North Carolina State University

Scientific Data Management Center for Enabling Technologies

Project Report (DE-FC02-07) ER25809

Period: 11/15/2006 - 11/14/2012

NC State PI:

Mladen A. Vouk
Computer Science, Box 8206 North Carolina State University Raleigh, NC 27695
vouk@ncsu.edu
919-515-7886

SDM Center Comprehensive Report July 2006 - June 2011, and NCSU ext. Jul 2011 - Nov 2012

http://sdmcenter.lbl.gov

Table of Contents

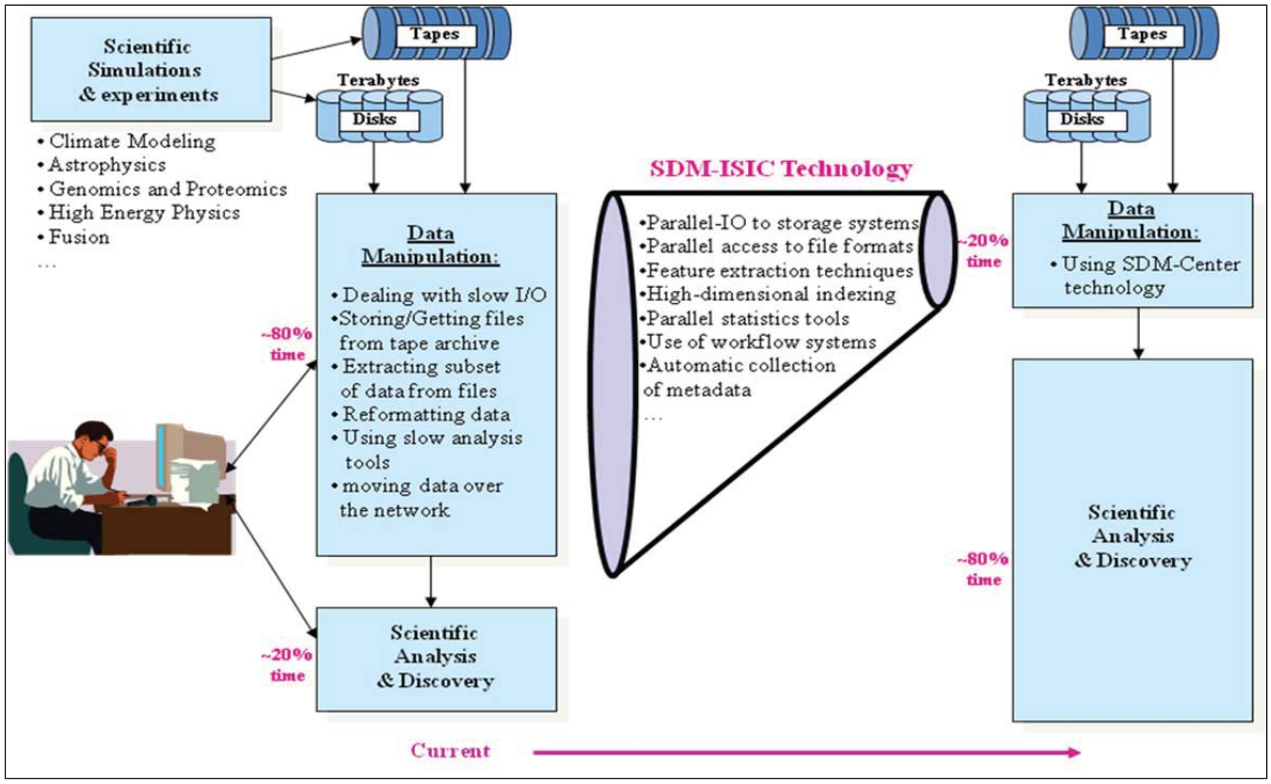
SDM Center Comprehensive Report

October 2006 - June 2011

http://sdmcenter.lbl.gov

SDM center Mission and goals

Managing scientific data has been identified by the scientific community as one of the most important emerging needs because of the sheer volume and increasing complexity of data being collected. Effectively generating, managing, and analyzing this information requires a comprehensive, end-to-end approach to data management that encompasses all of the stages from the initial data acquisition to the final analysis of the data. Fortunately, the data management problems encountered by most scientific domains are common enough to be addressed through shared technology solutions. Based on community input, we have identified three significant requirements. First, more efficient access to storage systems is needed. In particular, parallel file system and I/O system improvements are needed to write and read large volumes of data without slowing a simulation, analysis, or visualization engine. These processes are complicated by the fact that scientific data are structured differently for specific application domains, and are stored in specialized file formats. Second, scientists require technologies to facilitate better understanding of their data, in particular the ability to effectively perform complex data analysis and searches over extremely large data sets. Specialized feature discovery and statistical analysis techniques are needed before the data can be understood or visualized. Furthermore, interactive analysis requires techniques for efficiently selecting subsets of the data. Finally, generating the data, collecting and storing the results, keeping track of data provenance, data post-processing, and analysis of results is a tedious, fragmented process. Tools for automation of this process in a robust, tractable, and recoverable fashion are required to enhance scientific exploration. The goals of the center are shown schematically in the Figure below.

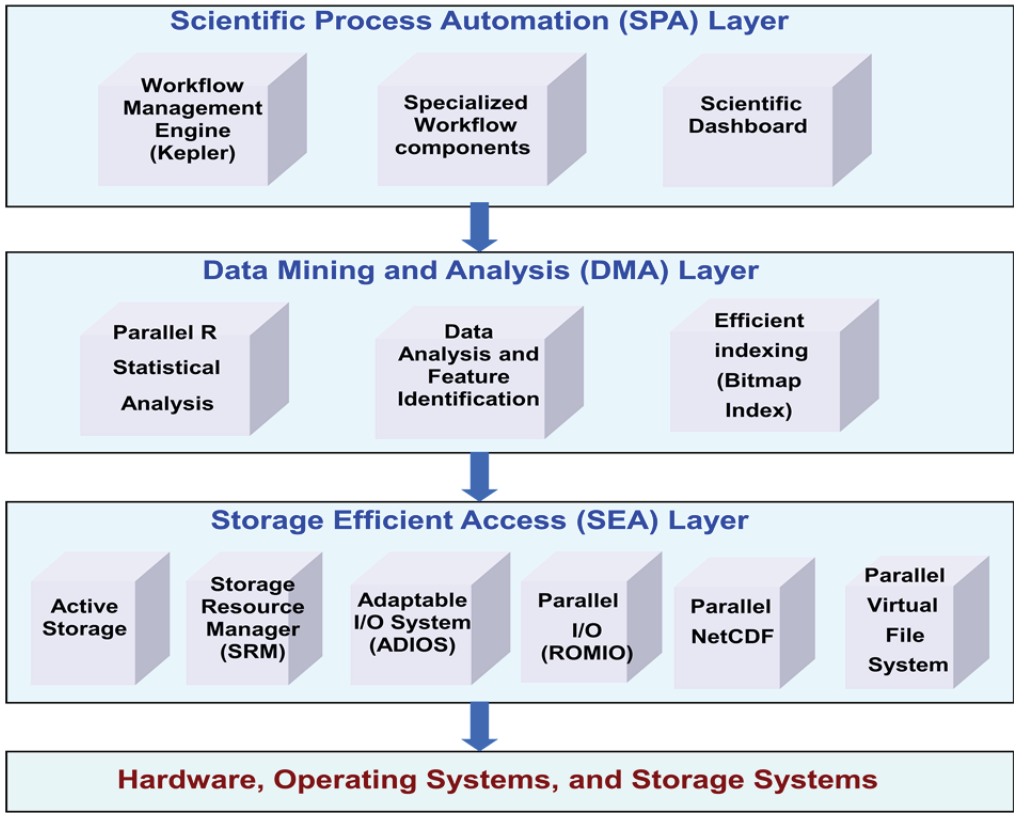


The SDM center was established under the SciDAC program to address these issues. The SciDAC-1 Scientific Data Management (SDM) Center succeeded in bringing an initial set of advanced data management technologies to DOE application scientists in astrophysics, climate, fusion, and biology. Equally important, it established collaborations with these scientists to better understand their science as well as their forthcoming data management and data analytics challenges. Building on our early successes, we have greatly enhanced, robustified, and deployed our technology to these communities. In some cases, we identified new needs that have been addressed in order to simplify the use of our technology by scientists. This report summarizes our work so far in SciDAC-2.

Our approach is to employ an evolutionary development and deployment process: from research through prototypes to deployment and infrastructure. Accordingly, we have organized our activities in three layers that abstract the end-to-end data flow described above. We labeled the layers (from bottom to top):

- Storage Efficient Access (SEA)
- Data Mining and Analysis (DMA)
- Scientific Process Automation (SPA)

The SEA layer is immediately on top of hardware, operating systems, file systems, and mass storage systems, and provides parallel data access technology, and transparent access to archival storage. The DMA layer, which builds on the functionality of the SEA layer, consists of indexing, feature identification, and parallel statistical analysis technology. The SPA layer, which is on top of the DMA layer, provides the ability to compose scientific workflows from the components in the DMA layer as well as application specific modules. The organization of the center and technologies included in each layer are shown below.



Organization of the report

This report consists of the following sections, organized according to the three layers, as follows.

The **Storage Efficient Access** (SEA) area includes the following activities: (1) file system benchmarking and application I/O behavior; (2) parallel I/O infrastructure evolution; (3) application interfaces and data models; and (4) next-generation I/O software technologies.

The **Data Mining and Analysis** (DMA) area includes the following activities: (1) high-performance statistical computing; (2) efficient searching and filtering in data-intensive scientific applications; and (3) Feature extraction and tracking in scientific applications.

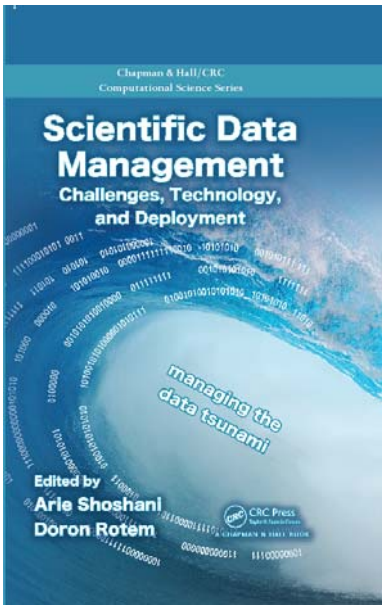
The **Scientific Process Automation** (SPA) area includes the following activities: (1) workflow development; (2) provenance collection; (3) generic actors; and (4) workflow fault tolerance.

In addition to the three sections covering progress in the three focus areas above, we include an additional section that describes our efforts in providing an **Framework for Integrated End-to-end SDM Technologies and Applications** (FIESTA) that uses multiple SDM center technologies for a specific SciDAC Fusion application project, called CPES (Center for Plasma Edge Simulation). The technologies used are from all three areas, and include workflow, analysis, I/O speed up, data movement technologies, and visual data analysis. The FIESTA was designed in collaboration with application users to provide them with sophisticated and powerful capabilities accessed through intuitive web interfaces. While FIESTA was designed in response to the CPES project, it was developed as a general framework that can be used in other application domains. We are currently engaging Combustion and Astrophysics scientists who expressed interest in using this framework as well.

Details of progress in each of the three areas, SEA, DMA, and SPA, as well as FIESTA, follow. This is followed by a **publications** and references section, and outreach, tutorials, invited talks, and theses, in *Appendix 1*. The SDM center has have developed numerous **collaborations** with application projects and other centers and institutes. This is summarized in a color-coded table form in *appendix 2*, as well as a summary description of the collaborative tasks. But, first we describe selected highlights that are discussed in these sections in more detail.

Selected Highlights of Achievements

- **The book on Scientific Data Management was published.** Members of the SDM center edited and contributed chapters to the book entitled "Scientific Data Management: Challenges, Existing Technology, and Deployment" [SR09]. In six out of thirteen chapters, the lead authors are members of the SDM center, and additional members contributed to the content of these chapters.
- **A book on Scientific Data Mining was published.** The book "Scientific Data Mining: A Practical Perspective," was authored by a member of the SDM center [Kam09].
- **Textbook titled "Practical Graph Mining with R" written entirely by students to be published by Chapman & Hall/CRC Press under the Data Mining and Knowledge Discovery Series**, was delivered in January 31, 2011, and will be published in 2012 (http://www.csc.ncsu.edu/news/1071). The lead editor and co-editors of the book are members of the SDM center.
- **High productivity in the SDM center, a large number of papers published.** During the last 5 years (January 2007 – July 2011)



members of the SDM center published **173 papers** (see publication list), and organized and presented numerous tutorials, invited talks, or invited sessions. Papers published prior to 2007, can be found in the SDM center web site: http://sdmcenter.lbl.gov.

- **Parallel NetCDF successfully used in production.** Parallel NetCDF (PnetCDF), designed, built, and supported by SDM center members, is now used in several production codes. It has been successfully used by the large-scale NCAR Community Atmosphere Model (CAM) and Global Cloud Resolving Model (GCRM). According to Yu-heng Tseng, Department of Atmospheric Sciences National Taiwan University, "Parallel NetCDF indeed solves a big problem on the large scale computing." A new parallel netCDF file format to support larger than 4 GB array size has been developed and tested extensively. The new format, called the "CDF-5" format, allows storage of variables of effectively unlimited size in the netCDF format. See details in Section 1.3.
- **400% I/O improvement achieved for collective I/O patterms on the Lustre parallel file system.** We enhanced I/O efficiency for the Lustre file system by as much as 400% by introducing a novel technique called partitioned collective IO (ParColl). ParColl partitions parallel processes into subgroups, each carrying out smaller, yet aggregated IO operations. This technique is important in part because it does not require a change in file format. See details in Section 1.2.
- **New high performance driver for Lustre developed and integrated into popular packages.** We have developed a Lustre driver for MPI-IO that enables higher performance by better tuning access to avoid performance pitfalls with Lustre file system on the Cray machines. This driver has been integrated into MVAPICH version 1.0, which is a popular MPI implementation for InfiniBand clusters, and MPICH2 version 1.0.7, which serves as the basis of most vendor implementations. Through these distributions, this technology will enhance I/O performance for a variety of applications using MPI-IO directly or through such high-level I/O libraries as HDF5 and PnetCDF. See details in Section 1.2.
- **ADIOS speeds up I/O on Cray XT, InfiniBand clusters, and IBM Blue Gene/P.** We developed the Adaptable I/O system (ADIOS), a componentization of the I/O layer. It provides an easy-to-use programming interface that abstracts the I/O metadata information and data structures from the source code into an external XML file. ADIOS allows users to only change the declaration of the transport methods in the XML file without any source code modification. ADIOS contains a new file format, bp (binary-packed), that is highly optimized for I/O for checkpoint operations. ADIOS has been integrated into the Chimera, GTC, GTS, XGC-1, and S3D codes, and can achieve up to 31GB/s on a 40 GB/s file system. ADIOS has improved I/O for the XGC-1 code on 128K processors by 40x. In one case, comparing Chimera with parallel HDF5 with ADIOS using bp, the improvement gain was over 1000x on 8192 cores. See more details in sections 1.3 and 4.
- **FastBit indexing technology received R&D 100 award.** FastBit is a very efficient indexing technology for accelerating database queries on massive datasets. FastBit has been proven to be theoretically optimal; it performs 50-100 times faster than any known indexing method based on its use of our patented compression method. It can search over multi-variable, scientific data where attributes have high cardinality (number of possible values). These unique characteristics made it useful in a variety of scientific applications. Our implementation was packaged in 2007 and released under an open-source license, and has attracted a lot of interest in multiple scientific applications, as well as new areas, such as network traffic data analysis, and query-based visualization. It received the prestigious R&D 100 award in 2008. See details in Section 2.2.

- **A huge (1000 fold) speedup of particle search for the Laser Wakefield Particle Accelerator project attained with FastBit**. We used FastBit to speed up the operations of searching and tracking particles in Laser Wakefield Particle Accelerator (LWPA) project (joint with VACET center). By

replacing an existing IDL based analysis program with a FastBit based program, we observed a three orders of magnitude speedup (from 300 seconds to 0.3 seconds) in the first test run. In another application, FastBit open source technology was used (without involving the FastBit developers) in a software called TrixX-BMI. It has been reported that FastBit enabled screening libraries of ligands 12 times faster than the state of art screening tools. See details in Section 2.2.

- **Another 1000 fold speedup was also achieved with specialized FastBit structures for Gyrokinetic Fusion region identification.** We developed a specialized bitmap index FastBit structure by directly utilizing the mesh structure of Gyrokinetic Transport Code (GTC) for simulating the magnetically confined fusion plasma. Consequently, we were able to improve the speed of identifying regions of interest by nearly 1000-fold (see details in section 2.2).
- **An accurate tool for classification of orbits in Poincare plots developed and deployed.** Software for the automatic classification of orbits in Poincare plots was developed and deployed for use by PPPL and other fusion scientists, solving a long-standing problem for scientists in this domain. It takes as input the coordinates of the points in an orbit and assigns to it one of four class labels based on the shape traced out by the points. Key challenges to solving this problem were the extraction of robust features representing the orbits and the creation of a high-quality training set. The cross-validation error rate using ensembles of decision trees is less than 4% and the code "works quite well" in the words of a physicist using it, who has recommended it to colleagues. This software would replace tedious manual labeling of the orbits, which is often error-prone and subjective. See details in section 2.3.
- **Parallel R (pR) for high performance statistical computing delivered super-linear scaling.** We have developed parallel R (pR) middleware for an easy-to-use almost-zero-overhead plug-in of parallel analysis functions written in compiled languages into a widely used open source R statistical environment. pR delivered a super-linear scalability in terms of the number of processors and improved the performance of the state-of-the-art technology by an average factor of 37. Its RScaLAPACK library is distributed as an RPM package across different Linux distributions and in more than 30 countries world-wide through the R's CRAN distribution site. Parallel R forms a server-side analysis engine, with a select set of analysis routines in the Dashboard web application. The initial set of routines was identified based on their frequent usage by climate and fusions communities. pR is discussed further in Section 2.1.
- **A new tool for automatic discovery of front detection developed.** We created an analytical methodology for automatic discovery of turbulent patterns, namely front detection and tracking, both in space and time, in the electrical potential fluctuation by plasma turbulence data from the XGC fusion simulations. The tool uses Automatic Parallelization of Data-Parallel Statistical Computing Codes (see details in section 2.1).
- **Orders of magnitude (100000 fold) speedup achieved for All Pairs Similarity Search (APSS).** The scalable algorithm was developed with specialized indices and heuristic optimization over data sets with millions of records in high dimensional spaces. We developed an open source library of algorithms for fast, incremental, and scalable all pairs similarity searches (see details in section 2.1).
- **Deeper insights into fusion data achieved.** The initial analysis by the SDM center of coherent structures in Gyrokinetic Simulation of Energetic Particle (GSEP) SciDAC center's fluid data is providing previously unexplored insights into the statistics of the structures in the ion heat flux variable. It discovered that there are some small structures with negative ion heat flux that need further investigation to determine if they are due to noise or physics (see details in section 2.3).
- **ProRata enabled systems biology studies in various DOE energy and environment applications.** We have brought to production our open source ProRata robust statistical software with the GUI for high-throughput quantitative shotgun proteomics. ProRata has been downloaded more than 1,000

times and has been used by the DOE Bioenergy centers and Genomics:GTL projects to predict the composition of the cellulosomal complex for biomass degradation and to perform genome-scale functional annotation of ethanol-producing bacteria, to infer metabolic aromatic compound degradation pathways by hydrogen-producing bacteria, and to understand the composition of complex microbial communities from environmentally-hazardous sites. ProRata is discussed in Section 2.3.

- **The Kepler developers hosted and collaborated with the ITER team.** The SDM Center collaborated with the ITER European Integrated Tokamak Modelling project team at the Institute of Fusion Research, France. This group has selected Kepler, the workflow system developed by the SDM center, for their workflow development, and visited twice to coordinate their work with the Kepler development team. During the visits by ITER project team members in 2007 and 2008, we shared our developments of various components of interest to the ITER teams. Kepler development is discussed in Section 3.1.
- **New workflow reliability and fault tolerance tools were developed in Kepler**. Development of workflow reliability and fault tolerance included development of a new model, and of the specifications and pilot versions of a Kepler-based alternative actor for workflow recovery. (See section 3.5)
- **Automatic provenance capture framework developed.** Run-time provenance capture scripts and automatic data-base feed have been developed to be used with the Kepler workflow system. This includes system provenance on the setup of simulation programs, the workflow provenance, and data provenance that captures the history of each data file. The data provenance is now used to find files of interest and move them to the user machine directly from the dashboard, which is used to monitor simulations, and support remote analysis of simulation data. See details in Section 3.4 and Section 4.
- **Integrated framework for real-time monitoring of large-scale simulations eliminated unnecessary computations.** The SDM center has developed an integrated framework, currently being used in production runs by the Center for Plasma Fusion Edge Simulation (CPES) scientists. The technologies provided by the center include the Kepler workflow system, a dashboard, provenance tracking and recording, parallel analysis capabilities, and SRM-based data movement. The Dashboard now has fast visualization with Web access, and other features, including the ability to compare images from multiple time-steps (shots), and display movies composed from multiple images by the workflow system. This integrated system has been used to perform simulation monitoring in real-time, as well as complex code-coupling tasks. Monitoring includes dynamic generation of graphs and images posted on the Dashboard. In a recent review of the CPES project, all the reviewers gave the project the highest possible grade of "excellent." See details in Section 4.
- **The FIESTA framework applied to a new code for real-time code monitoring**. Multiple fusion codes, as well as codes in other domains have been using workflow technology using our FIESTA framework. The workflow ties the results of the code runs into the Dashboard through the VisIT command line interface, making the results quickly available to scientists (see details in section 4).

1. Storage Efficient Access (SEA)

The core I/O functionality on today's high-performance computing (HPC) systems consists of a collection of I/O software that provides a convenient and efficient interface to the available I/O hardware. The projects in this layer focus on this core I/O functionality, and they have two complementary goals. First, we develop and support a collection of highly-scalable and freely available I/O software components that are used in production applications by scientists, and we actively engage the community to help application scientists better understand how to use these tools. Second, through our interactions with the community we identify specific deficiencies in functionality, performance, and usability that we then

work to address. Successful improvements are subsequently integrated into production releases, ensuring that these benefits are made widely available.

Overall, our work can be placed in four categories, discussed in the following sections:

- File system benchmarking and application I/O behavior
- Parallel I/O infrastructure evolution
- Application interfaces and data models
- Next-generation I/O software technologies

1.1 File System Benchmarking and Application I/O Behavior

The high peak rates of HPC I/O systems simply do not translate into adequate sustained performance for computational science applications. The root cause of this performance gap is the mismatch between the requirements of the system's applications and the capabilities of I/O hardware and software. Systematic evaluation of both I/O system capabilities and application requirements provides much needed insight into the efficient use of existing systems and help guide the design of next-generation I/O systems.

The objective of this work is to study file system characteristics that have significant impacts to the parallel I/O operations and evaluate the relative performance of the file systems available to important SciDAC applications on DOE compute platforms. The performance, functionality, and scalability of MPI-IO, parallel netCDF, and HDF5 are critical for many applications

Accomplishments

We instrumented the Lustre file system to log I/O operations for the purpose of understanding I/O patterns and tracking down I/O performance degradation using common benchmarks. These logs are used to track I/O accesses to individual Lustre servers and observe performance on a finer granularity than possible with simple MPI-IO logs. Using this capability, we have extensively characterized the parallel I/O performance on the Jaguar supercomputer [YOV+07]. We have examined the best stripe sizes over Jaguar and showed that the file distribution pattern across the Data Direct Networks (DDN) storage couplets can dramatically impact performance, resulting in a factor of two difference for certain operations [YOC+08]. In addition, we have also examined the scalability of metadata- and data-intensive operations. Our results have demonstrated that, for parallel file open, the shared file mode has the best scalability compared to the separate file mode. Moreover, we have investigated the performance impacts of parallel I/O techniques for handling small and non-contiguous I/O, including data sieving and collective I/O. We have also documented that, without specific optimizations, the performance of writes is hindered by internal Lustre lock contention [YVO08].

We also obtained the S3D application I/O kernel from our collaborators, Jacqueline Chen at Sandia National Laboratories, Ramanan Sankaran and Scott Klasky at ORNL. S3D is a parallel turbulent combustion application using a direct numerical simulation solver developed at Sandia National Laboratories. The S3D's I/O is originally programmed in Fortran I/O functions and each process writes all its sub-arrays to a separate file at each checkpoint. This approach can result in thousands to millions of files from a single production run.

We implemented three I/O methods, including MPI I/O, parallel netCDF, and parallel HDF5. All three methods write the arrays into a shared file in their globally canonical order. This approach reduces the number of files to one per checkpoint. We tested on Jaguar using up to 30,720 process cores (the total number of cores on Jaguar was then 31,328), and based on our analysis, we demonstrated that it is possible to improve the scalability of a representative application S3D by optimizing its I/O access pattern. The aggregated I/O bandwidth of S3D can be sustained to very large scale. For example, we demonstrated that there is a 15% bandwidth improvement by controlling the file distribution pattern in the

original S3D program. In addition, replacing the original file per process implementation with a shared file can avoid a 49% bandwidth drop across 8192 processes because of the reduction of time spent in parallel file creation [YVO08].

We have constructed two benchmarks, HPIO [CAL+06] and S3aSim [CFL+06], for evaluating the scalability of parallel I/O with overlapping/non-overlapping, and contiguous/non-contiguous patterns. Using these benchmarks as well as BTIO (NASA benchmark), S3D I/O (combustion code), and FLASH I/O (astrophysics code), we evaluated shared-file I/O performance on the Lustre and GPFS file systems. As on the Jaguar system, we observed significant performance degradation due to the file system lock contention. Our experiments show that if the I/O requests are carefully aligned with file system lock boundaries, performance can be improved, in some cases by an order of magnitude or more. See graph in Figure 1.

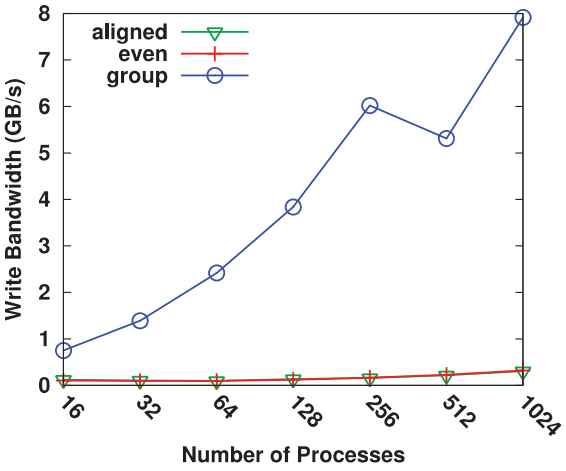


Figure 1: Collective I/O and centralized locking

We have worked with the ALCF team to deploy the Darshan I/O characterization tool, developed under base funding (FWP 56792), on our Blue Gene system. Darshan has been integrated into the software stack such that the I/O of most applications is automatically and transparently characterized. After months of successful use, we selected a two month period and performed a detailed analysis of the workloads seen during this time. We further used the data to identify an I/O-intensive application in need of improvement, the HSCD combustion physics code, and rapidly implemented I/O optimizations for the code. This work won the best paper award at MSST 2011 [CHA+11].

1.2 Parallel I/O Infrastructure Evolution

Multiple parallel file system options are now available, and most HPC systems now include a rudimentary I/O software stack. However, the performance of the I/O stack on many systems is much lower than possible given the hardware available. As HPC systems scale and application complexity increases, extracting the highest possible performance from the I/O hardware is critical to the overall effectiveness of the system. The objective of this work is to improve the state of parallel I/O support for HPC. The Parallel Virtual File System (PVFS) and ROMIO MPI-IO implementations are in wide use and provide key parallel I/O functionality. This work builds on these two components by enhancing them in order to ensure these capabilities continue to be available as systems continue to scale. In addition to improvements to these tools, special attention is paid to Cray systems using the Lustre parallel file system.

Accomplishments

Many advances in the **PVFS parallel file system** [CLR+00] project were facilitated by SDM support. Three PVFS releases were made between 11/2006 and 1/2007, including many bug fixes, Myricom MX and Portals communication drivers, and a new file distribution mechanism. Functionality was also added to PVFS to allow control of layout of files, facilitating research being performed in active storage at PNNL. This functionality was also rolled into a release.

On the research side, we have been implementing tracing in PVFS and ROMIO along with collaborators Aroon Nataraj and Al Malony (U. of Oregon, FASTOS Extreme OS Project) and Kwan-Liu Ma (UC Davis, SciDAC Ultravis Institute) with the goal of performance visualization of the entire I/O software stack on the Blue Gene/P system. Initial visualizations have been generated.

Because of the central role Cray systems play in DOE compute infrastructure, we have placed extra emphasis on understanding and improving I/O performance on the Cray XT systems, in particular when using the Lustre file system. To better understand I/O on the system, we built an alternative MPI-IO package for the Cray XT and Lustre, starting with the ROMIO implementation, called the **Opportunistic and Adaptive MPI-IO Library over Lustre (OPAL)** [YVC07, YVC+07]. Using OPAL we have profiled the internal processing of collective I/O operations on the Cray XT and uncovered opportunities for improvements. We developed a technique called partitioned collective IO (ParColl) that partitions parallel processes into subgroups, each carrying out smaller, yet aggregated I/O operations. In doing so, this partitioning reduces the costs of collective operations, and improves the scalability of collective IO by as much as 416% on 1024 processes [YV08]. Moreover, experiments on Linux clusters suggest that the technique can be applied to a wide range of platforms. We have collaborated with the National Center of Computational Sciences and have deployed our Cray/Lustre optimized library on Jaguar as a contributed alternative package at Oak Ridge National Laboratory, and we have worked with the MVAPICH team to have our specific improvements integrated into this popular distribution.

Through collaboration with the Argonne Leadership Computing Facility (ALCF) we have ensured that the I/O system on the Blue Gene/P system will meet performance and reliability goals. This includes aiding in the specification of the storage hardware, porting and deployment of PVFS at large scale, and working with IBM to solve a significant functionality problem in early versions of their MPI-IO software for the system. We implemented a lock-free driver for the Blue Gene that enables PVFS use, improved the scalability of some metadata operations, and integrated IBM's changes back into the ROMIO source tree.

We incorporated successful research efforts into the production **ROMIO MPI-IO library** [TGL99], including Lustre-specific improvements, file domain, and strided I/O optimizations. We have also developed additional test cases to exercise new code paths in the IBM Blue Gene/P MPI-IO implementation resulting from their work to fix limitations due to a 32-bit pointer on the system.

We have also completed prototypes of MPI-IO atomic mode and shared file pointer operations using MPI RMA and point-to-point communication. These techniques provide an option for avoiding file system locking overhead once MPI-2 features become available on leadership computing platforms [LRT07].

To further address the drawback of using block-based file lock protocol, we have implemented a scalable distributed lock management method that provides true byte-range locking granularity. We used S3D I/O and S3aSim benchmarks to evaluate several lock strategies, and observed the improvement of locking throughput up to between one to two orders of magnitude in performance. We also developed a hybrid two-phase locking protocol to improve the non-contiguous I/O performance [CRL+07].

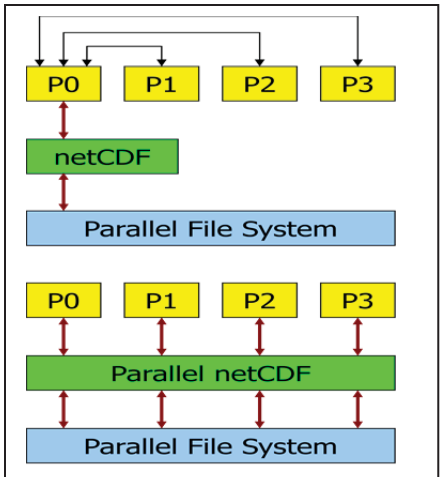


Figure 2: Parallel netCDF

1.3 Application Interfaces and Data Models

In order to make applications more nimble with respect to their I/O behavior, more effort must be spent on the applications and the interfaces that they use to interact with the I/O system. The objective of this work is to improve the usability and observed

I/O throughput for applications using parallel I/O by enhancements to or replacements for popular application interfaces to parallel I/O resources. This task was added in response to a perceived need for improved performance at this layer. Because of its popularity in the scientific community we have focused on the NetCDF interfaces, and in particular on a parallel interface to NetCDF files.

Accomplishments

Significant work has gone into making the **Parallel netCDF (PnetCDF)** [LLC+03] software ready for production. The original idea of PnetCDF is illustrated in the Figure 2. Instead of having all processors communicate with the parallel file system though a single processor, PnetCDF lets each processor communicate directly with parallel file system.

We moved to using SVN and Trac to manage the Parallel netCDF source tree, facilitating greater community involvement. We also improved the software to better operate on Blue Gene/L, Blue Gene/P, and SiCortex systems. Support for increasingly large datasets has become a critical issue for PnetCDF. The original UCAR netCDF format supports variables up to 2 GBytes (due to 32-bit limitations) without special work, but our users are beginning to surpass this limit. We have developed an extension (that uses 64-bits for sizes), named the "CDF-5" format, that allows us to store variables of effectively unlimited size. We have synchronized these changes with the serial netCDF team so that serial tools can interoperate.

PnetCDF is now used in several production codes. Recently, it has been successfully used by the large-scale Global Cloud Resolve Model (GCRM) and the community climate system model (CCSM). GCRM is the global atmospheric circulation and climate simulation application developed at Pacific Northwest National Laboratory and the Colorado State University. CCSM is a climate simulation framework that consists of component models such as atmosphere, land, ocean, and sea-ice, developed at the University Corporation for Atmospheric Research (UCAR). Recently, an optimization is developed to enable data aggregation for multiple, small-sized requests that can better utilize I/O bandwidth on modern parallel computers [GLC+09]. In collaboration with PNNL, we also developed a new I/O method based on this new optimization into GCRM and conducted a performance evaluation on the Cray XT4 parallel computer at NERSC. A significant performance improvement is observed over the current best I/O method. With the total data amounts of 15, 61, and 243 GB for 640, 1280, and 2560 processes respectively, Figure 3 shows up to 140% improvement in term of write bandwidths. The file system peak performance at the time these experiments were carried out had 12 GB per second write bandwidth. These results were presented in the Workshop on High-Resolution Climate Modeling 2010 [PSA+10].

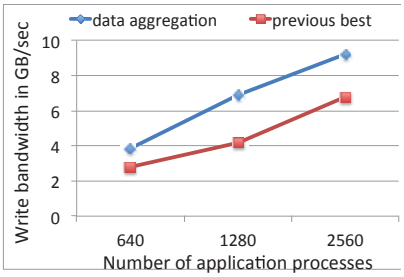


Figure 3. Evaluation of GCRM I/O performance with Parallel NetCDF data aggregation.

A prototype of **data analysis capabilities in PnetCDF** has been implemented. We are implementing this new data analysis API on top of the exiting PnetCDF data access API to provide users with range query, statistical and other content-based data analysis functions on netCDF datasets, without any detailed understanding of underlying data organization from end users.

We continued our collaboration with V. Pascucci (U. Utah) on parallel I/O in their IDX multi-resolution data format, extending the support and performing a study with J. Chen (SNL) team's S3D code. This data format facilitates interactive visualization of very large scale data sets. This work is being completed with a potential paper target of Cluster 2011. A paper detailing early results was presented at the Parallel Data Storage Workshop (PSDW), held in conjunction with SC10 [KPV+10].

1.4 The Adaptable I/O System (ADIOS)

Over the last decade there has been a substantial increase in the capability of leadership computing systems, with the Top500 machines seeing a many-fold increase from a few TFlops to more than a PFlop for the current generation of HPC systems. This increase in compute power has not been mirrored in I/O performance, which instead has suffered by comparison and even decreased in absolute terms. This imbalance, coupled with the increase in data volumes produced by applications, has resulted in a new bottleneck in the performance of these systems.

The Adaptable I/O System (ADIOS), was developed as a grass roots effort by the computational science community to address the issue of making I/O "easy-to-use", and allow application scientists to write extreme amounts of data (PB's/day) with very little impact on their calculations.

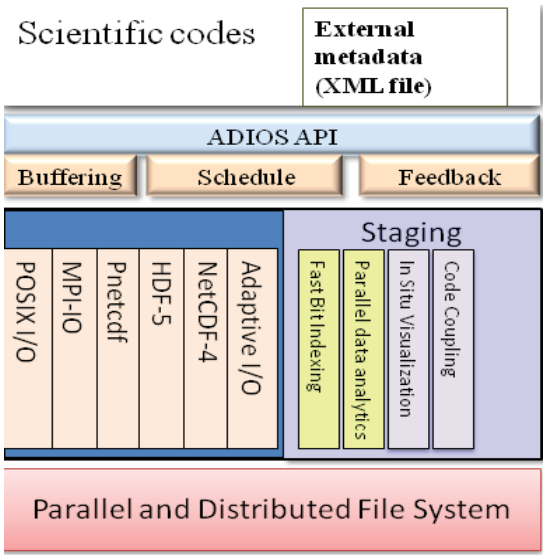


Figure 4. The ADIOS abstraction layer.

ADIOS research is driven mostly by the performance requirements of application scientists, where the adoption of new technologies is usually constrained due to the lack of usability in achieving the desired performance in an easy-to-use system. ADIOS provides an easy-to-use programming interface, which is as simple as Fortran file I/O statements. The conceptual architecture of ADIOS and its abstraction layer is shown in Figure 4. Abstracting the I/O metadata information and data structures from the source code into an external XML file can reduce code pollution and create the connection between high-level APIs and the underlying I/O implementation details, as well as other technical descriptions, such as buffering and scheduling, and asynchronous data movement. Because of this separation, different research groups can write I/O methods in ADIOS to allow for extreme performance, as shown in Figure 5. For example, the previous I/O implementations in the S3D simulation and the SCEC code, running on 96K cores, and 30K cores on the XT5 at ORNL, produce over 10X performance increases compared to collective MPI-IO. ADIOS also allows the output to be in the ADIOS-BP format, as well as HDF5, and parallel Netcdf-4, and ASCII. The ADIOS-BP format was created for extreme scale computing, and allows for extreme performance in reading data on HPC systems. In recent results, it often achieves 2 – 100X performance improvements in reading performance.

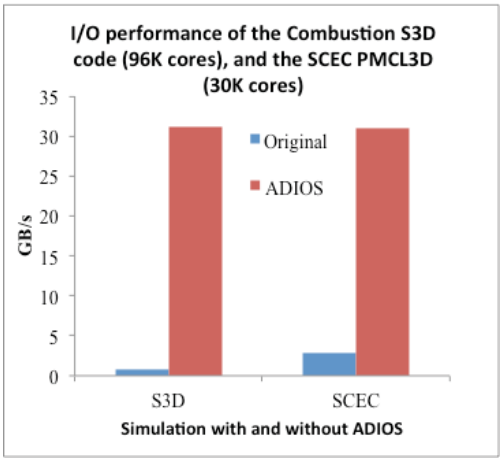


Figure 1: ADIOS Methods demonstrate optimized performance across multiple applications

1.5 Next-Generation I/O Software Technologies

Some challenges for future I/O systems call for the development of altogether new software technologies. One focus for software development is on the creation of a collaborative file caching system for use in HPC environments. Such a system would take advantage of small portions of memory on a collection of machines to generate a cache of sufficient size to enable aggregation and reorganization of I/O operations from HPC applications. A second focus of our work is in improving the analysis capabilities of HPC systems. A promising technology for improving analysis is **active storage**, which provides the ability to perform data processing on the storage nodes of modern file systems. We will discuss our successes in both of these areas in this section.

Accomplishments

A **distributed collaborative I/O caching system** has been implemented [LCC+07, LCC+07b]. A global, distributed metadata cache maintains state on where data is located in the cache, and a least recently used (LRU) policy is adopted for cache eviction. To enforce the atomicity of all file system read and write calls, we adopted a two-phase, multiple-reader single-writer, locking policy. Locks are managed by the entity responsible for metadata on that region. In some specific cases, where it is semantically correct, we enable multiple writers to improve concurrency. For testing and evaluation purposes, we have hooked the caching system into ROMIO for both the UFS and PVFS drivers, enabling testing on a wide variety of parallel file systems.

Two mechanisms for collaborative caching have been implemented. The first relies on I/O threads to drive the caching system. In this mode additional threads, spawned on compute nodes, are responsible for caching and I/O. The second mechanism relies on MPI-2 process spawning calls. This approach can be used to set up a distinct set of caching processes, called "cache servers" (where this MPI-2 functionality is available).

Based on the observation that the majority of the I/O patterns presented in scientific applications are write-only and do not require any software control for parallel data consistency or concurrent access atomicity, we developed a two-stage write-behind method for improving the performance of parallel write-only operations [LCC+07a]. This approach redistributes and combines non-contiguous file requests between processes to generate large contiguous I/O requests. Similarly, we have experimented with using POSIX asynchronous I/O routines as a mechanism for flushing dirty cache data during idle periods.

Testing and development were performed on several parallel machines: Tungsten running Lustre file system and the IBM cluster running GPFS file system at NCSA, Jazz running PVFS file system at ANL, and Ewok running Lustre at ORNL. We use several I/O benchmarks for performance evaluation: the NASA BTIO benchmark, the FLASH application I/O kernel, and S3D application I/O kernel. Our experiments show that with 3% of compute nodes allocated as I/O delegates, I/O bandwidth improvements range from 40% to 250%. When the number of I/O delegates is 10% of the application nodes, we observed up to 500% improvement [NLC+08].

Starting with an initial kernel-level **active storage prototype** [FFG+05], we identified several factors that degrade performance and investigated synchronization options between remote client and active storage processing component. To address some of these performance issues with the kernel space implementation, we developed a new user space implementation of active storage. Based on preliminary experiments and experience, it appears to be faster and more flexible than the original version, and it is also more portable: it operates on both Lustre and PVFS. We have decided to focus our efforts on the new user-space approach. We have evaluated this prototype [PN07], focusing on its use in out-of-core applications.

We are working on performance model for active storage. The purpose of developing such an analytical model for active storage is to enable prediction of expected performance benefits before deploying in a user application. The performance model we are developing includes system parameters: network bandwidth, number of server and compute nodes, disk bandwidth and information about the compute nodes user parallel job would run on.

Our recent work has focused on developing support for striped files [PN08]. Our previous implementations have been unable to deal with striped files, i.e., files whose data is spread across several nodes. Striping files is typically performed in parallel file systems to improve aggregate I/O bandwidth. In the current design we launch a processing component per storage node used by the matching file, and make every processing component process only the file chunks stored in its own node. If the processing components write to an output file, the output file will have to be created with the same striping pattern as the input file, and every processing component will have to write to only the file chunks stored in its

node. The experimental results on a Lustre file system show that our implementation for striped files can reduce the network traffic to near zero.

In addition, we have been working on adopting Active Storage to deal with data files with specific complex formats, such as netCDF or HDF5, that are very common for data exchange in some scientific applications. In particular, in collaboration with the SciDAC GCRM (global cloud resolving model) project we have been working with netCDF climate data.

1.6 Outreach

We take outreach very seriously. We have presented 11 tutorials on topics related to storage and parallel I/O in the last three years, including six full-day tutorials at the SC conference series. We organized and held a symposium at ANL for PVFS researchers and developers, helping everyone catch up with each other's work and coordinate future efforts. We actively participate in DOE Exascale workshops and other application-oriented meetings to help educate the community on I/O best practices.

We have worked closely with Garth Gibson of Carnegie Mellon University and the SciDAC Petascale Data Storage Institute to help in using PVFS as the foundation for class projects in parallel file systems. So far PVFS has been used in two courses as the basis for work in distributed directory storage in high-performance file systems and in alternative data organizations on local storage. Both efforts have resulted in student publications [PGL+07, PST+08].

We presented the keynote talk, "Storage in an Exascale World," at SNAPI 2010 and presented "Making the Most of the Software Stack" at the TACC Extreme Scale I/O and Data Analysis Workshop. We also started the Interfaces and Abstractions for Scientific Data Storage workshop series in 2009, with the third workshop to be held later this year.

2. Scientific Data Mining and Analysis (DMA)

The Data Mining and Analysis (DMA) layer provides the data-understanding technologies necessary for efficient and effective analytics of complex scientific data. This is accomplished through the development and deployment of the three core technologies:

- High performance parallel statistical computing
- Efficient searching and filtering in data-intensive scientific applications
- Feature extraction and tracking for scientific applications

2.1 High performance parallel statistical computing

Our overarching goal was to scale-up the existing data analysis tools, both in capability and capacity, to bridge the gap between scientific data production and data processing rates—a well-recognized need by the DOE scientific applications [Mou04]. The challenge is that *many of the current data analysis routines are written in non-parallel languages such as R, Matlab, and IDL and do not scale to massive data sets*. To address this challenge, our approach was based on using a statistical package, called *R* [IG96]. We provided a middleware between R's high-level data analytics language and a generic, parallel, optimized, portable computational analytics engine. We further used those tools to provide application scientists with the targeted knowledge discovery capabilities for complex scientific data, such as:

- *Analysis of fluctuation energy distribution*–the observed composite energy signal is distributed in space and time. We focused on discovery of turbulent patterns in the $dphi^2$ XGC energy data, where $dphi^2$ is the square of electrical potential fluctuation by turbulence.

- *Mapping and tracking the evolution of turbulence in space and time*–we developed a methodology for *multi-resolution* (across space and time) analysis of turbulence, especially, through front-tracking to establish such dynamics.

Accomplishments

Parallel R (pR) for high performance statistical computing delivered super-linear scaling

The goal of this activity was to provide an easily extensible mechanism to add third party parallel analysis capabilities to *R*. We advanced our prototype *Parallel R (pR)* system on several performance metrics. The ***RScaLAPACK*** library [YSB+05] was matured to a production level, and distributed as an RPM package of various Linux distributions [RPM09] as well across >37 mirror sites through R's CRAN network [YSB+05]. *RScaLAPACK* replaced many data analysis routines in *R* that are based on LAPACK [ABD+99] linear algebra solvers with the corresponding parallel, optimized and portable ScaLAPACK [BCC+97] routines.

We extended *RScaLAPACK* to provide a lightweight, easy-to-use ***pR*** middleware interface [SAC+07, POS07] that bridged the R statistical environment with parallel third party data mining and analysis libraries written in compiled languages such as MPI C/C++/Fortran. Similar to RScaLAPACK, the *pR* interface to parallel analysis functions mimics the serial functions' interfaces, which is a single function call. In contrast to *RScaLAPACK*, which was designed specifically for *ScaLAPACK*, the *pR* architecture was abstracted to allow parallel third-party analysis functions to integrate with the *R* environment without requiring major modifications to either *pR* or to the external third-party libraries. In addition, *pR* offered the efficiency and scalability of the underlying parallel third party analysis codes, with a few microseconds overhead induced by *pR* middleware. In contrast to *R*, its performance on a single processor has been improved by a factor of 30, and it has shown a super-linear scalability compared to a hypothetical ideal scaling with the number of processors if *R* were run in parallel. Unlike competing technologies such as *Rmpi* [Yu09], which is an *R* wrapper around C++ implementations of MPI, *pR* offers a number of advantages: (a) its average factor of execution time improvement is 37; (b) it does not require the *R* end-user to have knowledge of parallel computing; and (c) it enables one to employ third-party parallel data mining codes, unlike *Rmpi* that requires re-implementation of these codes in *R* scripting language with potentially significant performance degradation besides being an error-prone process. *pR* is available upon request from samatovan@ornl.gov.

We also extended the capability of RScaLAPACK library to support openMPI back-end in response to multiple users' requests, eased RScaLAPACK's installation via improved autoconf, provided processor grid manipulation routines, provided both static and dynamic MPI library support, etc. The updated version was released on the R's CRAN web-site.

In addition to data parallelism enabled by *pR*, we explored task parallelism support in R. By leveraging our *taskPR* [SBY05], we pursued two complementary directions: (a) embarrassingly parallel execution mode and (b) automatic out-of-order execution mode by applying compiler parallelization approaches to automatically parallelize *R* scripting codes without requiring end-user modifications [MLS07, LMY+07, LMY+11].

We also made parallel analysis capabilities available as part of more complex scientific workflows (see Section 3). We pursued this effort along the three complementary directions: (a) to enable analysis capabilities with the Dahsboard Web application (see Section 4); (b) to provide web services based parallel statistical computing capability within Kepler workflow engine; and (c) to enable social network based knowledge annotation [BGKS08]. These efforts were in collaboration with the SPA team. Leveraging *pR*'s analysis routines, the Dashboard web application (see Section 4) was upgraded to incorporate a data analysis layer. We evaluated the analysis functions usage by SciDAC scientists, including fusion and climate scientists, and designed a library of analytical routines to expose on the Dashboard.

Web Services have become a critical part of today scientific research domain. They support loosely coupled service oriented architecture to provide scalable interoperable systems. Since Kepler has been a very effective tool for managing scientific workflows, we extended data analysis capabilities of Kepler to web-services enabled data analysis. This new feature of accessing data analytics routines through a web-service enabled users to benefit from the functionality and high performance of parallel *R* as part of a more complex scientific workflow. We co-taught the SPA's Supercomputing 2007 tutorial and presented a module on the remote *R* enabled Kepler actor.

A third activity was to deploy this parallel analysis infrastructure across various DOE scientific applications. Working with a number of SciDAC application scientists, we observed that one of the key obstacles is the lack of infrastructure that couples data analytics capabilities with data management infrastructure. For example, the growing amount of climate data is currently made available for download through the DOE Earth Science Grid (ESG) portal, yet the analysis of this data is typically done on each institutional infrastructure. Moreover, the current ESG portal provides a convenient interface for users to download small subsets of data, but is not designed for large-scale data transfers. We addressed this need through the ESG Download Project (EDP). EDP provides a convenient mechanism for ESG customers to generate results leveraging massive amounts of previously inaccessible data. We worked with Marcia Branstetter (ORNL) from the DOE Climate Modeling SciDAC, and the ESG group, and applied the Data Mover Lite (DML) tool used in ESG. The SDM center produced a simple command-line interface that invoked the use of the ESG portal meta-database and the DML tool to enable large-scale data transfers of the climate simulation data from the Earth System Grid (ESG) portal to ORNL systems.

Automatic Spatio-Temporal Turbulent Front Detection and Evolution in Fusion Plasma

Few would argue that fusion energy has been the Holy Grail of renewable energy efforts. The grand challenge is to produce more energy through a fusion reaction than that required to initiate the process in a reactor. A key bottleneck is the turbulence, or unstable motion, of the fusion plasma. Turbulence influences the degree of energy lost by plasma during the fusion process; therefore, controlling the turbulence is critical to viable energy production. Discovery of dynamic turbulent patterns and trends from the data produced by a computer-simulated fusion reaction offers a potential to reveal ways to control the turbulence. Yet, it presents a challenge: how to effectively and efficiently analyze the massive amounts of data, which is inherently complex, noisy, and high-dimensional. To address this challenge, we created an analytical methodology for automatic discovery of turbulent patterns, namely front detection and tracking, both in space and time, in the electrical potential fluctuation by plasma turbulence data from the XGC simulations (see Figure 6) [SSC+10]. This work was conducted in collaboration with Dr. C.S. Chang, NYU. This process can potentially predict the structure, dynamics, and function of fusion plasma turbulence. It could also enable similar analyzes required in other disciplines, such as astrophysics and oceanography.

One strategy was the one of *reduced, yet informative, data representation* for the target data analysis task. For a fixed time-step, t_0, we approximated with line segments in a spatial region around the point of interest, r. The points corresponding to the fronts are the points, where the line segments change their slope from the direction almost parallel to the *x*-axis (green) to the direction almost parallel to the *y*-axis (blue and red) (see Fig. 6.a). Such an approach only required the slope and intercept of the approximating line segments for a few sequential sliding windows. The other steps of the end-to-end front detection and tracking process (see **Error! Reference source not found.**b) have been local, by nature, and have utilized *pRapply*() method for a multi-node multi-core parallel execution with an ideal speed-up, as described next.

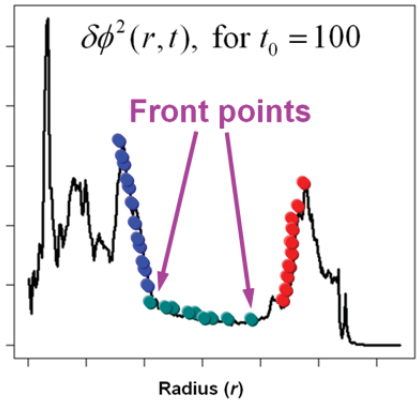


Fig. 1. Front detection and tracking in fusion simulation data.

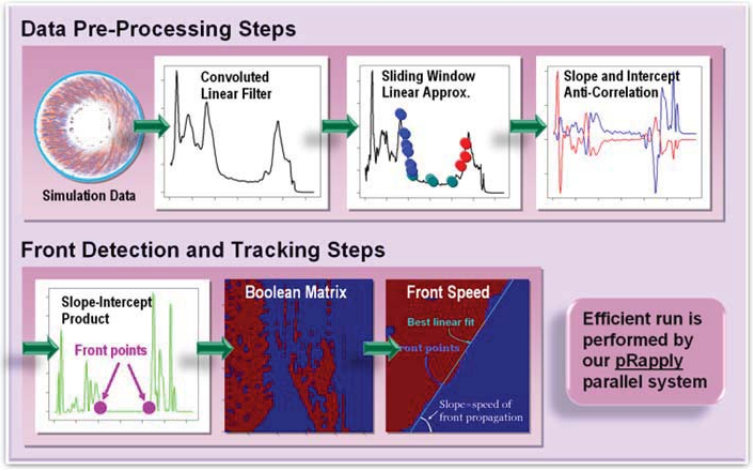


Fig. 2. A multi-step turbulent front detection and tracking process

Automatic Parallelization of Data-Parallel Statistical Computing Codes with pRapply in Hybrid Multi-Node Multi-Core HPC Environments

We extended the *pR* middleware [BH+09] with *pRapply*() function for the *R* open-source statistical environment to support automatic parallelization of data-parallel tasks in multi-node, multi-core, and hybrid environments [BK+09, SKB+10]. *pR* requires few or no changes to existing serial codes, offers a linear speed-up with the increasing number of processors, and yields over 50% end-to-end execution time improvements in our tests, compared to the commonly used *snow R* package. We released *pRapply*() as open source software.

Fast, Incremental, and Scalable All Pairs Similarity Search

Searching pairs of similar data records is an operation required for many data mining techniques like clustering and collaborative filtering. As the scale of the data has been increasing to several millions or billions of records in a high dimensional space, enabling fast and incremental similarity search over such data sets has become a formidable task. To address this challenge, we developed an open source library of algorithms for fast, incremental, and scalable all pairs similarity searches through improved indexing, systematic heuristic optimizations, and parallelization.

First, we designed a sequential algorithm for all pairs similarity search (APSS) that involves finding all pairs of records having similarity above a specified threshold. Our proposed fast matching technique speeds up APSS computation by using novel tighter bounds for similarity computation and indexing data structure. It offers the fastest solution known to-date with up to 6X speed-up over the state-of-the-art existing APSS algorithm.

We further addressed the incremental formulation of APSS problem, where APSS is performed multiple times over a given data set while varying the similarity threshold. The goal is to avoid redundant computations across multiple invocations of APSS by storing history of computation during each APSS. Depending on the similarity threshold variation, our proposed history binning and index splitting techniques achieve speed-ups from 2X to over 10^5 X over the state-of-the-art APSS algorithm. To the best of our knowledge, this is the first work that addresses this problem.

Finally, we designed scalable parallel algorithms for APSS that take advantage of modern multi-processor, multi-core architectures to further scale-up the APSS computation. Our proposed index sharing technique divides the APSS computation into independent tasks and achieves ideal strong scaling behavior on shared memory architectures. We also propose a complementary incremental index sharing

technique, which provides a memory-efficient parallel APSS solution while maintaining almost linear speed-up. Performance of our parallel APSS algorithms remains consistent for datasets of various sizes. To the best of our knowledge, this is the first work that explores parallelization for APSS. We demonstrate the effectiveness of our techniques using real-world million record data sets.

2.2 Efficient searching and filtering in data-intensive scientific applications

In the DOE data management workshop report [Mou04], most applications articulated critical needs to efficiently find important data based on search conditions over data values. This section addresses the technology for efficient searching and filtering of large-scale scientific multivariate datasets with hundreds of searchable attributes to deliver the most relevant data to the appropriate analysis tools, such as those in Sections 2.1. More specifically, our goal is to develop a scalable searching tool for scientific data and to integrate this tool with data analysis and data storage environments.

Accomplishments

We have developed an extremely efficient indexing technology for accelerating database queries on massive datasets, called **FastBit**. The FastBit indexing software grew out of years of basic research in scientific data management. The core of the software is a set of novel bitmap indexes that are optimized for scientific data. For the majority of scientific data, the existing data is not modified. *FastBit* indexes take advantage of this fact and optimize bitmap indexes using a combination of techniques, described briefly below. We have shown that *FastBit* is theoretically optimal and can answer queries 10 – 100 times faster than the most popular commercial indexing methods. There are three sets of orthogonal techniques in *FastBit*: binning, encoding and compression. The binning techniques allow us to work with scientific data that contains a very large number of different values. Performances of other bitmap indexes deteriorate significantly as the number of distinct values increases; we have developed a bin-based clustering technique to overcome this shortcoming [WSS08]. This allowed us to deal with scientific data values that are typically integers or floating point numbers. A significant part of the performance advantage stems from a new compression technique we have invented and patented, called Word-Aligned Hybrid (WAH) compression that provides a factor of 10-20X performance improvement over any known bitmap indexing method. Further improvements were achieved by a number of advanced bitmap encoding techniques that determine how the data is represented in the bitmaps. We have explored and developed a combination of two-level encoding methods that led to an additional five-fold improvement [WSS07]. The combined effect of these methods has produced unprecedented speedup in searching very large datasets, permitting certain applications, such as visual analytics, to perform searches interactively in real-time. We have compared FastBit with other design choices of implementing bitmaps, and have shown its superiority to other methods [OOW07]. In 2008, the FastBit software was recognized with an R&D 100 Award for its innovative technologies and its contribution to the wider community.

FastBit has been enhanced so that various operations important to exploration of scientific data can be performed efficiently on the bitmaps directly. In addition to cell identification based on specified conditions on the variables, FastBit can perform Region Growing (i.e. connecting neighboring cells into regions), and Region Tracking (i.e. tracking the evolution of regions over time). An example of applying this technique to combustion data is shown in Figure 7. The top left of the figure shows the progress of "flame fronts" in a combustion simulation.

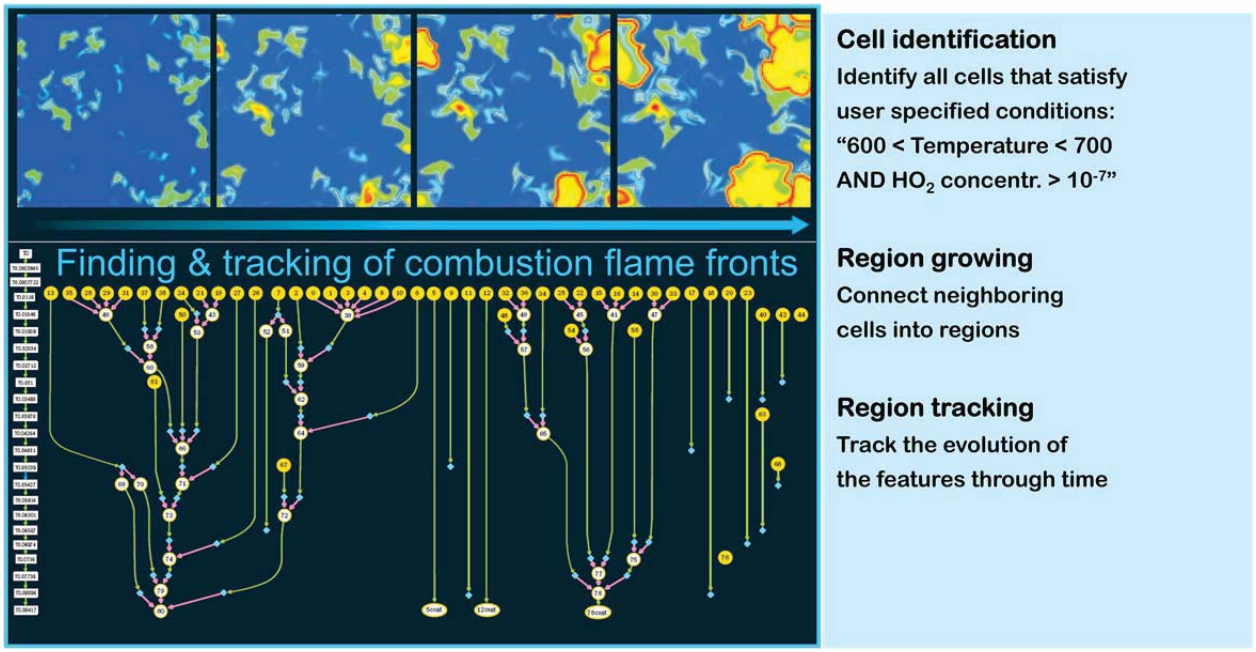


Figure 7: Using FastBit indexes to identify flame fronts in a combustion simulation

By applying Cell Identification (conditions that identify the flame front), region growing, and finding overlaps of regions to identify region growing, it was possible to construct quickly the graph below. This graph shows how combustion kernels (yellow dots on the top) combine over time (going down in the diagram) to form flame fronts.

FastBit has been used in a number of applications. The release of FastBit in 2007 as open source, and the nurturing and support of users' group, has facilitated outreach to various application domains. We describe below a couple of representative applications that use the FastBit technology.

- ***Query Driven Visualization.*** One of the most important applications of FastBit is the Query Driven Visualization (QDV) developed in collaboration with the SciDAC Visualization and Analytics Center of Enabling Technologies (VACET). In a number of on-going projects, FastBit is used as the software to speed up their data accesses. An example where FastBit is used is in the tracking of particles in the simulations of the Laser Wakefield Particle Accelerator [RPW+08]. In this particular test, the data set contains about 90 million particles. Using FastBit to track as many as 25 million particles is much faster than using the custom code that directly works with the raw data. Typically, much fewer particles are tracked in an analysis. As the number of tracked particles decreases, the relative advantage of FastBit increases because a smaller part of the bitmap indexes is needed in order to locate the particles. In this test, using FastBit is as much as 10,000 times faster when tracking a few hundred particles, a use case common in most realistic data analysis scenarios. Even if the number of particles tracked is a few millions, FastBit performs 1,000 times faster than scanning the data. This kind of performance is essential for real-time interaction, which cannot be achieved by using analytical software packages such as IDL. Figure 8 is a schematic diagram showing the use of FastBit in an interactive manner to select subsets of particles by specifying condition, to compute histograms, and to track particles over time steps.

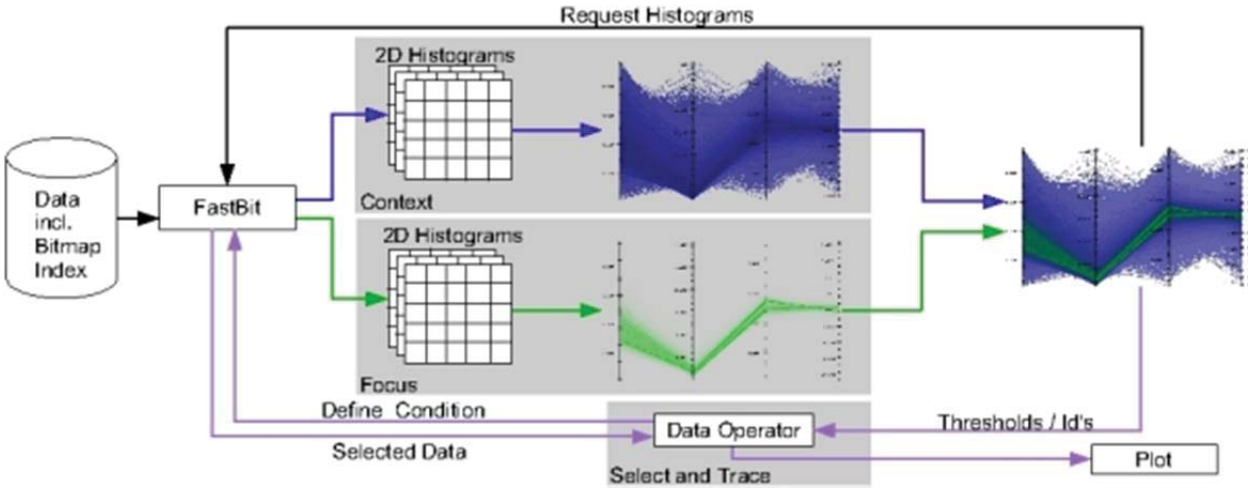


Figure 8: Use of FastBit indexes for exploring particle properties in an accelerator design

- ***Indexing of toroidal meshes.*** We have also been working with a fusion application to accelerate the exploration of coherent spatial objects known as regions of interest. By incorporating the special mesh structure used in the gyrokinetic transport code (GTC) for simulating magnetically confined fusion plasma, we are able to turn the FastBit search results into regions of interest much more efficiently than earlier approaches. Preliminary testing shows that the new approach can identify regions nearly 1000 times faster than the commonly used approach. A paper on this work was recently published [WSJ+11].
- ***Molecular Docking.*** Since the release of the FastBit software, a number of independent efforts using FastBit have sprung up around the world. The use of FastBit to accelerate molecular docking in TrixX-BMI is a striking example of such an effort. Jochen Schlosser and Matthias Rarey from University of Hamburg applied FastBit searching capability to match the ligands with cavities in proteins (i.e., molecular docking). In their tests, they found that the version using FastBit is on average 12 times faster than the previous version of the molecular docking software, which is a widely used in commercial product. When FastBit is used to process additional pharmacophore constraints, TrixX-BMI is measured to be hundreds of times faster. Other applications where FastBit has been applied include combined numerical and text queries [SCW+08], and data warehouses and OLAP [SW07].

Another technology developed and deployed is **efficient indexing and I/O strategies for analytical processing of scientific datasets.** Unlike basic searching techniques, such as FastBit, we address here the scenario where the query is an analytical request. For example, in the biology domain, the query sequence needs to be compared (via compute-intensive sequence alignment algorithm) against the database of known sequences so that ``similar'' sequences to the query sequence get retrieved. The exponentially growing size of the database of records and the complexity of analytical requests present a significant database management challenge. In SciDAC-I, we prototyped ***mpiBlast-pio*** library that took advantage of parallel and collective I/O techniques in Blast, a widely used NCBI bioinformatics sequence analysis tool, to scale *mpiBlast* [BLAST09] to thousands of processors [LMC+05]. In SciDAC-2, we robustified *mpiBlast-pio* and brought it to the production level open source software distributed from the *mpiBlast* web-site. We also developed an open source framework for fast matching and dynamic indexing for all pairs similarity searches. Our techniques achieved >6 times speed-ups compared to the state-of-the-art algorithms, while saving up to 32% memory requirements. The incremental version in response to threshold changes provided an additional 5x improvement.

We also focused on **adaptive scheduling technologies for parallel, scientific, data-intensive, web services-based analytical query processing.** Scientific web services often possess data models and query workloads quite different from commercial ones and are much less studied. Individual queries have to be processed against portions of a large, common dataset in parallel by multiple server nodes. Existing scheduling policies from traditional environments (i.e. cluster web servers and supercomputers) consider only the data or the computation aspect alone and are therefore inadequate for this new type of both data- and compute-intensive workload. We systematically investigated adaptive scheduling for scientific web services, by taking into account parallel computation scalability, data locality, and load balancing [LML+08, LML+07]. We demonstrated that intelligent resource allocation and scheduling are crucial in improving the overall performance of a parallel sequence database search server. Also, no single static strategy works best for all request workloads or all resources settings. In response, we developed several dynamic scheduling techniques that automatically adapt to the request workload and system configuration in making scheduling decisions. Our experiments show the combination of these techniques delivers a *several-fold* improvement in average query response time across various workloads.

2.3 Feature extraction and tracking for scientific applications

As the data from scientific simulations, observations, and experiments approach the petascale and beyond, we need to extract features to fully realize the benefits of our advanced computational and data collecting abilities. This area focuses on the development and application of analysis techniques to data from scientific simulations, observations, and experiments. We use techniques from several disciplines, including image and video processing, machine learning, statistics, and pattern recognition, to find useful information in massive, complex data sets [Kam08b]. Our goal is two-fold – to use data mining techniques to understand scientific phenomena and, as appropriate, to deploy our solution for use by application scientists. We have worked with a number of application projects, initiated by the SDM center or at the request of the domain scientists.

Accomplishments

The problems we focused on were driven by applications scientists. Each problem presented different challenges, and required different techniques. The challenge was not only to discover the combination of techniques that addressed the problem at hand, but also to discover new approaches for previously unsolved problems. This could only be achieved by working closely with the application scientists, understanding their problems, providing solutions, and iterating the process. We had great success in addressing several problem classes as described below.

Poincaré plots are an important tool for understanding data in Fusion science applications. A Poincaré plot is composed of orbits, each of which consists of several points created when a particle, moving around a toroid, intersects a poloidal plane. The shape of the orbit depends on the starting point of the particle. Our task is to assign an orbit to one of four classes – island chain, quasiperiodic, separatrix, and stochastic. This is currently done manually, a process which is tedious, error-prone, and subjective. Our early work, based on fitting second order polynomials to the points, appeared promising, but was sensitive to the choice of parameters, and did not easily extend to stochastic orbits, and did not lend itself to a simple extraction of rules for classification. Another approach, called KAM, proposed in the context of dynamical systems was suggested by our collaborators, Neil Pomphrey and Don Monticello (PPPL). This used graph-based features to represent the orbit, which was classified using heuristic rules. Our experiences with KAM indicated that it was not suitable for the characteristics of our data from simulations, which were not only noisy, but also had very thin lobes in the separatrix and island orbits. These experiences indicated a solution more suitable for our data. Through extensive experimentation, we **developed a system for classification of Poincaré plots**. This technique extracts robust features which were scale, rotation, and translation invariant. We addressed several major challenges, such as: i) improving the quality of the training data by varying the class labels with the number of points; ii) applying techniques from spatial statistics to identify locally stochastic orbits; iii) incorporating

appropriate scaling to handle orbits with thin lobes; iv) exploiting the alignment of peaks and valleys to capture local variation along the orbit; and v) using wavelet analysis to represent the multi-scale structure of orbits. After several iterations we obtained an error rate of ~4% using a patented algorithm for ensembles of decision trees. To the best of our knowledge, this is the first time an accurate, automated solution has been developed for this problem. Josh Breslau (PPPL) found the code worked quite well and is distributing it to M3D users. As the problem is of broad interest in the fusion community, it is being deployed for use by others at PPPL and MIT. The code will also form part of the workflows being developed by the SPA team. A journal paper summarizing the approach is in progress. An example of the classification of these Poincare plots achieveing 96% accuracy is show in figure 9.

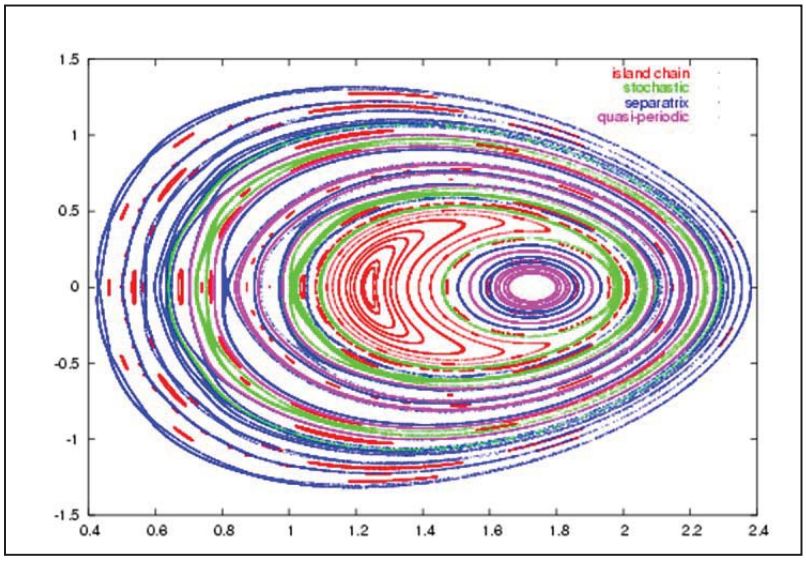


Figure 9: classification of different types of orbits in Poincare plots

A second project involves **Blob tracking in experimental images**. We are working with Fusion physicists Stewart Zweben and Ricardo Maqueda at PPPL to analyze high-speed, high-resolution images of plasma from NSTX to understand edge turbulence. This involves the segmentation, characterization, and tracking of coherent structures, known as blobs, in the image sequences. There are several challenges: the images are noisy; the theory behind edge turbulence is poorly understood and cannot influence what the scientists expect to see in the data; and the images in a sequence are varied, with both bright and faint blobs, as well as bright blobs with extended faint tails. It is non-trivial to come up with an algorithm which, with a single set of parameters, will perform well across all images in a sequence. Using sample images, we investigated several algorithms to de-noise the data, remove the ambient intensity, and identify the blobs. Our results were presented in a paper at the SPIE conference [LK07]. The more robust techniques are now being applied to full sequences to understand how well they handle the variation in the sequence. Once the blobs are identified, they can be tracked relatively easily using a simple overlapping between frames [GK08].

The third project is collaboration with the GSEP SciDAC (Zhihong Lin, PI). The analysis goal is to identify coherent structures in GSEP simulation fluid and particle data and to understand the non-linear interactions between the two. This is difficult as: (i) there is no definition of coherent structures; (ii) they vary extensively over time making it difficult to identify robust algorithms; (iii) the fluid data are on a twisted toroidal mesh while the particle data are unstructured, making existing algorithms inapplicable; and (iv) the data are currently in terabytes, with petabytes expected in the future. In FY09, we had implemented an initial algorithm that used several variables to identify the structures in the fluid data. Discussions with Zhihong Lin and Yong Xiao indicated that the approach also made sense from the physics viewpoint. The data from the smaller Ion Temperature Gradient (ITG) simulation were then analyzed at select time steps to extract statistics on the structures. This simulation has 32 toroidal planes with 40,000 grid points per plane. The initial results for a plane were very interesting, prompting the analysis of all planes at a time step. The preliminary conclusions from this analysis were i) that the event size distribution needs further analysis to confirm the type of distribution and ii) there are some small structures with negative ion heat flux that need further investigation to determine if they are due to noise

or physics. To investigate these issues, a larger ITG simulation was run, with 64 poloidal planes and 600,000 grid points per plane. The results were consistent with the smaller data set.

These results prompted a comparison with the Collisionless Trapped Electron Mode (CTEM) simulation to understand the physics better. The distribution of the ion heat flux in CTEM was different from ITG, requiring a different algorithm to identify the structures. We also found that there were a lot more negative structures which alternated with positive flux structures along a flux surface (see Figure 10). Both the statistics on the structures and a visual tracking of the structures over time indicated that from an analysis viewpoint, these structures were not due to noise.

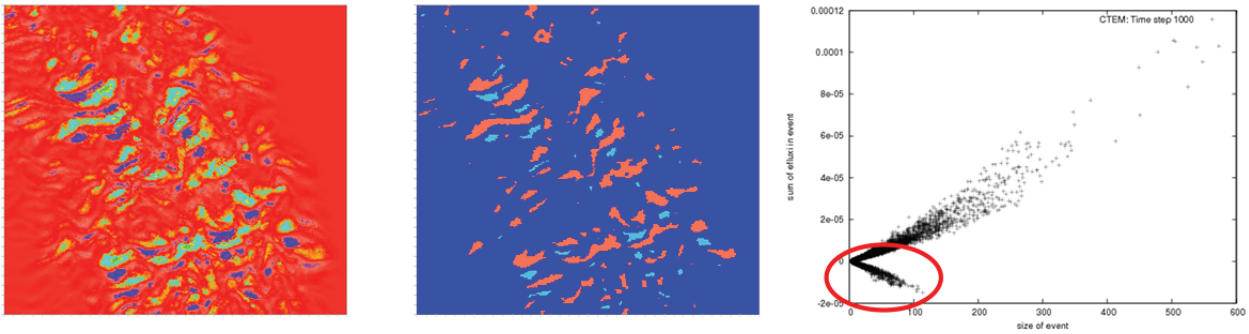


Figure 10: (a) A subset of the data on a2-D poloidal plane from the CTEM simulation showing the ion heat flux variable. (b) The structures in (a) with the positive structures in red and negative ones in blue. (c) A plot of the sum of ion heat flux in a structure vs. the size of the structure, clearly showing the large number of structures with negative flux.

The issue is being further investigated by GSEP physicists. The results on ITG simulations were presented in a poster at the 2010 Sherwood Fusion Theory Conference in April 2010 [KXL10], while the results of both ITG and CTEM analysis were presented at the GSEP Annual Meeting. This work is partially supported by the GSEP SciDAC Center.

Another activity in support of complex analysis tasks is enabling systems biology research in DOE BER/OASCR mission projects. In SciDAC-1, we developed an easy-to-use GUI-based software, called ***ProRata***, for **robust quantification of protein abundance** in high-throughput shotgun proteomics data [PKT+06, PKM+06]. Since then, ProRata has been downloaded more than 1000 times by a broad life sciences community; it has been featured by the *Journal of Proteome Research* [JPR06] and the *SciDAC Review* magazine [SGU+07]; it has been demonstrated in several conferences. In SciDAC-2, we robustified the *ProRata* software, added various features requested by the users, and outreached the use of ProRata by DOE Bioenergy:Genomics:GTL researchers. Among several collaboration activities, we highlight below the ones that have had scientific impact on DOE mission applications: (i) Enabled discovery and quantitative characterization of novel subunits in cellulosome complex critical for biomass degradation by bacterium, *Clostridium thermocellum* [RPH+09] in collaboration with the DOE ORNL Bioenergy center; (ii) Enabled the reconstruction of aromatic compound degradation pathways in a hydrogen producing bacterium, *R. palustris* [POL+08] in collaboration with DOE GTL [Project name]; (iii) Improved functional annotation of the ethanol producing bacterium, *Z. mobilis* [YHL+09] in collaboration with the DOE ORNL Bioenergy center; and (iv) first quantitative proteogenomic analysis of highly productive extremely acidophilic microbial communities [BPV+09].

A more recent activity is the Analysis of the zonal flow dynamics in gyrokinetic XGC1 simulation data. In collaboration with the SciDAC CPES Fusion Center (PI: Prof. C.S. Chang), our team has developed the *pR* analysis pipeline to **study three dimensional turbulence behavior** in time in the wavelet space. The use of this software to a special full-function (full-f) gyrokinetic code XGC1 to study ITG turbulence across the magnetic separatrix in divertor geometry has provided the following insights [CKD+09]: (a) the collisionless zonal flow grow in time in the density slope region, which is likely due to the stronger

GAM activities, with the inverse cascade of GAM energy into the zonal flows in the absence of collisions; and (b) the separatrix and X-loss effects do not change the well-known collisional damping effect of the zonal flows in the edge plasma.

2.4 Training and Outreach

Our training and outreach spanned a wide spectrum of activities. We presented our research findings at various national and international conferences, including invited talks at Supercomputing 2008 conference [SHB+09]. We summarized some of our findings as book chapters in the "Scientific Data Management" book, edited by A. Shoshani and D. Rotem [SR09, KWK+09, OW09]. We organized international conferences, such as the SIAM Data Mining conference, which, along with our editorial responsibilities [GKK08, Kam08a], allowed us to influence the broader scientific and technical communities in our areas of expertise.

We actively participated in a series of DOE Exascale workshops [HZS07], including a 2009-series such as ``Data Analysis, Management and Visualization in Fusion Energy Sciences at Extreme Scale." We co-organized the DOE/NSF workshop on "Mathematics for Analysis of Petascale Data" [KCC+08], the DOE OBER/OASCR workshop on "Genomics GTL Knowledgebase" [GFS09]. In the training arena, we contributed data analysis modules to the Supercomputing 2007 tutorial.

Some of our technologies, such as pR, are being taught as part of the undergraduate- and graduate-level curriculum on Automated Learning at NCSU, Computer Science Department; with the parallel data mining codes developed by the students using pR through their course projects. FastBit is attracting a growing user community with an active discussion mailing list and a number of enthusiastic contributors from around the world. Many research efforts presented in this report are the result of many PhD and MS students' theses. Our work on FastBit, for example, is sparking various research efforts around the country. During 2007, right after the public release of FastBit, it contributed significantly to two PhD theses from UC Berkeley and UIUC [SMW+08, RSW+07]. We are also aware of research efforts deriving from FastBit technologies or utilizing the software from other universities and private companies.

We worked closely with a number of DOE projects that has resulted in a number of joint publications or software usages including: (a) SciDAC Ultra-scale Visualization Institute (PI: K-L Ma) [SJH+07, SBH+08]; (b) SciDAC VACET Center (PIs: C. Johnson and W. Bethel) [OPW+08]; (c) SciDAC CPES Center (PI: C.S. Chang) [CKD+09]; (d) DOE Bioenergy Centers and GTL projects [RPH+09, POL+08, YHL+09, BPV+0, BGB+09].

3. Scientific Process Automation (SPA)

Effectively generating, managing, and analyzing scientific data requires a comprehensive, end-to-end approach that encompasses all stages from the initial data acquisition to the final analysis of the data. As part of the SPA thrust area, we are developing a suite of tools and frameworks that integrate into a robust and auditable system for automation of scientific processes to enhance and speed up scientific discovery. Our technologies provide run-time management of the workflow processes, provenance collection, and analysis and display of results. This has led to the deployment of production workflows that allow scientists to a) monitor, in near-real-time, complex tasks such as the execution of large simulation codes, and b) facilitate complex analyses of the process metadata and of the simulation results. This has resulted in significant savings in scientists' time, in more efficient use of resources, and in a more cost-effective scientific discovery process overall.

Workflow technologies have a long history in the database and information systems communities [GHS95]. Similarly, the scientific community has developed a number of problem-solving environments, most of them as integrated solutions [HRG+00]. Component-based solution support systems are also

proliferating [CL02, CCA06]. Scientific workflows merge advances in all these areas to automate support for sophisticated scientific problem-solving [LAB+06, LG05, DOE04, ABB+03, BVP00, VS97, SV96]. We use the term scientific workflow as a blanket term describing a series of structured activities and computations (called workflow components or actors) that arise in scientific problem-solving as part of the discovery process. This description includes the actions performed (by actors), the decisions made (control-flow), and the underlying coordination, such as data transfers (dataflow) and scheduling, required to execute the workflow. In its simplest case, a workflow is a linear sequence of tasks, each one implemented by an actor. An example of a scientific workflow is: transfer a configuration file to a large cluster, run a simulation passing this file as an input parameter, transfer the results of the simulation to a secondary system (e.g. a smaller cluster), select a known variable, and generate a movie showing how this variable evolves over time. Scientific workflows can exhibit and exploit data-, task-, and pipeline-parallelism. In science and engineering, process tasks and computations often are large-scale, complex, and structured with intricate dependencies [DOE04, DBN+96, EBV95, Elm66].

Over the past five years, our activities have both established Kepler as a viable scientific workflow environment and demonstrated its value across multiple science applications. We have published over 70 peer reviewed papers on the technologies highlighted in this short paper and have given Kepler tutorials at SC06, SC07, SC08, and SciDAC 2007. Our outreach activities have allowed scientists to learn best practices and better utilize Kepler to address their individual workflow problems.

Our contributions to advancing the state-of-the-art in scientific workflows have focused on the following areas. Progress in each of these areas is described in subsequent sections.

- **Workflow development.** The development of a deeper understanding of scientific workflows "in the wild" and of the requirements for support tools that allow easy construction of complex scientific workflows;
- **Generic workflow components and templates.** The development of generic actors (i.e. workflow components and processes) which can be broadly applied to scientific problems;
- **Dashboard development.** The development of a one-stop-shopping workflow monitoring and analytics dashboard;
- **Provenance collection and analysis**. The design of a flexible provenance collection and analysis infrastructure within the workflow environment; and
- **Workflow reliability and fault tolerance.** The improvement of the reliability and fault-tolerance of workflow environments.

3.1 Workflow development

The original base-line contribution of the SPA team has been to co-found the Kepler project – an open source workflow support environment [http://www.kepler-project.org]. Kepler is now a widely accepted scientific workflow development and execution environment that powers a number of research and production projects all over the world. The SPA researchers and engineers continue to regularly contribute to Kepler. We are constantly working with the Kepler Core team [e.g., AJB+04, LAB+06, GBA+07] to enhance Kepler at all levels including the user interface, documentation, and tutorials. Our work has led to a significant reduction in the effort required to generate real workflows [ABC+06, SAC+07]. We also actively partner with science teams to transfer technology to their projects and develop and deploy their workflows. This provides real-life case-studies which are then used to enhance Kepler requirements, to identify Kepler enhancements, generic functionalities, and canonical generic workflow solutions, and to improve user interfaces.

Accomplishments

As active participants, and founding members, in the Kepler research community we **contribute to the development of the Kepler environment**. Specifically, we participate in Kepler and Ptolemy workshops that have produced significant enhancements in the underlying workflow environment, and we have

identified and implemented new requirements for scientific workflows, fixed bugs, and improved the overall software development environment, as well as execution-time interfaces and data collection practices [e.g., BMR+08, CA08, NV08, SAC+07, VAB+07]. Some of the specific tools and additions to Kepler are discussed in sections that follow (e.g., provenance recorder, generic actors, ADIOS). An example of using Kepler for simulation monitoring and steering is shown in Figure 11. This particular workflow processes files representing time-steps as they are generated by simulations, processes them for display on the dashboard (described in section 3.3), and in parallel invokes a program (ELITE) which can determine if the simulation is stable.

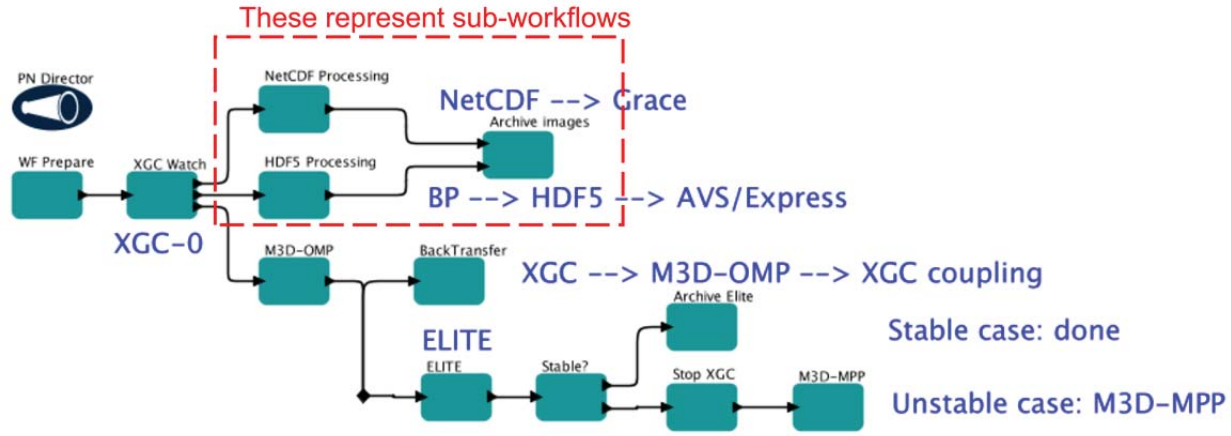


Figure 11: a workflow for monitoring the correctness of a simulation in real-time.

If the simulation is considered unstable, it stops the run (see box "stop XCG") thus saving compute resources, and calls another program (M3D-MPP) to recomputed new parameters for the next simulation run. This example shows the power of using workflows for near real-time simulation monitoring, generating summaries, graphs and images "on-the-fly" which can greatly help scientists follow up and understand their simulations behavior quickly.

Another critical component of the SPA activities is **transfer of workflow technology to science teams**. We have deployed workflows to a number of science teams to a) help introduce Kepler into their domains, and b) feedback lessons learned into Kepler to improve the Kepler environment. This includes:

- **ScalaBLAST** (Bioinformatics): This workflow performs a comparison of N submitted genomes against M genomes stored in a library. This requires generating the NxM individual comparisons, submitting the results on a cluster, and aggregating the results.
- **Groundwater Modeling and Analysis:** A set of Kepler-based scientific workflows have been constructed to support subsurface flow and transport modeling using the STOMP (Subsurface Transport Over Multiple Phases) simulation. The high-level groundwater modeling workflow involved specific computational tasks including clustering, multivariate interpolation, subsurface flow and transport simulation, and data visualization. Additional low-level workflows were developed to support data staging and simulation job submission. Furthermore, an iterative workflow was designed to collect several input variable ranges and perform a parameter study using fixed code and combinations of input variables values.
- **The Atmospheric Radiation Measurement (ARM) Program**: The ARM program deploys multiple "Value Added Products" (VAPs) which derive scientifically meaningful information from the original raw data sets using a complex combination of data transformations and scientific models. Unfortunately, most of these VAPs are defined through scripts, with no provenance tracking and limited fault tolerance. We developed a demonstration workflow for one of their most complex VAPs which not only provided an improved execution platform but enabled tracking of data

provenance for the first time [CGS+09]. After completion of this demonstration infrastructure, responsibility for the maintenance of this workflow and additional workflow development was transferred to the ARM development team. Over the past year, they have continued to explore how provenance information could be effectively utilized to meet their programmatic requirements [SHC+10].

- **Center for Plasma Edge Simulation (CPES)**: This fusion project has been a main user of our technologies providing both requirements and feedback. As the first project using Kepler for monitoring and post-processing of High Performance Computing applications, our actors for SSH and job management found immediate use. eSiMon has been used as the front-end to fusion scientists and the content was created by Kepler workflows. Provenance tracking was first used in the fusion monitoring workflow and by fusion users of eSiMon. The above tools comprise EFFIS, the End-to-end Framework for Fusion Integrated Simulation [CKP+10]. The integrated simulation is driven by a Kepler "coupling" workflow. The plasma state in the simulation by the XGC0 code on a supercomputer is constantly monitored, which involves a data conversion step using another fusion code (M3D) and a parameter-study using yet another code (Elite) on an analysis cluster. If the plasma state is found to be unstable, the XGC0 simulation is stopped and a combined XGC0 and M3D magneto-hydrodynamic simulation is started to go through the turbulent phase of the fusion reaction. After that, XGC0 is continued again alone. Additionally, the workflow creates plots and 2D visualizations from each code's output at every timestep for eSiMon, which can be used for on-line monitoring of the coupled simulation as well as for post-processing runs.
- **S3D** (Jackie Chen, Combustion) – Three types of workflows were identified: Preparation, Production, and Post-Production workflows. The preparation workflows are designed to save unnecessary faulty runs on the supercomputer. Before accepting a particular build of S3S for production use, it is necessary to run a series of test cases. The production workflows are similar to CPES runs, but, of course, the diagnostic plots to be generated during the simulation are different. The post-processing workflows are designed to use the dashboard as an index to archived runs in order to facilitate collaboration with their visualization partners.
- **CHIMERA** CHIMERA is a multi-dimensional radiation hydrodynamics code designed to study core-collapse supernovae. The code is made up of three essentially independent parts: hydrodynamics, nuclear burning, and a neutrino transport solver combined within an operator-split approach. Two-dimensional multi-physics simulations of core collapse supernovae have been performed with CHIMERA and three-dimensional simulations are underway. Collaboration with the SPA team involves insertion of ADIOS into CHIMERA for I/O speed-up, and automation of CHIMERA workflows.
- **CCSM** (climate): This recent effort with the National Center for Atmospheric Research (NCAR) is helping develop Kepler-driven provenance for the Community Climate System Model (CCSM[1]). CCSM belongs to an elite category of computer-based simulation models known as general-circulation models. The automatic capturing of provenance is an essential part of this activity, especially as the number and volume of simulations is expected to significantly grow.

We expect, and have seen, that over time, the science teams will assume responsibility for the maintenance of existing workflows and creation of new ones. Thus, we do not view our development of specific workflows as an ongoing obligation (or as a limit on the utilization of the tools, since other teams are developing workflows without our direct support) but rather a proven approach for deploying new technology through a win-win scenario.

In addition to our direct engagement with certain science teams, we have been **disseminating information on workflow technologies to the broader community** through a series of publications (see Publications), tutorials and presentations. In particular, we have given well attended tutorials at the last

[1] http://www.ccsm.ucar.edu/, http://www.esmf.ucar.edu/

three Supercomputing conferences (SC06, SC07, and SC08), the SciDAC 08 conference, and the 2007 meeting of the NCCS. We have also provided tutorials for interested groups, upon request.

3.2 Generic workflow components and templates

Many workflows contain sections with very similar functionality but subtle differences in how that functionality is obtained. For example, transferring a file between machines, submitting a job to a batch processing system, monitoring the execution of a running job, and remotely executing a command are found in almost all of the scientific workflows we have developed. However, the actual implementation of these capabilities varies dramatically depending on features such as the specific machine configurations (e.g., which batch processor is used), the security requirements (e.g. ssh or rsh, certificates or one-time-passwords), and the workflow requirements (e.g. failover options, fault tolerance requirements, validation options). Tools that provide the instantiation of a case-specific workflow or workflow component from more general templates or components would substantially improve the ability to reuse existing solutions, leading to greater productivity when developing workflows.

Accomplishments

We have developed a large number of **actors and actor packages** to support SDM-type workflows, including the following: A *ProcessFile* actor that allows to execute external commands on a remote machine and supports checkpointing of successful commands (useful for "smart resume" in case a restart of the workflow is necessary) and logging of errors; an *SSH-FileWatcher* that allows to list and watch remote directories with a set of "file-masks;" this is a frequently used step when monitoring loosely-coupled, remote processes. These and various actors have been bundled into an ssh and job management package for remote execution, file transfer, and job submissions to various job management systems (PBS, Condor, SGE, and LoadLeveler). Another set of actors has been developed to deal with web service-oriented workflow steps; these include: *WSWithComplexTypes* (permits sending and receiving of structured web service parameters), various "shim actors" that allow conversion between strings, records, and XML structures, and a number of array manipulation actors.

We have also investigated several alternative approaches to development of **generic actors**, including the separation of the control flow and data flow aspects of the workflow. Our current approach is to use data-driven code to flexibly adapt to the current environment by intelligently selecting the most appropriate execution path based on a combination of user input and available resources. We are also investigating the development of a broader workflow context resource, which would allow effective sharing of information (such as login certificates) between actors within a single, executing workflow. One of the key capabilities that have been used in several of our workflows is the movement of files between machines; for example, transferring information from the LCF machine to Ewok at ORNL on the CPES workflow. As a result, our first generic actor capability has been the deployment of a **GenericFileTransfer** actor. This actor can utilize a variety of protocols, including ftp, scp, and SRM-lite, to transfer a file between two computers. It also automatically confirms that the transfer completed properly (i.e. there were no errors and the resulting files are the same size) and if an error was encountered, retransmits the file up to a specified number of times.

Currently we have a **genericSSH** actor that can establish connection with SSH servers using a password, phassphrase, passcode, or an SSH certificate. We are currently working towards extending this functionality to support GSI-OpenSSH server. This requires implementing GSI authentication in addition to the existing authentication methods. We have also deployed a **genericJobLaunch** actor that encapsulates all the tasks involved in job launching into a single actor. This will enable users to use a single actor and configure its parameters instead of replicating the sub workflow and configuring its individual components.

We have developed a set of pilot **Kepler workflow templates** that we are evaluating. Similar to generic actors in their advantages, workflow templates can then be used, with only minor modifications, to

implement new workflows. The current set is concerned with file creation monitoring and data transfers during long-term workflow runs with fusion codes. The intent is to create sufficiently generic components that can be used in long-term workflows in other disciplines. Templates have been presented and discussed as part of the Kepler tutorial at the Supercomputing 2008. We have also developed a new approach to workflow design called Collection-Oriented Modeling and Design (COMAD) and implemented a new COMAD director. We have demonstrated that workflows in this modeling paradigm tend to be easier to understand and maintain [MBZ+08].

3.3 Dashboard development

The emergence of leadership class computing is creating a tsunami of data from petascale simulations. Results are typically analyzed by dozens of scientists. In order for the scientists to digest the vast amount of data being produced from the simulations and auxiliary programs, it is critical to automate the effort to manage, analyze, visualize, and share this data. One aspect of automation is to provide an easy-to-use web-based mechanism to monitor the progress of simulations, and view and compare the results generated with the use of the workflow system. A second aspect is to leverage the collective knowledge and experiences through a scientific social network. This can be achieved through a combination of parallel back-end services, provenance capturing, and an easy-to-use front-end tool. The SDM Center eSiMon (electronic Simulation Monitoring) Dashboard is one such tool. An example of using eSiMon in a fusion project is shown in Figure 12.

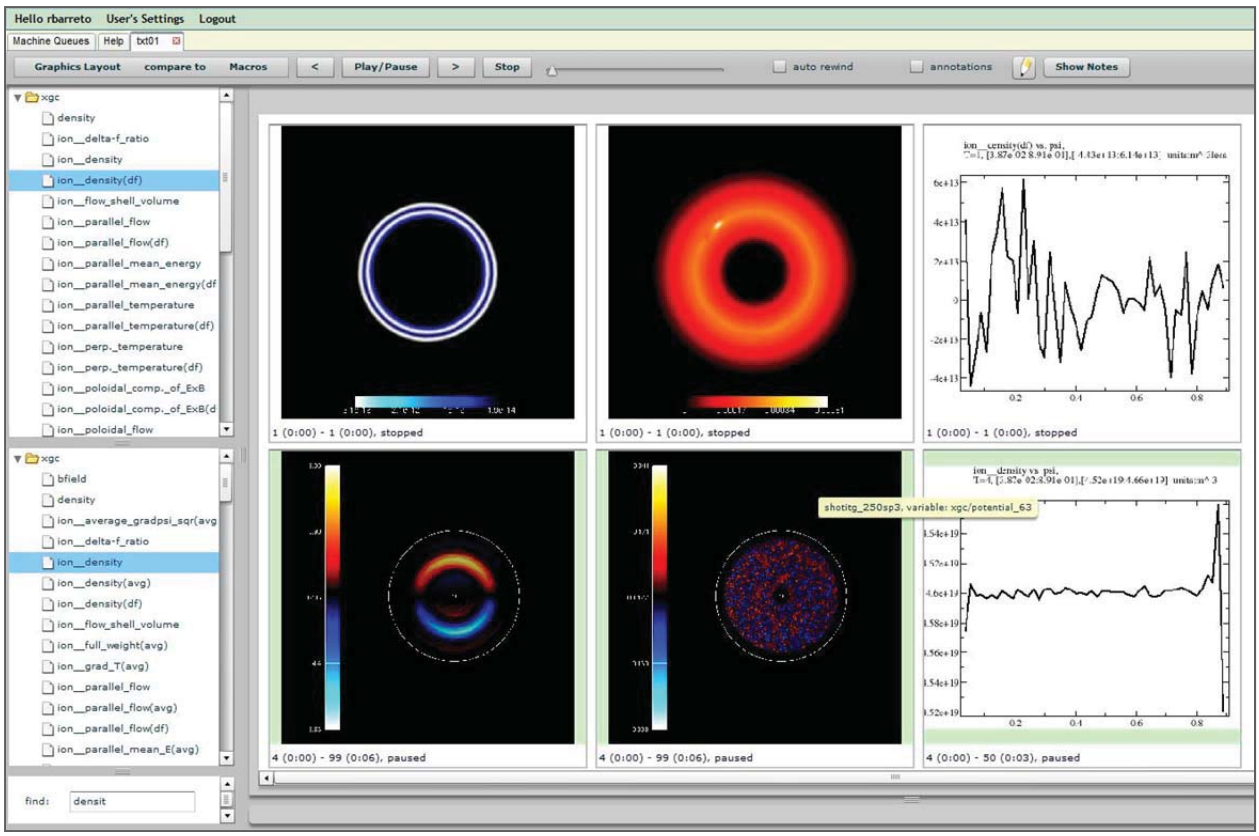


Figure 12: Use of a dashboard to observe results of simulation through web interface provides scientists easy to use interface for data exploration

Accomplishments

A second-generation "Dashboard" solution has been released [BKM09]. It is used in CPES production runs at ORNL. The eSiMon Dashboard is a web-based tool that provides scientists "one-stop-shopping" access to their simulations, the associated provenance information, and run-time monitoring and run-time and off-line sharing and analysis of their data. For example, they can view the status of the super-computers they or their collaboration group are running on, and provided they are authorized to do so they can interact with the simulation and their own or others' datasets. In designing the tool we have been careful to be inclusive. For example, in the context of simulation monitoring, a focus on visualization may exclude experimentalists, performance analysts or even the theoretical scientists. Therefore, the tool offers options. Physicists are not necessarily concerned with which visualization or analysis software is used to generate results. They want to think in terms of variables, physical entities and phenomena. Therefore, the main challenge in designing the user interface was allowing them to collectively focus and exchange information on their science. We believe we have been successful in doing so.

Our user-centered Dashboard is a scientific social network currently being used by physicists to monitor their simulations as well as their colleagues' and teammates' simulations. They can follow the data lineage and perform analysis on simulation results generated on leadership class computers. Scientific data is organized and presented as scientific variables on the dashboard, raising the focus of the researchers from the storage of the data (files) to its scientific meaning. The eSiMon Dashboard makes use of the provenance tracking system to enable collaborative analysis operations on the scientific data. eSiMon has a feature-rich and yet efficient graphical user interface, it is coupled with the provenance tracking system, and it supports a range of analysis features. It allows the scientist to identify the raw data files that contain the variables and request that these are sent to their own systems, in case they wish to perform their own analysis locally. We are continuously exploring new ways to enhance the existing collaborative features on the Dashboard.

The generality of the eSiMon Dashboard design can be seen as the same Dashboard design is applied to another application (combustion) shown in Figure 13. Here the goal is to understand how flame fronts progress over time.

3.4 Provenance collection and analysis

In scientific applications, effectively managing data provenance is extremely important [MGBM05, GMF+05, CFS+05, SPG05]. Data provenance [MGBM05] can be thought of as the complete processing history of a data product, for example, actor identification and invocation parameters (or application codes launched by those actors), properties such as the time, location, and userid of invocation, relevant environment and configuration parameters. This information needs to be persistent and permanently associated with a data product so that its provenance is readily available. It should also be searchable, so that data with certain provenance can be easily identified – for example, if a bug is identified and corrected, the provenance can help identify which runs should be repeated.

Accomplishments

We have **extended Kepler to support obtaining and recording provenance information** through a facility called “provenance recorder.” This extension allows a workflow to automatically collect provenance information at a user-defined level of granularity [VAB07, CA08]. This information is recorded in a database allowing easy searching and retrieval. However, in the distributed, heterogeneous environment utilized by scientific computing, simply recording provenance information from the workflow is insufficient. In a typical scientific workflow, while the workflow engine directs the activity, significant data manipulation occurs on derived products which are never fully incorporated into the

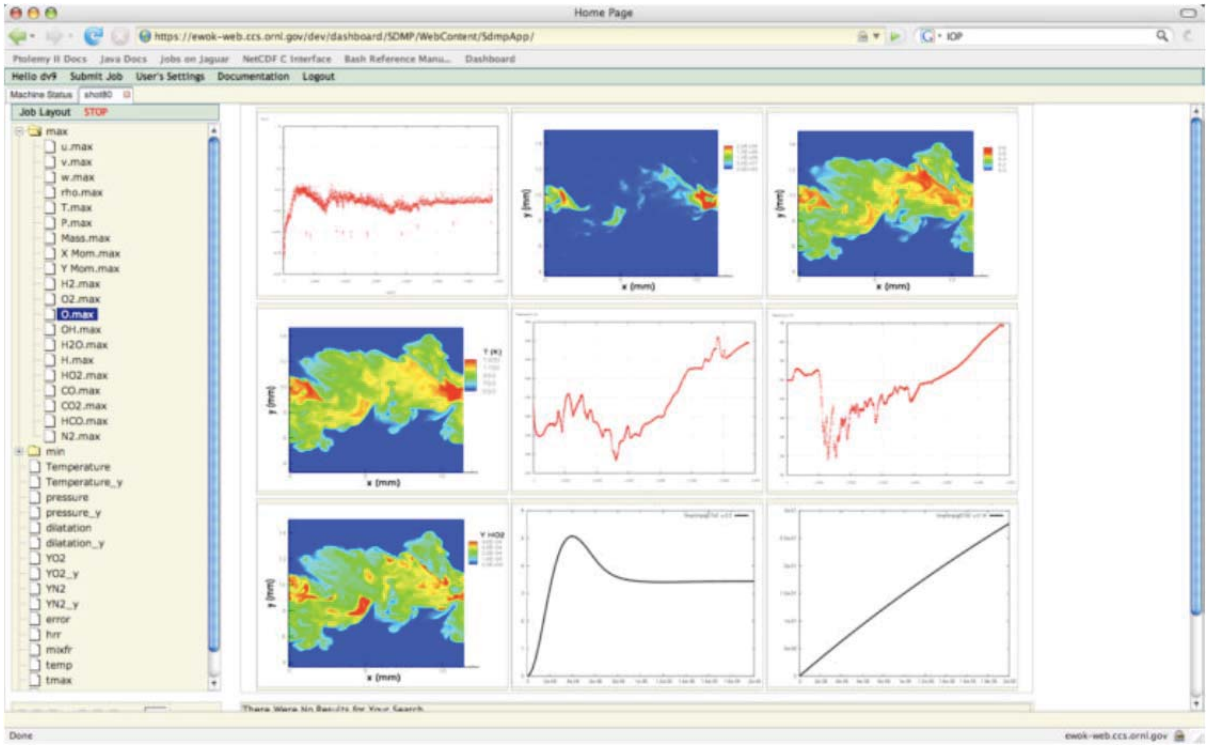


Figure 13: the use of the Dashboard for a combustion simulation

workflow environment. For example, the workflow engine may request the creation of a movie from a simulation output, but only the file names, not the underlying data, move through the workflow environment. In order to ensure the recording of the full data provenance, we have enabled the provenance capture environment to allow recording of events external to the workflow – including the compilation of the underlying simulation codes. This provides a more complete record of the workflow execution, increasing the ability of scientists to reliably replicate the results.

We have developed and released several versions of the **provenance recorder** within Kepler. This has provided an entirely new set of capabilities and made Kepler even more valuable. For example, one of the primary reasons CCSM developers are interested in Kepler is to enable provenance capture in their workflows. We distinguish four types of provenance: data, process, workflow and system.

The provenance recorder is located within the Kepler framework and tracks all events of interest generated by workflow execution. These events include actor executions, token transfers between actors, and any additional information that an actor may choose to provide. Because a large, long-running workflow may create thousands of events, the most recent version of the recorder allows the workflow developer to identify which information is recorded. This significantly reduced the information flow load and improves the performance during both workflow execution and during evaluation. The current official release of Kepler (v2.1.0/30-Sep-10, http://www.kepler-project.org) includes the provenance recorder.

Provenance information has multiple purposes, and thus requires multiple levels of granularity. For example, while a workflow is executing, significant provenance information may be collected to restart the workflow in case of error (see Fault Tolerance section below). However, once the workflow is complete, only a small fraction of that provenance may be required for validation purposes. In order to effectively reduce the amount of provenance information we store, we are actively working on approaches for archiving and managing provenance information based on its usage requirements. In

addition to the workflow engine, external processes must have efficient access to the provenance store. This ensures that relevant information generated outside of the workflow engine, such as from a process spawned by Kepler, can be effectively recorded. It also allows tools such as the Dashboard to utilize this information without having to execute a workflow. However, that raises the matter of security and privacy of the provenance information. That information may sometimes be at least as valuable as the raw data. Our provenance framework solves that problem in part through its access mechanisms [NV08], but more work is needed. Provenance recording capabilities of Kepler and some open issues are described in [VAB07, NV08].

3.5 Workflow reliability and fault tolerance

There are two basic forms of run-time fault-tolerance: forward-recovery (e.g., failure masking and redundancy based failover), and backward-recovery (e.g. check-pointing) [LB90, MV06, Vou05]. Exception handling is a very traditional way of managing run-time problems [MV96, Vou05]. It is also used in the workflow-oriented environments [HA96, CCPP99]. Exception handling can involve forward-recovery, backward-recovery, or graceful termination. More recently the web services community has recognized the need for some form of standardized fault-tolerance in the service provisioning through replication [SPPJ06]. An important component is collection of sufficient amount of meta-data (provenance information) about the workflow to enable fault-tolerance actions. It is the provenance information that is being collected through our provenance recorder that has the capability of providing meta-data needed to detect and locate workflow run-time issues.

Accomplishments

We have implemented a Kepler-based fault-tolerance (FT) framework that leverages collected provenance information to provide options for forward recovery as well as backward recovery of the workflows [MCA+10, YAC+10]. The FT framework addresses the majority of the known issues and faults in current production workflows and is composed of 3 major components:

- An error Handling Layer, which monitors the components beyond the workflow engine's direct control, such as visualization services, and reacts accordingly when an error is detected
- A Contingency actor, which provides a recovery block mechanism within the workflow.
- Checkpointing and Smart resume capabilities for when the above 2 mechanisms fail are not sufficient to prevent the workflow from terminating abnormally.

The Contingency actor handles faults at the Workflow Layer. This actor contains a primary sub-workflow, and possibly additional sub-workflows. During execution, data read by the actor is transferred first to the input of the primary sub-workflow. The sub-workflow contains the primary task to be executed, and if an error occurs, the Contingency actor may re-execute the primary sub-workflow or transfer the original input data to a different sub-workflow. The choice of which sub-workflow to execute is governed by a finite state machine. Additionally, the Contingency actor may be configured to pause between re-executions of a sub-workflow.

The extension of the model beyond the workflow control plane allows us to catch and process problem signals from environments that Kepler has no direct control over [MCA+10, YAC+10]. The key concept in that context is operational profiles – the frequency of usage of different Kepler and other operations, and their relationship to run-time failures. Operational profiles are an essential part of software reliability engineering. Typically they are created from the software requirements, and through customer reviews. Creation of operational profiles often is laborious and requires human intervention. Our approach builds an operational profile based on the actual usage from execution logs. The difficulty in using execution logs is that the amount of data to be analyzed is extremely large (more than a million records per day in many applications). Our solution constructs operational profiles by identifying all the possible clustered

sequences of events (patterns) that exist in the logs. This is done very efficiently using suffix arrays data structure [NWV09, NVW+08].

4 Framework for Integrated End-to-end SDM Technologies and Applications

SDM center computer scientists actively collaborate with fusion scientists from the Center for Plasma Edge Simulations (CPES). The main theme of this collaborative effort is to provide enabling technologies for complex coupled simulations running from petascale computers. The technologies being developed by the SDM center for CPES are fully applicable to other projects as well, and in fact, the success of many of our techniques has led to adoption by some of the biggest data-producing codes in the DOE (*e.g.*, (*e.g.*,CHIMERA for astrophysics simulations and S3D for combustion simulations). Furthermore, there is already a lot of interest from other FSP projects in the technologies developed within CPES, as well as collaboration with ITER simulation software developers in France.

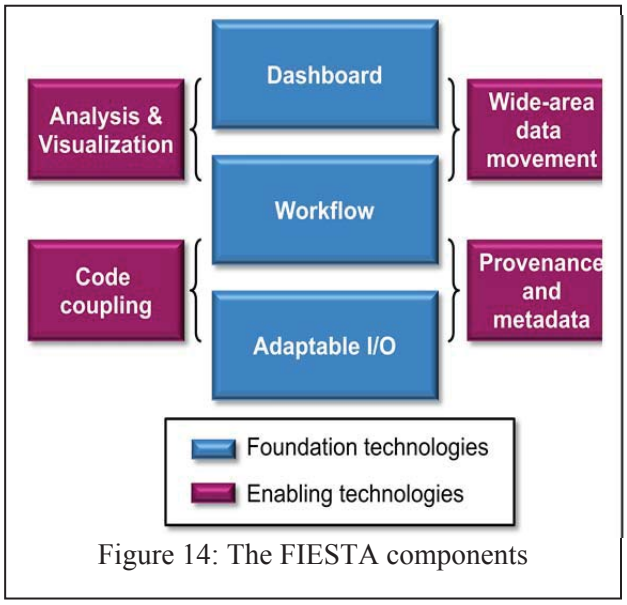


Figure 14: The FIESTA components

Accomplishments

In order to facilitate progress in the physics research, our team had to develop and use innovative techniques for high-performance I/O, fault-tolerant workflow technology, and delivery of visualization/analysis to the desktop for the growing community of CPES users. The architecture of framework for Integrated End-to-end SDM Technologies and Applications (FIESTA) is shown in the figure 14. The key aspect of all our technologies has been hiding the complexity from users to yield an easy-to-use, easy-to-reapply, scalable and robust computational framework. Our technologies have been applied to simulations using over ten-thousand processors, helping scientists monitor their simulations during runs and obtain scalable high-performance I/O. The main technologies that we have created and used are: (i) a componentized I/O method capable of scaling to tens of thousands of processors (**ADIOS – ADaptable I/O System**), (ii) standard scientific workflow templates for petascale simulation monitoring and coupling of codes (using **the Kepler framework**), (iii) a **Dashboard** where scientists can quickly access information about their runs and the machines they are running on, and (iv) **advanced visualization and analysis** techniques that can be accessed from both the dashboard and external tools, (v) use of **provenance** to identify original files from which visualization products were generated, and (vi) **SRM-lite**, providing fast methods to move these files from the simulation to local and remote resources. The overall approach is to place highly annotated, fast, easy-to-use I/O methods in the code, which can be monitored and controlled; have a workflow engine record all of the information; apply analysis and visualize this on a dashboard; move desired data to the user's site; and have metadata and provenance reported to a database

One of the main goals of the computer science work in CPES is to facilitate running workflows that will perform dynamic monitoring of simulations and the dynamic analysis of the time steps in order to track the simulations. This required the use and development of special components (called "actors") of the Kepler workflow system. This work exceeded our original goals in that such workflows were not only developed for the stated goals and are now deployed in production runs, but also incorporate many other tasks, such as generation of images and movies, data transformation on-the-fly as the data is generated, and combining of multiple small files into a larger file for archiving. The workflows developed not only

support “monitoring workflows” as originally envisioned, but also support “coupling workflows” across machines.

Over the course of the project, we have found some technology to be missing and have had to create it ourselves. One example of this is the development of a Dashboard, where fusion scientists can visualize results of simulations, while the simulations are running. This was developed, enhanced with efficient visualization technology, and is now used in production runs.

The second technology is for the improvement of I/O speed for dumping data from the compute engines. For this purpose, we have used the ADIOS framework, described in section 1.3. Since one of the goals under the CPES work was to facilitate code coupling, ADIOS also allows us to componentize this layer, so that we can move data quickly from different components inside of the CPES framework and communicate directly with our coupling and monitoring workflows.

More recently, we have added a third activity of providing wide-area file movement to the physicist’s site through requests made on the Dashboard, using Storage Resource Manager (SRM) technology that is used in the SDM center. This activity, referred to as SRM-lite (which is a client version of SRM) already started and is now part of our future goals.

We have also provided a support for parallel data analysis capabilities that are accessible through the Dashboard’s web portal and utilize our pR technology. An experimental set of 15 analysis routines were added to the Dashboard. The initial set of routines was selected based on our discussions with climate and fusion simulation scientists and analysis of their domain-specific application-based packages, such as CDAT, GKV, and vugyro.

Our focus has been in innovative techniques using I/O pipelines in staging, for in situ analysis, re-organization, visualization, and I/O. At the heart of our research has been the ADIOS research platform, which allows researchers the ability to invent new methods into the I/O stack, and immediately use these in major simulations which run on the all of the major computer platforms. In fact, our codes are used in many major codes including: GTS, GTC, XGC-1, XGC-0, M3D-MPP, M3D-OMP, M3D-K, Pixie3D, GTC-P, Chimera, S3D, HFODD, PAMR; along with many new codes which scientist are just starting to use with ADIOS in the USA, Europe, and Asia. It is important to note that the ADIOS team includes both partners within the SDM Center, specifically ORNL, LBNL, and NCSU, but also external collaborators including Auburn, Georgia Tech, Sandia, and Rutgers.

Overall research highlights by the ADIOS team include.

1. Using the staging area inside of ADIOS, researchers can now use a Service Oriented Architecture to couple codes. We have worked with the CPES SciDAC project, to couple the XGC-0, M3D-OMP, M3D-MPP code together. Using Kepler we then use the provenance tracking capability to track the provenance, and to move files that are output from the coupled codes to the Elite simulation, which runs on a different platform (first codes run on jaguar at OLCF, and the Elite code runs on ewok at the OLCF).
2. We have used codes embedded code, as part of the I/O stream, to create in situ workflows, which contain this embedded code, and we can execute this code on other nodes, during the simulation.
3. We have worked on scheduled movement of the I/O services, and we can reduce the impact of the I/O pipelines/ (disk activity) by paying careful attention to the data movement for I/O, and the data movement for local data movement inside of the HPC system.
4. We have worked on integrating file-based code coupling, alongside memory based code code, to work in concert with one-another. This has been demonstrated in several Fusion Simulation Project meetings.
5. We have coupled together more in situ visualization techniques inside the ADIOS staging area, thereby shielding the visualization code from the simulation code. Each code acts as a service, and uses the ADIOS I/O interface for the coupling mechanism.

6. We have demonstrated the value of an Application Log File format (ALF), in it's ability to place data on storage targets, thereby increasing concurrency for I/O patterns common for codes in reading.
7. We have worked closely with the S3D team to further help scale their code to 98K cores, having an I/O impact <1% to their calculation, using our sub-filing technique; including open, read, and close.
8. We have worked on new ADIOS methods which are capable of increasing the QoS for simulations, demonstrating this for "real-codes" such as XGC-1, allowing them to scale to over 240K cores.
9. We have worked closely with the DMA team, to increase the Statistics generated in ADIOS, automatically for the scientist. This includes such statistics as min/max, average, and standard deviation across each time step separately.
10. We have worked on optimal methods for the IBM BGP at the ALCF, using the GTC-P code, allowing their code to get "near-optimal" performance on the ALCF file system.

Publications and references

Publications are papers published by the SDM center staff since the beginning of SciDAC-2.

References are documents by others or selected publications by SDM center staff during SciDAC-1, or before the SDM center's inception in 2001 **(**All SDM center publications are posted on http://sdm.lbl.gov/sdmcenter/publications.php).

Publications (173) - Also see Appendix 4 for additional NCSU publications

2011

- [CC11] T. Critchlow, G. Chin Jr., "Supercomputing and Scientific Workflows Gaps and Requirements", A short paper The Fifth IEEE International Workshop on Scientific Workflows (SWF 2011) published within the Proceedings of the Seventh IEEE World Conference on Services (Services 2011). Washington, DC, July 2011.
- [CHA+11] Philip Carns, Kevin Harms, William Allcock, Charles Bacon, Robert Latham, Samuel Lang, and Robert Ross. Understanding and improving computational science storage access through continuous characterization. In Proceedings of 27th IEEE Conference on Mass Storage Systems and Technologies (MSST 2011), May 2011. Received **Best Paper** Award.
- [CSC+11] G. Chin Jr., C. Sivaramakrishnan, T. Critchlow, K. Schuchardt, A.H.H. Ngu, "Scientist-Centered Workflow Abstractions via Generic Actors, Workflow Templates, and Context-Awareness for Groundwater Modeling and Analysis", The Fifth IEEE International Workshop on Scientific Workflows (SWF 2011) published within the Proceedings of the Seventh IEEE World Conference on Services (Services 2011). Washington, DC, July 2011.
- [DCKP11] C. Docan, J. Cummings, S. Klasky, M. Parashar, "'Moving the Code to the Data - Dynamic Code Deployment using ActiveSpaces", to appear in IPDPS 2011.
- [GCP+11] A. Gándara, G. Chin Jr., P. Pinheiro da Silva, S. White, C. Sivaramakrishnan, T. Critchlow, "Knowledge Annotations in Scientific Workflows: An Implementation in Kepler", In Proceedings of the 23rd Scientific and Statistical Database Management Conference (SSDBM), Portland OR, July 2011.
- [JRR+11] R. Joseph, G. Ravunnikutty, S. Rnka, E. D Azevedo, S. Klasky, Efficient GPU implementation for Particle in Cell algorithm, to appear IPDPS 2011.
- [LPG+11] J. Lofstead, M. Polte, G. Gibson, S. Klasky, R. Oldfield, J. Bent, A. Manzanares, Q. Liu, N. Podhorszki, M. Wingate, M. Wolf, "Data Districts: Data Organization for High End-to-End Performance of Extreme Scale I/O", submitted to FAST 2011.
- [LMY+11] Jiangtian Li, Xiaosong Ma, Srikanth Yoginath, Guruprasad Kora, Nagiza F. Samatova, Transparent runtime parallelization of the R scripting language, Journal of Parallel and Distributed Computing (JPDC), Volume 71, Issue 2, February 2011, Pages 157-168.
- [LSE+11] Sriram Lakshminarasimhan, Neil Shah, Stephane Ethier, Scott Klasky, Rob Latham, Rob Ross, and Nagiza F. Samatova, Compressing the Incompressible with ISABELA: In-situ Reduction of Spatio-Temporal Data, *Distinguished Paper*, In 17th International European Conference on Parallel and Distributed Computing (Euro-Par 2011), Bordeaux, France, Aug. 2011.
- [NJC+11] A. H. Ngu, A. Jamnagarwala, G. Chin Jr., C. Sivaramakrishnan, T. Critchlow, "Kepler Scientific Workflow Design and Execution with Contexts", International Journal of Computers and Their Applications (IJCA) Special Issue on Scientific Workflows, Provenance and Their Applications, In Press.

- [TKA+11] Yuan Tian, Scott Klasky, Hasan Abbasi, Jay Lofstead, Ray Grout, Norbert Podhorszki, Qing Liu, Yandong Wang, Weikuan Yu. EDO: Improving Read Performance for Scientific Applications Through Elastic Data Organization. IEEE Cluster 2011. Austin, TX.
- [TKL+11] Y. Tian, S. Klasky, J. Lofstead, R. Grout, N. Podhorszki, Q. Liu, Y. Wang, W. Yu, "Data reordering Using Hilbert Space Filling Curve to Improve the Read Performance for Scientific Applications", submitted to IPDPS 2011.
- [TVL+11] Yuan Tian, Jeffrey S. Vetter, Honggao Liu, Scott Klasky, Weikuan Yu. neCODEC: Nearline Data Compression for Scientific Applications. Submitted to CCGrid 2011.
- [TY11] Y. Tian, W. Yu. Enabling Petascale Data Analysis for Scientific Applications through Data Reorganization, ICS 2011, Poster, Tucson, AZ, May 2011 (First Prize of ACM Student Research Competition in ICS'11).
- [TYV+11] Y. Tian, W. Yu, J. Vetter, J. Gao, RXIO: Design and Implementation of High Performance RDMA-capable GridFTP, Journal of Computers and Electrical Engineering, Special Issue of Emerging Computing Architectures and Systems, Nov 2011.
- [WSJ+11] Kesheng Wu, Rishi R. Sinha, Chad Jones, Stephane Ethier, Scott Klasky, Kwan-Liu Ma, Arie Shoshani, Marianne Winslett, Finding Regions of Interest on Toroidal Meshes, Journal of Computational Science & Discovery, 2011.
- [YWK+11] Weikuan Yu, K. John Wu, Wei-Shinn Ku, Cong Xu, Juan Gao. BMF: Bitmapped Mass Fingerprinting for Fast Protein Identification. IEEE Cluster 2011. Austin, TX.
- [ZDP+11] Fan Zhang, Ciprian Docan, Manish Parashar and Scott Klasky. Enabling Multi-Physics Coupled Simulations within the PGAS Programming Framework, 11th International Symposium on Cluster, Cloud and Grid Computing (ccGrid), May, 2011.
- [ZHM+11] Daniel Zinn, Quinn Hart, Timothy McPhillips, Bertram Ludaescher, Yogesh Simmhan, Michail Giakkoupis and Viktor K. Prasanna. Towards Reliable, Performant Workflows for Streaming-Applications on Cloud Platforms, 11th International Symposium on Cluster, Cloud and Grid Computing (ccGrid), May, 2011.

2010

- [AAC+10] Ilkay Altintas, Manish K. Anand, Daniel Crawl, Adam Belloum, Paolo Missier, Carole A. Goble, Peter M.A. Sloot. Understanding Collaborative Studies Through Interoperable Workflow Provenance. The third International Provenance and Annotation Workshop (IPAW2010). Submitted.
- [ABA+10] Manish Kumar Anand, Shawn Bowers, Ilkay Altintas, Bertram Ludaescher. Approaches for Exploring and Querying Scientific Workflow Provenance Graphs. The third International Provenance and Annotation Workshop (IPAW2010). Submitted.
- [ABL10] Manish Kumar Anand, Shawn Bowers, Bertram Ludäscher: Techniques for efficiently querying scientific workflow provenance graphs. EDBT 2010: 287-298
- [AKS10] H. Abbasi, S. Klasky, K. Schwan, M. Wolf, "Extracting Information ASAP!", PDSI 2010.
- [CKP+10] Cummings, Klasky, Podhorszki, Barreto, Lofstead, Schwan, Docan, Parashar, Sim, Shoshani, "EFFIS: and End-to-end Framework for Fusion Integrated Simulation", PDP 2010, http://www.pdp2010.org/.
- [DCK+10] C. Docan, J. Cummings , S. Klasky , M. Parashar, N. Podhorszki, F. Zhang, "Experiments with Memory-to-Memory Coupling for End-to-End fusion Simulation Workflows", ccGrid2010, IEEE Computer Society Press 2010.
- [DPK10] C. Docan, M. Parashar and S. Klasky, DataSpaces? : An Interaction and Coordination Framework for Coupled Simulation Workflows, Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (HPDC 2010), Chicago, Illinois, USA June, 2010.

- [DVA+10] P. Dreher, M. Vouk, S. Averitt, E. Sills, “An Open Source Option for Cloud Computing in Education and Research,” 2010, to appear.
- [IVY10] Harini Iyer, Mladen Vouk and Ustun Yildiz, Automation of Scientific Workflow Construction Using Templates & Patterns, In 1st International Workshop on. Workflow Approaches to New Data-centric Science. held in conjunction with SIGMOD’2010 Submitted.
- [Kam10a] C. Kamath, "Understanding wind ramp events through analysis of historical data," IEEE PES Transmission and Distribution Conference, New Orleans, April 2010.
- [Kam10b] C. Kamath, "Using simple statistical analysis of historical data to understand wind ramp events," LLNL Technical report LLNL-TR-423242, February 2010.
- [KLJ+10] S. Klasky, Qing Liu, Jay Lofstead, Norbert Podhorszki, Hasan Abbasi , CS Chang, Julian Cummings, Divya Dinakar, Ciprian Docan, Stephane Ethier , Ray Grout , Todd Kordenbrock, Zhihong Lin , Xiaosong Ma , Ron Oldfield, Manish Parashar, Alexander Romosan , Nagiza Samatova, Karsten Schwan, Arie Shoshani , Yuan Tian, Matthew Wolf, Weikuan Yu , Fan, Zhang , Fang Zheng, “ADIOS: powering I/O to extreme scale computing”, SciDAC 2010.
- [KPV+10] Sidharth Kumar, Valerio Pascucci, Venkatram Vishwanath, Philip Carns, Robert Latham, Tom Peterka, Michael Papka, and Robert Ross. Towards parallel access of multi-dimensional, multiresolution scientific data. In Proceedings of 2010 Petascale Data Storage Workshop, November 2010.
- [KXL10] C. Kamath, Y. Xiao, and Z. Lin, “Analysis of structures and event size statistics in plasma turbulence: Preliminary results,” Sherwood Fusion Theory Conference, Seattle, April 2010.
- [Latham10] Robert Latham, “Parallel netCDF,” in the Encyclopedia of Parallel Computing, David Padua, editor, Springer, 2010.
- [LMF+10] Lin, H., Ma, X., Feng, W., Samatova, N.F., Coordinating computation and I/O in massively parallel sequence search." IEEE Transactions on Parallel and Distributed Systems, 22(4): 529-543, 2011, doi http://doi.ieeecomputersociety.org/10.1109/TPDS.2010.101.
- [LZZ+10] J. Lofstead, F. Zheng, Q. Liu, S. Klasky, R. Oldfield, T. Kordenbrock, K. Schwan, M. Wolf, “Managing Variability in the IO Performance of Petascale Storage Sytems”, accepted ACM/IEEE SC 2010 Conference (SC’10), 2010.
- [MCA+10] Pierre Mouallem, Daniel Crawl, Ilkay Altintas, Mladen Vouk and Ustun Yildiz, A Fault-Tolerance Architecture for Kepler-based Distributed Scientific Workflows, In 22nd International Conference on Scientific and Statistical Database Management, (SSDBM’2010).
- [Nag10] M.Nagappan, "Analysis of Execution Log Files". Accepted at the Doctoral Symposium track of ICSE 2010, May 4th Cape Town SA
- [NJC+10] A. Ngu, A. Jamnagarwala, G. Chin Jr. , C. Sivaramakrishnan, T. Critchlow, “Context-Aware Scientific Workflow Systems Using KEPLER”, To appear in the International Journal of Business Process Integration. 2010.
- [NV10a] M. Nagappan, M.A. Vouk, "Abstracting Log Lines to Log Event Types for Mining Software System Logs." Accepted as short paper in the 7th IEEE Working Conference on Mining Software Repositories (MSR), 2-3, May, 2010, Cape Town, South Africa.
- [NV10b] M. Nagappan, M.A. Vouk, "Adaptive Logging: A Case Study of Logs from a Cloud Computing Environment". Submitted to 9th IEEE International Symposium on Network Computing and Applications (IEEE NCA10), 15-17th July, Cambridg,e MA, USA.
- [PSA+10] B. Palmer, K. Schuchardt, A. Koontz, R. Jacob, R. Latham, and W. Liao. IO for High Resolution Climate Models. Workshop on High-Resolution Climate Modeling, 2010.
- [Ross10] Robert Ross, “Parallel File Systems,” in the Encyclopedia of Parallel Computing, David Padua, editor, Springer, 2010.
- [SHT+10] E Stephan, T Halter, T Critchlow, P Pinheiro Da Silva, and L Salayandia. "Using Domain Requirements to Achieve Science-Oriented Provenance ." Short paper in The 3rd International Provenance and Annotation Workshop (IPAW'2010). June 2010.

- [SKR10] A. Shoshani, S. Klasky, R. Ross, "Scientific Data Management Challenges and Approaches in the Extreme Scale Era", SciDAC 2010.
- [SKB+10] N. Shah, G. Kora, P. Breimyer, Y. Shpanskaya, N.F. Samatova, pR: Enabling Automatic Parallelization of Data-Parallel Tasks and Interfacing Parallel Computing Libraries in R with Application to Fusion Reaction SimulationsThe R User Conference 2010, July 20-23, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, USA.
- [SSC+10] N. Shah, Y. Shpanskaya, C.-S. Chang, S.-H. Ku, A.V. Melechko, and N.F. Samatova, Automatic and statistically robust spatio-temporal detection and tracking of fusion plasma turbulent fronts with parallel R, SciDAC 2010, July 11-15, Chattanooga, TN.
- [TKP+10] R. Tchoua, S. Klasky, N. Podhorszki, B. Grimm, A. Khan, E. Santos, C.T. Silva, P. Mouallem, M. Vouk. "Collaborative Monitoring and Analysis for Simulation Scientists", In Proceedings of The 2010 International Symposium on Collaborative Technologies and Systems (CTS 2010), 2010
- [VSD10] M.A. Vouk, E. Sills, P. Dreher, "Integration of High-Performance Computing into Cloud Computing Services, Handbook of Cloud Computing, Ed. B. Furht, to appear, 2010.
- [WKK+10] Jianwu Wang, Prakashan Korambath, Seonah Kim, Scott Johnson, Kejian Jin, Daniel Crawl, Ilkay Altintas, Shava Smallen, Bill Labate, Kendall N. Houk. Theoretical Enzyme Design Using the Kepler Scientific Workflows on the Grid. Accepted by 5th Workshop on Computational Chemistry and Its Applications (5th CCA) at International Conference on Computational Science (ICCS 2010).
- [WSS10] K. Wu, A. Shoshani, and K. Stockinger. Analyses of Multi-Level and Multi-Component Compressed Bitmap Indexes. ACM TODS v35, Article 2, 2010
- [XHZ+10] Y. Xiao, I. Holod, W. L. Zhang, S. Klasky, Z. H. Lin, "Fluctuation characteristics and transport properties of collisionless trapped electron mode turbulence", Physics of Plasmas, 17, 2010.
- [YAC+10] Ustun Yildiz, Ilkay Altintas, Daniel Crawl, Pierre Mouallem and Mladen Vouk, "Fault-Tolerance in Dataflow-based Scientific Workflow Management", Submitted to SWF 2010
- [YPM_10] Ustun Yildiz, Pierre Mouallem, Mladen Vouk, Daniel Crawl and Ilkay Altintas, Fault-Tolerance in Dataflow-based Scientific Workflow Management, In 4th IEEE International Workshop on Scientific Workflows (SWF@ICWS 2010). Submitted.
- [YTV10] W. Yu, Y. Tian, and J. S. Vetter, Efficient Zero-Copy Noncontiguous I/O for Globus on InfiniBand? , Proceedings of the Third International Workshop on Parallel Programming Models and Systems Software for High-end Computing. Held in Conjunction with ICPP10, San Diego, CA, 2010.
- [YV10] Weikuan Yu and Jeffrey Vetter, "Initial Characterization of Parallel NFS Implementations," the Sixth International Workshop on System Management Techniques, Processes, and Services (SMTPS10), Atlanta, GA, April 2010.
- [YWX+10] Weikuan Yu, Kesheng Wu, Cong Xu, Arie Shoshani, Wei-Shinn Ku, BMF: Bitmapped Mass Fingerprinting for Fast Protein Identification, Auburn University Technical Report AU-CSSE-PASL/10-TR0, 2010. Available at http://pasl.eng.auburn.edu/pubs/pasl-2010-11-bmp.pdf
- [ZAD+10] F. Zheng, H. Abbasi, C. Docan, J. Lofstead, Q. Liu, S. Klasky, M. Parashar, N. Podhorszki, K. Schwan, M. Wolf, "PreDatA - Preparatory Data Analytics on Peta-Scale Machines", IPDPS 2010, IEEE Computer Society Press 2010.
- [ZSSL10] Zinn Daniel, Shawn Bowers, Sven Köhler, Bertram Ludäscher: Parallelizing XML data-streaming workflows via MapReduce . J. Comput. Syst. Sci. 76(6): 447-463 (2010)

2009

- [ABL09] Manish Kumar Anand, Shawn Bowers, Bertram LudÃ¤scher, A navigation model for exploring scientific workflow provenance graphs. SC-WORKS 2009

- [ABML09a] M. Anand, S. Bowers, T. McPhillips, and B. Ludaescher, Exploring Scientific Workflow Provenance Using Hybrid Queries over Nested Data and Lineage Graphs, In 21st Intl. Conf. on Scientific and Statistical Database Management (SSDBM), New Orleans, 2009, 237 – 254.
- [ABML09b] Manish Kumar Anand, Shawn Bowers, Timothy M. McPhillips, Bertram LudÃ¤scher: Efficient provenance storage over nested data collections. EDBT 2009: 958-969.
- [AWE+09] Hasan Abbasi, Matthew Wolf, Greg Eisenhauer, Scott Klasky, Karsten Schwan, Fang Zheng, DataStager? : Scalable Data Staging Services for Petascale Applications, High Performance Distributed Computing (HPDC) 2009.
- [BH+09] Paul Breimyer, William Hendrix, Guruprasad Kora, Nagiza F. Samatova, "pR: Lightweight, Easy-to-Use Middleware to Plugin Parallel Analytical Computing with R," IKE 2009: 667-673.
- [BK+09] Paul Breimyer, Guruprasad Kora, William Hendrix, Neil Shah, Nagiza F. Samatova, "pR: Automatic parallelization of data-parallel statistical computing codes for R in hybrid multi-node and multi-core environments," IADIS AC (2) 2009: 28-32.
- [BKP+09] R. Barreto, S. Klasky, N. Podhorszki, P. Mouallem, M. Vouk: "Collaboration Portal for Petascale Simulations", 2009 International Symposium on Collaborative Technologies and Systems, (CTS 2009), pp. 384-393, Baltimore, Maryland, USA, May 2009. doi: 10.1109/CTS.2009.5067505.
- [CGS+09] J. Chase, I. Gorton, C. Sivaramakrishnan, J. Almquist, A. Wynne, G. Chin, T. Critchlow, "Kepler + MeDICi -Service-Oriented Scientific Workflow Applications", In Proceedings of International Conference on Web Services (ICWS 2009), Los Angeles, CA, July 2009.
- [CKD+09] C S Chang, S Ku, P Diamond, M Adams, R Barreto, Y Chen, J Cummings, E D'Azevedo, G Dif-Pradalier, S Ethier, L Greengard, T S Hahm, F Hinton, D Keyes, S Klasky, Z Lin, J Lofstead, G Park, S Parker, N Podhorszki, K Schwan, A Shoshani, D Silver, M Wolf, P Worley, H Weitzner, E Yoon and D Zorin, "Whole-volume integrated gyrokinetic simulation of plasma turbulence in realistic diverted-tokamak geometry" IOP conference series SciDAC, 2009.
- [Cri09] T. Critchlow. "Scientific Process Automation Improves Data Interaction", Invited Cover Article for Scientific Computing. September 2009.
- [GBA+09] Antoon Goderis, Christopher Brooks, Ilkay Altintas, Edward A. Lee, Carole A. Goble: Heterogeneous composition of models of computation. Future Generation Comp. Syst. 25(5): 552-560 (2009).
- [GLC+09] Kui Gao, Wei-keng Liao, Alok Choudhary, Robert Ross, and Robert Latham. Combining I/O Operations for Multiple Array Variables in Parallel NetCDF? . In the Proceedings of the Workshop on Interfaces and Architectures for Scientific Data Storage, held in conjunction with the IEEE Cluster Conference, New Orleans, Louisiana, September 2009.
- [IBC+09] Florin Isaila, Francisco Javier Garcia Blas, Jesus Carretero, Wei-keng Liao, and Alok Choudhary. A Scalable Message Passing Interface Implementation of an Ad-Hoc Parallel I/O System. In the International Journal of High Performance Computing Applications, October 5, 2009.
- [Kam09] C. Kamath, Scientific Data Mining: A Practical Perspective, SIAM, Philadelphia, PA, 2009.
- [KGH+09] Wesley Kendall, Markus Glatter, Jian Huang, Thomas Peterka, Robert Latham, and Robert Ross. Terascale data organization for discovering multivariate climatic trends. In Proceedings of Supercomputing, November 2009.
- [LAB+09] B. LudÃ¤scher, I. Altintas, S. Bowers, J. Cummings, T. Critchlow, E. Deelman, D. D. Roure, J. Freire, C. Goble, M. Jones, S. Klasky, T. McPhillips, N. Podhorszki, C. Silva, I. Taylor, and M. Vouk. In A. Shoshani and D. Rotem, editors Scientific Process Automation and Workflow Management, Scientific Data Management: Challenges, Existing Technology, and Deployment, Computational Science Series, chapter 13. Chapman & Hall/CRC, 2009.
- [LBM09] B. Ludaescher, S. Bowers, and T. McPhillips. M. T. Ã–zsu and L. Liu, editors, Scientific Workflows, Encyclopedia of Database Systems. Springer, 2009

- [LCL+09] Samuel Lang, Philip Carns, Robert Latham, Robert Ross, Kevin Harms, and William Allcock, "I/O Performance Challenges at Leadership Scale," Proceedings of Supercomputing, November 2009.
- [LWM+09] Scientific Workflows: Business as Usual? , B. Ludaescher, M. Weske, T. McPhillips, S. Bowers. In 7th Intl. Conf. on Business Process Management (BPM), Ulm, Germany, 2009.
- [LZK+09] Jay Lofstead, Fang Zheng, Scott Klasky, Karsten Schwan, Adaptable, Metadata Rich IO Methods for Portable High Performance IO, IEEE International Parallel & Distributed Processing Symposium (IPDPS) 2009.
- [MBZL09] Timothy M. McPhillips, Shawn Bowers, Daniel Zinn, Bertram LudÃ¤scher: Scientific workflow design for mere mortals. Future Generation Comp. Syst. 25(5): 541-551 (2009).
- [MVK+09] Pierre Mouallem, Mladen Vouk, Scott Klasky, Norbert Podhorszki and Roselyne Barreto: “Tracking Files Using the Kepler Provenance Framework”, Proceedings of 21st International Conference on Scientific and Statistical Database Management, SSDBM’09, LNCS 5566, pp. 273-282, New Orleans, LA, USA, June 2009.
- [NWV09] M. Nagappan, K. Wu, M.A. Vouk,. 2009. "Efficiently Extracting Operational Profiles from Execution Logs using Suffix Arrays." 20th International Symposium on Software Reliability Engineering, 16-19 Nov, 2009, Mysuru, India. pp. 41 - 50.
- [RCG+09] Robert Ross, Alok Choudhary, Garth Gibson, and Wei-Keng Liao. Parallel data storage and access. In Arie Shoshani and Doron Rotem, editors, Scientiï¬ c Data Management: Challenges, Technology, and Deployment. Chapman & Hall/CRC, 2009.
- [RCM09] Robert Ross, Philip Carns, and David Metheney. Parallel ï¬ le systems. In Yupo Chan, John Talburt, and Terry Talley, editors, Data Engineering: Mining, Information and Intelligence. Springer, October 2009.
- [RGC+09] Oliver RÃ¼bel, Cameron G R Geddes, Estelle Cormier-Michel, Kesheng Wu, Prabhat, Gunther H Weber, Daniela M Ushizima, Peter Messmer, Hans Hagen, Bernd Hamann, Wes Bethel, Automatic beam path analysis of laser wakefield particle acceleration data. 2009 Comput. Sci. Disc.
- [RPH+09] B. Raman, C. Pan, G. B. Hurst, M. Rodriguez Jr, C. K. McKeown, P.K. Lankford, N. F. Samatova, and J. R. Mielenz, Impact of Pretreated Switchgrass and Biomass Carbohydrates on Clostridium thermocellum Cellulosome Composition, Public Library of Science, PLoS? ONE, 2009.
- [SR09] A. Shoshani and D. Rotem (Editors), Scientific Data Management: Challenges, Technology, and Deployment, Chapman & Hall/CRC Computational Science Series, December 2009.
- [STK+09] E. Santos, J. Tierny, A. Khan, B. Grimm, L. Lins, J. Freire, V. Pascucci, C. Silva, S. Klasky, R. Barreto, N. Podhorszki, “Enabling Advanced Visualization Tools in a Web-Based Simulation Monitoring System”, in IEEE International Conference on eScience 2009.
- [WCA+09] Jianwu Wang, Daniel Crawl, Ilkay Altintas. Kepler + Hadoop – A General Architecture Facilitating Data-Intensive Applications in Scientific Workflow Systems. In Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science (WORKS09) at Supercomputing 2009 (SC2009) Conference. ACM 2009, ISBN 978-1-60558-717-2.
- [YDV09] Weikuan Yu, Oleg Drokin, Jeffrey S. Vetter, Design, Implementation, and Evaluation of Transparent pNFS on Lustre, IEEE International Parallel & Distributed Processing Symposium (IPDPS) 2009.
- [YGC09c] U. Yildiz, A. Guabtni, and A. H. H. Ngu, Business versus Scientific Workflow: A Comparative Study, 2009. Research Report, Department of Computer Science, University of California, Davis, CSE-2009-3, http://www.cs.ucdavis.edu/research/tech-reports/2009/CSE-2009-3.pdf.
- [YGN09a] U. Yildiz, A. Guabtni, and A. H. H. Ngu, Towards scientific workflow patterns," in Proceedings of the 4th Workshop on Workflows in Support of Large-Scale Science, In conjunction with Super Computing, SC (USA), ACM Press, 2009.
- [YGN09b] U. Yildiz, A. Guabtni, and A. H. H. Ngu, Business versus scientific workflows: A comparative study," in Proceedings of the IEEE Third International Workshop on Scientific

Workflows, SWF (In conjunction with 7th IEEE International Conference on Web Services (ICWS 2009)), (USA), IEEE Computer Society Press, 2009.
- [ZBML09] Daniel Zinn, Shawn Bowers, Timothy M. McPhillips, Bertram LudÃ¤scher: X-CSR: Dataflow Optimization for Distributed XML Process Pipelines. ICDE 2009: 577-580
- [ZSML09b] Daniel Zinn, Shawn Bowers, Timothy M. McPhillips, Bertram LudÃ¤scher: Scientific workflow design with data assembly lines. SC-WORKS 2009

2008

- [AAH+08] E. W. Anderson, J. Ahrens, K. Heitmann, S. Habib, and C. Silva, Provenance in Comparative Analysis: A Study in Cosmology, Computing in Science and Engineering, 10(3):30-37, 2008
- [BGKS08] Breimyer, P., Green, N., Kumar, V., Samatova, N.F., BioDEAL: Biological data-evidence-annotation linkage system, Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2008), Philadelphia, PA, USA, Nov. 7-9, 2008.
- [BHF+08] A. Baptista, B. Howe, J. Freire, D. Maier, and C. Silva, Scientific Exploration in the Era of Ocean Observatories, AIP/IEEE Computing in Science and Engineering, pp. 53-58, May/June 2008 (Vol. 10, No. 3).
- [BMR+08] Shawn Bowers, Timothy McPhillips, Sean Riddle, Manish Anand, Bertram Ludaescher, Kepler/pPOD: Scientific Workflow and Provenance Support for Assembling the Tree of Life, in the proceedings of the International Provenance and Annotation Workshop (IPAW) June 17-18, 2008, Salt Lake City, published in "Provenance and Annotation of Data and Processes, editors J. Freire, D. Koop and L. Moreau, Springer-Verlag, Berlin Heidelberg, 2008, 70-77.
- [BPL08] Shawn Bowers, Timothy M. McPhillips, Bertram Ludaescher: Provenance in collection-oriented scientific workflows. Concurrency and Computation: Practice and Experience 20(5): 519-529 (2008)
- [CA08] Daniel Crawl and Ilkay Altintas, A Provenance-Based Fault Tolerance Mechanism for Scientific Workflows, in the proceedings of the International Provenance and Annotation Workshop (IPAW) June 17-18, 2008, Salt Lake City, published in "Provenance and Annotation of Data and Processes, editors J. Freire, D. Koop and L. Moreau, Springer-Verlag, Berlin Heidelberg, 2008, pp. 152-159
- [CFS+08] S. P. Callahan, J. Freire, C. E. Scheidegger, C. Silva, and Huy T. Vo, Towards Provenance-Enabling ParaView, in the proceedings of the International Provenance and Annotation Workshop (IPAW) June 17-18, 2008, Salt Lake City, published in "Provenance and Annotation of Data and Processes, editors J. Freire, D. Koop and L. Moreau, Springer-Verlag, Berlin Heidelberg, 2008, pp. 120-127
- [CKC+08] C S Chang, S Klasky, J Cummings, R. Samtaney, A Shoshani, et al.: "Towards a first-principles integrated simulation of tokamak edge plasma", Journal of Physics: Conference Series 125 (2008) 012042. Conference on Scientific Discovery through Advanced Computing (SciDAC), Seattle, WA, USA, 13-17 July 2008.
- [CSC+08] Chase, J., Schuchardt, K., Chin Jr., G., Daily, J., and Schiebe, T., Iterative Workflows in Numerical Simulations, Proceedings of the IEEE 2008 Second International Workshop on Scientific Workflows (SWF 2008).
- [EKA+08] T. Ellkvist, D. Koop, E. W. Anderson, J. Freire, and C. Silva, Using Provenance to Support Real-Time Collaborative Design of Workflows, in the proceedings of the International Provenance and Annotation Workshop (IPAW) June 17-18, 2008, Salt Lake City, published in "Provenance and Annotation of Data and Processes, editors J. Freire, D. Koop and L. Moreau, Springer-Verlag, Berlin Heidelberg, 2008, pp. 266-279

- [FKS+08] J. Freire, D. Koop, E. Santos, and C. Silva, Provenance for Computational Tasks: A Survey, Computing in Science and Engineering, 10(3):11-21, 2008, http://www.cs.utah.edu/~juliana/pub/freire-cise2008.pdf
- [FS08] Juliana Freire and Claudio T. Silva, Towards Enabling Social Analysis of Scientific Data, CHI Social Data Analysis Workshop, CHI 2008, April 5 – April 10, 2008, Florence, Italy http://www.cs.utah.edu/~juliana/pub/freire-sda-chi2008.pdf
- [GBA+08] Antoon Goderis, Christopher Brooks, Ilkay Altintas, Edward A. Lee and Carole Goble, Composing Heterogeneous Models of Computation in Kepler and Ptolemy II, accepted for publication in Future Generation Computer Systems (FGCS) in 2008. Early version published as a Technical Report, No. UCB/EECS-2007-139 EECS Department University of California Berkeley, November 27, 2007
- [GK08] A. Gezahegne and C. Kamath, "Tracking non-rigid structures in computer simulations," IEEE International Conference on Image Processing, San Diego, October 2008, pp. 1548-1551.
- [GKK08] A. Goodman, C. Kamath, V. Kumar, "Data analysis in the 21st century," Editorial, Statistical Analysis and Data Mining, Vol. 1, Issue 1, Feb 2008, pp 1-3.
- [Kam08a] C. Kamath, "Application-driven data analysis", Editorial, Statistical Analysis and Data Mining, Vol 1, Issue 5, 2008.
- [Kam08b] C. Kamath, "Sapphire: Experiences in Scientific Data Mining,", SciDAC 2008, Journal of Physics Conference Series 125, 012094, July 2008.
- [Kam08c] C. Kamath, "Scientific data mining – why is it difficult?" Invited presentation, Workshop on Algorithms for Modern Massive Data Sets, Stanford University, June 2008.
- [KCC+08] W. P. Kegelmeyer, R. Calderbank, T. Critchlow, L. Jameson, C. Kamath, J. Meza, N. Samatova, and A. Wilson, "Mathematics for Analysis of Petascale Data: Report on a Department of energy Workshop", June 2008.
- [KFS08] David Koop, Juliana Freire, and Claudio T. Silva, Querying and Re-Using Workflows with VisTrails Carlos E. Scheidegger, Proceedings of ACM SIGMOD International Conference on Management of Data, 2008, to appear. http://www.cs.utah.edu/~juliana/pub/vistrails-sigmod2008.pdf
- [KVP+08] Scott Klasky, Mladen Vouk, Manish Parashar, Ayla Kahn, Norbert Podhorszki, Roselyne Barreto, Deborah Silver, Steve Parker, "Collaborative Visualization Spaces for Petascale Simulations," Proceedings of the 2008 International Symposium on Collaborative Technologies and Systems (CTS 2008), 10-23, May, 2008, pp.
- [LKA+08] L. Lins, D. Koop, E. W. Anderson, S. P. Callahan, E. Santos, C. E. Scheidegger, J. Freire, and C. Silva, "Examining Statistics of Workflow Evolution Provenance: A First Study," Statistical and Scientific Database Management, 20th International Conference, SSDBM 2008, Bertram Ludäscher, Nikos Mamoulis (Eds.), Hong Kong, China, July 9-11, 2008, Proceedings. Lecture Notes in Computer Science 5069 Springer 2008, ISBN 978-3-540-69476-2, pp. 573-579, http://www.cs.utah.edu/~juliana/pub/vistrails-ssdbm2008.pdf
- [LML+08] Lin H., Ma X., Li J., Yu T., Samatova N.F., Adaptive Request Scheduling for Parallel Scientific Web Services, Proceedings of the 20th International Conference on Scientific and Statistical Database Management (SSDBM '08), Hong Kong, Jul 9-11, 2008.
- [LPA+08] Bertram Ludaescher, Norbert Podhorszki, Ilkay Altintas, Shawn Bowers, Timothy M. McPhillips: From computation models to models of provenance: the RWS approach. Concurrency and Computation: Practice and Experience 20(5): 507-518 (2008)
- [LSK+08] J. Lofstead, K. Schwan, S. Klasky, N. Podhorszki, and C. Jin, "Flexible IO and Integration for Scientific Codes", 2008 Workshop on Challenges of Large Applications in Distributed Environments.
- [MBZ+08] T. McPhillips, S. Bowers, D. Zinn, B. Ludaescher, Scientific Workflow Design for Mere Mortals, . Future Generation Computer Systems, 2008
- [Nag08] Mei Nagappan, An Overview of Provenance Collection in Workflow Systems and a Privacy Policy Model to Share the Information, April 11-22, High Performance and Applications Conference,

Oak Ridge, TN, April 11-12, 2008. http://www.nccs.gov/user-support/training-education/workshop-archives/high-performance-computing-and-applications-conference-2008/

- [NBH+08] Anne H. H. Ngu, Shawn Bowers, Nicholas Haasch, Timothy M. McPhillips, Terence Critchlow, "Flexible Scientific Workflow Modeling Using Frames, Templates, and Dynamic Embedding," Statistical and Scientific Database Management, 20th International Conference, SSDBM 2008, Bertram Ludäscher, Nikos Mamoulis (Eds.), Hong Kong, China, July 9-11, 2008, Proceedings. Lecture Notes in Computer Science 5069 Springer 2008, ISBN 978-3-540-69476-2, pp. 566-572
- [NLC+08] Arifa Nisar, Wei-keng Liao, and Alok Choudhary. Scaling Parallel I/O Performance through I/O Delegate and Caching System In the Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis , Austin, Texas, November 2008.
- [NV08] Meiyappan Nagappan, and Mladen Vouk, A Privacy Policy Model for Sharing of Provenance Information in a Query Based System, in the proceedings of the International Provenance and Annotation Workshop (IPAW) June 17-18, 2008, Salt Lake City, published in "Provenance and Annotation of Data and Processes, editors J. Freire, D. Koop and L. Moreau, Springer-Verlag, Berlin Heidelberg, 2008, pp. 62-69.
- [PN08] Juan Piernas, Jarek Nieplocha, "Efficient Management of Complex Striped Files in Active Storage", Proc. Europar'08. 2008.
- [POL+08] Pan, C., Oda, Y., Lankford, P.K., Zhang, B., Samatova, N.F., Pelletier, D.A., Harwood, C.S., Hettich, R.L., Characterization of anaerobic catabolism of p-coumarate in Rhodopseudomonas palustris by integrating transcriptomics and quantitative proteomics, Mol. Cell. Proteomics, vol. 7, no. 5, pp. 938-48, 2008, PMID: 18156135.
- [RPW+08] O. Rubel, Prabhat, K. Wu, H. Childs, J. Meredith, C.G.R. Geddes, E. Cormier-Michel, S. Ahern, G.H. Weber, P. Messer, H. Hagen, B. Hamann, and E.W. Bethel. High Performance Multivariate Visual Data Exploration for Extremely Large Data. SC08. 2008.
- [SBH+08] Samatova, N.F., Breimyer, P., Hendrix, W., Schmidt, M.C., Rhyne, T.-M., An Outlook into Ultra-Scale Visualization of Large-Scale Biological Data, Supercomputing 2008, Ultra-Scale Visualization Workshop (invited).
- [SCW+08] K. Stockinger, J. Cieslewicz, K. Wu, D. Rotem, and A. Shoshani. Using Bitmap Indexing Technology for Using Bitmap Indexing Technology for Combined Numerical and Text Queries. New Trends in Data Warehousing and Data Analysis, Annals of Information Systems. Vol 3. Pages 1-23. 2008. Tech Report LBNL-61768
- [SKS+08] C. Scheidegger, D. Koop, E. Santos, H. Vo, S. Callahan, J. Freire, and C. Silva, Tackling the Provenance Challenge One Layer at a Time, Concurrency And Computation: Practice And Experience, 20(5):473--483, 2008
- [SLA+08] E. Santos, L. Lins, J. P. Ahrens, J. Freire, and C. Silva, A First Study on Clustering Collections of Workflow Graphs, in the proceedings of the International Provenance and Annotation Workshop (IPAW) June 17-18, 2008, Salt Lake City, published in "Provenance and Annotation of Data and Processes, editors J. Freire, D. Koop and L. Moreau, Springer-Verlag, Berlin Heidelberg, 2008, pp. 160-173.
- [SMW+08] R. Sinha, M. Winslett, K. Wu, K. Stockinger, and A. Shoshani. Adaptive Bitmap Indexes for Space-Constrained Systems. ICDE 2008, pages 1418--1420.
- [ST08] C. Silva and J. Tohline, Guest Editorial: Special Issue on Computational Provenance, AIP/IEEE Computing in Science and Engineering May/June 2008 (Vol. 10, No. 3), pp. 9-10, , http://csdl2.computer.org/comp/mags/cs/2008/03/mcs2008030009.pdf
- [SVK+08] C. E. Scheidegger, H. T. Vo, D. Koop, J. Freire, and C. Silva. "Querying and Re-Using Workflows with VisTrails," ACM SIGMOD 2008, pp. 1251-1254
- [WSS08] K. Wu, K. Stockinger, and A. Shoshani. Breaking the Curse of Cardinality on Bitmap Indexes. SSDBM. 2008. Tech Report LBNL-173E

- [YOC+08] Weikuan Yu, Sarp Oral, Shane Canon, Jeffrey Vetter, Ramanan Sankaran. Empirical Analysis of a Large-Scale Hierarchical Storage System . The 14th European Conference on Parallel and Distributed Computing (Euro-Par 2008). August 2008. Spain.
- [YV08] W. Yu, J.S. Vetter. ParColl: Partitioned Collective IO on the Cray XT. The 37th International Conference on Parallel Processing. 2008.
- [YVO08] Weikuan Yu, Jeffrey S. Vetter, H. Sarp Oral, Performance Characterization and Optimization of Parallel I/O on the Cray XT. 22nd IEEE International Parallel and Distributed Processing Symposium (IPDPS'08). April 2008. Miami, FL.
- [Zin08] Modeling and Optimization of Scientific Workflows, Daniel Zinn, EDBT PhD Workshop, Nantes, France, 2008

2007

- [ACC+07] Ilkay Altintas, George Chin, Daniel Crawl, Terence Critchlow, David Koop, Jeff Ligon, Bertram Ludaescher, Pierre Mouallem1, Meiyappan Nagappan, Norbert Podhorszki, Claudio Silva, MladenVouk, “Provenance in Kepler-based Scientific Workflow Systems,” Poster # 41, at Microsoft eScience Workshop Friday Center, University of North Carolina, Chapell Hill, NC, October 13 - 15, 2007, pp. 82
- [BCK+07] Roselyne Barreto, Terence Critchlow, Ayla Khan, Scott Klasky,, Leena Kora, Jeffrey Ligon, Pierre Mouallem, Meiyappan Nagappan, Norbert Podhorszki, Mladen Vouk, “Managing and Monitoring Scientific Workflows through Dashboards, Poster # 93, at Microsoft eScience Workshop Friday Center, University of North Carolina, Chapell Hill, NC, October 13 - 15, 2007, pp. 108
- [CRL+07] A. Ching, R. Ross, W.-K. Liao, L. Ward, A. and Choudhary, “Noncontiguous Locking Techniques for Parallel File Systems,” SC2007, November 2007.
- [DEL+07] Provenance in scientific workflow systems, S. Davidson, S. Boulakia, A. Eyal, B. Ludaescher, T. McPhillips, S. Bowers, M. Anand, and J. Freire, IEEE Data Eng. Bull 30(4):44-50, 2007.
- [GBA+07] Antoon Goderis, Christopher Brooks, Ilkay Altintas, Edward A. Lee and Carole Goble. Composing Different Models of Computation in Kepler and Ptolemy II, Proc. of the 2nd Int. Workshop on Workflow Systems in e-Science (WSES 07) in conjunction with the Int. Conference on Computational Science (ICCS) 2007, Beijing, China, May 27-30, 2007
- [LCC+07] Wei-keng Liao, Avery Ching, Kenin Coloma, Alok Choudhary, and Lee Ward. An Implementation and Evaluation of Client-side File Caching for MPI-IO. In the Proceedings of the 21th International Parallel and Distributed Processing Symposium (IPDPS), Long Beach, California, March 2007.
- [LCC+07a] Wei-keng Liao, Avery Ching, Kenin Coloma, Alok Choudhary, and Mahmut Kandemir. Improving MPI Independent Write Performance Using A Two-Stage Write-Behind Buffering Method. In the Proceedings of the Next Generation Software (NGS) Workshop, held in conjunction with the 21th International Parallel and Distributed Processing Symposium (IPDPS), Long Beach, California, March 2007.
- [LCC+07b] Wei-keng Liao, Kenin Coloma, Alok Choudhary, and Lee Ward, “Cooperative Client-side File Caching for MPI Applications,” In the International Journal of High Performance Computing Applications, Volume 21, Number 2, pp. 144-154, 2007.

- [LCC+07c] Wei-keng Liao, Avery Ching, Kenin Coloma, Arifa Nisar, Alok Choudhary, Jackie Chen, Ramanan Sankaran, and Scott Klasky. Using MPI File Caching to Improve Parallel Write Performance for Large-Scale Scientific Applications. In the Proceedings of the ACM/IEEE Conference on Supercomputing, November 2007.
- [LK07] N. S. Love and C. Kamath, "Image analysis for the identification of coherent structures in plasma," Applications of Digital Image Processing, XXX, SPIE Conference 6696, San Diego, August 2007.]
- [LKAS07] J. Lofstead, S. Klasky, H. Abbasi, and K. Schwan, "Adaptable IO System (ADIOS) for Scientific Codes", Supercomputing 2007 Petascale Data Storage Workshop.
- [LML+07] Lin H, Ma X, Li J, Samatova NF, Ting Y, Processor and Data Scheduling for Online Parallel Sequence Database Servers, August 2007: NCSU TR-2007-23.
- [LMY+07] Li J, Ma X, Yoginath S, Kora G, Samatova NF, Automatic, Transparent Runtime Parallelization of the R Scripting Language, January 2007: NCSU TR-2007-3.
- [LRT07] R. Latham, R. Ross, and R. Thakur, "Implementing MPI-IO Atomic Mode and Shared File Pointers Using MPI One-Sided Communication," Int'l Journal of High Performance Computing Applications, (21)2:132–143, Summer 2007.
- [MLS07] Ma, X.; Li, J.; Samatova, N.F., Automatic Parallelization of Scripting Languages: Toward Transparent Desktop Parallel Computing, Proceedings of IEEE/ACS International Conference on Parallel and Distributed Processing Symposium (IPDPS 2007), pp. 1-6, 26-30 March 2007, doi: 10.1109/IPDPS.2007.370488.
- [OOW07] E. O'Neil, P. O'Neil and K. Wu. Bitmap Index Design Choices and Their Performance Implications. In IDEAS 2007. [LBNL-62756]
- [OR07] Ekow Otoo and Doron Rotem, Parallel Access of Out-Of-Core Dense Extendible Arrays, Cluster Computing, Austin, Texas, 2007.
- [PLK07] Workflow automation for processing plasma fusion simulation data, N. Podhorszki, B. Ludaescher, S. Klasky, 2nd Workshop on Workflows in Support of Large-Scale Science (WORKS 07).
- [PLK07a] Archive Migration through Workflow Automation, N. Podhorszki, B. Ludaescher, S. Klasky, Intl. Conf. on Parallel and Distributed Computing and Systems (PDCS), November 1921, 2007, Cambridge, Massachusetts.
- [PN07] Juan Piernas, Jarek Nieplocha, Evan J. Felix. "Evaluation of Active Storage Strategies for the Lustre Parallel File System". Proceedings of the Supercomputing'07 Conference, November, 2007.
- [POS07] Park BH, Ostrouchov G, Samatova NF., Sampling streaming data with replacement, Comput. Stat. Data Anal., vol. 52, no. 2, pp. 750-762, Oct 2007, PMID: 18304937.
- [RSW+07] F. Reiss, K. Stockinger, K. Wu, A. Shoshani, and J. M. Hellerstein. Enabling Real-Time Querying of Live and Historical Stream Data. In SSDBM 2007. [LBNL-61080]
- [SAC+07] Arie Shoshani, Ilkay Altintas, Alok Choudhary, Terence Critchlow, Chandrika Kamath, Bertram Ludäscher, Jarek Nieplocha, Steve Parker, Rob Ross, Nagiza Samatova, Mladen Vouk , :SDM Center Technologies for Accelerating Scientific Discoveries," SciDAC 2007 Proceedings, Dec 2007, Journal of Physics, Conference Series, Vol. 78, paper #012068, 5 pages.
- [SAC+07a] Arie Shoshani, Ilkay Altintas, Alok Choudhary, Terence Critchlow, Chandrika Kamath, Bertram Ludaescher, Jarek Nieplocha, Steve Parker, Rob Ross, Nagiza Samatova, Mladen Vouk,

Scientific Data Management: Essential Technology for Accelerating Scientific Discoveries, CTWatch Quarterly, Volume 3, Number 4, November 2007.

- [SFC07] C. Silva, J. Freire, and S. P. Callahan, C. Silva, J. Freire, and S. P. Callahan, Provenance for Visualizations: Reproducibility and Beyond, IEEE Computing in Science and Engineering, 9(5):82-89, 2007
- [SGU+07] Samatova NF, Gorin A, Uberbacher E, Karpinets T, Park BH, Pan C, Straatsma TP, Cannon W, Resat H, Lins RD, Oehmen C, BioPilot: Data-driven computing for biological systems. SciDAC Review, v. 5, p. 10-25, Fall 2007.
- [SJH+07] Sisneros, R., Jones, C., Huang, J., Gao, J., Park, B.H., Samatova, N.F., A multi-level cache model for run-time optimization of remote visualization, IEEE Trans Vis Comput. Graph, vol. 13, no. 5, pp. 991-1003, Sep-Oct 2007, PMID: 17622682.
- [SW07] K. Stockinger, K. Wu. Bitmap Indices for Data Warehouses. In Wrembel R., Koncilia Ch.: Data Warehouses and OLAP: Concepts, Architectures and Solutions. Idea Group, Inc. [LBNL-59952].
- [VAB+07] M. A. Vouk, I. Altintas, R. Barreto, J. Blondin, Z.Cheng, T. Critchlow, A. Khan, S. Klasky, J. Ligon, B. Ludaescher, P. A. Mouallem, S. Parker, N. Podhorszki, A. Shoshani, C. Silva: "Automation of Network-Based Scientific Worflows" published “Grid-based Problem Solving Envirionments, IFIP, Volume 239, eds. Gaffney. PW, Pool JCT (Boston: Springer), pp 35-61, 2007.
- [WSS07] K. Wu, K. Stockinger and A. Shoshani. Performance of Multi-Level and Multi-Component Compressed Bitmap Indexes. LBNL Tech Report LBNL-60891.
- [YOV+07] Weikuan Yu, Sarp Oral, Jeffrey Vetter, Richard Barrett. Efficiency Evaluation of Cray XT Parallel IO Stack. Cray User Group Meeting (CUG 2007), May 2007. Seattle, WA.
- [YVC+07] Weikuan Yu, Jeffrey Vetter, R. Shane Canon, Song Jiang. Exploiting Lustre File Joining for Effective Collective IO . Seventh IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2007), May 2007.
- [YVC07] Weikuan Yu, Jeffrey S. Vetter, R. Shane Canon, OPAL: An Open-Source MPI-IO Library over Cray XT. International Workshop on Storage Network Architecture and Parallel I/O (SNAPI'07). September 2007. San Diego, CA.

References

- [ABB+03] Altintas I., S. Bhagwanani, D. Buttler, S. Chandra, Z. Cheng, M. Coleman, T. Critchlow, A. Gupta, W. Han, L. Liu, B. Ludaescher, C. Pu, R. Moore, A. Shoshani, M. Vouk, "A Modeling and Execution Environment for Distributed Scientific Workflows," Proc. 15th IEEE International Conference on Scientific and Statisical Database Management (SSDBM 2003).
- [ABD+99] LAPACK: E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen., LAPACK Users' Guide, Third Edition. ed. Philadelphia: SIAM, 1999.
- [AJB+04] Altintas, E. Jaeger, C. Berkley, M. Jones, B. Ludascher, and S. Mock, “Kepler: An Extensible System for Design and Execution of Scientific Workflows, 16th Intl. Conf. on Scientific and Statistical Database Management (SSDBM), Santorini, Greece, 2004 http://users.sdsc.edu/~ludaesch/Paper/ssdbm04-kepler.pdf
- [BCC+97] ScaLAPACK: L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley, ScaLAPACK User's Guide, 1997.
- [BLAST09] mpiBLAST: Open-Source Parallel BLAST, http://www.mpiblast.org/

- [BVP00] Balay R.I, Vouk M.A., Perros H., "Performance of Network-Based Problem-Solving Environments," Chapter 18, in Enabling Technologies for Computational Science Frameworks, Middleware and Environments, editors Elias N. Houstis, John R. Rice, Efstratios Gallopoulos, Randall Bramley, Hardbound, ISBN 0-7923-7809-1, 2000
- [CAL+06] Avery Ching, Alok Choudhary, Wei-keng Liao, Lee Ward, and Neil Pundit. Evaluating I/O characteristics and methods for storing structured scientific data. In Proceedings of the International Parallel and Distributed Processing Symposium, April 2006.
- [CCA06] http://www.cca-forum.org/, accessed February 2006
- [CCPP99] Fabio Casati, Stefano Ceri, Stefano Paraboschi, and Giuseppe Pozzi, "Specification and Implementation of Exceptions in Workflow Management Systems," ACM Transactions on Database Systems 24(3), Sept. 1999
- [CFL+06] Avery Ching, Wu-chun Feng, Heshan Lin, Xiaosong Ma, and Alok Choudhary. Exploring I/O strategies for parallel sequence database search tools with S3aSim. In Proceedings of the International Symposium on High Performance Distributed Computing, June 2006.
- [CL02] I. Crnkovic and M. Larsson (editors), Building Reliable Component-Based Software Systems, Artech House Publishers, ISBN 1-58053-327-2, 2002, http://www.idt.mdh.se/cbse-book/
- [CLR+00] P. Carns, W. Ligon III, R. Ross, and R. Thakur, "PVFS: A Parallel File System For Linux Clusters," Proceedings of the 4th Annual Linux Showcase and Conference, Atlanta, GA, October 2000, pp. 317–327.
- [DBN+96] R.L. Dennis, D.W. Byun, J.H. Novak, K.J. Galluppi, C.C. Coats, and M.A. Vouk, "The Next Generation of Integrated Air Quality Modeling: EPA's Models-3," Atmospheric Environment, accepted, in print, expected 1996.
- [DOE04] R. Mount et al., Department of Energy, Office of Science report, "Data Management Challenge". Nov 2004, http://www.er.doe.gov/ascr/Final-report-v26.pdf
- [EBV95] Elmaghraby S.E., Baxter E.I., and Vouk M.A., "An Approach to the Modeling and Analysis of Software Production Processes," Intl. Trans. Operational Res., Vol. 2(1), pp. 117-135, 1995.
- [Elm66] Elmaghraby S.E., "On generalized activity networks," J. Ind. Eng., Vol. 17, 621-631, 1966
- [FFG+05] Evan J. Felix, Kevin Fox, Kevin Regimbal, Jarek Nieplocha. "Active Storage Processing in a Parallel File System". 6th LCI International Conference on Linux Clusters: The HPC Revolution. Chapel Hill, North Carolina, on April 26, 2005.
- [GHS95] D. Georgakopoulos, M. Hornick, and A. Sheth, "An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure,"Distributed and Parallel Databases, Vol. 3(2), April 1995.
- [HA98] Claus Hagen, Gustavo Alonso, "Flexible Exception Handling in the OPERA Process Support System," ICDCS 1998, pp. 526-533
- [HRG+00] Elias N. Houstis, John R. Rice, Efstratios Gallopoulos, Randall Bramley (editors), Enabling Technologies for Computational Science Frameworks, Middleware and Environments, Kluwer-Academic Publishers, Hardbound, ISBN 0-7923-7809-1, 2000.
- [IG96] R. Ihaka and R. Gentleman. R: A Ranguage For Data Analysis And Graphics. Journal of Computational and Graphical Statistics, 5:299-314, 1996.
- [JPR06] ProRata TOOLbox, Journal of Proteome Research, Vol. 5, No. 11, 2006
- [LAB+06] Scientific Workflow Management and the Kepler System, B. Ludaescher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao, Concurrency and Computation: Practice & Experience, 2006.
- [LB90] J.C. Laprie, and C. Beounes, "Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures", IEEE Computer Society Press, Volume 23, Issue 7, Pages: 39 – 51, July 1990.

- [LCC+06] Wei-keng Liao, Kenin Coloma, Alok Choudhary, Lee Ward, Eric Russell, and Neil Pundit. Scalable Design and Implementations for MPI Parallel Overlapping I/O. In the IEEE Transactions on Parallel and Distributed Systems, vol. 17, no. 11, pp. 1264-1276, Nov., 2006.
- [LG05] B. Ludaescher and C. A. Goble, editors. ACM SIGMOD Record, Special Section on Scientific Workflows, volume 34(3), September 2005.
- [LLC+03] J. Li, W. Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale, "Parallel netCDF: A High-Performance Scientific I/O Interface," Proceedings of SC2003, Phoenix, AZ, November 2003.
- [LMC+05] Lin H, Ma X, Chandramohan P, Geist A, Samatova NF, Efficient Data Access for Parallel BLAST, Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005), pp. 72, 04-08 April 2005, doi: 10.1109/IPDPS.2005.190.
- [Mou04] Pierre Mouallem, "Fault Tolerance and Reliability in Scientific Workflows," MS Thesis, NC State University, 2004
- [Mou04] Richard Mount, editor, The Office of Science Data-Management Challenge,Report from the DOE Office of Science Data-Management Workshops, March–May 2004, http://www-user.slac.stanford.edu/rmount/dm-workshop-04/Final-report.pdf.
- [MV96] D.F. McAllister, and M.A. Vouk, "Software Fault-Tolerance Engineering," Chapter 14 in Handbook of Software Reliability Engineering, McGraw Hill, pp. 567-614, January 1996.
- [PGL+07] Swapnil V. Patil, Garth A. Gibson, Sam Lang, and Milo Polte, "GIGA+: scalable directories for shared file systems," Parallel Data Storage Workshop in conjunction with Supercomputing, 2007.
- [PKM+06] Pan C, Kora G, McDonald WH, Tabb DL, VerBerkmoes NC, Hurst GB, Pellletier DA, Samatova NF, Hettich RL., ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation, Anal Chem., vol. 78, no. 20, pp. 7121-31, Oct 2006, PMID: 17037911.
- [PKT+06] Pan C, Kora G, Tabb DL, Pelletier DA, McDonald WH, Hurst GB, Hettich RL, Samatova NF., Robust estimation of peptide abundance ratios and rigorous scoring of their variability and bias in quantitative shotgun proteomics, Anal Chem., vol. 78, no. 20, pp. 7110-20, Oct 2006, PMID: 17037910
- [PST+08] Milo Polte, Jiri Simsa, Wittawat Tantisiriroj, Garth Gibson, "Fast Log-Based Concurrent Writing of Checkpoints," Petascale Data Storage Workshop, November 2008.
- [RPM09] RPM RScaLAPACK Download, http://rpm.pbone.net/index.php3/stat/4/idpl/4766700/com/R-RScaLAPACK-0.5.1-9.fc7.i386.rpm.html
- [SBY05] N. Samatova, D. Bauer, and S. Yoginath, The taskPR Package, pp. 6, May 2005, http://cran.r-project.org/doc/packages/taskPR.pdf
- [SPPJ06] J. Salas, F. Perez, M. Patia-Martinez, R. Jiminez-Peris, "WS-Replication: A Framework for Highly Available Web Services," WWW Conf., Edinburgh, Scotland, May 2006.
- [SS07] H. Shan and J. Shalf, "Using IOR to Analyze the I/O Performance of HPC Platforms," 2007 Cray User Group Conference, Seattle WA, May 2007.
- [SV96] Singh M.P., Vouk M.A., "Scientific workflows: scientific computing meets transactional workflows," Proceedings of the NSF Workshop on Workflow and Process Automation in Information Systems: State-of-the-Art and Future Directions, Univ. Georgia, Athens, GA, USA; 1996, pp.SUPL28-34.
- [TGL99] Rajeev Thakur, William Gropp, and Ewing Lusk, "On Implementing MPI-IO Portably and with High Performance," in Proc. of the Sixth Workshop on I/O in Parallel and Distributed Systems, May 1999, pp. 23-32.
- [Vou05] Mladen A Vouk, "Software Reliability Engineering of Numerical Systems," Chapter 13, in Accuracy and Reliability in Scientific Computing, Editor: Bo Einarsson, ISBN 0-89871-584-9, SIAM, 2005, pp 265-300.

- [VS97] Vouk M.A., and M.P. Singh, "Quality of Service and Scientific Workflows," in The Quality of Numerical Software: Assessment and Enhancements, editor: R. Boisvert, Chapman & Hall, pp.77-89 , 1997.
- [YSB+05] Yoginath S, Samatova NF, Bauer D, Kora G, Fann G, Geist A, RScaLAPACK: High-performance parallel statistical computing with R and ScaLAPACK, Proceedings of the 18th International Conference on Parallel and Distributed Computing Systems (PDCS-2005), Sep 12-14, 2005, Las Vegas, Nevada.
- [Yu09] H. Yu. The Rmpi Package. http://cran.r-project.org/web/packages/Rmpi/index.html, 2009.

Appendix 1: Tutorials, training, thesis, outreach, invited presentations

SDM center

Invited plenary talk: A. Shoshani, "Scientific Data Management: Essential Technology for Data-Intensive Science", SIAM Computational Science & Engineering (CSE'07), February 2007.

Invited to organize session: A. Shoshani, "Data Analysis, Management and Visualization" in DoE workshop on Fusion Energy Sciences at Extreme Scale, March, 2009. Invited participant from the SDM center on the panel were: A. choudhary, C. Kamath, S. Klasky, M. Vouk, and N. Samatova.

Invited poster and short paper: A.Shoshani et al, "SDM center technologies for accelerating scientific discoveries," SciDAC 2007 Conference, http://www.iop.org/EJ/abstract/1742-6596/78/1/012068.

SEA

Invited talk: R. Ross, "Preparing for Exascale: Understanding HPC Storage Systems," Workshop on Interfaces and Abstractions for Scientific Data Storage (IASDS), Heraklion, Crete, Greece, September 2010.

Invited talk: R. Ross, "Data Models and Data Analysis at Exascale," High-End Computing File Systems and I/O Conference, Arlington, VA, August 2010.

Invited talk: Robert Latham, "Parallel I/O in Practice," CScADS Workshop on Leadership-class Machines, Petascale Applications, and Performance Strategies, Snowbird, UT, July 2010.

Invited talk: Robert Latham, "Parallel I/O in Practice," Big Data for Science Workshop, Virtual School of Computational Science and Engineering, July, 2010.

Invited talk: R. Ross, "Scientific Computing at Extreme Scale," University of Connecticut, Storrs, CT, June 2010.

Invited talk: R. Ross, "Storage in an Exascale World," IEEE International Workshop on Storage Network Architecture and Parallel I/Os (SNAPI), Incline Village, NV, May 2010.

Invited talk: R. Ross, "Applications, Data, and the Future of Storage in Computational Science," SCI Institute, University of Utah, Salt Lake City, UT, May 2010.

Invited talk: Robert Latham, "Middleware Libraries for Parallel I/O," Carnegie-Mellon University, Pittsburgh, PA, April 2010.

Invited talk: Robert Latham, "Making the Most of the I/O Software Stack," Extreme Scale I/O and Data Analysis Workshop, Austin, TX, March 2010.

Invited talk: Robert Ross, "Input/Output (I/O) in Computational Science," University of Chicago, Chicago, IL, February 2010.

Invited talk: Alok Choudhary, "Detailed Analysis of I/O Traces of Large Scale Applications", The International Conference on High-Performance Computing, India, Dec. 2009.

Tutorial: Robert Latham, Robert Ross, Marc Unangst, and Brent Welch, "Parallel I/O in Practice," SC 2009, Portland, OR, November 2009.

Tutorial: William Gropp, Ewing Lusk, Robert Ross, and Rajeev Thakur, "Advanced MPI," SC2009, Portland, OR, November 2009.

Tutorial: W. Gropp, E. Lusk, R. Ross, R. Thakur, "Advanced MPI," SC2008, Austin, TX, November 2008.

Tutorial: R. Latham, R. Ross, M. Unangst, and B. Welch, "Parallel I/O in Practice," SC2008, Austin, TX, November 2008.

Tutorial: R. Latham and R. Ross, "Parallel I/O for High Performance Computing," FAST 2008, San Jose, CA, February 2008.

Tutorial: W. Gropp, E. Lusk, R. Ross, R. Thakur, "Advanced MPI," SC2007, Reno, NV, November 2007.

Tutorial: R. Latham, W. Loewe, R. Ross, R. Thakur, "Parallel I/O in Practice," SC2007, Reno, NV, November 2007.

Tutorial: R. Latham and R. Ross, "Parallel I/O: Not Your Job," CScADS Workshop on Petascale Architectures and Performance Strategies, Snowbird, UT, July 2007.

Invited talk: R. Latham, R Ross, R Thakur "High Performance Parallel Data and Storage Management", SIAM Computational Science & Engineering (CSE'07), February 2007.

Tutorial: R. Latham and R. Ross, "Parallel I/O in Practice," SciDAC 2007 Tutorials Workshop, Boston, MA, June 2007.

Tutorial: R. Latham, W. Loewe, R. Ross, R. Thakur, "Parallel I/O in Practice," SC2006, Tampa, FL, November 2006.

Tutorial: E. Lusk, W. Gropp, R. Ross, and R. Thakur, "Advanced MPI: I/O and One-Sided Communication," SC2006, Tampa, FL, November 2006.

Tutorial: R. Latham and R. Ross, "Parallel I/O in Practice," Cluster 2006, Barcelona, Spain, September 2006.

Tutorial: R. Ross and J. Worringen, "High-Performance Parallel I/O," EuroPVM/MPI 2006, Bonn, Germany, September 2006.

DMA

Invited talk: C. Kamath, "Mining Science Data", SIAM Computational Science & Engineering (CSE'07), February 2007.

Invited talk: C. Kamath, "Scientific data mining – why is it difficult?" Invited presentation, Workshop on Algorithms for Modern Massive Data Sets, Stanford University, June 2008.

Invited talk: N.F. Samatova, "From Data to Mathematical Models in Systems Biology," DOE ASCAC-BERAC Committee Meeting, October, 2007.

Invited talk: N.F. Samatova, "An Outlook into Ultra-Scale Visualization of Large-Scale Biological Data," Ultra-Viz Workshop, Supercomputing, 2008.

Invited talk: N.F. Samatova and A. Shoshani, "Scientific Data Management Center: Technologies and Applications," DOE OBER/OASCR Genomics:GTL Workshop, Feb. 2008.

Invited poster and short paper: N.F. Samatova, "High performance statistical computing with parallel R: Applications to biology and climate," SciDAC 2006 Conference.

Invited to organize workshop: W. P. Kegelmeyer, R. Calderbank, T. Critchlow, L. Jameson, C. Kamath, J. Meza, N. Samatova, and A. Wilson, "Mathematics for Analysis of Petascale Data: Report on a Department of Energy Workshop", June 2008.

Invited to organize workshop session: Samatova N.F., "Data Integration" in DOE Genomics:GTL Systems Biology Knowledgebase Workshop chaired by Gregurick S, Fredrickson J, Stevens R., May 2008.

Invited to organize workshop session: Samatova N.F., "Exascale Challenges for Biology" in DOE Exascale Workshop chaired by Horst S, Zacharia T, Stevens R, April, May, June, 2007.

Tutorial: P. Breimyer, contributed a module to "Introduction to scientific workflow management and the Kepler system," tutorial by Altintas; Ilkay, Mladen Vouk, Scott Klasky, Norbert Podhorszki, Daniel Crawl presented at Supercomputing 2007, Reno, NV, 11 Nov 07, Complete materials are available on-line at: http://sdm.ncsu.edu/tutorial/ or http://groups.google.com/group/spa-tutorial

SPA

Tutorial: Provided the Fusion Simulation Project hands-on training. Slides for the training are available at http://users.nccs.gov/~sklasky/effis.pdf. January 2009

Invited talk: T. Critchlow, I. Altintas, S. Klasky, M. Vouk, S. Parker, B. Ludaescher, "Accelerating Scientific Exploration Using Workflow Automation Systems", SIAM Computational Science & Engineering (CSE'07), February 2007.

Invited talk: M.A. Vouk, "Automation of Large-scale Network-Based Scientific Workflows using Kepler: Tools and Case Studies," IFIP Working Conference 9 on Grid-based Problem Solving Environments: Implications for Development and Deployment of Numerical Software, IFIP Working Group 2.5 on Numerical Software, Prescott, AZ, July 17-21, 2006

Invited talk: Ilkay Altintas, et al, Accelerating the scientific exploration process with scientific workflows, SciDAC 2006 conference, http://www.iop.org/EJ/abstract/1742-6596/46/1/065.

Invited Talk: Mladen Vouk, "Automation of Large-scale Network-Based Scientific Workflows using Kepler: Tools and Case Studies," Computer Science Seminar, UNC Charlotte, Sept. 1, 2006

Workshop Presentation: Mladen Vouk, "Automation of Large-scale Network-Based Scientific Workflows," Microsoft eScience, The Johns Hopkins University, Baltimore, Maryland, October 13 - 15, 2006, p21

Tutorial: Ilkay Altintas, Bertram Ludaescher, Scott Klasky, Mladen A. Vouk. "Introduction to scientific workflow management and the Kepler system,", Proceedings of the 2006 ACM/IEEE conference on Supercomputing, Tampa, Florida, Nov 06, Article No. 205, 2006, ISBN:0-7695-2700-0

Tutorial: Scott Klasky, Mladen A. Vouk, Ilkay Altintas, Jeff Ligon, Pierre Moualem, Mei Nagappan, "Scientific Data Workflows," National Center for Computational Sciences (NCCS) 2007 Users Meeting at Oak Ridge National Laboratory, March 27-29, 2007

Tutorial: Altintas; Ilkay, Mladen Vouk, Scott Klasky, Norbert Podhorszki, Daniel Crawl, "Introduction to scientific workflow management and the Kepler system," presented at Supercomputing 2007, Reno, NV, 11 Nov 07, Complete materials are available on-line at: http://sdm.ncsu.edu/tutorial/ or http://groups.google.com/group/spa-tutorial

Tutorial: Timothy McPhillips, Sean Riddle, Manish Anand, Scientific Workflow Support for AToL Data Analysis & Management, Bertram Ludaescher, Shawn Bowers, NSF/AToL meeting, New Orleans, March 8-9, 2008

Invited Talk: Mladen Vouk, "Cloud Computing – Issues, Research and Implementations," the 30th International Symposium on Information Technology Interfaces (ITI) 2008, June 23-26, Cavtat, Croatia

Tutorial: Ilkay Altintas, Mladen Vouk, Scott Klasky, Norbert Podhorszki, "Introduction to scientific workflow management and the Kepler system," SciDaC 2008, Seattle, WA, 18-Jul-08.

Tutorial: Ilkay Altintas, Mladen Vouk, Scott Klasky, Norbert Podhorszki, Daniel Crawl, "Introduction to scientific workflow management and the Kepler system," Supercomputing 2008, Austing, TX, Nov 08

Ph.D and M.S. Thesis (involving SDM center technologies)

- Frederick Ralph Reiss, Data Triage, EECS Department, University of California, Berkeley, Technical Report No. UCB/EECS-2007-79, June 1, 2007, http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-79.pdf
- Rishi Rakesh Sinha, Indexing Scientific Data, Graduate College of the University of Illinois at Urbana-Champaign, 2007, http://usha.cs.uiuc.edu/~rsinha/thesis_skt.pdf
- Zhengang Cheng, "Verifying Commitment Based Business Protocols and their Compositions: Model Checking using Promela and Spin," Ph.D, North Carolina State University, 2006 http://www.lib.ncsu.edu/pubweb/www/ETD-db/web_root/collection/available/etd-08092006-005135/unrestricted/etd.pdf
- Jiangtian Li, "Towards Transparent Parallel Statistical Data Processing on Multi-core Computers" Ph.D, North Carolina State University, 2008.
- Heshan Lin, "High Performance Parallel and Distributed Genomic Sequence Search," Ph.D, North Carolina State University, 2009.
- Chongle Pan, "An Integrated Experimental and Computational Approach to Proteomics: Scaling from High Resolution Qualitative Analysis to Quantitative Measurements with Confidence Evaluation," Ph.D, University of Tennessee, Knoxville, 2006.
- Chandra Mohan, "Efficient In-Database Analytics through Embedding MySQL into R," M.S., North Carolina State University, 2008.

Appendix 2:
Collaboration with Application Projects and other Centers and Institutes

Application Domains	Workflow Technology (Kepler)	Metadata and Provenance	Data Movement and Storage	Indexing (FastBit)	Parallel I/O (pNetCDF, etc.)	Parallel Statistics (pR, …)	Feature Extraction	Active Storage	ADIOS
Climate Modeling (Drake)	monitoring				pNetCDF	pMatlab			
Astrophysics (Mezzacappa)	monitoring, code-couple	dashboard							Accelarate I/O
Combustion (Jackie Chen)	monitoring, code-couple	dashboard	DataMover-Lite	flame front	Global Access	pMatlab	tranient events		Accelarate I/O
Combustion (Bell)			DataMover-Lite						
Fusion (PPPL)							poincare plots		
Fusion (CPES)	monitoring, code-couple	Dashboard	DataMover-Lite	Toroidal meshes		pR	Blob tracking		Accelarate I/O
High Energy Physics	Lattice-QCD		SRM, DataMover	event finding					
Groundwater Modeling	production workflows								
Accelarator Science (Ryne)					MPIO-SRM				
Biology	ScalaBlast					ProRata		ScalaBlast	
Climate Cloud modeling (Randall)					pNetCDF			cloud modeling	
Fusion (RF) (Bachelor)						paralleize GVK	poincare plots		
Laser Wakefield accelarotors				visual analytics					
Climate (CCSM-NCAR)									
other activities	CCSM model runs	provenance capture						integrate with Luster	
Centers & institutions									
Open Science Grid			SRM-tester						
Earth System Grid			SRM and DML						
Petascale Storage Institute (Gibson)					Posix-IO				
Vis Institute (Ma)				query-based vis	put parallel I/O in Vis	pR			
Vis Center (Bethel)				query-based vis		pR			

Legend		currently in progress		problem identified		interest expressed

Appendix 3: Details of activities described in applications vs. technologies table

Below we summarized the activities in the cells of the scientific applications vs. SDM center technologies shown in the table above. The summary is organized by technologies, and for each technology the tasks for each application project are described.

Workflow Technology tasks

- ***Application: Fusion (CPES)***

  Code Coupling: A first workflow demonstrating the coupling of XGC-0 and M3D has been developed. Next steps: Scientists want to couple other codes (e.g. ELITE and NIMROD) for improved accuracy and performance reasons. The goal is to use these workflows in production.

  Monitoring and Archiving: We have developed a workflow which on the fly (i) moves simulation output data to a secondary (remote) resource, (ii) processes (converts) data, (iii) creates images from the data, and (iv) archives the results.

- ***Application: Groundwater Modeling***

  Five different workflows associated with multi-scale simulations of subsurface biogeochemical processes have been identified as potential candidates to be represented and modeled as computational workflows. Of these five, the first workflow under implementation is one focused on a continuum simulation of flow and transport for two non-reacting tracers.

- ***Application: Combustion***

  We have worked with the S3D team to understand the basic workflow requirements. We have started to work with them to run their netcdf output and run this through a series of services which have been constructed through our alliance with the CPES SciDAC project. This includes: splitting netcdf files with infinite time dimensions to one time dimension, joining the files on another system, creating grace and png files from the data, and finally running an avs/express offscreen rendering server to produce 2d colormap images. In order to avoid costly large-scale runs with untested code, we have developed a "preparation workflow" that automatically checks out the latest simulation code from a repository, builds ("makes") the application, and runs it with a number of test cases. We are planning to extend the workflow to handle other domains (e.g. Fusion) in the future.

- ***Application: Biology***

  We are investigating using a simple web interface to define and execute multiple ScalaBLAST (large-scale sequence homology comparisons) workflows and summarize the resulting data set, providing computational biologists a novel and efficient data management capability.

Metadata and Provenance tasks

- ***Application: Combustion***

  We have discussed a need for a simplified tool for day-to-day tracking, analysis, and graphing of simulations that is integrated with the workflow and simulation tracking systems, and are currently exploring the possibility of extending our web-based data management and query tools for use with this project. This will leverage similar work that we are doing with CPES but will require adaptation to the grids used by the combustion simulations, and may require additional analysis tools to be integrated.

- ***Application: Astrophysics (TSI) and Fusion (CPES)***

  Provenance: Considerable progress has been made on unification of the provenance approaches. General classification is in place (process, data, workflow and system) and we are working on the

general solution. Details are at http://www.vistrails.org/index.php/SDM_Provenance. Some astrophysics and fusion specific data schemas are also in place. SDM center main contact: Mladen Vouk.

- ***Applications: Astrophysics (TSI) and Fusion (CPES) and combustion***

  Dashboard: Dashboard activities are progressing very fast. We have a prototype for CPES that is quite sophisticated. The group has regular teleconferences related to tasks and design. The architecture is now solidifying around a data-base centered repository with remote feeds and real-time updates of the job progress and states.

- **Application: Climate**

  We are working with scientists from NCAR to help develop Kepler-driven workflows and provenance for the Community Climate System Model (CCSM). CCSM belongs to an elite category of computer-based simulation models known as general-circulation models. The automatic capturing of provenance is an essential part of this activity, especially as the number and volume of simulations is expected to significantly grow.

Data Movement and Storage tasks

- ***Application: High Energy Physics***

  Storage Resource Managers (SRMs) have been used for several years by High Energy Physics projects. In cooperation with the Open Science Grid (OSG), we continue to support the STAR project in its use of our SRM. This includes its use for large scale robust data movement activity, as well as its use for dynamic data analysis tasks.

- ***Application: Fusion (CPES)***

  We have used SRM-Lite for this project as well in two different ways. The first is for a user to pull files into their workstation of laptop. For this purpose SRM-Lite has a GUI that shows progress of the transfer. The second way is for SRM-Lite to be used by a Kepler actor – future work.

- ***Collaboration: with Open Science Grid***

  LBNL has developed a test-suite for SRMs used extensively by OSG to test the compatibility and adherence to the SRM specification of several SRM implementations in the US and Europe. This work is funded by the OSG.

- ***Collaboration: with Earth System Grid***

  LBNL has been providing SRM software as well as SRM-Lite for several years now. The latest version provided is a new implementation, called the Berkeley Storage Manager (BeStMan). This work continues to evolve. This work is funded by the ESG.

- ***Application: Combustion***

  We have built a new workflow for migrating an archive from one mass storage to another. This workflow enhances earlier work with concurrent transfers over the network. It was successfully used to migrate a 10TB INCITE archive from NERSC to ORNL within 11 days. The data migration workflow has mechanisms to deal with failures, i.e., allows the user to continue the migration even after some intermediate steps have failed (e.g., due to network problems).

Indexing Technology tasks

- ***Application: Combustion***

  We had previously applied Fastbit to develop software for flame front identification, region growing, and region tracking. We also developed a simple GUI application for displaying and tracking features in 2D combustion data. The application has moved on to 3D simulations and requires more sophisticated visualization.

- ***Application: Fusion (CPES)***

  The goal is to use Fastbit technology for searching over data in toroidal meshes. We have identified the problem and the algorithms that could potentially address the problem. We are implementing the algorithms to study the actual performance characteristics.

- ***Application: High-Energy Physics***

  In order to have the broadest impact in this community, we plan to integrate FastBit with the popular ROOT framework. The STAR software team is willing to help with ROOT expertise and manpower for testing. Work scheduled to start by June, 2007.

- ***Collaboration: with the Visualization Center (VACET)***

  We have integrated and deployed FastBit software for visualization applications. This includes extending HDFpart with FastBit indexing as well as real-time analysis of Laser Wakefield Particle Accelerator simulation data.

- ***Collaboration: with UltraScale Visualization Institute (USVI)***

  We have developed a special version of FastBit for indexing data from toroidal meshes. This code also generates regions that are selected by conditions on the variables. In collaboration with USVI, we are working on using this software for real-time explorations of toroidal data from Fusion simulations.

Parallel I/O Technologies tasks

- ***Application: Climate (CCSM)***

  The Community Climate System Model (CCSM) groups are interested in using PnetCDF as a mechanism for improving I/O performance for their large-scale simulations. We are routinely participating in concalls with NCAR and others, and PnetCDF is now an output format for the POP ocean code. Main application contact: John Drake.

- ***Application: Climate (GCRM)***

  The Global Cloud Resolving Model (GCRM) group at PNNL uses netCDF as a storage format. We have been working with their developers to use PnetCDF for better scalability on large systems. Main application contact: Bruce Palmer

- ***Application: Combustion (Jackie Chen)***

  This group is interested in improving overall I/O performance. NWU obtained an I/O kernel and developed approaches to store simulation data in a canonical format that eliminates most post-processing prior to analysis.

- ***Application: Materials (QBOX)***

  This group is interested in improving I/O performance for the QBOX code on the IBM BlueGene systems. We have performed initial experiments at ANL to better understand their I/O patterns. Main contact: Guilia Galli.

- ***Collaboration: with Petascale Data Storage Institute (PDSI)***

  We are interacting with the PDSI to further specify and prototype POSIX I/O extensions for High End Computing (HEC). We have had numerous meetings and email discussions on this topic.

- ***Collaboration: with UltraScale Visualization Institute (USVI)***

  We are discussing I/O concerns with participants in the USVI. We hope to apply parallel I/O techniques in visualization codes that will be used to view petascale simulation data.

- ***Collaboration: with Universal Nuclear Energy Density Functional (UNEDF)***

  We are discussing I/O concerns with participants in the UNEDF. Our goal is to devise I/O approaches to make full utilization of leadership-class machines.

Feature Extraction tasks

- ***Application: Combustion (TSTC)***

  The goal is to develop robust techniques for quantitative identification and tracking of transient events in combustion simulation data. The purpose is to understand the process of ignition, extinction, and re-ignition.

- ***Application: Fusion (CPES)***

  The goal is to characterize and track the blobs in high-resolution, ultra-high-speed images from the gas-puff diagnostic on the NSTX. The purpose is to contribute to the success of devices such as ITER by improving the understanding of the coherent structures and validating or invalidating theories.

  Application: Fusion (RF)

  The goal is the classification and characterization of Poincare plots for simulation and experimental data. The purpose is to use the simulations to drive the experiments and use the experiments to validate the simulations. The package developed achieves high accuracy classification, and is now being used.

- ***Application: Fusion (GPS)***

  The goal is tracking of blobs in a high-dimensional particle simulations. The techniques have been developed and collaboration with the scientists is continuing.

- Application: Fusion (GSEP)

  The goal is to identify coherent structures in fluid and particle data and understand their interactions. The purpose is to gain insights into the effects of energetic particles on the performance of burning plasmas. [To be colored as "in progress"]

- Application: Renewable energy

  The goals are to understand the effects of increased wind energy on the power grid and improve the forecast of energy generated by wind farms through the identification of sensors important to the forecast as well as wind-driven anomalous events on the system. [To be colored as "in progress"]

High Performance Statistical Analysis tasks

- ***Application: Combustion***

  We initiated a dialog on providing parallel Matlab interface to her S3D library. Jackie handed over to us the parts of her Fortran90 library that deals with I/O and would like to get a plug-in of this library into parallel Matlab environment so that the subsequent analysis and visualization capabilities of parallel Matlab could be utilized. She has assigned her PhD student (David Lignel) to help us in this task. Application contact: Jackie Chen.

- ***Application: Fusion (CPES) and Collaboration with the Visualization Center***

  This task is in collaboration with George Ostrouchov and Sean Ahern. The goal is to parallelize their data analysis routines written in R using our parallel R platform. They need to handle data consisting of billion of particles and sequential R is limited for this task.

- ***Application: Climate***

  The spherical harmonic transform is a critical computational kernel of the dynamics portion of spectral atmospheric weather and climate codes. John and his team currently develop and use Matlab library for computing spherical harmonic transforms to solve simple partial differential equations on the sphere. We identified a strategy on how to parallelize this library for them so that it could be applied to more realistic problem sizes using parallel Matlab. Application contact: John Drake

- ***Application: Climate***

Assessment of global climate change impacts requires increasingly finer spatial and temporal resolutions from existing Earth Systems Modeling predictions. Given a fine resolution observational data and a course grain resolution simulation data, statistical downscaling could be applied to learn statistical relationships that link large-scale simulation results with fine grain regional observations. We develop a parallel Matlab library to support that. The library includes a number of components that are routinely used by climate community such as EOF, CCA, MLR, filtering routines. Application contacts: John Drake and George Ostrouchov

- ***Application: Nanoscience (DOE CNMS Center***

  This group is simulating an electron beam induced deposition process using Matlab library. We are providing parallelism to this simulation framework using parallel Matlab to bring the required efficiency. Applicatiuon contact: Philip Rack.

- ***Application: Biology (DOE Genomics:GTL projects)***

  We provide quantitative proteomics capabilities with ProRata. Application of our technologies to a number of problems in GTL community has been demonstrated. Specifically, in collaboration with B. Hettich and Carol Harwood, we reconstructed aromatic compound degradation pathways in a hydrogen producing bacteria using ProRata. Joint paper is under review. Also, we applied ProRata to quantifying the abundance of microbial communities in several DOE contaminated sites. Joint paper is being written and interesting hypotheses are generated about the presence of virus in the community that significantly changed the structure of the communities in one of the two sites. Application contact: R. Hettich, J. Banfield, C. Harwood, M. Buchanan.

- ***Collaboration: with UltraScale Visualization Institute (USVI)***

  We are discussing heterogeneous information analysis and visualization issues with participants in the USVI (Kwan-Lu Ma and Juan Huang). The primary application area is biology. We discuss issues of uncertainty representation in biological networks. We also worked with them on interactive remote visualization. The multi-cache framework with adaptive adjustment of cache parameters using statistical analysis and parameter optimization techniques has been developed and jointly published with Dr. J. Huang.

Appendix 4: NCSU Extension Results (2011-2012)

Publications

1. [DV12] Patrick Dreher, Mladen A. Vouk, “Utilizing Open Source Cloud Computing Environments to Provide Cost Effective Support for University Education and Research,” in Cloud Computing for Teaching and Learning: Strategies for Design and Implementation," IGI Global, Editor Li Chao, 2012, pp 32-49
2. [LAB+10] Bertram Ludaescher, Ilkay Altintas, Shawn Bowers, Juian Cummings, Terence Critchlow, Ewa Deelman, David De Roure, Juliana Freire, Carole Goble, Matthew Jones, Scott Klasky, Timorthy McPhillips, Norber Pdohorszki, Claudio Silva, Ian Taylor, Mladen Vouk, “Scientific Process Automation and Workflow Management, Chapter 13 in *Scientific Data management – Challenges, Technology and Deployment*, Editors: Arie Shoshani and Doron Rotem, CRC Press, 2010, pp, 467-507.
3. [MCA+10] P. Mouallem, D. Crawl, I. Altintas, M. Vouk and U. Yildiz. "A Fault-Tolerance Architecture for Kepler-based Distributed Scientific Workflows". SSDBM 2010, LNCS 6187, pp. 452-460, June 2010
4. [NMV11] Nagappan, M., Murphy, B., Vouk, M.A., "Which Code Construct Metrics are Symptoms of Post Release Failures?", The 2nd International Workshop on Emerging Trends in Software Metrics (WETSoM 2011), 24 May 2011, Honolulu, Hawaii, USA, pp. 65-68.
5. [NPV11] Nagappan, M., Peeler, A., Vouk, M.A., "Modeling Cloud Failure Data: A Case Study of the Virtual Computing Lab,” The ICSE 2011 Software Engineering For Cloud Computing Workshop, May 22, 2011, Honolulu, HI, USA, pp. 8-14.
6. [NV10[Nagappan, M., Vouk, M.A., "Adaptive Logging: A Case Study of Logs from a Cloud Computing Environment,” In the Fast Abstracts track of 21st International Symposium on Software Reliability Engineering, 1-4 Nov, 2010, San Jose, California.
7. [NV10] Nagappan, Meiyappan; Vouk, Mladen A.; , "Abstracting log lines to log event types for mining software system logs," Mining Software Repositories (MSR), 2010 7th IEEE Working Conference on , vol., no., pp.114-117, 2-3 May 2010, doi: 10.1109/MSR.2010.5463281 (http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5463281&isnumber=5463276)
8. [TKP+10] R. Tchoua, S. Klasky, N. Podhorszki, B. Grimm, A. Khan, E. Santos, C. Silva, P. Mouallem and M. Vouk. "Collaborative Monitoring and Analysis for Simulation Scientists," CTS 2010, ISBN 978-1-4244-6620-7, pp. 235-244, 2010
9. [VM11] Vouk M. and Mouallem P., "On High-Assurance Scientific Workflows," IEEE 13th Conference on High Assurance Systems Engineering 2011, Boca Raton, November 2011, pp. 73-82.
10. [Vou12] M.A Vouk, “A Note on Uncertainty in Real-Time Analytics” in R. Boisvert and A. Dienstfrey, editors, Proceedings of the IFIP WoCo 10 on Uncertainty Quantification in Scientific Computing, IFIP WG 2.5 on Numerical Software, August 2011, and in Springer, *COMPUTER SCIENCE UNCERTAINTY QUANTIFICATION IN SCIENTIFIC COMPUTING*, IFIP Advances in Information and Communication Technology, 2012, Volume 377/2012, 312-318, DOI:10.1007/978-3-642-32677-6_20
11. [VSD10] Mladen A. Vouk, Eric Sills, and Patrick Dreher, “Integration of High-Performance Computing into Cloud Computing Services,” Ch 11 in Handbook of Cloud Computing, editors B. Furht and A. Escalante, Springer, ISBN: 978-1-4419-6523-3, pp. 255-276, 2010
12. [YMV+10] Ustun Yildiz, Pierre Mouallem, Mladen Vouk, Daniel Crawl, Ilkay Altintas, "Fault-Tolerance in Dataflow-Based Scientific Workflow Management," services, pp.336-343, 2010 6th World Congress on Services, 2010, July 5-10, Miami, Florida.

.

NCSU Talks, Presentations, Tutorials

1. M. Vouk, H. Iyer, J. Ligon, P. Mouallem, M. Nagappan, "Some SDM Technologies – from Giga to Exa)," All-Hands-Meeting, Scientific Data Management Center, DOE, Chicago, October 2010, Presentation.
2. M. Vouk, "Cloud Computing for Exascale World," All-Hands-Meeting, Scientific Data Management Center, DOE, Chicago, October 2010, Presentation.
3. M. Vouk, "Petascale Analytics using VCL," in the Petascale Data Analytics on Clouds: Trends, Challenges, and Opportunities (PDAC-10) workshop organized in Cooperation with ACM/IEEE SC10, November 14, 2010. New Orleans, LA, USA.
4. M. Vouk, "Workflow Management Support for Uncertainty Analysis," The 2010 Workshop on Verification, Validation, and Uncertainty Analysis in High-Performance Computing (VVUHPC 2010), 14 November 2010, New Orleans, LA, USA, Collocated with the 23rd Supercomputing Conference (SC10)
5. M. Vouk, "Cloud Computing for Higher Education (VCL - A Cloud That Works), 19 January 2011, Tennessee Board of Regents CIO Meeting, Nasville, TN, Invited Talk
6. M. Vouk, "Exascale Computing in the Cloud," University Computing Center (SRCE), 40th Aniversary Invited Talk, Zagreb, Croatia, 28th March 2011, Keynote, http://www.srce.unizg.hr/o-srcu/povijest-srca/dansrca-2011/dan-srca-2011-program/exascale-computing-in-the-cloud/, http://www.srce.unizg.hr/o-srcu/povijest-srca/dansrca-2011/videogalerija/exascale-computing-in-the-cloud/
7. M. Vouk, "Creating a Cloud Computing Solution for NC State University," at IBM CIO Leadership Exchange in San Francisco on April 5-6, 2011, Invited Talk
8. M. Vouk,, invited presentation in "Exploring the Next Frontier in WAC/WID: A Multi-University, NSF-Sponsored Project to Enable Engineering Faculty to Teach Writing in a Four-Year Sequence of Technical Courses." *(#1197), organized by P. Anderson, M. Carter, G. Gannod, and M. Gustaffson,* Conference on College Composition and Communication (CCCC), April 7, 2011, Atlanta, GA
9. M. Vouk, invited participation on Panel on "Is Integration of Communication and Technical Instruction Across the SE Curriculum a Viable Strategy for Improving the Real-World Communication Abilities of Software Engineering Graduates?" organized by Gerald C. Gannod, Janet E. Burge, Paul V. Anderson, and Andrew Begel at 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE&T 2011), 22-24 May, 2011, pp. 525-529
10. M. Vouk, "Clouds that work," presentation to Monterrey Tech, Mexico, delegation, 17-Jun-2011, Raleigh.
11. M. Vouk, "Clouds that work," presentation to Osaka , Japan, delegation, 30-Jun-2011, Raleigh.
12. M. Vouk, A. Peeler, J. Thomson, A. Kurth, "VCL Level1 Tutorial", Raleigh, 8-July 2011
13. M. Vouk, A. Peeler, J. Thomson, A. Kurth, "VCL Level2 Tutorial", Raleigh, 15-July 2011
14. M. Vouk, A. Peeler, J. Thomson, A. Kurth, "VCL Level3 Tutorial", Raleigh, 25-27-July 2011
15. M. Vouk, "From Grid to Cloud – the VCL Journey," LaGRID 2011, Boca Raton, FL, 4th November 2011 (Keynote)
16. S. Mills, M. **Vouk**, M. Mayer, Cloud Computing Panel at the Smarter Computing Executive Summit, 4-6 October 2011, Pinehurst, NC, (http://ibmreferencehub.com/STG/smarter_computing_pinehurst2011/)
17. M. Vouk, "Are High-Assurance Computing Clouds Possible?" HASE11, Boca Raton, FL, 11th November 2011 (Keynote)
18. M. Vouk, "VCL and vCentennial – Future Today," presentation to the COE IES and Tech Incubator, 22 February 2012.
19. M. Vouk, "Clouds in Education," at World of Innovation Conference, Wroclaw, Poland, 3rd April 2012 (Invited Talk).

20. M. Vouk, A. Peeler, A. Kurth, J. Thompson, E. Sills, “VCL Experience and Some Suggested Best Practices:
NC State,” presentation at Clemson University, 16-April-2012.
21. M. Vouk, “Constructing Next Generation Clouds,” ICACON 2012, RTP, 19-20 April, 2012 (Invited Talk)
22. M. Vouk, A. Peeler, A. Kurth, J. Thompson, E. Sills, “VCL Experience and Some Suggested Best Practices:
NC State,” presentation to a delegation from The Federal University of Technology, Minna, Nigeria (http://www.futminna.edu.ng/), 2-May-2012.
23. M. Vouk,“BigData” Management and Analytics in the Cloud in the Contenxt of the Materials Genome Initiative,” NIST Workshop on Materials Genome Initiative, NIST, Washington, DC, 14-15 May, 2012 (Invited Talk)
24. M. Vouk, “Developing and Implementing Curricula that Fully Integrate Technical and Writing Instruction: The Program Director’s Perspective,” 11th International WRITING ACROSS THE CURRICULUM CONFERENCE (IWAC), June 7-9, 2012, Coastal Georgia Center, Savannah, GA
25. M. Vouk, “BigData,” Panel at ITI 2012, Cavtat, Croatia, 25-28 June 2012 (http://iti.srce.unizg.hr/index.php/ITI/index/pages/view/venue) – panel Chair.
26. M. Vouk and L. Williams, "An Investigation of Scientific Principles Involved in Software Security Engineering." presented at the quarterly Science of Security Lablet meeting, Raleigh, NC, 17-Jul-2012
27. M. Vouk, “vCentennial – Future Now,” invited presentation given at the NCSU Hunt Library Technical Advisory Board meeting is coming up next Monday, August 13, 2012.
28. M. Vouk, “Future Now – Cloud Computing in Education: Practice, Research, Education and Beyond the Clouds,” Workshop on ‘Beyond the Cloud: Cloud Computing Technologies’, EBTIC and Khalifa University, Abu Dhabi, October 2, 2012 (Invited Talk)
29. M. Vouk et al., Panel: “The Fundamentals of Establishing Successful Public and Private Sector Partnerships with Academia.” at the 14th IEEE International High Assurance Systems Engineering Symposium (HASE), October 25-27, 2012, Omaha, Nebraska
30. M. Vouk, “We Now Have Clouds! (How do we walk on them?)”, MCNC Day, November 16th, 2012, Elon College, NC

Ph.D and M.S. Thesis (involving SDM center technologies)

1. Jeff Ligon (NCSU Ph.D., 2010, "The Use of Locally Invertible Convolutional Encoders for Encryption")
2. Harini Iyer, (NCSU M.S., 2010, "Automation of Scientific Workflow Construction using Templates and Patterns")
3. Meiyappan Nagappan, (NCSU Ph.D., 2011, "A Framework for Analyzing Software System Log Files")
4. Pierre Mouallem, (NCSU Ph.D., 2011, "A Fault Tolerance framework for Kepler-based Distributed Scientific Workflows")
5. Nikhil Talpallikar, (NCSU M.S., 2012, "High-Performance Cloud Computing: VCL Case Study") - Chair [Record] [Thesis (PDF)]
6. Georgy Mathew Kallumkal, (NCSU M.S., 2013, “A Micro-cloud Model for Adaptable High Performance Computing)