Sudharshan S. Vazhkudai

Oak Ridge National Laboratory
vazhkudaiss@ornl.gov, http://users.nccs.gov/~vazhkuda, Ph: 865-258-8986

1. CURRENT POSITIONS

Group Leader, Technology Integration, Oak Ridge National Laboratory (ORNL http://www.ornl.gov)

- *Occupation: R&D Management*

Joint Faculty, The University of Tennessee (UTK http://www.utk.edu)

2. PROFILE

- ***Experience***: 15 years of experience in the US Department of Energy's (DOE) national lab system, working in the supercomputing and extreme-scale data center programs.
- ***Education***: Doctorate in computer science with a focus on distributed storage and data management.
- ***Leadership***: Lead a group of 23 R&D staff members to build and deploy solutions for the nation's premier supercomputing center, the Oak Ridge Leadership Computing Facility (OLCF http://www.olcf.ornl.gov). OLCF is home to the world's No. 3 supercomputer, Titan, and the fastest storage system, Spider, providing billions of core hours to a scientific user base from academia, government and industry, to perform breakthrough research in climate, materials, alternative energy sources and energy storage, chemistry, nuclear physics, astrophysics, quantum mechanics, and the gamut of scientific inquiry. As a member of OLCF management, I contribute to the strategic direction of the center, and lead my group in several areas such as file and storage systems, non-volatile memory (NVRAM), data management, system architecture, networking and distributed systems.
- ***Technology Innovation and Creativity***:
  - Lead the TechInt group in the *development and deployment* of the following for OLCF:
    - ✓ *Systems Software* for the Titan supercomputer, Titan (27 petaflop heterogeneous CPU/GPU machine with 299,008 CPU cores and 18,688 GPUs), in the areas of scheduling, runtime environment and reliability.
    - ✓ *Storage Systems* such as the deployment of Spider, a Lustre-based parallel file system (PFS) (1 TB/s I/O throughput, 40 PB and 1 Billion files), the development of associated I/O tools and services, the development of a large-scale disk-cache/tape-based archival storage system software, HPSS (50 PB, with 61 million files), the development of multi-tiered storage abstracting distributed SSDs and PFS, tagging/annotations-based storage and data analytics workflow-aware storage.
    - ✓ *Solutions for future deep memory hierarchies*, such as NVRAM-based memory extension, distributed SSD-based burst buffer and active processing.
    - ✓ *Data/Metadata Management and Analytics* by federating and correlating metadata and the wealth of logs from HPC center resources (e.g., data, jobs, users, publications and subsystems) using graph and mining techniques, to answer numerous data disposition questions, discover new knowledge pathways, and optimize data center operations.
  - Led the development of a distributed computation and data management solution for the US DOE's experimental neutron scattering facility, the Spallation Neutron Source (SNS), a billion-dollar national infrastructure that produces petabytes of data.
  - 60+ research papers in highly selective peer-reviewed conferences/journals, with around 1732 citations, an h-index of 19 and an i10-index of 32.
  - Having worked in computer science R&D, in the leadership of the nation's premier supercomputing center, and in architecting solutions for the SNS experimental facility uniquely positions me to deal with the challenges posed by future extreme-scale systems.
- ***Budget Management***: Manage a multimillion-dollar annual budget.
- ***Track Record Delivering Products***: Experience in taking research ideas to products, under tight constraints. Software products deployed on national infrastructure, with thousands of users.
- ***Mentoring & Team Building***: Ability to attract and retain talent. Successful in building dynamic teams to achieve both short/long-term goals. Received an "*Outstanding Mentor Award*" from ORNL.
- Experience dealing with sensitive vendor and customer relationships.

3. EDUCATIONAL EXPERIENCE

Doctor of Philosophy, Computer Science, University of Mississippi, May 2003

Dissertation: Bulk Data Transfer Forecasts and the Implications to Grid Scheduling

Highlight: Thesis work conducted in part at Argonne National Laboratory (ANL, a DOE lab), on the Globus grid toolkit, through Wallace Givens and dissertation fellowships.

Master of Science, Computer Science, University of Mississippi, December 1998

Thesis: Performance Oriented Distributed OS-Evolutionary Steps towards a Distributed Linux

Bachelor of Engineering, Computer Science, Karnatak University (India), June 1996

4. WORK EXPERIENCE

OAK RIDGE NATIONAL LABORATORY (US DOE Lab) **Oak Ridge, TN**

2012 – present *Group Leader*, Technology Integration, Center for Computational Sciences

2003 – 2012 *Research Scientist*, Computer Science and Mathematics Division

THE UNIVERSITY OF TENNESSEE **Knoxville, TN**

2010 – present *Joint Faculty*, Joint Institute of Computational Sciences

2005 – 2006 *Adjunct Assistant Professor*, Computer Science

ARGONNE NATIONAL LABORATORY (US DOE Lab) **Argonne, IL**

2000 – 2002 *Givens Fellow/Doctoral Fellowship*, Math and Computer Science Division

WIPRO INFOTECH **Bangalore, India**

1996 – 1997 *Design Engineer, R&D*

5. R&D MANAGEMENT

- **GROUP LEADER, TECHNOLOGY INTEGRATION, ORNL**

Mission: The Technology Integration (TechInt http://techint.nccs.gov) group is charged with delivering new technologies in supercomputing and extreme-scale storage systems for the OLCF. The technology scope includes PFS, NVRAM, system architecture, archival storage and data management. TechInt is responsible for: delivering system software solutions for the Titan supercomputer (world's No. 3 machine); developing and deploying solutions for the deep-storage/memory hierarchy (e.g., Spider, the world's fastest PFS, NVRAM-based distributed store and the HPSS archival storage); and technologies for the future 200 petaflop Summit machine. We further evaluate emerging technologies in the above areas, and provide the systems programming to seamlessly integrate them into the infrastructure as they are adopted.

Composition: 23 research staff (direct/matrixed), with 5-30+ years of experience in systems R&D.

Role: Involves architecting of technology solutions, project/people/budget management, maintaining vendor/customer relationships, and contributing to the strategic direction of OLCF.

- Set a vision for the group, identify R&D directions to pursue by ascertaining gaps in the systems software/hardware stack, conceive project ideas, and conceptualize the basic high-level system design.
- Work with cross-functional teams such as operations, scientific user base and user assistance to gather requirements for products, and develop specifications.
- Build teams for projects, prioritize and assign tasks to staff, monitor progress of deliverables, and mentor staff on technical roadblocks.
- Hands-on involvement in architecting solutions.
- Responsible for the deployment and user outreach of products.
- Employee performance management, appraisals, hiring, retention and compensation planning.
- Manage a multimillion-dollar budget (staff and hardware).
- Responsible for relationship management with vendors, users and funding agencies.
- Publish high-quality papers, contribute to program reviews, and develop grant proposals.

Growth: I have expanded the group to include more areas such as NVRAM, system architecture, and data management, which involved successfully lobbying for a need to grow in these areas to senior management as well as writing grant proposals to funding agencies.

Impact of group's R&D:

- Design and deployment of the Spider storage system, one of the world's fastest file systems.
- Software products have been deployed in OLCF (Titan, Spider and other clusters), other DOE labs, industry (e.g., IBM, Intel) and several sites worldwide, and being used by thousands of users.
- Our tools have resulted in the efficient use of supercomputer time (a precious, rigorously peer-reviewed commodity), helped large-scale applications improve their I/O and be more resilient.
- Our work has been published in high-quality venues, and cited in industry blogs on storage systems.

Thrust Areas: Leading projects in several thrust areas, towards a robust supercomputing infrastructure.

A. The Deep Storage Hierarchy: *Led the development of short-term scratch PFS, associated tools/services and long-term archival storage for curation. The storage solutions deployed are being used by over a thousand users for their day-to-day activities at the OLCF HPC center.*

- ***The Spider Lustre-based Parallel File System Development and Deployment*:** (i) *Design, deployment, end-to-end tuning and operation* of the HPC center-wide Lustre storage system, *Spider* (1 TB/s, 40 PB, 20,160 2TB disks, 288 OSS and 2016 OSTs), and its integration into Titan's 18,688 compute nodes and several data analysis clusters via scalable I/O routing **[SpiderPFS:SC14]**. (ii) *Highly parallel tools development* for standard file system operations. These tools are being used to operate on petabytes of data and 1 Billion files. (iii) *Lustre PFS* code contributions in areas such as balanced I/O placement based on OST load **[BPIO:ICPADS14]**, temporal replication **[TemporalReplication:ISC09]**, novel data recovery **[IOSched:SC07, Recovery:ICPP08]**, and use of accelerators for distributed RAID **[GPURAID:MASCOTS12]**. (iv) *A model-driven provisioning tool* to assist storage system designers and administrators reconcile key figures of merit (cost, capacity, performance, disk size, rebuild times, redundancy), answer what-if scenarios, and determine the relative importance of spare parts in minimizing data unavailability, both during initial system provisioning and continuous operations. Validated the tool with field observed data from Spider operations **[ModelSpider:SC15]**.
- ***Scalable I/O Monitoring, Signature Extraction and Data Scheduling:*** (i) A monitoring infrastructure by tracking the workload on Spider's backend storage controllers (72), and exposing the throughput and load statistics via a database for higher-level I/O-aware tools (e.g., file system partition selection). (ii) *Automatic extraction of application I/O signatures* from noisy Spider storage backend I/O logs, using a rich suite of statistical techniques **[IOSIG:FAST14, IOAID:SC16]**. (iii) *A coordinated scheduling* mechanism to coincide data staging, computation and data offloading in an attempt to view the PFS as a cache **[IOSched:SC07, ScratchasCache:ICS09]**.
- ***TagIt: An integrated search and discovery service for file systems.*** Led the design and development of TagIt, a novel data management service framework to annotate, tag, index, and perform filtering operations on files, and tightly integrated them into a shared-nothing GlusterFS distributed file system **[TagIt:ATC16]**. TagIt manages: a scalable and consistent metadata index database (MySQL) inside the file system, exploiting readily available resources; enables advanced tagging that allows users to mark from entire collections of files down to specific portions of a file; enables the association of operators to a tag for pre-processing, filtering or automatic metadata extraction, which is seamlessly offloaded to file servers in a load-aware fashion.
- ***Multi-Tiered Storage*:** Leading an effort to abstract the multiple storage tiers seen in HPC centers, e.g., a distributed SSD-based burst buffer, PFS, warm archive and deep archival storage. Designing scalable NoSQL database catalogs, unified namespace, and policy-based data migration to present a virtual, on-demand storage hierarchy for application workflows.
- ***HPSS Archival Storage*:** HPSS is a two-decade partnership among several DOE labs and IBM. (i) Oversee ORNL activities on the HPSS disk/tape storage software development. (ii) Led the development of a data quality assurance framework to verify the integrity of the archive (61 million files). (iii) Contributed to an advanced technology working group, providing a strategic roadmap for the evolution of HPSS for the next decade. (iv) Member of the HPSS executive committee that decides both the tactical and strategic direction of HPSS.

B. Deep Memory Hierarchy (SSDs and NVRAM): *Devising NVRAM solutions for next-generation supercomputers, from three perspectives, namely resilience, memory extension, and active processing.*

- ***Fault Tolerance—SSD-based Burst Buffer Storage*:** The 200 petaflop Summit machine in 2018 will have compute-node local SSDs on its ~ 5K nodes. Developing a FUSE-based, relaxed POSIX distributed file system, *SymphonyFS*, that provides a unified namespace between the node-local SSDs and the center-wide GPFS file system so user applications see a single mount point for N-N and N-1 checkpoint writes. This will enable the seamless use of the node-local SSD for users.
- ***Memory Extension—NVMalloc*:** (i) Led the development of a runtime library for applications to use block-addressable SSDs in a byte-addressable fashion, as a secondary memory partition, and explicitly allocating variables therein **[NVMalloc:IPDPS12]**. (ii) Developing mechanisms to specify the level of persistence desired for variables at allocation time, while the NVMalloc library seamlessly checkpoints and migrates data between the deep-memory tiers. (iii) Application variable usage profiling to provide placement suggestions across deep-memory tiers.
- ***Active Processing*:** (i) *Active Flash*: Developed mechanisms to push computation into the flash device

controller where the data already resides, facilitating in-situ analysis, and minimizing energy consumed due to data movement. Built an Active Flash emulation using the OpenSSD development kit **[ActiveFlash:FAST13, TrueOutOfCore:MSST12]**. (ii) *Analysis-aware Storage System (AnalyzeThis)*: Built an analytics workflow and provenance aware storage appliance using an array of active flash devices, proposing novel analysis object semantics to tie together the data and the analytics to be (or was) performed on the data. Analysis awareness is deeply ingrained in each and every layer of the storage, elevating data analyses as first-class citizens in the storage system. **[AnalyzeThis:SC15]** (*highlighted in DOE's ASCR Discovery*). (iii) *A Programmable Shared-Memory System for Processing-In-Memory Devices (AnalyzeThat)*: Developed a rich PIM-aware Data Structure (PADS), a key-value data container encapsulation that integrally ties together the data, analyses and the runtime needed to interface with and place data on the PIM device array.

- ***Lifetime Analysis*:** (i) Profiled the data production rate of Titan's 18,688 compute nodes over four years (~11GB/node) to project the rate for Summit nodes in 2018, and using this to determine the Summit node-local SSD's required drive writes/day. (ii) Studying the effects of SSD wear when used for different workloads, such as checkpointing, draining, memory extension and active analysis.

C. Data, Logs, Metadata Management and Mining: *Leading the development and deployment of solutions for large-scale data search, metadata extraction, derivation and indexing, data curation and log analysis for the production OLCF environment.*

- ***Constellation Science Graph Network for Scalable Data Discovery*:** Constellation federates metadata from the OLCF resource fabric (stat metadata from ~1 Billion files from Spider PFS/HPSS, millions of jobs metadata from the scheduler, thousands of users/groups, publications and systems), and captures them in a custom-built in-memory graph; builds links or associations among resources (vertices in the graph) by correlating the metadata to infer hidden relationships (e.g., linking data to jobs, extracting keywords from publications to link together relevant publications or publications, jobs and data); graph traversals and high-performance indexes external to the graph enable searches. The stat metadata index is built using Hbase and Spark queries on PFS stat/job metadata; we also build an hierarchical index by extracting metadata from within the datasets themselves, and create more metadata from base metadata. Based on this graph engine, we can discover data, suggest related data products/papers of interest, identify popular datasets via pagerank, study and create new user-specified "tags" that tie together resources for quick retrieval and sharing **[ConstellationGraph:BigData16]**.
- ***DOIs for HPC Data*:** Developed a Digital Object Identifier workflow to curate important data products from Titan jobs for long-term storage. DOI provides a way to cite data products and processes.
- ***GUIDE—Data Analytics on large-scale Logs to Optimize HPC Data center Operations:*** Led the creation of a scalable Splunk-based infrastructure to aggregate and analyze huge amounts of logs to present a "*window into and optimize HPC data center operations*." Logs include the following. (i) Spider PFS: I/O bandwidth data from controllers, Lustre-level profiling data, file size distribution of the 1 Billion files and health logs from 2016 OSTs comprising of 20,160 disks, (ii) Titan CPU/GPU RAS data from 18,688 nodes, (iii) Moab job scheduler and node allocation data, (iv) interconnect congestion data from the 9600 routers, and (v) HPSS archival storage usage and file size distribution data from 61 Million files. Developed higher-level services using a variety of analytics techniques such as the correlation of logs, data mining, statistical and visual analytics, which are being used to identify hotspots, debug performance bottlenecks and trend analysis for future resource provisioning.

D. Computer System Architecture and Resilience: *Developed software systems for the efficient use of the Titan's 299,008 CPU cores and 18,688 GPUs in the areas of runtime, scheduling and resilience.*

- ***Functional Partitioning Runtime for Many-core nodes*:** Leading the development of a runtime environment, Functional Partitioning (FP), to partition a many-core node such that an end-to-end application (simulation + data analysis tasks) can be scheduled on the same node, in-situ, alongside the application's simulation job for better end-to-end performance. We exploit the CPU/GPU cores and memory that are under-utilized by the application's simulation, and put them to use for the application's own end-to-end data analysis tasks **[FP:SC10]**.
- ***Scheduling*:** To minimize the runtime variability of jobs and reduce node allocation fragmentation, we developed a dual-ended scheduling policy for the Moab scheduler on Titan as opposed to the first-fit policy. Our algorithm schedules large jobs top down and small jobs bottom up, thereby minimizing the fragmentation for large, long-running jobs that is caused by small, short-lived jobs. Further, we also modified the scheduling of nodes to a job to prioritize the *z* dimension (nodes within a rack), followed by the *x* dimension (nodes in the racks that are in the same row offer better communication bandwidth)

and then the *y* dimension (columns) **[TitanScheduling:SC16]**. Thousands of jobs are benefitting due to this technique. Received a ***Significant Event Award***, a prestigious lab-wide recognition.

- ***Supercomputer Failure Analysis*:** Analysis of temporal failure characteristics of Titan's 299,008 CPUs and 18,688 GPUs to understand trends in machine failure, MTBF, single bit errors, double bit errors, off the bus errors and temperature correlation to failure **[TitanGPUReliability:HPCA15]**. This study was the first of its kind for a large-scale GPU deployment. Based on the insights, devised checkpointing advisory tools **[LazyChkpt:DSN14]** that are saving millions of core hours for production jobs, devised an energy-based scheduling algorithm to schedule large node-count GPU jobs that stress the GPUs more to be scheduled at the bottom of the rack where it is much cooler than the top of the rack, or cordoned off nodes with frequent failures.
- ***Interconnect Congestion Analysis*:** Developed and deployed a lightweight, scalable mechanism to monitor the router traffic on Titan's interconnect fabric (9600 routers). Tool being used to analyze interference among jobs.
- ***Heterogeneous Processor Architectures and End-to-end Computing*:** (i) Studied how end-to-end workflows (simulations + data analysis) perform on heterogeneous processor architectures such as big processors (AMD Opterons, XEONs), Coprocessors (GPUs, MIC) and little processors (ARM, ATOM), considering several competing dimensions such as performance, on-chip scalability, energy-efficiency, and error resiliency to provide insights on the design of future heterogeneous systems for end-to-end workflows. (ii) Procured an ARM-based cluster, and making it available to a community of users to study the architecture's viability for exascale.

- **RESEARCH SCIENTIST, COMPUTER SCIENCE RESEARCH, ORNL**

Led several storage and data management projects, resulting in software releases and deployments.

A. FreeLoader Cache using Distributed Storage Scavenging: (i) Developed a distributed storage infrastructure, built using space contributions from commodity desktops, presenting it as a client-side cache for data-intensive applications **[Freeloader:SC05]**, (ii) Developed techniques for throttling resource usage on desktops **[GovernorThrottling:ICAC05]** (iii) Developed novel data caching and recovery schemes via prefix caching and suffix patching **[PrefixCaching:ICS06**, **DataAvailability:JOG09]**, and (iv) Explored the use of FreeLoader as a distributed checkpoint storage for volunteer computing jobs on workstations **[stdchkFS:ICDCS08]**.

B. Scientific Data Management for SNS: Lead architect for a distributed computation and data management solution for O(100TB) from the Spallation Neutron Source (SNS), a billion dollar facility, catering to a large user community **[SNS:PPPJ07]**. Received a ***Significant Event Award***.

C. **Staged Data Transfer**: Developed decentralized tools to expedite wide-area data delivery between HPC centers and end-user sites, using intermediate nodes, thereby alleviating the last-mile problem **[ResultDeliver:TPDS11, CloudDeliver:IPDPS11]**.

- **JOINT FACULTY, THE UNIVERSITY OF TENNESSEE**

Worked with students and postdocs on several problems in storage and data management.

A. Distributed NVRAM/DRAM-based staging Storage: Led an NSF grant to construct a distributed, intermediate storage system, built from a job's own allocation of nodes (SSDs and DRAM tiers), in order to conduct in-situ data analysis **[StagingStore:ICDCS11]**.

B. Technical lead and Co-PI on an NIH effort to build a data repository so that National Cancer Institute (NCI) funded centers can aggregate, annotate, search, index, and download data.

- **DOCTORAL THESIS** (in part at Argonne National Lab)**:** Developed middleware for the orchestration of bulk data transfers in the Globus Data Grid by building replica selection **[ReplicaSelection:CCGRID01]**, regression methods for transfer predictions **[RegressionPrediction:JHPCA03]** and distributed data downloads **[TransferCoallocation:JOG04]**.
- **MASTERS THESIS:** Developed a distributed OS for Linux **[DistributedLinux:FGCS02]**.

6. <u>SELECTED PUBLICATIONS</u> (60+ papers: http://users.nccs.gov/~vazhkuda/publications.html)

1. **[ConstellationGraph:BigData16]** S. S. Vazhkudai, et. al., "Constellation: A Science Graph Network for Scalable Data and Knowledge Discovery in Extreme-Scale Scientific Collaborations," *Proceedings of the IEEE Workshop on Big Data Metadata and Management*, Washington DC, December 2016.
2. **[TitanScheduling:SC16]** C. Zimmer, S. Gupta, S. Atchley, S.S. Vazhkudai, and C. Albing, "A Multi-faceted Approach to Job Placement for Improved Performance on Extreme-Scale Systems," *Proceedings of Supercomputing 2016 (SC16): 29th IEEE/ACM Int'l Conference on High Performance Computing, Networking, Storage and Analysis*, Salt Lake City, UT, November 2016.

3. **[IOAID:SC16]** Y. Liu, R. Gunasekaran, X. Ma, and S.S. Vazhkudai, "Server-side Log Data Analytics for I/O Workload Characterization and Coordination on Large Shared Storage Systems," *Supercomputing 2016 (SC16)*, Salt Lake City, UT, November 2016.
4. **[TagIt:ATC16]** H. Sim, Y. Kim, S.S. Vazhkudai, G.R. Vallee, S. Lim, and A.R. Butt, "TagIt: An Integrated Search and Discovery Service for Extreme-Scale File Systems," *(Poster) Proceedings of the USENIX Annual Technical Conference (ATC)*, Denver, CO, June 2016.
5. **[AnalyzeThis:SC15]** H. Sim, Y. Kim, S.S. Vazhkudai, et. al., "AnalyzeThis: An Analysis Workflow-Aware Storage System," *Supercomputing 2015 (SC)*, Austin, TX, November 2015.
6. **[ModelSpider:SC15]** L. Wan, F. Wang, S. Oral, D. Tiwari, S.S. Vazhkudai, Q. Cao, "A Practical Approach to Reconciling Availability, Performance, and Capacity in Provisioning Extreme-scale Storage Systems," *Supercomputing 2015 (SC)*, Austin, TX, November 2015.
7. **[TitanGPUReliability:HPCA15]** D. Tiwari, et. al., "Understanding GPU Errors on Large-scale HPC Systems and the Implications for System Design and Operation," *21st IEEE High Performance Computer Architecture (HPCA)*, San Francisco, CA, February, 2015.
8. **[BPIO:ICPADS14]** F. Wang, et. al., "Improving Large-scale Storage System Performance via Topology-aware and Balanced Data Placement," *20th IEEE Int'l Conference on Parallel and Distributed Systems (ICPADS)"*, Hsinchu, Taiwan, December 2014.
9. **[SpiderPFS:SC14]** S. Oral, et. al., "Best Practices and Lessons Learned from Deploying and Operating Large-Scale Data-Centric Parallel File Systems," *Supercomputing 2014 (SC)*, New Orleans, LA, November 2014.
10. **[LazyChkpt:DSN14]** D. Tiwari, S. Gupta, S.S. Vazhkudai, "Lazy Checkpointing: Exploiting Temporal Locality in Failures to Mitigate Checkpointing Overheads on Extreme-Scale Systems," *44th IEEE/IFIP Conference on Dependable Systems and Networks (DSN)*, Atlanta, GA, June 2014.
11. **[IOSIG:FAST14]** Y. Liu, R. Gunasekaran, X. Ma, S.S. Vazhkudai, "Automatic Identification of Applications I/O Signatures from Noisy Server-Side Traces," *12th USENIX Conference on File and Storage Technologies (FAST)*, Santa Clara, CA, February 2014.
12. **[ActiveFlash:FAST13]** D. Tiwari, S. Boboila, S.S. Vazhkudai, et. al., "Active Flash: Towards Energy-Efficient, In-Situ Data Analytics on Extreme-Scale Machines," *11th USENIX Conference on File and Storage Technologies (FAST)*, San Jose, CA, February 2013.
13. **[GPURAID:MASCOTS12]** A. Khasymski, M.M. Rafique, A.R. Butt, S.S. Vazhkudai, D.S. Nikolopoulos, "On the Use of GPUs in Realizing Cost-Effective Distributed RAID," *IEEE Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS)*, Washington, D.C., August 2012.
14. **[NVMalloc:IPDPS12]** C. Wang, S.S. Vazhkudai, et. al., "NVMalloc: Exposing an Aggregate SSD Store as a Memory Partition in Extreme-Scale Machines," *26th Int'l Parallel & Distributed Processing Symposium (IPDPS)*, Shanghai, China, May 2012.
15. **[TrueOutOfCore:MSST12]** S. Boboila, Y. Kim, S.S. Vazhkudai, et. al., "Active Flash: Out-of-core Data Analytics on Flash Storage," *28th Mass Storage Systems and Technologies* (MSST), Monterey, CA, April 2012.
16. **[ResultDeliver:TPDS11]** H. Monti, A.R. Butt, S.S. Vazhkudai, "Timely Result-Data Offloading for Improved HPC Center Scratch Provisioning and Serviceability," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, Vol. 22, No. 8, pp. 1307-1322, August 2011.
17. **[StagingStore:ICDCS11]** R. Prabhakar, S. S. Vazhkudai, et. al., "Provisioning a Multi-Tiered Data Staging Area for Extreme-Scale Machines," *31st Int'l Conference on Distributed Computing Systems (ICDCS)*, Minneapolis, MN, June 2011.
18. **[CloudDeliver:IPDPS11]** H. Monti, A.R. Butt, S.S. Vazhkudai, "CATCH: A Cloud-based Adaptive Data Transfer Service for HPC," *25th IEEE International Parallel & Distributed Processing Symposium* (IPDPS), Anchorage, Alaska, May 2011.
19. **[FP:SC10]** M. Li, S. S. Vazhkudai, et. al., "Functional Partitioning to Optimize End-to-End Performance on Many-core Architectures," *Supercomputing (SC)*, New Orleans, LA, November 2010.
20. **[DataAvailability:JOG09]** X. Ma, S.S. Vazhkudai, Z. Zhang, "Improving Data Availability for Better Access Performance: A Study on Caching Scientific Data on Distributed Desktop Workstations," *Journal of Grid Computing*, Vol. 7, No. 4, pp. 419-438, December 2009.
21. **[ScratchasCache:ICS09]** H. Monti, A.R. Butt, S.S. Vazhkudai, "/Scratch as a Cache: Rethinking HPC Center Scratch Storage," *23rd ACM Int'l Conference on Supercomputing (ICS)*, Yorktown Heights, NY, June 2009.

22. **[TemporalReplication:ISC09]** C. Wang, Z. Zhang, X. Ma, S.S. Vazhkudai, F. Mueller, "Improving the Availability of Supercomputer Job Input Data Using Temporal Replication," *Int'l Supercomputing Conference*, Hamburg, Germany, June 2009.
23. **[Recovery:ICPP08]** C. Wang, Z. Zhang, S. S. Vazhkudai, et. al., "On-the-fly Recovery of Job Input Data in Supercomputers," *37th Int'l Conference on Parallel Processing* (ICPP), Portland, OR, September 2008.
24. **[stdchkFS:ICDCS08]** S.A. Kiswany, M. Ripeanu, S. S. Vazhkudai, A. Gharaibeh, "stdchk: A Checkpoint Storage System for Desktop Grid Computing," 28th Conference on Distributed Computing Systems (ICDCS), Beijing, China, June 2008.
25. **[IOSched:SC07]** Z. Zhang, C. Wang, S. S. Vazhkudai, X, Ma, G. Pike, F. Mueller, J.W. Cobb, "Optimizing Center Performance through Coordinated Data Staging, Scheduling and Recovery," *Supercomputing (SC)*, Reno, NV, November 2007.
26. **[SNS:PPPJ07]** S. S. Vazhkudai, et. al., "A Java-based Science Portal for Neutron Scattering Experiments," *Principles and Practices of Programming in Java*, Lisbon, Portugal, September, 2007.
27. **[PrefixCaching:ICS06]** X. Ma, S. S. Vazhkudai, et. al., "Coupling Prefix Caching and Collective Downloads for Remote Data Access," *20th Int'l Conference on Supercomputing (ICS)*, pp. 229-238, Cairns, Australia, June 2006.
28. **[Freeloader:SC05]** S. Vazhkudai, X. Ma, V. Freeh, et. al., "FreeLoader: Scavenging Desktop Storage Resources for Scientific Data," *Supercomputing (SC)*, Seattle, WA, November 2005.
29. **[GovernorThrottling:ICAC05]** J. Strickland, et. al., "Governor: Autonomic Throttling for Aggressive Idle Resource Scavenging," *2nd IEEE Int'l Conference on Autonomic Computing (ICAC)*, Seattle, Washington, 2005.
30. **[TransferCoallocation:JOG04]** S. Vazhkudai, "Distributed Downloads of Bulk, Replicated Grid Data," *Journal of Grid Computing*, No. 2, pp. 31-42, 2004.
31. **[RegressionPrediction:JHPCA03]** S. Vazhkudai, et. al., "Using Regression Techniques to Predict Large Data Transfers," *Journal of High Performance Computing Applications - Special Issue on Grid Computing: Infrastructure and Applications*, Vol 17. No. 3, pp. 249-268, 2003.
32. **[DistributedLinux:FGCS02]** S. Vazhkudai, et. al., "PODOS - The Design and Implementation of a Performance Oriented Linux Cluster," *Journal of Future Generation Computer Systems - Special Issue on Cluster Computing*, 2002.
33. **[ReplicaSelection:CCGRID01]** S. Vazhkudai, S. Tuecke, I.T. Foster, "Replica Selection in the Globus Data Grid," *Cluster Computing and the Grid* (CCGRID), Brisbane, Australia, May 2001.

9. SELECTED PROGRAM DEVELOPMENT ACTIVITIES

Initiated several multi-institutional projects by obtaining external grants from DOE, NSF, and NIH.

1. L. Ramakrishnan, D. Agarwal, S.S. Vazhkudai, M. Franklin, C. Aragon, "Usable Data Abstractions for Next-Generation Scientific Workflows," DOE ASCR Scientific Data Management Program Announcement LAB 14-1043, 09/2014-08/2017 (Role: Co-PI).
2. S.S. Vazhkudai, et. al., "Dynamic Staging Architecture for Accelerating I/O Pipelines," *NSF High-End Computing Research Activity (HECURA)*, CCF-0937827, 04/2010-03/2013 (Role: PI).
3. X. Ma, Y. Zhou, V.W. Freeh, S. Vazhkudai, "Application-adaptive I/O Stack for High End Computing," *NSF HECURA*, CCF-0621470, FY 2007-2009 (Role: Co-PI).
4. S. Vazhkudai, et. al., "Storage Virtualization: An Integrated Approach to Machine-Room Storage Management," *DOE ORNL LDRD*, FY 2007-2008 (Role: PI).

10. AWARDS

- Significant Event Awards (2012 and 2015) – from ORNL for innovative data management for SNS and novel scheduling algorithm for the Titan supercomputer, respectively.
- Outstanding Mentor Award (Feb 2008) - from the Oak Ridge Institute for Science and Education.
- Wallace Givens and Doctoral Fellowship (2000 - 2002) - Argonne National Laboratory.

11. PROFESSIONAL SERVICE and SYNERGISTIC ACTIVITIES

- Served on several conference committees: SC, ICS, ICDCS, IPDPS, TPDS, JPDC, DSN and TOS.
- Educational outreach: advised 4 postdocs; co-advised 10 doctoral students, 4 masters students and several interns from different universities.