2017 IEEE 37th International Conference on Distributed Computing Systems

# DuoFS: An Attempt at Energy-Saving and Retaining Reliability of Storage Systems

Bing Jiao¶, Xiaomin Zhu†, Xiaojun Ruan‡, Xiao Qin§, and Shu Yin*‖,
¶Hunan University, China, Email: a.i.jiaobing@gmail.com
†National University of Defense Technology, China, Email: xmzhu@nudt.edu.cn
‡West Chester University of Pennsylvania, PA 19383, Email: xruan@wcupa.edu
§Auburn University, AL 36849, Email: xqin@auburn.edu
‖ShanghaiTech University, Shanghai, China, Email: yinshu@shanghaitech.edu.cn
*Corresponding Author

***Abstract*—As issues of the Energy Wall and the Reliability Wall become unavoidable, it is a demanding and challenging task to reduce energy consumption in large-scale storage systems in modern data centres while retaining acceptable systems reliability. Most energy conservation techniques inevitably have adverse impacts on the parallel disk systems. To address the reliability issues of energy-efficient parallel storage systems, we propose a reliable energy-efficient storage system called DuoFS, which aims at improving both energy efficiency and reliability of parallel storage systems by seamlessly integrating HDDs and SSDs. With the help of the middleware layer, DuoFS can distribute popular data to SSD-based nodes and put HDD-based nodes into the low-power mode under light workload conditions without modification of the parallel systems.**

***Index Terms*—hybrid parallel storage system, PVFS, PLFS, energy- efficient, reliability**

## I. Introduction

Traditional energy-saving techniques in disk array systems do not take into account adverse impacts of energy conservation mechanisms on system reliability. Existing solutions tend to frequently spin up and down hard disk drives (HDDs) or replace HDDs with solid state disks (SSDs) [1]. These solutions may cause severe mechanical malfunction of HDDs and lead to permanent data loss. Moreover, heavy write I/O loads shorten the lifetime of SSDs due to the limited numbers of erasure cycles of flash memory [2]. We propose a hybrid parallel storage system call DuoFS aiming to improve reliability and energy-efficiency of storage systems by introducing a middleware I/O layer so as to apply the hybrid disk technique without modification of existing file systems. At the heart of DuoFS is the hybrid storage technique that exploits flash-based storage devices low power consumption as well as tradition disk-based storage appliances large capacity and long lifetime to improve energy efficiency and reliability of parallel storage systems under fluctuate I/O workloads.

The benefits of DuoFS are as follows:

- DuoFS offers a better balance among energy consumption, reliability, and hardware budget under a fluctuating system workload;
- As DuoFS re-organizes the I/O workload using a middleware layer, it can be implemented without additional kernel modification of existing file systems.
- Since DuoFS mounts multiple file systems underneath, it makes the storage architecture expandable as the systems go up-scaling.

## II. DuoFS: A Hybrid Storage System

### A. Framework

The overview framework of DuoFS is expressed in Fig. 1. The DuoFS consists of four major components– a Data Marker, an I/O Re-organizer, a File System Selector, and underlying file systems. The I/O access pattern will be firstly analyzed to identify the data popularly so that the data will be categorized into two: Hot and Cold data, indicating the frequently accessed data and less frequently accessed data. The Hot/Cold Data Marker attaches the popularity information to the data. By reading the popularity information, the lower I/O Re-organizer will rearrange the I/O into two groups– hot and cold– accordingly. The File System Selector uses the popularity information to determine the destination file system underlying– hot data on the File System-Hot, which mounts flash-based storage nodes while cold data on the File System-Cold, which mounts hard disk storage nodes. The DuoFS is designed especially to fit the write-once-read-multiple-times scenarios, where frequently read data is cached to File System-Hot to achieve better performance while the rest of data is stored on File System-Cold and is pushed to idle mode to saving energy.

We develop a proof-of-concept prototype of DuoFS using PLFS [3] on the parallel virtual file system (PVFS). We choose two backends via PLFS: one serves as the directory holding hot data and is mounted to the File System-Hot while the other holds cold data and is mounted to the File System-Cold.

### B. Preliminary Results under FIO Tests

Fig. 2 illustrates the IOPS tests of the DuoFS using the FIO benchmark in addition to the energy consumption evaluations. Noted that the "SSD-Only" indicates the situation that only SSD-based nodes are involved while the HDD-based nodes are managed to low-power mode by the File System-Cold. The purpose of presenting the "SSD-Only" is to provide a full flash-based storage case. Our argument is that the

1063-6927/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDCS.2017.79



IEEE computer society

Authorized licensed use limited to: Auburn University. Downloaded on November 17,2026 at 23:04:37 UTC from IEEE Xplore. Restrictions apply.

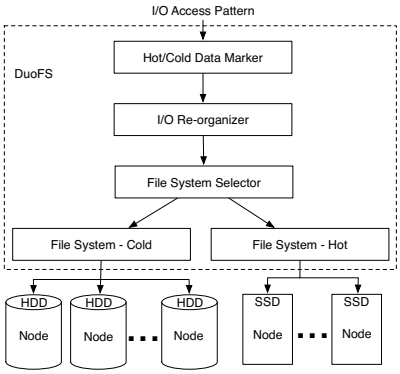


Fig. 1: An Overview of DuoFS Framework

hybrid storage mechanisms such as DuoFS balance the price-performance ratio better due to the significant price different between SSDs and HDDs.

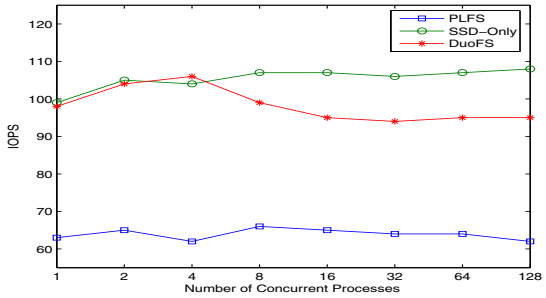


Fig. 2: IOPS Comparison under FIO Testing

Figs. 3 plots the energy consumption of PLFS and DuoFS under the FIO benchmark. We can observe that when the number of processes is smaller than 32, the DuoFS can save more than 60% of the energy compare to the traditional PLFS.

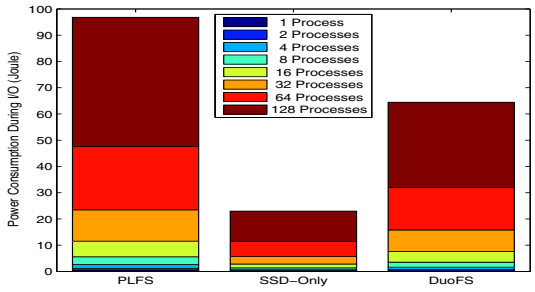


Fig. 3: Energy Consumption Comparison under FIO Testing

We apply MREED [4] to quantify the system reliability of our proposed DuoFS. Compared to the energy consumption results in the aforementioned section, the DuoFS only trades nearly 5% of the reliability for more than 60% energy conservation (Fig. 4).

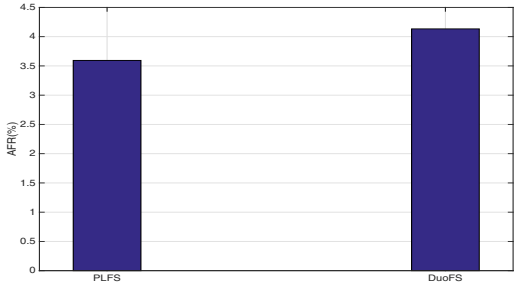


Fig. 4: Reliability Comparison between PLFS and DuoFS

## III. Conclusion

DuoFS presents a hybrid idea of applying two separated file systems (File System-Hot and File System- Cold) that mount different types of storage nodes. File System-Hot mounts SSD nodes to achieve better I/O performance and active power consumption where File System-Cold mounts HDD nodes to gain larger capacity and longer lifetime. Our ongoing work on this project is performed in the following few aspects. Firstly, a hot/cold data identifier should be carefully designed and implemented. Second, the power management policy should be more sophisticated. Thirdly, we intend to address the write issue carefully, to cover more I/O scenarios such as fault tolerance using checkpointing.

## Acknowledgment

Shu Yin's research is supported by the National Natural Science Foundation of China under Grant 61402158, the China Postdoctoral Science Foundation under Grant 2015M572708, and ShanghaiTech University under a start-up grant. Xiaomin Zhu's work is supported by the National Natural Science Foundation of China under Grant 61572511, and the Scientific Research Project of National University of Defense Technology under Grant ZL16-03-09. Xiao Qin's work is supported by the U.S. National Science Foundation under Grant CCF-0845257 (CAREER), CNS- 0917137 (CSR), and DUE- 0830831 (SFS).

## References

[1] S. He, X. H. Sun, and A. Haider, "Has: Heterogeneity-aware selective data layout scheme for parallel file systems on hybrid servers," in *2015 IEEE International Parallel and Distributed Processing Symposium*, May 2015, pp. 613–622.

[2] M. Maas, T. Harris, K. Asanovic, and J. Kubiatowicz, "Trash day: Coordinating garbage collection in distributed systems," in *Proceedings of the 15th USENIX Conference on Hot Topics in Operating Systems*, ser. HOTOS'15. Berkeley, CA, USA: USENIX Association, 2015, pp. 1–1. [Online]. Available: http://dl.acm.org/citation.cfm?id=2831090.2831091

[3] J. Bent, G. Gibson, G. Grider, B. McClelland, P. Nowoczynski, J. Nunez, M. Polte, and M. Wingate, "Plfs: A checkpoint filesystem for parallel applications," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, ser. SC '09. New York, NY, USA: ACM, 2009, pp. 21:1–21:12. [Online]. Available: http://doi.acm.org/10.1145/1654059.1654081

[4] S. Yin, X. Li, K. Li, J. Huang, X. Ruan, X. Zhu, W. Cao, and X. Qin, "Reed: A reliable energy-efficient raid," in *2015 44th International Conference on Parallel Processing*, Sept 2015, pp. 649–658.



Authorized licensed use limited to: Auburn University. Downloaded on November 17,2026 at 23:04:37 UTC from IEEE Xplore. Restrictions apply.