XSEDE Resources and Science Highlights

Jay Alameda
alameda@illinois.edu
Outline

• XSEDE Resources and Usage Breakdown
• Extended Collaborative Support Services
• XSEDE and Science Gateways
XSEDE Resources and Usage Breakdown

XSEDE
Extreme Science and Engineering Discovery Environment
XSEDE offers huge variety of resources

- Leading-edge distributed memory systems
- Very large shared memory systems
- High throughput systems
- Visualization engines
- Data analytics systems
- Accelerators and co-processors

Many scientific problems have components that call for use of more than one architecture.
XSEDE Usage Breakdown CY2015 by NSF Division

Total XD SUs Charged by NSF Directorate

- Mathematical and Physical Sciences: 233,940,866.0
- Engineering: 65,869,544.0
- Biological Sciences: 180,646,245.0
- Geosciences: 35,206,512.0
- Social, Behavioral, and Economic Sciences: 340,193.0
- Other: 1,576,237.0
- Humanities/Arts: 544.0

2015-11-01 to 2015-11-30 Src: XDCDB. Powered by XDMoD/Highcharts
XSEDE Usage Breakdown CY2015 by Field of Science

Total XD Sus Charged by Field of Science

- Physics: 15,292,602.0
- Biophysics: 74,846,751.0
- Materials Research: 53,314,228.0
- Gravitational Physics: 42,087,493.0
- Molecular Biosciences: 40,288,056.0
- Biochemistry and Molecular Structure and Function: 39,910,399.0
- Astronomical Sciences: 39,764,209.0
- Fluid, Particulate, and Hydraulic Systems: 23,567,175.0
- Elementary Particle Physics: 20,974,800.0
- Physical Chemistry: 16,347,480.0

2015-11-01 to 2015-11-30 Src: XDCDB. Powered by XDMoD/Highcharts
XSEDE Extended Collaborative Support

XSEDE
Extreme Science and Engineering Discovery Environment
Extended Collaborative Support Services
Strategic Objective - To Help Users Make More Productive Use of XSEDE Resources

• Support people who understand the discipline as well as the systems (perhaps more than one support person working with a project).

• 37 FTEs, spread over ~80 people at almost a dozen sites.
How do you get Extended Collaborative Support?

• You have to ask for it - it’s an allocated resource.

• You can always ask for it, i.e. midstream or even as part of a startup request.

• Lasts up to a year.

• Must have specific goal in mind - can’t just say I want additional programming support.
ECSS has 5 support services

• Support for
  – Research Teams
  – Novel and Innovative Projects
  – Community Capabilities
  – Gateways
  – Training and Outreach
Support for Research Teams
(led by Lonnie Crosby, NICS (U. of Tennessee))

• Optimization
  – Profiling
  – Scaling to higher core count
  – Improving IO
  – Porting to GPUs
  – Finding better solvers (what’s better often depends on the degree of parallelization)

• Visualization

• Workflows
Selected Highlight (1)

- **Large eddy simulations of extended wind farms, and direct numerical simulations of turbulent channel**
  - PI: Meneveau, Johns Hopkins
  - ECSS Project Team: Darren Adams and David Bock (NCSA)
- Interaction between large wind farms with multiple wind-turbines and the atmospheric boundary layer flow
- The ECSS goals were to provide assistance implementing two dimensional domain decomposition and parallel FFT as well as 3D visualization representations to assist in simulations and analysis of large wind farms.
- The work plan was modified part way through to abandon an approach that included developing a stand-alone MPI transpose and custom 2D FFT. Instead, the team modified the existing P3DFFT library to completely skip the FFT and transpose operations on the 3rd array index.

Performance of the 2D FFT in calculation the x and y derivatives. This was obtained from runs on the NICS Kraken system. The figure presents scaling of a 1,024³ mesh. The blue points are from the original 2D FFT implemented with 1D domain decomposition. The orange points are the new 2D MPI FFT scaled out to 8,192 processors.
Selected Highlight (2)

- Visualization representations
  - designing and developing a myriad of different techniques visualizing the changing vector field over time. Work focused specifically on various particle advection, tracing, and streamline development of high quality scenes integrating new particle visualization methods developed above with earlier existing volume rendering techniques to show multiple variables in a single representation.
FRED- Epidemiology Modeling Code

- Agent-based modeling system that uses synthetic populations derived from census data to capture geographic and demographic distributions.
- Recently used in real time by the Dept. of Health and Human Services in managing the H1N1 outbreak.
- Usually applied to a region. When scaled to the entire US, the serial implementation required approximately 96 hours and 540 GB of memory to complete a 100-day simulation.
- An ECSS consultant reduced this to 3-4 hours using less than 200 GB of memory.

15% serologic attack rate, no mitigation
Vaccinate a month before the peak
Support for Novel and Innovative Projects
(led by Sergiu Sanielevici, Pittsburgh Supercomputing Center)

• Pro-actively reaching out to communities new to advanced computing, e.g.
  – social science
  – social network analysis
  – language processing
  – genomics
  – digital humanities
  – economics
  – library science
  – public health

• Look for pioneers, work with them to craft the project, and then help them begin to execute it.
Large Genome Assemblies

- ECSS Staff working with leading researchers and code developers
- Largest ever metagenome assembly, using 3.5 TB RAM on PSC Blacklight

“I wouldn’t have been able to do anything on Blacklight without ECSS staff… (consultant) took a real interest and solved a lot of things that were hard for me. He found bugs in the software and got them resolved with the software authors. I’d worked for months and not made that progress. Without his expertise, I might have given up…”
Support for Community Codes
(led by John Cazes, Texas Advanced Computing Center)

Community codes – applications, tools, and libraries used by multiple research groups

• Deploying, hardening, and optimizing useful software systems
• Assisting users with community codes and tools
• Establishing relationships between XSEDE and developer communities
• Extending XSEDE documentation to cover community code implementations
Selected Highlight

• **Collaboration with the Broad Institute: Genomics Community Capabilities**

• Internal project to provide support to over 15 research groups

• Goal: Port and optimize the widely used genomics applications, Trinity and ALLPATHS-LG, to Blacklight and other XSEDE resources

• Challenges: Neither was designed for a parallel architecture

• Solution:
  – Adjusted the workflow to work with parallel filesystems
  – Collaborated with the Broad Institute and the National Center for Genome Analysis to include parallel optimizations in the code base

• Success: Enabled the largest Trinity job ever
  – 20 assemblies with 1.5 billion reads each
Trinity Pipeline

Inchworm
64 cores ~100 hours

Chrysalis
128 cores ~400 hours

Butterfly
64 cores ~50 hours
Support for Science Gateways
(led by Marlon Pierce, Indiana U)

• Science Gateways enable communities of users associated with a common discipline to use computational resources through a familiar and simpler interface e.g. a web interface.

• Examples of ECSS Science Gateway Support
  – Assist with job submission and data movement to XSEDE
  – Automation of scientific processes through workflows
  – Integration of XSEDE resources into a portal/gateway
  – Support with grid security and community accounts

• More details to come -
Support for Training, Education, and Outreach
(led by Jay Alameda, NCSA)

• Collaboration with XSEDE Training, Education, and Outreach Services

• Training support
  – On-line course development and update
  – Synchronous training (live, web, ...)
  – Collaboratively identify new training areas

• Education
  – Online course support

• Outreach
  – Conferences, speakers bureau, campus visits
Education: Applications of Parallel Computers

• Adapted from Jim Demmel’s CS267 Course
  – Video, Slides in Cornell’s Virtual Workshop
  – Quizzes in every class

• 3 Homework (programming) problems
  – ECSS staff helped port to XSEDE resources
  – ECSS staff part of support team for students
  – Autograding by Berkeley team

• Capped enrollment at 300 students
  – Limit reached in < 3 days

• Evolved into a Small Private Online Course (SPOC), in collaboration with a number of campuses.
XSEDE and Science Gateways
Using XSEDE Resources Through Your Web Browser

XSEDE
Extreme Science and Engineering Discovery Environment
What Is a Science Gateway?
Science Gateways

- Science gateways provide Web-based access to XSEDE.
  - Help researchers and students run well-known scientific application codes.
  - Coming: provide access to online data sets
- Gateways are community provided resources
- Gateways are domain specific
  - Chemistry, phylogenetics, neuroscience, biophysics, …
What Are Some Example Science Gateways?
Some Prominent XSEDE Gateways

- **CIPRES**: phylogenetic research.
  - Prominent applications: RAxML, BEAST, MrBayes

- **SEAGrid**: computational chemistry, material science, and engineering
  - Prominent applications: Gaussian, LAMMPS, NWChem

- **Neuroscience Gateway**: computational neuroscience.
  - Prominent applications: NEURON, Freesurfer, Brian

- **CyberGIS**: geospatial research
  - Prominent applications: TauDEM, pPySAL, pRasterBlaster

Full listing: https://www.xsede.org/gateways-listing
Do People Use Science Gateways?
More users access supercomputers via gateways than from the command line in 2014.
Gateways Submit over 20,000 jobs/month to XSEDE supercomputers.
CIPRES: Supporting Research and Education

- Most popular science gateway in XSEDE
  - ~40% of all XSEDE users
- In use on 6 continents
- Cited in major journals (Cell, Nature, PNAS)
- Used at major research institutions (Stanford, Harvard, Yale)
- Used by ~76 researchers for curriculum delivery
- Supports hundreds of publications every year (1570 to date)
- Used in 80% of EPSCoR states
- Used by a 15-year-old high school student who won the Massachusetts state science fair with no support from SDSC staff
How Do I Get Started? Can I Get Help? Can I Build My Own Gateway?
More Information

• Contact XSEDE Science Gateways Program Manager
  – Marlon Pierce, marpierc@iu.edu
  – Marlon loves to get email about gateways.

• XSEDE Gateways Overview: https://www.xsede.org/gateways-overview

• Community Mailing List: gateways@xsede.org
  – Send “subscribe gateways” to majordomo@xsede.org
Thanks for listening and welcome to XSEDE!