

Intisari

Hadoop Mapreduce merupakan framework yang populer digunakan untuk mengolah *Big Data*. Namun terdapat persoalan yang ditemui pengguna maupun pengembang aplikasi Hadoop MapReduce, di antaranya adalah optimasi performa melalui konfigurasi Hadoop dan mengelola *resource* secara efektif. Untuk mencapai performa yang optimal dan pengelolaan *resource* yang efektif, diperlukan pendalaman mengenai konfigurasi Hadoop, di mana terdapat lebih dari 200 parameter konfigurasi yang dapat memengaruhi performa Hadoop MapReduce.

Berbagai penelitian telah dilakukan untuk mengatasi permasalahan tersebut seperti pengujian pengaruh konfigurasi-konfigurasi tertentu terhadap kinerja MapReduce atau melakukan otomasi terhadap konfigurasi sesuai kebutuhan pengguna, namun penelitian-penelitian tersebut belum meneliti pengaruh konfigurasi pembagian *resource* terhadap kinerja Hadoop MapReduce. Pada penelitian ini dilakukan analisis pengaruh pembagian *resource* berupa jumlah container per node terhadap kinerja Hadoop MapReduce dengan indikator waktu, CPU, dan memori beserta menganalisis keterkaitan antara metrik-metrik yang lebih rinci antar indikator-indikator tersebut. Penelitian ini dilakukan dengan mengeksekusi Job MapReduce dengan konfigurasi container per node yang berbeda-beda dan juga diteliti pada lingkungan alokasi RAM yang berbeda-beda.

Ditemukan bahwa tidak terdapat pola tertentu pada waktu eksekusi Job MapReduce seiring naiknya jumlah container per node, akan tetapi ditemukan keterkaitan antara indikator-indikator kinerja dan ditemukan angka jumlah container per node yang relatif optimum yaitu 4 container per node.

Kata kunci : *Hadoop, MapReduce, YARN, pembagian resource, container per node.*

Abstract

Hadoop MapReduce is a populer framework used to process BiG Data. Yet there are numbers of problem that users and Hadoop MapReduce application developers face, mainly performance optimization through Hadoop configuration and effectively managing resource. To achieve optimized performance and effective resource management, in depth understanding on Hadoop configuration is needed, as there are more than 200 configuration parameters which can impact Hadoop MapReduce's performance.

A number of research have been done to resolve the problem mentioned such as testing the effect of certain configuration on MapReduce's performance or automating configuration based on the user's need, yet these research have not investigate the effect of resource partition configuration on Hadoop MapReduce. In this research, analysis is done on the effect of resource partitioning using container per node number on Hadoop MapReduce with time, CPU, and memory used as indicator, and how their subindicators interact with each other. This research is done by excuting MapReduce job with different numbers of contaner per node and done in different RAM allocation environment.

It is found that there is no pattern of MapReduce Job execution time as the number of container per node increase, but it is also found that there are relations between it's performance's indicator and there is a relatively optimum number of container per node, which is 4 container per node.

Keywords : *Hadoop, MapReduce, YARN, resource partition, container per node.*