

Evaluasi dan analisis capacity scheduler dan fair scheduler pada teknologi big data hadoop framework = Evaluation and analysis capacity scheduler and fair scheduler in hadoop framework on big data technology / Diyanatul Husna

Diyanatul Husna, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20433096&lokasi=lokal>

Abstrak

ABSTRAK

Apache Hadoop merupakan framework open source yang mengimplementasikan MapReduce yang memiliki sifat scalable, reliable, dan fault tolerant. Scheduling merupakan proses penting dalam Hadoop MapReduce. Hal ini dikarenakan scheduler bertanggung jawab untuk mengalokasikan sumber daya untuk berbagai aplikasi yang berjalan berdasarkan kapasitas sumber daya, antrian, pekerjaan yang dijalankan, dan banyaknya pengguna. Pada penelitian ini dilakukan analisis terhadap Capacity Scheduler dan Fair Scheduler. Pada saat Hadoop framework diberikan 1 pekerjaan dengan ukuran data set 1,03 GB dalam satu waktu. Waiting time yang dibutuhkan Capacity Scheduler dan Fair Scheduler adalah sama. Run time yang dibutuhkan Capacity Scheduler lebih cepat 6% dibandingkan Fair Scheduler pada single node. Sedangkan pada multi node Fair Scheduler lebih cepat 11% dibandingkan Capacity Scheduler. Pada saat Hadoop framework diberikan 3 pekerjaan secara bersamaan dengan ukuran data set (1,03 GB) yang sama dalam satu waktu. Waiting time yang dibutuhkan Fair Scheduler lebih cepat dibandingkan Capacity Scheduler yaitu 87% lebih cepat pada single node dan 177% lebih cepat pada multi node. Run time yang dibutuhkan Capacity Scheduler lebih cepat dibandingkan Fair Scheduler yaitu 55% lebih cepat pada single node dan 212% lebih cepat pada multi node. Turnaround time yang dibutuhkan Fair Scheduler lebih cepat dibandingkan Capacity Scheduler yaitu 4% lebih cepat pada single node, sedangkan pada multi node Capacity Scheduler lebih cepat 58% dibandingkan Fair Scheduler. Pada saat Hadoop framework diberikan 3 pekerjaan secara bersamaan dengan ukuran data set yang berbeda dalam satu waktu yaitu data set 1 (456 MB), data set 2 (726 MB), dan data set 3 (1,03 GB) dijalankan secara bersamaan. Pada data set 3 (1,03 GB), waiting time yang dibutuhkan Fair Scheduler lebih cepat dibandingkan Capacity Scheduler yaitu 44% lebih cepat pada single node dan 1150% lebih cepat pada multi node. Run time yang dibutuhkan Capacity Scheduler lebih cepat dibandingkan Fair Scheduler yaitu 56% lebih cepat pada single node dan 38% lebih cepat pada multi node. Turnaround time yang dibutuhkan Capacity Scheduler lebih cepat dibandingkan Fair Scheduler yaitu 12% lebih cepat pada single node, sedangkan pada multi node Fair Scheduler lebih cepat 25,5% dibandingkan Capacity Scheduler

ABSTRACT

Apache Hadoop is an open source framework that implements MapReduce. It is scalable, reliable, and fault tolerant. Scheduling is an essential process in Hadoop MapReduce. It is because scheduling has responsibility to allocate resources for running applications based on resource capacity, queue, running tasks, and the number of user. This research will focus on analyzing Capacity Scheduler and Fair Scheduler. When hadoop framework is running single task. Capacity Scheduler and Fair Scheduler have the same waiting time. In data set 3 (1,03 GB), Capacity Scheduler needs faster run time than Fair Scheduler which is

6% faster in single node. While in multi node, Fair Scheduler is 11% faster than Capacity Scheduler. When hadoop framework is running 3 tasks simultaneously with the same data set (1,03 GB) at one time. Fair Scheduler needs faster waiting time than Capacity Scheduler which is 87% faster in single node and 177% faster in muliti node. Capacity Scheduler needs faster run time than Fair Scheduler which is 55% faster in single node and 212% faster in multi node. Fair Scheduler needs faster turnaround time than Capacity Scheduler which is 4% faster in single node, while in multi node Capacity Scheduler is 58% faster than Fair Scheduler. When hadoop framework is running 3 tasks simultaneously with different data set, which is data set 1 (456 MB), data set 2 (726 MB), and data set 3 (1,03 GB) in one time. In data set 3 (1,03 GB), Fair Scheduler needs faster waiting time than Capacity Scheduler which is 44% faster in single node and 1150% faster in muliti node. Capacity Scheduler needs faster run time than Fair Scheduler which is 56% faster in single node and 38% faster in multi node. Capacity Scheduler needs faster turnaround time than Fair Scheduler which is 12% faster in single node, while in multi node Fair Scheduler is 25,5% faster than Capacity Scheduler