

Perbandingan Kinerja Block Storage Ceph dan ZFS di Lingkungan Virtual

Performance Comparison of Ceph and ZFS Block Storage in Virtual Environment

Faza Abdani Auni Robbi, Agung Budi Prasetyo, Eko Didik Widiyanto*)

Departemen Teknik Komputer, Fakultas Teknik, Universitas Diponegoro
Jl. Prof. Soedarto, SH, Kampus Undip Tembalang, Semarang, Indonesia 50275

Cara sitasi: F. A. A. Robbi, A. B. Prasetyo, and E. D. Widiyanto, "Perbandingan Kinerja Block Storage Ceph dengan ZFS," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 1, 2019. doi: 10.14710/jtsiskom.7.1.2019.7-11, [Online].

Abstract - The growth of data requires better performance in the storage system. This study aims to analyze the comparison of block storage performance of Ceph and ZFS running in virtual environments. Tests were conducted to measure their performances, including IOPS, CPU usage, throughput, OLTP Database, replication time, and data integrity. Testing was done using 2 node servers with a standard configuration of the storage system. Server virtualization uses Proxmox on each node. ZFS has a higher performance of reading and writing operation than Ceph in IOPS, CPU usage, throughput, OLTP and data replication duration, except the CPU usage in writing operation. The test results are expected to be a reference in the selection of storage systems for data center applications.

Keywords - block storage performance; Ceph; Proxmox virtual environment; ZFS

Abstrak – Berkembangnya data memerlukan kinerja sistem penyimpanan yang lebih baik. Penelitian ini bertujuan menganalisis perbandingan kinerja block storage dalam menjalankan lingkungan virtual antara sistem penyimpanan Ceph dan ZFS. Parameter perbandingan meliputi IOPS, penggunaan CPU, throughput, database OLTP, waktu replikasi, dan Integritas data. Pengujian dilakukan menggunakan 2 node server dengan konfigurasi standar pada setiap sistem penyimpanannya. Virtualisasi server menggunakan proxmox pada setiap node. ZFS mempunyai kinerja yang lebih unggul daripada Ceph dalam operasi pembacaan dan penulisan berdasarkan IOPS, penggunaan CPU, throughput, OLTP dan waktu replikasi data, kecuali di penggunaan CPU untuk operasi penulisan. Hasil pengujian ini diharapkan dapat menjadi acuan dalam pemilihan sistem penyimpanan untuk aplikasi pusat data.

Kata Kunci – kinerja block storage; Ceph; lingkungan virtual Proxmox; ZFS

I. PENDAHULUAN

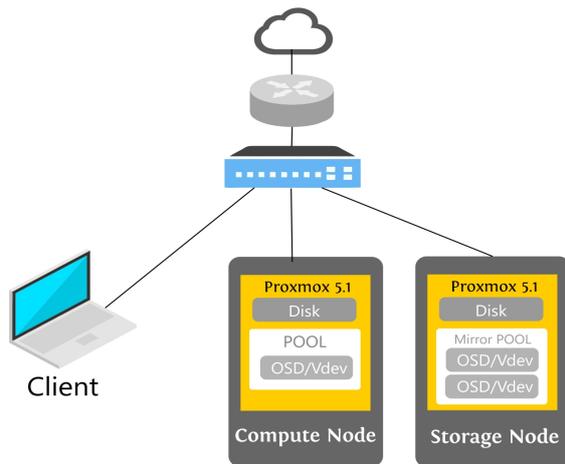
Perkembangan teknologi pusat data saat ini mengarah pada virtualisasi dan komputasi awan (*cloud computing*) [1], [2]. Perkembangan tersebut juga membuat data yang tersimpan semakin bertambah banyak [3]. Pusat data ini memerlukan penyimpanan yang sangat handal untuk mengakomodasi kebutuhan tersebut [4].

Media penyimpanan pada pusat data sewaktu-waktu dapat mengalami kerusakan karena banyak faktor, seperti kegagalan sumber daya, cacat pabrik, dan usia pakai [3]. Data pada media penyimpanan tersebut dapat hilang dan menimbulkan kerugian akibat kerusakan fisik media penyimpanan. Masalah tersebut dapat diminimalkan dengan membuat sistem penyimpanan yang bersifat *redundant* dan mendukung integritas data yang disimpan, misalnya dengan menggunakan Ceph (sistem penyimpanan terdistribusi yang *scalable*) yang mampu mendukung redundansi data [5]. Saat ini, sistem Ceph banyak digunakan di lingkungan *cloud* terutama sebagai *back-end block storage* seperti Openstack yang menyediakan penyimpanan yang *scalable*.

Ceph menyimpan data ke dalam blok-blok data menggunakan RBD (*RADOS Block Device*). Selain mendukung redundansi data, dengan konfigurasi yang tepat Ceph dapat meningkatkan kinerja sistem penyimpan *back-end* tervirtualisasi [6]. Ceph ini telah diimplementasikan untuk membantu pengelolaan dan pertukaran data menggunakan *Software Defines Networking* (SDN) yang tidak memerlukan perangkat khusus pada sistem penyimpanan, salah satunya adalah OSIRIS [7]. OSIRIS menghubungkan peneliti dari berbagai kampus untuk mengkolaborasikan data dengan mudah tanpa membangun infrastruktur sendiri.

Selain Ceph, dalam virtualisasi telah banyak digunakan beragam sistem penyimpanan lain, seperti ZFS, GlusterFs, HDFS, dan Lustre [8]. Sistem file ZFS dikembangkan oleh Oracle sebagai pengganti *file system* tradisional dan *volume manager* yang terdapat dalam sebagian besar sistem operasi berbasis Linux atau Unix. ZFS dapat menyediakan kapasitas yang besar dan kinerja yang tinggi serta jaminan integritas data dan pengelolaan yang sederhana [9].

*) Penulis korespondensi (Eko Didik Widiyanto)
Email: didik@live.undip.ac.id



Gambar 1. Arsitektur sistem penyimpanan

Uji kinerja sistem file secara terpisah telah dilakukan dalam [6] untuk Ceph serta [9] dan [10] untuk ZFS. Pengujian menunjukkan bahwa Ceph dan ZFS mempunyai kinerja yang baik sebagai sistem penyimpanan. Uji perbandingan kinerja antar sistem file juga telah dilakukan. [8] membandingkan kinerja beragam sistem file terdistribusi menggunakan *map reduce*, yaitu Glusterfs, Lustre, Ceph dan HDFS. [11] membandingkan kinerja sistem file GlusterFS dan Ceph serta [12] membandingkan ZFS dan LVM di lingkungan virtual.

Berbeda dengan [8] yang menggunakan tolak ukur *map reduce* serta [11] dan [12] yang membandingkan tipe sistem file berbeda, kajian ini bertujuan membandingkan kinerja sistem penyimpanan *block storage* berbasis Ceph dan ZFS dalam lingkungan virtual menggunakan Proxmox. Penelitian ini memilih lingkungan virtual yang mengarah ke cloud privat. Kinerja yang dibandingkan meliputi IOPS (*Input/output Operations Persecond*), penggunaan CPU, *throughput*, nilai OLTP (*Online Transaction Processing*), waktu replikasi, dan integritas data.

II. METODE PENELITIAN

Desain sistem untuk mengukur kinerja *block storage* ini dilakukan dengan menggunakan 2 server, 1 switch, dan 1 *client*, seperti yang ditunjukkan dalam Gambar 1. Spesifikasi server yang digunakan ditunjukkan oleh Tabel 1. Masing-masing server menjalankan lingkungan virtual Proxmox VE 5.1.

Dua server tersebut digunakan sebagai *compute node* dan *storage node*. Versi sistem yang dibandingkan adalah Ceph RBD (*Rados Block Device*) versi luminous dan ZFS versi 0.7.8-1 yang ditunjukkan pada Tabel 2.

Antara 2 *node* dan 1 *client* terhubung dengan switch melalui *gigabit network ethernet card*. *Compute node* merupakan *node* dimana VM dipasang dan menjadi tempat Proxmox VE berjalan (sebagai *hypervisor*) untuk menjalankan VM yang ada. *Storage Node* dijadikan sebagai *node* penyimpanan utama.

Pada *compute node* terdapat Proxmox VE 5.1 yang terpasang pada *harddisk* 500 GB. Sebuah SSD (*Solid*

Tabel 1. Spesifikasi server yang digunakan

Komponen	Server	
	Compute Node	Storage Node
Prosesor	Intel core I5-2540M @ 2.6 GHz upto 3.3 Ghz	Pentium (R) Dual Core E5700 @ 3.00 GHz
RAM	6 GB	4GB
Penyimpanan	SSD 128 GB, HDD 500GB	2 X HDD 1 TB 7200 RPM
NIC	Gigabit Ethernet	Gigabit Ethernet
Perangkat Lunak	Proxmox VE 5.1	Proxmox VE 5.1

Tabel 2. Spesifikasi sistem penyimpanan

	Sistem File 1	Sistem File 2
Nama Sistem penyimpanan	Ceph	ZFS
Format image	RBD	ZFS
Versi	Luminous	ZFS 0.7.8-1

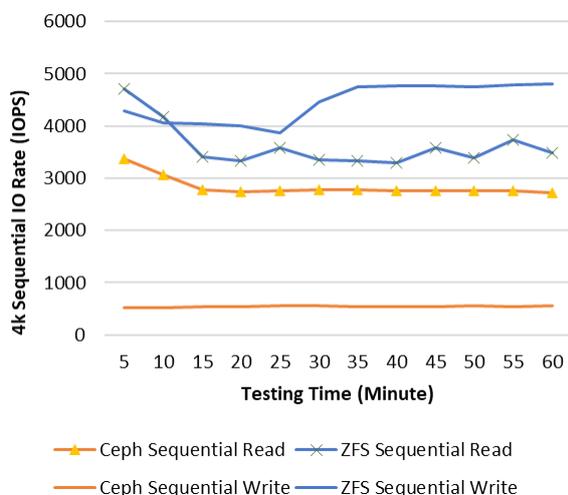
State Disk) berukuran 120 GB sebagai tempat penyimpanan agar saat pengujian tidak terjadi *bottleneck* antara *harddisk* yang terdapat pada *storage node* dengan *compute node*.

Pada *storage node* terdapat 3 buah *harddisk*, 1 *harddisk* berukuran 160 GB sebagai tempat Proxmox 5.1 dan 2 *harddisk* berukuran 1 TB. Masing-masing *harddisk* 1 TB menjadi OSD/Vdev yang bersifat *mirroring* yang dijadikan sebagai objek pengujian kinerja, waktu replikasi data dan integritas data. Dua OSD/Vdev tersebut di atur sebagai *mirroring* dengan menjadikannya sebagai sebuah *pool*. Switch yang digunakan sebagai penghubung antara *compute node* dengan *storage node* yaitu Mikrotik RB951G dengan menggunakan konfigurasi standar.

Sistem penyimpanan terdiri dari 2 OSD/Vdev yang masing-masing berupa 1 *harddisk*. 2 OSD/Vdev tersebut terdapat pada *storage node*. Dua OSD/Vdev ini bersifat *redundant* dimana antara 1 OSD/Vdev dengan OSD/Vdev yang lain saling tereplikasi (*mirroring*) sehingga apabila terjadi kerusakan pada salah satu OSD/Vdev, maka terdapat *backup* data lain pada OSD/Vdev *mirroring*-nya.

Sistem mengecek keadaan *harddisk* yang terdapat pada *pool*. Apabila terjadi kerusakan pada *harddisk*, sistem masih dapat berjalan, akan tetapi hanya dengan sebuah *harddisk*. Hal ini dikarenakan sistem yang bersifat *redundant*. *Harddisk* yang tidak dilakukan penggantian maka dapat mengakibatkan kegagalan sistem. Apabila tidak terjadi kerusakan, maka sistem berjalan normal.

Apabila dilakukan pergantian *harddisk* yang rusak, maka sistem diharapkan melakukan replikasi data. Replikasi data yang dilakukan yaitu antara sebuah *harddisk* normal dengan *harddisk* baru yang menggantikan *harddisk* yang rusak.



Gambar 2. Nilai IO 4k pembacaan dan penulisan sekuensial

III. HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan membandingkan beberapa parameter berupa IOPS, *throughput*, penggunaan CPU, *database OLTP*, waktu replikasi, dan integritas data.

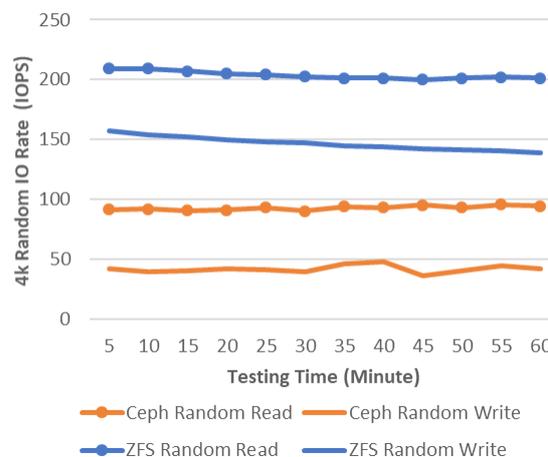
A. IOPS (Input/output Operation Per Second)

Pengujian IO dilakukan menggunakan *file* 4k dengan tujuan menguji kinerja sistem dalam melakukan penulisan dan pembacaan file berukuran kecil baik secara sekuensial maupun acak (*random*). Pengujian dilakukan pada VM windows 7 menggunakan Iometer sebagai *benchmark tool*. Iometer dikonfigurasi dengan memilih target *disk* yang sebelumnya sudah dilakukan *mount* pada VM. *Maximum disk size* yang diberikan sebesar 100000000 *sector* dengan 32 *outstanding I/O* per target. Pengujian dilakukan selama 60 menit.

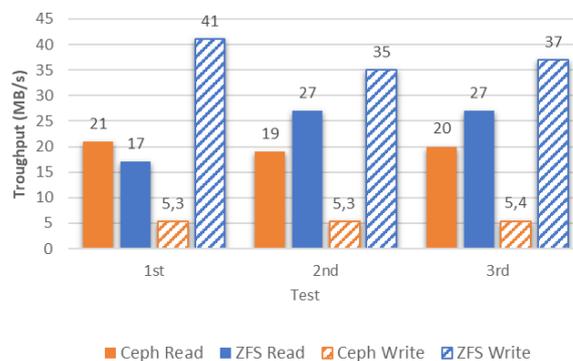
Gambar 2 menunjukkan hasil pengujian IOPS 4k pembacaan dan penulisan sekuensial dari kedua sistem. Nilai rata-rata 4k pembacaan sekuensial pada Ceph adalah 2838 IOPS, sedangkan untuk ZFS adalah 3612 IOPS. Nilai rata-rata 4k penulisan sekuensial pada Ceph sebesar 540 IOPS, sedangkan ZFS 4445 IOPS. Nilai IOPS sistem penyimpanan ZFS lebih besar dibandingkan dengan sistem penyimpanan Ceph.

Gambar 3 menunjukkan nilai rata-rata IOPS pembacaan acak Ceph yaitu sebesar 92,7, sedangkan nilai rata-rata IOPS pembacaan acak ZFS sebesar 203 IOPS. Nilai rata-rata IOPS penulisan acak Ceph sebesar 41 IOPS, sedangkan ZFS sebesar 146 IOPS. Nilai IOPS Pengujian 4k pembacaan dan penulisan acak nilai ZFS lebih unggul dibanding dengan Ceph.

Kinerja pembacaan dan penulisan sekuensial Ceph yang dihasilkan jauh lebih tinggi daripada [6] yang menghasilkan nilai pembacaan sekuensial sebesar 736 IOPS, sedangkan nilai penulisan sekuensial 41,5 IOPS. Kinerja ini diuji dengan parameter Ceph *default*. Pembacaan acak Ceph yang dihasilkan lebih rendah daripada [6] sebesar 212 IOPS, sedangkan penulisan



Gambar 3. Nilai IO 4k pembacaan dan penulisan acak



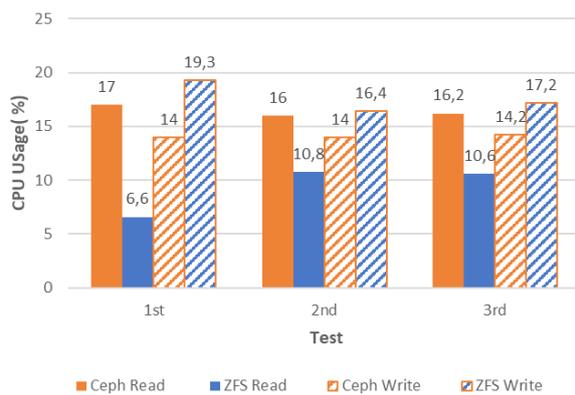
Gambar 4. Throughput pembacaan dan penulisan

lebih cepat, yaitu sebesar 12 IOPS. Penelitian tersebut menggunakan perangkat berbeda dan menjalankan 9 VM aktif dan berfokus pada penyetelan parameter Ceph. Penyetelan parameter Ceph yang tepat dapat mempercepat pembacaan acak dalam Ceph seperti dalam [6].

B. Throughput dan Penggunaan CPU

Pengujian *throughput* dan penggunaan CPU dilakukan dengan mengirimkan file dengan ekstensi ISO sebanyak 44 file dengan total file 25,35 GB. Pengujian dilakukan dalam VM windows 7. Percobaan yang dilakukan sebanyak 3 kali. Target *disk* yang diujikan berdasarkan *storage node*. Jika dilakukan penulisan, maka yang terjadi adalah penulisan ke *storage node* dari *compute node*. Jika dilakukan pembacaan, maka yang terjadi adalah pembacaan pada *storage node* dan *compute node*. Setiap akan melakukan pengujian, dilakukan *restart server* agar tidak terjadi *cache* pada RAM dan *harddisk*.

Terdapat perbedaan yang tidak terlalu jauh pada *throughput* pembacaan seperti yang ditunjukkan pada Gambar 4 dimana ZFS mempunyai *throughput* sedikit lebih besar dibandingkan dengan CEPH. Dari 3 kali



Gambar 5. Penggunaan CPU pembacaan dan penulisan

pengujian nilai *throughput* pembacaan rata-rata Ceph sebesar 20 MB/s dan ZFS 23,6 MB/s, sedangkan untuk penulisan terdapat perbedaan yang jauh, yaitu *throughput* rata-rata penulisan pada Ceph sebesar 5.3 MB/s, sedangkan ZFS sebesar 37 MB/s.

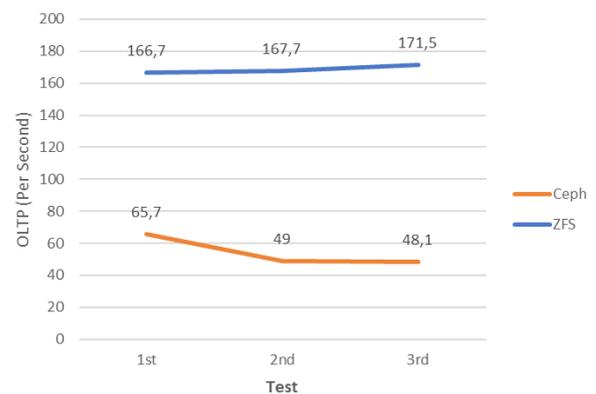
Kinerja *throughput* ZFS yang dihasilkan lebih kecil daripada [12] yang memperoleh nilai *throughput* pembacaan pada ZFS sebesar 50 MB/s, sedangkan untuk penulisan sebesar 60 MB/s. Perbedaan ini disebabkan pengujian *throughput* dalam [12] tidak dilakukan pada VM.

Kinerja *throughput* Ceph yang dihasilkan lebih rendah daripada [11] yang menguji kinerja Ceph pada VM menggunakan virtio sebagai *driver* KVM dengan penyimpanan terdistribusi. Nilai *throughput* pembacaan Ceph yang diperoleh adalah sebesar 60 MB/s, sedangkan untuk *throughput* penulisan sebesar 38 MB/s. Perbedaan ini disebabkan dalam [11] menggunakan Sata dengan penyimpanan tunggal dalam satu *node server*, sedangkan penelitian ini dilakukan operasi penulisan dan pembacaan dilakukan antar 2 *node* terpisah.

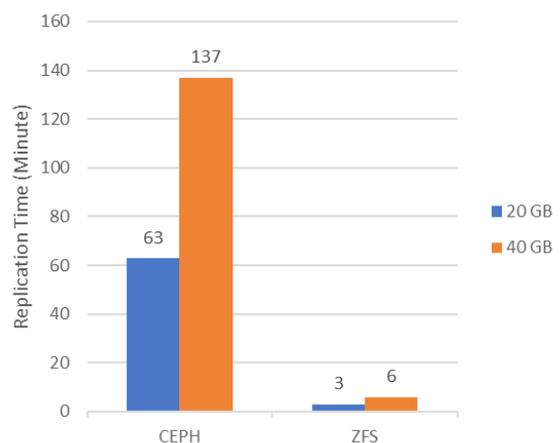
Penggunaan CPU pembacaan *storage node* dari 3 kali pengujian yang dilakukan, penggunaan CPU pada sistem penyimpanan Ceph lebih besar dari ZFS (Gambar 5). Penggunaan CPU rata-rata pada Ceph adalah 16,4 persen, sedangkan pada ZFS 9,3 persen berbeda 7 persen. Pada penggunaan CPU penulisan, sistem penyimpanan ZFS menggunakan sumber daya CPU yang lebih besar dibanding dengan Ceph. Penggunaan CPU rata-rata pada Ceph adalah 14 persen, sedangkan pada ZFS 17,6 persen berbeda 3.5 persen. Dalam operasi pembacaan, Ceph membutuhkan CPU lebih banyak daripada ZFS, sedangkan dalam operasi penulisan sebaliknya.

C. Akses Database

Pengujian *database* dilakukan menggunakan sysbench pada VM Debian 8. Nilai yang dibandingkan adalah nilai OLTP per detik dari tiap sistem penyimpanan. Pengujian OLTP dilakukan pada



Gambar 6. Nilai OLTP per detik



Gambar 7. Waktu replikasi data

database Mysql. Pengujian dengan membuat sebanyak 1000000 baris dengan waktu pengujian 1 menit. Pengujian dilakukan dengan 3 kali pengujian pada tiap sistem penyimpanan.

Gambar 6 menunjukkan hasil nilai OLTP dalam 3 kali pengujian. Nilai OLTP Ceph mempunyai nilai berkisar 3 kali lebih rendah dibanding dengan ZFS. Nilai rata-rata OLTP dari Ceph sebesar 54 OLTP per detik, sedangkan untuk ZFS 168 OLTP per detik. Ceph mengalami penurunan dari percobaan pertama dan kedua, sedangkan ZFS mengalami kenaikan. Nilai maksimum OLTP pada Ceph sebesar 65,7 OLTP, sedangkan untuk ZFS sebesar 171,5 OLTP. Nilai minimum OLTP pada Ceph sebesar 48,1 OLTP, sedangkan untuk ZFS 171,5 OLTP. Dari hal tersebut, ZFS mempunyai kinerja yang lebih baik dalam operasi *database*.

D. Replikasi Data

Pengujian waktu replikasi data bertujuan untuk mengetahui waktu yang diperlukan oleh sistem untuk membangun kembali *mirror disk* setelah terjadi *disk failure* pada salah *harddisk* setelah dilakukan penggantian *harddisk* yang baru. Pengujian dilakukan dengan melepas salah satu *harddisk* secara bergantian kemudian menghapus isi partisi, kemudian melakukan

mount kembali *harddisk* tersebut ke dalam *node*. Pengujian waktu replikasi data dilakukan pada *storage node*. Pengujian dilakukan dengan 2 ukuran dari kapasitas yang terpakai yaitu 20 GB dan 40 GB.

Waktu replikasi Ceph mempunyai selisih yang sangat jauh dibanding dengan ZFS seperti yang ditunjukkan oleh Gambar 7. Ceph memerlukan waktu 63 menit untuk melakukan replikasi data 20 GB, sedangkan ZFS hanya memerlukan waktu 3 menit. Replikasi dengan ukuran 40 GB Ceph menghabiskan waktu 137 menit, sedangkan ZFS hanya menghabiskan waktu 6 menit.

E. Integritas Data

Pengecekan integritas data dilakukan untuk memastikan bahwa sistem penyimpanan dapat memberikan integritas data ketika salah satu *disk* terjadi *harddisk failure*. Pengecekan integritas data dilakukan dengan membandingkan nilai *hash file* sebelum terjadi *harddisk failure* dengan setelah terjadi *harddisk failure*. Pengecekan integritas dilakukan menggunakan algoritma SHA256 dengan *file* berekstensi *jpg*. SHA256 dipilih karena mempunyai akurasi yang lebih tinggi dengan nilai *hash* 256 bit tidak seperti algoritma *checksum* berbasis *md5* yang mempunyai nilai *hash* 128 bit. Pengecekan dilakukan pada VM Debian yang sebelumnya sudah dibuat.

Pengecekan awal dilakukan ketika server berjalan normal. *File* gambar *jpg* yang ada menunjukkan nilai *hash sha256 c1e7253b12fc64877b17036d1cb133b17475a4b1d5d8d40100e3f6d299b70282*. Pengecekan kedua dilakukan ketika proses pergantian *harddisk* dan replikasi data selesai dilakukan. Dari pengecekan kedua, nilai *hash* yang didapatkan mempunyai nilai yang sama, yaitu *7253b12fc64877b17036d1cb133b17475a4b1d5d8d40100e3f6d299b70282*. Pengujian integritas dilakukan pada kedua sistem penyimpanan dan menunjukkan hasil yang sama. Keduanya menunjukkan kemampuan integritas data yang baik ketika terjadi *harddisk failure* sesuai dengan [5], [9].

IV. KESIMPULAN

Dari pengujian yang telah dilakukan, hampir dari semua pengujian sistem penyimpanan ZFS lebih unggul dibanding dengan Ceph, baik itu IOPS, penggunaan CPU, *throughput*, OLTP dan waktu replikasi data, kecuali dalam penggunaan CPU pembacaan pada saat dilakukan penulisan dimana ZFS mempunyai persentase yang lebih tinggi yang menunjukkan penggunaan sumber daya CPU ZFS lebih besar dibanding dengan Ceph. Pengujian integritas data menghasilkan hasil yang sama yaitu Ceph dan ZFS sama-sama memiliki integritas data yang baik.

DAFTAR PUSTAKA

[1] P. Kedia, R. Nagpal, and Tejinder Pal Singh, "A Survey on Virtualization Service Providers,

Security Issues, Tools and Future Trends," *International Journal of Computer Applications*. Noida, Uttar Pradesh, pp. 36–42, 2013.

[2] H. Singh and D. Seehan, "Current Trends in Cloud Computing A Survey of Cloud Computing Systems," *International Journal of Electronics and Computer Science Engineering*, vol. 1, pp. 1215–1219, 2012.

[3] N. Vurukonda and B. T. Rao, "A Study on Data Storage Security Issues in Cloud Computing," *Procedia Computer Science*, vol. 92, pp. 128-135 2016.

[4] K. Singh, *Learning Ceph*. Birmingham, UK: Packt Publishing Ltd, 2015.

[5] S. A. Weil, S. A. Brandt, E. L. Miller, and D. D. E. Long, "Ceph: A Scalable, High-Performance Distributed File System," in *7th Symposium on Operating Systems Design and Implementation*, Washington, USA, Nov. 2006, pp. 307-320.

[6] S. Meyer and J. P. Morrison, "Impact of Single Parameter Changes on Ceph Cloud Storage Performance," *Scalable Computing*, vol. 17, no. 4, pp. 285–298, 2016.

[7] S. McKee, E. Kissel, B. Meekhof, M. Swany, C. Miller, and M. Gregorowicz, "OSIRIS: A Distributed Ceph Deployment using Software Defined Networking for Multi-Institutional Research," *IOP Journal of Physics: Conference Series*, vol. 898, 2017.

[8] M. Vaidya and S. Deshpande, "Critical Study of Performance Parameters on Distributed File Systems Using MapReduce," *Procedia Computer Science*, vol. 78, pp. 224-232, 2016.

[9] *Solaris™ ZFS and Red Hat Enterprise Linux Ext3: File System Performance*, Sun Microsystems, Santa Clara, CA, USA, 2007. [online]. Available: http://www.cs.utexas.edu/users/dahlin/Classes/GradOS/papers/zfs_linux.pdf

[10] E. D. Widiyanto, A. B. Prasetijo, and A. Ghufroni, "On the Implementation of ZFS (Zettabyte File System) Storage System," in *2016 3rd International Conference on Information Technology, Computer and Electrical Engineering*, Semarang, Indonesia, Oct. 2016, pp. 408-413.

[11] J. J. Johari, M. F. Khalid, M. Nizam, M. Mydin, and N. Wijee, "Comparison of Various Virtual Machine Disk Images Performance on GlusterFS and Ceph Rados Block Devices," in *the 3rd International Conference on Informatics & Applications*, Kuala Terengganu, Malaysia, 2014, pp. 1-7.

[12] V. Phromchana, N. Nupairoj, and K. Piromsopa, "Performance Evaluation of ZFS and LVM (with ext4) for Scalable Storage System," in *2011 8th International Joint Conference on Computer Science and Software Engineering*, Nakhon Pathom, Thailand, May 2011, pp. 250-253.