
Marlon Pierce
Science Gateway and Workflow Program Lead

XSEDE
Extreme Science and Engineering Discovery Environment
XSEDE Is...

Infrastructure at your Service

Allocatable Brains


http://www.geneticsandsociety.org/img/original/Brain1.jpg
SDSC Comet: Long Tail Computing

• Modest core counts: full bisection bandwidth up to Comet island (1,728 cores)
• 128 GB DRAM/node (5.3 GB/core): single node shared memory apps and MPI codes with large per-process memory footprint
• AVX2: Codes with vectorizable loops. Any application with significant performance gain relative to Sandy Bridge or Ivy Bridge (AVX)
• SSDs: Computational chemistry, finite elements. Apps that generate large numbers of small temporary files (finance, QM/MM)
• GPU nodes: Molecular dynamics, linear algebra, image and signal processing.
• Large memory nodes: de novo genome assembly, visualization of large data sets, other large memory apps
• Science Gateways: Gateway-friendly environment with local gateway hosting capability, flexible allocations, scheduling policies for rapid throughput, heterogeneous workflows, and virtual clusters for software environment
• High performance virtualization: workloads with customized software stacks, especially those that are difficult to port or deploy in standard XSEDE environment
TACC Wrangler: Data Intensive Computing

- Many different ways of working with Wrangler
  - Data management via Globus Publication and iRODS
  - Data analysis systems or Data Extraction, Translation, and Loading
  - Databases (SQL and noSQL, PostGIS, GraphDBs)
  - Hadoop/Spark based processing with Flash based HDFS backend
  - Out of core computations on the Flash storage
Bridging to Nontraditional HPC Users and Enabling HPC + Big Data Workflows

Leveraging PSC’s expertise with shared memory, *Bridges* will feature 3 tiers of large, coherent shared-memory nodes – 12TB, 3TB, and 128GB – to support a uniquely flexible and user-friendly environment:

- **Interactivity** is the feature most frequently requested by nontraditional HPC communities and for doing data analytics and testing hypotheses.

- **Gateways and tools for gateway building** will provide easy-to-use access to *Bridges’* HPC and data resources, reaching large numbers of users who aren’t programmers.

- **Database and web server nodes** will provide persistent NoSQL and relational databases to enable data management, workflows, and distributed applications.

- **High-productivity programming languages & environments** (R, Python, MATLAB, Java, Hadoop, etc.) will let users scale familiar applications and workflows.

- **Virtualization** will allow users to bring their particular environments for portability, reproducibility, and security and provide interoperability with clouds.

- **Campus bridging** will streamline interoperation with campus resources and enable burst offload capability through a pilot project with Temple University.

Interest from new communities is already very high: examples include the digital humanities, machine learning, statistics, genomics, and radio astronomy.
- Geographically Distributed OpenStack-based Cloud, 0.5 PetaFLOPS
- High-speed connections to Internet2 and local connections to Wrangler disk storage at IU and TACC
- iPlant Atmosphere-based API and User Environment.
- Globus-based large scale file movement
- Quarry Gateway VM Hosting available today!
### Some XSEDE Support Areas

<table>
<thead>
<tr>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Gateways</td>
<td>Helps building and running community science portals, web services over XSEDE.</td>
</tr>
<tr>
<td>Scientific workflows</td>
<td>Helps using a wide range of general purpose scientific workflow tools.</td>
</tr>
<tr>
<td>Novel, Innovative Projects</td>
<td>Helps new communities that can strongly benefit from the use of XSEDE’s ecosystem of advanced digital services.</td>
</tr>
<tr>
<td>Community codes</td>
<td>Installs and supports community developed applications on XSEDE resources.</td>
</tr>
<tr>
<td>Data analytics applications</td>
<td>Supports XSEDE users developing algorithms utilizing data analytics and mining software and technologies.</td>
</tr>
<tr>
<td>Research Team Support</td>
<td>Optimizes application codes, improves work and data flows, and increases the effective use of XSEDE.</td>
</tr>
</tbody>
</table>
How Can I Help You?

marpierc@iu.edu
Questions

• What kinds of problems do you have that you wish other people would solve?
• What kinds of infrastructure, services and support do you need?
• Based on what you have heard, would you use XSEDE?
• Is XSEDE a good fit for supporting the EarthCube Testbed Activity?
Our reach will forever exceed our grasp, but, in stretching our horizon, we forever improve our world.
CyberGIS Center for Advanced Digital and Spatial Studies

CyberGIS Software Ecosystem
- CyberGIS Gateway and Applications
- CyberGIS Toolkit
- GISolve Middleware

Science Communities
- Advanced cyberinfrastructure
- Agriculture
- Climate change impact assessment
- Emergency management
- Geographic information science
- Geography and spatial sciences
- Geosciences

User Communities
- Biologists
- Geographers
- Geoscientists
- Social scientists
- General public
- Broad GIS users

TauDEM Application on CyberGIS Gateway
Hierarchical workflow with mix of large MPI jobs (top) and large number of serial tasks (post processing)

The size of the workflow depends on the number of sites and sampling frequency

For each of the 200 sites in the input map, generate a hazard curve

Each site has a sub-workflow with 820,000 tasks

SGT (Strain Green Tensor - MPI jobs) output: **15.6 TB** (40 * 400 GB)

Final outputs: **500 million files** (820000/site x 600 sites) **5.8 TB** (600 * 10 GB)
XSEDE Extended Collaborative Support Services

• Get expert staff as well as computer allocations
  – 37 FTEs, spread over ~80 people at almost a dozen sites.
  – Many PhDs or equivalent experience

• Wide range of areas,
  – Performance analysis
  – Petascale optimization techniques
  – Novel and Innovative Projects
  – Building Science Gateways.
  – Supporting Scientific Workflows
XSEDE Offers Variety

• See https://www.xsede.org/resources/overview
• Large, highly scalable parallel computing resources
• Very large shared memory systems
• High throughput systems (OSG)
• Visualization engines
• Accelerators: GPUs, MICs
• Science clouds

XSEDE is not a Cloud, but it is IaaS: Infrastructure at your Service