論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Design, Optimization and Evaluation of a Hierarchical Filesystem
著者(和文)	徐天棋
Author(English)	Tianqi Xu
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第11025号, 授与年月日:2018年12月31日, 学位の種別:課程博士, 審査員:松岡 聡,増原 英彦,遠藤 敏夫,脇田 建,額田 彰
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第11025号, Conferred date:2018/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	数理・計算科学　専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of （ Philosophy ）
学生氏名： Student's Name	徐 天棋	指導教員（主）： Academic Supervisor(main)	特任教授　松岡 聡
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Large-scale clusters have been utilized to advance scientific discoveries for decades with their state-of-the-art computational performance. By running in large-scale, the run time of applications that handle big-data can be reduced from days and years to hours. The computational performance has been growing dramatically fast, which help to handle such large-scale applications more efficiently. However, the performance of the storage system can hardly catch up with the performance increment of the computation, which makes the storage system a bottleneck in the large-scale clusters. Moreover, with the increasing amount of data being produced every day to be analyzed, the performance gap between the computational and the storage system exacerbated.

In this thesis, we conduct three works to address such challenges in two major kinds of clusters for running large-scale data-intensive applications, the HPC centers and Clouds. Although by providing high performance instances, such as the HPC instance on Amazon AWS, the Cloud matches the computational performance with the HPC centers, the performance of the storage systems is still far behind the performance of the parallel file system in HPC centers. The major Cloud storage systems are built on object storage technologies, for better load-balancing, scalability and availability. However, the object storage introduces additional limitations to the workload, such as the N-1 write and consistency issues, which limits the performance of applications, especially for the large-scale data-intensive processing. Such limitations make the storage the bottleneck of the system while executing large-scale data-intensive applications, degrades the performance of the applications, and costs users more due to the Cloud pricing policy. To help to improve the performance of the storage on the Cloud, we propose a software-level burst buffer system, called CloudBB. CloudBB utilizes instances as a temporary buffer space to buffer the I/O data from other instances. By buffering the intermediate data within a high-performance network, CloudBB accelerates both the read and write operations from the applications. Moreover, we propose a Master-Worker and Key-Value architecture to achieve both scalability while maintaining the high metadata performance. We implemented CloudBB on top of FUSE so that the applications need no code modification to use CloudBB. We evaluated our CloudBB on Amazon AWS Cloud environment with both micro-benchmarks and real applications. We observed significant performance improvement against the Amazon S3 storage.

Even though the HPC centers equipment with much higher performance parallel file systems compared to the Cloud, with the dramatic increment of the computational performance in HPC, the performance of the parallel file systems can hardly catch up, with multiple users sharing the same file system, the parallel file systems become the bottleneck in the HPC centers for executing large-scale data-intensive applications. Burst buffer systems have been deployed in the latest cutting-edge HPC centers to buffer the bursty I/O and improve the I/O performance. However, having burst buffer systems involves additional procedures including procurement, deployment and maintenance. Therefore, it is not economically nor logistically feasible to have a burst buffer system in every HPC center. Moreover, physically deployed burst buffer has its designed capacity and performance, which can hardly adapt to any demand changes. In order to solve such problems, we extend our CloudBB to the HPC environments with high performance network supports and propose a software- and user-level on-demand burst buffer system, HuronFS. Similar to the CloudBB, HuronFS is designed as a software burst buffer utilizing compute nodes with high performance network to buffer the I/O data, hence HuronFS can be easily deployed on any HPC centers. Furthermore, thanks to the Master-Worker and Key-Value architecture, the capacity and the performance of HuronFS can be adapted to the workload. We evaluated our HuronFS on two supercomputer systems and demonstrated that using HuronFS can achieve the performance of state-of-the-art parallel file systems and help users to improve the I/O performance in the HPC centers.

Utilizing the burst buffer systems has been proven to accelerate the I/O performance. To use a burst buffer, the users need to specify the configurations of burst buffer in the job scripts, i.e. buffer size. However, the configurations must be carefully determined to avoid experiencing poor performance or causing the low job throughput due to under-utilization.

To understand the performance impacts from different burst buffer configurations, we conduct a performance analysis with a trace-driven simulator. We collect I/O trace from real applications and simulate their performance under different configurations. We explore three different configurations aspects, the swap-in/out granularity; different buffer size; and different data replacement algorithms. The simulation results show that different applications have different requirements for burst buffer configurations and help to further optimize the performance using burst buffer.

Our studies of design optimization and evaluation of burst buffer in both HPC centers and Cloud helps to alleviate the performance gap between the computation and storage systems, helps to further understand and optimize the burst buffer performance to support the future large-scale data-intensive applications.