

INTISARI

OPTIMASI PERFORMA LAYANAN *INTERNATIONAL IMAGE INTEROPERABILITY FRAMEWORK* UNTUK PENYAJIAN CITRA *SLIDE* MIKROSKOPIS MENGGUNAKAN ORKESTRASI KLASTER DAN *DYNAMIC LOAD BALANCING*

Sugeng Dwi Cahyono

23/531008/PPA/06757

Penyajian citra digital *Whole Slide Imaging* (WSI) dalam skala besar menghadapi tantangan utama berupa tingginya *response time* dan keterbatasan *bandwidth* pada sistem penyimpanan berbasis *cloud*. Permasalahan ini semakin kompleks seiring meningkatnya permintaan akses simultan dan ukuran file yang besar. Penelitian ini mengusulkan optimalisasi layanan *International Image Interoperability Framework* (IIIF) melalui penerapan orkestrasi *cluster* dan *dynamic load balancing* berbasis *microservices*. Pengujian dilakukan dengan memvariasikan jumlah pengguna *concurrent*, jumlah *node cluster*, dan dan membandingkan algoritma *load balancer* seperti *Round Robin* (RR), *Dynamic Weighted Round Robin* (DWRR), dan *Dynamic Smooth Weighted Round Robin* (DSWRR).

Hasil penelitian menunjukkan bahwa peningkatan kapasitas container (2 CPU, 2 GB) mampu menurunkan *response time* hingga 84,98% (P50) dan 69,11% (P90), serta meningkatkan RPS hingga 267,91%. Skalabilitas horizontal dengan empat container menghasilkan performa paling stabil pada seluruh skenario ramp up. Selain itu, algoritma DWRR terbukti memberikan performa terbaik pada beban menengah hingga tinggi, dengan RPS tertinggi di seluruh pengujian serta latensi yang lebih rendah dibandingkan DSWRR dan RR. Dari aspek biaya, pendekatan horizontal scaling berbasis *cluster* terbukti lebih efisien dibandingkan vertical scaling, dengan estimasi penghematan sekitar 83,8% atau 6,18 kali lebih murah serta biaya operasional yang lebih elastis karena kapasitas dapat disesuaikan mengikuti fluktuasi beban tanpa mengunci biaya pada ukuran komputasi terbesar. Kombinasi antara peningkatan kapasitas baseline, penambahan jumlah container, serta penerapan algoritma DWRR terbukti efektif mengatasi tingginya *response time* pada proses *tiling IIIF* sekaligus meningkatkan efisiensi distribusi beban dan efisiensi biaya operasional secara signifikan.

Kata Kunci: *Whole Slide Imaging, Dynamic Load Balancer, International Image Interoperability Framework, Orkestrasi Cluster, Optimasi Response time*

ABSTRACT

OPTIMIZING THE PERFORMANCE OF INTERNATIONAL IMAGE INTEROPERABILITY FRAMEWORK TILE DELIVERY FOR WHOLE SLIDE IMAGING USING CLUSTER ORCHESTRATION

Sugeng Dwi Cahyono

23/531008/PPA/06757

Large-scale Whole Slide Imaging (WSI) delivery faces two dominant constraints: high response time and limited bandwidth in cloud-based storage environments. These constraints become more severe as concurrent access increases and file sizes grow. This study proposes an optimization of International Image Interoperability Framework (IIIF) services through cluster orchestration and microservices-based dynamic load balancing. Experiments were conducted by varying the number of concurrent users and cluster nodes, and by comparing load-balancing strategies, including Round Robin (RR), Dynamic Weighted Round Robin (DWRR), and Dynamic Smooth Weighted Round Robin (DSWRR).

The results show that increasing container capacity to 2 vCPU and 2 GB memory reduces response time by up to 84.98% at P50 and 69.11% at P90, while improving throughput (RPS) by 267.91%. Horizontal scaling with four containers delivers the most stable performance across all ramp-up scenarios. Among the evaluated algorithms, DWRR consistently performs best under medium-to-high workloads, achieving the highest RPS and lower latency than both DSWRR and RR. From a cost perspective, cluster-based horizontal scaling is more economical than vertical scaling, yielding an estimated cost reduction of approximately 83.8% about 6.18× cheaper. This advantage is reinforced by operational elasticity: capacity can be adjusted with workload fluctuations instead of being locked to a peak-sized compute instance. Overall, the combination of an optimized baseline configuration, four-container horizontal scaling, and DWRR load balancing effectively mitigates IIIF tiling latency under request bursts, while improving both load distribution efficiency and operational cost efficiency.

Keywords: *Whole Slide Imaging, Dynamic Load Balancer, International Image Interoperability Framework, Cluster Orchestration, Response Time Optimization*