XSEDE: a range of opportunities for students

John Towns
PI and Project Director, XSEDE
jtowns@ncsa.illinois.edu

XSEDE

Extreme Science and Engineering Discovery Environment

A bit about me…

- PI and Project Director of XSEDE
  - one of the leading projects globally in this space
- Director, Collaborative Cyberinfrastructure Program Office at NCSA
  - one of the premier computing centers in the world
- Co-Founder, Illinois Campus Cluster Program
  - filling the mid-range computing gap
- Inaugural Board of Directors, Compute Canada

How I got here…

- Someone who has taken their own path through life…
- Just a mid-western boy from Missouri
  - I don't see a lot of folks from Missouri in the circles I am in
- Always had an interest in science and technology
  - but I didn't know it was called "technology" back then
- Never really quite fit with my classmates
  - fortunately, I didn't care too much and my parents were always supportive
- Failed physicist
- Failed computational scientist
- I have found my way though...

XSEDE Vision

The eXtreme Science and Engineering Discovery Environment (XSEDE):

- enhances the productivity of scientists and engineers by providing them with new and innovative capabilities

and thus

- facilitates scientific discovery while enabling transformational science/engineering and innovative educational programs

XSEDE's Objectives

- Understand the cyberinfrastructure requirements of the science and engineering research and education community
- Create a cyberinfrastructure ecosystem to facilitate improved researcher productivity
- Provide a unique user friendly interface to the resources and services accessible via XSEDE
- Maximize researcher productivity

Motivation for XSEDE

- Scientific advancement across multiple disciplines requires a *variety of resources and services* and thus the availability of *comprehensive cyberinfrastructure* composed of *heterogeneous digital resources*.

- High-end computational science is better served if these capabilities leverage the aggregate expertise of a *small number of leading institutions* rather than being fully centralized at a single institution or being fully decentralized.

XSEDE Project Factoids: high order bits

- 5 year, 130M project
  - option for additional 5 years of funding upon major review after PY3
- No funding for major hardware
  - coordination, support and creating a national/international cyberinfrastructure
- ~120 FTE funded across 17 partner institutions
  - ~240 individuals funded
  - many more involved!!

What do you mean by “Advanced Digital Services?”

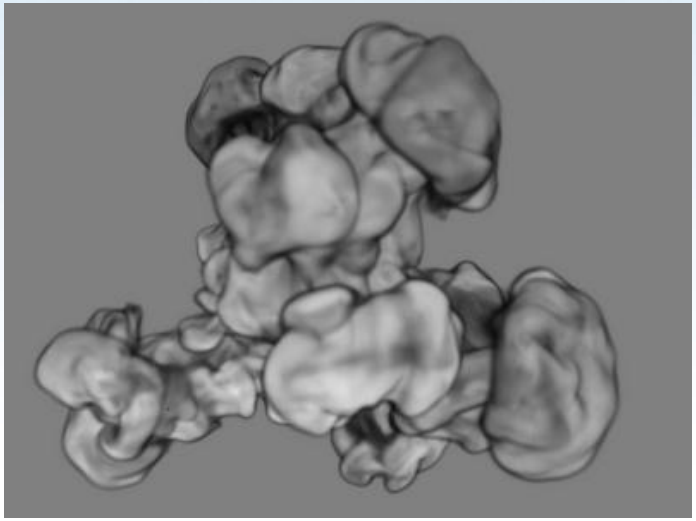
- Often use the terms “resources” and “services”
  - these should be interpreted very broadly
  - most are likely not operated by XSEDE
- Examples of resources
  - compute engines: HPC, HTC (high throughput computing), campus, departmental, research group, project, …
  - data: simulation output, input files, instrument data, repositories, public databases, private databases, …
  - instruments: telescopes, beam lines, sensor nets, shake tables, microscopes, …
  - infrastructure: local networks, wide-area networks, …
- Examples of services
  - collaboration: wikis, forums, telepresence, …
  - data: data transport, data management, sharing, curation, provenance, …
  - access/used: authentication, authorization, accounting, …
  - coordination: meta-queuing, …
  - support: helpdesk, consulting, ECSS, training, …
  - And many more: education, outreach, community building, …

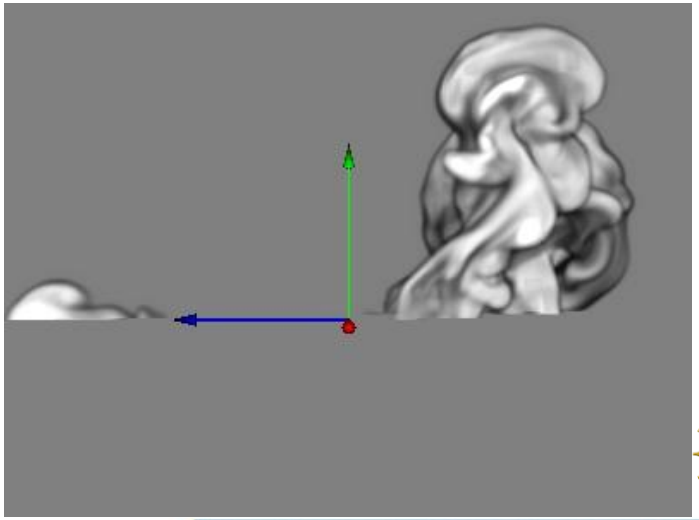
XSEDE Visualization and Data Analysis Activities

- Longhorn @ TACC
  - Dell XD Visualization Cluster, 2048 compute cores, 14.5TB memory, 512 GPUs
  - Longhorn Visualization Portal
    - internet gateway to the Longhorn
- Nautilus @ Univ of Tenn/RDAV
  - SGI Altix UV 1000, 1024 cores in 4 TB shared memory

Supernovae Visualizations

- Working with D. Townsley, University of Alabama, XSEDE ECSS Staff modified a custom volume renderer to represent both density and color using different variables from a single data-set (top). Also experimented with cut-away views of volume representation (bottom).

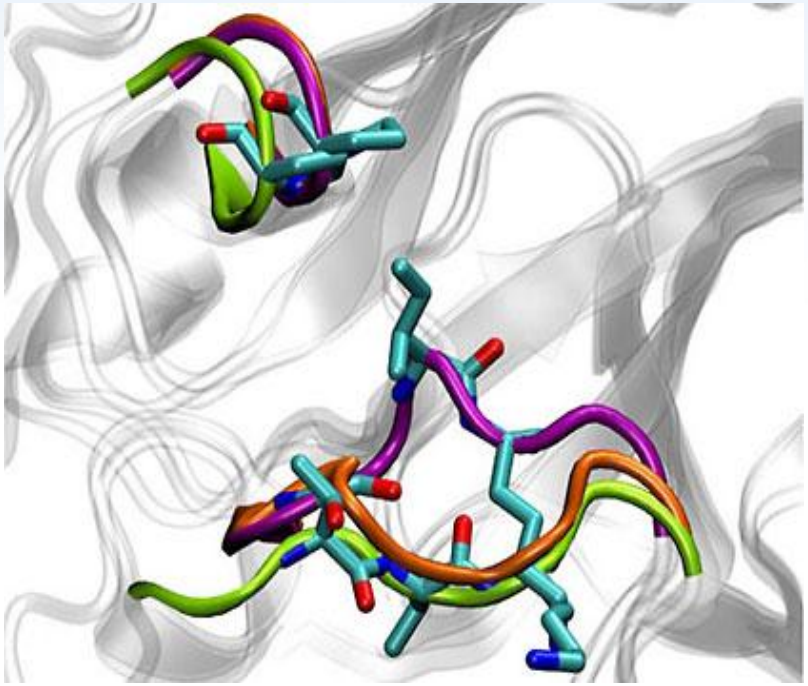




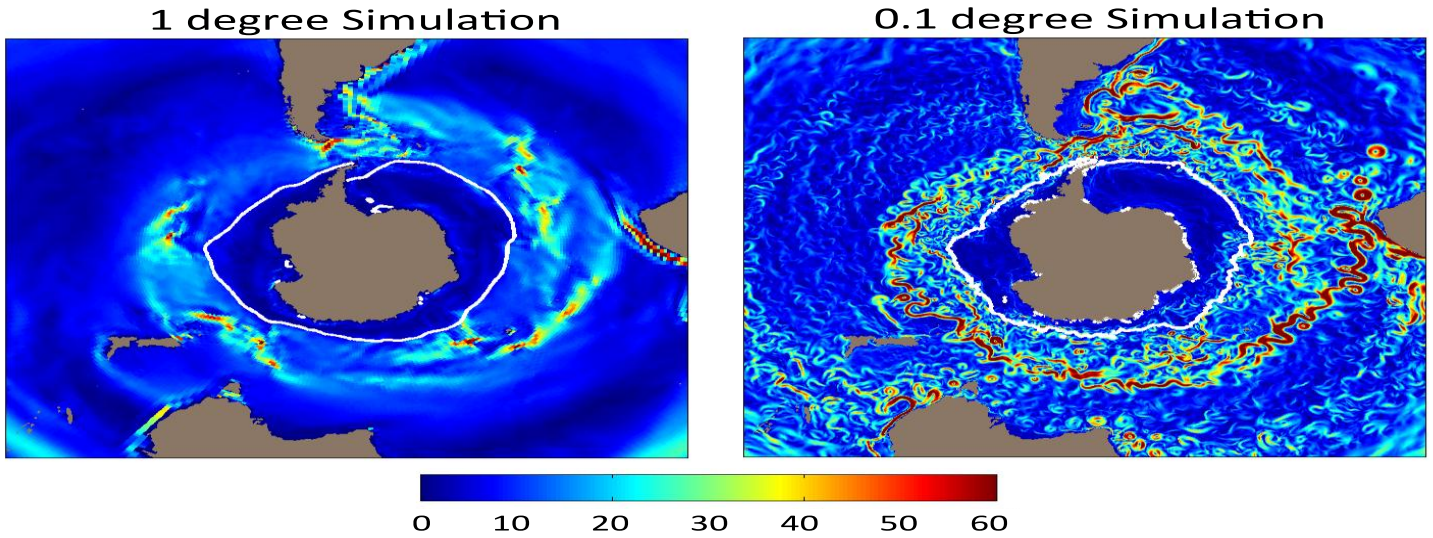
Customized flu strain therapy

Rommie Amaro, California-Irvine

The 150- and 430-loop structures are shown for 09N1 crystal structure (purple), 09N1 second-most dominant molecular dynamics (MD) cluster representative structure (green backbone), and VN04N1 crystal structure (orange), indicating that the pandemic N1 adopts an open 150-loop conformation. Gly147, Ile149, Lys150 and Pro431 are shown in stick representation. This simulation was conducted on SDSC's Trestles supercomputer.



Ozone effects on the Antarctic Ice Sheet



Cecilia Bitz, Univ of Washington. 10-km simulations of the Antarctic ice sheet using the Community Earth Systems Model (CESM) to determine if the Antarctic ozone depletion is causing sea ice to expand. Plotting surface temperature (degrees Celsius) response to depleting ozone over the second half of the 20th century. The low resolution response (left) is about twice as strong as high resolution (right).

XSEDE Data Activities

- Data Sharing and Data Management
  - critical capabilities to support research
- XSEDE-Wide File System (XWFS)
  - high performance wide area file system
- Global Federated File System (GFFS)
  - distributed file system sharing
- Globus Transfer (Globus Online)
  - managed data transfer

XWFS

- High performance wide area file system
  - focus on high end resources available via XSEDE
    - Track 2 and similar resources systems
      - HPC, data intensive, visualization
  - limited number of end points
  - high performance network connections via XSEDEnet
- Provide single file system spanning resources
  - ease file movement/sharing among high end resources
  - support workflows using multiple resources
- Technical evaluation conducted during first year of XSEDE
  - developing implementation based on IBM's Global Parallel File System (GPFS)
- Many challenges yet to be addressed

GFFS: accessing remote data via a file system mount

- Global directory structure mapped directly into the local operating system via FUSE mount
- Access XSEDE resources via the file system regardless of location
  - files and directories can be accessed by programs and shell scripts as if they were local files
  - jobs can be started by copying job descriptions into directories
  - one can see the jobs running or queued by doing an “ls”.
  - one can “cd” into a running job and access the working directory where the job is running directly
- Currently being deployed as part of XSEDE’s Campus Bridging activities
  - many challenges anticipated as we scale this to a level never attempted before

Three Use Cases Illustrate GFFS

- Accessing data at an NSF center from a home or campus
  - export directory at NSF center that you want to access
  - FUSE mount the XSEDE GFFS into your local file system
  - Create, Read, Update, and Delete files at the center from home
- Accessing data on a campus machine from an NSF center
  - export directory on campus file server into the GFFS
  - FUSE mount the GFFS on the login node at the center, or specify state-in/stage out in a job description
  - Create, Read, Update, and Delete files at home from the center
- Directly sharing data with a collaborator at another institution
  - export directory on campus file server into the GFFS
  - give your collaborator desired level of access (RWX)
  - collaborator FUSE mounts the GFFS their desktop
  - share files

Globus Transfer: Moving Big Data *Easily*

- The Challenge:
  - what should be trivial…

*"I need my data **over there** – at my _____" (supercomputing center, campus server, etc.)*

Data Source → **Data Destination**

  - can be painfully tedious and time-consuming

"GAAAH! %&@#&"

Data Source

! Config issues

! Firewall issues

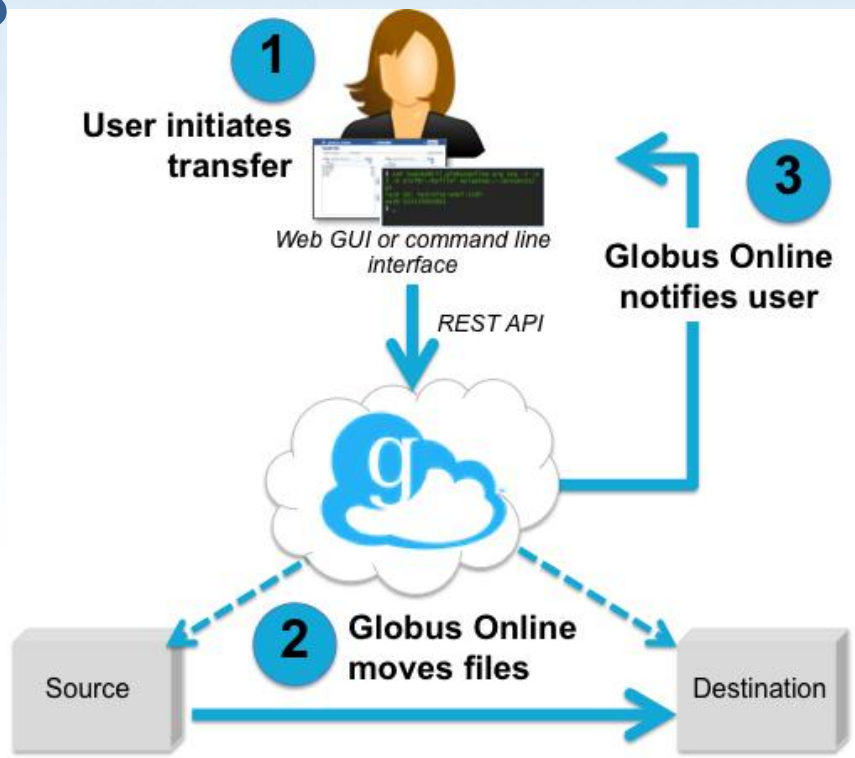
! Unexpected failure = manual retry

Data Destination

Adapted from slides from Steve Tuecke, UChicago

What is Globus Online?

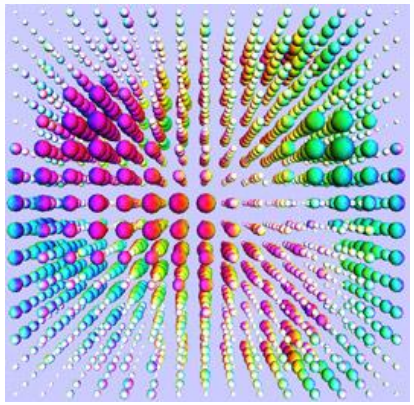
- Reliable file transfer.
  - Easy “fire-and-forget” transfers
  - Automatic fault recovery
  - High performance
  - Across multiple security domains
- No IT required.
  - Software as a Service (SaaS)
    - No client software installation
    - New features automatically available
  - Consolidated support & troubleshooting
  - Works with existing GridFTP servers
  - Globus Connect solves “last mile problem”
- Supports XSEDE, NERSC, ESnet, UCLA, U Michigan, U Miami, U Colorado, U Washington, Indiana U, …



Adapted from slides from Steve Tuecke, UChicago

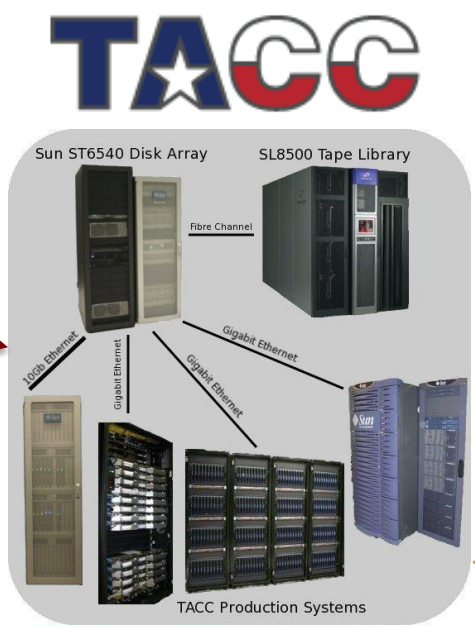
Case Study: Lattice QCD

- **Fast**: Reduced transfer times
- **Easy**: Fire-and-forget transfers
  - Automated retry
  - No file pre-staging
  - No complex infrastructure
  - Convenient CLI and GUI interfaces



"Globus Online frees up my time to do more creative work than typing scp commands or devising scripts to initiate and monitor progress to move many files."

Indiana University researcher moved ~6 TB from Oak Ridge to TACC in 2 days



"I moved 100 7.3 GB files tonight in about 1.5 hours. I am very impressed. I also like the new commands and help system."

Adapted from slides from Steve Tuecke, UChicago

OK… that's scary… and why do I care?

- Cyberinfrastructure is in transition
  - many years of experimentation and development
  - lots of lessons learned and very interesting initial implementations
  - working toward the creation of a production distributed environment
- Production services supporting a range of activities uncover many new challenges
  - can't just support the niche communities
  - much greater expectations of reliability, scalability, performance…
- XSEDE and the community need smart young minds to help solve this new generation of problems

XSEDE Scholars Program

- Learn more about high performance computing and XSEDE resources
- Network with cutting-edge researchers and professional leaders
- Belong to a cohort of student peers to establish a community of academic leaders
- Particular focus on African Americans, Hispanics, Native Americans, and women
- Will be accepting applications for the 2013-2014 cohort in the spring of 2013.
  - https://www.xsede.org/xsede-scholars-program

Student Programs for XSEDExy and OSG Summer Schools

- Annual program to support student participation
  - XSEDE annual conference
  - OSG Summer Schools
- XSEDE'13 opportunities
  - papers and posters
  - student programming competition
  - student program involvement and support
    - ~90 grads, undergrads and high school students
  - https://www.xsede.org/web/xsede13/students

XSEDE-PRACE Summer Schools:
HPC Challenges in Computational Sciences

- Annual joint summer school between XSEDE and PRACE
  - PRACE: Partnership for Advanced Computing in Europe
    - http://www.prace-ri.eu/
- Focus of Summer Schools
  - targeted at graduate students and postdocs
  - gain greater knowledge about HPC and its applications in multiple fields of science and engineering
  - foster new collegial friendships and partnerships among the international presenters and attendees
- Planning for 2013 event under way
  - will support ~60 attendees
  - 2012 event:
    - https://www.xsede.org/web/summerschool12/

Student Campus Champions
Under Development

- Offshoot of Campus Champions Program
- Main responsibilities of student campus champion
  - provide outreach on campus
  - assist users with accessing XSEDE resources
  - assist with user training on campus
  - attend annual meeting
  - to give and learn from the program
- Benefits to the students
  - create credibility
  - enhance leadership and entrepreneurship qualities

Training Opportunities

- XSEDE offers regular training opportunities
  - in-person and distributed events
  - https://www.xsede.org/web/xup/course-calendar

- Extensive online training offerings
  - https://www.xsede.org/web/xup/online-training

Leading toward careers…

- There are many paths to many opportunities
  - faculty position path turned out not to be for me
- Our community needs many different types of people
  - faulty researchers
    - architectures, networks, languages, compilers, algorithms, distributed systems, visualization, data…
  - technical staff
    - developers, performance engineers, system administrators, network engineers, storage engineers, visualization, data management, data analysis…
  - leaders and administrators
    - visionaries, program managers, project managers,…