BABI

PENDAHULUAN

1.1 LATAR BELAKANG

Ketidakseimbangan beban kerja server menjadi sebuah masalah yang sangat tidak diharapkan *Cloud Service Provider* (CSP). Ketidakseimbangan ini akan mengakibatkan penurunan kinerja komputasi dan menurunkan kualitas layanan. Ketika sumber daya mengalami kenaikan permintaan dan permintaan tersebut melebihi kapasitasnya, maka sumber daya pada *server* akan habis yang artinya tidak ada lagi *server* yang tersedia untuk menangani permintaan [1].

Load Balancing akan membagikan beban kerja ke setiap server dan memproses permintaan yang dikirimkan, hal ini akan meningkatkan kinerja server secara keseluruhan. Load balancing dapat mendistribusikan beban kerja server secara merata untuk mencapai Service Level Agreement (SLA) dan pemanfaatan sumber daya komputasi yang tepat. Penggunaan dan pembagian beban sumber daya yang tepat dapat meminimalkan konsumsi sumber daya. Hal ini akan membantu penyedia cloud dalam mempraktikkan konsep fault tolerance, scalability, high availability demi meningkatkan kualitas layanan komputasi awan [2]. Penelitian Saeid & Yahiya [3] tahun 2019 membuktikan bahwa menerapkan load balancing harus diterapkan pada server cloud, sehingga server akan tidak akan kelebihan beban dan sumber daya akan digunakan secara tepat. Namun penelitian [3] menggunakan private cloud yang tidak memiliki jaminan kualitas layanan pada cloud, sehingga layanan ini tidak memiliki jaminan waktu aktif server (uptime).

Microsoft Azure sebagai layanan public cloud memberikan jaminan bahwa instance yang tersedia pada tingkat Consumption Tier, Basic Tier, Standard Tier, and Premium Tier dapat melayani permintaan dengan waktu aktif (uptime) setiap bulannya sebesar 99.95% [4]. Berdasarkan permasalahan tersebut, dalam laporan ini membahas implementasi load balancing dengan backend berbasis IP pada Microsoft Azure.

1.2 RUMUSAN MASALAH

Rumusan masalah dalam laporan ini sebagai berikut :

- 1. Bagaimana penerapan *load balancing* dengan *backend* berbasis IP pada *Microsoft Azure* dalam membagikan beban kerja pada *server*?
- 2. Bagaimana performa *server* yang telah diterapkan *load balancing* dalam membagikan beban kerja pada *server* berdasarkan *throughput*, CPU *time* dan *request rate*?

1.3 TUJUAN

Tujuan dalam laporan ini sebagai berikut :

- 1. Mendapatkan hasil dalam penerapan *load balancing* dalam membagikan beban kerja pada *server*.
- 2. Menganalisis performa *server* dalam menangani permintaan berdasarkan *throughput*, CPU *time* dan *request rate*.

1.4 MANFAAT

- 1. Meminimalisir server cloud mengalami kelebihan beban.
- 2. Menjadi referensi dalam menerapkan *load balancing* pada *cloud computing*.
- 3. Menjadi referensi bagi instansi yang ingin menerapkan *load balancing* pada *Microsoft Azure*.