



Managing Globus Endpoints

Globus for System Administrators

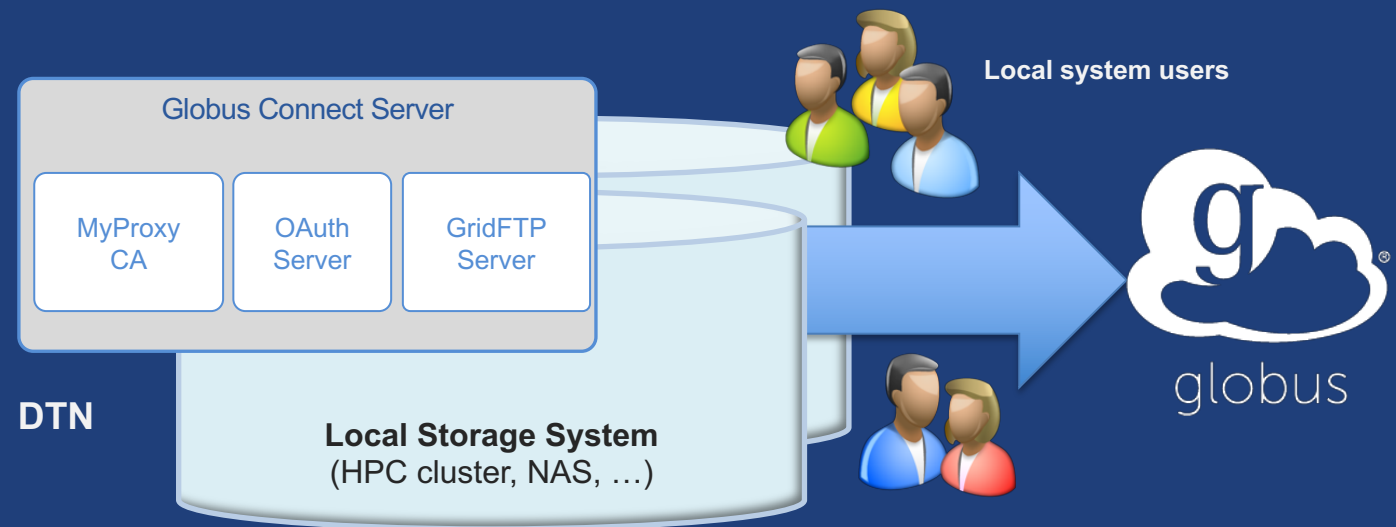
Vas Vasiliadis
vas@uchicago.edu

NYSERNet – May 1, 2018



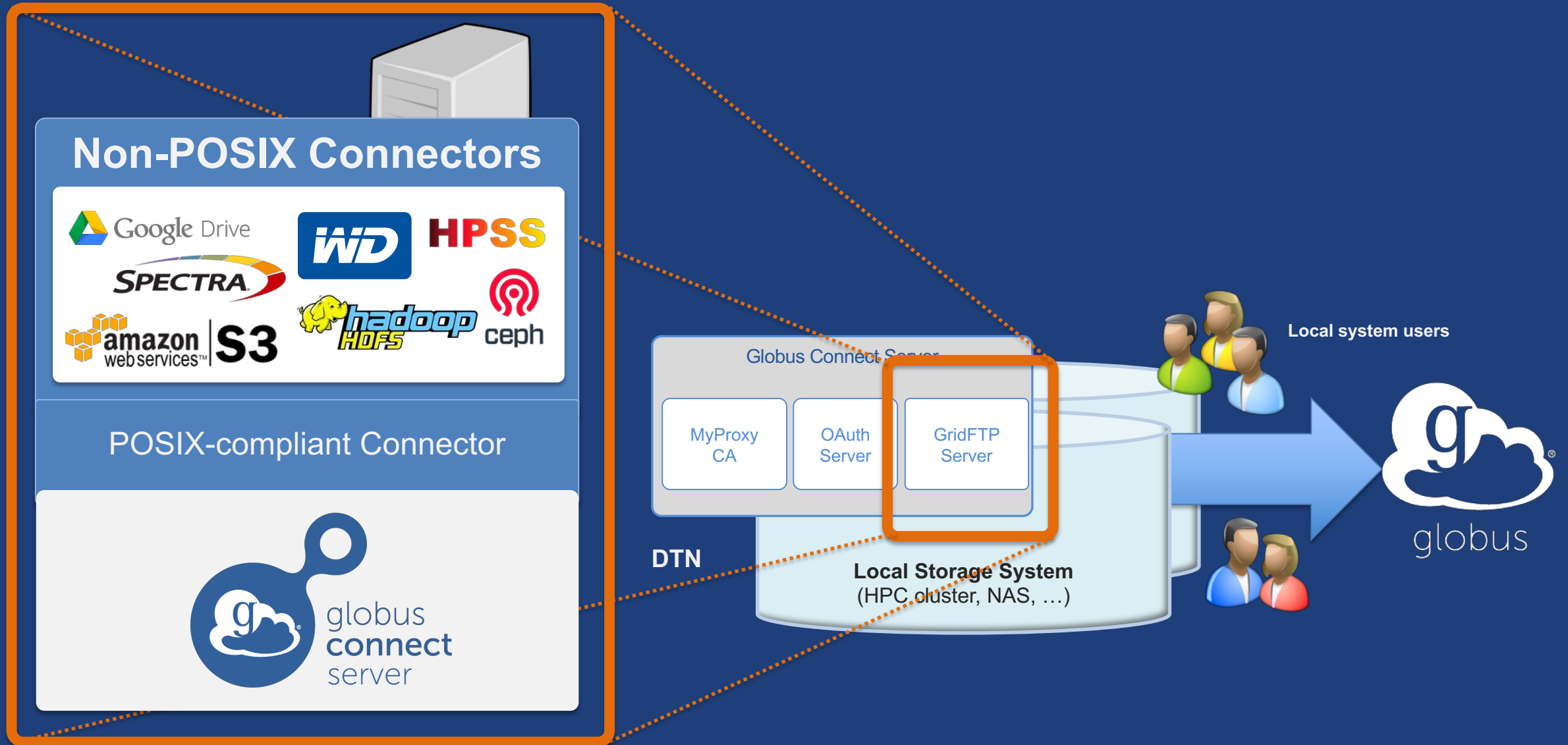
Globus Connect Server

- **Makes your storage accessible via Globus**
- **Multi-user server, installed and managed by sysadmin**
- **Default access for all local accounts**
- **Native packaging
Linux: DEB, RPM**



docs.globus.org/globus-connect-server-installation-guide/

Globus Connect Server

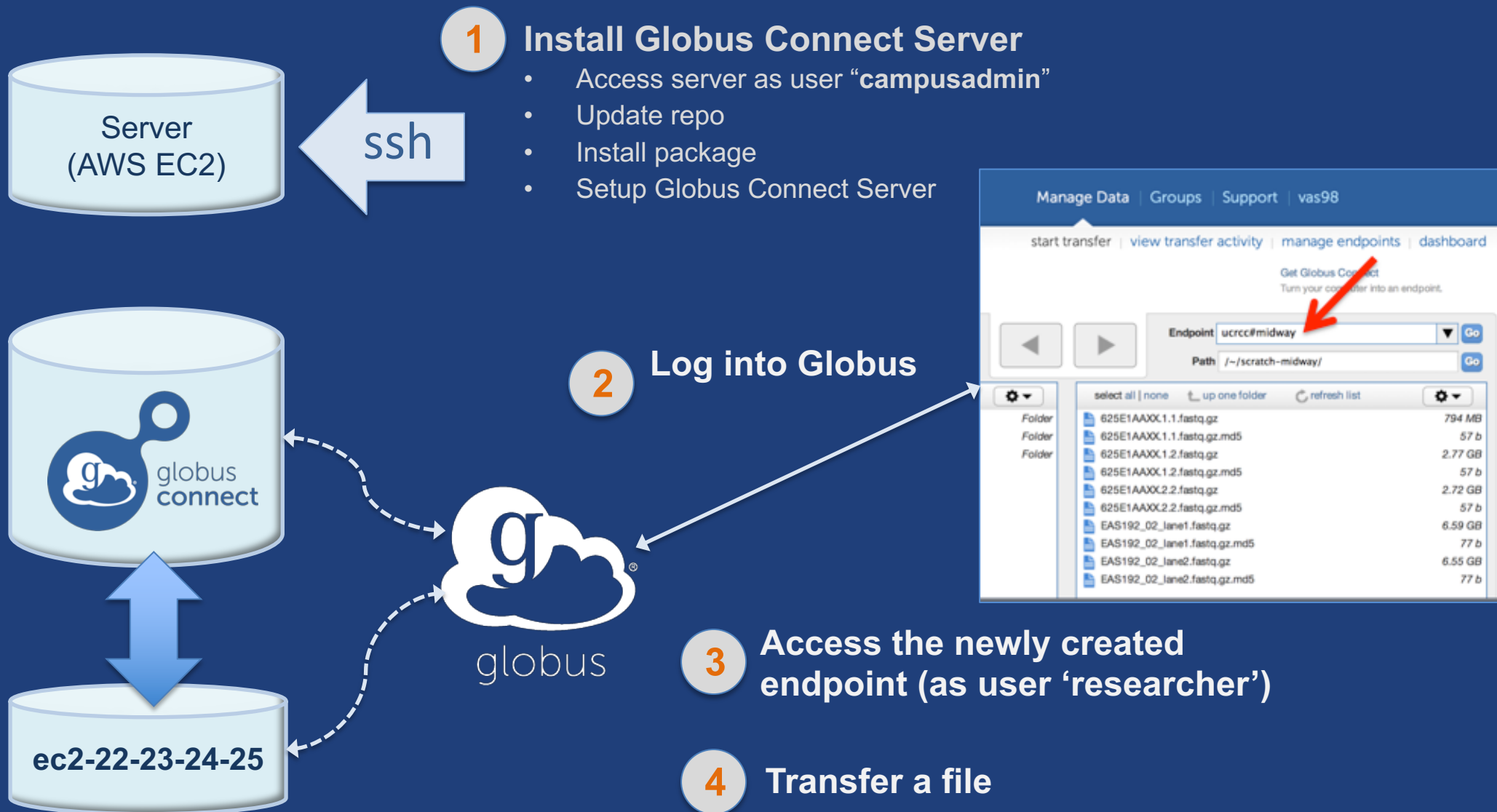


Creating a Globus endpoint on your server

- **In this example, Server = Amazon EC2 instance**
- **Installation and configuration of Globus Connect Server requires a Globus ID**
- **Go to `globusid.org`**
- **Click “create a Globus ID”**
 - Optional: associate it with your Globus account



What we are going to do:





Access your server

- **Get the IP address for your EC2 server (bit.ly/ec2ip)**
- **Log in as user 'campusadmin'**
`ssh campusadmin@<EC2_instance_IP_address>`
- **Please sudo su before continuing**
 - User 'campusadmin' has passwordless sudo privileges



Install Globus Connect Server

```
$ sudo su
$ curl -L0s 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 when prompted

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 host name in the Endpoint field
 - Log in as “researcher”; you will see the user’s home directory
- **Transfer files to/from a test endpoint (e.g. Globus Tutorial) and your EC2 endpoint**

Globus accounts and endpoint access

- **Globus account: Primary identity (+ Linked Identities)**
- **Endpoint initially accessible by creator**
- **Endpoint not visible?**
 - Primary identity is your institutional ID?
 - Link your Globus ID!



Configuring Globus Connect Server



Endpoint configuration

- **Globus service “Manage Endpoints” page**
- **DTN (Globus Connect Server) config**
 - `/etc/globus-connect-server.conf`
 - Standard .ini format: `[Section] Option = value`
 - To enable changes you must run:
`globus-connect-server-setup`
 - “Rinse and repeat”



Common configuration options

- **Manage Endpoints page**
 - Display Name
 - Visibility
 - Encryption
- **DTN configuration file**
 - RestrictPaths
 - IdentityMethod (CILogon, OAuth)
 - Sharing
 - SharingRestrictPaths



Exercise: Make your endpoint visible

- **Edit endpoint attributes**
 - Change the name to something useful, e.g. <your_name> EC2 Endpoint
 - For the “Visible To” attribute select “Public - Visible to all users”
- **Find your neighbor’s endpoint**
 - You can access it too 😊



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~/.*



Exercise: Restrict access

- **Set** `RestrictPaths=RW~,N~/archive`
- Run `globus-connect-server-setup`
- **Access your endpoint as 'researcher'**
- **What's changed?**



Enabling sharing on an endpoint

- In config file, set `Sharing=True`
- Run `globus-connect-server-setup`
- Use the CLI to flag as managed endpoint (also configurable via the web app)

* Note: Creation of shared endpoints requires a Globus subscription for the managed endpoint

Limit sharing to specific accounts

- `SharingUsersAllow =`
- `SharingGroupsAllow =`
- `SharingUsersDeny =`
- `SharingGroupsDeny =`



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`



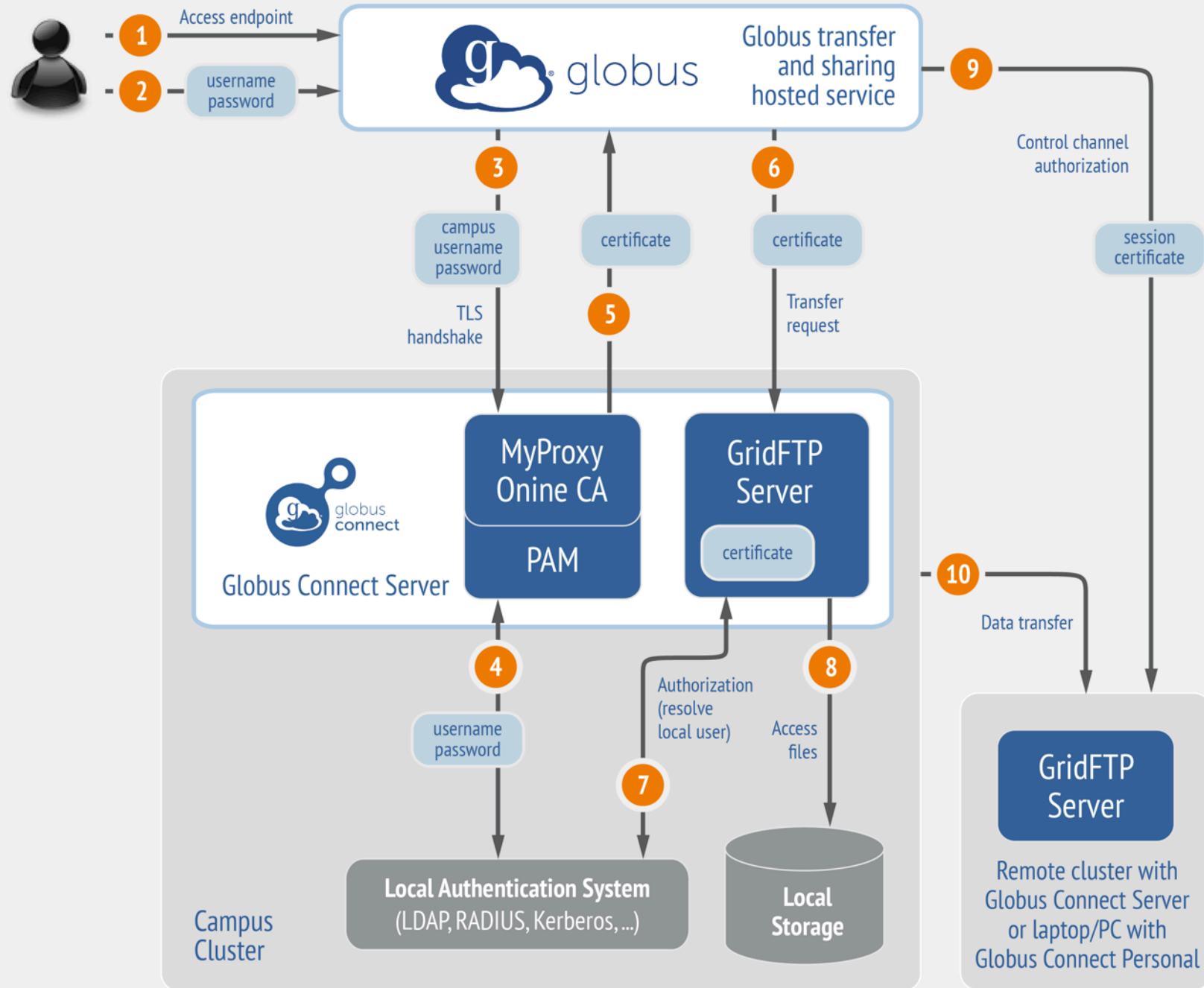
Accessing Endpoints

Ports needed for Globus

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



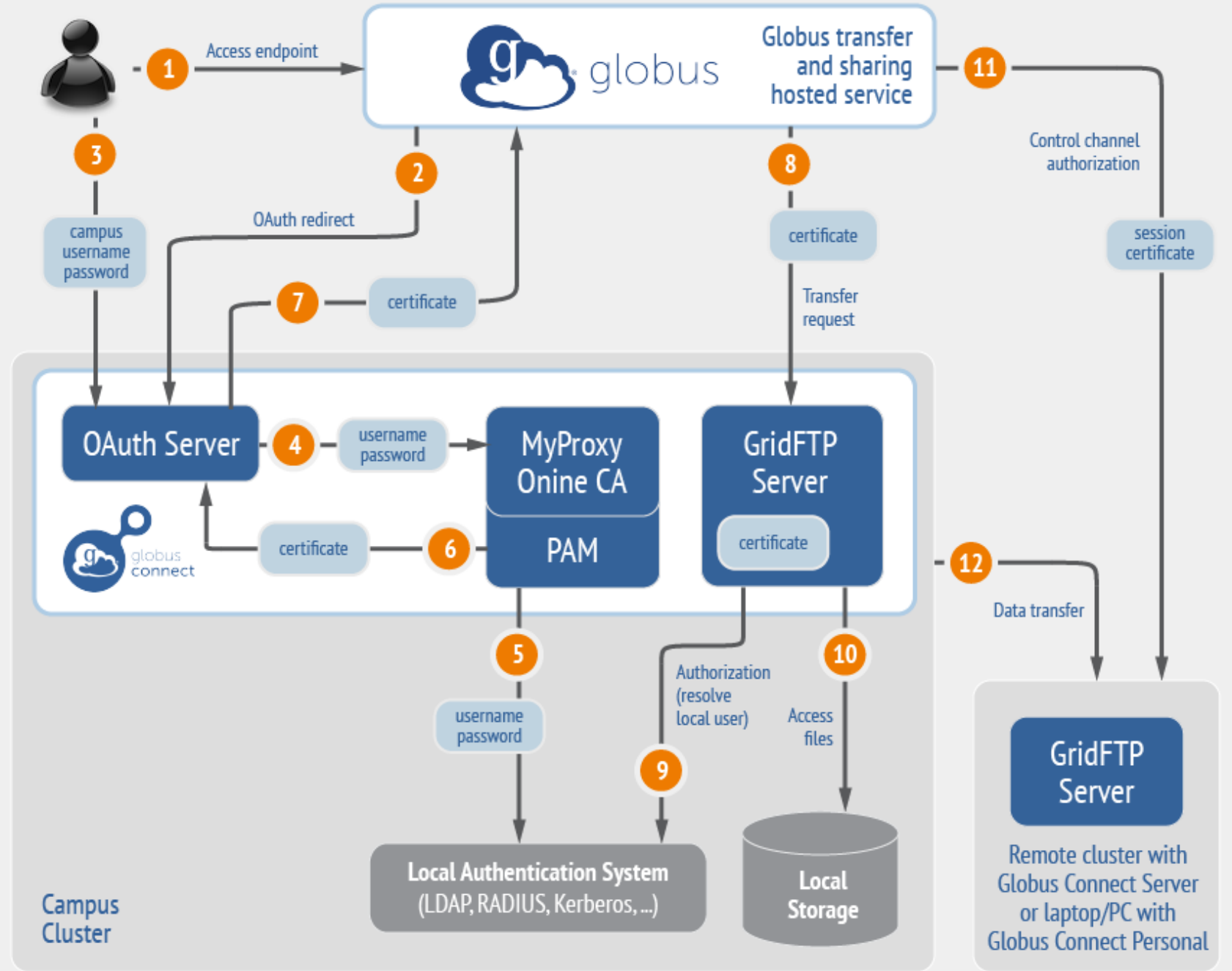
Endpoint activation using MyProxy



Default configuration
(avoid if at all possible)



Endpoint activation using MyProxy OAuth



Best practice configuration

Single Sign-On with InCommon/CILogon

- **Your Shibboleth server must release R&S attributes to CILogon—especially the ePPN attribute**
- **Local resource account names must match your institutional ID (InCommon ID)**
- **In `/etc/globus-connect-server.conf` set:**

```
AuthorizationMethod = CILogon
```

```
CILogonIdentityProvider =  
<institution_listed_in_CILogon_IdP_list>
```



Managed endpoints and subscriptions



Subscription configuration

- **Subscription manager**
 - Create/upgrade managed endpoints
 - Requires Globus ID linked to Globus account
- **Management console permissions**
 - Independent of subscription manager
 - Map managed endpoint to Globus ID
- **Globus Plus group**
 - Subscription Manager is admin
 - Can grant admin rights to other members



Creating managed endpoints

- **Required** for sharing, management console, reporting, ...
- **Convert existing endpoint to managed via CLI (or web):**
`globus endpoint update --managed <endpt_uuid>`
- **Must be run by subscription manager**
- **Important: Re-run endpoint update after deleting/re-creating endpoint**



Monitoring and managing Globus endpoint activity

Management console

- **Monitor all transfers**
- **Pause/resume specific transfers**
- **Add pause conditions with various options**
- **Resume specific tasks overriding pause conditions**
- **Cancel tasks**
- **View sharing ACLs**



Endpoint Roles

- **Administrator:** define endpoint and roles
- **Access Manager:** manage permissions
- **Activity Manager:** perform control tasks
- **Activity Monitor:** view activity



Demonstration:
Management console
Endpoint Roles
Usage Reporting



...on performance



Balance: performance - reliability

- **Network use parameters: concurrency, parallelism**
- **Maximum, Preferred values for each**
- **Transfer considers source and destination endpoint settings**

```
min(  
    max(preferred src, preferred dest),  
    max src,  
    max dest  
)
```

- **Service limits, e.g. concurrent requests**



Illustrative performance

Petascale DTN Project

November 2017

L380 Data Set

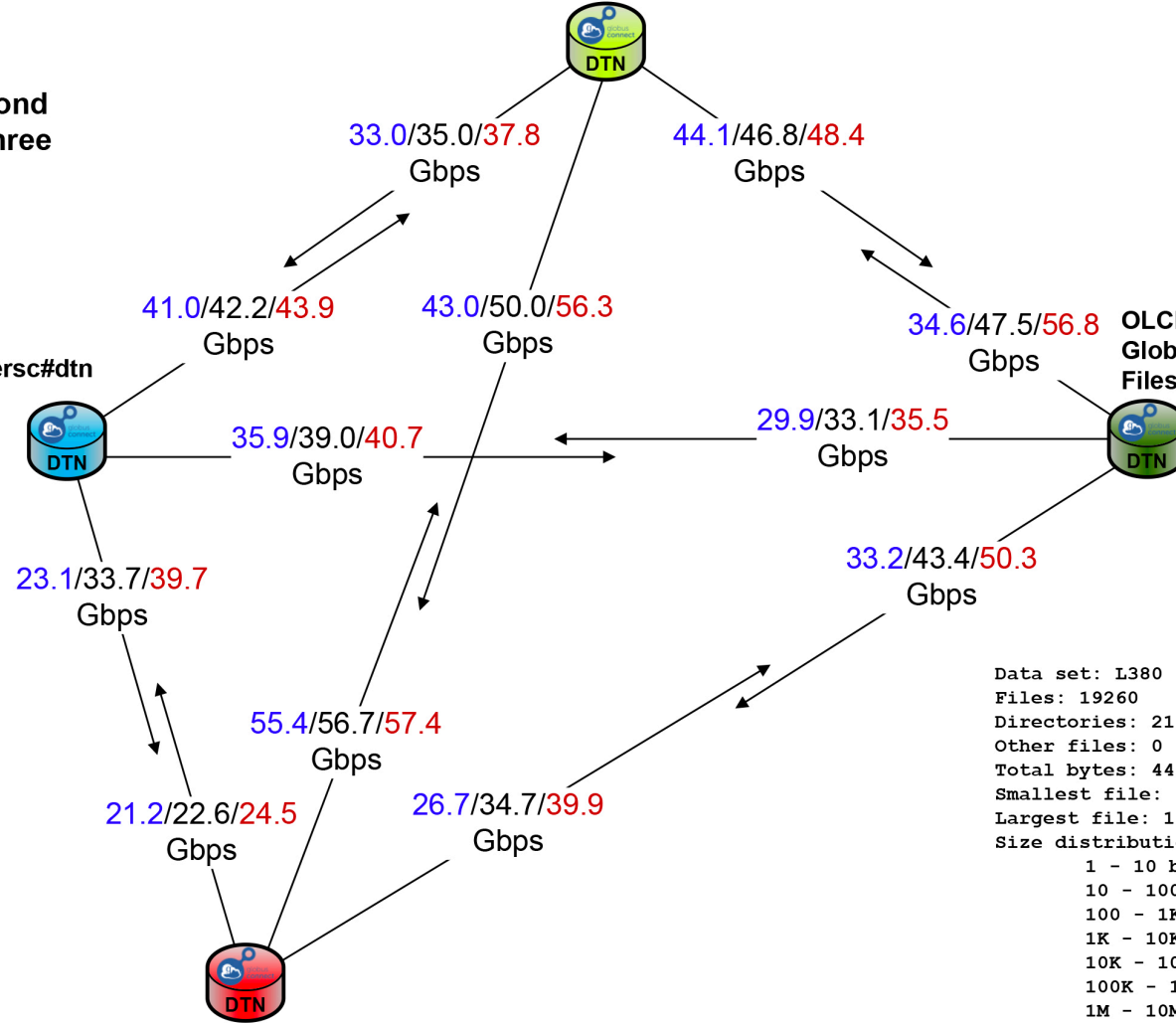
Gigabits per second
(min/avg/max), three transfers

NERSC DTN cluster
Globus endpoint: nersc#dtn
Filesystem: /project

ALCF DTN cluster
Globus endpoint: alcf#dtn_mira
Filesystem: /projects

OLCF DTN cluster
Globus endpoint: olcf#dtn_atlas
Filesystem: atlas2

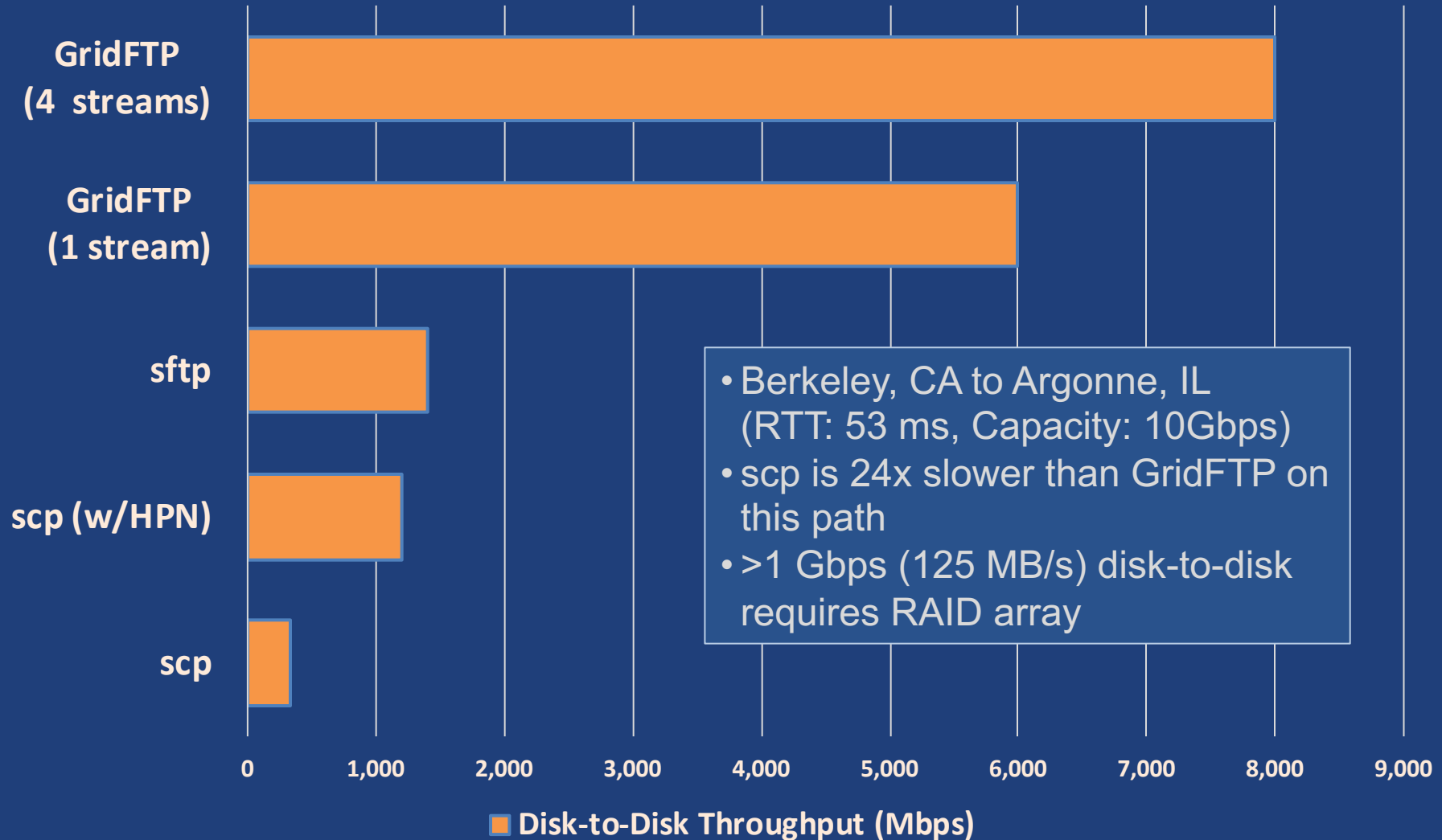
NCSA DTN cluster
Globus endpoint: ncsa#BlueWaters
Filesystem: /scratch



Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.4T bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313896248 bytes (11G bytes)
Size distribution:
1 - 10 bytes: 7 files
10 - 100 bytes: 1 files
100 - 1K bytes: 59 files
1K - 10K bytes: 3170 files
10K - 100K bytes: 1560 files
100K - 1M bytes: 2817 files
1M - 10M bytes: 3901 files
10M - 100M bytes: 3800 files
100M - 1G bytes: 2295 files
1G - 10G bytes: 1647 files
10G - 100G bytes: 3 files



Disk-to-Disk Throughput: ESnet Testing

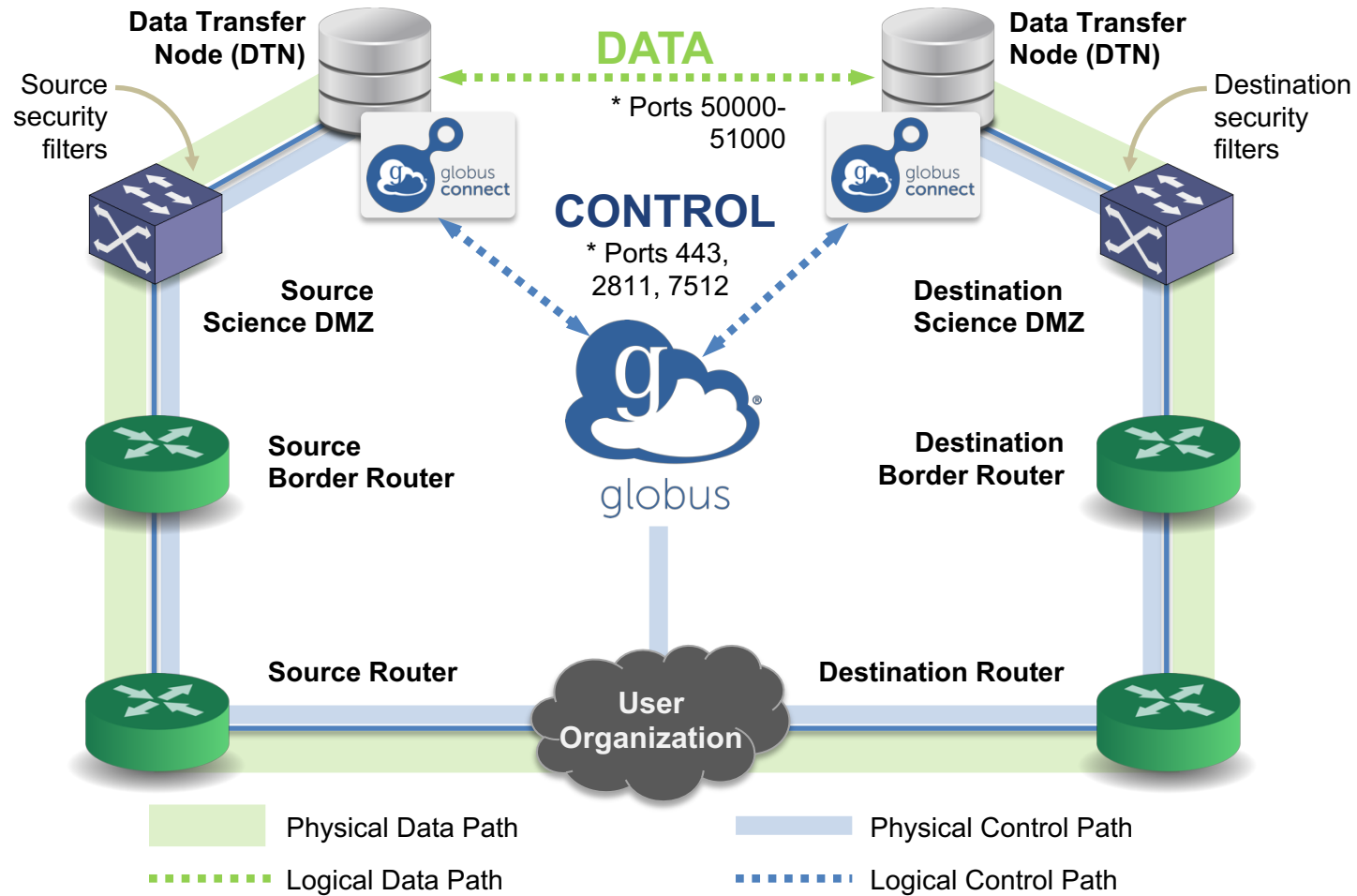




Deployment Scenarios



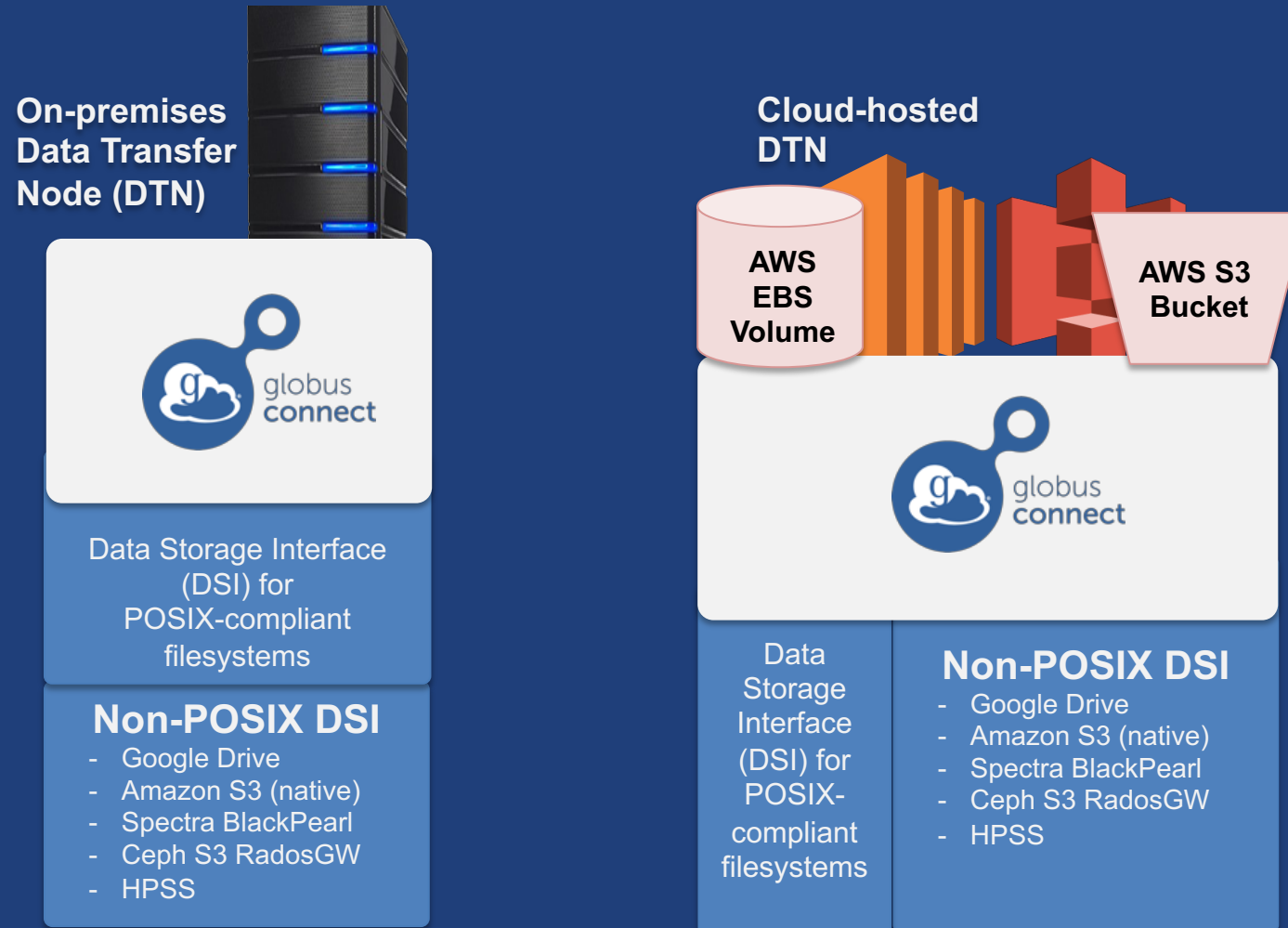
Best practice network configuration



* Please see TCP ports reference: https://docs.globus.org/resource-provider-guide/#open-tcp-ports_section

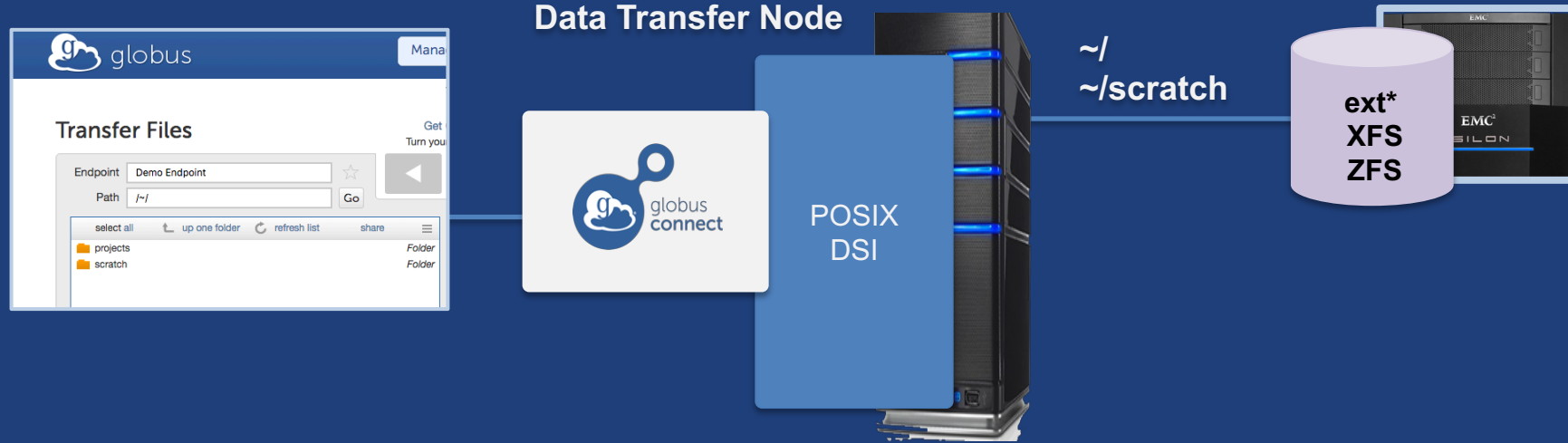


The Data Transfer Node



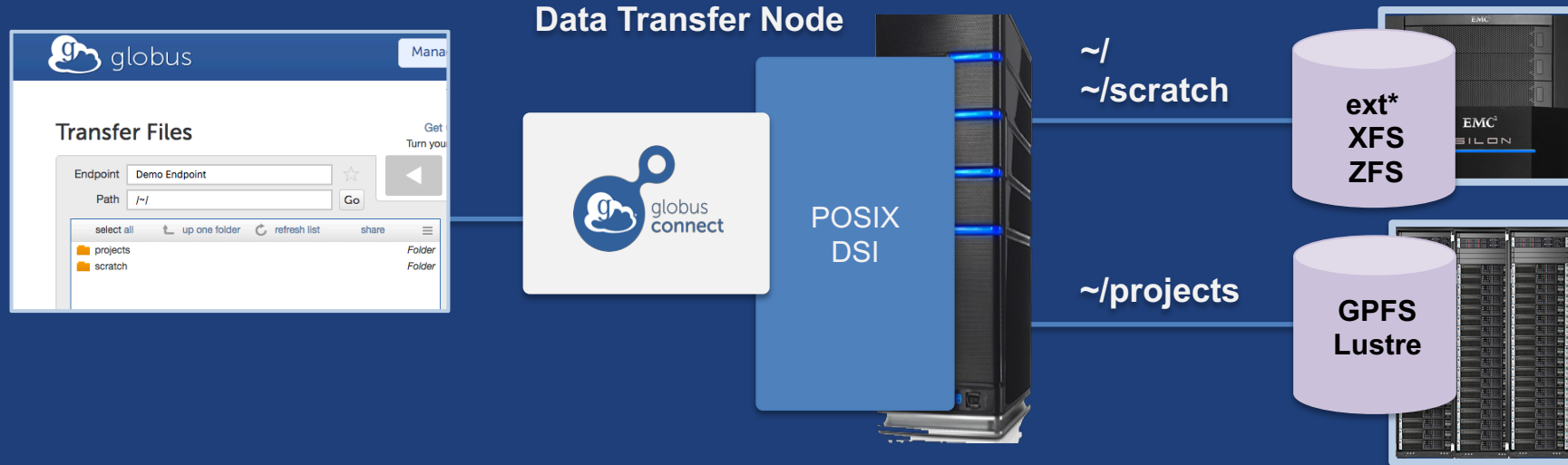


Multi-endpoint configuration



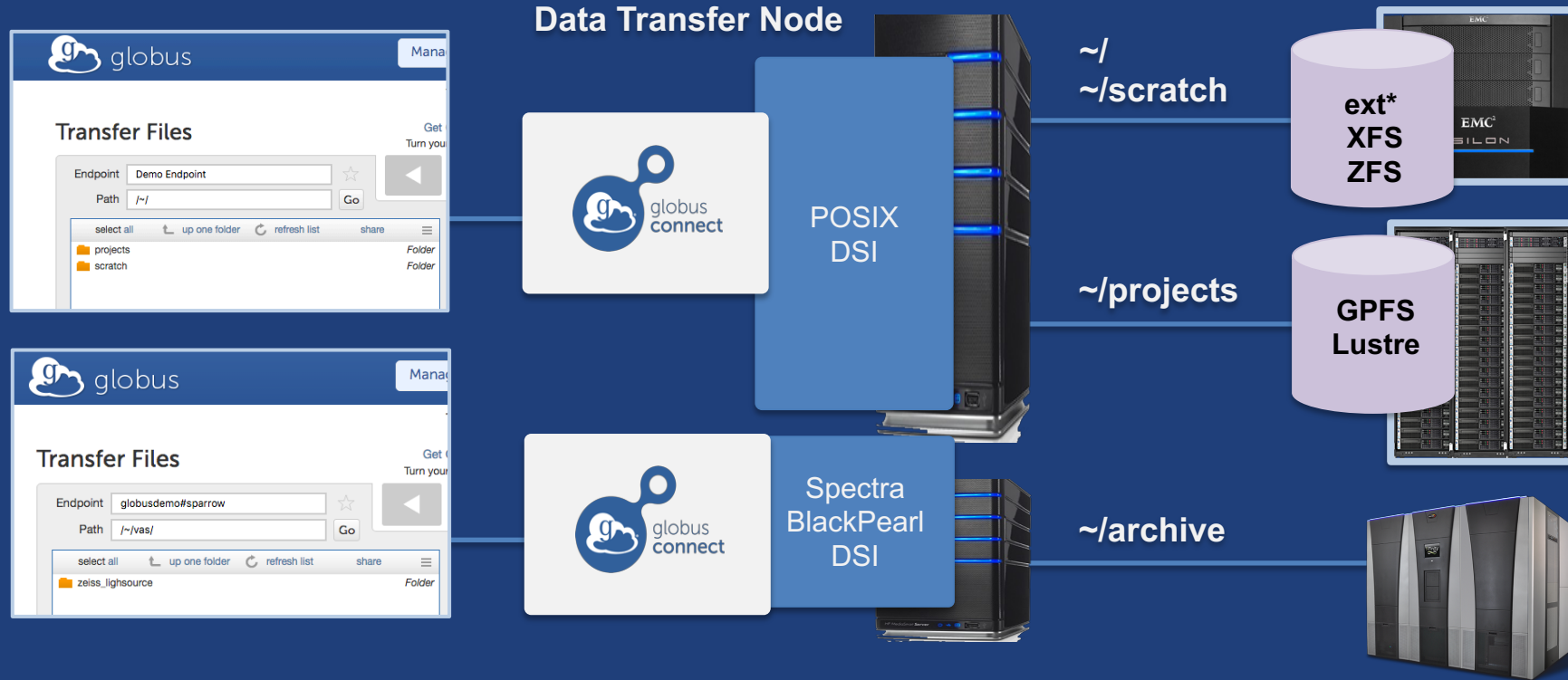


Multi-endpoint configuration



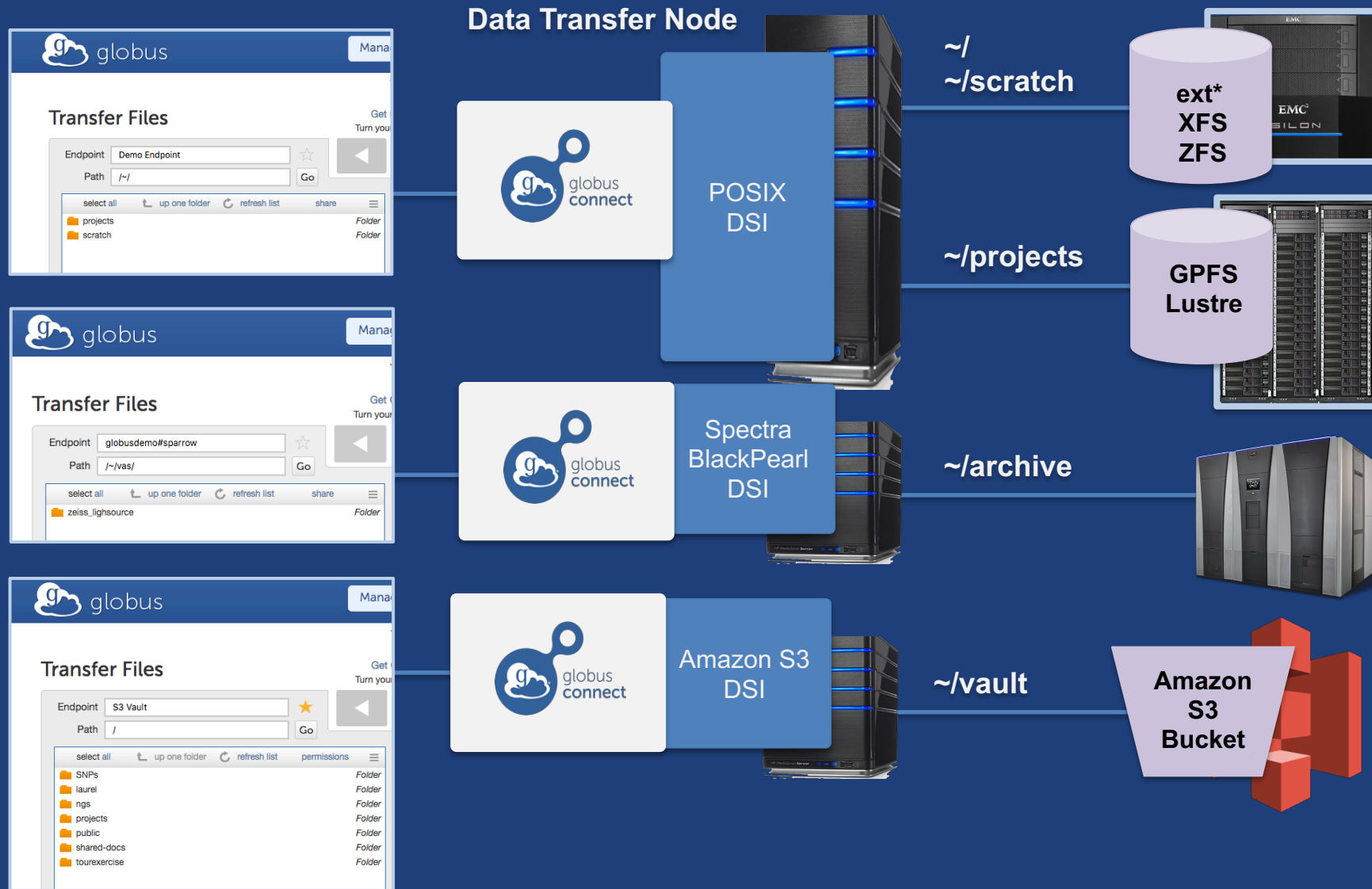


Multi-endpoint configuration

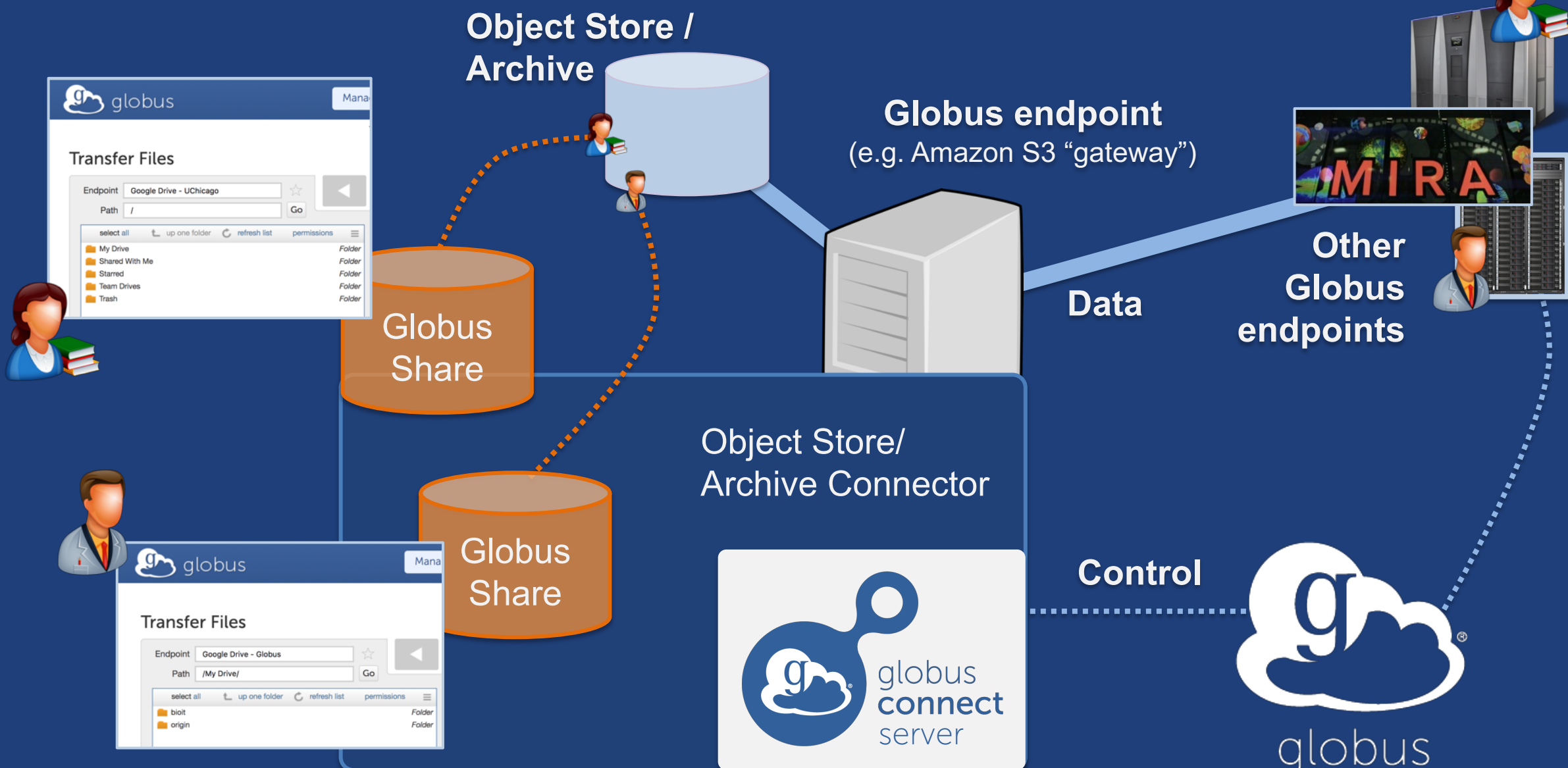




Multi-endpoint configuration



Deploying a premium connector gateway



- **Turnkey on-premise object storage**
- **Globus connector using S3 API**
- **Low TCO: Manufactures own drives**
- **Erasur coding**
- **Auto data integrity checks with self-healing**
- **Cloud-based systems management tools**
- **Data Forever: automatic migration to new tech**



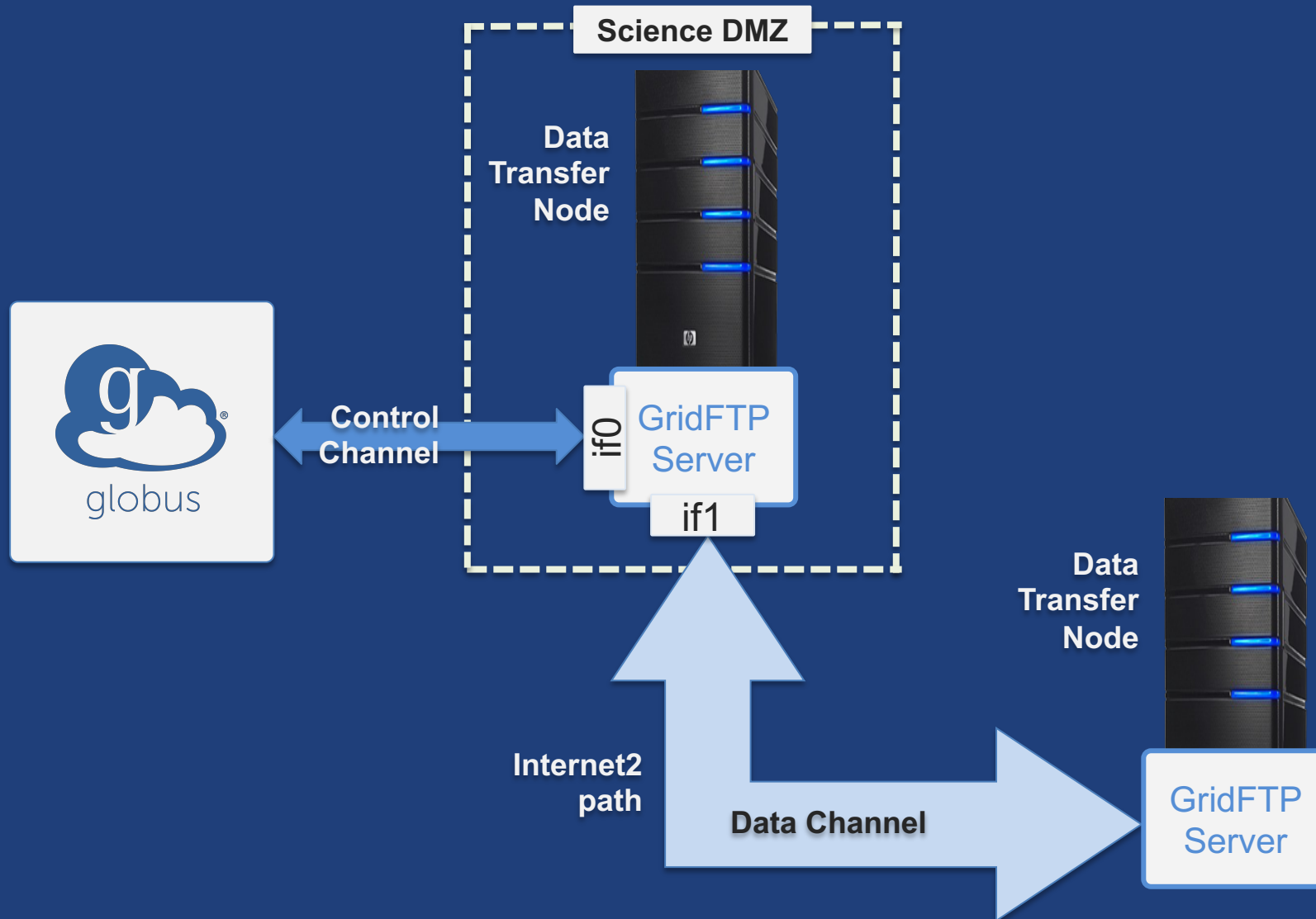
<https://docs.globus.org/premium-storage-connectors/wd-activescale/>



Network paths

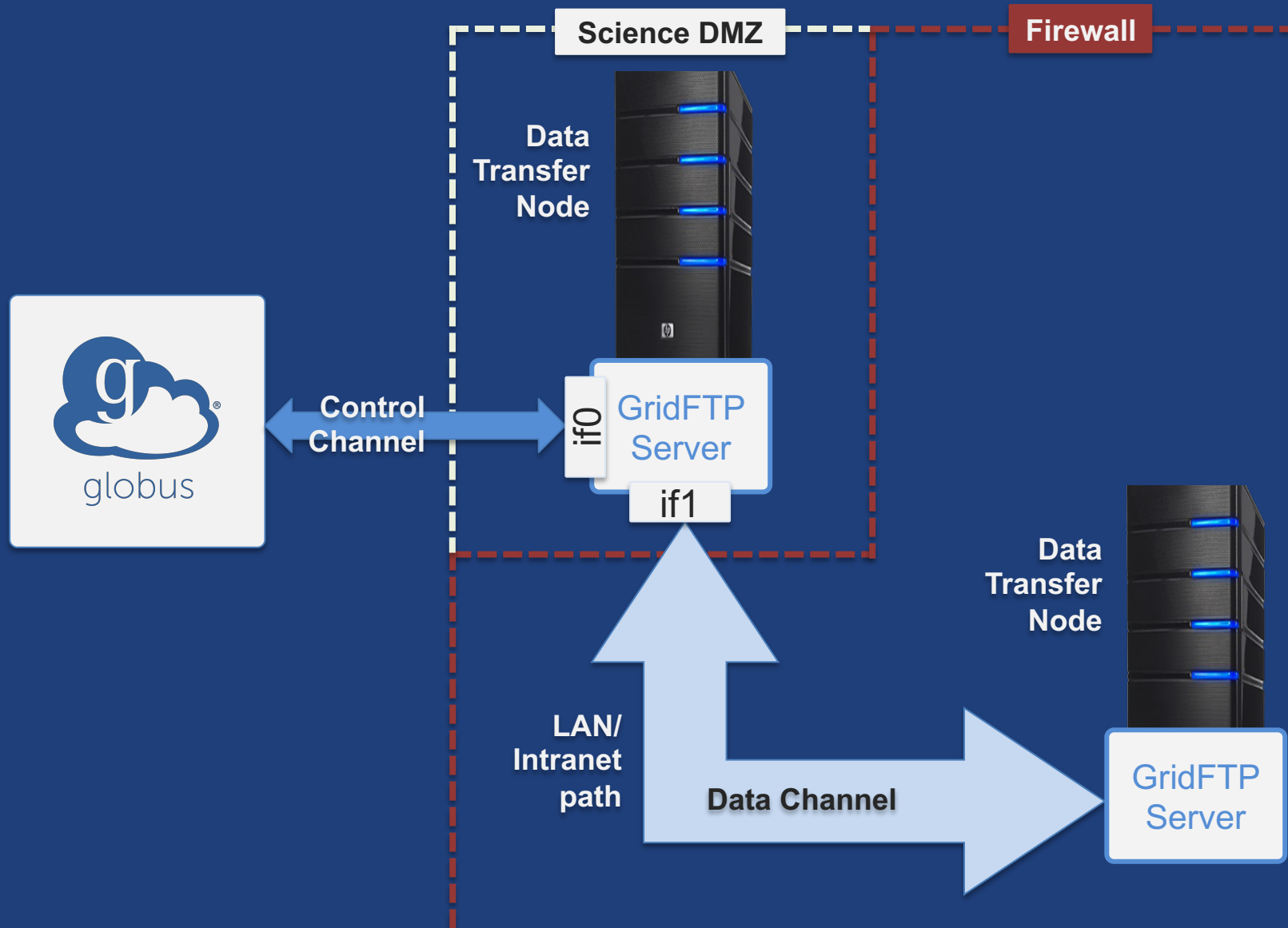
- **Separate control and data interfaces**
- **"DataInterface =" option in globus-connect-server-conf**
- **Common scenario: route data flows over Science DMZ link**

Dual-homed DTN – high speed data path





Dual-homed DTN – private network data path





Other Deployment Options



Encryption

- **Requiring encryption on an endpoint**
 - User cannot override
 - Useful for “sensitive” data
- **Globus uses OpenSSL cipher stack as currently configured on your DTN**
- **FIPS 140-2 compliance: ensure use of FIPS capable OpenSSL libraries on DTN**
 - <https://www.openssl.org/docs/fips/UserGuide-2.0.pdf>

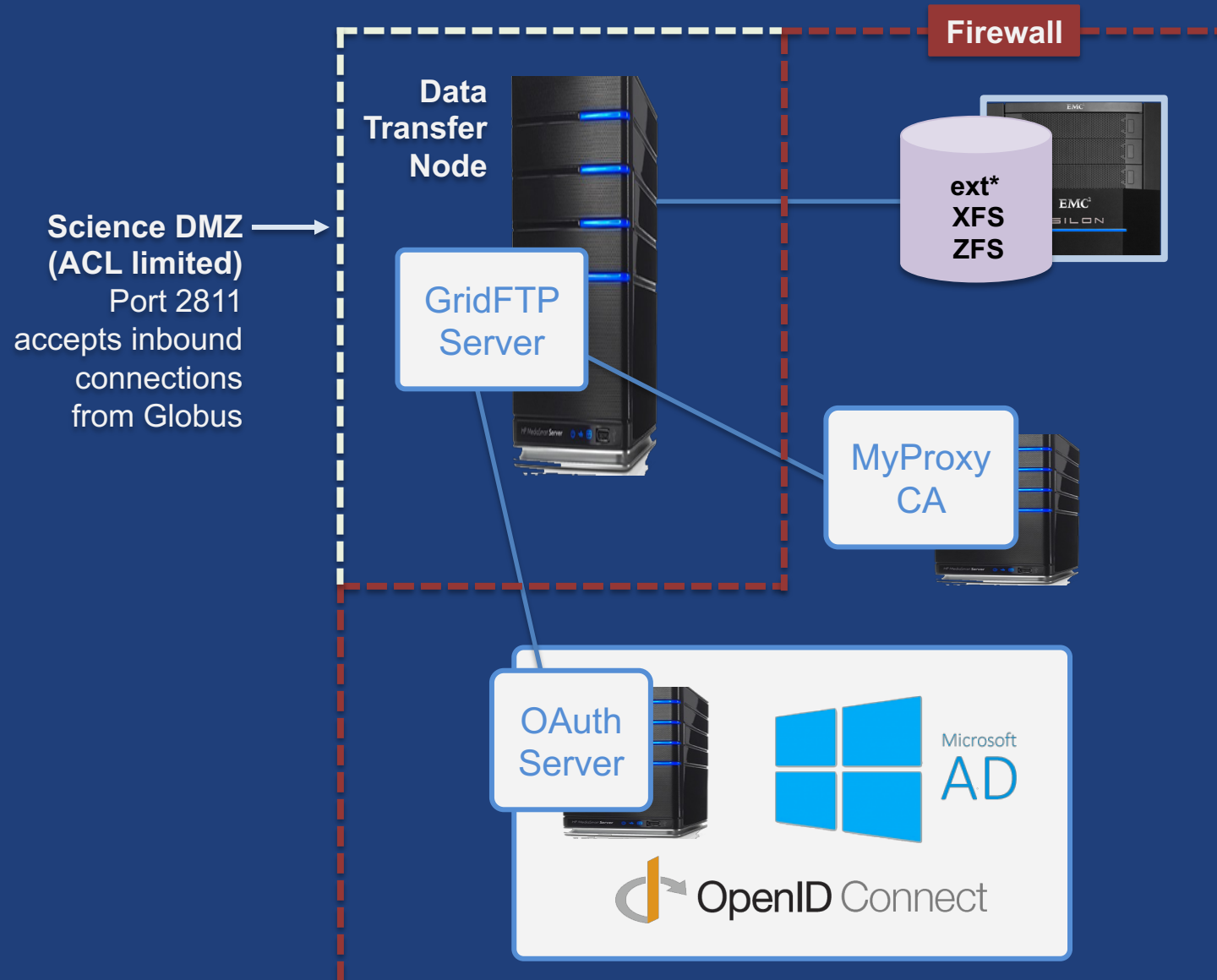


Distributing Globus Connect Server components

- **Globus Connect Server components**
 - globus-connect-server-io, -id, -web
- **Default: -io, -id and -web on single server**
- **Common options**
 - Multiple -io servers for load balancing, failover, and performance
 - No -id server, e.g. third-party IdP
 - -id on separate server, e.g. non-DTN nodes
 - -web on either -id server or separate server for OAuth interface



Distributing Globus Connect Server components





Setting up multiple `-io` servers

- **Guidelines**
 - Use the same `.conf` file on all servers
 - First install on the server running the `-id` component, then all others
- **Install Globus Connect Server on all servers**
- **Edit `.conf` file on one of the servers and set [MyProxy] Server to the hostname of the server you want the `-id` component installed on**
- **Copy the configuration file to all servers**
 - `/etc/globus-connect-server.conf`
- **Run `globus-connect-server-setup` on the server running the `-id` component**
- **Run `globus-connect-server-setup` on all other servers**
- **Repeat steps 2-5 as necessary to update configurations**



Example: Two-node DTN



On “primary” DTN node (34.20.29.57):
/etc/globus-connect-server.conf
[Endpoint] Name = **globus_dtn**
[MyProxy] Server = **34.20.29.57**



On other DTN nodes:
/etc/globus-connect-server.conf
[Endpoint] Name = **globus_dtn**
[MyProxy] Server = **34.20.29.57**



Globus Network Manager

For environments with super duper
special network constraints...
(a.k.a. "for the very brave")



Globus Network Manager

- **Information from GridFTP to facilitate dynamic network changes**
- **Callbacks during GridFTP execution on local DTN**
- **Supplements information available via Globus transfer API**



Globus Network Manager Callbacks

- **Pre-listen (binding of socket)**
- **Post-listen**
- **Pre-accept/Pre-connect (no Data yet)**
- **Post-accept/Post-connect (data in flight)**
- **Pre-close**
- **Post-close**



Network manager use cases

- **Science DMZ Traffic Engineering**
 - Use SDN to dynamically route data path
 - Control path uses traditional route
- **Automated WAN bandwidth reservation**
 - OSCARS, AL2S
- **Note: All this requires custom code**



Future directions



New Globus web app

app.globus.org



Protected data

- **NIST 800-171 Low**
- **High assurance endpoints**
 - User must authenticate with specific identity within a specified time period, with browser session and native app device instance isolation
 - Audit logging
 - Multi-factor authentication
- **For data that requires additional security**
 - HIPAA Personal Health Information (PHI) w/ BAA
 - Personally Identifiable Information (PII)
 - Sensitive but unclassified
- **Two additional subscription tiers**
 - **High assurance tier:** for all added security features
 - **BAA tier:** high assurance features plus BAA with UChicago
- **Initial release**
 - Transfer, sharing, web app, CLI (excludes publication, search, GlobusID)



New Storage Connectors

- **We continue to grow our connector set**
- **On near-term radar**
 - Box
 - Google Cloud Storage
- **Under consideration**
 - Microsoft Azure Blob Storage
 - Wasabi
 - Others?

 box

 Google Cloud

Microsoft Azure

 **wasabi**[™]
hot cloud storage

Globus Connect Server v5 - motivations

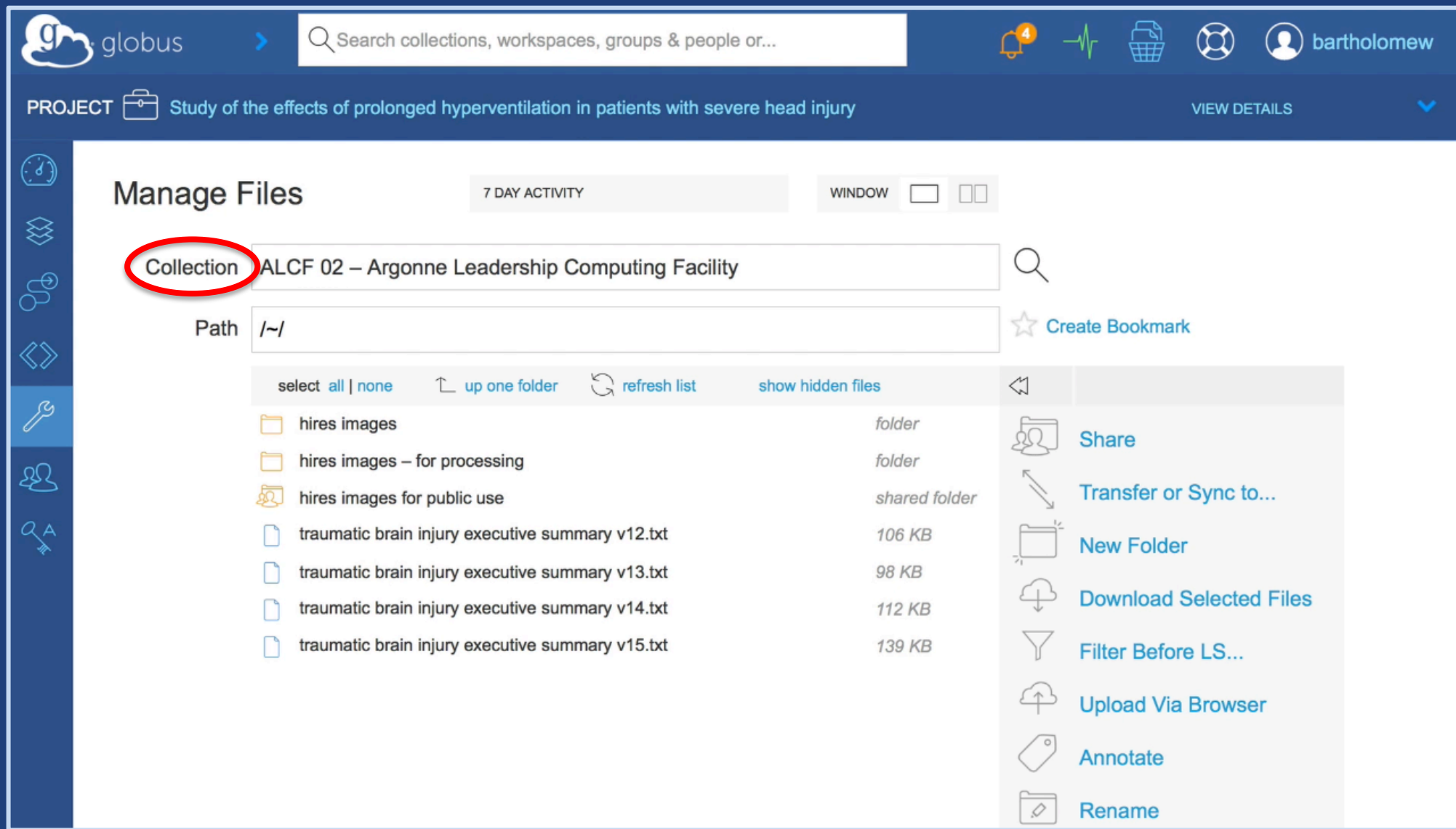
- **Facilitate automation of installation and upgrades**
- **Allow scale out deployment**
 - Across DTNs
 - Across multiple file system connectors
- **Reduce number of ports required**
- **Streamline user experience with use of Globus sharing**
- **Enhance user registration of credentials for cloud storage connectors**
- **Prepare foundation for next set of enhanced capabilities**



New features with Globus Connect Server v5

- **Collection model**
- **HTTPS access to storage**
- **Security improvements**
 - OAuth2 in GridFTP (no more X.509 user certificates or Myproxy!)
 - OpenID Connect identity provider
 - Credential expiration LoA policies
 - User credential management (e.g., for Google Drive, S3, Kerberos)
- **Kerberos protected file systems**
- **Directory listing with path expressions**

Collections: The evolution of endpoints



The screenshot displays the Globus interface for managing files. At the top, there is a search bar and a navigation bar with icons for notifications, activity, shopping cart, and user profile (bartholomew). Below this, a project banner reads "PROJECT Study of the effects of prolonged hyperventilation in patients with severe head injury" with a "VIEW DETAILS" link.

The main section is titled "Manage Files" and includes a "7 DAY ACTIVITY" filter and window controls. A search bar contains the text "Collection ALCF 02 - Argonne Leadership Computing Facility", with the word "Collection" circled in red. Below the search bar is a "Path" field containing "/~/". To the right of the path field is a "Create Bookmark" button.

Below the search and path fields is a toolbar with the following options: "select all | none", "up one folder", "refresh list", and "show hidden files".

The main content area displays a list of files and folders:

Item	Type	Size
hires images	folder	
hires images - for processing	folder	
hires images for public use	shared folder	
traumatic brain injury executive summary v12.txt		106 KB
traumatic brain injury executive summary v13.txt		98 KB
traumatic brain injury executive summary v14.txt		112 KB
traumatic brain injury executive summary v15.txt		139 KB

On the right side of the file list, there is a vertical menu of actions: "Share", "Transfer or Sync to...", "New Folder", "Download Selected Files", "Filter Before LS...", "Upload Via Browser", "Annotate", and "Rename".



Collection properties

- **Set of blobs (files), hierarchically named (folders)**
- **Rooted at a unique DNS name**
- **URL referenceable files, folders**
- **Accessible and manageable via:**
 - **HTTPS: client/server file access**
 - **GridFTP: async bulk transfer**
 - **REST API: advanced operations**
- **OAuth2 authentication and authorization via Globus Auth**
- **Collection-specific access policies**
- **Data is stored on a storage system, which determines storage policies such as durability and availability**
- **File change events**



Installation & configuration enhancements for v5

- **Setup with any identity (GlobusID not required)**
- **Automatable installation and configuration**
- **Configuration API, CLI, GUI**
- **Scale-out deployment without shared file system**
- **Backup / restore configuration to / from the cloud**
- **Multiple storage systems simultaneously**
- **Single port GridFTP (no ephemeral ports)**
- **Distributed as Docker containers**



Streamlined data sharing with v5

- **Remove friction of sharing**
 - Guest collections where possible, e.g., Google Drive
 - Hybrid collections: Mapped access to home & project folders, else guest access
- **Enhanced sharing permissions**
 - permission expiration
 - permissions on files (not just folders)
 - sharing via URL possession
- **Storage connectors: share from anywhere**



New capabilities built on collections and v5

- **Data search (early release available now by request)**
 - With access control
 - Schema agnostic
 - Custom indexes domain specific
- **Event driven actions for automation**
 - Replication of data (across storage tiers)
 - Metadata extraction and ingest to search
 - Run analysis pipelines



Join the Globus community

- Access the service: globus.org/login
- Create a personal endpoint: globus.org/app/endpoints/create-gcp
- Documentation: docs.globus.org
- Engage: globus.org/mailing-lists
- Subscribe: globus.org/subscriptions
- Need help? support@globus.org
- Follow us: [@globusonline](https://twitter.com/globusonline)



Support resources

- **Customer engagement team**
- **Globus documentation: docs.globus.org**
- **Helpdesk and issue escalation: support@globus.org**
- **Globus professional services team**
 - Assist with portal/gateway/app architecture and design
 - Develop custom applications that leverage the Globus platform
 - Advise on customized deployment and intergation scenarios



Open Discussion