

Analisis performa kecepatan mapreduce pada hadoop menggunakan TCP packet flow analysis = Execution time performance analysis of hadoop mapreduce using TCP packet flow analysis

Priagung Khusumanegara, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20387453&lokasi=lokal>

Abstrak

Komputasi terdistribusi merupakan salah satu kemajuan teknologi dalam mengolah data. Penggunaan komputasi terdistribusi memudahkan user untuk mengolah data menggunakan beberapa komputer yang secara fisik terpisah atau terdistribusi. Salah satu teknologi yang menggunakan konsep komputasi terdistribusi adalah Hadoop. Hadoop merupakan framework software berbasis Java dan open source yang berfungsi untuk mengolah data yang memiliki ukuran yang besar secara terdistribusi. Hadoop menggunakan sebuah framework untuk aplikasi dan programming yang disebut dengan MapReduce. Enam skenario diimplementasikan untuk menganalisa performa kecepatan MapReduce pada Hadoop. Berdasarkan hasil pengujian yang dilakukan diketahui penambahan jumlah physical machine dari satu menjadi dua physical machine dengan spesifikasi physical machine yang sesuai perancangan dapat mempercepat kecepatan rata-rata MapReduce. Pada ukuran file 512 MB, 1 GB, 1.5 GB, dan 2 GB, penambahan physical machine dapat mempercepat kecepatan rata-rata MapReduce pada masing-masing ukuran file sebesar 161.34, 328.00, 460.20, dan 525.80 detik. Sedangkan, penambahan jumlah virtual machine dari satu menjadi dua virtual machine dengan spesifikasi virtual machine yang sesuai perancangan dapat memperlambat kecepatan rata-rata MapReduce. Pada ukuran file 512 MB, 1 GB, 1.5 GB, dan 2 GB, penambahan virtual machine dapat memperlambat kecepatan rata-rata MapReduce pada masing-masing ukuran file sebesar 164.00, 504.34, 781.27, dan 1070.46 detik. Berdasarkan hasil pengukuran juga diketahui bahwa block size dan jumlah slot map pada Hadoop dapat mempengaruhi kecepatan MapReduce.

.....

Distributed computing is one of the advance technology in data processing. The use of distributed computing allows users to process data using multiple computers that are separated or distributed physically. One of technology that uses the concept of distributed computing is Hadoop. Hadoop is a Java-based software framework and open source which is used to process the data that have a large size in a distributed manner. Hadoop uses a framework for application and programing which called MapReduce. Six scenarios are implemented to analyze the speed performance of Hadoop MapReduce. Based on the study, known that the additional the number of physical machines from one to two physical machines with suitable specifications design can speed up the average speed of MapReduce. On file 512 MB, 1 GB, 1.5 GB, and 2 GB size additional the number of physical machines can accelerate MapReduce average speed on each file size for 161.34, 328.00, 460.20, and 525.80 seconds. Meanwhile, additional the number of virtual machines from one to two virtual machines with suitable specifications design can slow down the average speed of MapReduce. On file 512 MB, 1 GB, 1.5 GB, and 2 GB size, additional the number of virtual machines can slow down the average speed of each MapReduce on a file size for 164.00, 504.34, 781.27, and 1070.46 seconds. Based on the measurement result is also known that the block size and number of slot maps in Hadoop MapReduce can affect speed.