



Open Science Grid

Submit locally, run globally.

The Open Science Grid (OSG) provides the national fabric of distributed high throughput computing, serving researchers across a wide variety of scientific disciplines.

Over 125 institutions
sharing resources

Over 500,000 research
computing jobs a day

Over 925,000,000 CPU
hours in the last year



Scalable, shared resources for:

- *individual researchers*
- *multi-institutional collaborations*
- *institutions and their research computing providers*

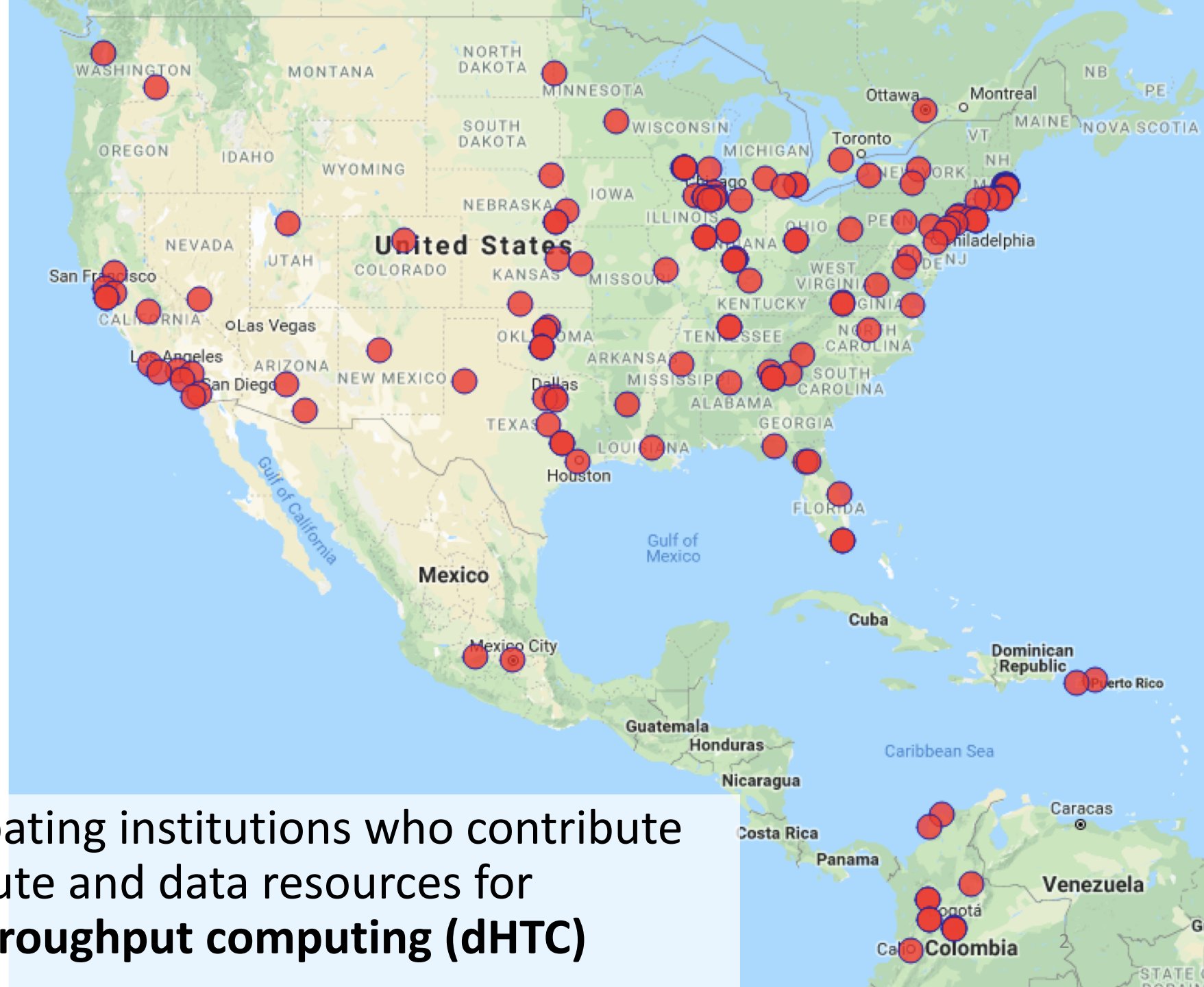
opensciencegrid.org



Open Science Grid

What is the Open Science Grid?

a consortium of participating institutions who contribute
shareable compute and data resources for
***distributed* high-throughput computing (dHTC)**





Open Science Grid

Engaging the Research CI Ecosystem

Proactive, personalized facilitation and support for:

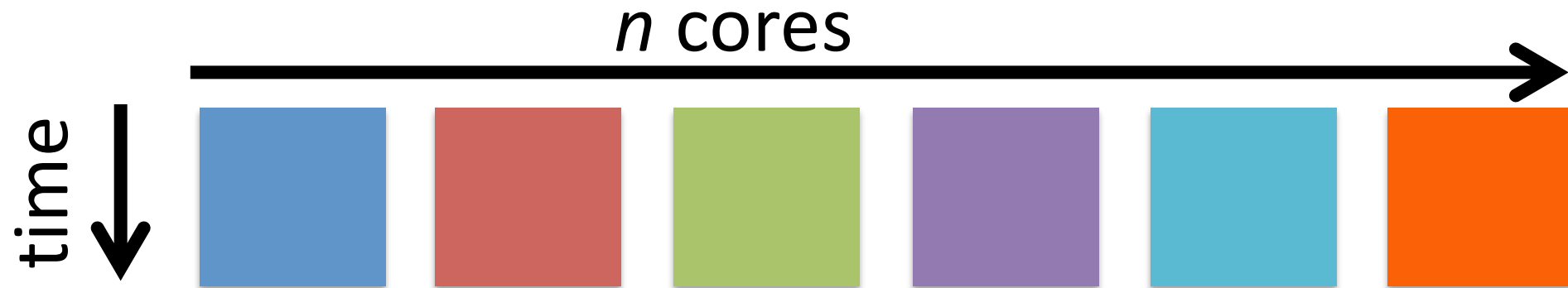
- Individual researchers via **OSG Connect**
- Institutions and large collaborations
 - Organization-specific **submit points**
 - **Contributing institutional resources** via OSG
 - **Data federation** across OSG sites
 - Learning from OSG's HTC Facilitation Community
- **Presentations/Training** at community meetings
- OSG-hosted education and training (OSG User School)



HPC vs HTC: An Analogy



High Throughput Computing (HTC)



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

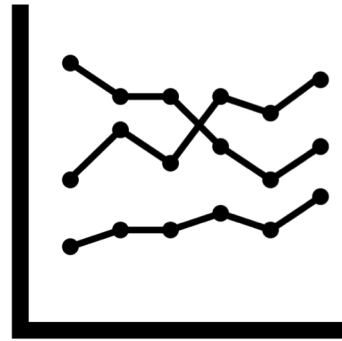
- Scheduling: only need **1 CPU core, each** (shorter wait, faster to peak)
- Easier recovery from failure
- No special programming required
- Number of concurrently running jobs is *more* important
- CPU speed and homogeneity are *less* important

Is it HTC-able?

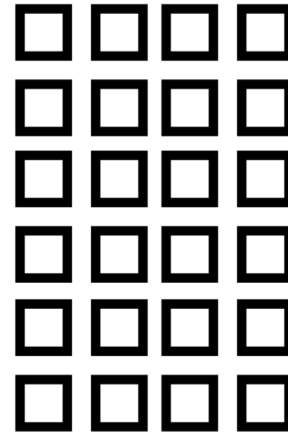
Is your problem divisible into lots of pieces?



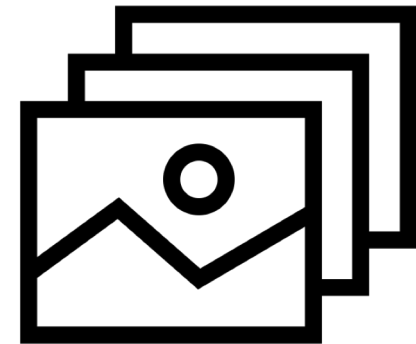
RNA/DNA
read mapping
(and other genomics)
(and other text analysis)



statistical model
optimization



parameter
optimization



multi-image/sample
analysis

Is it OSG-able?

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

	Ideal Jobs! (up to 10,000 cores across jobs, per user!)	Still Very Advantageous!	Less-so, but maybe
cores (GPUs)	1 (1; non-specific type)	<8 (1; specific GPU type)	>8 (or MPI) (multiple)
Walltime	<12 hrs* *or checkpointable	<24 hrs* *or checkpointable	>24 hrs
RAM	<few GB	<10s GB	>10s GB
Input	<500 MB	<10 GB	>10 GB
Output	<1 GB	<10 GB	>10 GB
Software	<i>‘portable’ (pre-compiled binaries, transferable, containerizable, etc.)</i>	<i>most other than</i> →	<i>Licensed software; non-Linux</i>

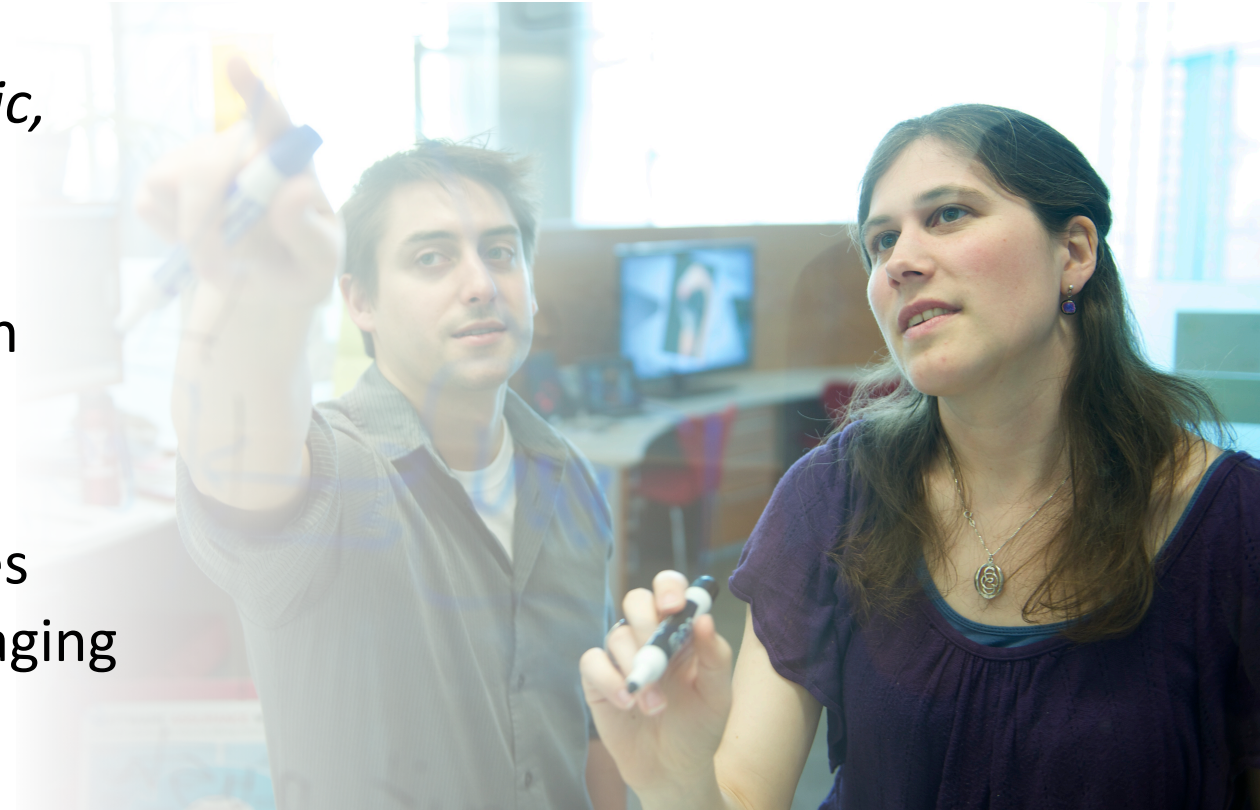


Open Science Grid

For individual researchers: **OSG Connect**

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

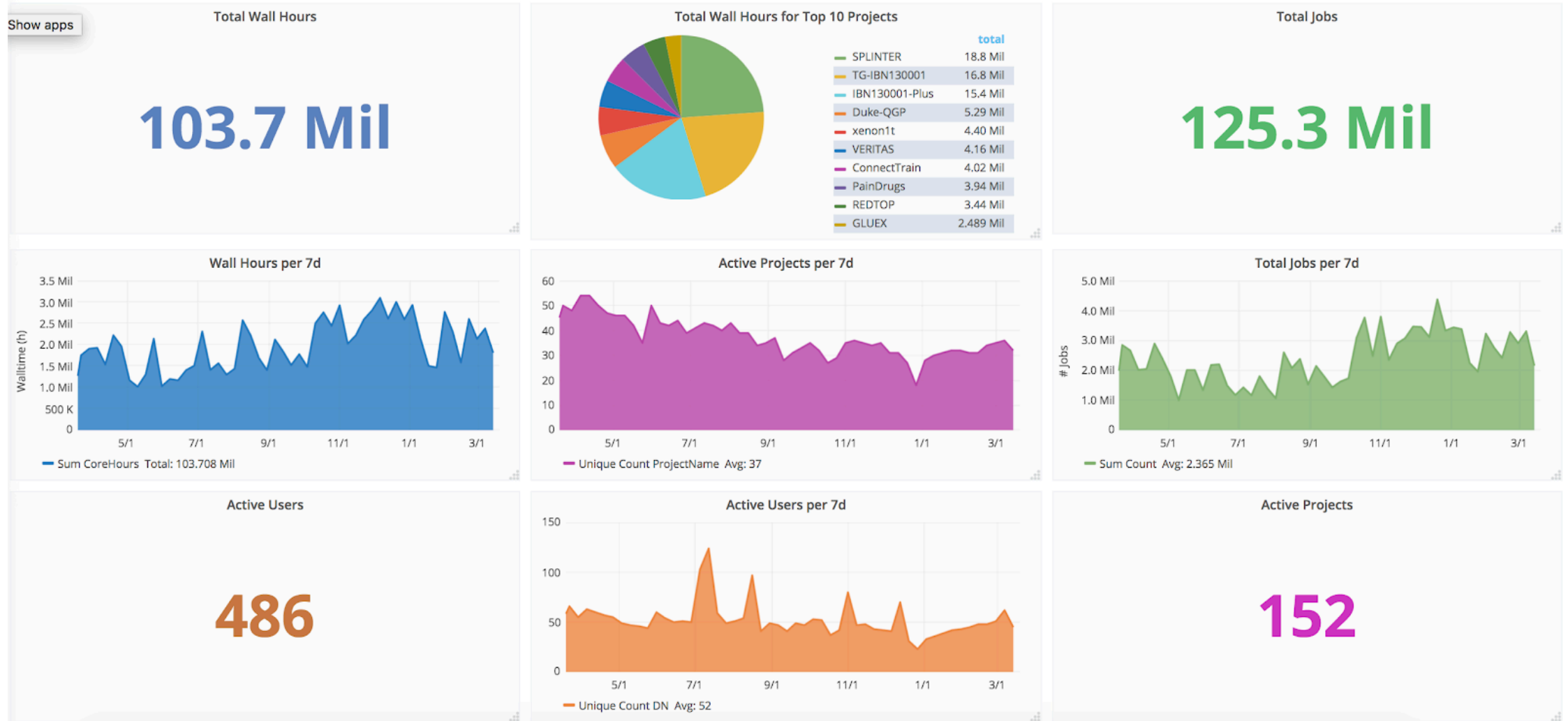
- osgconnect.net > “Sign Up”
- *available to researchers at any U.S. academic, government, or non-profit organization*
- includes:
 - initial consultation with an OSG Research Computing Facilitator
 - online documentation and examples
 - access to OSG's central software modules
 - (roughly) unlimited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





Open Science Grid

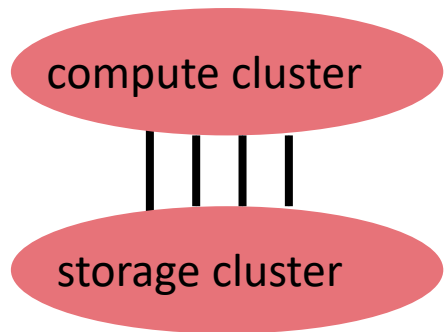
For individual researchers: OSG Connect





Open Science Grid

For institutions: OSG Integration Services

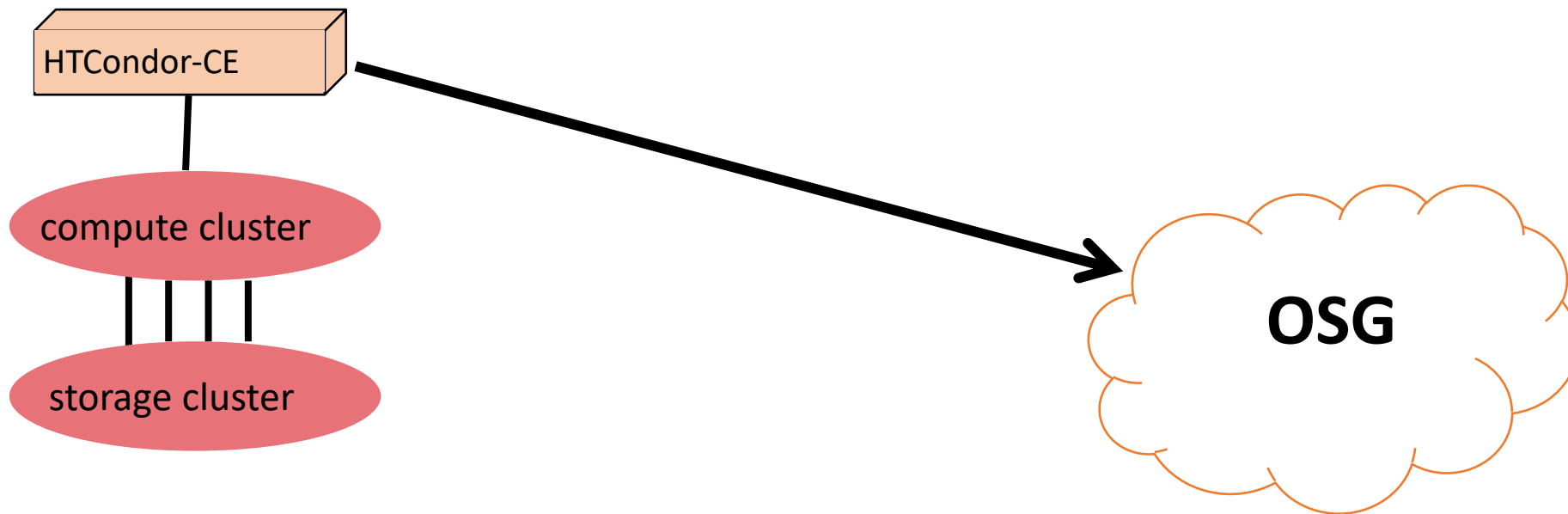


Campuses *at least* operate the stuff in red.
OSG can operate everything in orange.



Open Science Grid

For institutions: OSG Integration Services

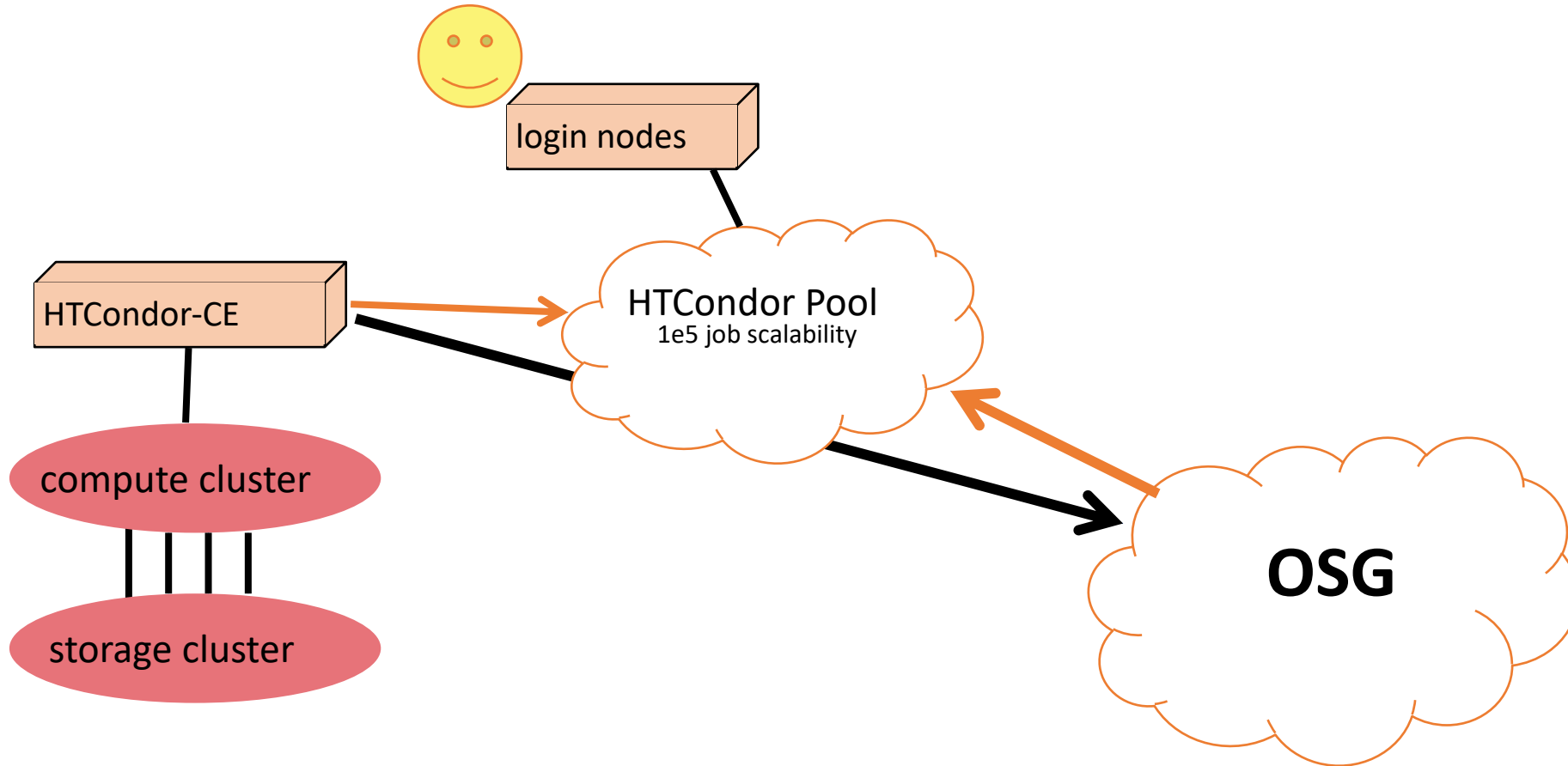


Campuses at least operate the stuff in red.
OSG can operate everything in orange.



Open Science Grid

For institutions: OSG Integration Services

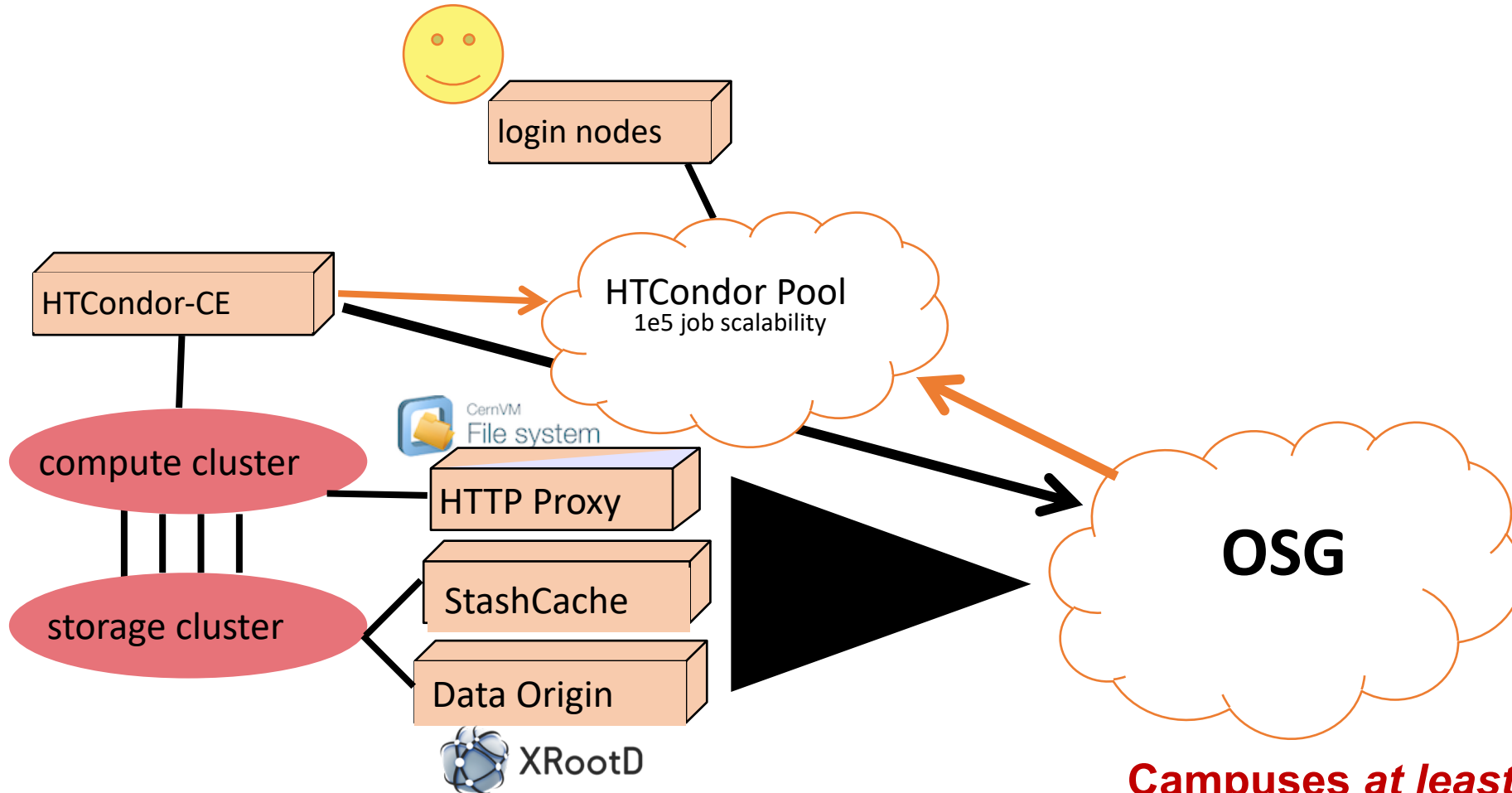


Campuses at least operate the stuff in red.
OSG can operate everything in orange.



Open Science Grid

For institutions: OSG Data Federation

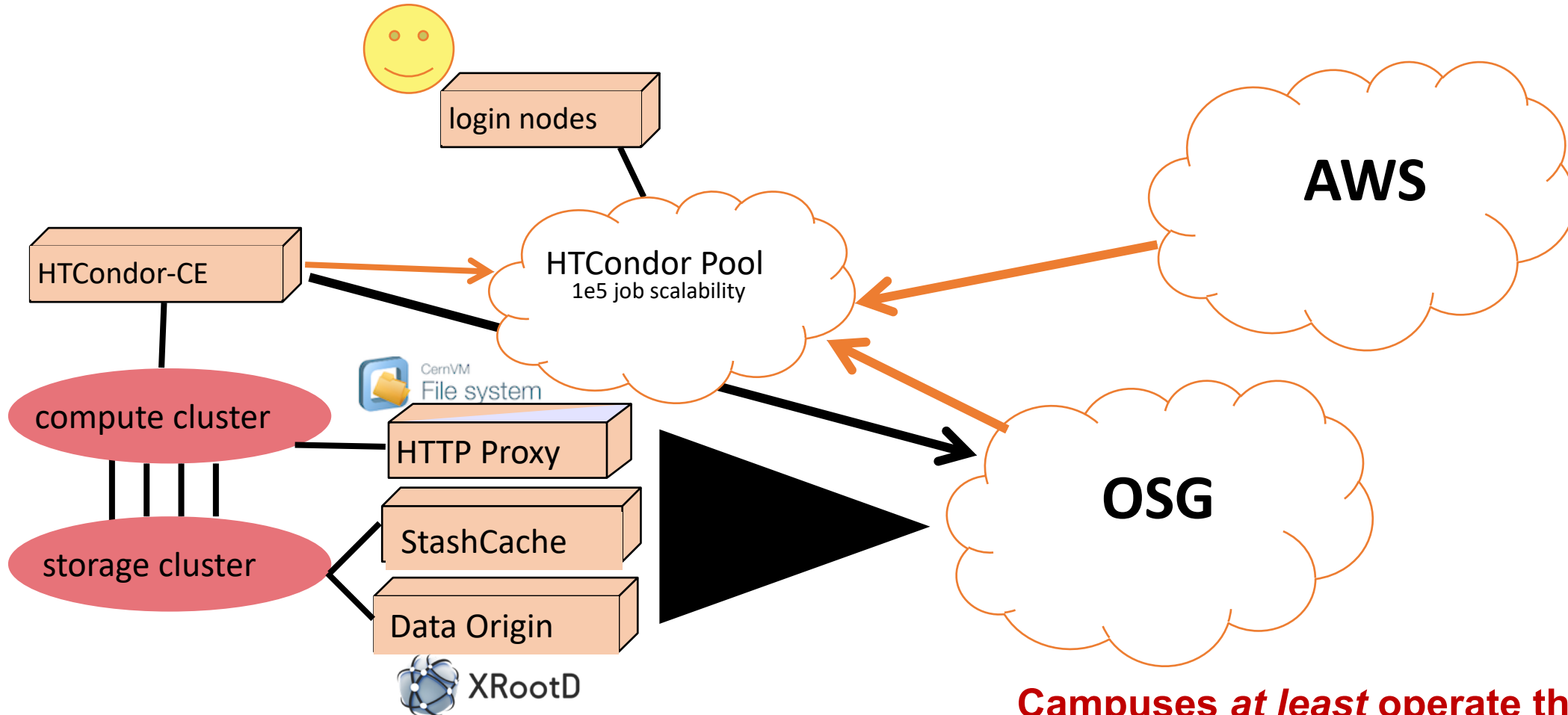


Campuses at least operate the stuff in red.
OSG can operate everything in orange.



Open Science Grid

For institutions: OSG Cloud Integration

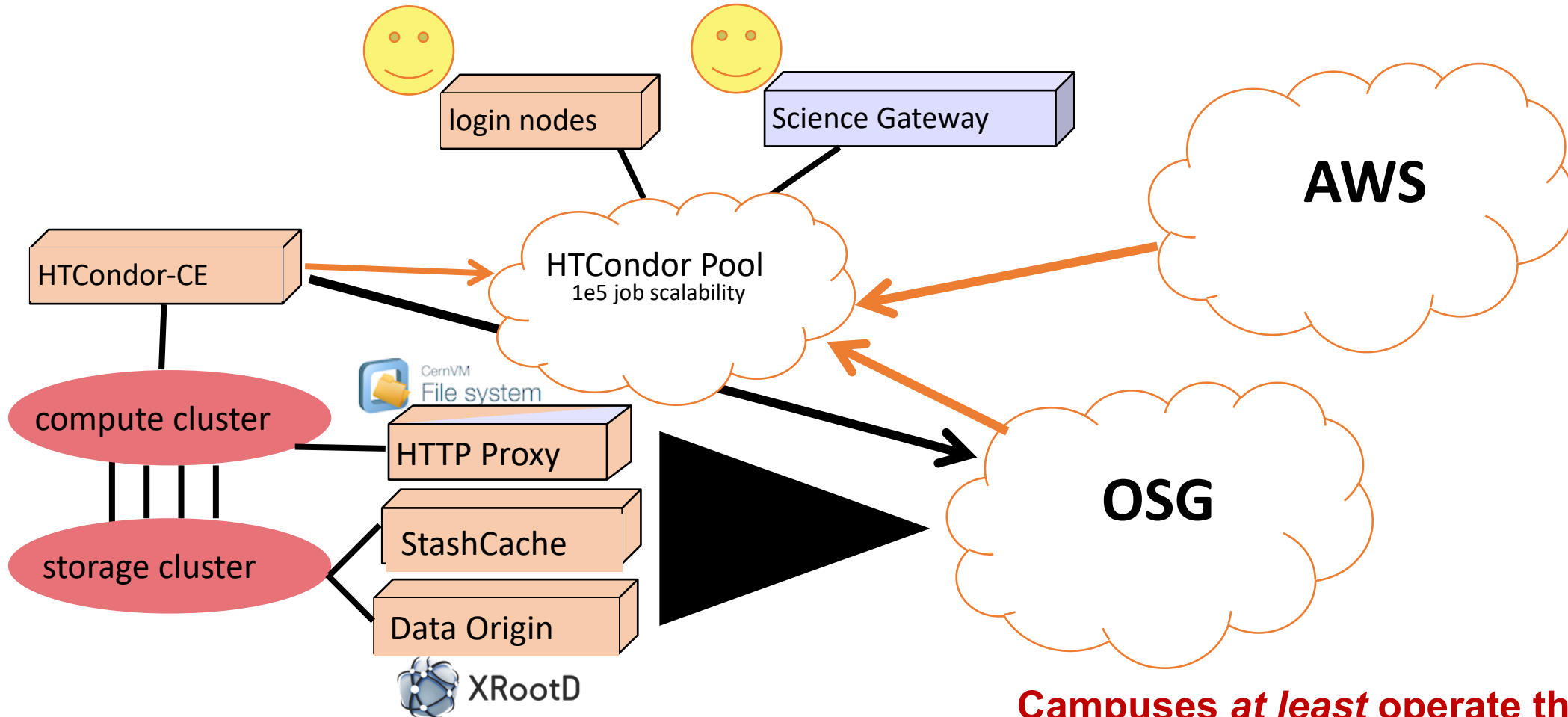


Campuses at least operate the stuff in red.
OSG can operate everything in orange.



Open Science Grid

For institutions: OSG for Gateways



Campuses at least operate the stuff in red.
OSG can operate everything in orange.

Future Directions for
**NSF ADVANCED
COMPUTING
INFRASTRUCTURE**

to Support U.S. 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



Open Science Grid

Submit locally, run globally.

Questions?

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