Article

International Journal of
HIGH PERFORMANCE
COMPUTING APPLICATIONS

The International Journal of High Performance Computing Applications
1–15

DOI: 10.1177/1094342016631610
hpc.sagepub.com
SAGE

Enhancing hybrid parallel file system through performance and space-aware data layout

Shuibing He[1,2], Yan Liu[3], Yang Wang[4], Xian-He Sun[5] and Chuanhe Huang[1]

Abstract

Hybrid parallel file systems (PFSs), which consist of solid-state drive servers (SServer) and hard disk drive servers (HServer), have recently attracted growing attention. Compared to a traditional HServer, an SServer consistently provides improved storage performance but lacks storage space. However, most current data layout schemes do not consider the differences in performance and space between heterogeneous servers and may significantly degrade the performance of the hybrid PFSs. In this article, we propose performance and space-aware (PSA) scheme, a novel data layout scheme, which maximizes the hybrid PFSs' performance by applying adaptive varied-size file stripes. PSA dispatches data on heterogeneous file servers not only based on storage performance but also storage space. We have implemented PSA within OrangeFS, a popular PFS in the high-performance computing domain. Our extensive experiments with representative benchmarks, including IOR, HPIO, MPI-TILE-IO, and BTIO, show that PSA provides superior I/O throughput than the default and performance-aware file data layout schemes.

Keywords

Parallel I/O system, parallel file system, data layout, solid-state drive, hybrid I/O system

1. Introduction

Many large-scale applications in science and engineering have become more and more data intensive (Kandemir et al., 2008). For example, Table 1 shows the data requirements of a few representative applications at Argonne National Laboratory in 2012 (Latham et al., 2013). The generated data of these applications reach several terabytes per year. Such large data requirements are putting unprecedented pressure on computer input/output (I/O) systems to store data effectively. In the meanwhile, storage devices have a slower performance improvement than central processing units during the past three decades. While processor speeds have increased nearly by 50% each year, the access latency of a single hard disk drive (HDD) has only reduced by roughly 7% (Hennessy and Patterson, 2011). As a result, I/O system has become the major performance bottleneck for many applications in high-performance computing (HPC) domain.

To accommodate growing volumes of data, parallel file systems (PFS), such as PVFS (Carns et al., 2000), OrangeFS (Orange File System), Lustre (Microsystems, 2007), and GPFS (Schmuck and Haskin, 2002), have been developed to achieve high aggregate I/O throughput by leveraging the parallelism of multiple HDD-based file servers. However, due to the diversity of data access patterns, it is not always enough to improve I/O performance by simply adding more storage servers. For example, in the face of noncontiguous small requests, PFSs still perform poorly (He et al., 2014b) because of the low degree of server parallelism as well as the frequent disk head movements on each server due to random accesses (Song et al., 2011b). Hence, fully utilizing the underlying file servers is still a challenging task.

[1]State Key Laboratory of Software Engineering Computer School, Wuhan University, Wuhan, Hubei, China
[2]State Key Laboratory of High Performance Computing National University of Defense Technology, Changsha, Hunan, China
[3]College of Computer Science and Electronic Engineering, Hunan University, China
[4]Shenzhen Institute of Advanced Technology, Chinese Academy of Science, Shenzhen, China
[5]Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA

Corresponding author:
Shuibing He, State Key Laboratory of Software Engineering Computer School, Wuhan University, Wuhan, Hubei, China.
Email: heshuibing@gmail.com