2-8-2015

CSR: Small: Collaborative Research: SANE: Semantic-Aware Namespace in Exascale File Systems

Yifeng Zhu
Principal Investigator; University of Maine, Orono, zhu@eece.maine.edu

RSR Award Detail

Research Spending & Results

Award Detail

Awardee:	UNIVERSITY OF MAINE SYSTEM
Doing Business As Name:	University of Maine
PD/PI:	Yifeng Zhu (207) 581-2499 zhu@eece.maine.edu
Award Date:	08/20/2011
Estimated Total Award Amount:	$ 190,947
Funds Obligated to Date:	$ 190,947 FY 2011=$190,947
Start Date:	09/01/2011
End Date:	08/31/2014
Transaction Type:	Grant
Agency:	NSF
Awarding Agency Code:	4900
Funding Agency Code:	4900
CFDA Number:	47.070
Primary Program Source:	040100 NSF RESEARCH & RELATED ACTIVIT
Award Title or Description:	CSR: Small: Collaborative Research: SANE: Semantic-Aware Namespace in Exascale File Systems
Federal Award ID Number:	1117032
DUNS ID:	186875787
Parent DUNS ID:	071750426
Program:	COMPUTER SYSTEMS
Program Officer:	M. Mimi McClure (703) 292-5197 mmcclure@nsf.gov

Awardee Location

Street:	5717 Corbett Hall
City:	ORONO
State:	ME
ZIP:	04469-5717
County:	Orono
Country:	US
Awardee Cong. District:	02

Primary Place of Performance

Organization Name:	University of Maine
Street:	5717 Corbett Hall
City:	ORONO
State:	ME
ZIP:	04469-5717
County:	Orono
Country:	US
Cong. District:	02

Abstract at Time of Award

Explosive growth in volume and complexity of data exacerbates the key challenge facing the management of massive data in a way that fundamentally improves the ease and efficacy of their usage. Exascale storage systems in general rely on hierarchically structured namespace that leads to severe performance bottlenecks and makes it hard to support real-time queries on multi-dimensional attributes. Thus, existing storage systems, characterized by the hierarchical directory tree structure, are not scalable in light of the explosive growth in both the volume and the complexity of data. As a result, directory-tree based hierarchical namespace has become restrictive, difficult to use, and limited in scalability for today's large-scale file systems.

This project investigates a novel semantic-aware namespace scheme to provide dynamic and adaptive namespace management and support typical file-based operations in Exascale file systems. The project leverages semantic correlations among files and exploits the evolution of metadata attributes to support customized namespace management, with the end goal of efficiently facilitating file identification and end users data lookup. This project provides significant performance improvements for existing file systems in Exascale file systems. Since Exascale file systems constitute one of the backbones of the high-performance computing infrastructure, the semantic-aware techniques also benefits a great number of scientific and engineering data-intensive applications. This project strengthens the ongoing development of high performance computing infrastructures at both UNL and UMaine. The project enhances undergraduate and graduate education at both participating institutions and outreach to K-12 in UMaine via an ongoing NSF-funded ITEST program.

Publications Produced as a Result of this Research

Y. Hua, J. Hong, Y. Zhu, D. Feng, X. Lei "SANE: Semantic-Aware Namespace in Ultra-large-scale File Systems" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, v.25, 2014, p.1328-1338

Y. Deng, Y. Hu, X. Meng, Y. Zhu "Predictively booting nodes to minimize performance degradation of a power-aware web cluster" JOURNAL OF CLUSTER COMPUTING, v.17, 2014, p.1309

M. Yu, J. Wan, Y. Zhu, C. Xie "A New Parity-Based Migration Method to Expand RAID-5" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, v.1, 2013, p.

Hui Tian, Hong Jiang, Ke Zhou, Dan Feng "Transparency-orientated Encoding Strategies for Voice-over-IP Steganography" THE COMPUTER JOURNAL, v.55, 2012, p.702

Bo Mao, Hong Jiang, Lei Tian, Suzhen Wu, Dan Feng, Jianxi Chen, and Lingfang Zeng "HPDA: A Hybrid Parity-based Disk Array for Enhanced Performance and Reliability" ACM TRANSACTIONS ON STORAGE, v.8, 2012, p.1

Bo Mao, Hong Jiang, Suzhen Wu, Yinjin Fu and Lei Tian "SAR: SSD Assisted Restore Optimization for Deduplication-based Storage Systems in the Cloud" 7TH IEEE INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGE (NAS'12), v. , 2012, p.

Dongyuan Zhan, Hong Jiang, Sharad Seth "Locality & Utility Co-optimization for Practical Capacity Management of Shared Last Level Caches" PROCEEDINGS OF THE 26TH INTERNATIONAL CONFERENCE ON SUPERCOMPUTING (ICS'12), v. , 2012, p.

J. Wang, R. Hua, Y. Zhu, C. Xie, P. Wang and W. Gong "C-IRR: An Adaptive Engine for Cloud Storage Provisioning Determined by Economic Models with Workload Burstiness Consideration" PROCEEDINGS OF 2012 INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGES (NAS'12), v. , 2012, p.

J. Yue and Y. Zhu "Making Write Less Blocking for Read Accesses in Phase Change Memory" PROCEEDINGS OF THE 20TH IEEE INTERNATIONAL SYMPOSIUM ON MODELLING, ANALYSIS, AND SIMULATION OF COMPUTER AND TELECOMMUNICATION SYSTEMS (MASCOTS'12), v. , 2012, p.

J. Wang, R. Hua, Y. Zhu, J. Wan, C. Xie and Y. Chen "RO-BURST: A Robust Virtualization Cost Model for Workload Consolidation over Clouds" PROCEEDINGS OF 2012 IEEE/ACM INTERNATIONAL SYMPOSIUM ON CLUSTER, CLOUD AND GRID COMPUTING (CCGRID'12), v. , 2012, p.

Jian Hu, Hong Jiang, Lei Tian and Lei Xu "GC-ARM: Garbage Collection-Aware RAM Management for Flash based Solid State Drives" PROCEEDINGS OF THE 7TH IEEE INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGE (NAS'12), v. , 2012, p.

Jian Hu, Hong Jiang and Prakash "Understanding Performance Anomalies of SSDs and Their Impact in Enterprise Application Environment" PROCEEDINGS OF THE 12TH JOINT ACM SIGMETRICS/PERFORMANCE CONFERENCE (SIGMETRICS¿??2012, POSTER), v. , 2012, p.

Yu Hua, Hong Jiang, Yifeng Zhu, Dan Feng, and Lei Tian "Semantic-Aware Metadata Organization Paradigm in Next-Generation File Systems" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, v.23, 2012, p.337

Zhichao Yan, Hong Jiang, Dan Feng, Lei Tian and Yujuan Tan "SUV: A Novel Single-Update Version-Management Scheme for Hardware Transactional Memory Systems" PROCEEDINGS OF THE 26TH IEEE INTERNATIONAL PARALLEL & DISTRIBUTED PROCESSING SYMPOSIUM (IPDPS¿??12), v. , 2012, p.

Wen Xia, Hong Jiang, Dan Feng, Lei Tian, Min Fu and Zhongtao Wang "P-Dedupe: Exploiting Parallelism in Data Deduplication System" PROCEEDINGS OF THE 7TH IEEE INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGE (NAS'12), v. , 2012, p.

L. Lin, Y. Zhu, J. Yue, Z. Cai and B. Segee "Hot Random Off-loading: A Hybrid Storage System With Dynamic Data Migration" PROCEEDINGS OF THE 19TH ANNUAL MEETING OF THE IEEE INTERNATIONAL SYMPOSIUM ON MODELING, ANALYSIS AND SIMULATION OF COMPUTER AND TELECOMMUNICATION SYSTEMS (MASCOSTS'11), v. , 2011, p.

J. Yue, Y. Zhu, Z. Cai, and L. Lin "Energy Efficient Buffer Cache Replacement for Data Servers" PROCEEDINGS OF THE 2011 INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGE (NAS 2011), v. , 2011, p.

Project Outcomes Report

Disclaimer

The project aimed to improve the performance and energy efficiency of I/O operations of large-scale cluster computing platforms.

A challenging issue in performance evaluation of parallel storage systems through trace-driven simulation is to accurately characterize and emulate I/O behaviors in real applications. The correlation study of inter-arrival times between I/O requests, with an emphasis on I/O intensive scientific applications, shows the necessity to further study the self-similarity of parallel I/O arrivals. We have analyzed several I/O traces collected in large-scale supercomputers and concluded that parallel I/Os exhibit statistically self-similar like behaviors. We have implement a memory access series generator in which the inputs are the measured properties of the available memory trace series. Experimental results show that this model can faithfully capture the complex access arrival characteristics of memory workloads, particularly the heavy-tail characteristics under both Gaussian and non-Gaussian workloads.

Another challenging issue in large-scale cluster servers is the power consumption. Prior studies have shown most memory space on data servers is used for buffer caching and thus cache replacement becomes critical. Temporally concentrating memory accesses to a smaller set of memory chips increases the chances of free riding through DMA overlapping and also enlarges the opportunities for other ranks to power down. We have designed a new power and thermal-aware buffer cache replacement algorithm. Our simulation results based real-world TPC-R I/O trace show that our algorithm can save up to 12.2% energy with marginal degradation in the cache performance. In addition, row accesses in memory chips are not only very slow in response but also cost significant amount of energy. The interleaved access from different process segments destroys access locality seen at process segment. To address this, we design a new memory architecture that adds a small cache in memory controller to recover accesses locality and a new cache management scheme that exploits the semantic information of memory access requests to better capture the access locality.

To increase the performance of I/O systems, we have designed and implemented a hybrid storage system that dynamically allocates or migrates data between SSD and hard disks in order to achieve the optimal performance gain. We designed a hybrid storage architecture that treats the SSD as a by-passable cache to hard disks, and developed an online algorithm that judiciously exchanges data between the SSD and the disks. Our basic principle is to place hot and randomly accessed data on the SSD, and other data, particularly cold and sequentially accessed data on hard disks. Our hybrid storage system, called Hot Random Off-loading (HRO), is implemented as a simple and general user-level layer above conventional file systems in Linux and supports standard POSTIX interfaces, thus requiring no modifications to underneath file systems or users applications. This prototype is comprehensively evaluated by using a commodity hard disk and SSD.

This project helps improve the energy-efficiency of cluster computers and promote green computing. Large-scale cluster computers consume large amounts of electrical power. Prior research founds storage systems and computer memory consumes a significant portion of electrical power. This research helps improve the energy efficiency of memory and storage systems by incorporating energy-aware buffer cache replacement algorithms and by integrating energy-efficient SSD with conventional hard disk drives.

We have published more than ten research papers in top conferences and journals to disseminate our research results and findings. In addition, this project has successfully trained two Ph.D. students. This project has also trained two Ph.D. students and 1 M.S. student at the University of Maine.

This project has also help integrate the research result into our curriculum. New concepts of self-similarity workload modeling, energy-aware buffer cache management, and hybrid storage systems have been incorporated into one undergraduate and graduate level courses.

Last Modified: 02/08/2015
Modified by: Yifeng Zhu

For specific questions or comments about this information including the NSF Project Outcomes Report, contact us.

My Desktop
Prepare & Submit Proposals
Proposal Status
Proposal Functions
Awards & Reporting
Notifications & Requests
Project Reports
Submit Images/Videos
Award Functions
Manage Financials
Program Income Reporting
Federal Financial Report History
Financial Functions
Grantee Cash Management Section Contacts
Administration
User Management
Research Administration
Lookup NSF ID

Preview of Award 1117032 - Final Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1117032
Project Title:	CSR: Small: Collaborative Research: SANE: Semantic-Aware Namespace in Exascale File Systems
PD/PI Name:	Yifeng Zhu, Principal Investigator
Recipient Organization:	University of Maine
Project/Grant Period:	09/01/2011 - 08/31/2014
Reporting Period:	09/01/2013 - 08/31/2014
Submitting Official (if other than PD\PI):	Yifeng Zhu Principal Investigator
Submission Date:	02/08/2015
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	Yifeng Zhu

Accomplishments

*** What are the major goals of the project?**

The objective of the project is to provide dynamic and adaptive namespace management and support typical file based operations in Exascale file systems. The project plans to leverage semantic correlations among files and exploit the evolution of metadata attributes to support customized namespace management, with the end goal of efficiently facilitating file identification and end users data lookup.

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities: We have conducted research work on large-scale namespace management and metadata management. The major activities are designing, implementing, and evaluating semantic-aware namespace and metadata management schemes to improve performance and searchability in Exascale file systems. Keeping semantic-awareness and application awareness in mind, we have proposed a number of novel approaches from different angles, including per-file flat and small namespace (SANE), locality aware data aggregation (ANTELOPE), membership representation for large data sets (DBA), scalable searchable file systems (VSFS), Parity-based Migration in expansion of RAID (PBM), Two-Stage Write for new emerging storage-class memory, and Active Journaling (AJ).

Specific Objectives: The specific objectives of this project are to: (1) demonstrate the existence of application semantics and necessity of utilizing semantics in largescale namespace and metadata management; (2) propose distributed and scalable namespace and metadata management approaches; (3) implement and evaluate proofofconcept prototype systems; (4) investigate how to speed up I/O performance, particular metadata operations, by using active journaling; (5) investigate how to deploy emerging memory devices, such as Phase-Change Memory (PCM) to build storage-class memory.

Significant Results: We showed that the proposed schemes including SANE, ANTELOPE, DBA, and AJ significantly advance the stateoftheart namespace and metadata management for ultralarge file systems. In order to support the proposed schemes, a number of key techniques are proposed and developed. The proposed schemes and key techniques enable file system developers, domain experts, and applications to fast and accurately lookup files in very largescale file sets. A versatile searchable file system, VSFS, is developed to provide a transparent, accurate and realtime file-search service through a POSIXcompatible file system namespace. In addition, we have designed a new writing strategy to accelerate the write performance, which cause performance bottleneck since write operations are slow and they block all pending read operations.

Key outcomes or Other achievements: Our key research findings have been published or accepted by top computer conference and journals. Our HPCA'13 paper was one of the four candidates for Best Paper Award. At the UMaine site, we have trained one postdoc research fellow, and two PhD students.

*** What opportunities for training and professional development has the project provided?**

At UMaine, this collaborative project is training two Ph.D. students and one post-doctoral research fellow.

*** How have the results been disseminated to communities of interest?**

We have published multiple papers in top conferences and journals, including DATE'13, HPCA'13, IEEE TPDS, IEEE TC, Supercomputing'12, Cluster'12, NAS'12, IPCCC'12, MASCOTS'12, and CCGrid'12. Our HPCA'13 paper was one of the four candidates for Best Paper Award.

Products

Books

Book Chapters

Conference Papers and Presentations

Inventions

Journals

M. Yu, J. Wan, Y. Zhu, C. Xie (2013). A New Parity-Based Migration Method to Expand RAID-5. *IEEE Transactions on Parallel and Distributed Systems*. 1 . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Y. Deng, Y. Hu, X. Meng, Y. Zhu (2014). Predictively booting nodes to minimize performance degradation of a power-aware web cluster. *Journal of Cluster Computing*. 17 (4), 1309. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Y. Hua, J. Hong, Y. Zhu, D. Feng, X. Lei (2014). SANE: Semantic-Aware Namespace in Ultra-large-scale File Systems. *IEEE Transactions on Parallel and Distributed Systems*. 25 (5), 1328-1338. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Licenses

Other Products

Other Publications

Patents

Technologies or Techniques

Thesis/Dissertations

Websites

Participants/Organizations

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Zhu, Yifeng	PD/PI	1
Yue, Jianhui	Postdoctoral (scholar, fellow or other postdoctoral position)	6
Cai, Zhao	Graduate Student (research assistant)	9
Rahman, Tania	Graduate Student (research assistant)	6

Full details of individuals who have worked on the project:

Yifeng Zhu
Email: zhu@eece.maine.edu

Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Supervised graduate students and post-doc research fellows who worked on this project. Integrate research into courses PI taught.

Funding Support: None

International Collaboration: Yes, China
International Travel: Yes, China - 0 years, 0 months, 14 days; France - 0 years, 0 months, 7 days

---

Jianhui Yue
Email: jianhui.yue@maine.edu
Most Senior Project Role: Postdoctoral (scholar, fellow or other postdoctoral position)
Nearest Person Month Worked: 6

Contribution to the Project: Dr. Yue is working on storage-class memory systems to accelerate the I/O operations.

Funding Support: Dr. Yue is supported by this grant.

International Collaboration: No
International Travel: No

Zhao Cai
Email: cai.zhao@maine.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 9

Contribution to the Project: Mr. Cai is working on the active journaling.

Funding Support: Mr. Cai is partially supported by this grant.

International Collaboration: No
International Travel: No

Tania Rahman
Email: tania.rahman@maine.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 6

Contribution to the Project: Mr. Rahman is working on (1) active journaling, and (2) leveraging the emerging new storage-class memory devices to speed up the I/O operations, particularly the metadata performance.

Funding Support: Mrs. Rahman is supported by this grant.

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?

Name	Type of Partner Organization	Location
Huazhong University of Science and Technology	Academic Institution	Wuhan, China

Full details of organizations that have been involved as partners:

Huazhong University of Science and Technology

Organization Type: Academic Institution
Organization Location: Wuhan, China

Partner's Contribution to the Project:
Collaborative Research

More Detail on Partner and Contribution: The PI is collaborating with them on metadata management and has published several papers with them as a co-author.

What other collaborators or contacts have been involved?

NO

Impacts

What is the impact on the development of the principal discipline(s) of the project?

We have known the necessity and usefulness of exploitation of application semantics in designing namespace and metadata management schemes for largescale file systems.

(1) To overcome the limitations of the traditional treebased namespace scheme, SANE introduces a new naming methodology based on the notion of semanticaware perfile namespace, which exploits semantic correlations among files, to dynamically aggregate correlated files into small, flat but readily manageable groups to achieve fast and accurate lookups. The semantic correlations and file groups identified in SANE can also be used to facilitate file prefetching and data deduplication, among other systemlevel optimizations. Extensive tracedriven experiments on our prototype implementation validate the efficacy and efficiency of SANE.

(2) In order to bridge the gap between network architecture and data placement in largescale cloud data centers. ANTELOPE is proposed to leverage offline precomputation to improve online query performance. The main contributions of ANTELOPE include: (1). ANTELOPE uses a semanticaware partial materialization as a suitable tradeoff between the construction efficiency and the query accuracy by precomputing the related, rather than all, aggregates; (2). ANTELOPE aggregates data with strong locality into the same or adjacent servers, thus reduces network traffic through highlevel switches and meanwhile producing no decrements upon the quality of cloud services; (3) ANTELOPE makes use of an LSH scheme to efficiently identify data locality with acceptable complexity such that data with strong locality can be aggregated and placed in the same or adjacent servers.

(3) The existing filesearch solutions for HPC analytics are either poorly scalable for largescale systems, or lack a well-integrated interface to allow applications to easily use them for critical tasks. A Versatile Searchable File System, VSFS, which provides a transparent, accurate and realtime filesearch service through a POSIXcompatible file system namespace that can be integrated into any HPC/Big Data legacy code without modifications. Additionally, to support realtime file search, VSFS uses a DRAMbased distributed architecture to perform realtime file indexing. Moreover, a versatile index scheme is designed to adapt to the various forms of HPC datasets. The results of our VSFS prototype evaluation show that VSFS is scalable in a typical HPC environment. It achieves significantly better fileindexing and filesearch performance than the popular SQL/NoSQL solutions, while it only introduces negligible I/O overhead.

(4) We have also studied how to exploit new and emerging storage and nonvolatile memory devices to speed up the metadata performance. One of the key challenges of these new devices, such as phase-change memory (PCM), is their inferior write performance. To improve the write performance of PCM, we have proposed a new write scheme, called two-stage-write, which leverages the speed and power difference between writing a zero bit and writing a one bit. Writing a one takes longer time but less electrical current than writing a zero. We propose to divide a write into stages: in the write-0 stage all zeros are written at an accelerated speed, and in the write-1 stage, all ones are written with increased parallelism, without violating power constraints. We also present a new coding scheme to improve the speed of the write-1 stage by further increasing the number of bits that can be written to PCM in parallel.

(5) Although journaling reduces the recovery time after a system crash, this technique suffers from write-twice overhead, particularly for metadata. At running time, updates are saved in the journal temporarily and written to original locations later. This write-twice overhead degrades the system performance to a considerable degree. We propose a new journaling technique, called Active Journaling (AJ), which allows the journal accessible during running time, exploits the concealed replication functionality of journaling for disk layout tuning techniques, offsets the write twice overhead, and makes journaling overall a performance boost instead of a continuous performance drag. We have implemented Active Journaling in Linux. Using the non-journaled file system ext2 as the baseline, our experiments show that AJ improves overall performance up to 35.37 ± 9.48%. Compared with ext3 journaling file system, AJ enhances overall performance up to 29.84±5.63% (ordered mode), 29.82±4.44% (writeback mode), and 39.89 ± 7.43% (journal mode).

(6) Random accesses are generally harmful to performance in hard disk drives due to more dramatic mechanical movement. We design, implement, and evaluate Hot Random Off-loading (HRO), a self-optimizing hybrid storage system that uses a fast and small SSD as a by- passable cache to hard disks, with a goal to serve a majority of random I/O accesses from the fast SSD. HRO dynamically estimates the performance benefits based on history access patterns, especially the randomness and the hotness, of individual files, and then uses a 0-1 knapsack model to allocate or migrate files between the hard disks and the SSD. HRO can effectively identify files that are more frequently and randomly accessed and place these files on the SSD. We implement a prototype of HRO in Linux and our implementation is transparent to the rest of the storage stack, including applications and file systems. We evaluate its performance by directly replaying three real-world traces on our prototype. Experiments demonstrate that HRO improves the overall I/O throughput up to 39% and the latency up to 23%.

(6) Areal density scaling in magnetic hard drives is in jeopardy as magnetic particles become unstable when they are sufficiently small. Shingled recording holds great promise to mitigate the problem of density scaling cost-effectively by overlapping data tracks. However, this innovative technology suffers severely from slow small writes since a small write needs to read adjacent tracks and then write them back in a special order to eliminate inter-track interference. This prevents shingle recording from being widely adapted in practice.

We have proposed and implemented hybrid wave-like shingled recording disk system (HWSR) to improve both the performance and the capacity of a shingled recording disk. HWSR containsthree different storage media: memory, SSD, and hard disk. The memory has a very small capacity to reduce the overall cost. It is used to buffer hot writes. The SSD is used as a disk cache to improve random read performance. HWSR consists of three key components: (1) a new data layout based on segmentation for shingled recording disk to reduce random write amplification; (2) a new shingled track layout named wave-like shingled recording (WSR) to further improve its capacity; (3) a new replacement policy based on least write amplification that effectively reduces the miss rate and data rewritten amount.

Experimental evaluation on our prototype evaluation under a variety of I/O intensive workloads show that HWSR system improves the performance of small writes significantly. Results show that our design is not sensitive to many design parameters such as the block size and the segment size. While our new data layout slightly increase the average seek time as sequential data blocks are stored on adjacent tracks, such degradation can be effectively hide by the memory buffer and the SSD cache.

What is the impact on other disciplines?

The research project has started to integrate into the US CMS research facility at the University of Nebraska (UNL). US CMS is a collaboration US scientists participating in the Compact Muon Solenoid (CMS) experiment at the Large

Hadron Collider (LHC) at CERN in Geneva, Switzerland. UNL's US CMS Tier-2 site is a child site of the Tier-1 site at Fermi Nation Laboratory (FNAL). CMS sites employ dCache, a distributed storage data caching system, to support data access and transfer. We have started to prototype SAM2 toolkit, in particular the prefetching algorithms, into dCache to improve the I/O performance. In addition to CMS Tier-2 facility, UNL's Research Computing Facility (RCF), the primary computation resource at UNL, will benefit from our SAM2 toolkit. RCF includes (1) a distributed-memory supercomputer named Prairiefire that has 256 AMD Opteron processors capable of 88.5Gflops and (2) a shared-memory supercomputer from SGI named Homestead that contains 32 500MHz MIPS processor.

At the University of Maine, Dr. Yifeng Zhu has collaborated with researchers in Marine Science and Earth Sceinces to alleviate the I/O bottleneck for their simulations. The climate model developed at UMaine is I/O intensive. We have been collaborating to utilize parallel I/O to speed up the data-intensive science applications. Due to the success of this project, PI have been collaborating with faculties in Earth Science and Bioengineering on big data research projects. Two collaborative proposals have been submitted to NSF.

What is the impact on the development of human resources?

At UMaine, this collaborative research project has supported two graduate-students, and one post-doctoral research. Our post-doc has received the best paper nominations in HPCA'13 and he is also planning to enter the academia.

At the University of Maine, the research findings and concepts have been incorporated into two innovative NSF-funded education programs, directed by Dr. Yifeng Zhu, to provide college undergraduates as well as middle-school teachers and their students' firsthand experiences in scientific computing. (1) The Supercomputing Undergraduate Program in Maine (SuperMe), funded by NSF, is an opportunity for 10 UMaine undergraduate students to spend the summer conducting the kind of sophisticated, meaningful scientific research that is usually reserved for more advanced students. (2) With a separate NSF grant, another three-year program aims to integrate supercomputer modeling into the Maine middle-school science curriculum. Called Inquiry-based Dynamic Earth Applications of Supercomputing (IDEAS), the program will allow 20 middle-school teachers and 60 of their students each year to explore the myriad intricacies of UMaine's climate computer model by accessing the supercomputer with their state-issued laptops.

What is the impact on physical resources that form infrastructure?

Due to this research work, we have received equipment donation from LexisNexis, a leading global provider of content-enabled workflow solutions designed specifically for professionals in the legal, risk management, corporate, government, law enforcement, accounting, and academic markets.

What is the impact on institutional resources that form infrastructure?

Due to our research on ultra-large scale data storage systems, our research group along with other researchers in data science and engineering has been selected as one of signature and emerging Areas of excellence in research and education at the University of Maine. This designation is to form strategic and focused planning and resource allocation to preserve UMaine's national stature and impact in Maine.

What is the impact on information resources that form infrastructure?
Nothing to report.

What is the impact on technology transfer?
Nothing to report.

What is the impact on society beyond science and technology?
Nothing to report.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.