

BAB I

PENDAHULUAN

1.1 Latar Belakang

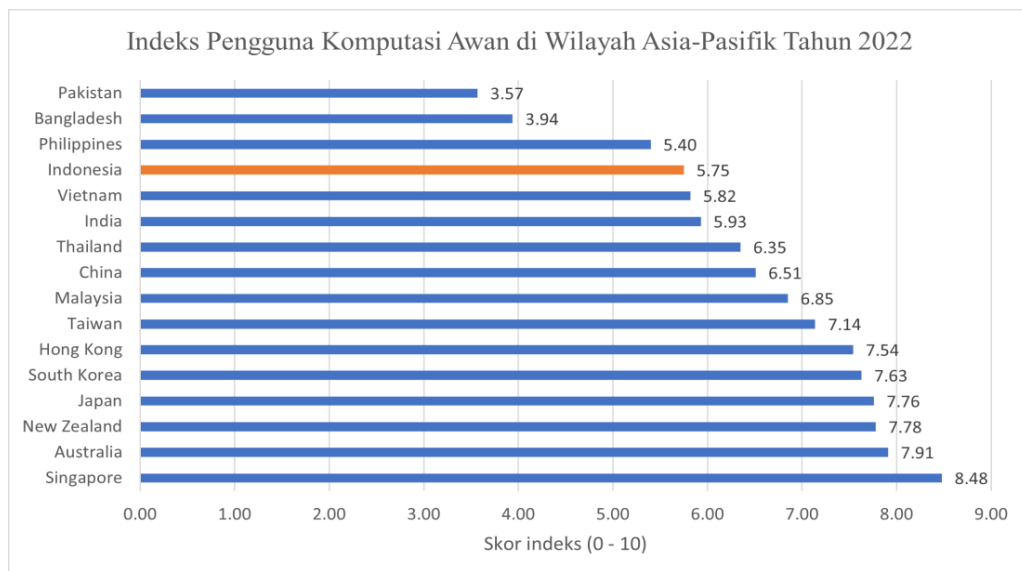
Pada era modern, kemajuan teknologi dan informasi yang pesat memberikan dampak terhadap perkembangan bidang jaringan, yang ditandai dengan munculnya teknologi virtualisasi, sebuah inovasi yang memiliki peran krusial dalam memfasilitasi pengembangan sistem jaringan yang lebih efisien dan fleksibel (Danuri, 2019). Teknologi virtualisasi sendiri mengacu pada pendekatan yang memungkinkan penggunaan sumber daya komputasi secara efisien melalui pembentukan versi virtual dari infrastruktur fisik (Data et al., 2020).

Seiring dengan perkembangan teknologi virtualisasi, muncul konsep baru yang bernama komputasi awan, yaitu suatu model layanan komputasi yang memungkinkan akses mudah dan *on-demand* ke sumber daya komputasi yang dapat dibagikan (seperti *server*, penyimpanan, basis data, jaringan, perangkat lunak, analisis, dan kecerdasan buatan) melalui internet (Apriliani dkk., 2022). Dalam komputasi awan, pengguna tidak perlu memiliki atau mengelola secara langsung infrastruktur fisik tersebut, sebaliknya hanya perlu menyewa sumber daya sesuai kebutuhan tersendiri dari penyedia layanan komputasi awan (Putri dkk., 2022).

Secara global, penyedia layanan komputasi awan seperti Google Cloud Platform (GCP), Amazon Web Service (AWS), dan Microsoft Azure berfungsi sebagai entitas yang menyediakan sumber daya untuk memenuhi kebutuhan pengguna di bidang teknologi. GCP memiliki keunggulan yang tidak dimiliki oleh- penyedia layanan komputasi lainnya. Sebagai contoh, GCP dapat melakukan *load balancing* lintas wilayah (*cross-region*) secara instan ketika *user* membutuhkan dua wilayah yang berbeda untuk memenuhi ketersediaan performa aplikasinya (*cloud.google.com*, 2023; Sekaran & Krishna, 2017). Selain itu,

dalam konfigurasi *load balancer*, *Content Delivery Network* (CDN) sudah terintegrasi dengan alat *back-end*, sehingga dapat mengurangi kompleksitas konfigurasi lebih lanjut, dan memudahkan pengguna untuk mengoptimalkan pengiriman konten *web* (Wang dkk., 2018).

Selanjutnya setiap penyedia layanan komputasi awan memiliki pusat data yang tersebar di berbagai lokasi geografis dan dibagi menjadi *region* serta *zones*, sebagai salah satu contoh *region*nya ialah Asia-Pasifik (Nazir dkk., 2022). *Region* Asia-Pasifik menjadi bagian integral yang berfungsi sebagai penyedia akses sumber daya komputasi di wilayah geografis yang melibatkan beberapa jumlah *zones* seperti "asia-southeast1a", "asia-southeast1b", dan "asia-southeast1c" (Jelidi dkk., 2019). Pada tahun 2022, berdasarkan data yang dipublikasikan oleh *Statista.com.*, (2022) di *region* Asia-Pasifik, Singapura mendominasi sebagai pengguna komputasi awan tertinggi yang di ikuti dengan Australia, New Zealand, Jepang dan lainnya. Data tersebut dapat dilihat pada gambar 1.1 yang mencerminkan indeks pengguna komputasi awan di wilayah Asia-Pasifik Tahun 2022.



Gambar 1.1 Indeks pengguna komputasi awan di wilayah Asia-Pasifik

Sumber : (*Statista.com.*, 2022)

Sahat Parulian, 2024

ANALISIS DAN IMPLEMENTASI INFRASTRUKTUR KOMPUTASI AWAN BERBASIS WEB DENGAN PEMANFAATAN LOAD BALANCING DAN AUTO-SCALING PADA GOOGLE CLOUD PLATFORM

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

Selain itu, berdasarkan data yang tercantum pada Gambar 1.1, Indonesia menduduki posisi ke-13 dari total 16 negara yang terletak di kawasan Asia-Pasifik, dengan skor indeks sebesar 5,7 data tersebut didukung dengan hasil survei *Centre for Strategic and International Studies* (CSIS) yang menyatakan sebanyak 69,8% lembaga di Indonesia belum menggunakan layanan komputasi awan di Tahun 2022 (*katadata.co.id.*, 2022). Hal ini menunjukkan bahwa pemanfaatan teknologi komputasi awan di Indonesia masih berada pada tingkat yang relatif rendah jika dibandingkan dengan negara-negara lain di kawasan ini. Rendahnya penggunaan teknologi komputasi awan dapat mempengaruhi terhadap kinerja sebuah *web* terhadap *server* (O. Nugroho & Hutagalung, 2020). Sebagai contoh nyata, berdasarkan laporan media massa digital, sebuah *web* yang dimiliki oleh Badan Usaha Milik Negara (BUMN) mengalami penurunan kinerja karena tingginya jumlah *traffic* terhadap *server* sehingga *web* tidak dapat diakses (*Kompas.com.*, 2022).

Tidak hanya itu saja sebuah *web* untuk Penerima Peserta Didik Baru (PPDB) milik dinas Pendidikan Kabupaten Klaten dan Kota Makassar mengalami *server down* ketika *user* ingin mengakses *web* tersebut, permasalahan ini terjadi diakibatkan kenaikan *traffic* terhadap *server* dalam waktu yang bersamaan. (*Detik.com.*, 2022.) Menurut peneliti, permasalahan ketidakmampuan *server* dalam menangani kenaikan jumlah *traffic* yang tinggi, dikarenakan infrastruktur *server* yang buruk dan kurang mendukung sehingga mengakibatkan penurunan performa *web* dalam menangani informasi kebutuhan *user* untuk jangka waktu lama. Sebab, berdasarkan penelitian yang dilakukan oleh (Riskiono & Darwis, 2020) bahwa performa *web* dapat dipengaruhi dengan dua metode yang terdapat dalam infrastruktur *server*, yaitu *load balancing* dan *auto-scaling*.

Selain itu, penelitian yang dilakukan oleh (Mada & Ladi, 2023) menyatakan bahwa dengan menggunakan *load balancing* pada mesin virtual yang memiliki spesifikasi 2 GB RAM dan 2 vCPU, mampu mendistribusikan *traffic* secara merata di antara beberapa sumber daya komputasi ketika jumlah *traffic* lebih dari

Sahat Parulian, 2024

ANALISIS DAN IMPLEMENTASI INFRASTRUKTUR KOMPUTASI AWAN BERBASIS WEB DENGAN PEMANFAATAN LOAD BALANCING DAN AUTO-SCALING PADA GOOGLE CLOUD PLATFORM

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

5.000 *request* dan *auto-scaling* berperan untuk menambah dan mengurangi jumlah mesin virtual yang digunakan pada saat *traffic* tinggi dan rendah. Oleh sebab itu, komputasi awan seharusnya dianggap sebagai salah satu alternatif untuk menciptakan infrastruktur yang optimal dalam mencegah penurunan performa *web* dengan penggunaan metode *load balancing* dan *auto-scaling* (Mada & Ladi, 2023). Maka dari itu, peneliti berkeinginan untuk melakukan penelitian terhadap infrastruktur layanan komputasi awan dengan menggunakan kelebihan yang dimiliki oleh Google Cloud Platform untuk menerapkan metode *load balancing* dan *auto-scaling* sebagai solusi untuk mengatasi tantangan dalam menangani *traffic* tinggi dan rendah kepada *server*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, maka dapat dikemukakan beberapa rumusan masalah dalam penelitian ini, yaitu:

- 1) Bagaimana rancangan infrastruktur komputasi awan dalam implementasi *web* untuk mengatasi *traffic* tinggi dan rendah?
- 2) Bagaimana performa metode *load balancing* dan *auto-scaling* dalam implementasi *web* guna menangani *traffic* tinggi dan rendah di lingkungan komputasi awan?
- 3) Bagaimana perbandingan biaya terhadap performa mesin virtual yang digunakan dalam lingkungan komputasi awan?

1.3 Batasan Masalah

Berdasarkan rumusan masalah yang sudah dipaparkan, terdapat cakupan batasan masalah yang dijelaskan sebagai berikut:

- 1) Terdapat banyak platform yang dapat digunakan untuk membuat infrastruktur komputasi awan, pada penelitian ini hanya menggunakan Google Cloud Platform.
- 2) Penelitian ini hanya berfokus pada implementasi rancangan infrastruktur dan performansi komputasi awan di sebuah *web* statis yang akan

Sahat Parulian, 2024

ANALISIS DAN IMPLEMENTASI INFRASTRUKTUR KOMPUTASI AWAN BERBASIS WEB DENGAN PEMANFAATAN LOAD BALANCING DAN AUTO-SCALING PADA GOOGLE CLOUD PLATFORM

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

terintegrasi dengan *auto-scaling* dan *load balancer*.

- 3) Tidak membahas keamanan komputasi awan.
- 4) Penggunaan sumber daya komputasi awan harus dimonitor untuk mengontrol besaran biaya yang dikeluarkan. Oleh karena itu, batasan anggaran perlu diperhatikan dengan seksama dalam penggunaan layanan komputasi awan pada suatu penelitian. Dalam konteks penelitian ini, batasan anggaran akan diarahkan pada alokasi tipe mesin virtual yang berbeda untuk setiap konfigurasinya. Dengan demikian, penelitian ini tidak akan menggunakan mesin dengan kapasitas yang melebihi batasan anggaran yang telah ditetapkan oleh peneliti.
- 5) *Traffic* rendah dalam penelitian ini sebesar 0 – 5000 *user* dan *Traffic* tinggi sebesar >10000

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah dan batasan masalah, terdapat beberapa tujuan dalam penelitian ini, yang dijelaskan sebagai berikut:

- 1) Mengetahui rancangan infrastruktur komputasi awan dalam implementasi *web* untuk mengatasi *traffic* tinggi dan rendah.
- 2) Menentukan performa *load balancing* dan *auto-scaling* dengan memperhatikan *down-time* saat menambahkan dan mengurangi *instance baru* ketika mengalami *traffic* tinggi dan rendah.
- 3) Menentukan biaya terhadap performa mesin virtual yang cocok dari pengimplementasi *web* yang menggunakan beberapa jenis mesin virtual berbeda.

1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat secara teoritis maupun manfaat secara praktis sebagai berikut:

Sahat Parulian, 2024

ANALISIS DAN IMPLEMENTASI INFRASTRUKTUR KOMPUTASI AWAN BERBASIS WEB DENGAN PEMANFAATAN LOAD BALANCING DAN AUTO-SCALING PADA GOOGLE CLOUD PLATFORM

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

1) Manfaat Teoritis

Penelitian ini diharapkan dapat menjadi referensi dan panduan bagi para insinyur komputasi awan atau peneliti lainnya yang tertarik untuk mengetahui kinerja infrastruktur komputasi awan dan performa *load balancing* dan *auto-scaling* dalam konteks komputasi awan atau penelitian serupa. Hal ini bertujuan untuk mendukung perancangan infrastruktur yang lebih optimal dengan memanfaatkan Google Cloud Platform.

2) Manfaat Praktis

Penelitian ini memiliki manfaat praktis, diantaranya:

- a) Bagi Perusahaan atau Organisasi yang ingin menerapkan *load balancing* dan *auto-scaling* untuk implementasi *web*, diharapkan melalui penelitian ini dapat menjadi acuan dalam merancang infrastruktur komputasi awan yang relevan dengan menggunakan Google Cloud Platform.
- b) Bagi Peneliti, diharapkan mampu menjadi sarana penerapan ilmu yang telah diperoleh selama aktivitas perkuliahan di Program Studi S1 Sistem Telekomunikasi Universitas Pendidikan Indonesia.

1.6 Struktur Organisasi Penulisan

Penelitian ini memiliki suatu struktur organisasi penulisan yang terdiri dari 5 (Lima) BAB meliputi:

1) Bab I Pendahuluan

Bab ini memuat suatu latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

2) Bab II Kajian Pustaka

Bab ini membahas landasan teori dan penelitian-penelitian relevan terdahulu.

3) Bab III Metode Penelitian

Bab ini menjelaskan seputar alur penelitian, bagan perancangan dan implemmtasi sistem, dan identifikasi kebutuhan penelitian.

Sahat Parulian, 2024

ANALISIS DAN IMPLEMENTASI INFRASTRUKTUR KOMPUTASI AWAN BERBASIS WEB DENGAN PEMANFAATAN LOAD BALANCING DAN AUTO-SCALING PADA GOOGLE CLOUD PLATFORM

Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

4) Bab IV Hasil dan Pembahasan

Bab ini mencakup hasil pengujian dari sistem dan bahasan yang didapatkan setelah melewati proses uji coba.

5) BAB V Kesimpulan

Bab ini memuat kesimpulan dari hasil pengujian sistem dan aplikasi beserta saran dan implikasi dari penelitian.