

Pengembangan Dashboard dan Analysis Tools untuk Infrastruktur Machine Learning Fakultas Ilmu Komputer Universitas Indonesia = Development of Dashboard and Analysis Tools for Machine Learning Infrastructure at Faculty of Computer Science, University of Indonesia

Hakim Amarullah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920535438&lokasi=lokal>

Abstrak

Proses training model membutuhkan sumber daya komputasi yang akan terus meningkat seiring dengan bertambahnya jumlah data dan jumlah iterasi yang telah dicapai. Hal ini dapat menimbulkan masalah ketika proses training model dilakukan pada lingkungan komputasi yang berbagi sumber daya seperti pada infrastruktur komputasi berbasis kluster. Masalah yang ditimbulkan terutama terkait dengan efisiensi, konkurensi, dan tingkat utilisasi sumber daya komputasi. Persoalan efisiensi muncul ketika sumber daya komputasi telah tersedia, tetapi belum mencukupi untuk kebutuhan job pada antrian teratas. Akibatnya sumber daya komputasi tersebut menganggur. Penggunaan sumber daya tersebut menjadi tidak efisien karena terdapat kemungkinan sumber daya tersebut cukup untuk mengeksekusi job lain pada antrian. Selain itu, pada cluster computing juga membutuhkan sistem monitoring untuk mengawasi dan menganalisis penggunaan sumber daya pada kluster. Penelitian ini bertujuan untuk menemukan resource manager yang sesuai untuk digunakan pada kluster komputasi yang memiliki GPU agar dapat meningkatkan efisiensi, implementasi sistem monitoring yang dapat membantu analisis penggunaan sumber daya sekaligus monitoring proses komputasi yang sedang dijalankan pada kluster, dan melayani inference untuk model machine learning. Penelitian dilakukan dengan cara menjalankan eksperimen penggunaan Slurm dan Kubernetes. Hasil yang diperoleh adalah Slurm dapat memenuhi kebutuhan untuk job scheduling dan mengatur penggunaan GPU dan resources lainnya pada kluster dapat digunakan oleh banyak pengguna sekaligus. Sedangkan untuk sistem monitoring, sistem yang dipilih adalah Prometheus, Grafana, dan Open OnDemand. Sementara itu, sistem yang digunakan untuk inference model adalah Flask dan Docker.

.....

The amount of computational power needed for the model training process will keep rising along with the volume of data and the number of successful iterations. When the model training process is conducted in computing environments that share resources, such as on cluster-based computing infrastructure, this might lead to issues. Efficiency, competition, and the level of resource use are the three key issues discussed. Efficiency problems occur when there are already computing resources available, yet they are insufficient to meet the demands of high-level workloads. The power of the machine is subsequently wasted. The utilization of such resources becomes inefficient because it's possible that they would be adequate to complete other tasks on the front lines. A monitoring system is also necessary for cluster computing in order to track and assess how resources are used on clusters. The project seeks to set up a monitoring system that can assist in analyzing the usage of resources while monitoring the computing processes running on the cluster and locate a suitable resource manager to be utilized on a computing cluster that has a GPU in order to increase efficiency, also serve inference model in production. Slurm and Kubernetes experiments were used to conduct the investigation. The findings show that Slurm can handle the demands of job scheduling, manage the utilization of GPUs, and allow for concurrent use of other cluster resources. Prometheus,

Grafana, and Open OnDemand are the chosen monitoring systems. Else, inference model is using Flask and Docker as its system constructor.