

ABSTRAK

Ketersediaan aplikasi adalah faktor penting dalam keberjalanan aplikasi karena memiliki dampak langsung pada pengalaman pengguna. Penerapan *autoscaling* pada lingkungan *cloud computing* dapat membantu menjaga ketersediaan dan skalabilitas suatu aplikasi sekaligus meningkatkan efisiensi biaya operasional. Garvan News sebagai aplikasi publik menginginkan penerapan *autoscaling* pada server aplikasi mereka namun terkendala pada keterbatasan sumber daya manusia. Permasalahan tersebut dapat ditangani melalui pendekatan *Infrastructure As Code* (IAC) untuk meningkatkan konsistensi dan efisiensi dalam pengelolaan server aplikasi. Penerapan *autoscaling* dengan IAC dilakukan dengan menganalisis kebutuhan dan membuat desain arsitektur dan desain *stack*. Stack dibagi menjadi 3 bagian untuk kemudian diimplementasikan ke dalam lingkungan AWS secara berurutan. Kemudian dilakukan uji beban menggunakan Apache Jmeter terhadap aplikasi web untuk memperoleh beban maksimal yang dapat ditangani oleh AWS ketika terjadi lonjakan permintaan. *Autoscaling* AWS dapat menangani permintaan dengan baik ketika diuji dengan 25000 dan 40000 pengguna. Namun memiliki error percentage yang sangat tinggi saat dibebani dengan 50000 pengguna.

Kata Kunci : *autoscaling*, *infrastructure as code*, AWS, pengujian beban