What's Different About Overlay Systems?

Brian Lin

OSG Software Team

University of Wisconsin - Madison

Overlay Systems are Awesome!

Free resources when you need them? With the OSG doing the hard work? Yes, please!

What's the Catch?

Requires more infrastructure, software, set-up, management, troubleshooting...

Open Science Grid

*"You know you have a **distributed system** when the crash of a computer you've never heard of stops you from getting any work done."*

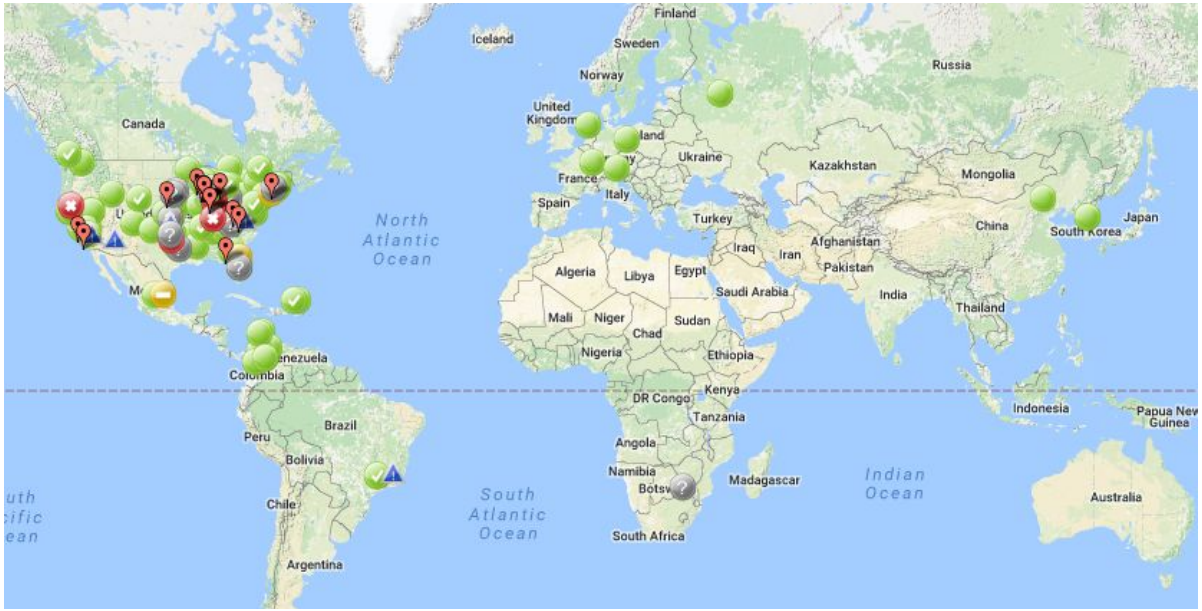
- Leslie Lamport

Open Science Grid

#1: Heterogenous Resources

Accounting for differences between the
OSG and your local cluster

Sites of the OSG



Source: http://display.opensciencegrid.org/

Heterogeneous Resources - Software

- Different operating systems (Red Hat, CentOS, Scientific Linux; versions 6 and 7)
- Varying software versions (e.g., at least Python 2.6)
- Varying software availability (e.g., no BLAST*)

Solution: Make your jobs more portable, OASIS (more in Wednesday's talks)

Heterogeneous Resources - Hardware

- CPU: Mostly single core
- RAM: Mostly < 8GB
- GPU: Limited #s but more being added
- Disk: No shared file system (more in Thursday's talks)

Solution: Split up your workflow to make your jobs more high throughput

Open Science Grid

#2: With Great Power Comes Great Responsibility

How to be a good netizen

Resources That You Don't Own

- Primary resource owners can kick you off for any reason
  - **Solution**: Implement self-checkpointing, HTC-ize your jobs
- No local sys admin relationship
- No sensitive data (again)!

Be a Good Citizen!

- Use of shared resources is a privilege
- Only use the resources that you request
- Be nice to your submit nodes

Solution: Test jobs locally and when you're done test them some more

#3: Slower Ramp Up

Leasing resources takes time!

Slower Ramp Up

- Adding slots: pilot process in the OSG
  vs slots already in your local pool
- Not a lot of time (~minutes) compared to most job runtimes (~hours)
- Small trade-off for increased availability

Open Science Grid

Thanks!

Questions?