The Scientific Data Management Center

Arie Shoshani (PI)

Lawrence Berkeley National Laboratory

Co-Principal Investigators

DOE Laboratories

ANL: Rob Ross
LBNL: Doron Rotem
LLNL: Chandrika Kamath
ORNL: Nagiza Samatova
PNNL: Terence Critchlow
Jarek Nieplocha

Universities

NCSU: Mladen Vouk
NWU: Alok Choudhary
UCD: Bertram Ludaescher
SDSC: Ilkay Altintas
UUtah: Claudio Silva

Centers/Institutes meeting, October 24-25, 2008

Problems and Mandate

- **Why is Managing Scientific Data Important for Scientific Investigations?**
  - **Sheer volume and increasing complexity of data being collected are already interfering with the scientific investigation process**
  - **Managing the data by scientists greatly wastes scientists effective time in performing their applications work**
  - **Data collection, storage, transfer, and archival often conflict with effectively using computational resources**
  - **Effectively managing, and analyzing this data and associated metadata requires a comprehensive, end-to-end approach that encompasses all of the stages from the initial data acquisition to the final analysis of the data**
- **Enable scientists to most effectively discover new knowledge by removing data management bottlenecks, and enabling effective data analysis**
  - **Improve productivity of data management infrastructure**
  - **Taking away the burden from scientists**
  - **Engaging Scientists, education**

Focus of SDM center

- high performance
  - fast, scalable
  - Parallel I/O, parallel file systems
  - Indexing, data movement
- Usability and effectiveness
  - Easy-to-use tools and interfaces
  - Use of workflow, dashboards
  - end-to-end use (data and metadata)
- Enabling data understanding
  - Parallelize analysis tools
  - Streamline use of analysis tools
  - Real-time data search tools
- Sustainability
  - robustness
  - Productize software
  - work with vendors, computing centers

- Establish dialog with scientists
  - Outreach,
  - partner with scientists,
  - education (students, scientists)

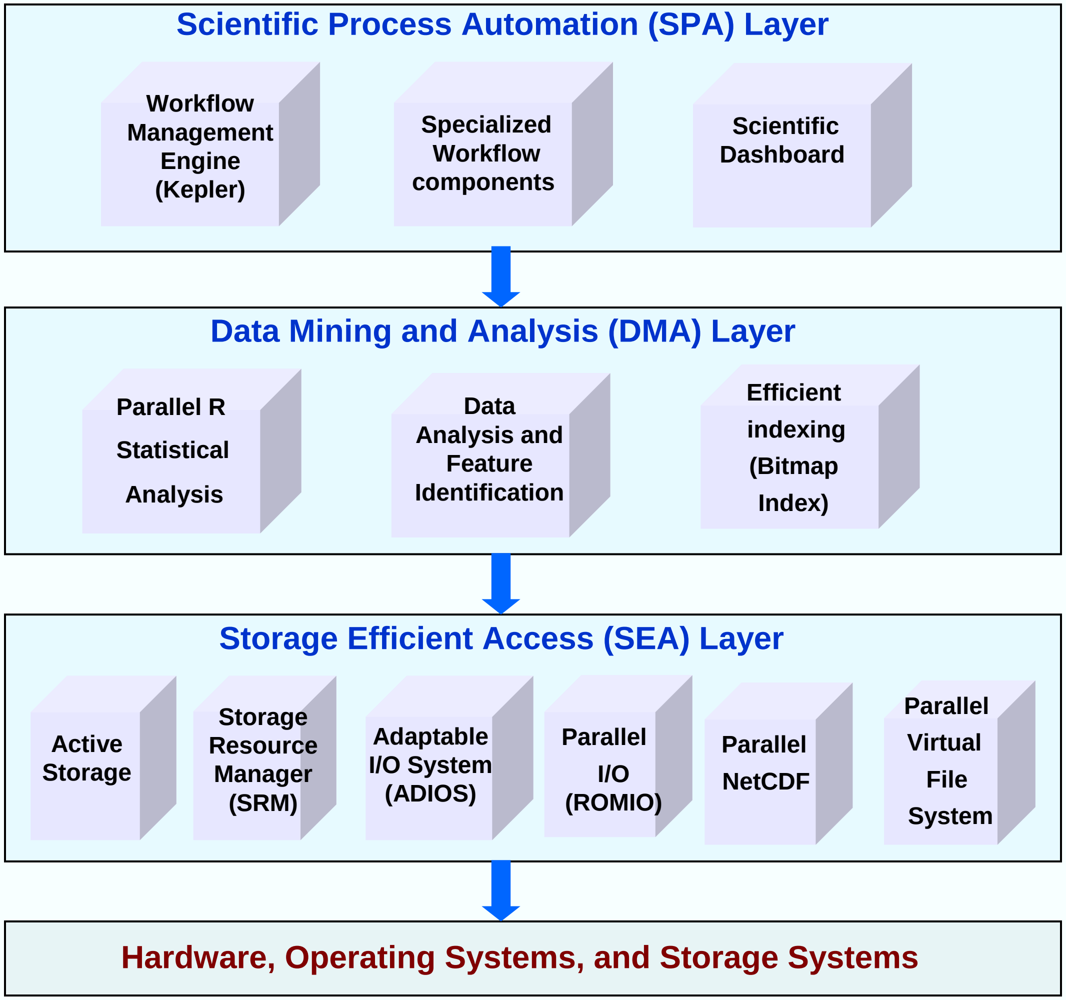
Organization of the center:
based on three-layer organization of technologies

Integrated approach:

- **To provide a scientific workflow and dashboard capability**
- **To support data mining and analysis tools**
- **To accelerate storage and access to data**

Benefits scientists by

- **Hiding underlying parallel technology**
- **End-to-end support of applications**
- **Permitting assembly of modules using workflow description tool**
- **Tracking data management tasks through web-based dashboards**



Results

✓ High Performance Technologies

Usability and effectiveness

Enabling Data Understanding

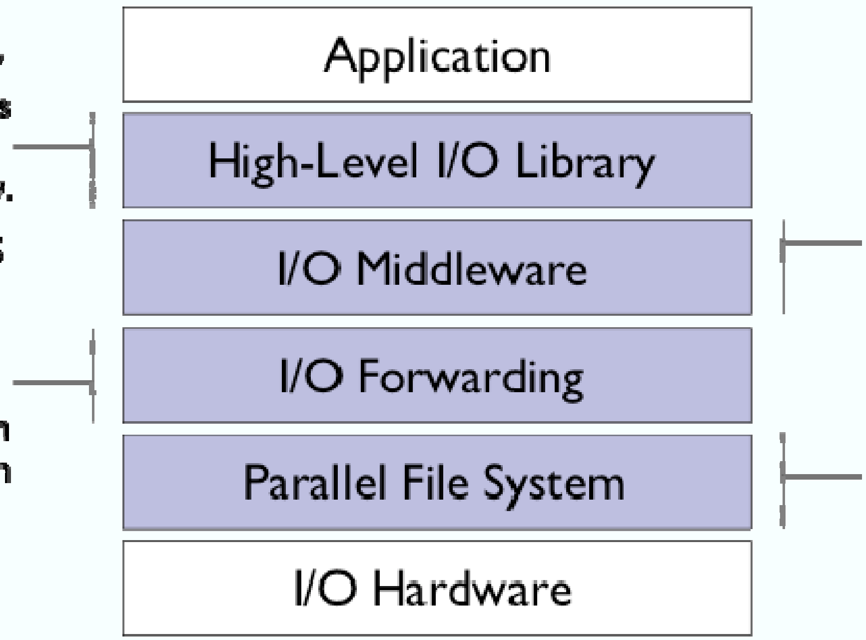
The I/O Software Stack

High-Level I/O Library
maps application abstractions onto storage abstractions and provides data portability.

HDF5, Parallel netCDF, ADIOS

I/O Forwarding
bridges between app. tasks and storage system and provides aggregation for uncoordinated I/O.

IBM ciod



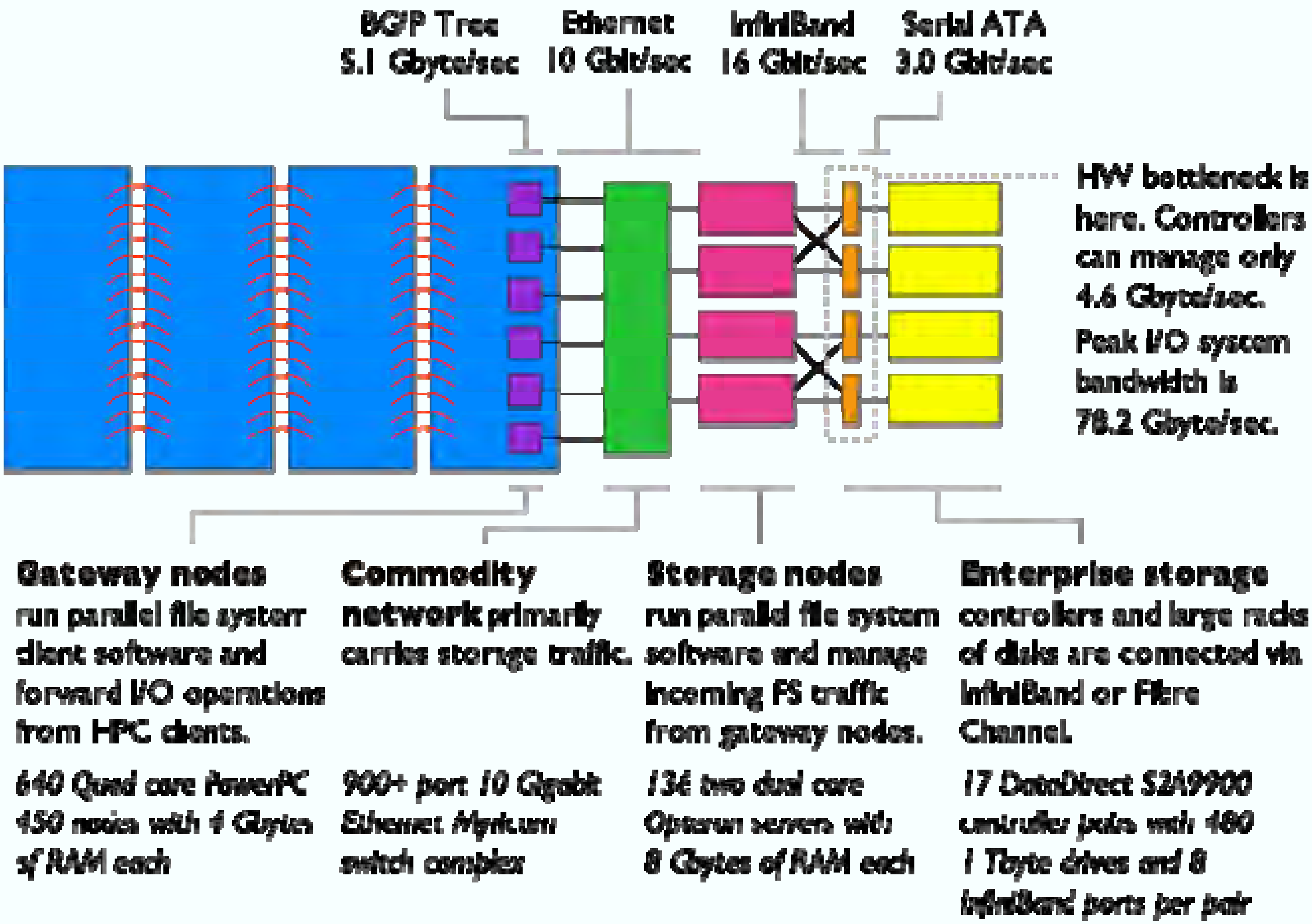
I/O Middleware
organizes accesses from many processes, especially those using collective I/O.

MPI-IO

Parallel File System
maintains logical space and provides efficient access to data.

PVFS, PanFS, GPFS, Lustre

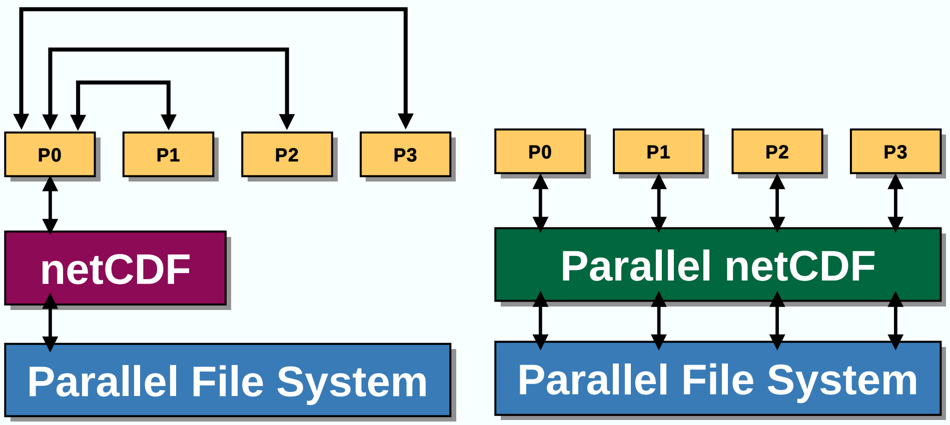
PVFS on IBM Blue Gene/P



Architectural diagram of the 557 TFlop IBM Blue Gene/P system at the Argonne Leadership Computing Facility.

Speeding data transfer with PnetCDF

Inter-process communication

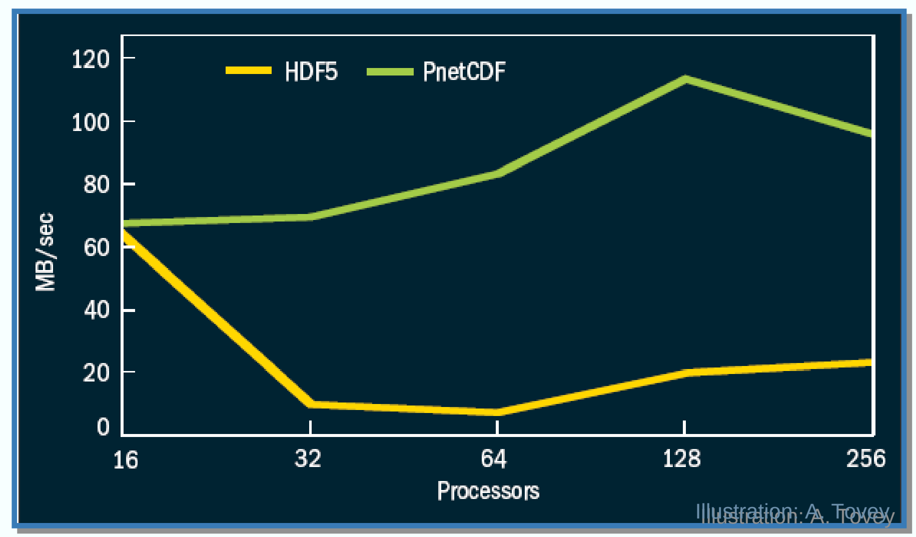


Enables high performance parallel I/O to netCDF data sets

Achieves up to 10-fold performance improvement over HDF5

Early performance testing showed PnetCDF outperformed HDF5 for some critical access patterns.

The HDF5 team has responded by improving their code for these patterns, and now these teams actively collaborate to better understand application needs and system characteristics, leading to I/O performance gains in both libraries.



Contacts: Rob Ross, ANL, Alok Choudhari, NWU

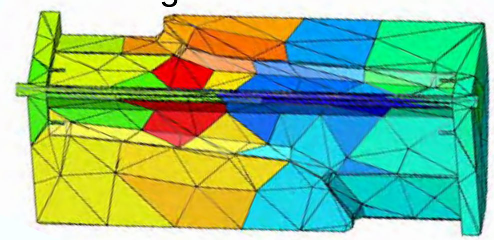
Improving IO in accelerator design simulation on Jaguar/Cray XT*

- **Application: SLAC accelerator design**
  - **Omega3P: simulation program that uses higher-order tetrahedral elements**
  - **Had bad reading patterns that do not scale**
  - **Use netCDF files**

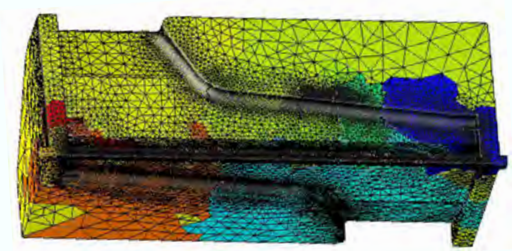
Before (in seconds)

N-CPUs	Writing Time	Solver Time
128	30.27	634.74
256	59.26	324.16
512	146.24	163.30
1024	340.15	94.86
2048	499.21	45.86
4096	965.64	26.08

Scaling from Regular meshes



To adaptive meshes



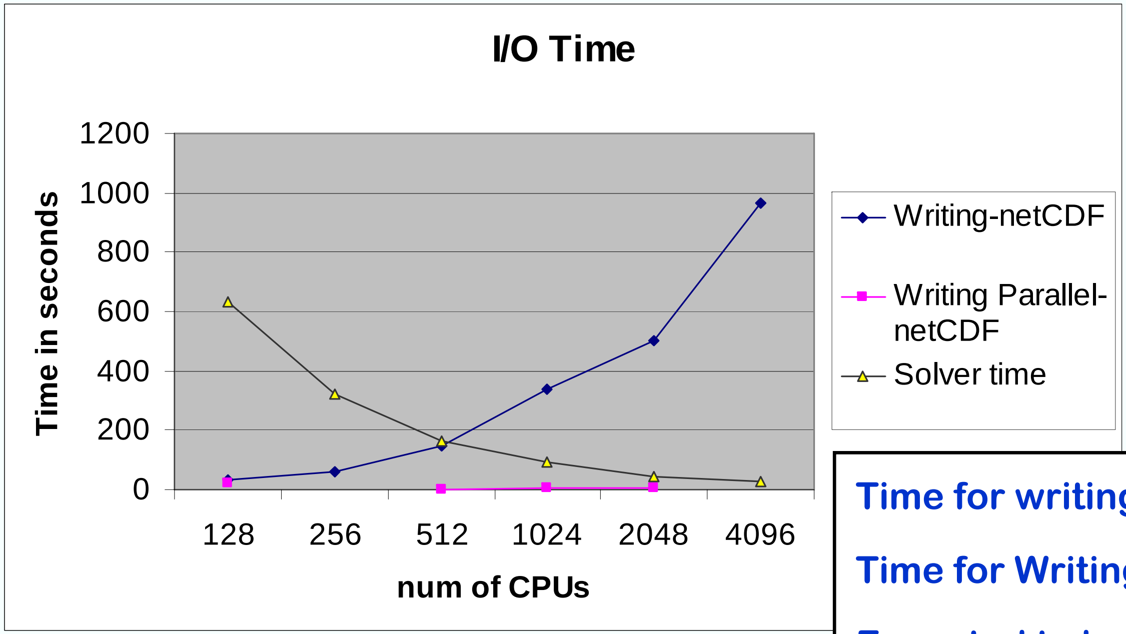
- **Time for Writing File >> Time for Solver !!!**

(*) Lie-Quan (Rich) Lee (SLAC) and Stephen Hodson (ORNL)

Using Parallel-netCDF instead of Netcdf and using MPI_Info

After **(in seconds)**

NCPUs	Writing Time	Solver Time
512	1.50	163.30
1024	3.27	94.86
2048	7.90	45.86



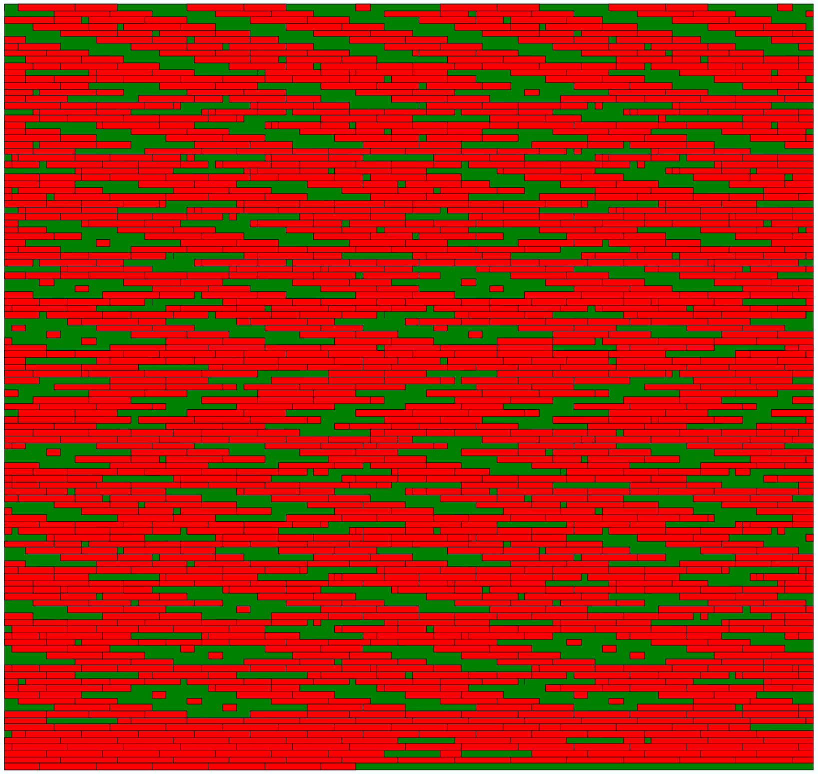
Time for writing data reduced 100 times

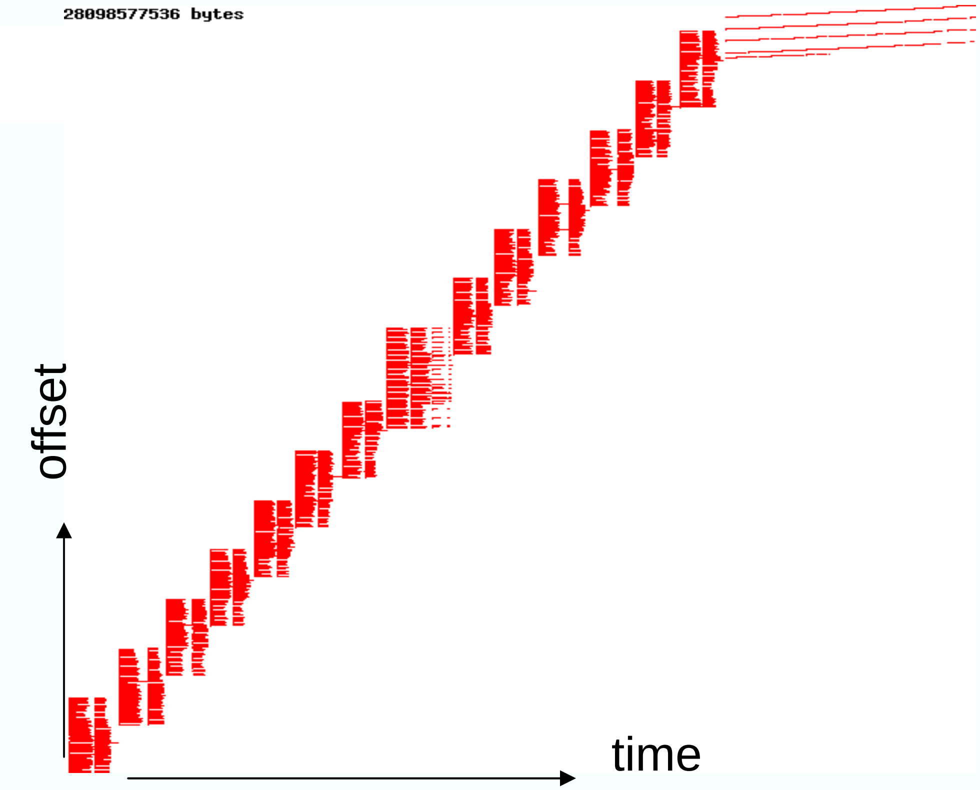
Time for Writing File << Time for Solver

Expected to behave better for larger problem sizes.

Contact: Alok Choudhari, NWU

Parallel netCDF (no hints)

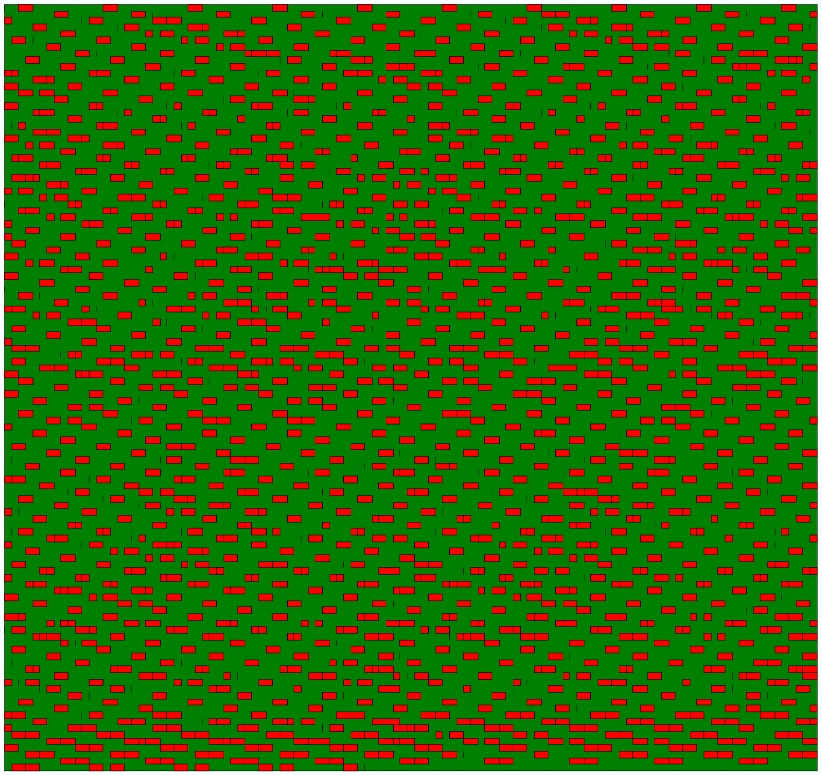


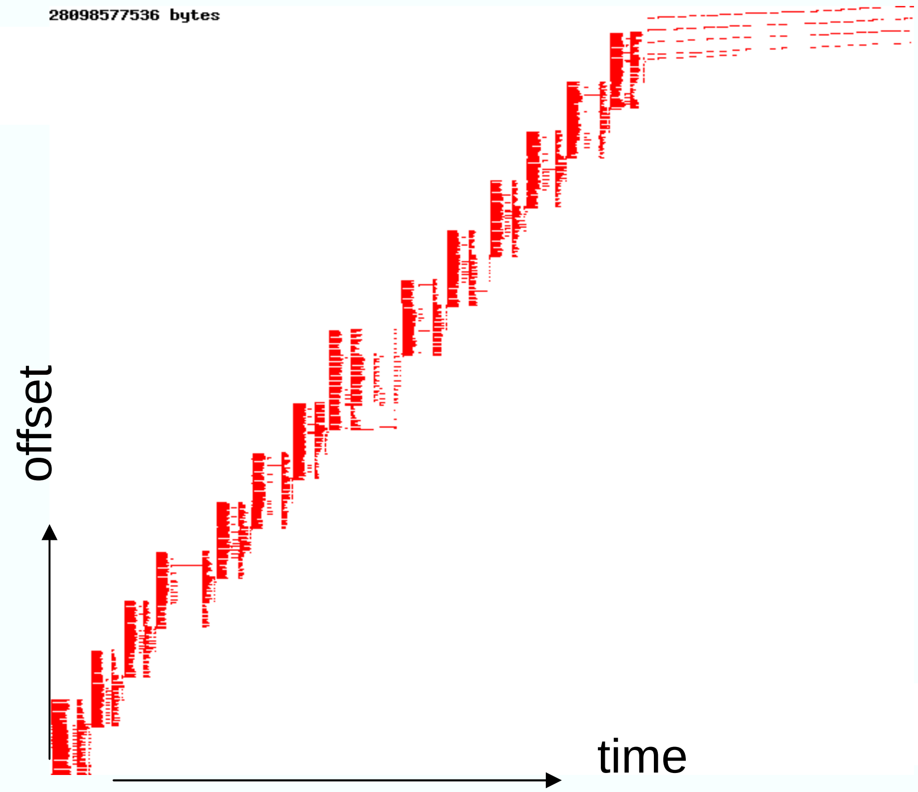


- Block depiction of 28 GB file
- Record variable scattered
- Reading in way too much data!

- Y axis larger here
- Default "cb_buffer_size" hint not good for interleaved netCDF record variables

Parallel netCDF (hints)

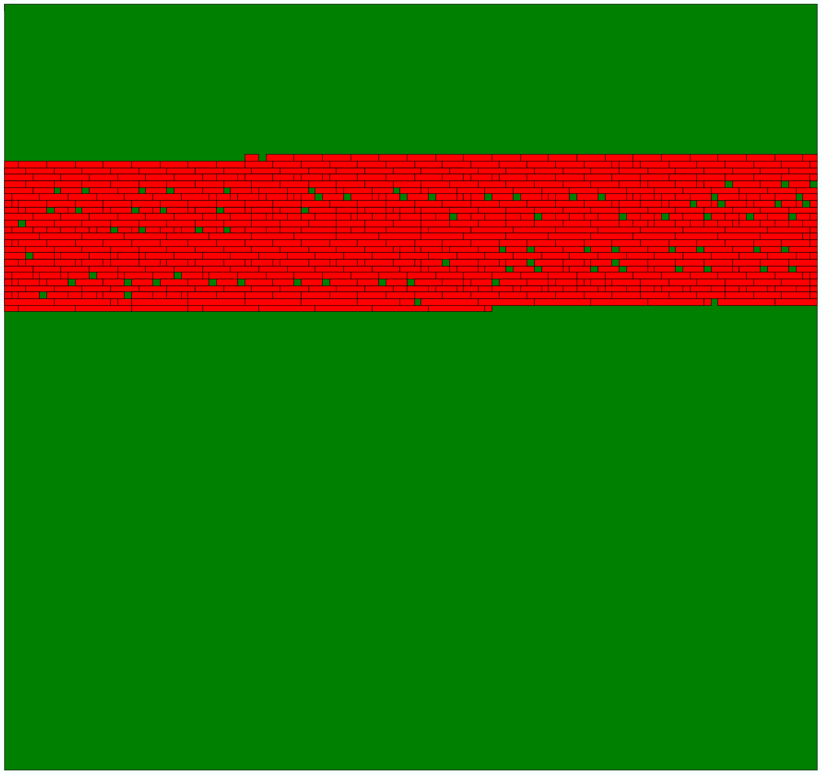


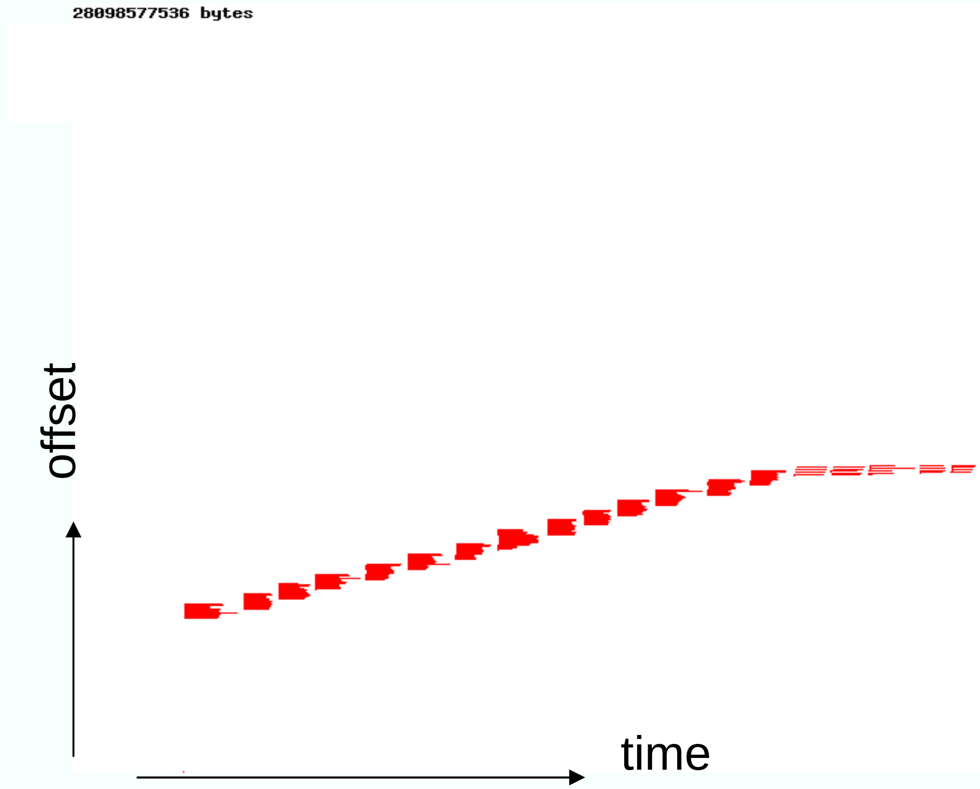


- With tuning, much less reading
- Better efficiency, but still short of MPI-IO

- Still some overlap
- "cb_buffer_size" now size of one netCDF record
- Better efficiency, at slight perf cost

Parallel netCDF (current SVN)





- Development effort to relax netCDF file format limits
- No need for record variables
- Data nice and compact like MPI-IO and HDF5

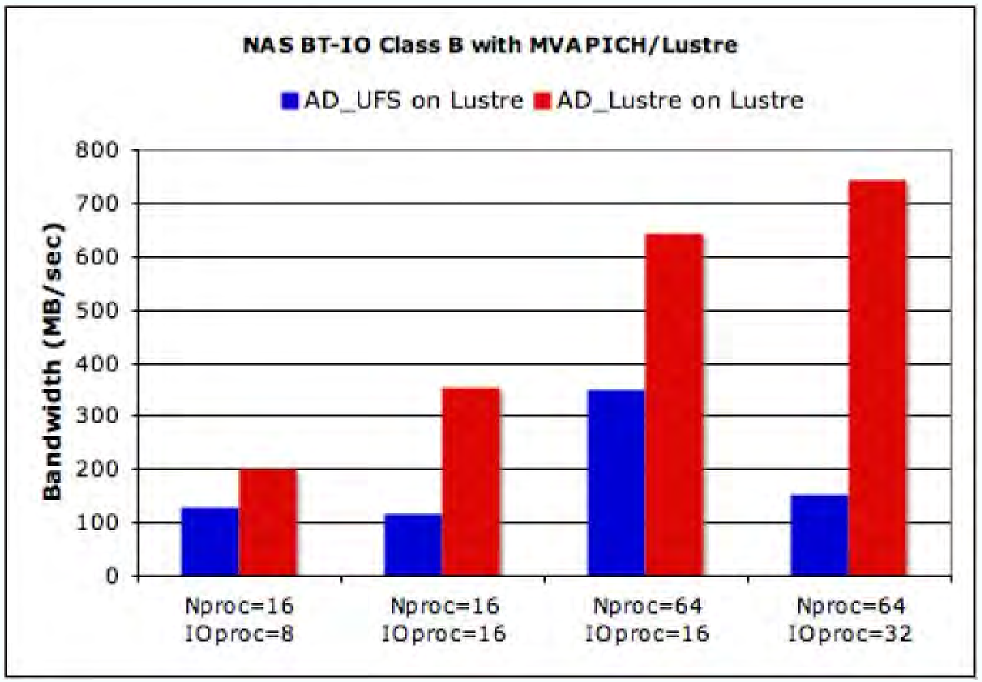
- Rank 0 reads header, broadcasts to others
  - Much more scalable approach
- Approaching MPI-IO efficiency
- Maintains netCDF benefits
  - Portable, self-describing, etc.

Contacts: Rob Ross, ANL, Alok Choudhari, NWU

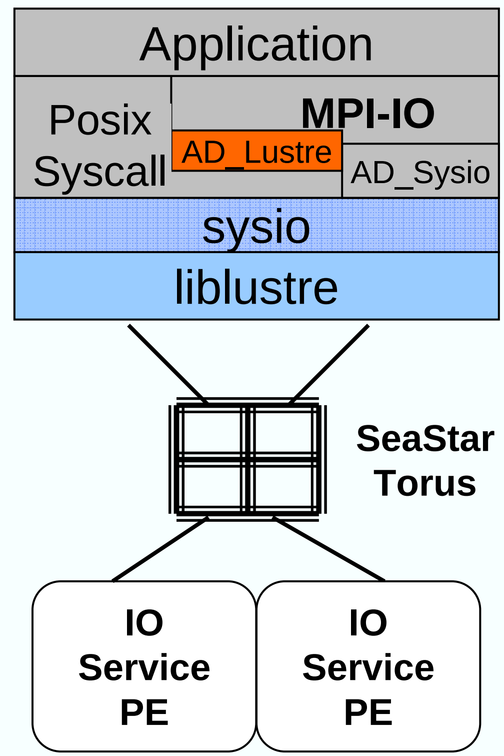
MPI-IO Driver for Lustre

- Available for Beowulf clusters and Cray XT
- Overcome the restriction of a proprietary MPI-IO stack on Cray XT
- Enabled arbitrary striping specification over Cray XT
- Lustre stripe-aligned file domain partitioning
- Released via MVAPICH-1.0 and MPICH2-1.0.7

Performance on an 80-node beowulf cluster



Software Diagram on Cray XT

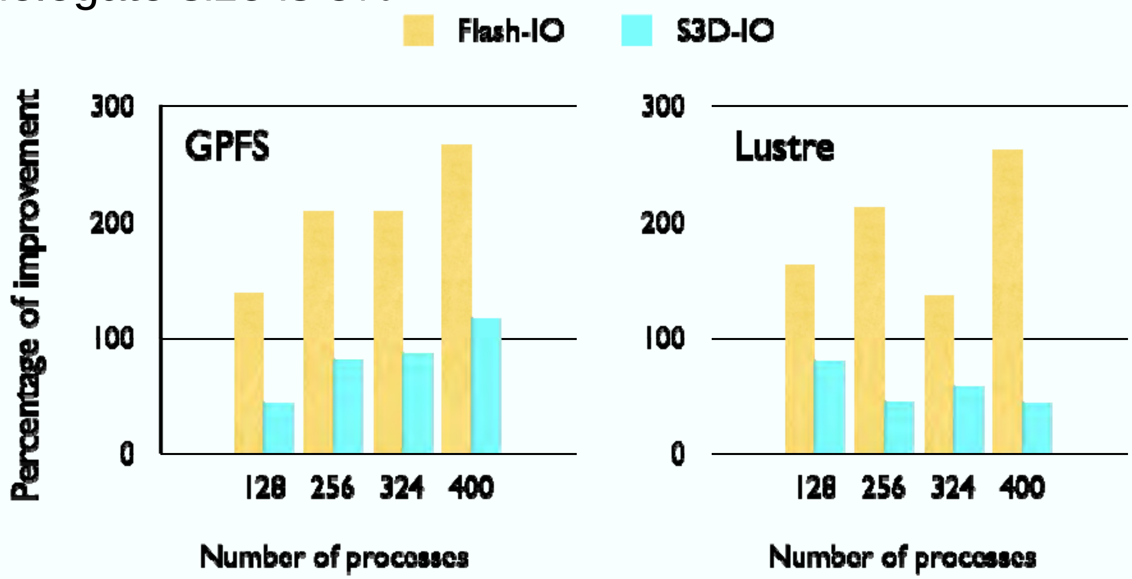


Contact: W. Yu, PNNL

Caching with I/O delegate

- **Allocate a dedicate group of processes to perform I/O**
  - **Uses a small percentage (< 10 %) of additional resource**
  - **Entire memory space at delegates can be used for caching**
  - **Collective I/O off-load**

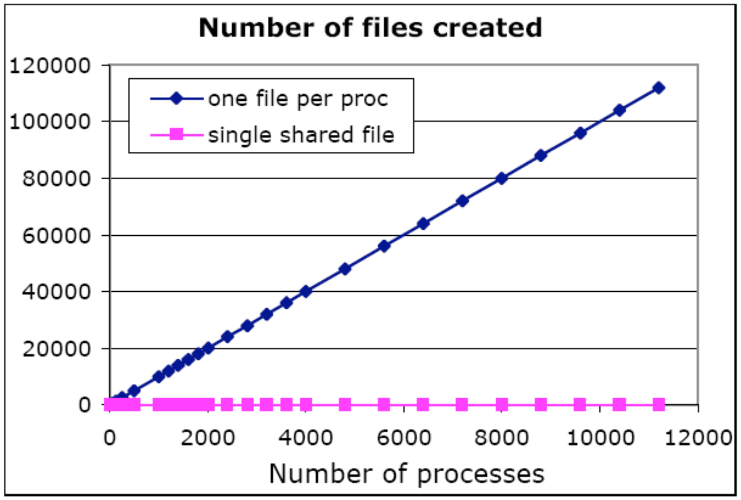
I/O delegate size is 3%

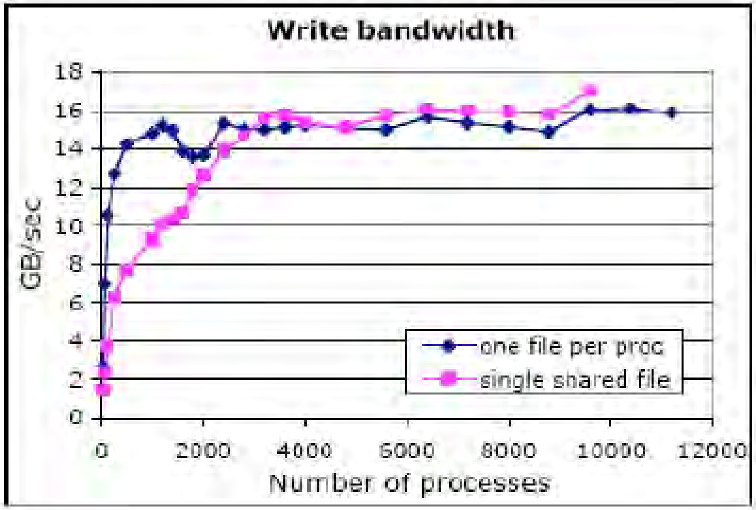


A. Nisar, W. Liao, and A. Choudhary. Scaling Parallel I/O Performance through I/O Delegate and Caching System. SC 2008.

S3D-IO on Cray XT Performance/Productivity

- **Problem:**
  - Number of files created are often generated per processor
  - Causes problems with archiving and future access
- **Approach**
  - Parallel I/O (MPI-IO) optimization
  - One file per variable during I/O
  - Requires multi-processor coordination during I/O
- **Achievement**
  - Shown to scale to 10s of thousands of processors on production systems
  - better performance but eliminating the need to create 100K+ files



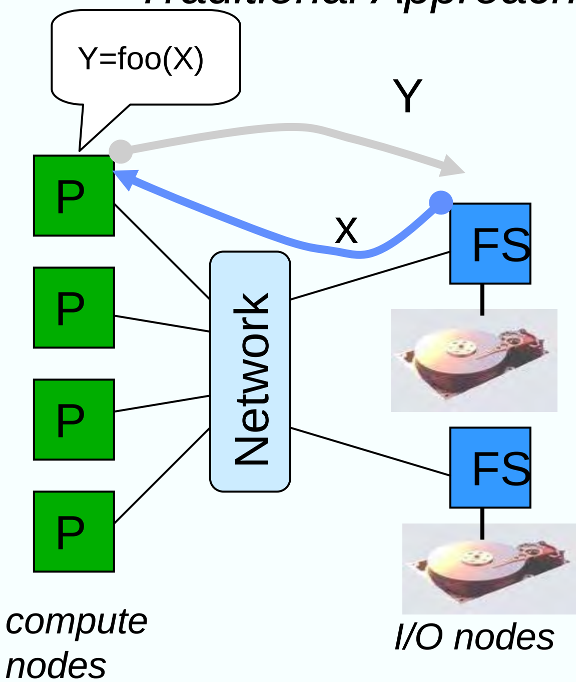


Contacts: Rob Ross, ANL, Alok Choudhari, NWU

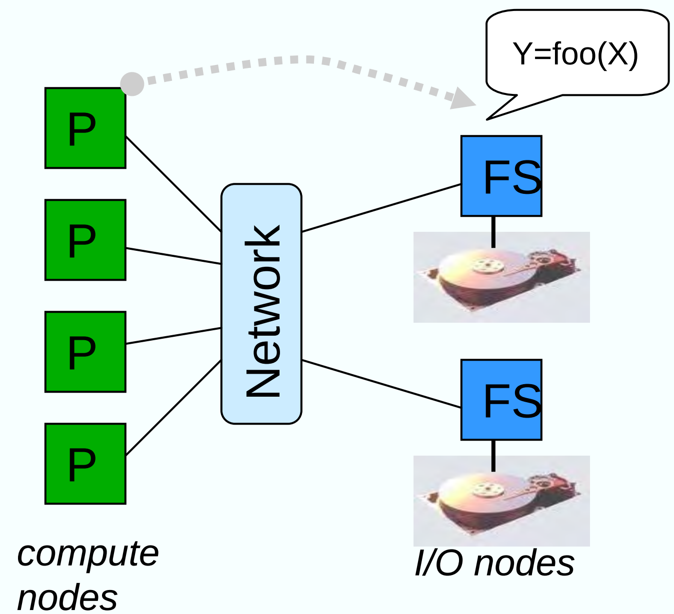
Active Storage in Parallel File Systems

- **Active Storage exploits the old concept of moving computing to the data source**
- **Avoids data movement across the network in parallel machine by allowing applications use compute resources on the I/O nodes of the cluster for data processing**
- **Active Storage efficiently deals with both striped and netCDF files, eliminating > 95% of the network traffic in climate applications**
- **Developed for Luster and PVFS file systems**

Traditional Approach



Active Storage



Contact: J. Nieplocha et. al, PNNL

Active Storage Application: High Throughput Proteomics

9.4 Tesla High Throughput Mass Spectrometer

1 Experiment per hour
5000 spectra per experiment
4 MByte per spectrum

Per instrument:
20 Gbytes per hour
480 Gbytes per day

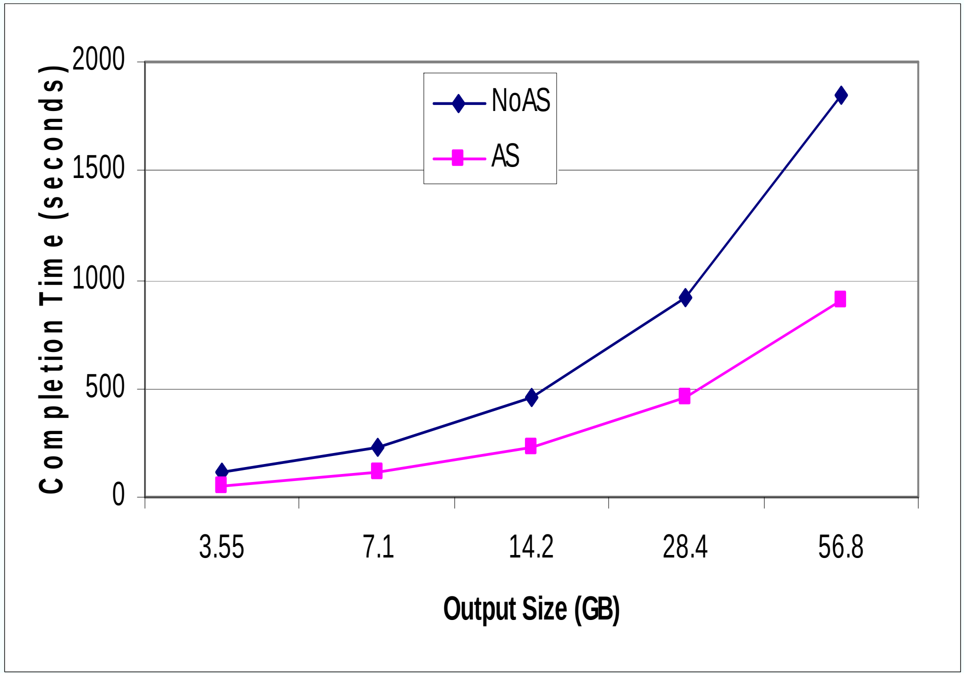
Next generation technology will increase data rates x200

Application Problem

Given 2 float input number for target mass and tolerance, find all the possible protein sequences that would fit into specified range

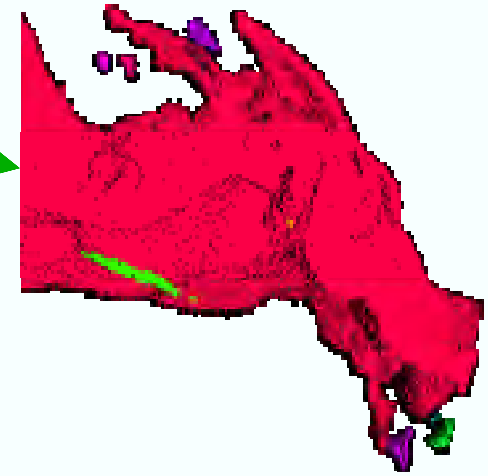
Active Storage Solution

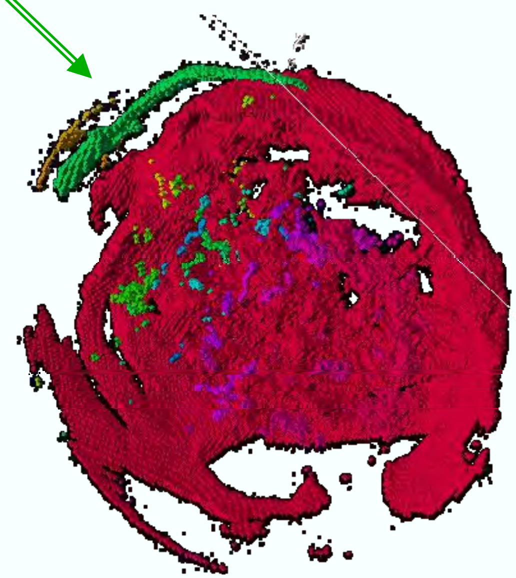
Each OST receives its part of the float pair sent by the client stores the resulting processing output in its Lustre OBD (object-based disk)



Searching Problems in Data Intensive Sciences

- Find the HEP collision events with the most distinct signature of Quark Gluon Plasma
- Find the ignition kernels in a combustion simulation
- Track a layer of exploding supernova





These are not typical database searches:

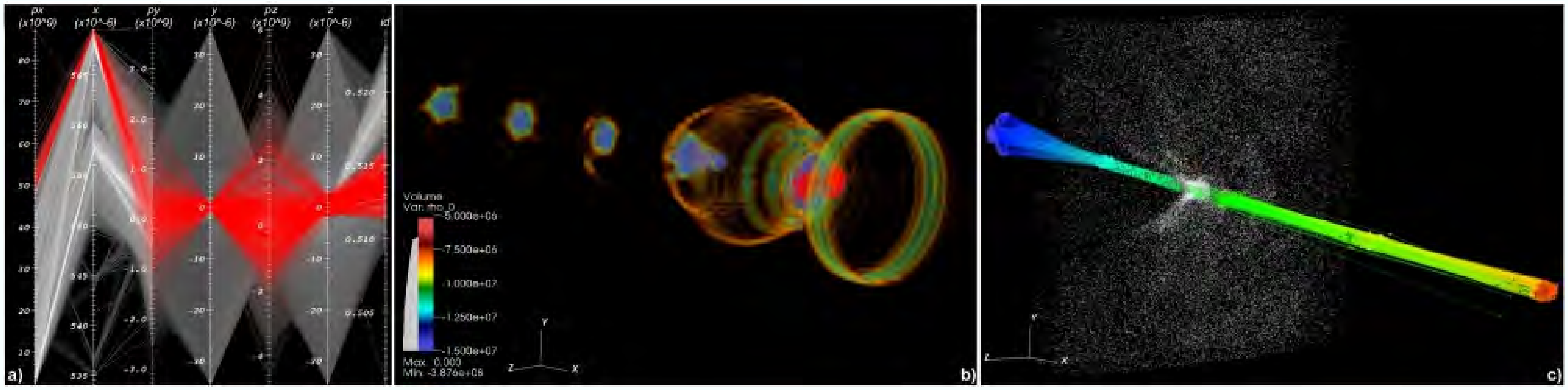
- Large high-dimensional data sets
  (1000 time steps X 1000 X 1000 X 1000 cells X 100 variables)
- No modification of individual records during queries, i.e., append-only data
- Complex questions: 500 < Temp < 1000 && CH3 > 10^{-4} && …
- Large answers (hit thousands or millions of records)
- Seek collective features such as regions of interest, histograms, etc.

- Other application domains:
  - real-time analysis of network intrusion attacks
  - fast tracking of combustion flame fronts over time
  - accelerating molecular docking in biology applications
  - query-driven visualization

FastBit: accelerating analysis of very large datasets

- Most data analysis algorithm cannot handle a whole dataset
  - Therefore, most data analysis tasks are performed on a subset of the data
  - Need: very fast indexing for real-time analysis
- FastBit is an extremely efficient compressed bitmap indexing technology
  - Can search billion data values in seconds
  - FastBit improves the search speed by 10x – 100x of times than best known indexing methods
  - Uses a patented compression techniques
- Size: FastBit indexes are modest in size compared to well-known database indexes
  - On average about 1/3 of data volume compared to 3-4 times in common indexes (e.g. B-trees)

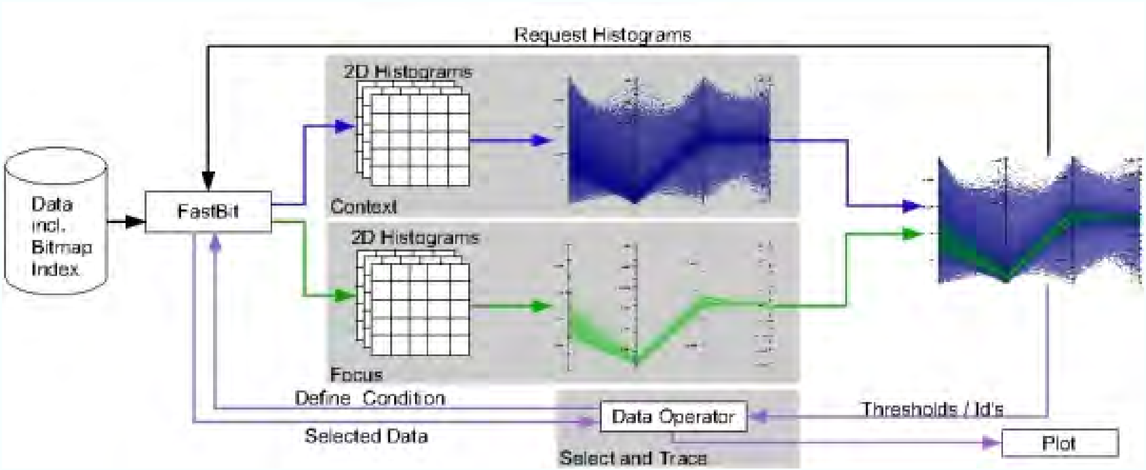
Query-Driven Visualization

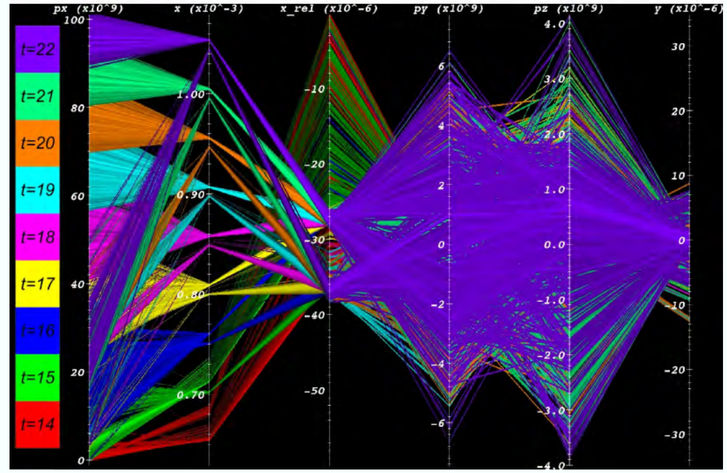


- **Collaboration between SDM and VACET**
  - **Use FastBit indexes to efficiently select the most interesting data for visualization**
- **Above example: laser wakefield accelerator simulation**
  - **VORPAL produces 2D and 3D simulations of particles in laser wakefield**
  - **Finding and tracking particles with large momentum is key to design the accelerator**
  - **Brute-force algorithm is quadratic (taking 5 minutes on 0.5 mil particles), FastBit time is linear in the number of results (takes 0.3 s, 1000 X speedup)**

Contact: John Wu, Wes Bethel, LBNL

Bin-Based Parallel Coordinate Display

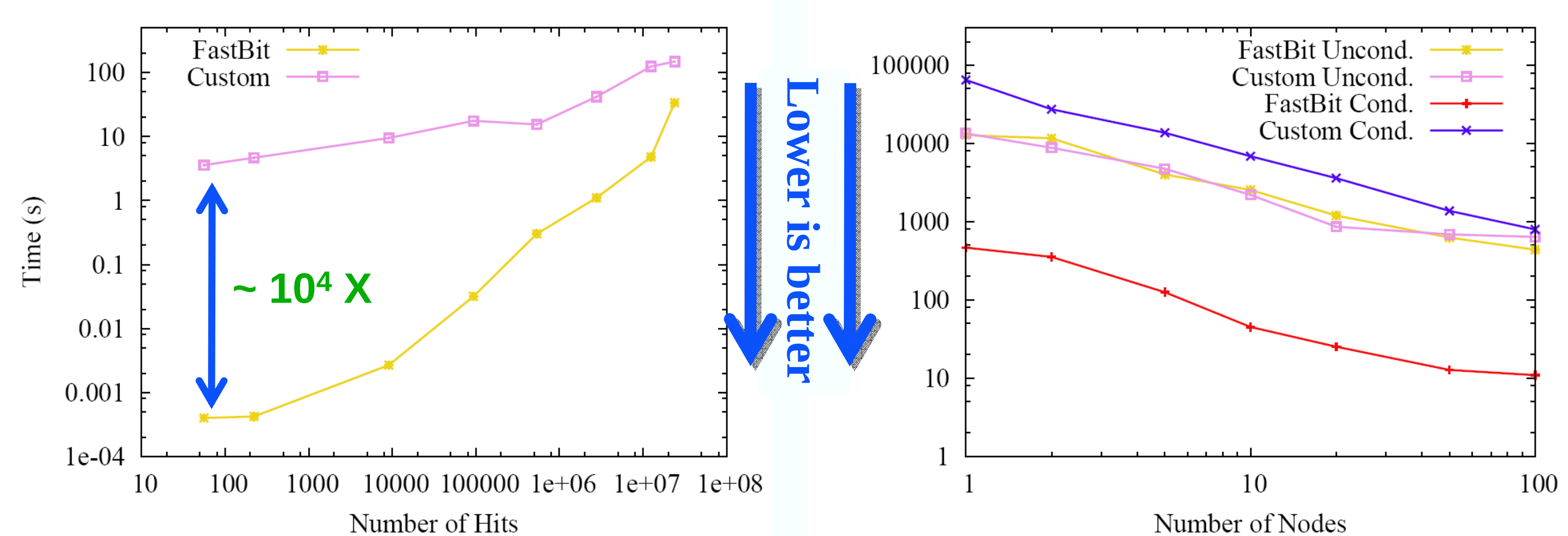




- Integrate FastBit with H5Part, a HDF5 package for particle physics data
- Use FastBit to compute histograms efficiently
- Bin-based parallel coordinate display reduces the number of lines displayed on screen, reduces visual clutter, reduces response time
- FastBit further speeds up the response time further

Contact: John Wu, Wes Bethel, LBNL

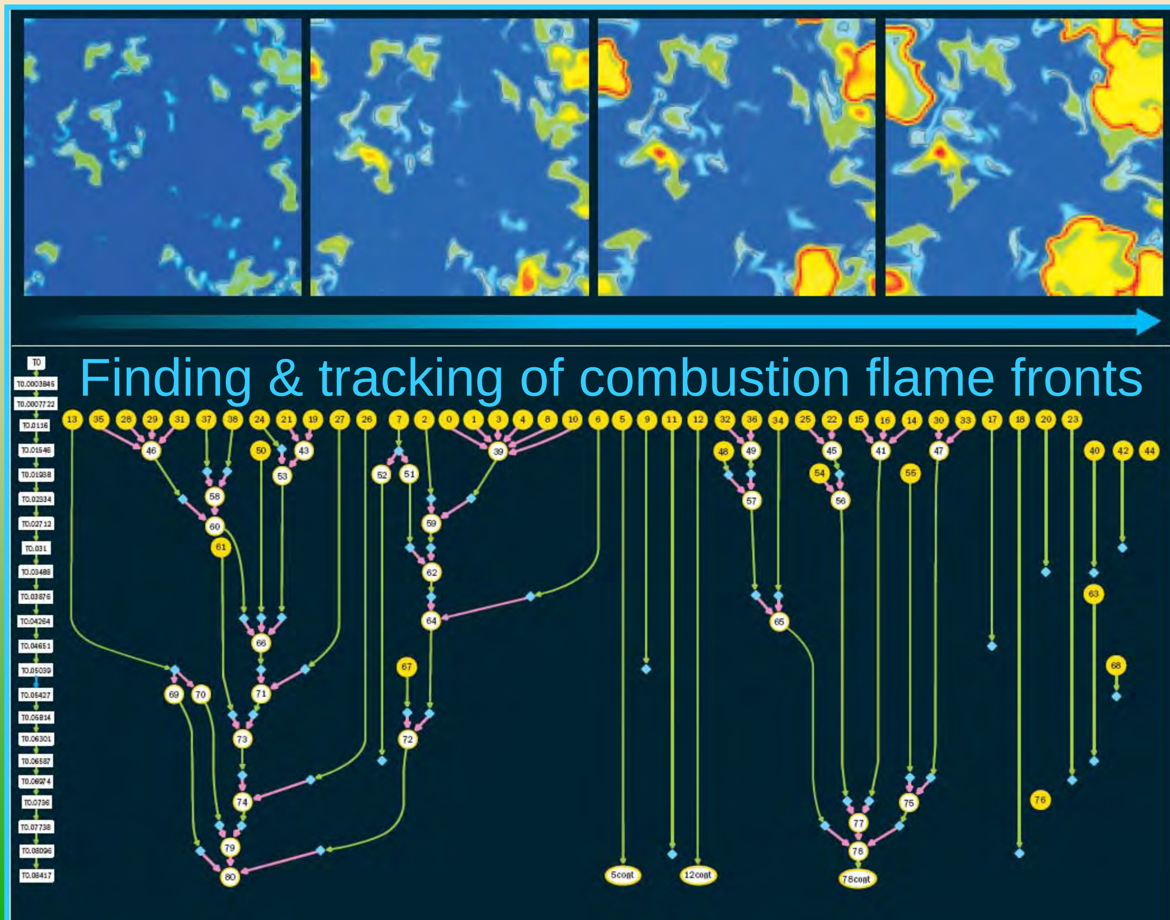
FastBit Speeds up Historgraming



- **Time needed to compute desired histograms**
- **Custom code that directly uses the raw data directly**
- **FastBit can be 1000 X faster than the custom code (left)**
- **FastBit maintains the performance advantage on a parallel system**

Flame Front Tracking in Combustion Simulation using FastBit

Searching for regions that satisfy particular criteria is a challenge. FastBit efficiently finds regions of interest.



Cell identification

Identify all cells that satisfy user specified conditions: "600 < Temperature < 700 AND HO_2 concentr. > 10^{-7}"

Region growing

Connect neighboring cells into regions

Region tracking

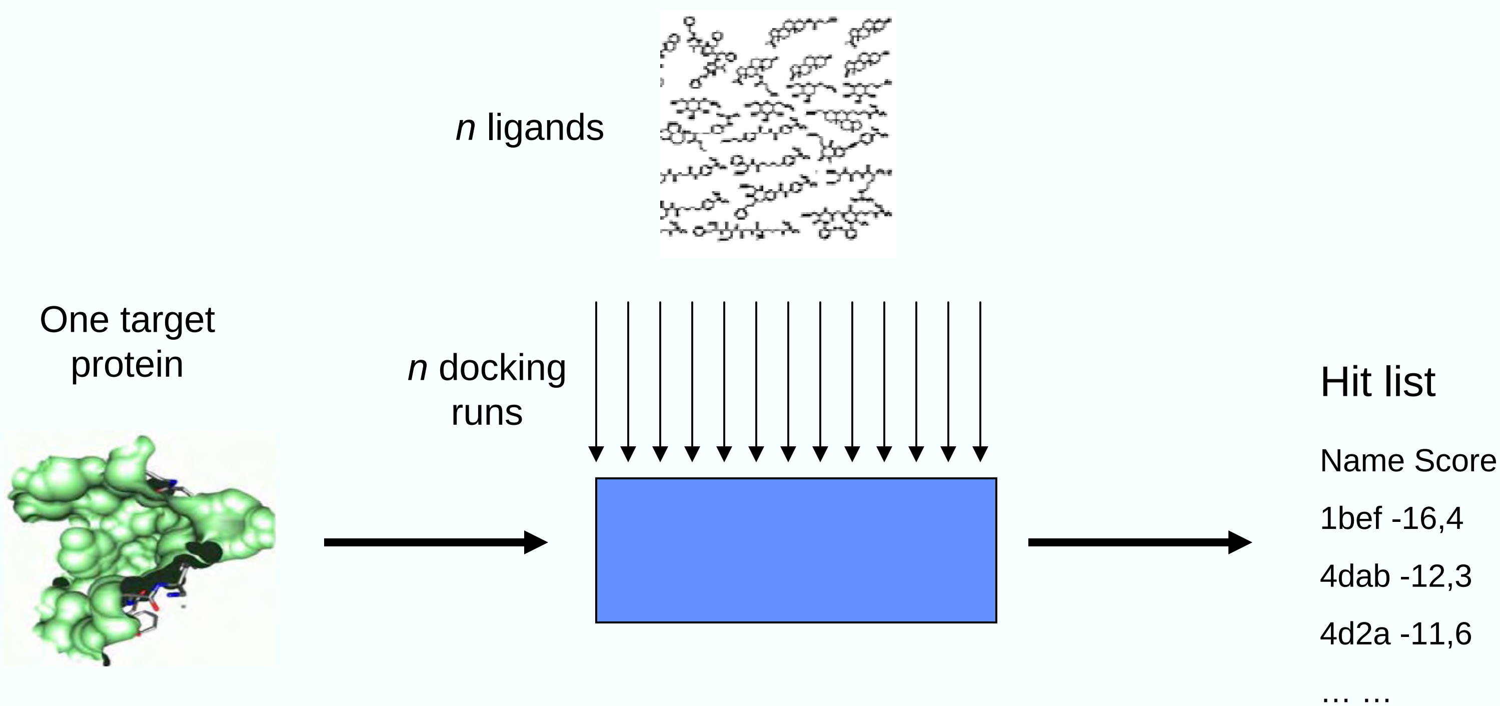
Track the evolution of the features through time

SNL: Drs. J. Chen, W. Doyle NCSU: Dr. T. Echekki

Contact: John Wu, LBNL (kwu@lbl.gov)

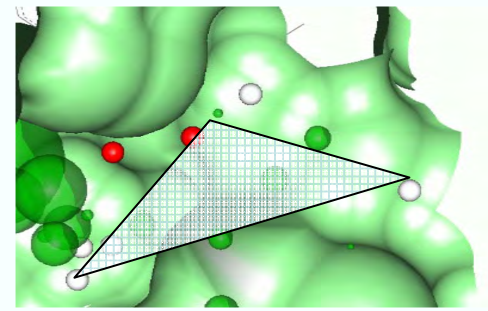
Use of FastBit for Molecular Docking

- **FastBit has been released as open-source**
- **Example of use by others**
  - **Jochen Schlosser [schlosser@zbh.uni-hamburg.de]**
    Center for Bioinformatics, University of Hamburg
- **Problem: Structure-based virtual screening, standard setup**



Use of FastBit for Molecular Docking

- **Specification of the descriptor as triangle geometry**
  - **Types of interaction centers**
  - **Triangle side lengths**
  - **Interaction directions**
  - **80 bulk dimensions**



- **Receptors**
  - **Receptor descriptors are generated similarly**
  - **Using complementary information where necessary**
- **Idea: Usage of pharmacophore constraints on receptor triangles**
  - **Reduces number of queries**
  - **Improved query selectivity because the pharmacophore tends to be inside the protein cavity**

Results

- TrixX-BMI is an efficient tool for virtual screening with average runtime in sub-second range
- **With pharmacophore constraints using FastBit, speedup 140 – 250**

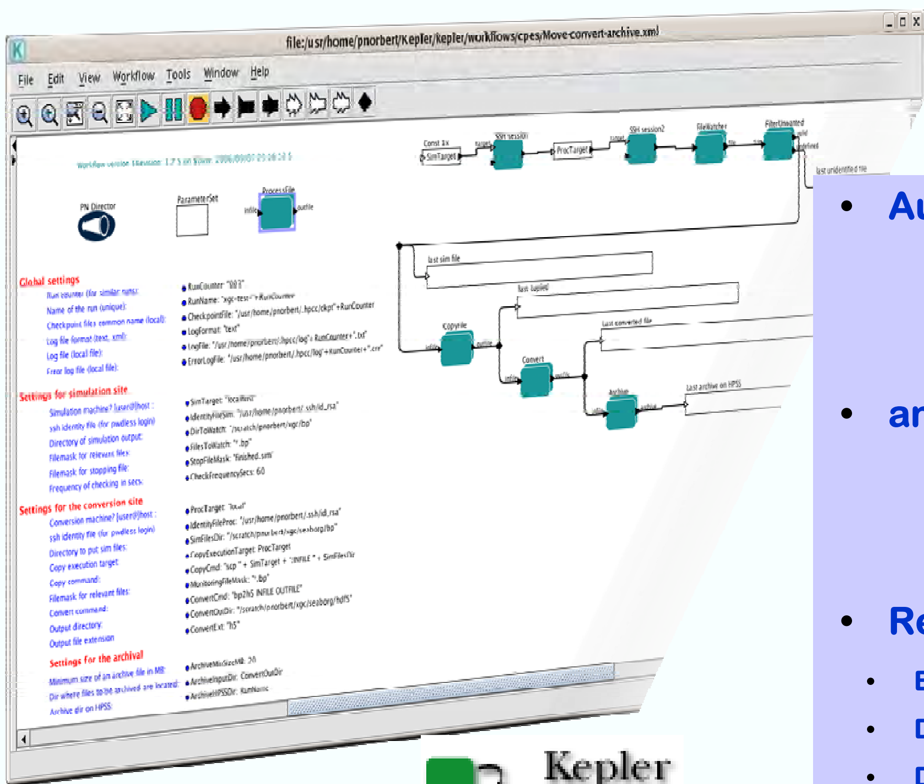
Results

High Performance Technologies

✓ Usability and effectiveness

Enabling Data Understanding

Workflow automation requirements in Fusion Center for Plasma Edge Simulation (CPES) project

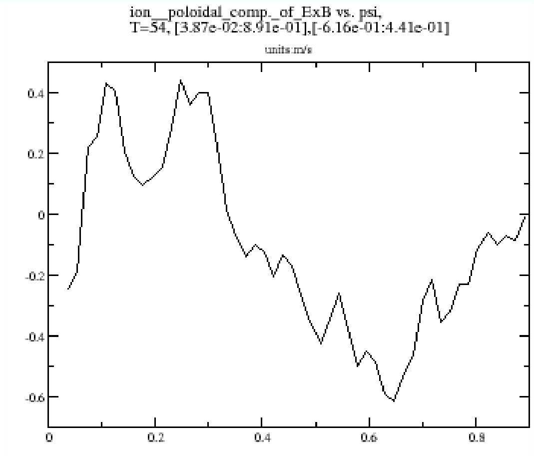


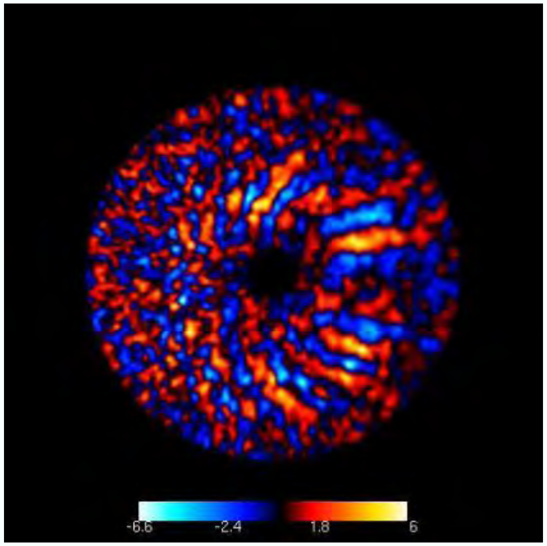
- **Automate the monitoring pipeline**
  - **transfer of simulation output to remote machine**
  - **execution of conversion routines,**
  - **image creation, data archiving**
- **and the code coupling pipeline**
  - **Run simulation on a large supercomputer**
  - **check linear stability on another machine**
  - **Re-run simulation if needed**
- **Requirements for Petascale computing**
  - **Easy to use**
  - **Dashboard front-end**
  - **Dynamic monitoring**
  - **Parallel processing**
  - **Robustness**
  - **Configurability**

Contact: Scott Klasky, et. al, ORNL

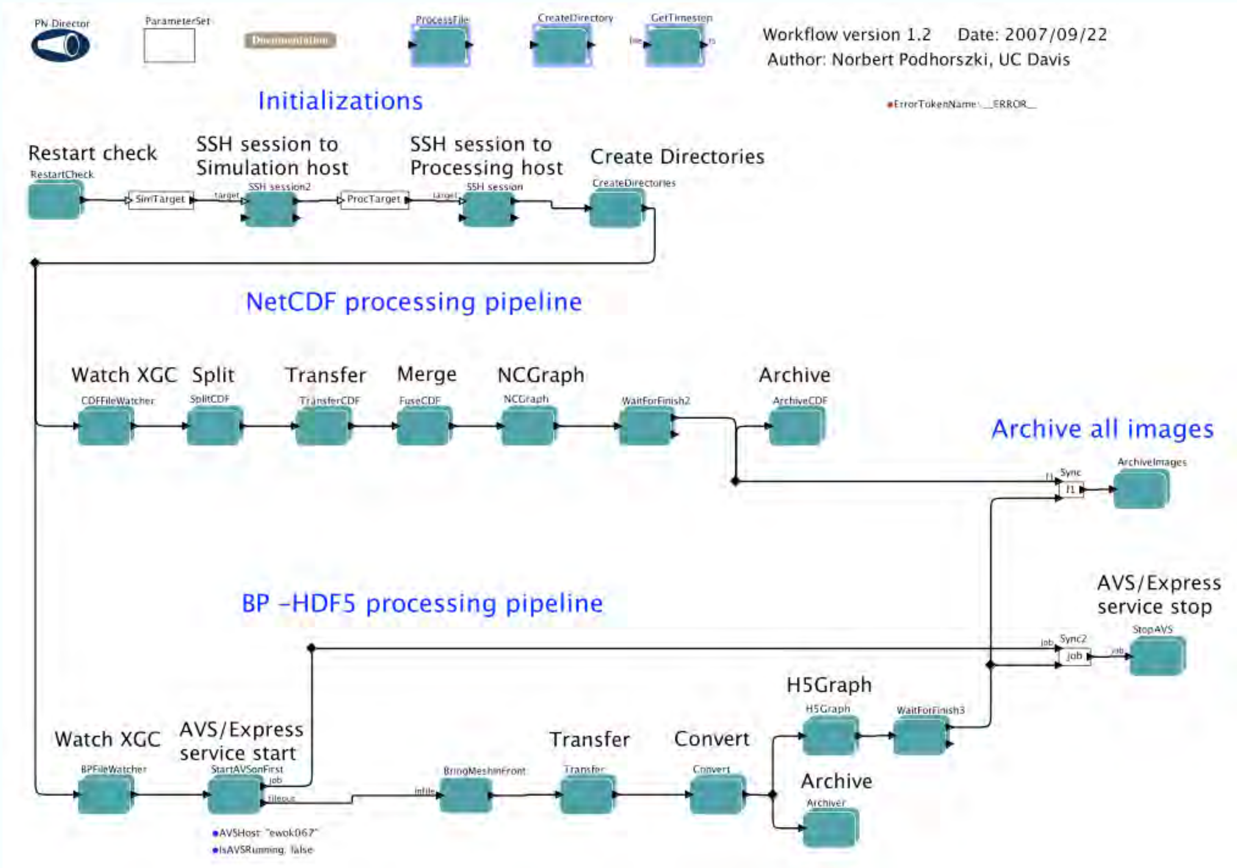
Real-Time Monitoring a simulation Plus archiving

- **NetCDF files**
  - **Transfer files to e2e system on-the-fly**
  - **Generate plots using grace library**
  - **Archive NetCDF files at the end of simulation**
- **Binary files**
  - **Transfer to e2e system using *bbcp***
  - **Convert to HDF5 format**
  - **Start up AVS/Express service**
  - **Generate images with AVS/Express**
  - **Archive HDF5 files in large chunks to HPSS**
- **Generate movies from the images**
- **Stop simulation if it does not progress properly**



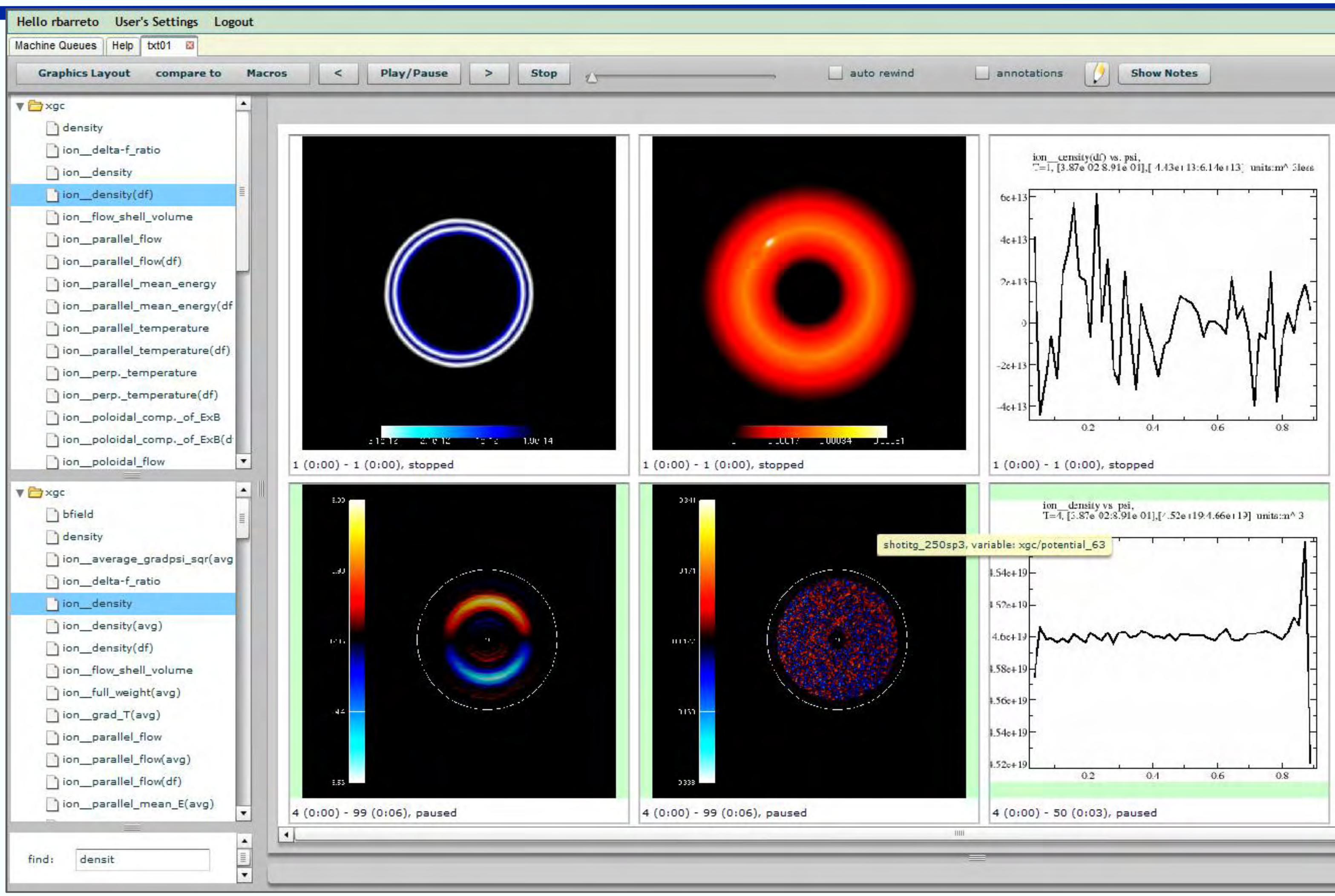


The Kepler Workflow



- **Kepler is a workflow execution system based on Ptolemy (open source from UCB)**
- **SDM center work is in the development of components for scientific applications (called actors)**

Real-time visualization and analysis capabilities on dashboard



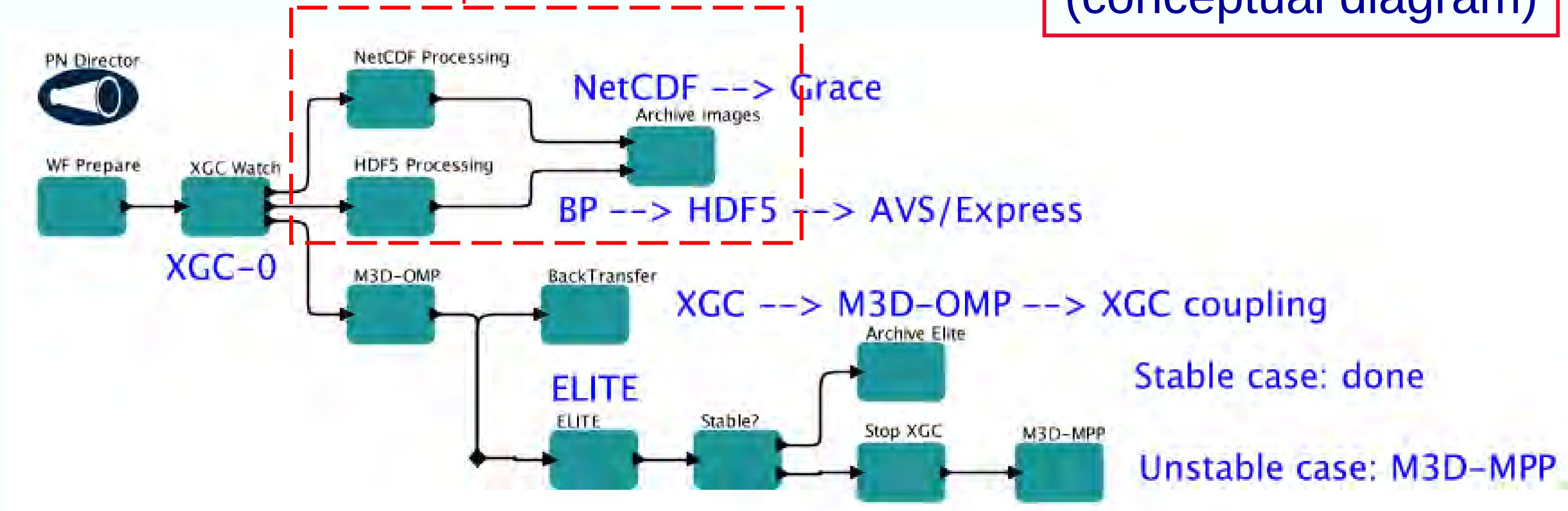
visualize and compare shots

Simulation Steering: Coupling XGC-0 and M3D Codes

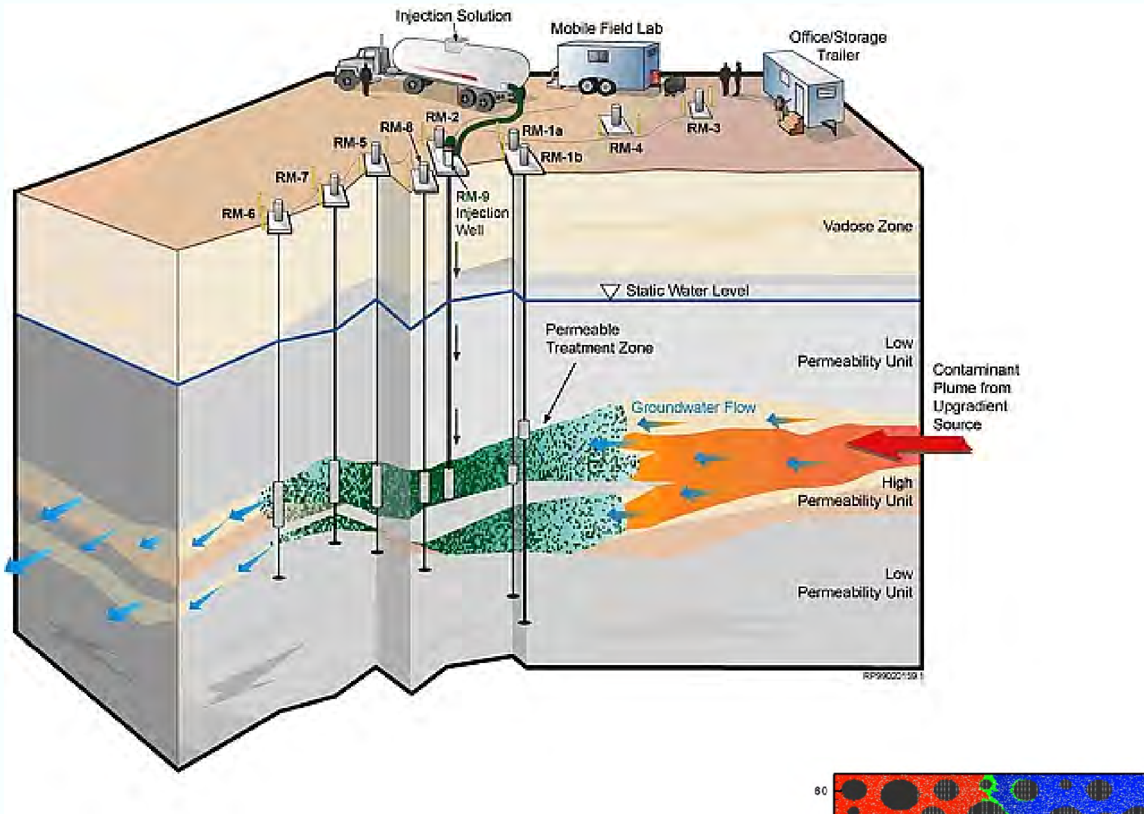
- **The processing loop transfers data regularly**
  - **from the machine that runs XGC-0 (jaguar)**
  - **to another machine (ewok)**
  - **for equilibrium and linear stability computations.**
- **If the linear stability test fails**
  - **a job is prepared and submitted to perform nonlinear parallel M3D-MPP computation.**

These represent sub-workflows

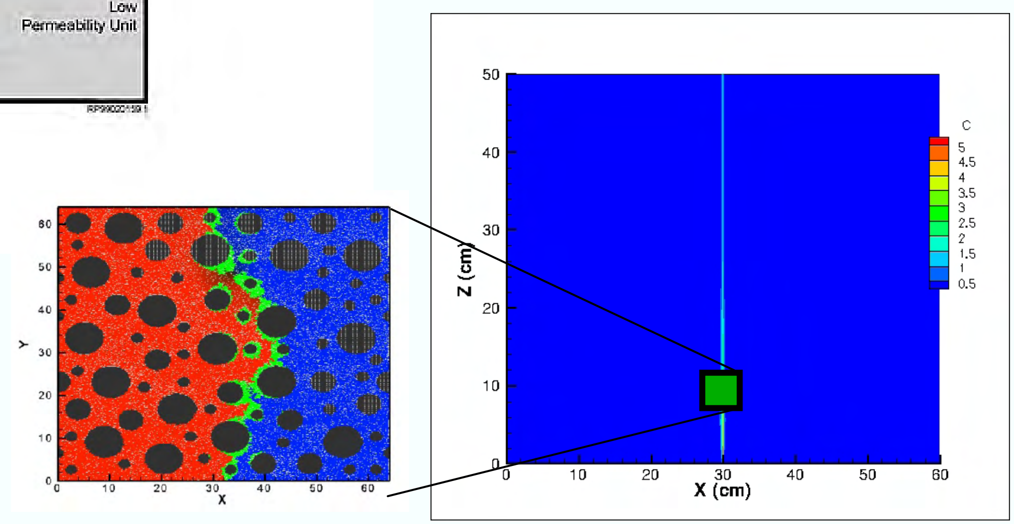
(conceptual diagram)



Using Kepler to Perform Parameter Studies in Subsurface Sciences



Hybrid Multiscale Modeling Benchmark Problem



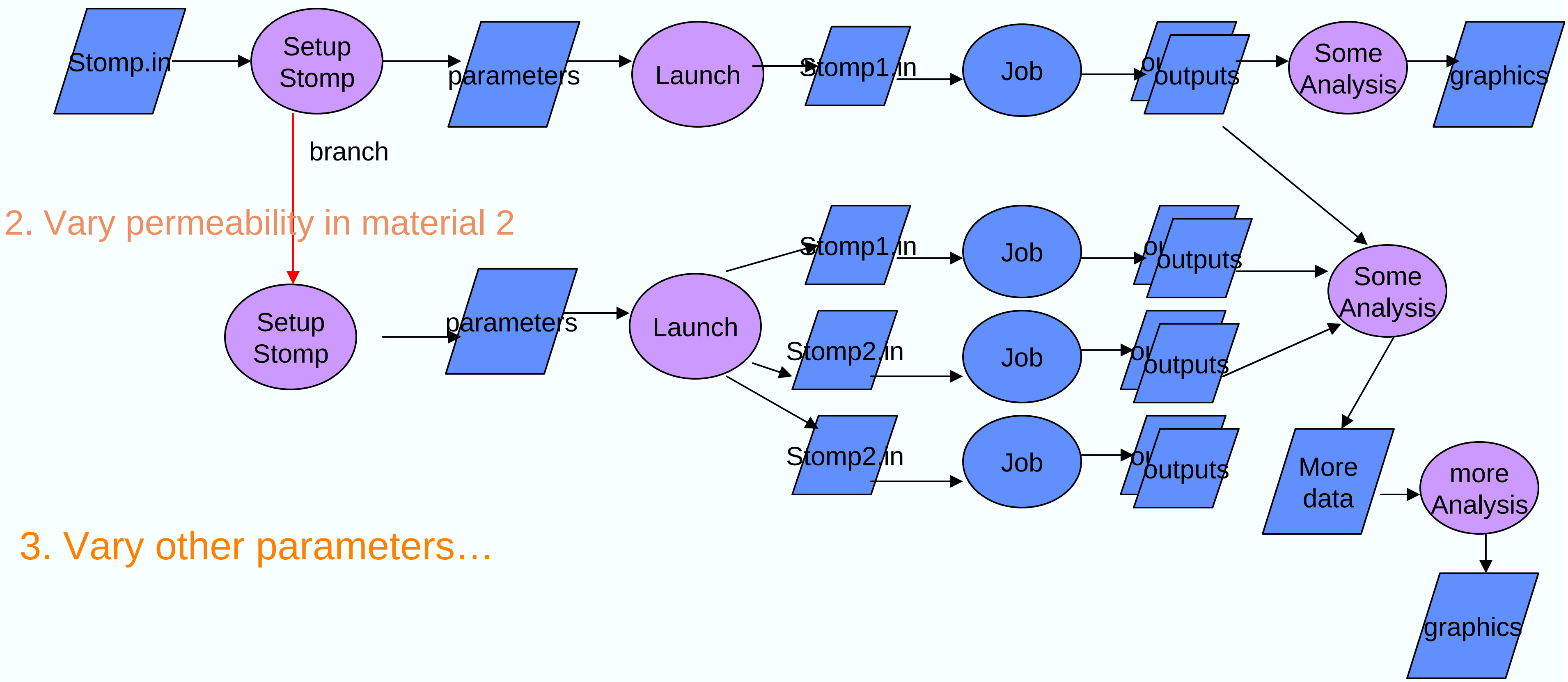
App Contact: Karen Schuchardt, PI, PNNL

SDM Contact: Terence Critchlow, PNNL

Workflow for parameter studies

User works within a “Study” where a Study can be represented as a graph of processes and data inputs/outputs. Some processes are triggered by the user, others appear as by-products of user actions.

1. Baseline computation



3. Vary other parameters…

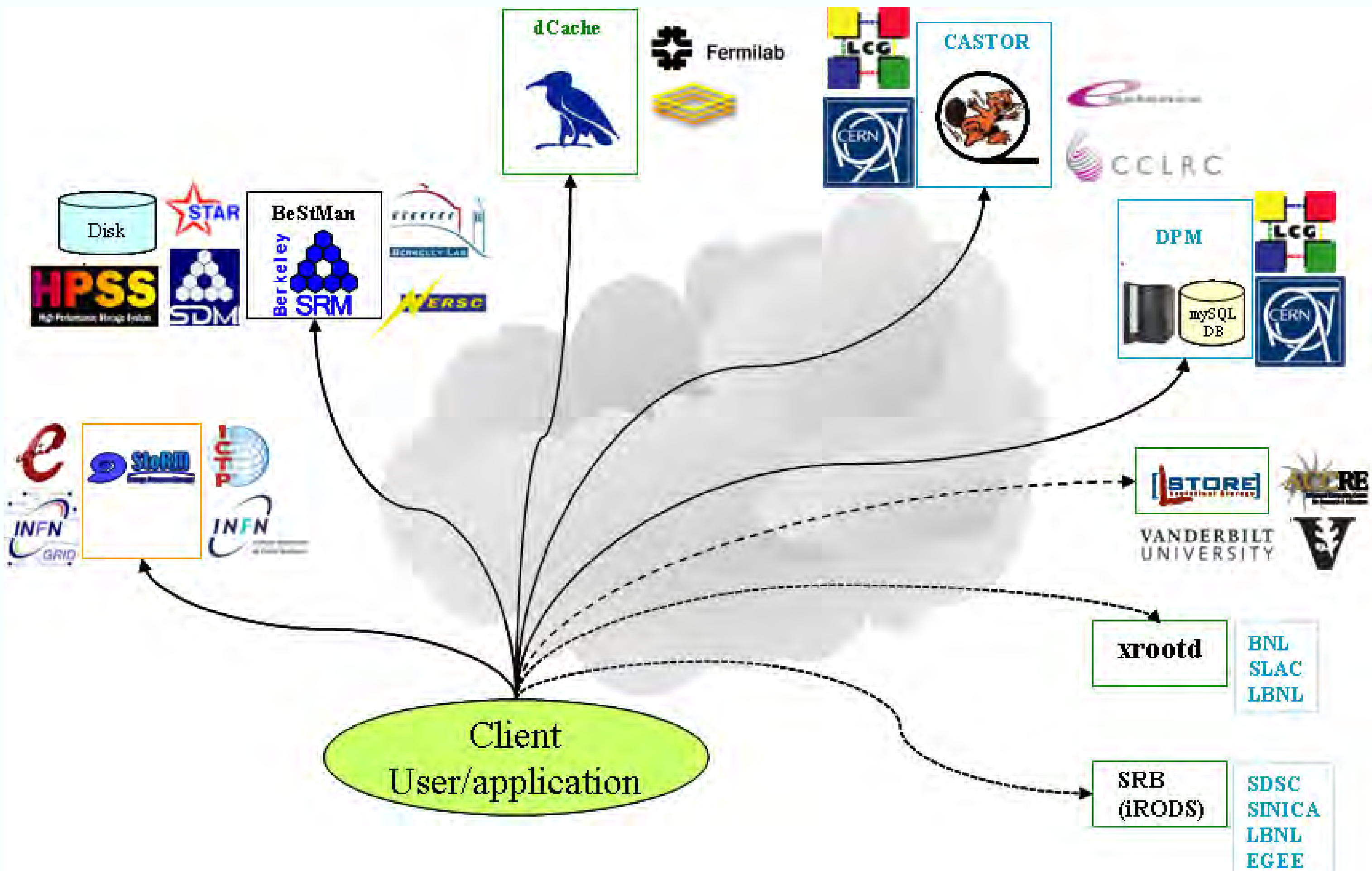
More Results

High Performance Technologies

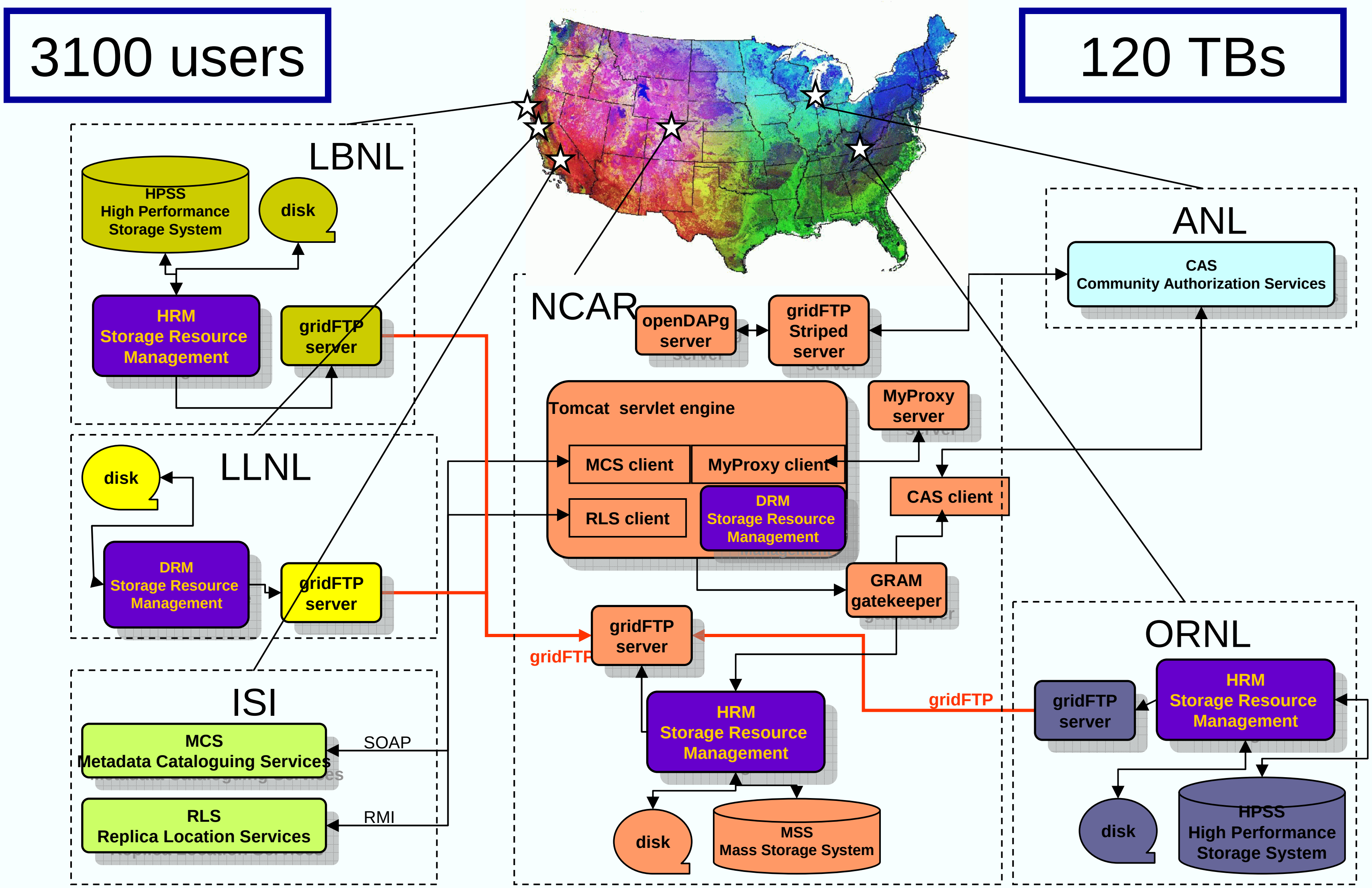
✓ **Usability and effectiveness**

Enabling Data Understanding

Storage Resource Managers (SRMs): Middleware for storage interoperability and data movement

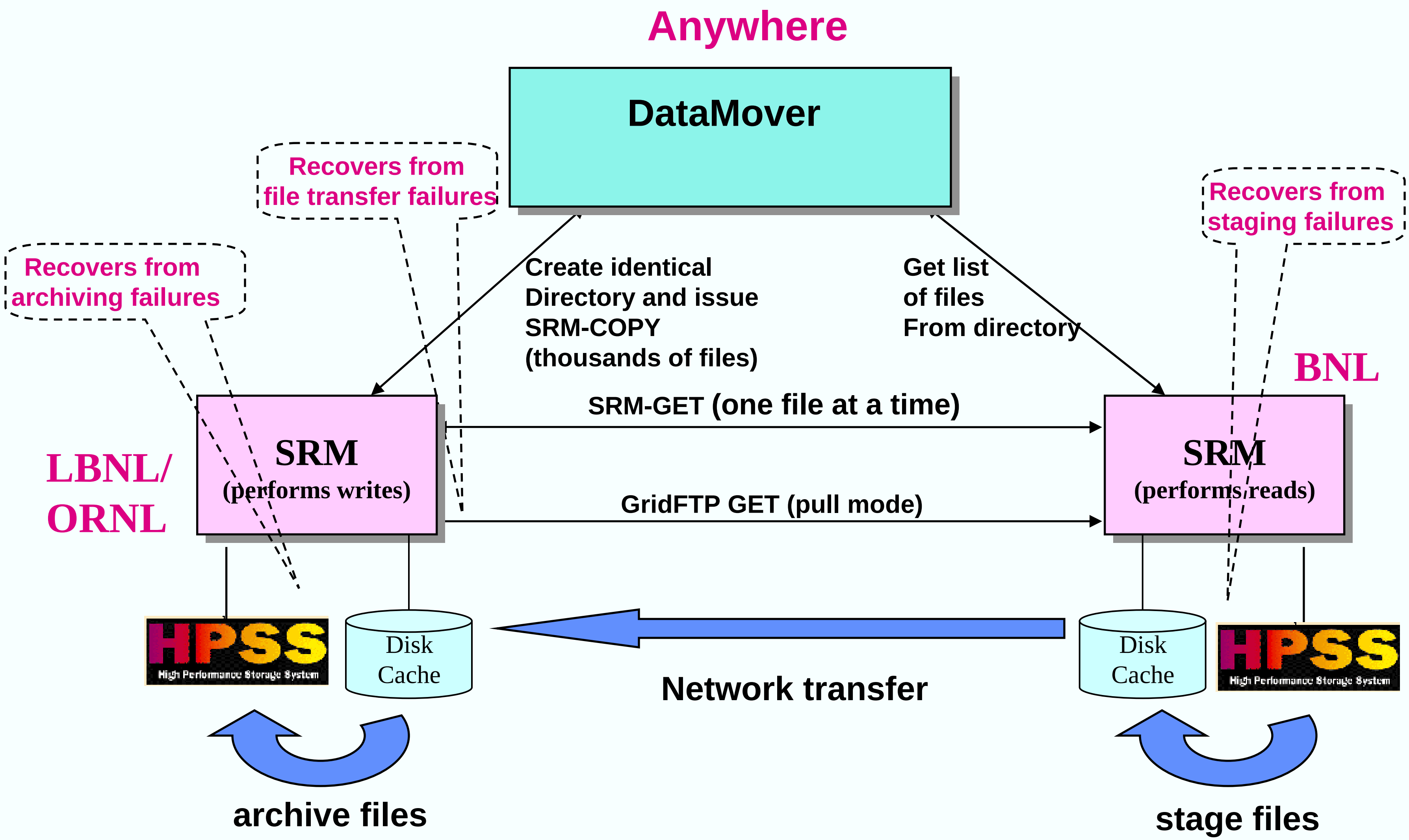


SRM use in Earth Science Grid



SDM Contact: A. Sim, A. Shoshani, LBNL

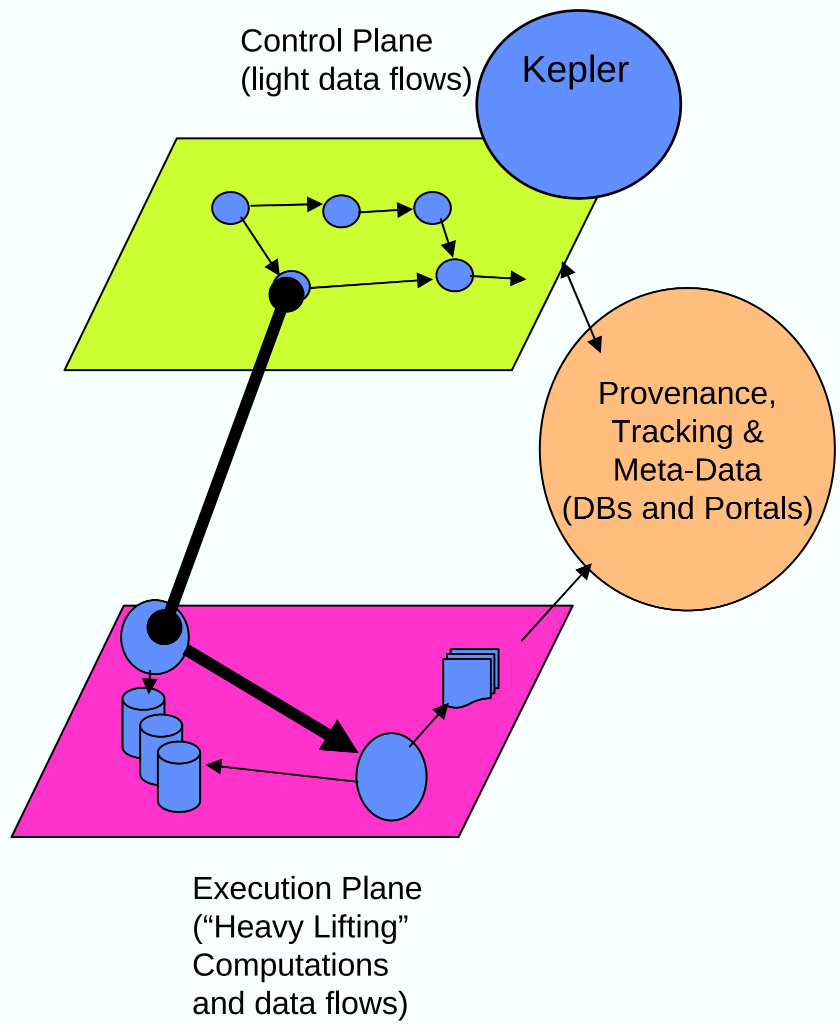
SRM as DataMover: Performs “rcp –r directory” on the WAN



50X *reduction in the error rates*, from 1% to 0.02% in the STAR project

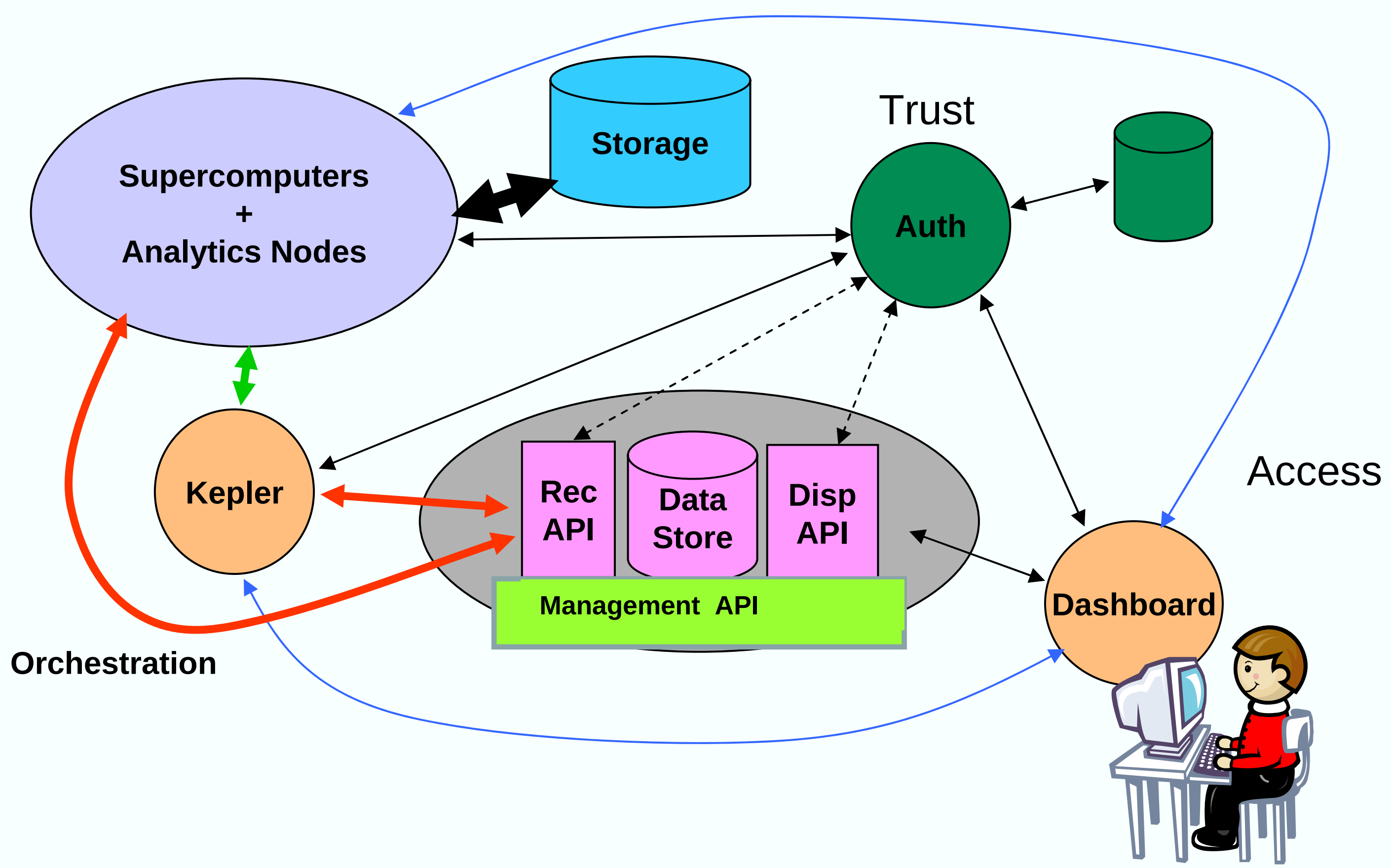
Capturing Provenance in Workflow Framework

- ***Process* provenance**
  - **the steps performed in the workflow, the progress through the workflow control flow, etc.**
- ***Data* provenance**
  - **history and lineage of each data item associated with the actual simulation (inputs, outputs, intermediate states, etc.)**
- ***Workflow* provenance**
  - **history of the workflow evolution and structure**
- ***System* provenance**
  - **Machine and environment information**
  - **compilation history of the codes**
  - **information about the libraries**
  - **source code**
  - **run-time environment settings**



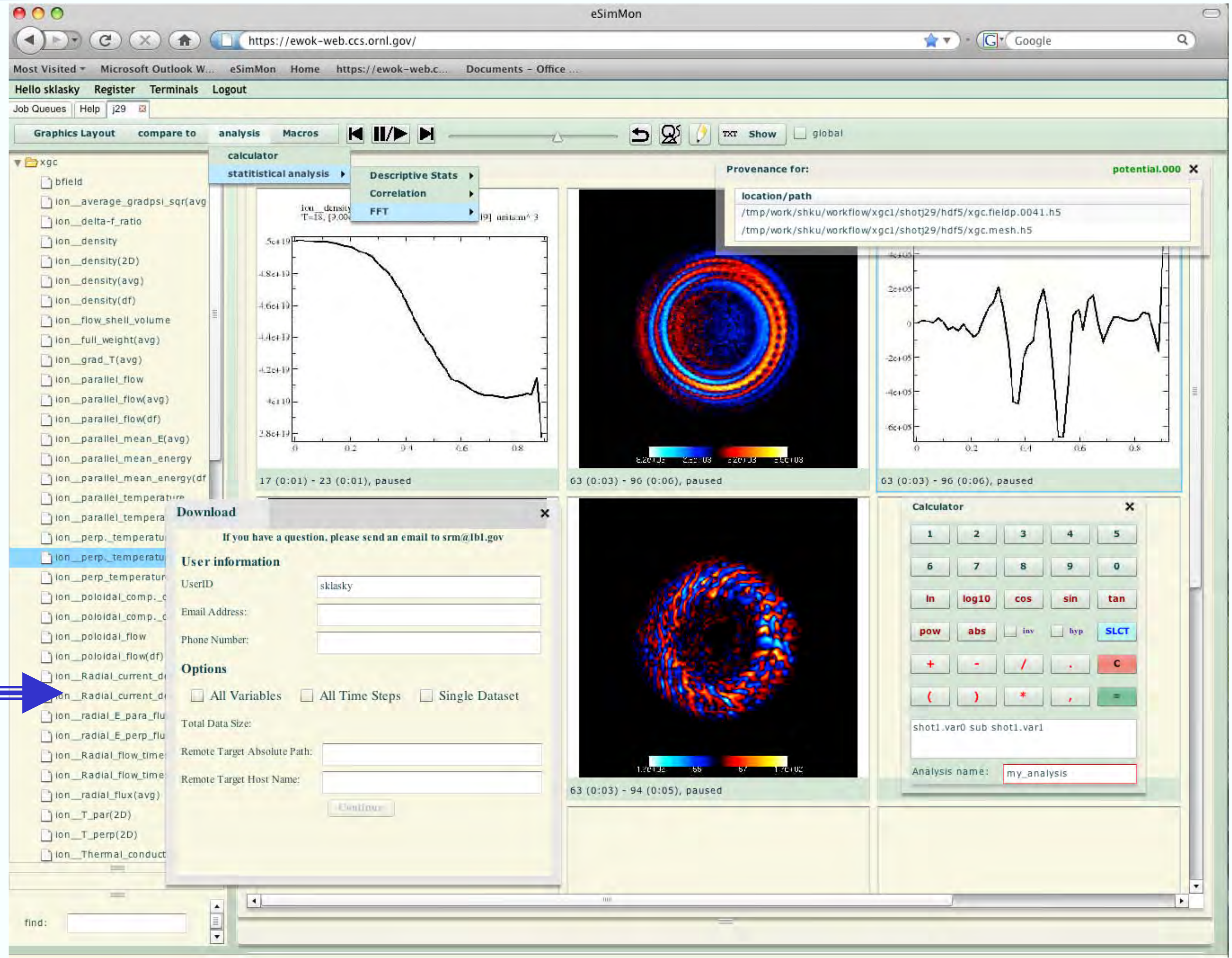
SDM Contact: Mladen Vouk, NCSU

FIESTA: Framework for Integrated End-to-end SDM Technologies and Applications



Provenance is captured in a data store and used by dashboard

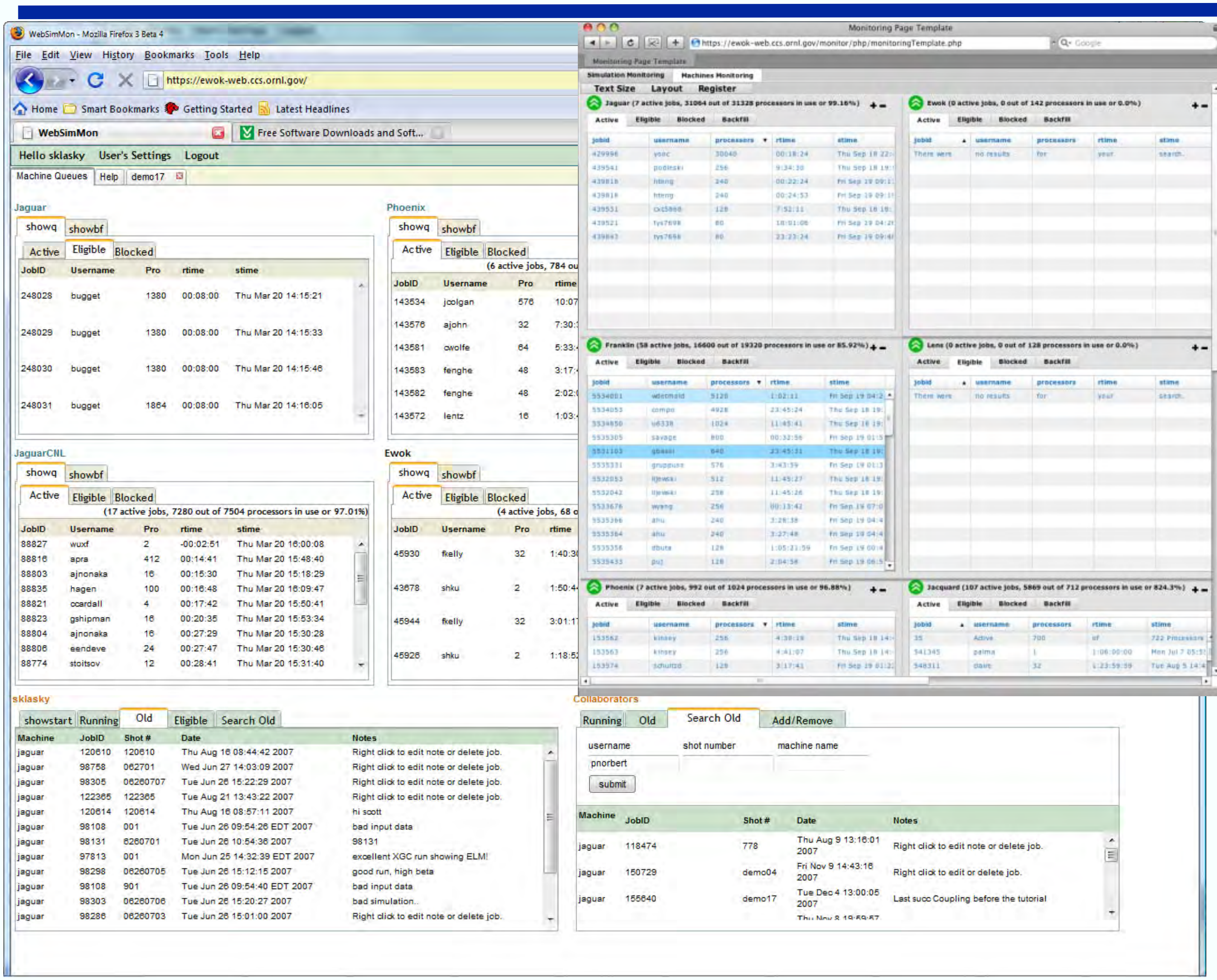
Dashboard uses provenance for finding location of files and automatic download with SRM



Download window

SDM Contact: Scott Klasky, ORNL

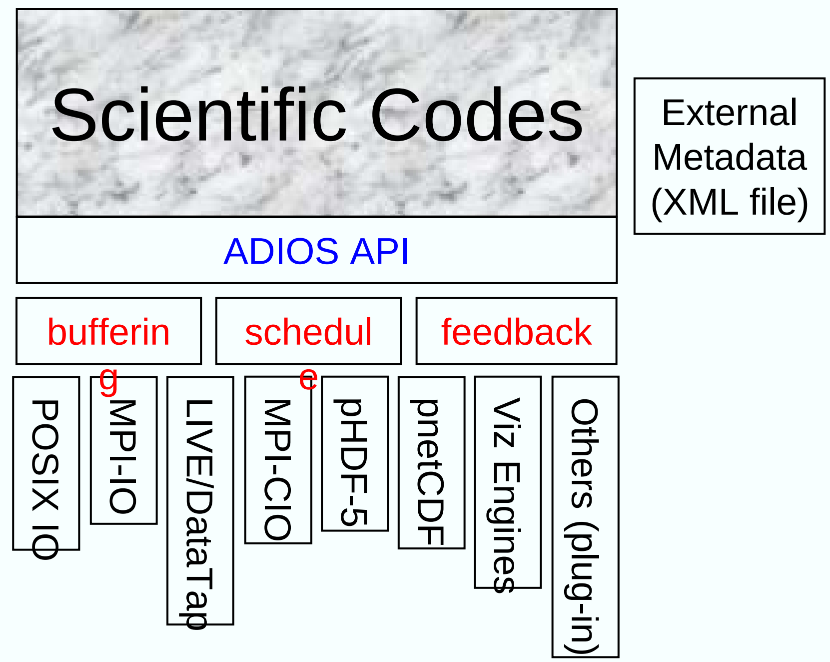
Dashboard is used for job launching and real-time machine monitoring



- Allow for secure logins with OTP.
- Allow for job submission.
- Allow for killing jobs.
- Search old jobs.
- See collaborators jobs.

Adaptable I/O system (ADIOS)

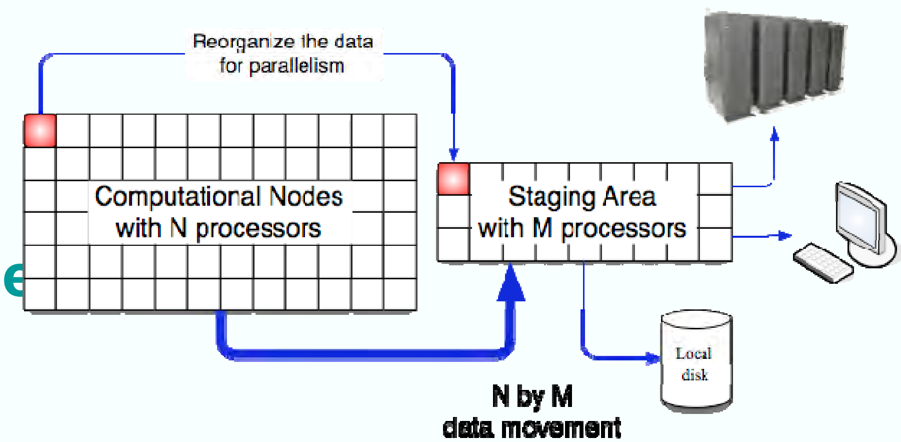
- **Allows plug-ins for different I/O implementations.**
- **Abstracts the API from the method used for I/O.**
- **Simple API, almost as easy as F90 write statement.**
- **Best practices/optimize IO routines for all supported transports "for free"**
- **Componentization.**
- **Thin API**
- **XML file**
  - **data groupings with annotation**
  - **IO method selection**
  - **buffer sizes**
- **Common tools**
  - **Buffering**
  - **Scheduling**
- **Pluggable IO routines**
- **Main advantages for users**
  - **No need to change code when running on various platforms**
  - **Change only external XML file**
  - **Asynchronous I/O saves computing cycles**



SDM Contact: Scott Klasky, ORNL

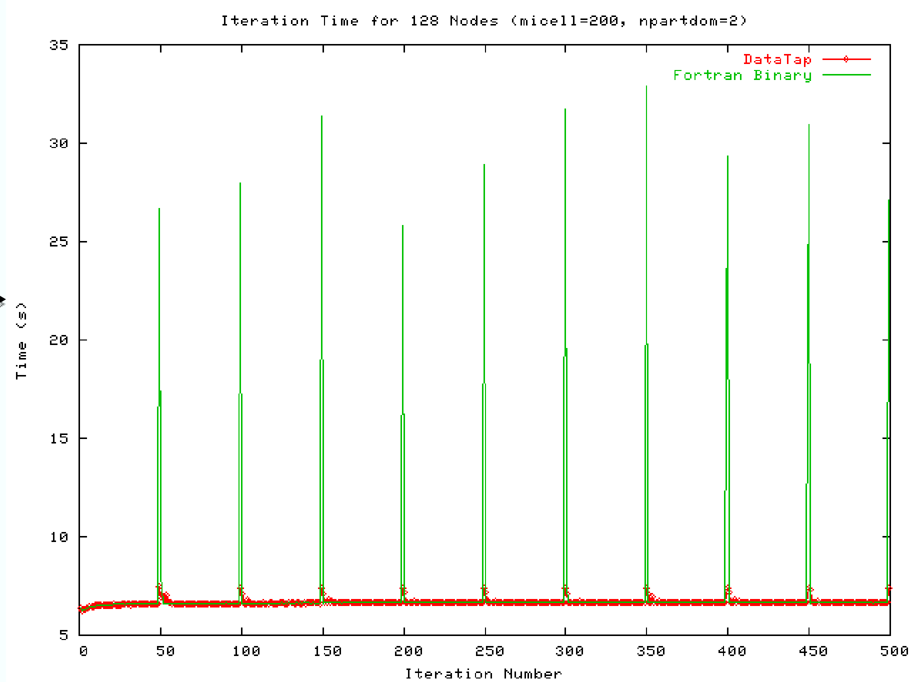
ADIOS Overview

- **ADIOS is an IO componentization, which allows us to**
  - **Abstract the API from the IO implementation**
  - **Switch from synchronous to asynchronous IO at runtime**
  - **provide fast IO at runtime**
- **Combines**
  - **Fast I/O routines**
  - **Easy to use**
  - **Scalable architecture (1000s cores) millions of proce**
  - **QoS**
  - **Metadata rich output**
  - **Visualization applied during simulations**
  - **Analysis, compression techniques applied during simulations**
  - **Provenance tracking**



Initial ADIOS performance.

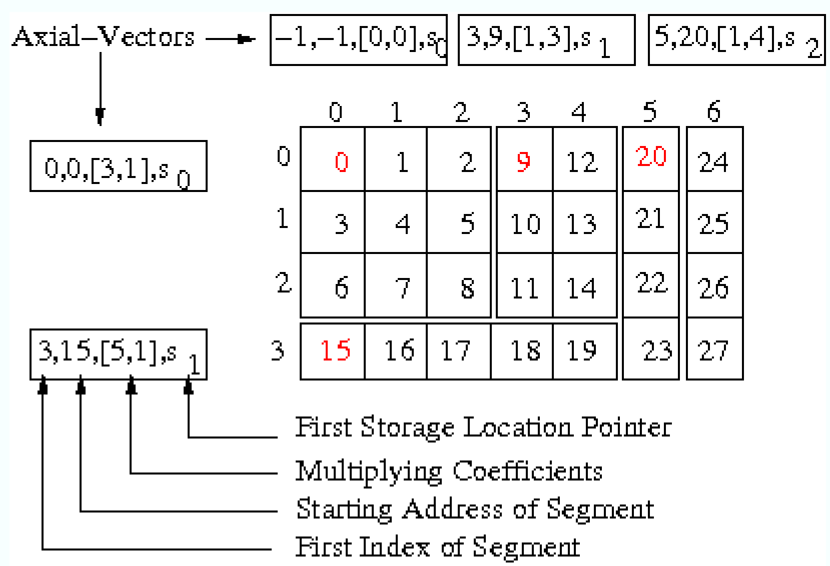
- **June 7, 2008: 24 hour GTC run on Jaguar at ORNL**
  - 93% of machine (28,672 cores)
  - MPI-OpenMP mixed model on quad-core nodes (7168 MPI procs)
  - three interruptions total (simple node failure) with 2 10+ hour runs
  - Wrote 65 TB of data at >20 GB/sec (25 TB for post analysis)
  - IO overhead ~3% of wall clock time.
  - Mixed IO methods of synchronous MPI-IO and POSIX IO configured in the XML file
- **DART: <2% overhead for writing 2 TB/hour with XGC code.**
- **DataTap vs. Posix**
  - 1 file per process (Posix).
  - 5 secs for GTC computation.
  - ~25 seconds for Posix IO
  - ~4 seconds with DataTap



SDM CENTER

Extendable Arrays

- **Dense arrays that grow dynamically by extent of dimensions, or number of dimensions need to be restructured. How can that be avoided?**
- **Example**
  - **A 2-D array initially dened as A[3][3] and then extended by 2 columns, then by 1 row, followed by 1 column and so on.**
- **Developed libraries**
  - **Inserting blocks**
  - **Reading any array sub-structure**
- **Sparse arrays**
  - **Developed new method for HDF5**
  - **Balanced Extendible Hashing**



SDM Contact: Ekow Otoo, LBNL

Results

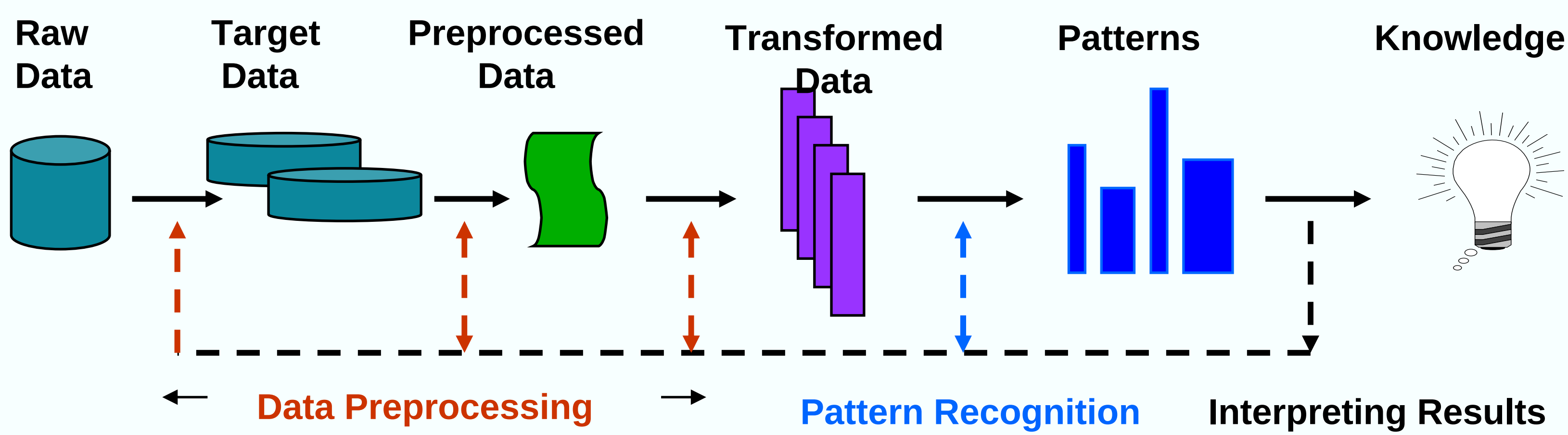
High Performance Technologies

Usability and effectiveness

✓ Enabling Data Understanding

Scientific data understanding: from Terabytes to a Megabytes

- **Goal: solving the problem of data overload**
  - **Use scientific data mining techniques to analyze data from various SciDAC applications**
  - **Techniques borrowed from image and video processing, machine learning, statistics, pattern recognition, …**



Data Fusion
Sampling
Multi-resolution analysis

De-noising
Object - identification
Feature- extraction
Normalization

Dimension- reduction

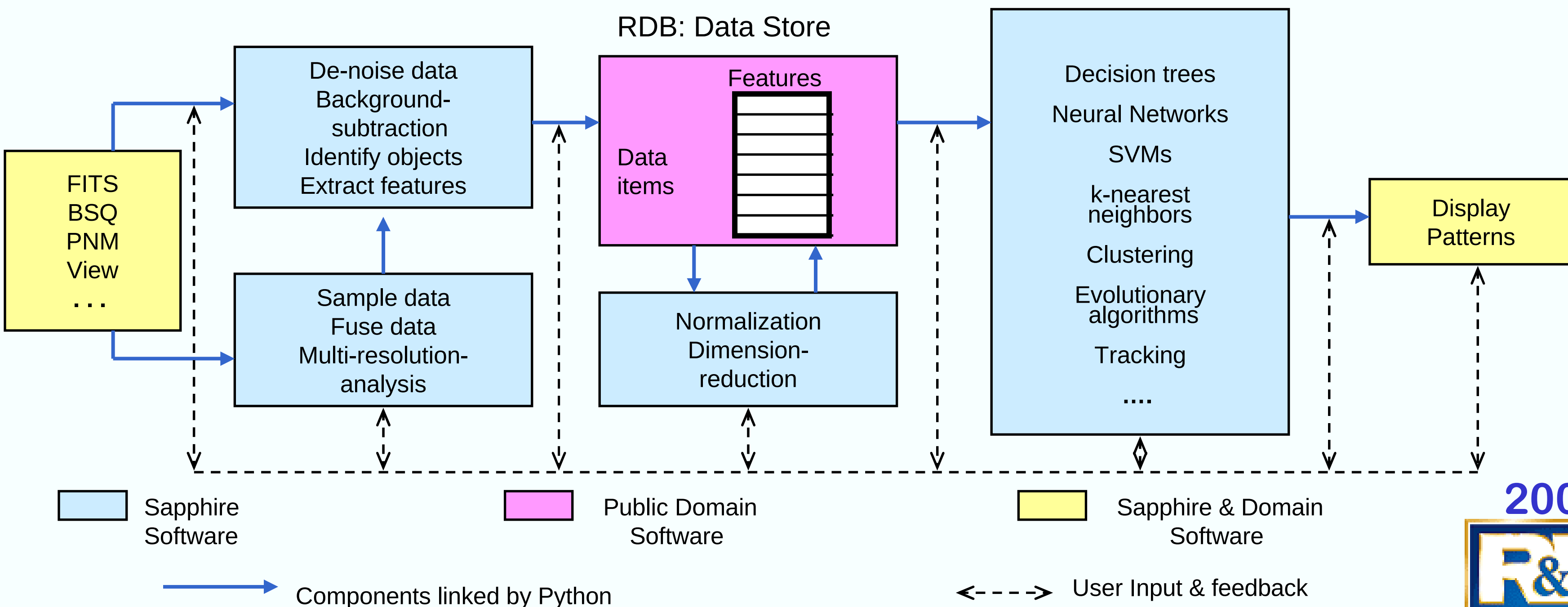
Classification
Clustering
Regression

Visualization
Validation

An iterative and interactive process

Sapphire: scientific data mining

- **research in robust, accurate, scalable algorithms**
- **modular, extensible software**
- **analysis of data from practical problems**
- **Leverage funding through DOE NNSA, LLNL LDRD, GSEP SciDAC project, and SDM SciDAC Center**

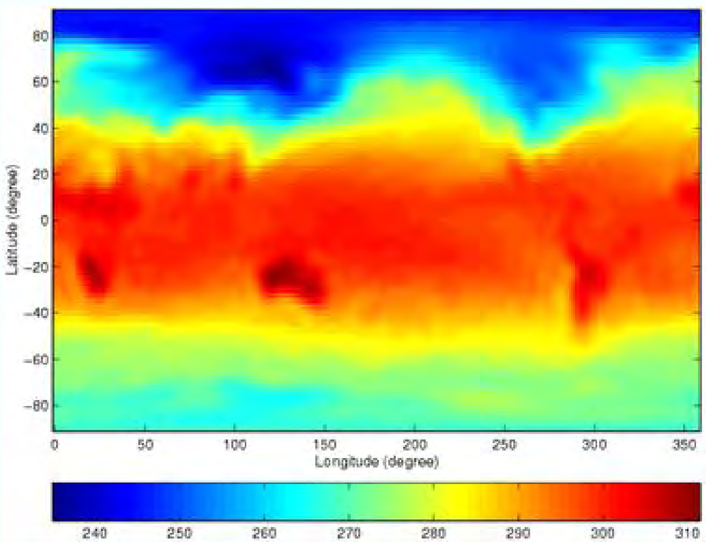


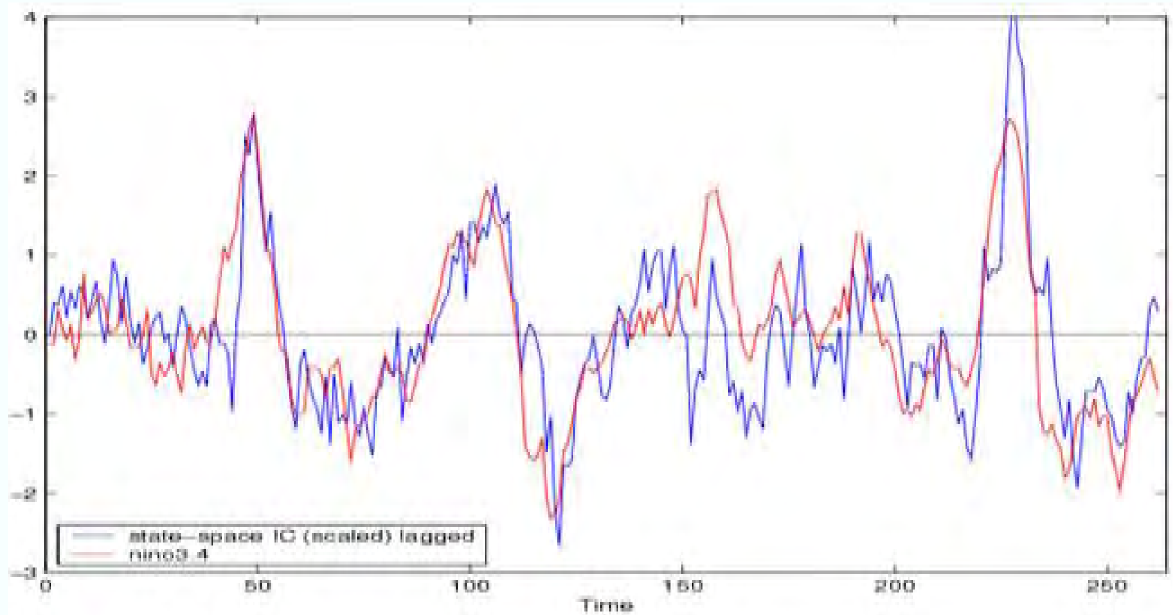
2006

SDM Contact: Chandrika Kamath, LLNL

Separating signals in climate data

- We used independent component analysis to separate El Niño and volcano signals in climate simulations
- Showed that the technique can be used to enable better comparisons of simulations

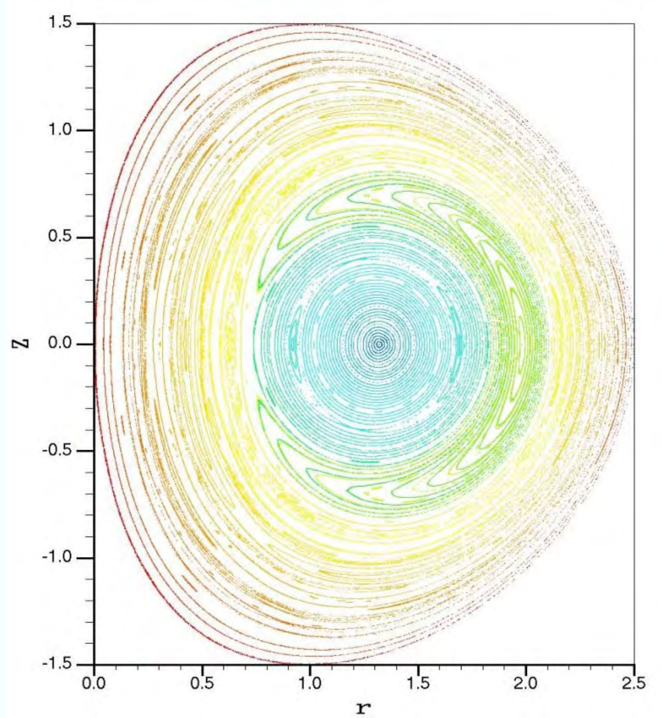




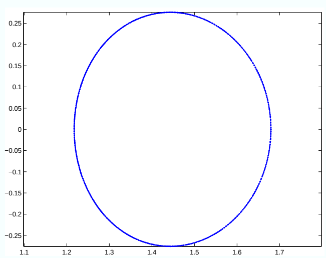
Collaboration with Ben Santer (LLNL)

Classification of puncture (Poincaré) plots for NCSX

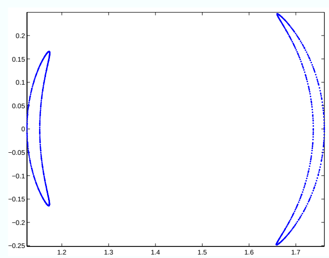
- Joint work with PPPL (Klasky, Pomphrey, Monticello)
- Classify each of the nodes: quasiperiodic, islands, separatrix
- Connections between the nodes
- Want accurate and robust classification, valid when few points in each node



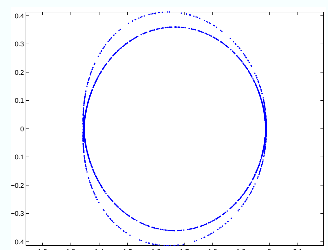
National Compact Stellarator Experiment



Quasiperiodic



Islands

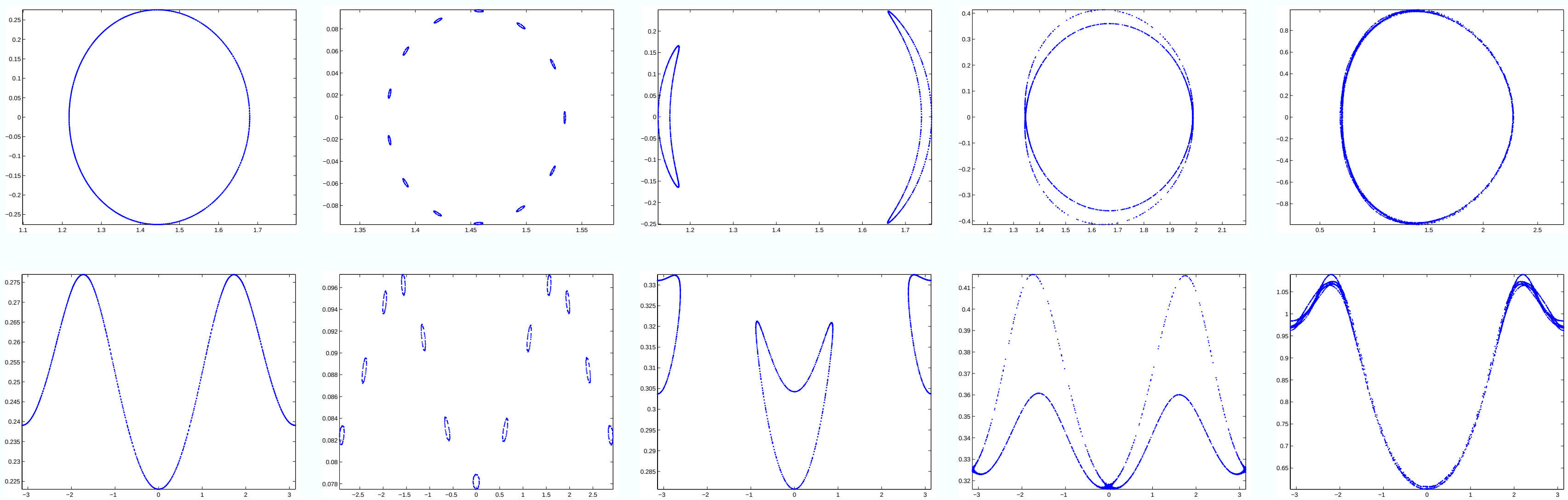


Separatrix

Collaboration with J. Breslau, N. Pomphrey, D. Monticello(PPPL), S. Klasky(ORNL)

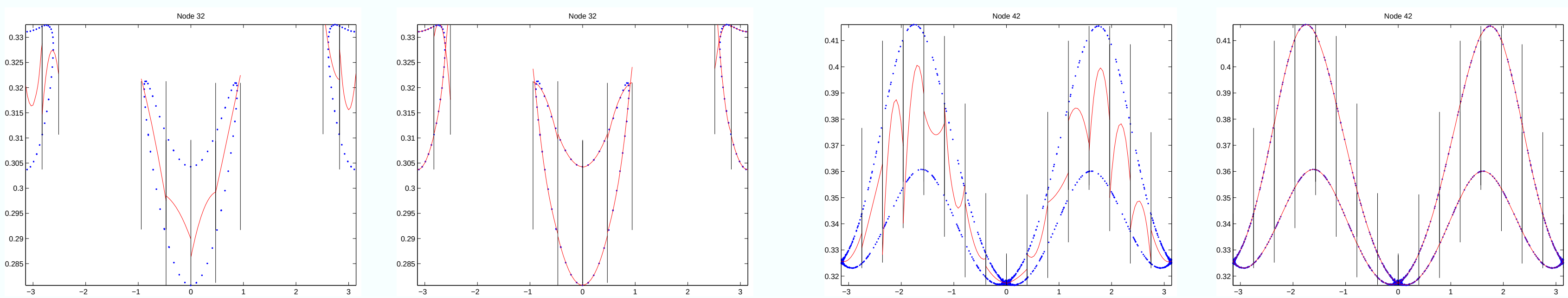
Polar Coordinates

- **Transform the (x,y) data to Polar coordinates (r,θ).**
- **Advantages of polar coordinates:**
  - **Radial exaggeration reveals some features that are hard to see otherwise.**
  - **Automatically restricts analysis to radial band with data, ignoring inside and outside.**
  - **Easy to handle rotational invariance.**



Piecewise Polynomial Fitting: Computing polynomials

- **In each interval, compute the polynomial coefficients to fit 1 polynomial to the data.**
- **If the error is high, split the data into an upper and lower group. Fit 2 polynomials to the data, one to each group.**

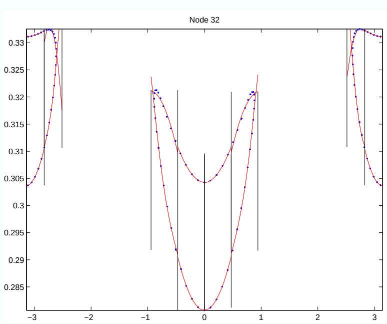


Blue: data. Red: polynomials. Black: interval boundaries.

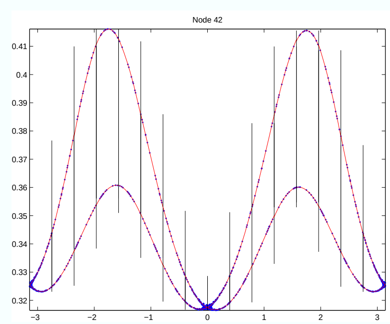
Classification

- The number of polynomials needed to fit the data and the number of gaps gives the information needed to classify the node:

	Number of polynomials	
Gaps	**one**	**two**
Zero	Quasiperiodic	Separatrix
> Zero		Islands



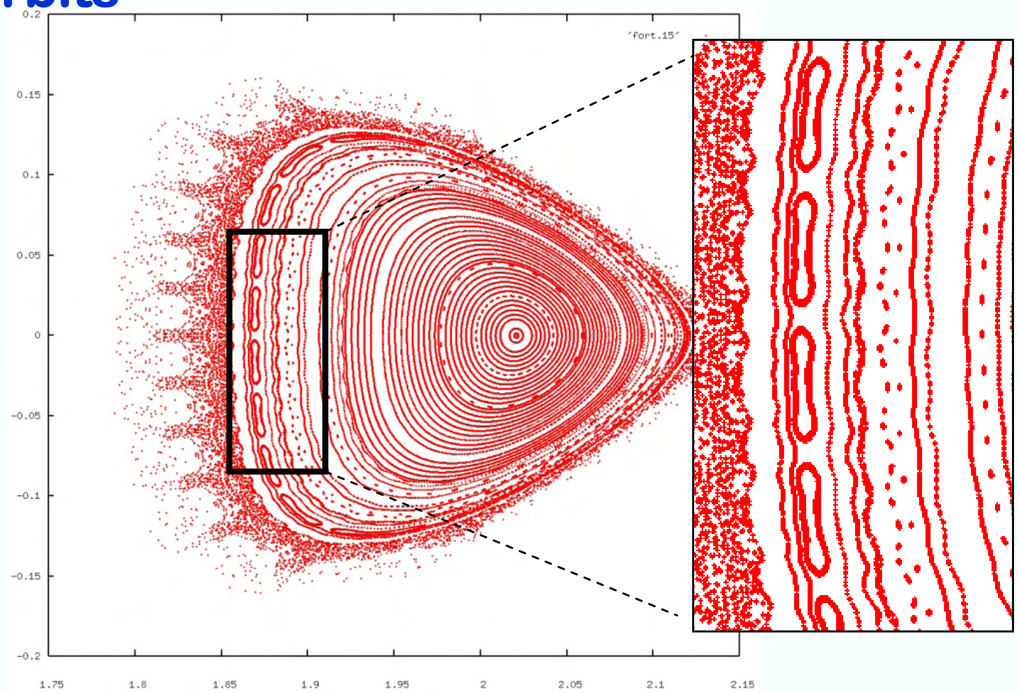
2 Polynomials
2 Gaps
→ Islands



2 Polynomials
0 Gaps
→ Separatrix

How do we extract representative features for an orbit?

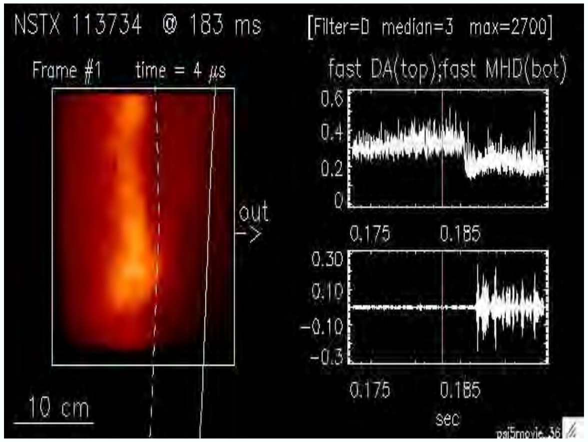
- Variation in the data makes it difficult to identify good features and extract them in a robust way
- Issues with labels assigned to orbits
- Next steps: characterizing island chains and separatrix orbits



Identifying missing orbits

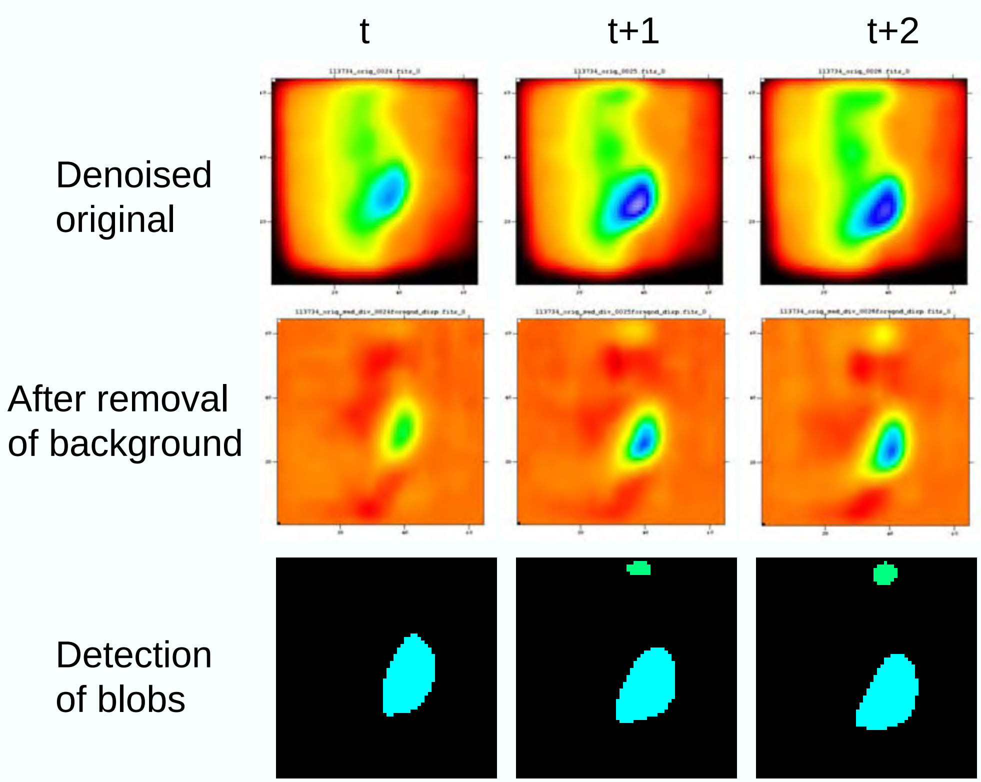
Understand the turbulence which causes leakage of the fusion plasma

- **Requirements for fusion – high temperature and confined plasma**
- **Fine-scale turbulence at the edge causes leakage of plasma from the center to the edge**
  - **Loss of confinement**
  - **Heat loss of plasma**
  - **Erosion or vaporization of the containment wall**



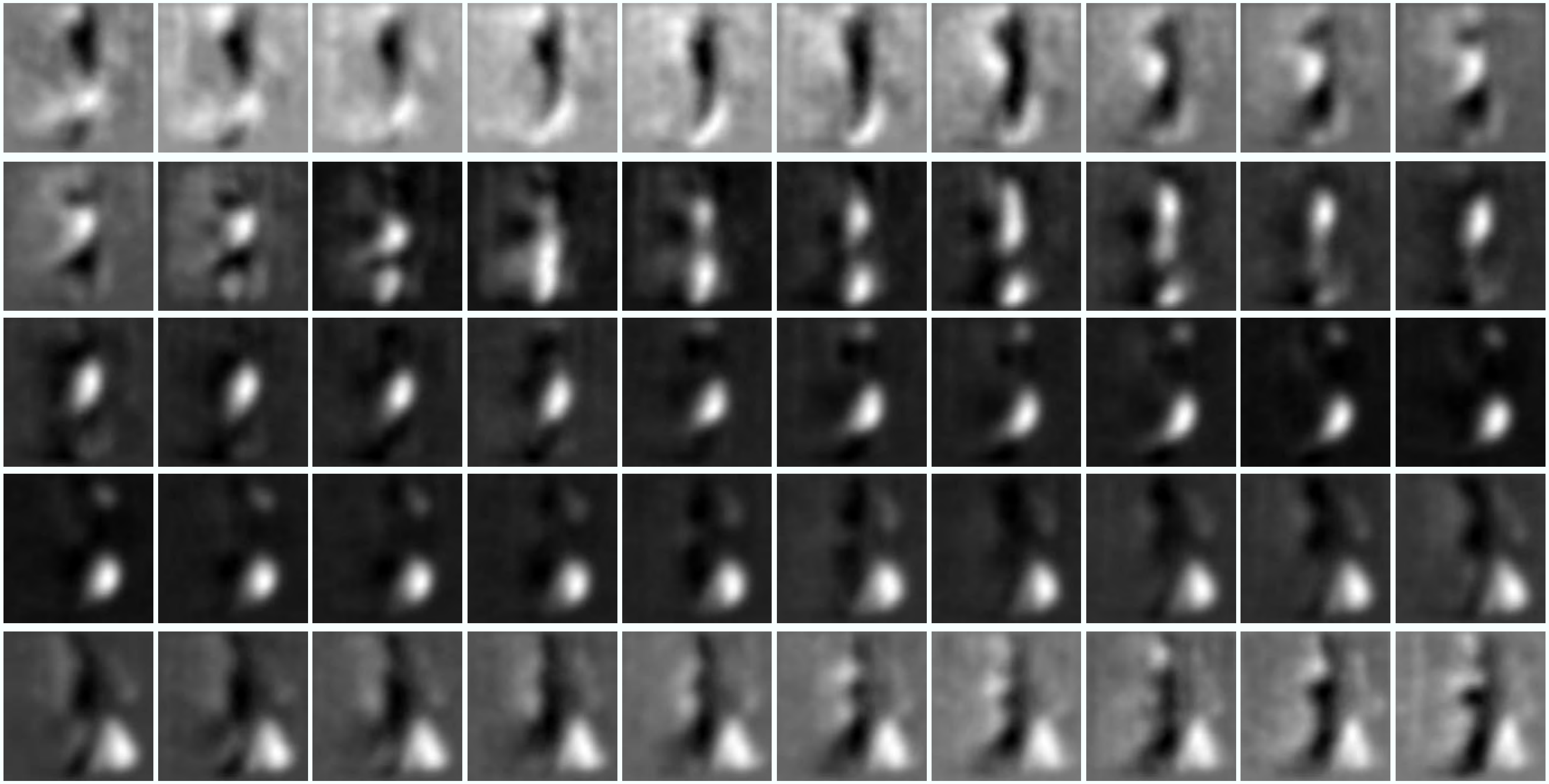
Tracking blobs in fusion plasma

- **Using image and video processing techniques to identify and track blobs in experimental data from NSTX to validate and refine theories of edge turbulence**



Collaboration with S. Zweben, R. Maqueda, and D. Stotler (PPPL)

Example frames to segment (sequence 113734: frames 1-50)

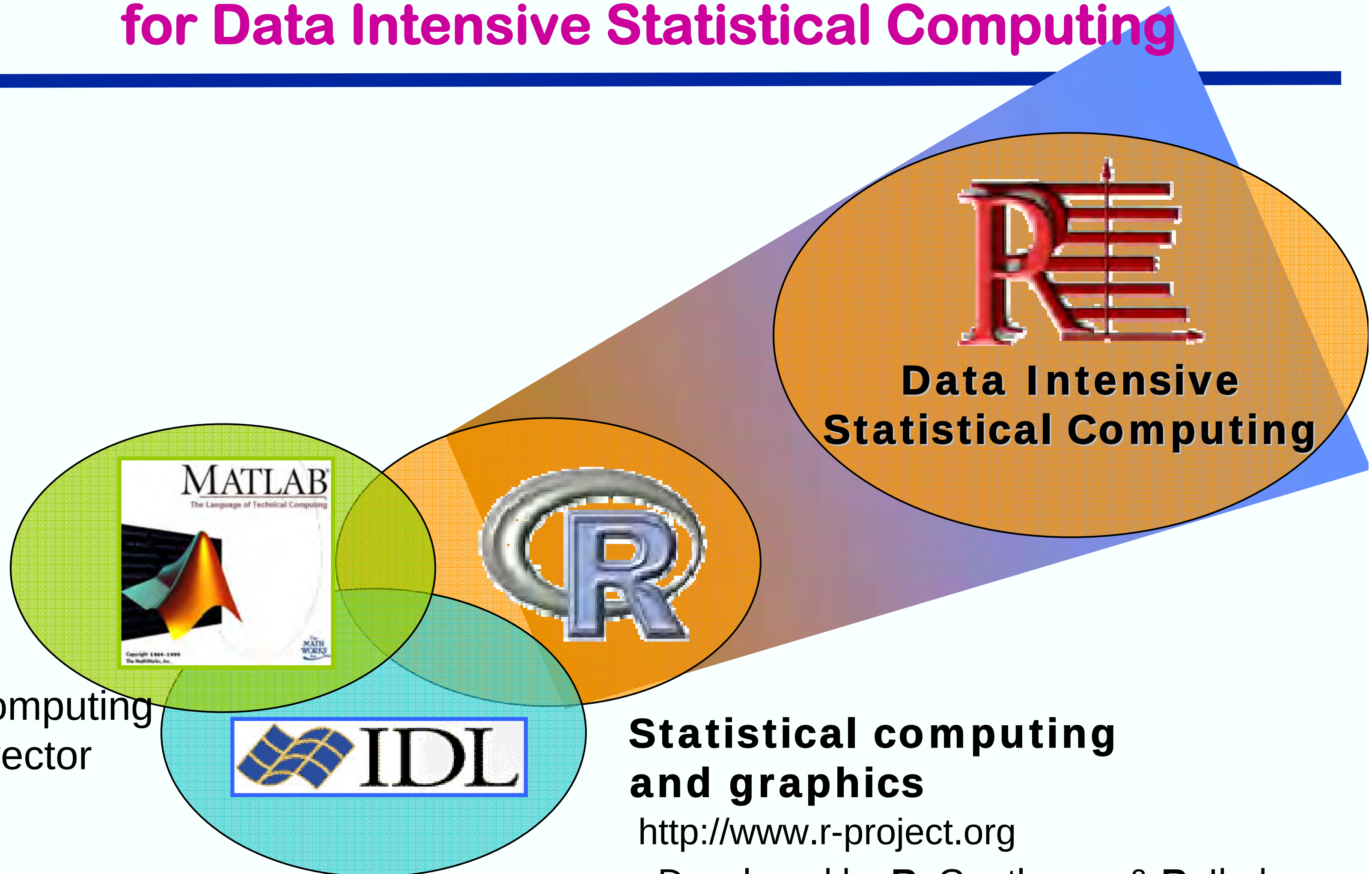


We are investigating several image segmentation methods

- **Techniques tried:**
  - Immersion-Based: basic immersion, constrained watershed, watershed merging
  - Region Growing: seeded region growing, seed competition
  - Model-Based: 2-D Gaussian fit
- **Challenges**
  - how do we select the parameters in an algorithm,
  - how do we handle the variability in the data especially for longer sequences,
  - how do the choices of algorithms and parameters influence the “science”, …
- **Why is this difficult?**
  - coherent structures are poorly understood empirically and not understood theoretically
  - no known ground-truth
  - noisy images
  - variation within a sequence

Work in progress

Parallel R (pR) Technology for Data Intensive Statistical Computing



•Technical computing
•Matrix and vector formulations

•Data Visualization and analysis platform
•Image processing, vector computing

Statistical computing and graphics

http://www.r-project.org

- Developed by **R**. Gentleman & **R**. Ihaka
- Expanded by community as open source
- Extensible via dynamically loadable libs

Contact: Nagiza Samatova, ORNL (samatovan@ornl.gov)

Task and Data Parallelism in pR

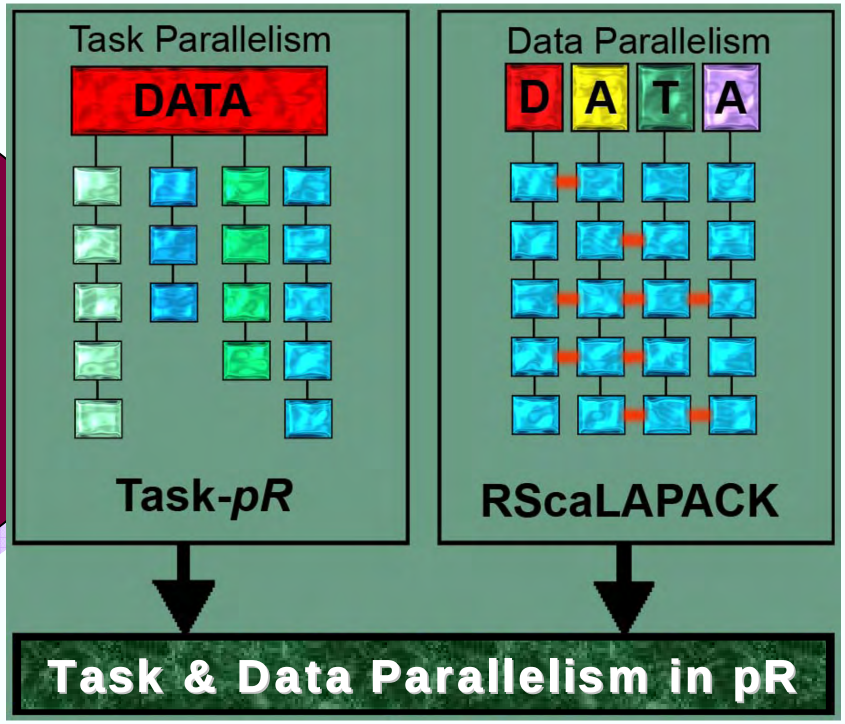
Goal: Parallel R (*pR*) aims:

(1) to automatically detect and execute *task-parallel* analyses;

(2) to easily plug-in *data-parallel* MPI-based C/Fortran codes

(3) to retain high-level of *interactivity, productivity* and *abstraction*

Task Parallelism **Data Parallelism**



Task-parallel analyses:

- Likelihood Maximization
- Re-sampling schemes: Bootstrap, Jackknife
- Markov Chain Monte Carlo (MCMC)
- Animations

Data-parallel analyses:

- *k*-means clustering
- Principal Component Analysis
- Hierarchical clustering
- Distance matrix, histogram, etc.

ProRata use in OBER Projects

TOOLbox

ProRata

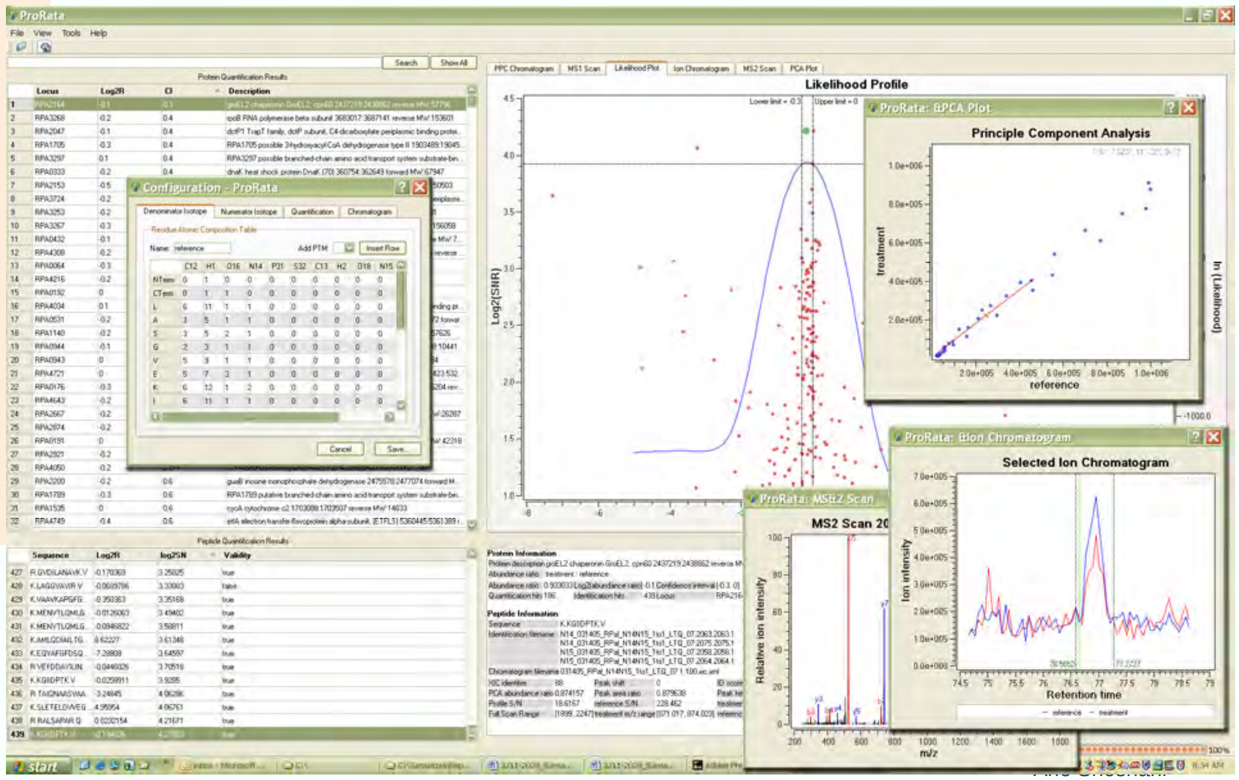
To detect quantitative protein differences between two conditions, researchers often perform stable-isotope labeling. The algorithms that currently are applied to the data, however, can make incorrect assignments because the S/N typically is low in these experiments. Also, the programs do not assess bias and variability. So, Nagiza Samatova, Robert Hettich, and colleagues at the Oak Ridge National Laboratory and the University of Tennessee developed a p… ves i… e ap- p… sta- b…

>1,000 downloads

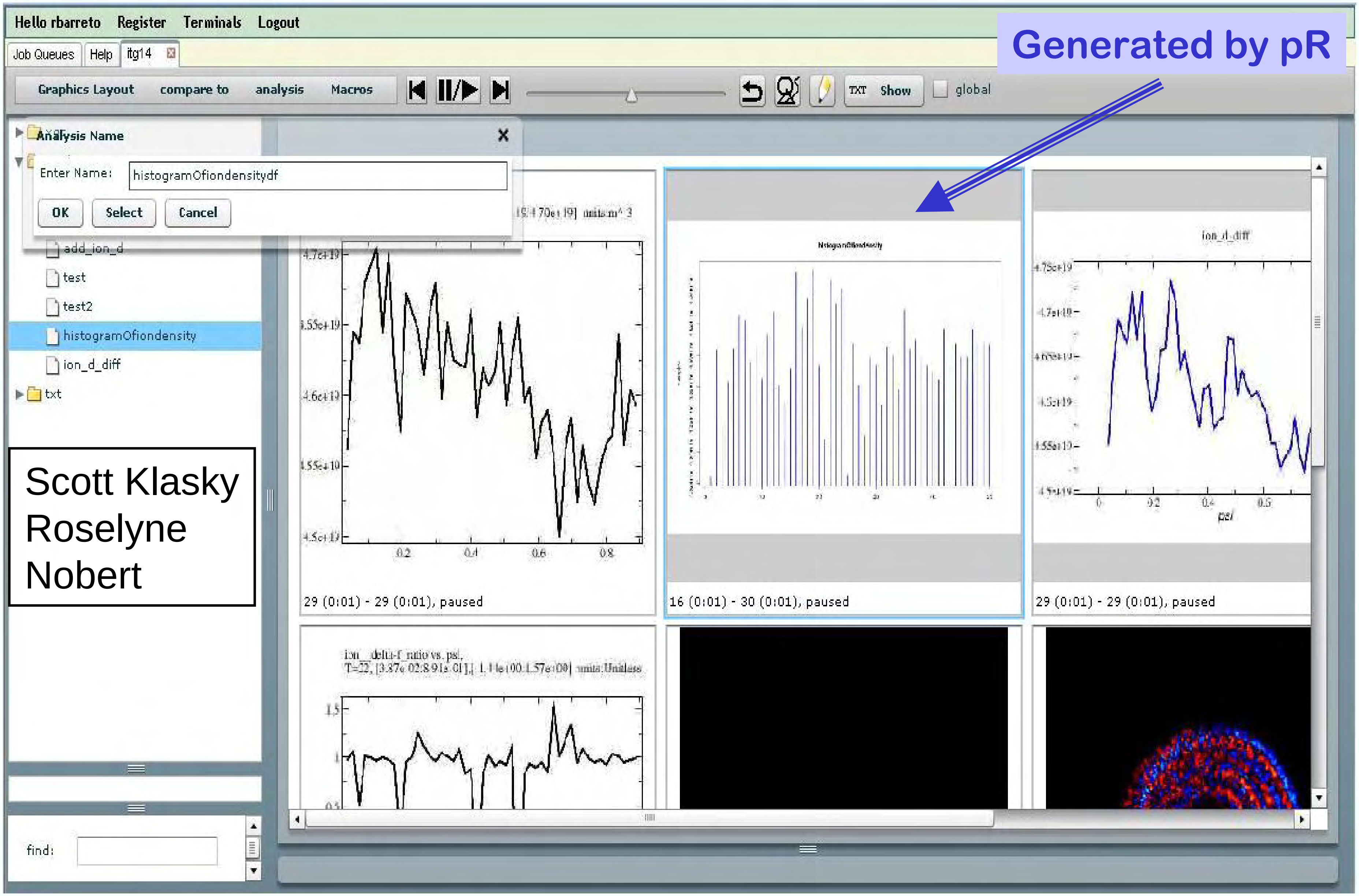
J. of Proteome Research
Vol. 5, No. 11, 2006

DOE OBER Projects Using ProRata:

- *Jill Banfield, Bob Hettich*: Acid Mine Drainage
- *Michelle Buchanan*: CMCS Center
- *Steve Brown, Jonathan Mielenz*: BESC BioEnergy
- *Carol Harwood, Bob Hettich*: MCP *R. palustris*



Dashboard Interface to pR



Generated by pR

Scott Klasky
Roselyne
Nobert

SDM center collaboration with applications

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
Climate Modeling (Drake)	workflow				pNetCDF	pMatlab		
Astrophysics (Blondin)	data movement	dashboard						
Combustion (Jackie Chen)	data movement	distributed analysis	DataMover-Lite	flame front	Global Access	pMatlab	tranient events	
Combustion (Bell)			DataMover-Lite					
Fusion (PPPL)							poincare plots	
Fusion (CPES)	data-move, code-couple	Dashboard	DataMover-Lite	Toroidal meshes		pR	Blob tracking	
Materials - QBOX (Galli)					XML			
High Energy Physics	Lattice-QCD		SRM, DataMover	event finding				
Groundwater Modeling	identified 4-5 workflows							
Accelarator Science (Ryne)					MPIO-SRM			
SNS	workflow	Data Entry tool (DEB)						
Biology	ScalaBlast					ProRata		ScalaBlast
Climate Cloud modeling (Randall)					pNetCDF			cloud modeling
Data-to-Model Coversion (Kotamathi)								
Biology (H2)								
Fusion (RF) (Bachelor)							poincare plots	
Subsurface Modeling (Lichtner)						Over AMR		
Flow with strong shocks (Lele)						conditional statistics		
Fusion (extended MHD) (Jardin)								
Nanoscience (Rack)						pMatlab		
other activities								integrate with Luster

currently in progress — problem identified — interest expressed

Arie Shoshani

SDM center collaboration with other centers/institutes

Centers & institutions	Workflow Technology (Kepler)	Metadata And provenance	Data Movement and storage	Indexing (FastBit)	Parallel I/O (pNetCDF, etc.)	Parallel Statistics (pR, …)	Feature extraction	Active Storage
Open Science Grid			SRM-tester					
Earth System Grid			SRM and DML					
Petascale Storage Institute					Posix-IO			
Vis Institute (Ma)				query-based vis	put parallel I/O in Vis	pR		
Vis Center (Bethel)	workflow in vis			query-based vis		pR		

currently in progress (orange) — problem identified (amber) — interest expressed (yellow)

Summary Remarks (1)

- SDM center has developed data management tools that provide
  - High performance
    - now at petascale, planning for exascale
    - across the I/O software stack
    - Specialized Indexing technologies
    - Parallel analysis tools
  - Usability and effectiveness
    - Developed FIESTA: a Framework for Integrated End-to-end SDM Technologies and Applications
    - Based on workflow and dashboard technologies
    - Provide real-time monitoring, repeated analysis, code coupling
    - Future: pre-production, post production (analysis) workflows
    - Integrate I/O efficient tools via common API
    - Future: Allow analysis pipeline where data is
    - Simple to use data movement tools
  - Enabling data understanding
    - Framework for use of multiple techniques – analysis pipeline
    - Parallel statistics tools, specialized for several application domains
    - Use if indexing in query-based visualization

Summary Remarks (2)

- SDM center spends much effort on
  - Sustainability and usability
    - Working with vendors on I/O and file systems– Cray, IBM
    - Working with data centers – ANL, ORNL, NERSC
    - Packaging and releasing open source products – PVFS, ROMIO, pNetCDF, FastBit, pR, Kepler, …
    - SDM center developed or enhanced many products that are in use today
    - Current SDM projects also looking to next generation of systems and applications - active storage, pNFS, architectures, I/O forwarding and aggregation, asynchronous I/O, parallel analysis tools, extendable arrays, …
  - Establishing contacts with scientists
    - Successfully collaborated with scientist from various disciplines: Fusion, Combustion, Astrophysics, groundwater, biology, climate, material science, …
    - Collaboration with other centers/institutes: Vacet (query-based Vis), PDSI (APIs for generic file systems), IUSV (efficient I/O for vis), ESG (SRM), OSG (SRM), CEDPS (SRM log analysis), PERI (through Dashboard).
    - Holding tutorials at SC and other conferences: PVFS, ROMIO, pNetCDF, Kepler, Sapphire, …
    - Educating students at: UCD, NCSU, NWU, Utah; postdocs at LBNL, ORNL, PNNL
  - Future
    - Focus on scaling, robustness, ease of use
    - Engaging additional scientists and applications, based on current successes
    - Identify problems based on above activities, and perform needed research

The END

Extra slides

Scientific Workflow Requirements

- Unique requirements of scientific WFs
  - Moving large volumes between modules
    - Tightlly-coupled efficient data movement
  - Specification of granularity-based iteration
    - e.g. In spatio-temporal simulations – a time step is a “granule”
  - Support for data transformation
    - complex data types (including file formats, e.g. netCDF, HDF)
  - Dynamic steering of workflow by user
    - Dynamic user examination of results
- Developed a working scientific work flow system
  - Automatic microarray analysis
  - Using web-wrapping tools developed by the center
  - Using Kepler WF engine
  - Kepler is an adaptation of the UC Berkeley tool, Ptolemy