Globus Endpoint Deployment and Administration

Vas Vasiliadis
University of Chicago,
Argonne National Laboratory

vas@uchicago.edu

GlobusWorld
April 20, 2016

Presentation material available at

globusworld.org/**tutorial2016**

Agenda

- Research data management challenges
- Globus: a high-level flyover
- File Transfer and Sharing: Accelerating and streamlining collaboration
- Data Publication: Enhancing reproducibility and discoverability
- Our sustainability challenge
- Globus campus deployment & intergation
- Deployment best practices: the Science DMZ
- Leveraging the Globus platform

Demonstration:
Command Line Interface (CLI)

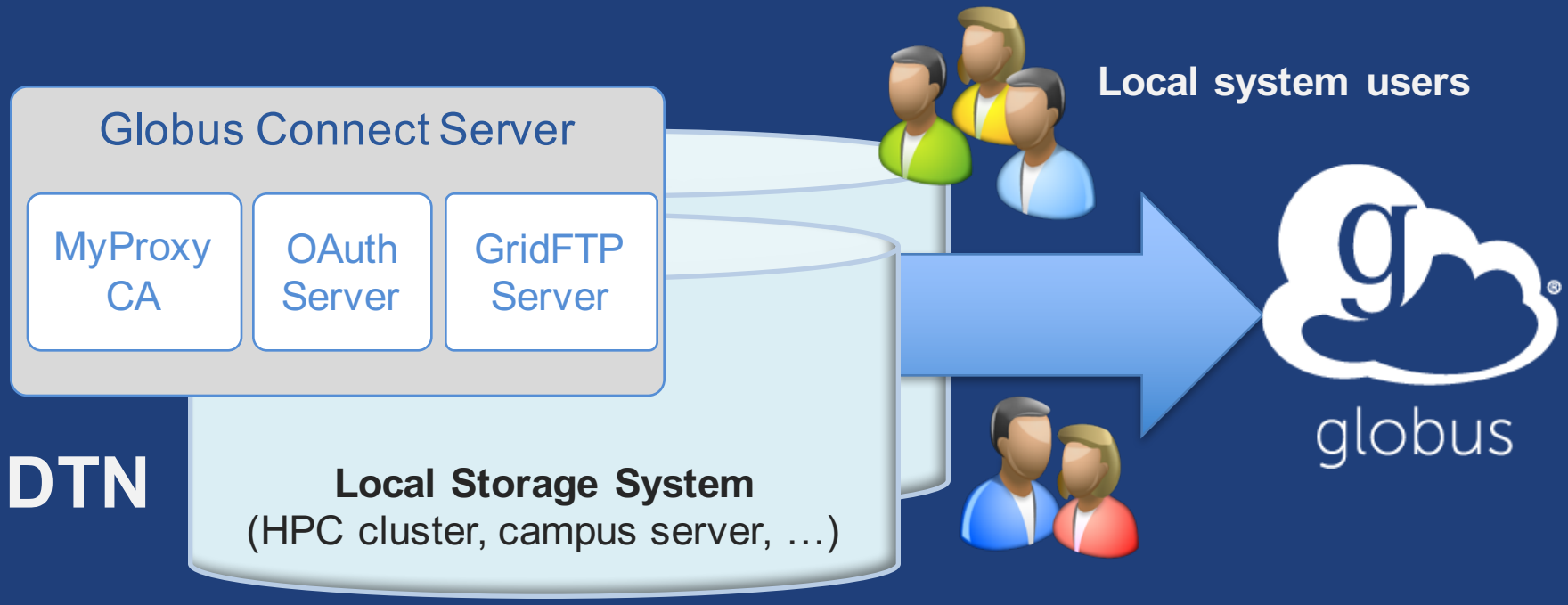
Exercise: Globus CLI

1. **Create a Globus ID**
2. **Go to: globusid.org/keys**
3. **Add your SSH key to your Globus ID**
   globusid.org > login > manage SSH and X.509 keys
4. `ssh <globusid>@cli.globusonline.org`
5. **Run `help` to see available commands**

Enabling your storage system:

Globus Connect Server

Globus Connect Server

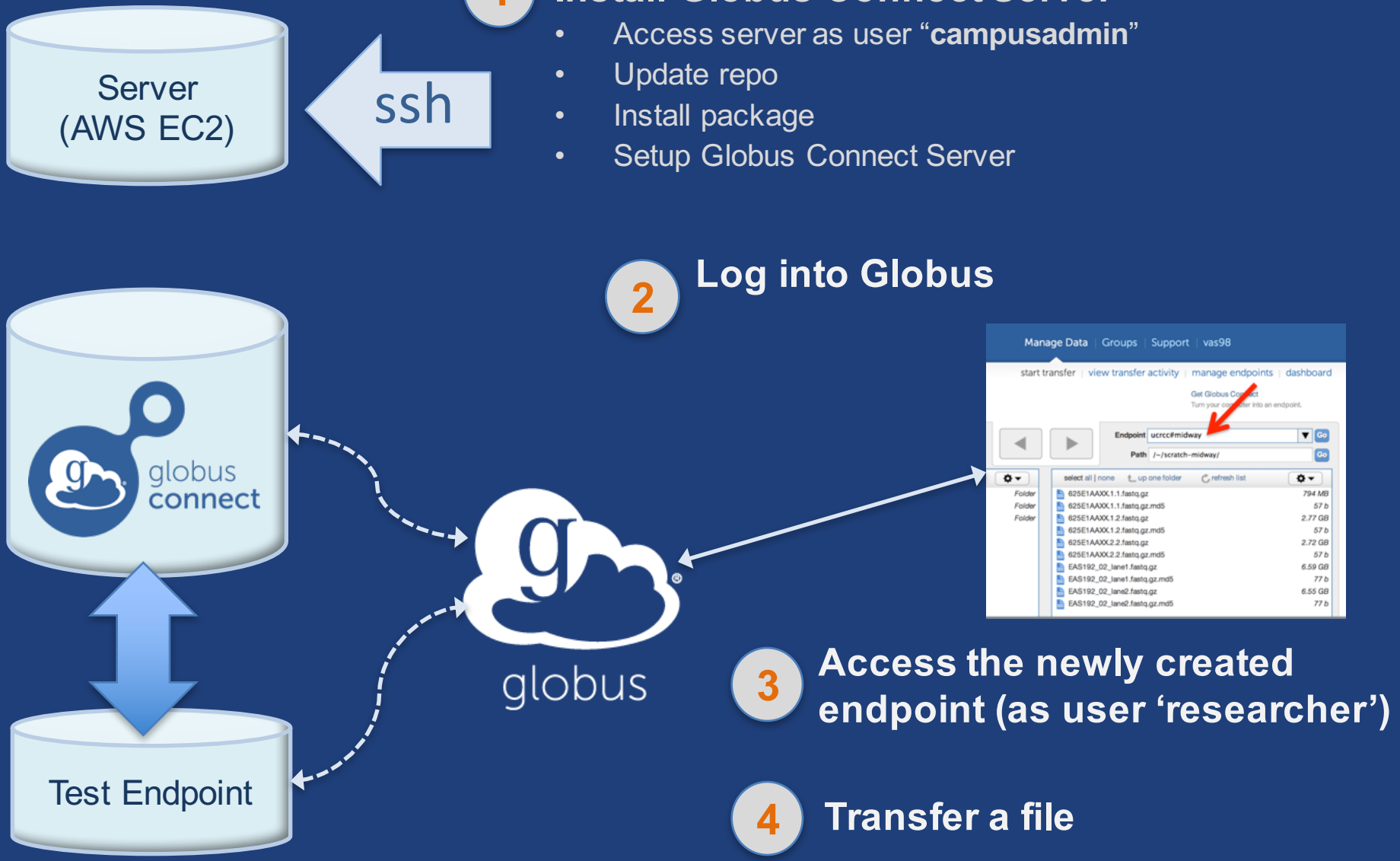


- **Create endpoint on practically any filesystem**
- **Enable access for all users with local accounts**
- **Native packages: RPMs and DEBs**

Demonstration

- **Creating a Globus endpoint on your storage system**
- **In this example, storage system = Amazon EC2 server**
- **Akin to what you would do on your DTN**

What we are going to do:



Access your host

- **Go to globusid.org and create a Globus ID if you don't have one already**
  - Optional: associate it with your Globus account
- **Your slip of paper has the host IP**
- **Log in as user 'campusadmin':**

  `ssh campusadmin@<EC2_instance_IP_address>`

  (password: **globus2016**)
- **NB: Please** `sudo su` **before continuing**
  - User 'campusadmin' has passwordless sudo privileges

Step 3: Install Globus Connect Server

(globus.org/events/oin-workshop/globus-tutorial)

```
$ sudo su
$ curl -LOs http://toolkit.globus.org/ftppub/globus-
connect-server/globus-connect-server-
repo_latest_all.deb
$ dpkg -i globus-connect-server-repo_latest_all.deb
$ apt-get update
$ apt-get -y install globus-connect-server
$ globus-connect-server-setup
```

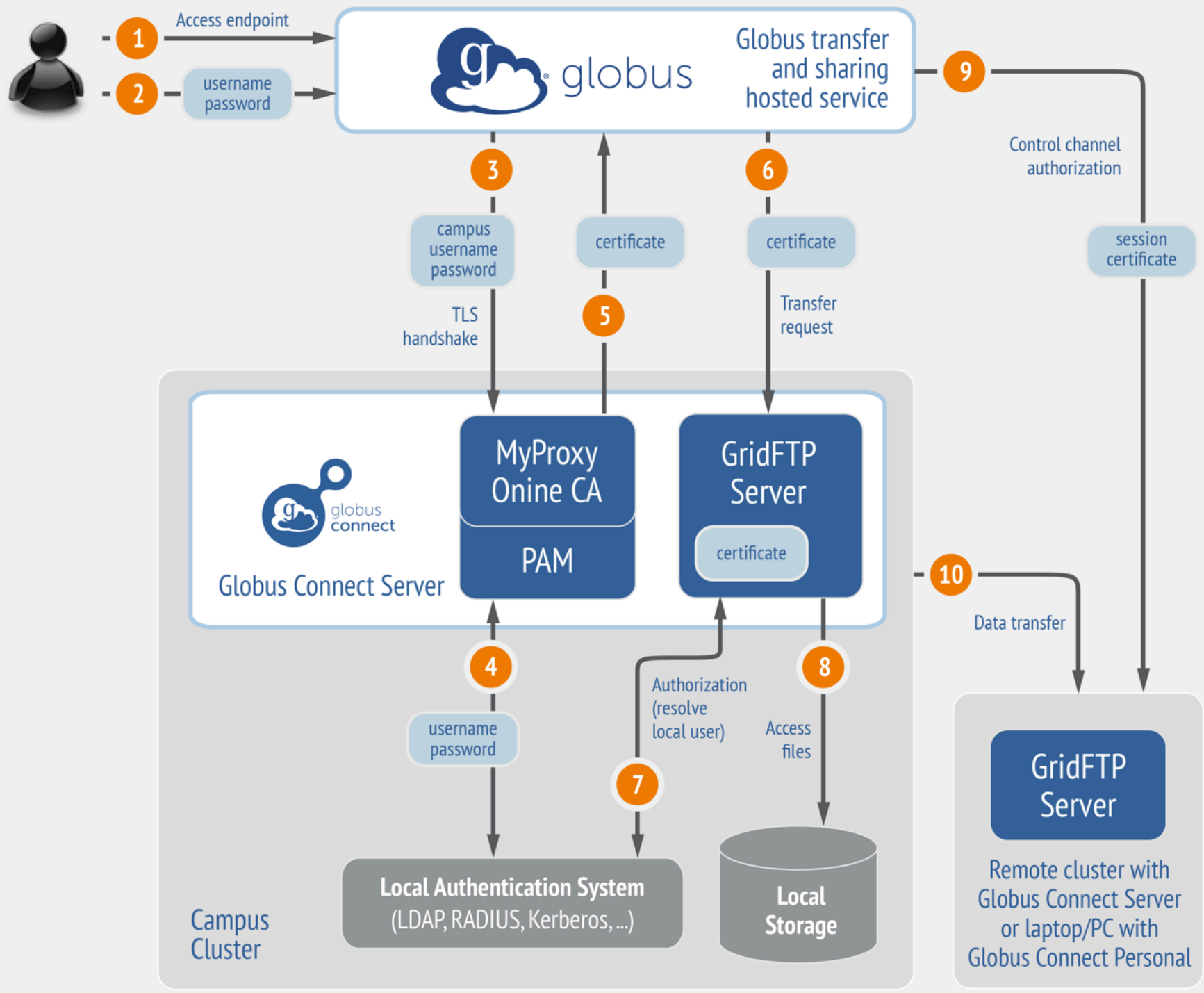
Use your Globus ID username/password here

You have a working Globus endpoint!

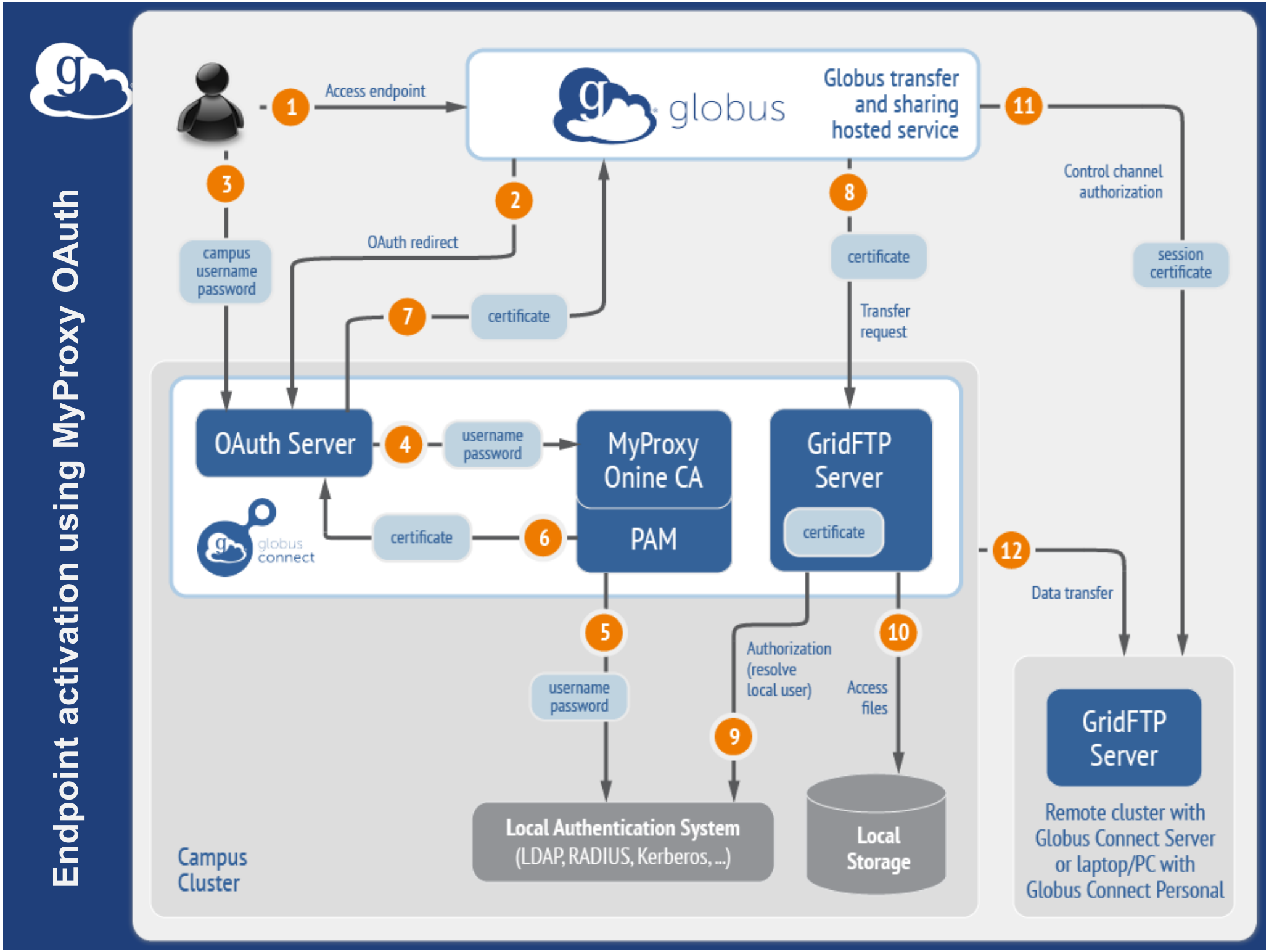
Access the Globus endpoint

- **Go to Manage Data → Transfer Files**
- **Access the endpoint you just created**
  - Search for your EC2 DNS name in the Endpoint field
  - Log in as user "**researcher**" (pwd: **globus2016**); you should see the user's home directory
- **Transfer files**
  - Between esnet#???-diskpt1 and your endpoint

Endpoint activation using MyProxy



Endpoint activation using MyProxy OAuth



Ports needed for Globus

- **Inbound: 2811 (control channel)**
- **Inbound: 7512 (MyProxy), 443 (OAuth)**
- **Inbound: 50000-51000 (data channel)**
- **If restricting outbound connections, allow connections from:**
  - 80, 2223 (used during install/config)
  - 50000-51000 (GridFTP data channel)
- **Futures: single-port GridFTP**

Configuring Globus Connect Server

- **Configuration options specified in:**
  `/etc/globus-connect-server.conf`
- **To enable changes you must run:**
  `globus-connect-server-setup`
- **“Rinse and repeat”**

Configuration file walkthrough

- **Structure based on .ini format**

```
[Section]
Option
```

- **Commonly configured options:**

```
Name
Public
RestrictedPaths
Sharing
SharingRestrictedPaths
IdentityMethod (CILogon, Oauth)
```

Changing your endpoint name

- **Edit `/etc/globus-connect-server.conf`**
- **Set `[Endpoint] Name = "myendpoint"`**
- **Run `globus-connect-server-setup`**
  - Enter your Globus ID & password when prompted
- **Access the endpoint in your browser using the new endpoint name**
  - You may need to refresh your browser to see the new name in the endpoint list

Making your endpoint public

- **Edit `/etc/globus-connect-server.conf`**
- **Set `[Endpoint] Public = True`**
- **Run `globus-connect-server-setup`**
- **Try accessing a neighbor's endpoint:**
  - You will be prompted for credentials…
  - …access endpoint as "**researcher**"

Enabling sharing on an endpoint

- **Edit: `/etc/globus-connect-server.conf`**
- **Uncomment `[GridFTP] Sharing = True`**
- **Run `globus-connect-server-setup`**
- **Go to the Transfer Files page**
- **Select the endpoint**
- **Create shared endpoints and grant access to other Globus users***

* Note: Shared endpoints may only be created on managed endpoints under a **Globus subscription**

Creating managed endpoints

- **Required for sharing, management console, reporting, etc.**
- **Convert existing endpoint to managed:**

```
endpoint-modify --managed-endpoint <endpoint_name>
```

- **Must be run by subscription manager, using the Globus CLI**
- **Important: Re-run endpoint-modify after deleting/re-creating endpoint**

Path Restriction

- **Default configuration:**
  - All paths allowed, access control handled by the OS
- **Use `RestrictPaths` to customize**
  - Specifies a comma separated list of full paths that clients may access
  - Each path may be prefixed by R (read) and/or W (write), or N (none) to explicitly deny access to a path
  - '~' for authenticated user's home directory, and * may be used for simple wildcard matching.
- **e.g. Full access to home directory, read access to /data:**
  - `RestrictPaths = RW~,R/data`
- **e.g. Full access to home directory, deny hidden files:**
  - `RestrictPaths = RW~,N~/.*`

Sharing Path Restriction

- **Restrict paths where users can create shared endpoints**
- **Use `SharingRestrictPaths` to customize**
  - Same syntax as `RestrictPaths`
- **e.g. Full access to home directory, deny hidden files:**
  - `SharingRestrictPaths = RW~,N~/.*`
- **e.g. Full access to public folder under home directory:**
  - `SharingRestrictPaths = RW~/public`
- **e.g. Full access to /proj, read access to /scratch:**
  - `SharingRestrictPaths = RW/proj,R/scratch`

Deployment Best Practice: Science DMZ

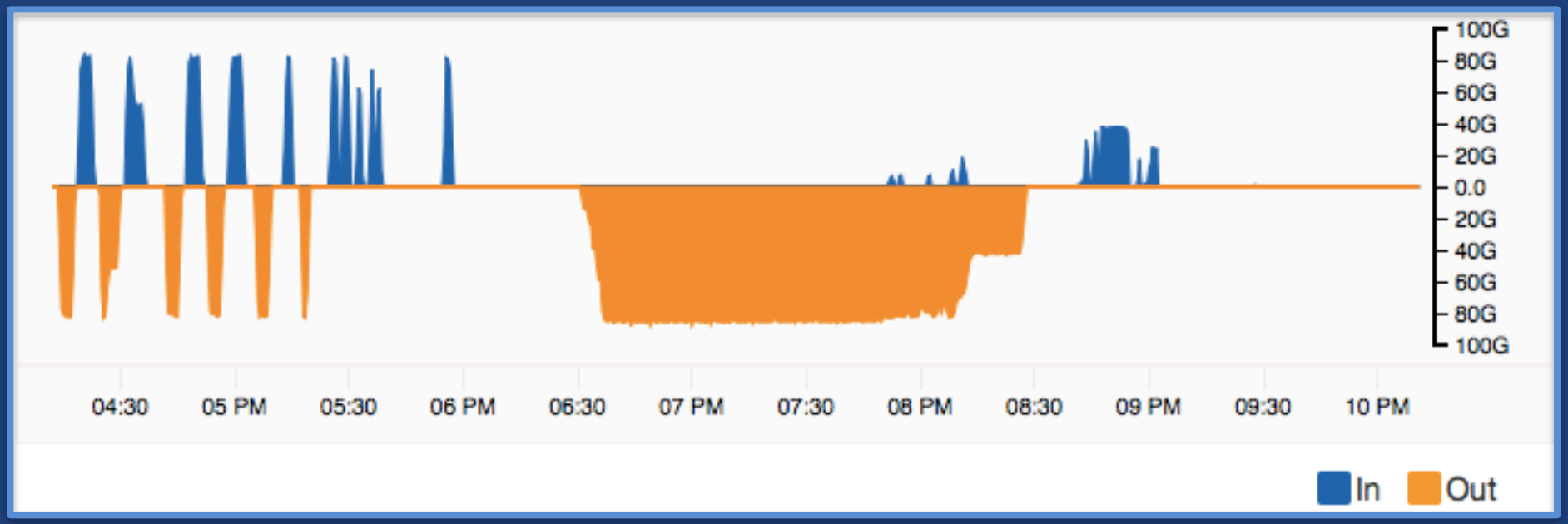
Balance: performance - reliability

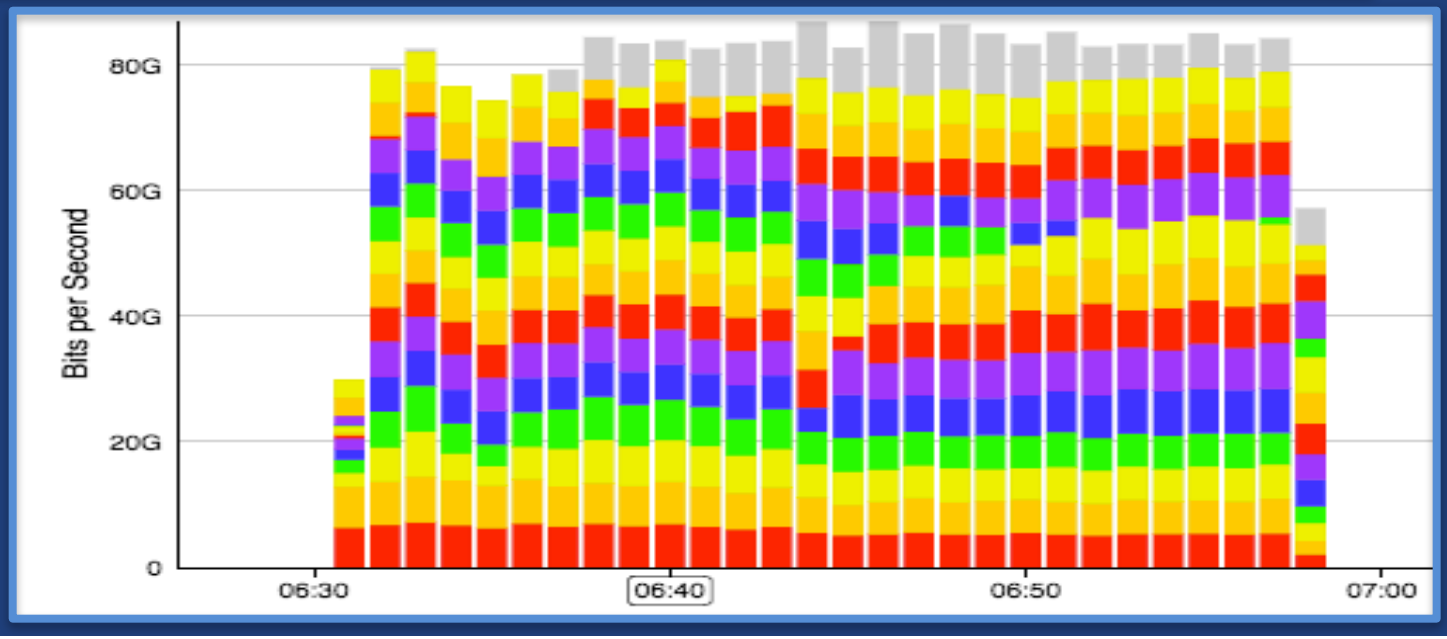
- **In-flight tuning based on transfer profile (#files, sizes)**
- **Request-specific overrides**
  - Concurrency
  - Pipeline depth
  - Parallelism
- **Endpoint-specific overrides; especially useful for multi-DTN deployments**
- **Service limits, e.g. concurrent requests**

Illustrative performance

- **10x scp throughput (typical)**
  - 20-100x demonstrated
- **On par/faster than UDP based tools (NASA JPL and anecdotal)**
  - Globus optimizations help on
  - UDP can outperform on high-latency links
- **AWS S3 uploads: ~3.5Gbps sustained, 5.2Gbps peak (10Gbps I/O)**

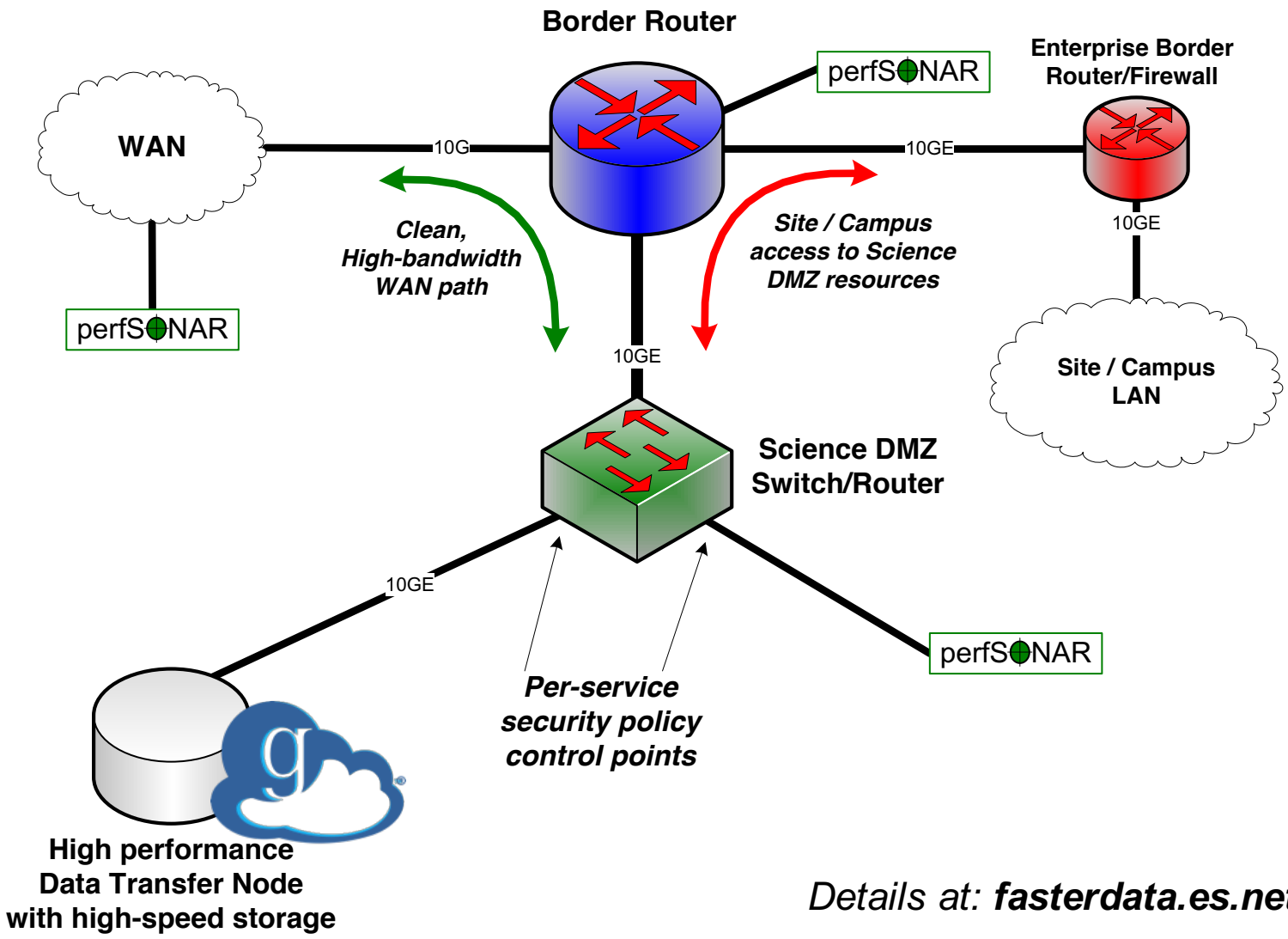
Extreme example: ~85 Gbps sustained disk-to-disk, Ottawa–New Orleans



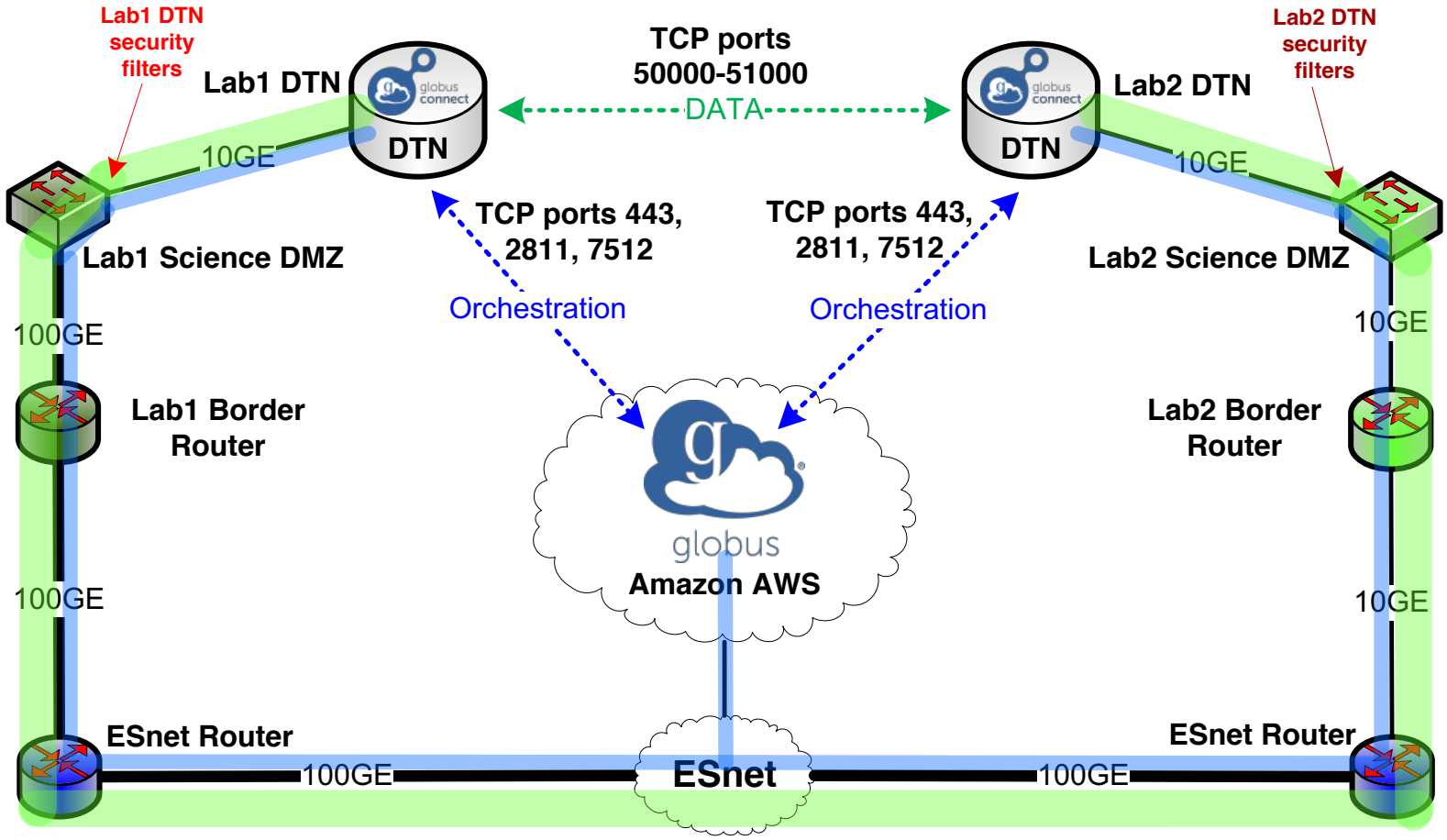


Raj Kettimuthu,
Argonne team
@SC14

Best-practice deployment



Network paths (two-DTN example)



Globus Platform-as-a-Service

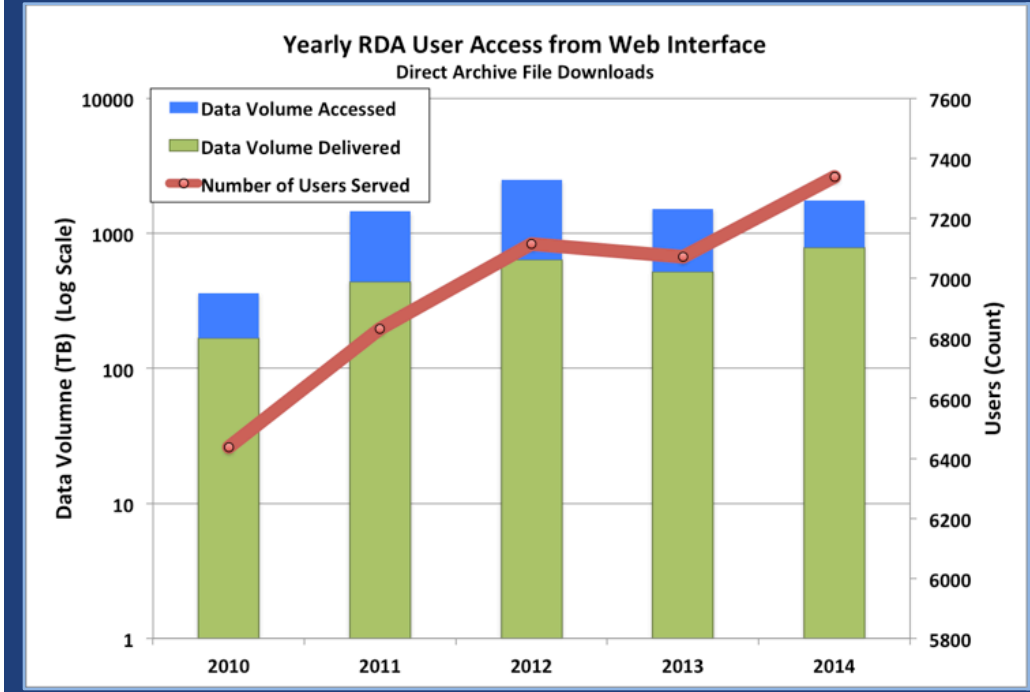
Globus PaaS at NCAR

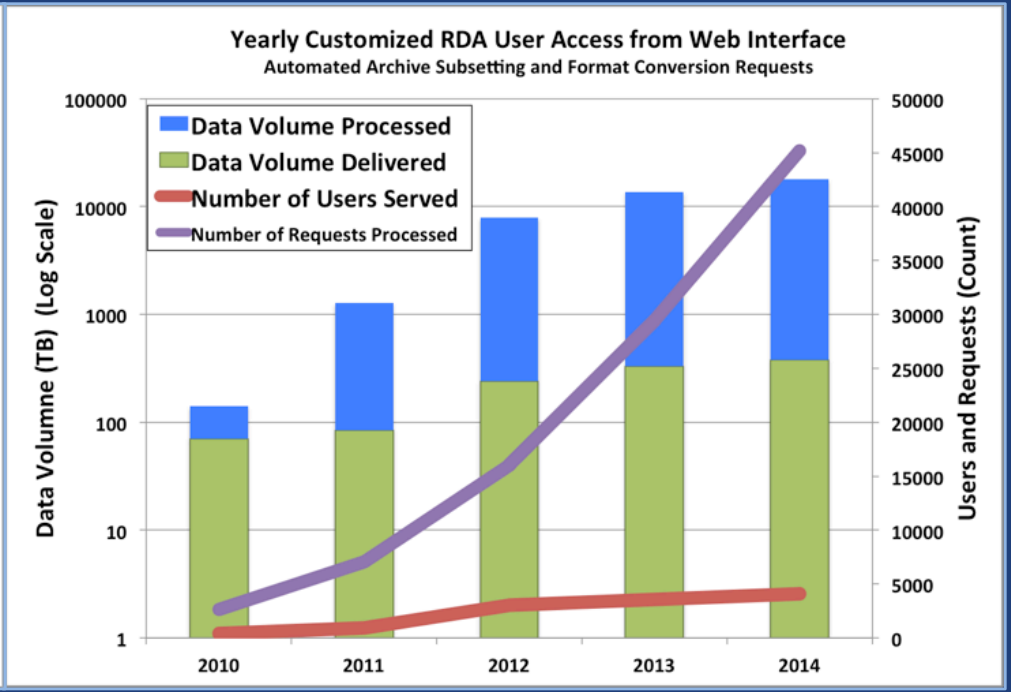
What is the RDA?

- **Free and open access to 600+ datasets for climate and weather research**
- **Worldwide usage**
- **Multiple data access pathways**
  - HTTP (wget, cURL, etc.)
  - OPeNDAP, WCS, WMS
  - Web services (CLI, API)
  - Analysis on HPC systems (NCAR users)

Courtesy of Thomas Cram, NCAR

RDA Usage





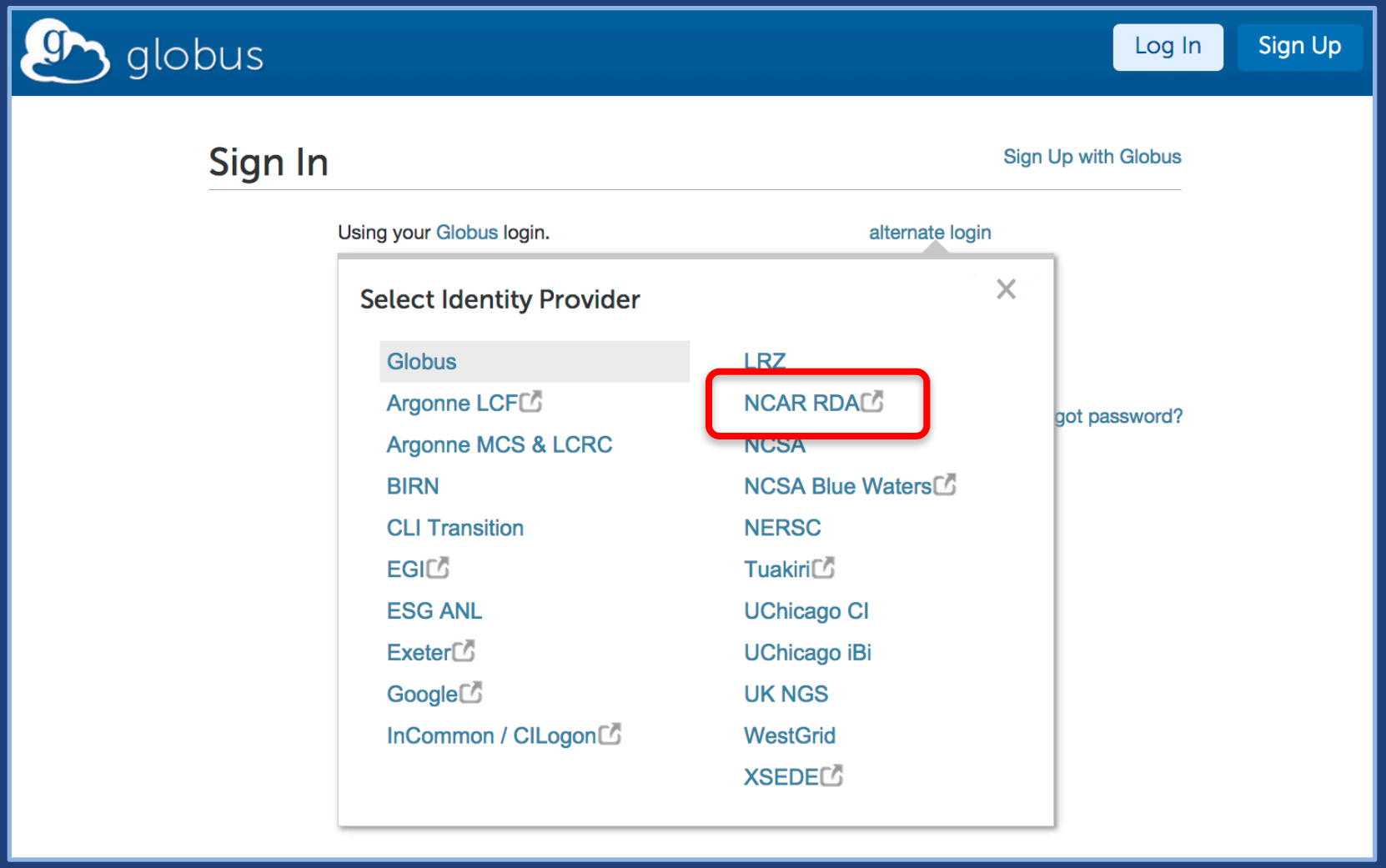
- **2014**
  - 17+ PB virtual processing
  - Web downloads: 7300 users, 750 TB served
  - 45,000 custom orders, 4000 users, 380 TB served

Courtesy of Thomas Cram, NCAR

Globus PaaS at NCAR

- **Single shared endpoint**
- **Data copied to subdirectories under endpoint source path**
- **Allow read permission to subdirectories under the shared endpoint**
- **ACLs managed programatically via Globus CLI**
- **Single sign-on using NCAR credentials**

RDA Alternate Identity login



Courtesy of Thomas Cram, NCAR

RDA Alternate Identity login

NCAR Research Data Archive (RDA)
MyProxy Client Authorization

Welcome to the NCAR RDA OAuth for MyProxy Client Authorization Page. The Client below is requesting access to your account. If you approve, please sign in with your RDA email/username and RDA password.

Client Information

Name: Globus Online
URL: https://www.globusonline.org

NCAR RDA Email/Username tcram@ucar.edu

NCAR RDA Password ••••••••••••

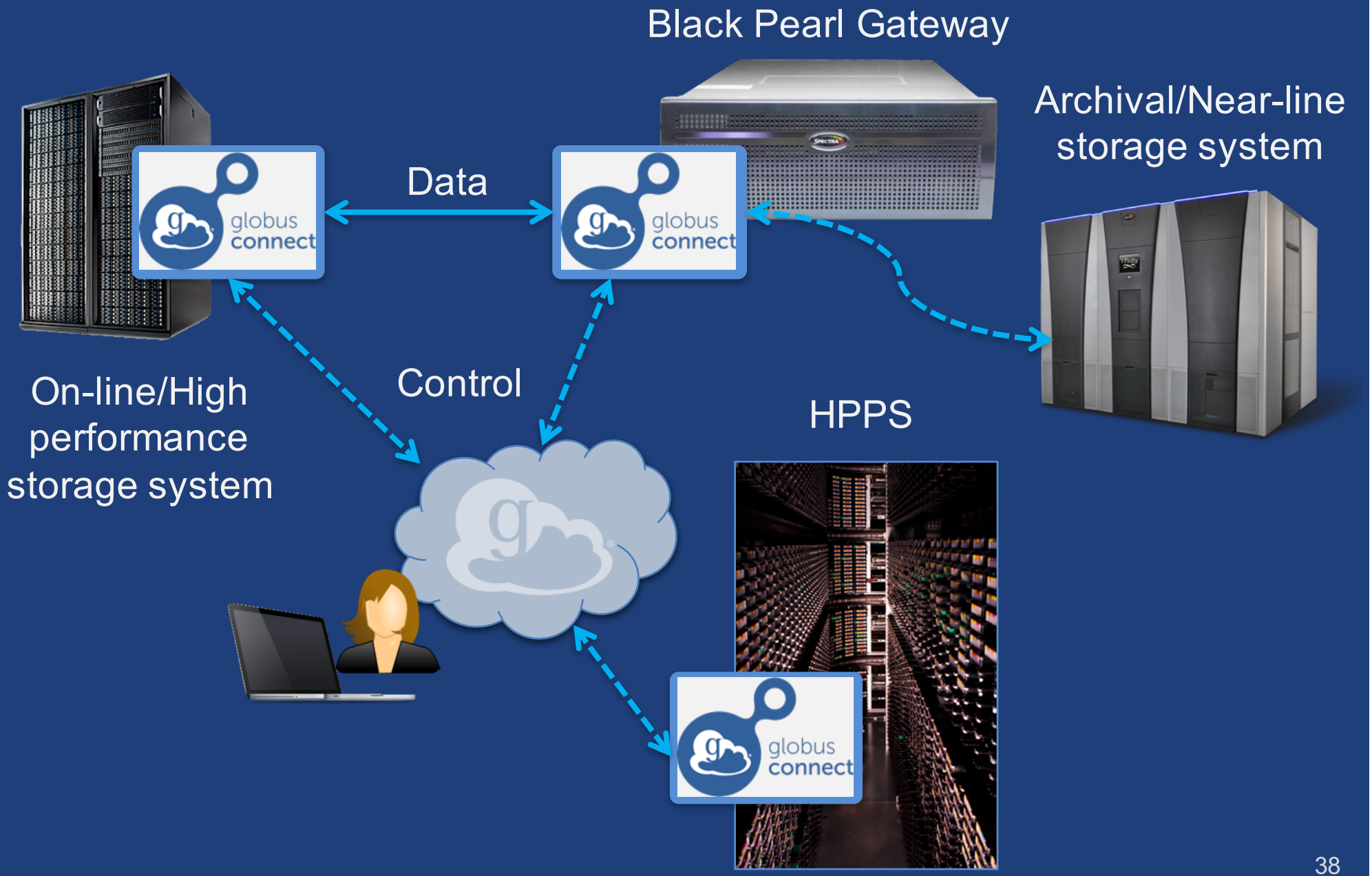
Sign In Cancel

Courtesy of Thomas Cram, NCAR

Based on widely used web standards

- **OAuth 2.0 Authorization Framework**
  - Globus Auth is an OAuth2 authorization server
- **OpenID Connect Core 1.0**
  - Globus Auth is an OIDC claims provider
- **Allows use of standard OAuth2 and OIDC libraries**
  - e.g., Google OAuth Client Libraries (Java, Python, etc.), Apache mod_auth_openidc

Bridging the storage hierarchy



Bridging the campus to AWS

Campus Cluster,
Sequencing Core

Analysis
Targets

globus
genomics

EC2

S3

Public
Datasets

Faculty Storage

Burst
Storage

globus

Digital Repository

Publication,
Archival

- Native S3 support; IAM cross-account access
- OpenStack Ceph support (Jetstream)

Our AWS operations infrastructure

Elastic Compute Cloud: Scalable runtime infrastructure for all Globus services

Virtual Private Cloud: Isolation and protection of the Globus production runtime environment

Simple Storage Service: Highly durable object store for all Globus static artifacts

Identity and Access Management: Fine-grained authN/authZ for operations/administration

Relational Database Service: Reliable, high-performance database for Globus backend services

Route 53: Highly available name resolution and routing for all Globus services

Simple Email Service: Simple notifications utility for Globus backend services

Coming soon

- **HTTP/HTTPS support**
- **OpenStack Ceph endpoints (Jetstream)**

Enable your storage system

- Signup: **globus.org/signup**
- Create endpoint: **globus.org/globus-connect-server**
- Need help? **support.globus.org**
- Subscribe to help us make Globus self-sustaining: **globus.org/provider-plans**
- Follow us: **@globusonline**