

# Performansi Pemrosesan *Big Data* pada Virtualisasi Berbasis *Container* dan *Hypervisor*

Muhammad Rashief<sup>1</sup>, Sidik Prabowo<sup>2</sup>, Siti Amatullah Karimah<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>rashiefm@student.telkomuniversity.ac.id, <sup>2</sup>pakwowo@telkomuniversity.ac.id,

<sup>3</sup>karimahsiti@telkomuniversity.ac.id

---

## 1. Pendahuluan

*Big Data* merupakan istilah yang mulai digunakan untuk menggambarkan proses penerapan kekuatan komputasi yang serius dan untuk kumpulan informasi yang sangat besar dan seringkali sangat rumit[1]. Kebutuhan akan kecepatan data yang besar membebankan tuntutan pada infrastruktur komputasi yang mendasarinya[2]. Hadoop adalah *framework* populer yang digunakan untuk analisis data tak terstruktur yang cepat dan hemat biaya[3]. Dimana dari beberapa *ecosystem* Hadoop, HDFS dan MapReduce merupakan inti dari komponen Hadoop, dimana HDFS dan MapReduce merupakan komponen untuk penyimpanan dan pemrosesan Hadoop[4]. Karena Hadoop digunakan untuk pemrosesan dan penyimpanan *dataset* yang sangat besar dan dirancang untuk bekerja pada node fisik, *deployment* dan *maintenance*-nya memiliki biaya yang mahal[5]. Dimana dengan mem-virtualisasi akan memberikan *flexibility* dan mengurangi biaya. Teknologi virtualisasi yang ada dibebankan secara kasar antara solusi berbasis container dan hypervisor[6]. Virtualisasi berbasis container memberikan kinerja yang mendekati aslinya, dimana container bekerja pada tingkat sistem operasi dan berbagi kernel yang sama dengan *host*[7].

### Latar Belakang

Saat ini hampir semua Hadoop *cluster* berjalan diatas *virtual machine*, semua perusahaan seperti Yahoo, dan Facebook menggunakan *virtual machine* untuk Hadoop *cluster information*. Dimana ketika meng-*install* OS pada *virtual machine* diharuskan meng-*install* keseluruhan *package* termasuk kernel, aplikasi dasar, dan lainnya[8]. Untuk mempresentasikan kedua virtualisasi tersebut, penelitian ini menggunakan Docker untuk container dan VMware untuk hypervisor dimana Docker saat ini adalah solusi container yang paling populer dan VMware adalah salah satu pimpinan di pasar hypervisor[6].

Pada penelitian terkait[6], peneliti mengatakan masih sedikitnya pembuktian kuantitatif untuk hipotesis yang diberikan dengan cara "apple to apple". Atas dasar tersebut, penelitian ini dilakukan dengan skenario yang berbeda dengan menjalankan Hadoop diatasnya, dan tingkat *dataset* yang berbeda.

### Topik dan Batasannya

Beberapa batas yang terdapat pada tugas akhir ini adalah :

- Hadoop yang dijalankan diatas container hanya menggunakan Docker.
- Hadoop yang dijalankan diatas hypervisor hanya menggunakan VMware.
- Hanya menggunakan satu *physical machine*.
- Menggunakan tiga jenis pengujian yaitu pengujian untuk *execution time*, *disk I/O*, dan *CPU utilization*.