing* menghasilkan kenaikan nilai *throughput* 22,3%, penurunan *response time* sebesar 73,79%, penurunan CPU *utilization* sebesar 25,52%, dan menghilangkan kemunculan *error*. Pada skenario dua dan tiga, didapati algoritma *least-connection* lebih unggul berdasarkan metrik *throughput* pada skenario 3 dengan perbedaan sebesar 0,18%, *response time* pada skenario 2 memiliki perbedaan sebesar 1,77%, dan CPU *utilization* pada skenario 2 dan 3 memiliki perbedaan sebesar 0,35% dan 1,89%. Dengan *round-robin* unggul dalam metrik *throughput* pada skenario 2 dengan perbedaan sebesar 0,19% dan *response time* pada skenario 3 dengan perbedaan sebesar 0,5%.

Kata kunci: *cloud computing, load balancing, hypervisor, HAProxy, web server*

ABSTRACT

Cloud computing is a technology that provides resources that can be accessed by anyone who has access without having a physical computer. In cloud computing, there is a problem where some virtual machines become overloaded and others have less. This problem occurs on a server at the Informatics Department of the Yogyakarta "Veteran" National Development University which has a web service. The quality of service from the server will decrease if the requests are made massively and in a short period. This situation occurs due to requests that cannot be handled and timeouts caused by resources that can no longer handle incoming requests.

Load balancing was implemented in the cloud computing architecture used on servers at the Informatics Department of the National Development University "Veteran" Yogyakarta. A comparison was also made of the round-robin and least-connection algorithms in load balancing to determine the performance of these two algorithms in handling high-traffic requests on the server.

Tests were carried out in three scenarios, without and using load balancing on two and three virtual machines. It was found that the use of load balancing resulted in an increase in throughput value of 22.3%, a decrease in response time of 73.79%, and a decrease in CPU utilization of 25.52%, and eliminated the occurrence of errors. It is found that in scenarios two and three the least-connection algorithm is superior based on throughput metrics in scenario 3 with a difference of 0.18%, response time in scenario 2 has a difference of 1.77%, and CPU utilization in scenarios 2 and 3 has a difference by 0.35% and 1.89%. The round-robin is superior in throughput metrics in scenario 2 with a difference of 0.19% and response time in scenario 3 with a difference of 0.5%.

Keywords: cloud computing, load balancing, hypervisor, HAProxy, web server