

Breakout 1: Facilities Access and Collaboratories

Application – Magnetic Fusion

Goal: Dramatically improve productivity of the 20 minute experimental pulse cycle for a tokamak (longer time between pulses for ITER and much more data). What today is done overnight or next week, do instead between pulses thereby enabling the scientist to make more intelligent setup choices for the next pulse.

Problem: Collecting pulse data, analyzing data, comparing experiment to simulations and allow remote people to provide feedback to experiment within the 20 min between pulse time. (30 pulses/day in an 8 hour day)

Support for collaboration

- Collaborative distributed visualization capabilities needed (tiled walls – high pixel count)
- Currently using electronic notebook – want more capabilities
- Remote monitoring of experiments/ final control is in control room
- Access grid for remote input to the control room and distributed working meetings (e.g. interactive code debugging).

Authentication and Authorization

- Site security versus available security mechanisms are a difficult issue
- Experimental repositories have traditionally been open with peer pressure ensuring appropriate use of data

Distributed storage

- MDSplus data repository has provided critical capabilities but has limitations when moving to real time (e.g., accessing recent data while new data is being added)
- Simulation data currently tends to be near the compute facilities
- Experimental data moves around and needs to go to many sites very quickly, as it is produced
- Vision – simulation data and experimental data both flow so that they can be migrated toward each other and compared

Data-grid issues

- Fast/immediate bulk data transfer (MDSplus client/server not files) is important
- Complex simulation codes are hard to maintain and keep running so don't really want to migrate between compute facilities – create node on the Grid.
- Need remote distributed computing capabilities
 - Resource discovery – not an issue
 - Distributed job management – not an issue -

Breakout 1: Facilities Access and Collaboratories

Application – Macro-Molecular Crystallography

Goal: Determine molecular structure of proteins

Problem: Collect images of a sample (10-80MB/image at the rate of 1-10/sec, 1000 images/data set peak, currently managing about 20% of this). On-line data collection could benefit from immediate feedback.

Support for collaboration

- Currently use secure shell – asynchronous through the data and then feedback. Some use of remote access collaboration now.
- Want –
 - Remote control – interacting at all decision steps (needs strong security, real-time guarantees on data)
 - Remote monitoring with local control with the remote people helping to set policy. Supervisory role for the remote sites
- Usually one group – and a small number of sites collaborating at any one time
- Tools – remote instrument access, access to files, videoconferencing, local compute resources

Authentication and Authorization

- Need easy to use but with strong protection of the data (drug companies)
- Single sign-on
- Migrating user community to different facilities, so need uniformity of access methods and identification

Distributed storage

- bulk file migration

Other issues – data rate will continue to increase with improvements to detectors

Application – Chemical Sciences/Nano-science

Goal: Develop predictive understanding of combustion and other reactive flow type applications.

Problem: Complex problem that combines experimental and simulated data

Support for collaboration

- Multi-disciplinary collaboration
- Data sharing and document sharing are the primary modes of collaboration
- Long-term goal is to get tighter coupling of collaborations than currently available from the collaboration meetings that are 1-2/ year
- Data uniformity and interoperability is a primary problem
- Data mining
- Documentation of community knowledge (annotation)

Breakout 1: Facilities Access and Collaboratories

Support for compute capabilities –

- Currently compute bound
- Large simulations/computational demand
- When computing on 20TF system, then data files will be 35TB

Authentication and Authorization

- Digital signatures/archive for verification of pedigree (including annotations)
- Common user identification/account access so users can migrate
- Security has been a barrier, it is too hard to use and scientists have not been willing to put up with it. (Any extra steps become a barrier.)
- Site policy/infrastructure is sometimes the problem

Distributed storage

- Currently storage at the site of the end users
- Still working on figuring out where the data should reside when it reaches terabyte sizes.
- Will likely be moving to more distributed storage
- Large data sets have limited life-times since they get replaced with better data

Data-grid issues

- Information infrastructure - Meta data is a critical component
- Resource discovery – very important capability – allow broader access

Application – High Energy Physics

Goal: Understand the fundamental foundations of matter (building blocks, particles, forces, theory of everything (standard model), why we exist, how universe formed ...).

Problem: Worldwide collaborations with strong political issues (distributed funding, national pride etc.), no single site has storage or computing to do all storage or computing.

Support for collaboration

- Need easier to use videoconferencing technologies for meetings. Collaboration meetings are big effort, but today's solutions are low tech, wireless, phone bridges, VRVS with PPT/www, H.323 ISDN interactive video, higher tech limited by ease of use, reliability etc. The low tech solutions have been more reliable so they are being used more.
- Want to be able to distribute 1GB in a few seconds
- Real time experiment control (today read only. Linear Collider would like to do control also).

Authentication and Authorization

- Cross-site access mechanisms
- Need certificates to allow access to resources (computing, data).
- Need recognition of keys between very different communities/countries for users (professors, students) around world (some in DoE proscribed countries).

Breakout 1: Facilities Access and Collaboratories

Distributed storage –

- Data grids currently proposed focus as solution for data access, job placement.
- Support for the distributed tier centers and ability to work closely with all International collaborators
 - Place HENP data and jobs at multiple sites with minimum delay.
 - "Raw" data sources limited number of sites (accelerators, major computing centers for MC). Analyzed data come from a wider diversity of sources.
 - Requires transparent data replication (resource discovery, optimum selection, caching ...) of raw and analyzed data at multiple sites in a hierarchical fashion.
 - Sites are worldwide, high performance (hundreds of Mbits/s today between Europe, Japan, & US) lower speed elsewhere (Russia, China, S. America, Caucasus, S.E. Europe, India, Pakistan in particular).
- Large data replications need TByte/day today, growing factor of 2/year.
 - Large data replications could be a scavenger service to avoid scheduling issues.
 - Small data set access (e.g. GB or less today) by individual physicist for interactive analysis (few seconds).
- Need faster file systems to keep up with the network

Other

- Community uses multiple networks (multiple service providers) in multiple countries.

Application – Astrophysics

Problem: Simulate supernova collapse

Support for collaboration

- Collaborative replay of visualization output of simulations – shared simultaneous visualization of multiple parameters for the same run.
- Workflow for processing and analysis of experimental data with feedback to the next night's observation. Distributed collaboration on the workflow.

Distributed storage

- Want to be able to in the future do a 3D simulation run is a 50TB file (currently store at the compute facilities)

Application – Global Accelerator Network

Goal: Collaborative operation/control of an accelerator. Accelerator that is not only built collaboratively but also run collaboratively.

Support for collaboration

- Videoconferencing/access grid
- Accelerator physics studies will also require interaction with simulations
- Remote monitoring and control of instruments

Breakout 1: Facilities Access and Collaboratories

- Ability to tie in simulations
- Authentication and Authorization
- Very important issue particularly due to safety concerns