Applied Mechanics and Materials Vols. 644-650 (2014) pp 1620-1624
Online available since 2014/Sep/22 at www.scientific.net

doi:10.4028/www.scientific.net/AMM.644-650.1620

Submitted: 18.07.2014
Accepted: 19.07.2014

A statistics-based data placement strategy for hybrid storage

Yuanhua Yang[1,a], Xianbin Xu[1,b], Shuibing He[1,c], Yuhua Wen[2,d]

[1]School of Computer, Wuhan University, Wuhan, 430072, China

[2]Department of Physics, Xiamen University, Xiamen 361005, China

[a]yangyuanhua123@163.com, [b]xbxu@whu.edu.cn, [c]Shuibing He, [d]yhwen@xmu.edu.cn

Keywords: hybrid storage, solid state disk, hard disk drive, data placement

Abstract. Although hard disk drives have been popular over several decades, there still exists the deficiency because of their slow speeds and high power consumptions. By contrast, flash-based solid state disks exhibit good performance and low power consumption. However, the limited lifetimes become a fatal flaw of solid state disks. In order to take full advantage of hard disk drives and solid state disks, we design a hybrid storage system to make them work in a complementary manner. Further, we propose a data placement scheme for this system to determine the data placement on the underlying solid state disks or hard disk drives based on the data access statistics. Experiment results show that the lifetime of solid state disks and the response time of the system can be significantly improved compared with the alone storage media.

I. Introduction

Although hard disk drives (HDDs) has been fashionable in personal computers for several decades, it still has a lot of shortcomings. First, the high energy consumption makes it difficult to improve the Revolutions Per Minute. Second, its storage density is difficult to be increased in some requirements. Third, the resistance is very bad due to its mechanical movement. More importantly, its slow response time of data accessing greatly decreases the performance of the storage system.

Besides HDDs, other technologies have also been developed to provide data storage, such as phase-change memory (PCM), STT-RAM, and NAND-flash. Especially, the NAND-flash based solid state drives (SSDs) has many advantages compared with HDDs. For example, SSDs have good resistance, low power consumption, and fast response time to accessing data. These advantages make SSDs widely applied in personal computer and many embedded devices. However, the characteristic of erasure-before-write of SSDs significantly degrades their reliability. In addition, the cost of proven technique of HDDs is lower than SSDs. Table 1 lists some properties of SSDs and HDDs.

Table 1: Lifetime, performance and cost comparison [1].

Media	Lifetime	Access time(us)	Cost($/GB)
HDDs	MTTF=1.2Mhr	<5500	3
SDDs	10K-1M Erase Cycles	Read<45, Write<200	25

As can be seen from Table 1, HDDs and SDDs have their respective advantages in performance, lifetime, and cost. First, the erasure number of each block in SSDs is typical about 10K-1M, which makes SSDs an unreliable device. Therefore, the lifetime of SSDs is restricted because of the limited times of block's erasure. Comparing with the MTTF of HDDs, SSD's lifetime is very shot. Second, the access speeds of SSDs is much faster than that of HDDs. Third, it is evident that the cost/GB of HDDs is cheaper than that of SDDs. How to balance the configuration of HDDs and SSDs is considerably important for the economic object. Finally, unlike traditional HDDS, write operations in SSDs are performed as out-of-place updates, which must handle the garbage collection (GC) and can severely degrades SSD's performance.

From the aforementioned analysis, we can know that SSDs and HDDs have their merits and shortcomings, respectively. But their properties can still complement such as accessing speed, cost,