

# BAB I

## PENDAHULUAN

### A. Latar Belakang

Dunia teknologi informasi setiap waktunya mengalami kemajuan yang semakin canggih sehingga membuat banyak kalangan dengan mudah untuk mengoperasionalkannya dalam bentuk apapun terlebih dengan adanya jaringan internet. Keunggulan penggunaan jaringan internet adalah kecepatan dalam menyampaikan informasi dengan biaya yang terjangkau dan cakupan komunikasi yang luas. Salah satu teknologi yang muncul akibat perkembangan internet dan kini ramai digunakan adalah *cloud computing* atau komputasi awan yang memungkinkan kita menyewa sumber daya teknologi informasi berupa *software*, *processing power*, dan *storage* yang dapat diakses dari manapun melalui internet (Abidah et al., 2020). Penyewaan *cloud computing* dapat melalui penyedia layanan *cloud* seperti *Google Cloud Platform (GCP)*, *Amazon Web Services (AWS)*, *Microsoft Azure*, *IDCloudhost*, *Niagahoster* dan yang lainnya kemudian membayar sesuai dengan kebutuhan yang digunakan atau istilahnya *on-demand* (Rashid & Chaturvedi, 2019). Kelebihan dari penggunaan *cloud computing* adalah dapat menghemat biaya investasi, menghemat waktu, operasional dan manajemen lebih mudah, meningkatkan *availability* serta ketersediaan data, dan menghemat biaya operasional pada saat *realibilitas* (Rumetna, 2018).

Manfaat yang ditawarkan oleh *cloud computing* membuat banyak instansi maupun perusahaan pada saat ini mulai berpindah dari penggunaan infrastruktur *on-premise* menjadi *cloud computing* untuk meningkatkan

produktivitas dan pelayanan. Layanan pada sebuah perusahaan sudah pasti akan diakses oleh banyak orang untuk mendapatkan informasi sehingga hal tersebut dapat mengakibatkan peningkatan *traffic* data yang signifikan pada *server* (Wijayanti, 2020). Permasalahan muncul apabila beban kerja yang diterima oleh *server* terlalu tinggi sehingga dapat mempengaruhi kinerja dari *server* yang digunakan dan yang paling parah dapat terjadi *server down*. Hal tersebut dapat mengganggu proses pertukaran data serta kenyamanan dari pengguna layanan sehingga dapat mempengaruhi daya tarik pelanggan organisasi, menurunkan pendapatan dan kehilangan reputasi (Jader et al., 2019).

Konsep HA (*High Availability*) merupakan jawaban dari permasalahan tersebut dimana HA merupakan sebuah pendekatan yang digunakan untuk meminimalisir kegagalan pada sebuah sistem atau *server*. Hal tersebut dapat dilakukan dengan adanya *server* duplikasi yang mampu mem-*backup server* agar ketika terdapat gangguan pada *server* maka beban kerja *server* akan digantikan oleh *server* yang lain (Iryani et al., 2022). Salah satu metode HA yang sering digunakan pada *cloud computing* adalah *load balancing*. *Load balancing* digunakan untuk membagi beban *traffic* antara jalur ke *server* yang satu dengan jalur ke *server* yang lainnya agar tidak terjadi *overload* pada salah satu jalur dan *traffic* dapat berjalan secara optimal sehingga dapat meningkatkan kualitas dan kestabilan dari layanan *server* dengan waktu tanggap yang kecil serta *throughput* yang maksimal (Pratama et al., 2021).

Sehingga dalam penelitian ini penulis bertujuan untuk melakukan analisis perbandingan kinerja dan *responsivitas* antara infrastruktur *cloud web*

*server* yang menggunakan *high availability* atau *load balancing* dengan *single deployment* tanpa *load balancing* pada lingkungan *cloud* GCP. Diharapkan dengan adanya penelitian ini dapat memberikan wawasan yang lebih baik tentang *high availability* terhadap kinerja *web server*, sehingga dapat menjadi panduan bagi administrator sistem dalam mengoptimalkan infrastruktur *web* mereka di lingkungan GCP.