

Abstract

Hadoop is a framework that commonly used for distributed processing and distributed storage in big data. Hadoop cluster processed and store the data on each node in the cluster. Hadoop cluster development need funds and high energy. Beside that, management of the cluster also challenging.

Container-based virtualization as LXC and OpenVZ are used to performs higher performances, nearly as the physical cluster. LXC will replace OpenVZ in future. Early research stated that LXC outperforms OpenVZ in I/O throughput and MapReduce Job execution time on top of unmanaged virtual machine and different I/O scheduler. In this research, I used Proxmox as management platform and same I/O scheduler on both (OpenVZ and LXC).

In Proxmox, the I/O throughput and execution time of Mapreduce Job in syslog parser case LXC gives results 5 seconds faster on OpenVZ. OpenVZ also give higher I/O throughput compared to LXC.

Kata kunci: big data, virtualization, hadoop, distributed computing, performance analysis, container-based