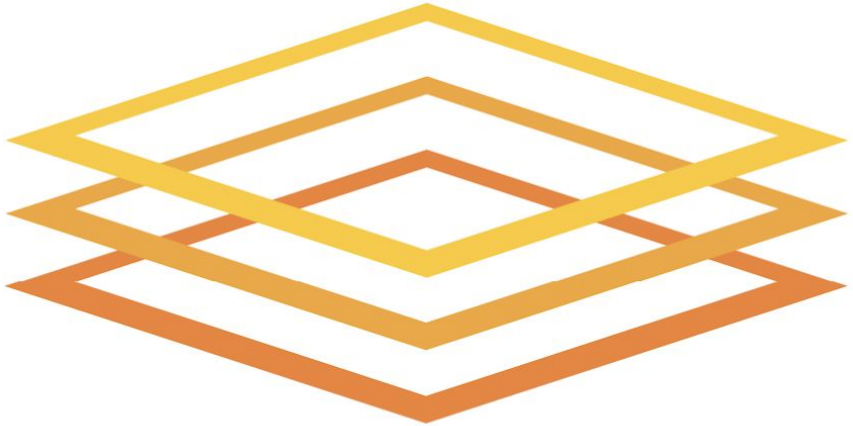


Open Science Grid

Submit locally, run globally.

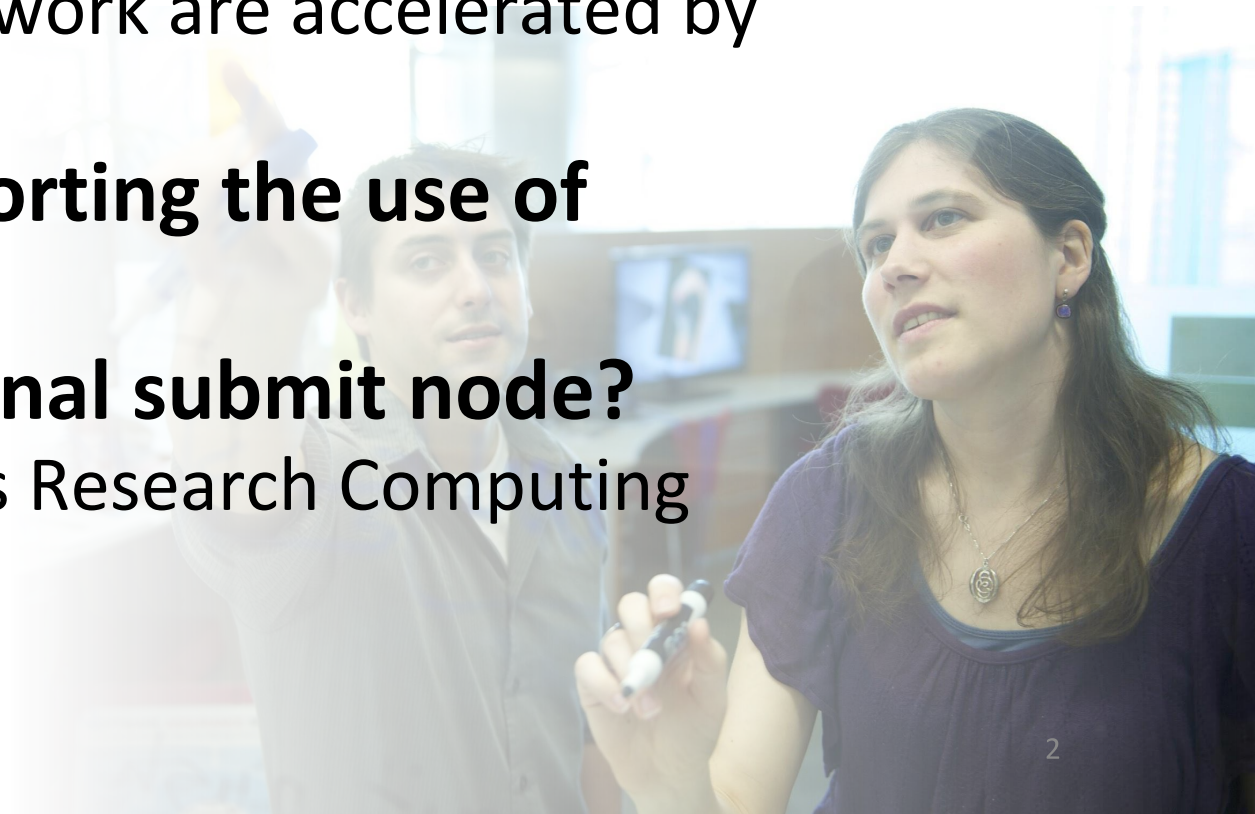
Using and Facilitating the Use of OSG

OSG Pre-Workshop, 2019 CC* PI Meeting

Lauren Michael, Research Facilitation lead (facilitation for campuses and individual research projects)

Using and Facilitating Use of OSG

- **How can your researchers interact with OSG?**
  - What types of computational work are accelerated by OSG?
- **What's different about supporting the use of HTC/OSG?**
- **When to pursue an institutional submit node?**
  - How can you learn from OSG's Research Computing Facilitators?



Open Science Grid

Engaging the Research CI Ecosystem

Proactive, personalized facilitation and support for:

- Individual researchers via **OSG Connect**
- Institutions and large collaborations
  - **Sharing institutional resources** via OSG
  - Organization-specific **submit points**
    - **Data federation** across OSG sites
  - Learning from OSG's **HTC Facilitation Community**

- **Presentations/Training** in HTC and OSG compute execution
- **HTC Facilitation Training**

For individual researchers: OSG Connect

Access to and support for using OSG's open submission point

- **osgconnect.net > "Sign Up"**
- *available to researchers at any U.S. academic, government, or non-profit organization*
- includes:
  - initial consultation with an OSG Research Computing Facilitator
  - online documentation and examples
  - access to OSG's central software modules
  - (roughly) unlimited scratch; space for staging large input (Stash); built-in data caching

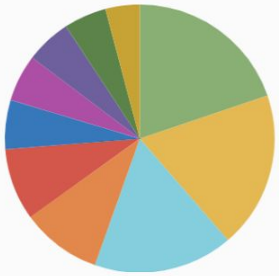
Open Science Grid

OSG Connect
(in the last year)

Total Wall Hours

149.3 Mil

Total Wall Hours for Top 10 Projects

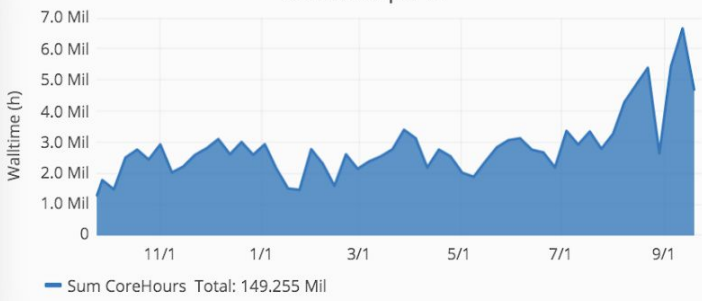


	total
SPLINTER	20.2 Mil
TG-IBN130001	19.2 Mil
IBN130001-Plus	17.1 Mil
Hawaii_Doetinchem	9.83 Mil
CpDarkMatterSimulation	8.85 Mil
chemml	6.04 Mil
molcryst	5.76 Mil
SourceCoding	5.57 Mil
PainDrugs	5.22 Mil
GLUEX	4.18 Mil

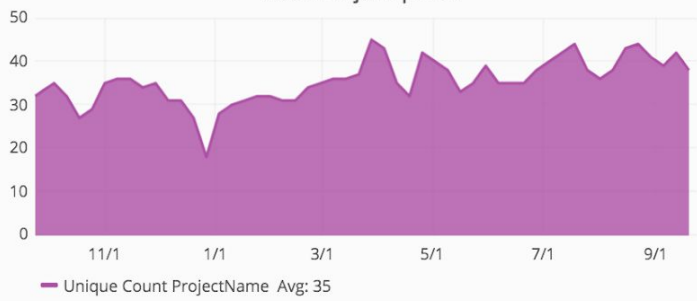
Total Jobs

136.7 Mil

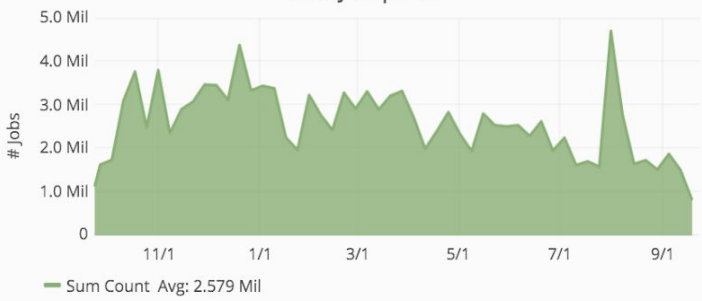
Wall Hours per 7d



Active Projects per 7d



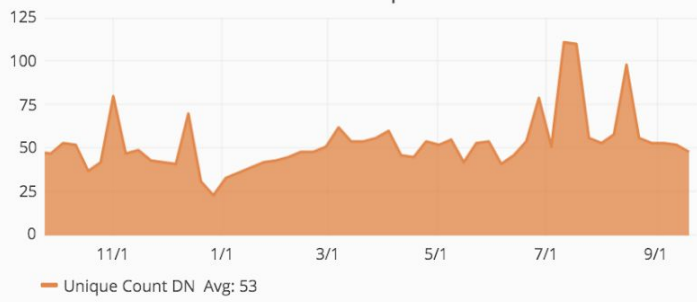
Total Jobs per 7d



Active Users

441

Active Users per 7d

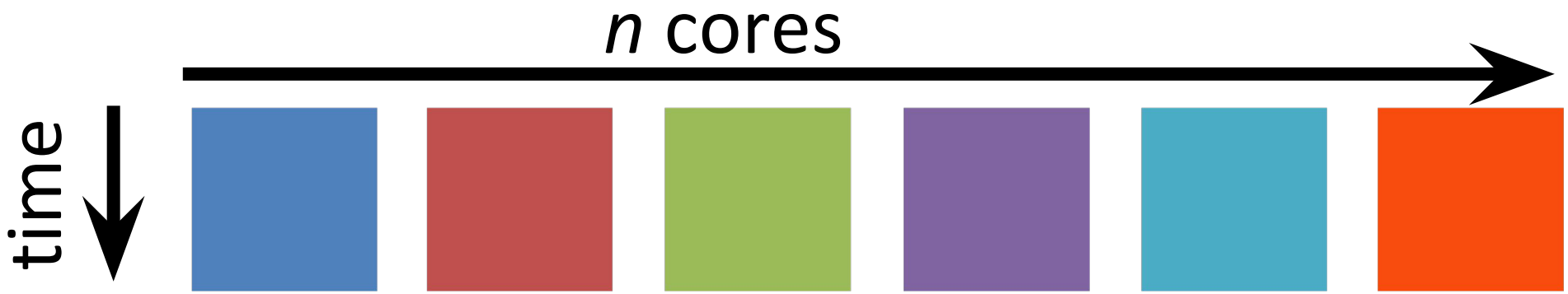


Active Projects

156

What is HTC?: An Analogy

High Throughput Computing (HTC)

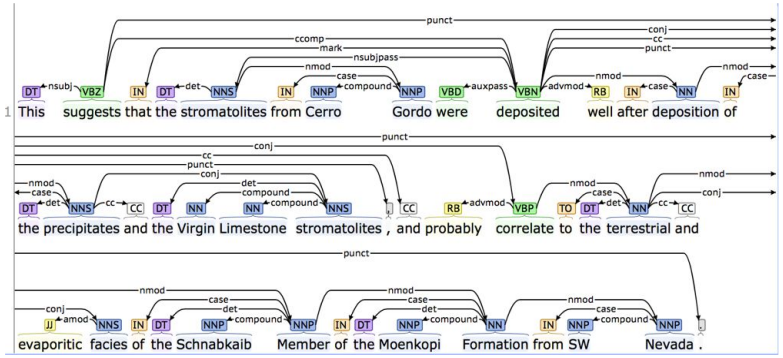


versus internal parallelism (multi-threading, MPI, etc.)

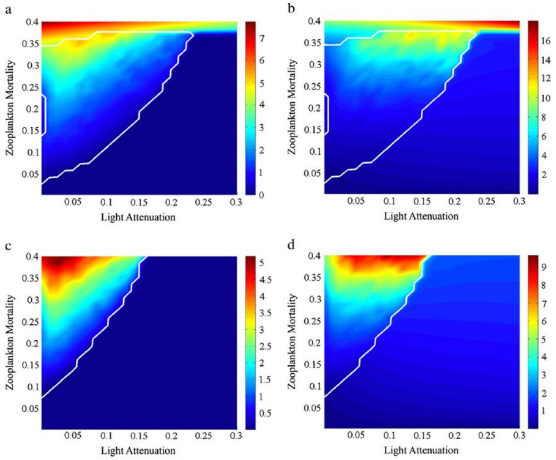
- Scheduling: only need **1 CPU core, each** (shorter wait; faster to peak)
- No special programming required
- Easier recovery from failure
- Number of concurrently-running jobs is *more* important
- CPU speed and homogeneity are *less* important

Is it HTC-able?

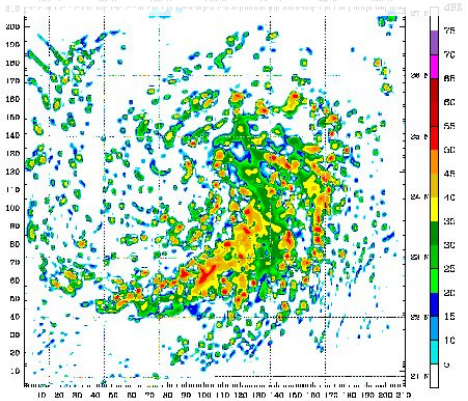
Is your problem divisible into lots of pieces?



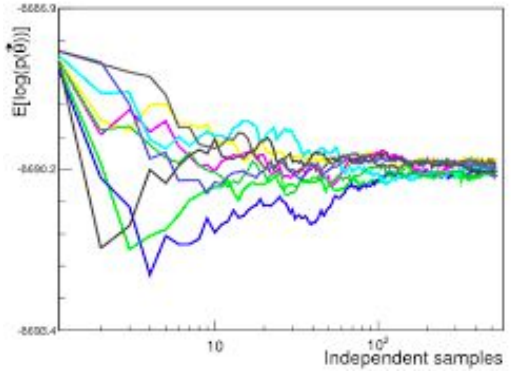
text analysis (most genomics …)



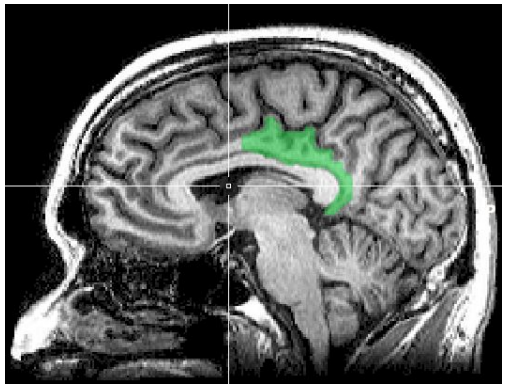
parameter sweeps



multi-start simulations



statistical model optimization (MCMC, numerical methods, etc.)



(multi-)image and sample analysis

https://www.nap.edu/catalog/21886/future-directions-for-nsf-advanced-computing-infrastructure-to-support-us-science-and-engineering-in-2017-2020

Future Directions for

In the middle of page 10...

... well-established peer of theory and experimentation. Increased capability has historically enabled new science, and many fields increasingly rely on **high-throughput** computing....

In the middle of page 13...

... Many fields increasingly rely on **high-throughput** computing that requires a greater aggregate amount of computing than a typical university can be expected to provide. Such applications can be run ...

In the middle of page 15...

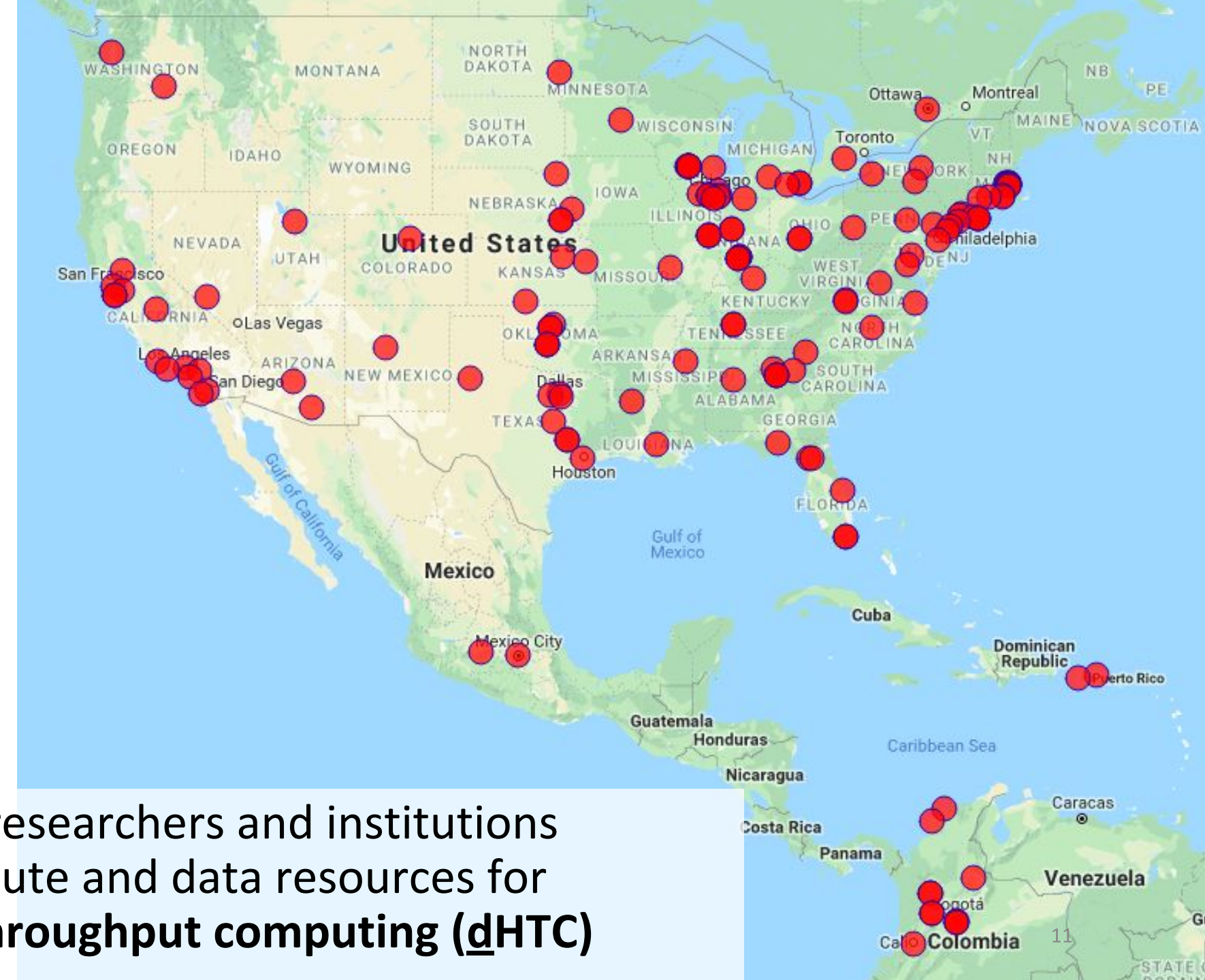
... require a single, large, tightly coupled parallel computer and (b) broaden the accessibility and utility of these large-scale platforms by allocating **high-throughput** as well as high-performance workflows to them....

in 2017–2020

https://www.nap.edu/catalog/21886/future-directions-for-nsf-advanced-computing-infrastructure-to-support-us-science-and-engineering-in-2017-2020

Open Science Grid

What is the **Open Science Grid**?



a consortium of researchers and institutions
who *share* compute and data resources for
***distributed* high-throughput computing (dHTC)**

Is it OSG-able?

Is your problem divisible into lots of “laptop-sized” pieces?

Per-Job Resources	**Ideal Jobs!** (up to 10,000 cores, per user!)	**Still Very Advantageous!**	**Probably not…**
cores (GPUs)	**1** (1; non-specific)	**<8** (1; specific GPU type)	**>8 (or MPI)** (multiple)
Walltime *(per job)*	**<10 hrs*** *or checkpointable	**<20 hrs*** *or checkpointable	**>20 hrs**
RAM *(per job)*	**<few GB**	**<10 GB**	**>10 GB**
Input *(per job)*	**<500 MB**	**<10 GB**	**>10 GB**
Output *(per job)*	**<1 GB**	**<10 GB**	**>10 GB**
Software	*‘portable’ (pre-compiled binaries, transferable, containerizable, etc.)*	*most other than □□□*	*licensed software; non-Linux*

Research Computing *Facilitation*

accelerating research transformations

proactive engagement

personalized guidance

teach-to-fish training

technology agnostic

upward advocacy

collaboration liaising

Research Computing Facilitators

The Missing Human Link in Needs-Based Research Cyberinfrastructure

ECAR

ECAR Research Bulletin | May 16, 2016

Lauren Michael, University of Wisconsin–Madison
Bruce Maas, University of Wisconsin–Madison

Overview

As the roles of core, campus-supported IT services for research have expanded—including the emergence of cloud-based models—the benefits of on-campus human support and user engagement have become increasingly apparent. Ongoing challenges in securing research funding reemphasize a need to demonstrate significant societal impact via effective and efficient investments. At the same time, many campus research computing providers still face challenges in engaging researchers represented in the "long tail" of computing needs, where potentially significant, compute-enabled transformations to scholarship have yet to be realized. The most common models for research computing resources may already meet the significant needs of well-established or "traditional" users, typically in the physical

For individual researchers: OSG Connect

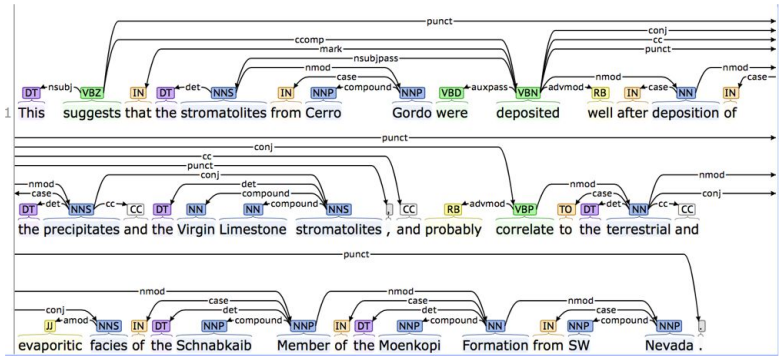
Access to and support for using OSG's open submission point

- **osgconnect.net > "Sign Up"**
- *available to researchers at any U.S. academic, government, or non-profit organization*
- includes:
  - **initial consultation with an OSG Research Computing Facilitator**
  - online documentation and examples
  - access to OSG's central software modules
  - (roughly) unlimited scratch; space for staging large input (Stash); built-in data caching

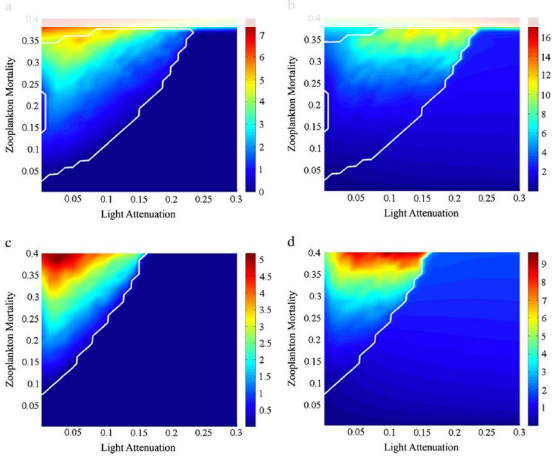
What's Different about OSG Facilitation?

1. **HTC is frequently new to users**
   - 'splitting up' work, optimizing throughput, etc.
   - many have HTC-able work and don't know (you may not, either)
2. **OSG job logistics are different than using local resources**
   - file transfer vs. shared filesystems
   - software portability vs. system-wide installation
   - inherent interruption/retry
   - testing and troubleshooting on non-local resources
3. **Potential scalability and research transformation!**

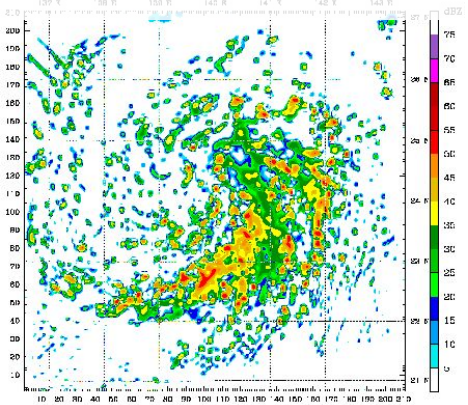
What's Different about OSG Facilitation?



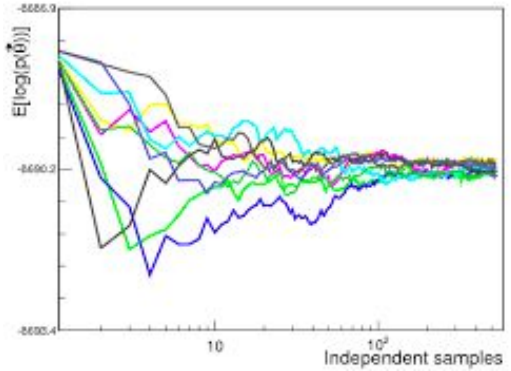
text analysis (most genomics …)



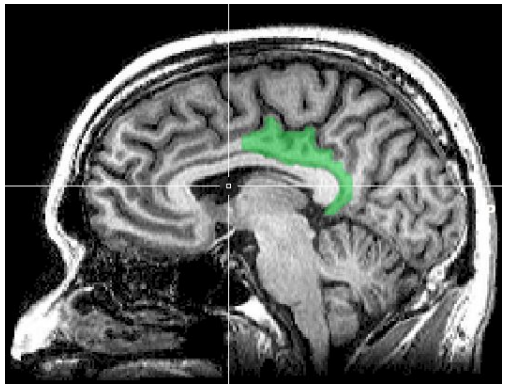
parameter sweeps



multi-start simulations



statistical model optimization (MCMC, numerical methods, etc.)



(multi-)image and sample analysis

Signs of HTC-able work

- Any mention of **numerous** samples, images, models, parameters, etc.
- Nearly anything **written by the researcher** (e.g. c/fortran, Python, R)
  - Break out of loops!
  - Common internal parallelism could really be HTC (e.g. Matlab's 'parfor', 'distributed server', etc.)
- Some community softwares that use multi-threading/multiprocessing (e.g. OpenMP)
  - many are simply looping over data portions or independent tasks
  - HTC-able: break up input (or 'parameter' space), turn off multi-threading, combine results
- **Long-running** jobs (especially if non-MPI); see above explanations

OSG Connect Facilitation Strategy

Open Science Grid

1. Meet with every potential new user.
2. Ask progressively for details.

OSG Connect Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**
   - *Tell me about your research …*

     … and how does computing fit in?
   - *What is your near-term bottleneck?*
   - *How are your running it now?*

     (the compute/data requirements?)

     (your computing background?)
   - ***How much/big would you like to run?***

OSG Connect Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**
3. **Set expectations.**

If you …

> execute your work ***this way***

> requiring ***these learning steps*** and ***this much work***

Then you …

> can reach your research goal after ***this much clock time*** and

> could really achieve ***this much more research outcome.***

OSG Connect Facilitation Strategy

Open Science Grid

1. **Meet with every potential new user.**
2. **Ask progressively for details.**
3. **Set expectations.**
4. **Follow up with the personalized plan for #3.**
5. **Make ongoing support <u>accessible</u>.**

<u>over</u>-emphasize your willingness to help!

When to pursue an institutional submit node?

*… like having a cluster where you don't administer the 'worker' nodes, but **still provide all of the user support**.*

The institution is (at least) responsible for:

- user *facilitation*
  - incl. software portability (jobs *may* use OSG modules), troubleshooting
- administering user authorization
- (some) HTCondor administration on the submit node
- administering/integrating any institutional data storage
  - (may exclude a 'hosted' submit node)

Open Science Grid

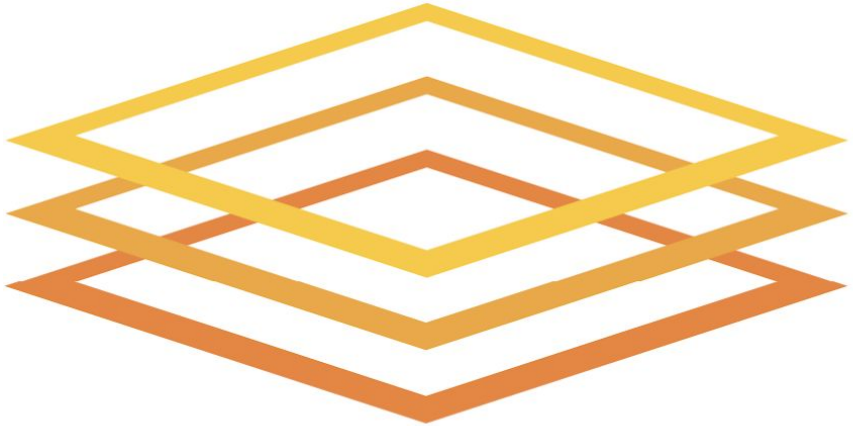
HTC Training

- **Trainings at Conferences and Community Events**
  - intro to OSG participation, HTC, and OSG compute execution; usually half-day
  - targeted for research computing staff, gateway developers, etc.
  - recent/upcoming opportunities:
    - Gateways 2019, OSG AHM 2019, Internet2 TechX 2018, RMACC 2018
- **OSG User School, annual (summer)**
  - Week-long education on HTC and OSG compute execution
  - Targeted for researchers *and* research computing staff

HTC Facilitation Training

Learn from OSG's Research Computing Facilitators

- **HTC Facilitation Shadowing**
  - Join OSG RCFs in facilitating use of **OSG Connect** by researchers from *your* institution
  - On-site shadowing at UW-Madison
- **HTC Facilitation Community (starting up!)**
  - Email list and virtual meet-ups to discuss HTC Facilitation
  - Meet-ups at the annual OSG All-Hands Meeting (spring) and OSG User School



Open Science Grid

Submit locally, run globally.

Questions?

- **help@opensciencegrid.org**
- Lauren Michael, lmichael@wisc.edu
  - Research Facilitation lead (facilitation for campuses and OSG Connect users)