What are the benefits of using Science Gateways

Amit Majumdar, Division Director
Data Enabled Scientific Computing Division

Wayne Pfeiffer, Distinguished Scientist

San Diego Supercomputer Center
University of California San Diego

SDSC Summer Institute 2016

Outline

- **Science Gateways**
  - Examples

- **Hands-on using a Science Gateway**

Science Gateways

- **An online community space for science and engineering research and education**

- **A web-based resource for accessing data, software, computing services and equipment specific to the needs of a science or engineering discipline**

HPC Resources

- **Available via XSEDE – Extreme Science and Engineering Discovery Environment**
  - NSF funded supercomputers, advanced support, services, allocation, EOT
- **You learned yesterday – how to**
  - Compile codes
  - Launch and manage jobs
  - Manage data on filesystems
  - Use HPC resources effectively
  - etc

Administrative and Technical tasks (barriers?)

- **Write allocation proposals (peer-reviewed) for supercomputer time every year**

- **Understand HPC machines, policies, complex OS/software**

- **Install and benchmark complex applications on HPC resources**

- **Different machines have different schedulers**

- **Understand and manage remote authentication**

- **Figure out data transfer, file systems, storage**

Science gateways

- **Easy web based user interface GUI**
  - Upload input files, models
  - Set application and HPC related parameters
  - Run jobs by the click of a button
- **Scientific applications already installed optimally on HPC resources at the backend**
- **Easily access, download output results**
  - Some provide post processing, viz
- **Some provide RESTful services**
- **Gateway team writes annual allocation proposal**

Catalyzes and democratizes computational science research for researchers and students from all universities, colleges and institutions

Search XSEDE...

HOME   ABOUT   USER SERVICES   RESOURCES   EDUCATION & OUTREACH   GATEWAYS   TECHNOLOGY DATABASE

Science Gateways

A Science Gateway is a community-developed set of tools, applications, and data that are integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community. Gateways enable entire communities of users associated with a common discipline to use national resources through a common interface that is configured for optimal use. Researchers can focus on their scientific goals and less on assembling the cyberinfrastructure they require. Gateways can also foster collaborations and the exchange of ideas among researchers.

How to Turn Your Project into a Science Gateway

1. Get an XSEDE allocation; Start-up and Educational allocations require only a one paragraph project description. For more information visit the Allocations section of the web site.

2. Register your project as an XSEDE Gateway

3. Build a portal

4. Set up your developer accounts by Adding users to an existing allocation. Also, set up your Gateway community accounts.

Using Existing Gateways

Gateways are independent projects, each with its own guidelines for access. Most gateways are available for use by anyone, although they usually target a particular research audience. XSEDE Science Gateways are portals to computational and data services and resources across a wide range of science domains for researchers, engineers, educators, and students. Depending on the needs of the communities, a gateway may provide any of the following features:

Science Gateways

Overview

Gateway Listing

For PIs

For Developers

Gateways Symposium

Gateway Applications

SDSC SAN DIEGO SUPERCOMPUTER CENTER

UNIVERSITY OF CALIFORNIA

XSEDE SGW

TITLE	FIELD OF SCIENCE	PORTAL HOMEPAGE
Diagrid	Advanced Scientific Computing	Visit Portal
UCI Complex Social Science Gateway	Anthropology	Visit Portal
Massive Pulsar Surveys using the Arecibo L-band Feed Array (ALFA)	Astronomical Sciences	Visit Portal
Center for Multiscale Modeling of Atmospheric Processes	Atmospheric Sciences	Visit Portal
Biodrugscore: A portal for customized scoring and ranking of molecules docked to the human proteome	Biochemistry and Molecular Structure and Function	Visit Portal
High-Resolution Modeling of Hydrodynamic Experiments with UltraScan	Biophysics	Visit Portal
ROSIE, The Rosetta Online Server that Includes Everyone	Biophysics	Visit Portal
Chem Compute	Chemistry	Visit Portal
Computational Chemistry Grid (GridChem)	Chemistry	Visit Portal
ParamChem Gateway	Chemistry	Visit Portal
User-Friendly Security Solutions for Grid Environments	Communications and Computational Systems	Visit Portal
OpenTopography	Earth Sciences	Visit Portal
Purdue Environmental Data Portal	Earth Sciences	Visit Portal
WaterHUB - Platform for water education, research, data access, partnership and collaboration	Earth Sciences	Visit Portal
Network for Earthquake Engineering Simulation	Earthquake Hazard Mitigation	Visit Portal
CMS gWMS	Elementary Particle Physics	Visit Portal
Network for Computational Nanotechnology and nanoHUB	Emerging Technologies Initiation	Visit Portal
Globus Online	Engineering Infrastructure Development	Visit Portal
CyberGIS Gateway	Geography and Regional Science	Visit Portal
The Earth System Grid	Global Atmospheric Research	Visit Portal

Network for Computational Nanotechnology and nanoHUB	Emerging Technologies Initiation	Visit Portal
Globus Online	Engineering Infrastructure Development	Visit Portal
CyberGIS Gateway	Geography and Regional Science	Visit Portal
The Earth System Grid	Global Atmospheric Research	Visit Portal
The iPlant Collaborative Agave API	Integrative Biology and Neuroscience	Visit Portal
MP-Complete	Materials Research	Visit Portal
VLab - Virtual Laboratory for Earth and Planetary Materials	Materials Research	Visit Portal
NIST Digital Repository of Mathematical Formulae	Mathematical Sciences	Visit Portal
Galaxy	Molecular Biosciences	Visit Portal
Integrated database and search engine for systems biology (IntegromeDB)	Molecular Biosciences	Visit Portal
ROBETTA: Automated Prediction of Protein Structure and Interactions	Molecular Biosciences	Visit Portal
Providing a Neuroscience Gateway	Neuroscience Biology	Visit Portal
SCEC Earthworks Project	Seismology	Visit Portal
Asteroseismic Modeling Portal	Stellar Astronomy and Astrophysics	Visit Portal
CIPRES Portal for inference of large phylogenetic trees	Systematic and Population Biology	Visit Portal
Computational Anatomy	Visualization, Graphics, and Image Processing	Visit Portal

Gateway users surpass login users in 2013
Automated user-counting in 2015
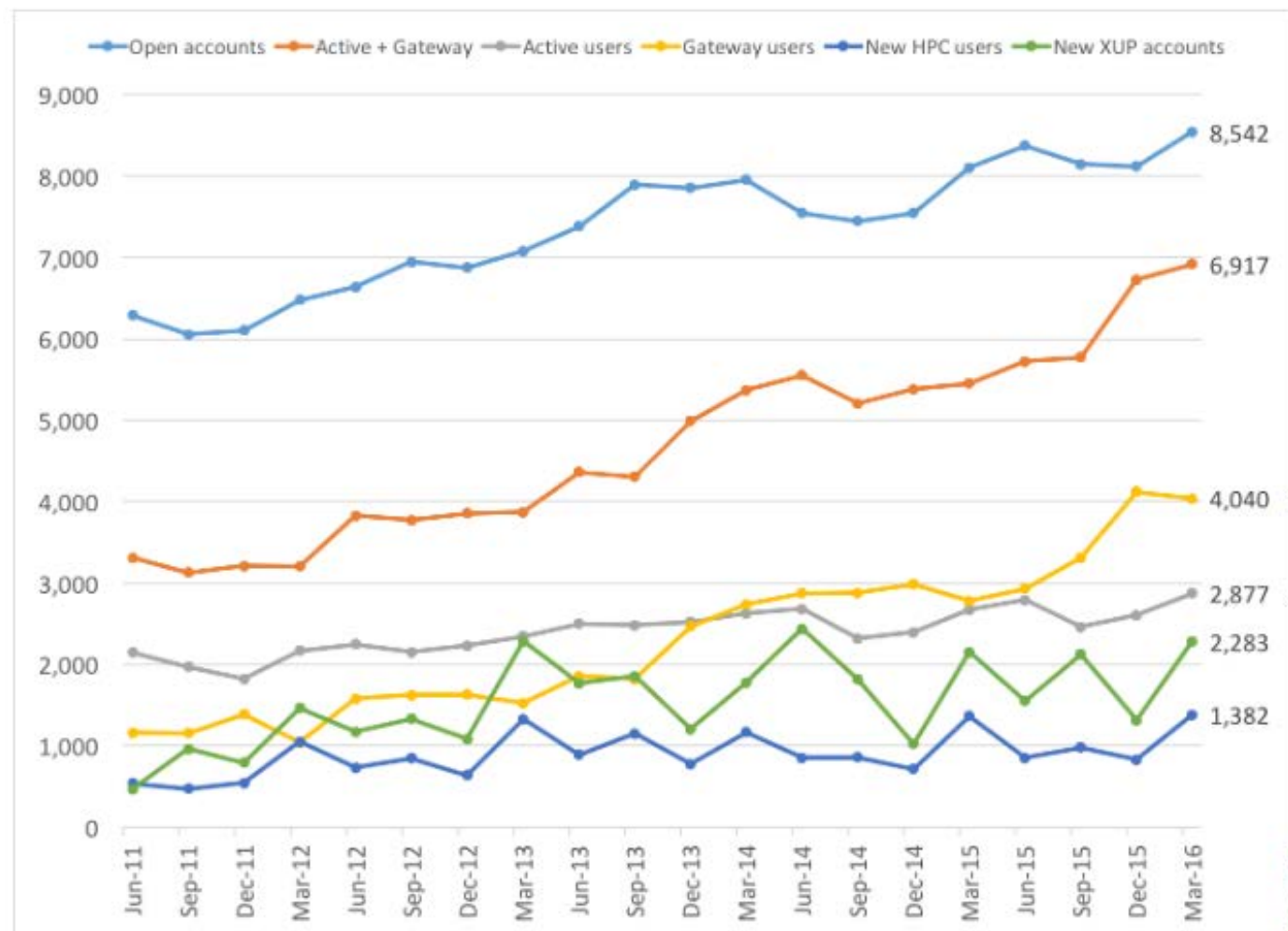
Source: David Hart

Tactics for Gateway Success:

Step 1: identify a user population in need

Step 2: commit to responding to user's needs

Step 3: let user behavior/needs drive
improvements

Step 4: manage challenges that threaten
productivity of high end users

Step 5: with limited resources, prioritization is
key

Step 6: stay in touch with your community

Step 7: embrace customer service

Nancy Wilkins-Diehr, SDSC – PI - http://sciencegateways.org/
Other institutions: Elizabeth City State in North Caronila, Indiana University,
University of Notre Dame, Purdue University, the Texas Advanced Computing Center at the
University of Texas, Austin, and the University of Michigan at Ann Arbor
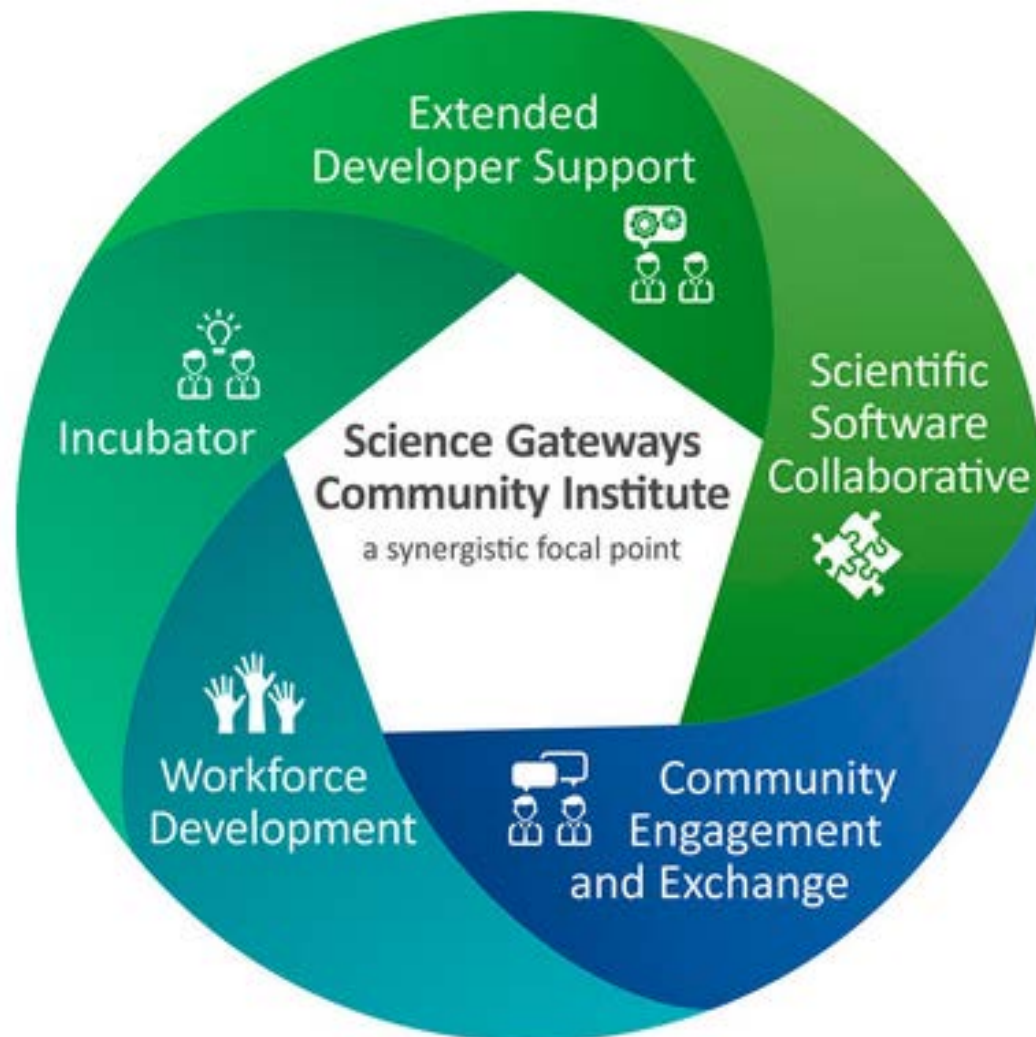
The five key areas for the Science Gateways Community Institute to increase the number, ease of use, and effective application of gateways to serve the greater research and engineering community. *Source: SDSC*

The advent of DNA sequencing lets scientists infer phylogenetic trees from multiple sequence alignments

Multiple sequence alignment is a matrix of taxa vs characters

```
                    .    .    .    ..... .    .    .
Human          AAGCTTCACCGGCGCAGTCATTCTCATAAT...
Chimpanzee     AAGCTTCACCGGCGCAATTATCCTCATAAT...
Gorilla        AAGCTTCACCGGCGCAGTTGTTCTTATAAT...
Orangutan      AAGCTTCACCGGCGCAACCACCCTCATGAT...
Gibbon         AAGCTTTACAGGTGCAACCGTCCTCATAAT...
```

Tree inference code, e.g., BEAST, MrBayes, or RAxML

Phylogeny is represented as tree with taxa at its tips

```
 /-------- Human
 |
 |---------- Chimpanzee
 +
 |    /---------- Gorilla
 |    |
 \---+             /----------------------------- Orangutan
     \------------+
                   \------------------------------------------------ Gibbon
```

SDSC develops, maintains, & hosts the NSF-funded CIPRES gateway, which runs phylogenetics codes via a browser interface; analyses span the entire tree of life

# The CIPRES gateway has been extremely popular and supports thousands of researchers around the world

- **>17,000 CIPRES users have run on NSF-funded supercomputers, including 3,152 in 4Q2015 or 47% of all active XSEDE users!**
- **>2,000 publications have been enabled by CIPRES use!**
- **US statistics from 2015**
  - 49 states + 2 territories + DC
  - 252 universities & colleges
  - 18 institutes
  - 22 museums, gardens, & zoos
  - 21 government agencies
  - 4 high schools
- **Non-US statistics from 2015**
  - 85 countries
  - 603 universities & colleges
  - 161 institutes
  - 80 museums, gardens, & zoos
  - 134 government agencies
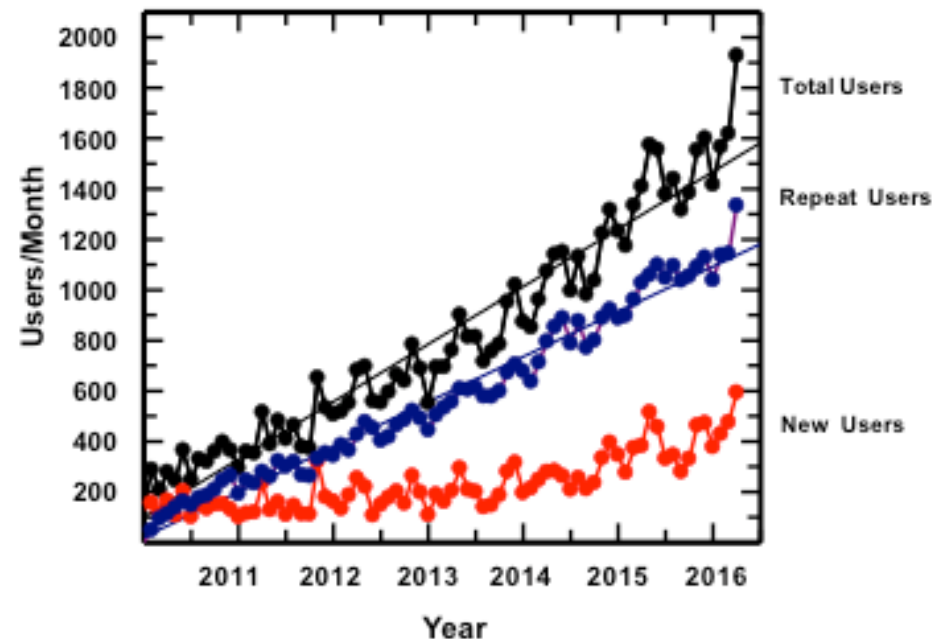
11 codes are supported by CIPRES on Comet & Gordon; most have modest scalability; some run for days

Code	Version	Language	Computer	Cores charged
BEAST	1.8.3	Java + C++	Comet	2, 4, or GPUs
BEAST2	2.3.2	Java + C++	Comet	1, 2, or 3
DPPDiv	1.0	C++	Gordon	16
FastTree	2.1.8	C	Comet	3
GARLI	2.0.1	C++	Comet	1 to 24
jModelTest2	2.1.6	Java + C	Gordon	8
MAFFT	7.187	C	Gordon	8
MrBayes	3.2.6	C + C++	Gordon	8 or 16
Migrate	3.6.11	C	Comet	1 to 72
Phylobayes	1.5a	C++	Gordon	64
RAxML	8.2.8	C	Comet	12, 24, or 48

BEAST phylogenetic tree inference for data sets with many patterns speeds up very well using K80 GPUs instead of Haswell cores on Comet

BEAST 1.8.2 on Comet
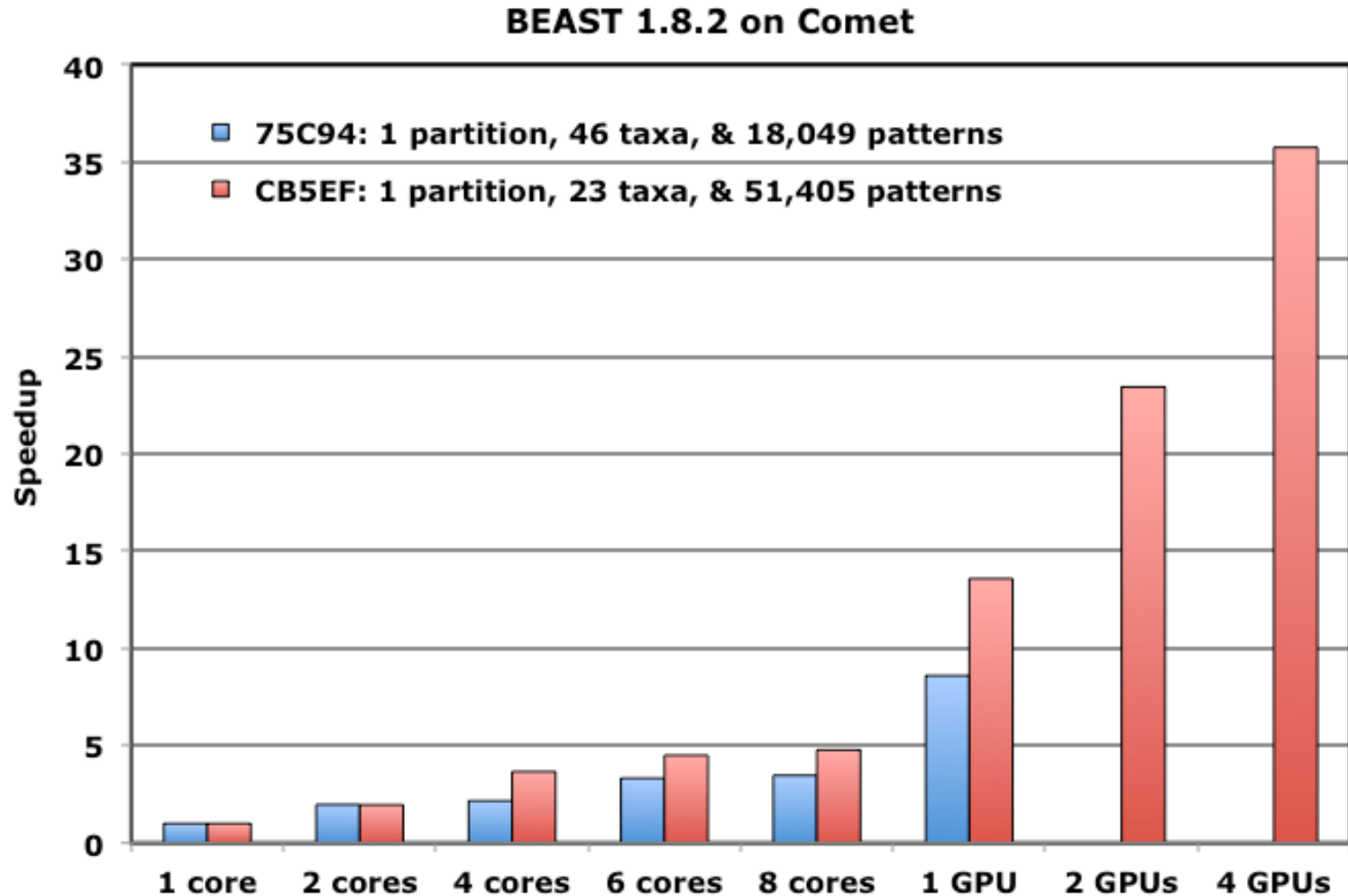
NSF funded Neuroscience Gateway at SDSC

- **NSG – in operation since early 2013 – nsgportal.org**
- **Built using the CIPRES gateway software**
- **NSG benefits the broader neuroscience research community in several ways, e.g.:**
- **Researchers can run larger complex neuronal networks, parameter sweep simulations, brain image processing tools**
- **Fully integrated The Virtual Brain (TVB) connectome pipeline workloads can be processed in parallel**
- **Easy access to widely used simulation tools such as:**
  Brian, NEST, NEURON, pGENESIS, PyNN, MOOSE, and FreeSurfer
- **Researchers from EU Human Brain Project providing optimization pipeline**
- **Access to new HPC resources – GPUs, Intel MICs**
- **Can be used by researchers with limited local (university-level) resources to address questions that require access to large scale, advanced systems**
- **Can be used by simulator developers to test, benchmark, and scale codes on large scale resources**
- **Can be used for classes, workshops, and tutorials**

BluePyOpt on Comet (1.1.27) ⓘ - Running BluePyOpt analyses

Brian on Stampede (2.0b2) ⓘ - Brian is a simulator for spiking neural networks

Brian on Comet (2.0b2) ⓘ - Brian is a simulator for spiking neural networks

The Virtual Brain Personalized Multimodal Connectome Pipeline on Comet () ⓘ - Connectome Pipeline on Comet

FREESURFER on Comet (5.3.0) ⓘ - Freesurfer tool on Comet

PyMOOSE (3.0.1 Gulab Jamun) ⓘ - Running Moose models on Comet

NEST on Stampede (2.6.0) ⓘ - Neural Simulation Technology using Python

NEST using Python on Comet (2.2.1) ⓘ - Neural Simulation Technology using Python

NEST on Stampede (2.6.0) ⓘ - Neural Simulation Technology

NEST on Comet (2.2.1) ⓘ - Neural Simulation Technology

NEURON7.3 Python on Stampede (7.3) ⓘ - Using Python to run NEURON 7.3

NEURON7.3 Python on Comet (7.3) ⓘ - Using Python to run NEURON 7.3

NEURON7.3 on Stampede (7.3) ⓘ - Latest NEURON simulation software package on Stampede

NEURON7.3 on Comet (7.3) ⓘ - Latest NEURON simulation software package on Comet

NEURON7.4 Python on Comet (7.4) ⓘ - Using Python to run NEURON 7.4

NEURON7.4 on Comet (7.4) ⓘ - Latest NEURON simulation software package on Comet

PGENESIS on Stampede (2.3) ⓘ - Parallel Genesis software

PGENESIS on Comet (2.3) ⓘ - Parallel Genesis software

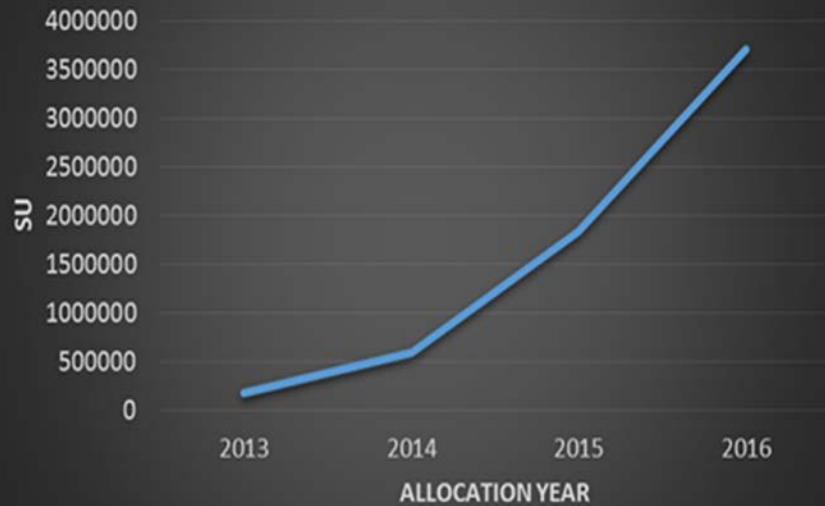**PyNN on Stampede** (0.7.5) ⓘ - Python package for simulator-independent specification of neuronal network models

PyNN on Comet (0.7.5) ⓘ - Python package for simulator-independent specification of neuronal network models
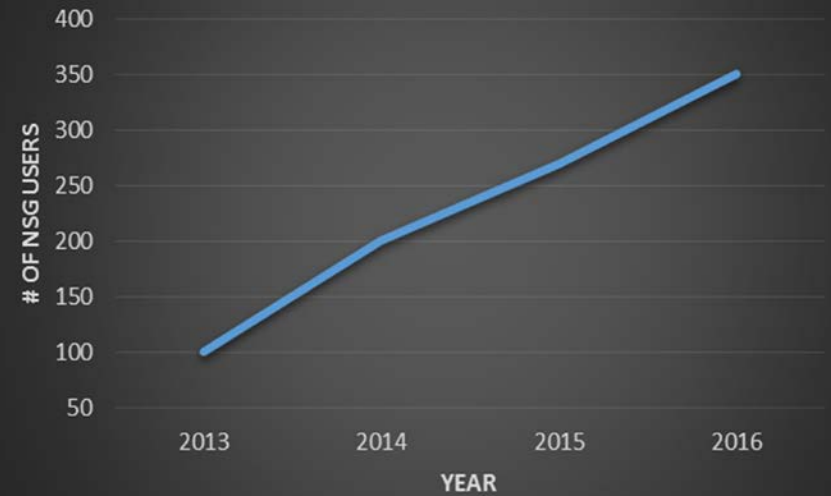
Python on Stampede (2.7.9) ⓘ - Running Python models

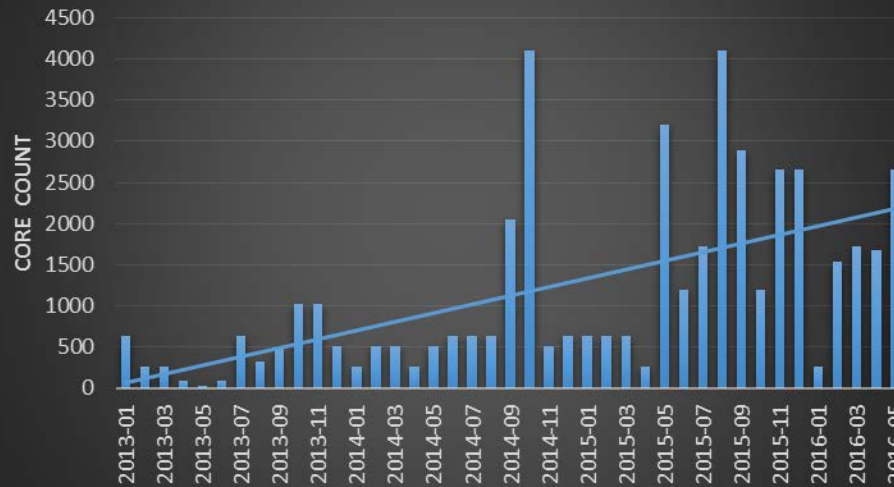**Python on Comet** (2.7.9) ⓘ - Running Python models

NSG Usage

Hands-on – to show how easy it is!

1. **Go to https://www.phylo.org and click on Use the CIPRES Science Gateway.**
2. **Login as a guest.**
3. **Upload the attached data set to Data in Guest Folder.**
4. **Go to Tasks to Create new task.**
5. **Select Input Data: i.e., the data set just uploaded.**
6. **Select Tool: BEAST on XSEDE.**
7. **Select Input Parameters. Use the default parameters with the following exceptions.**

   . **First row folks specify 400 patterns. That will have the job run on 2 cores.**
   . **Second row folks specify 4000 patterns. Then the job will run on 4 cores.**
   . **Other rows specify 4000 patterns and "always" instead of "dynamic" for beagle_scaling.**
      That will run on 4 cores, but be much slower than the default scaling.

8. **Enter Description: e.g., benchmark2.dynamic.4000patt .**

9. **Save parameters.**

10. **Save Task and enter an appropriate Description, e.g., benchmark2.dynamic.4000patt**

11. **Click on Run Task.**

12. **Click on View Status and then Intermediate Results while job is running or Output when job is done.**

13. **Then look at stdout.txt. Near the bottom, the time is output in seconds or minutes.**

Science Gateways - Summary

- **Allows anyone from anywhere to _easily_ access and use HPC (and data, instrument etc.) for computational science**
  - All users start out with some amount of core hours (depends on the gateway and the science)
  - If you graduate out of a SGW, you can write your own allocation proposal (gateway/XSEDE staff can help)
    - In many cases still use the gateway to charge to your allocation
- **It creates a cyberinfrastructure environment for the science community to enable**
  - Research
  - Education
  - Sharing of information and data