The globusworld TOUR

Globus Endpoint Administration

Rachana Ananthakrishnan - **rachana@globus.org**

Greg Nawrocki - **greg@globus.org**

Johns Hopkins University
April 11, 2019

globus

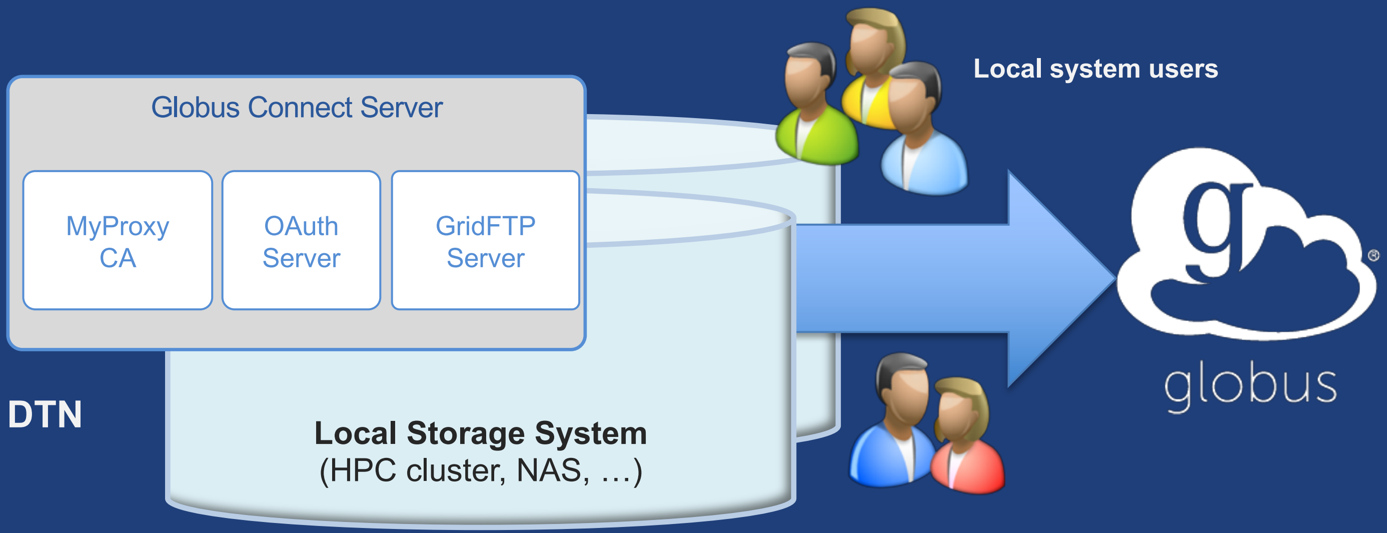
The DTN and Globus Connect Server

Current - Full feature set: GCS 4.x

Future – Some features currently available: GCS 5.x

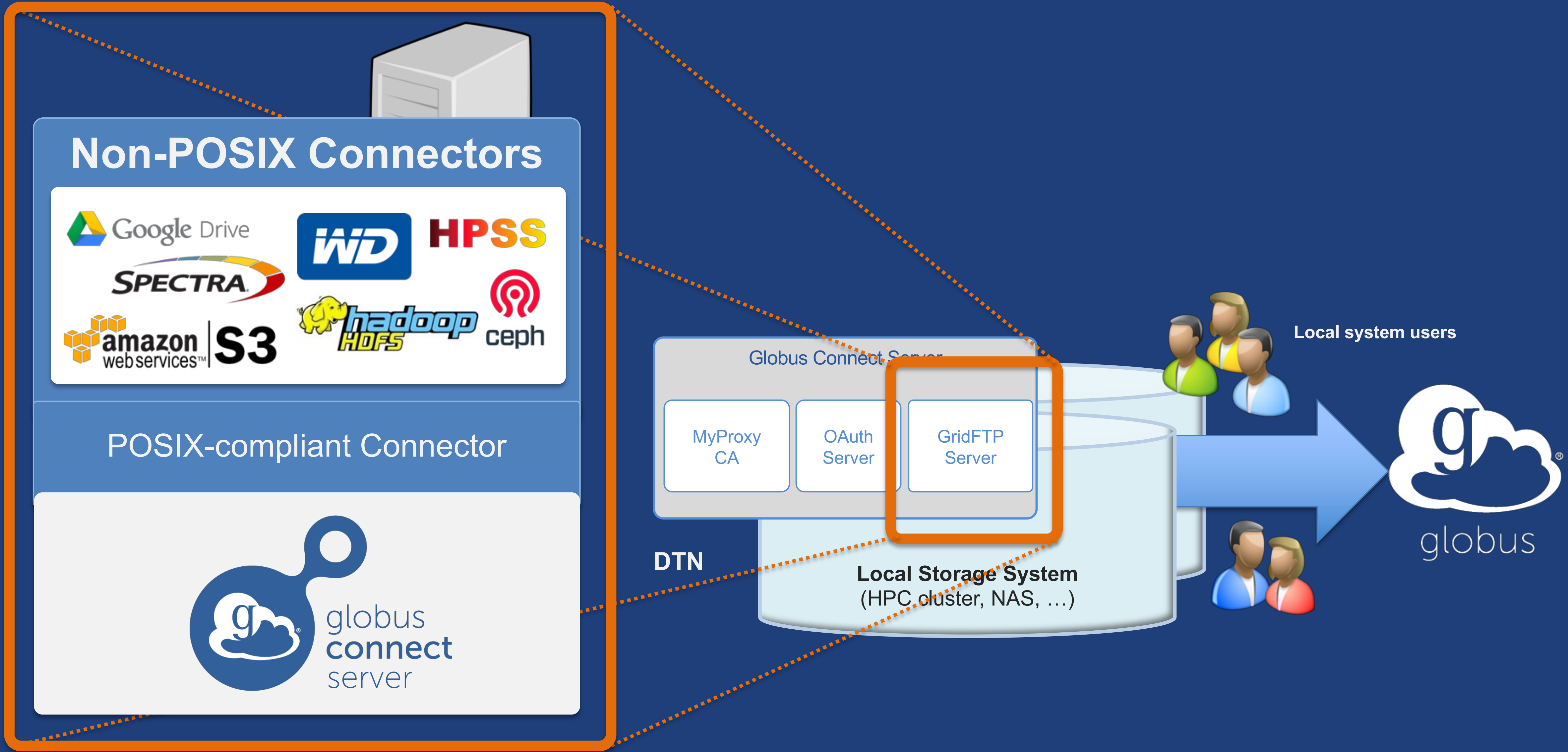
Globus Connect Server

- 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/

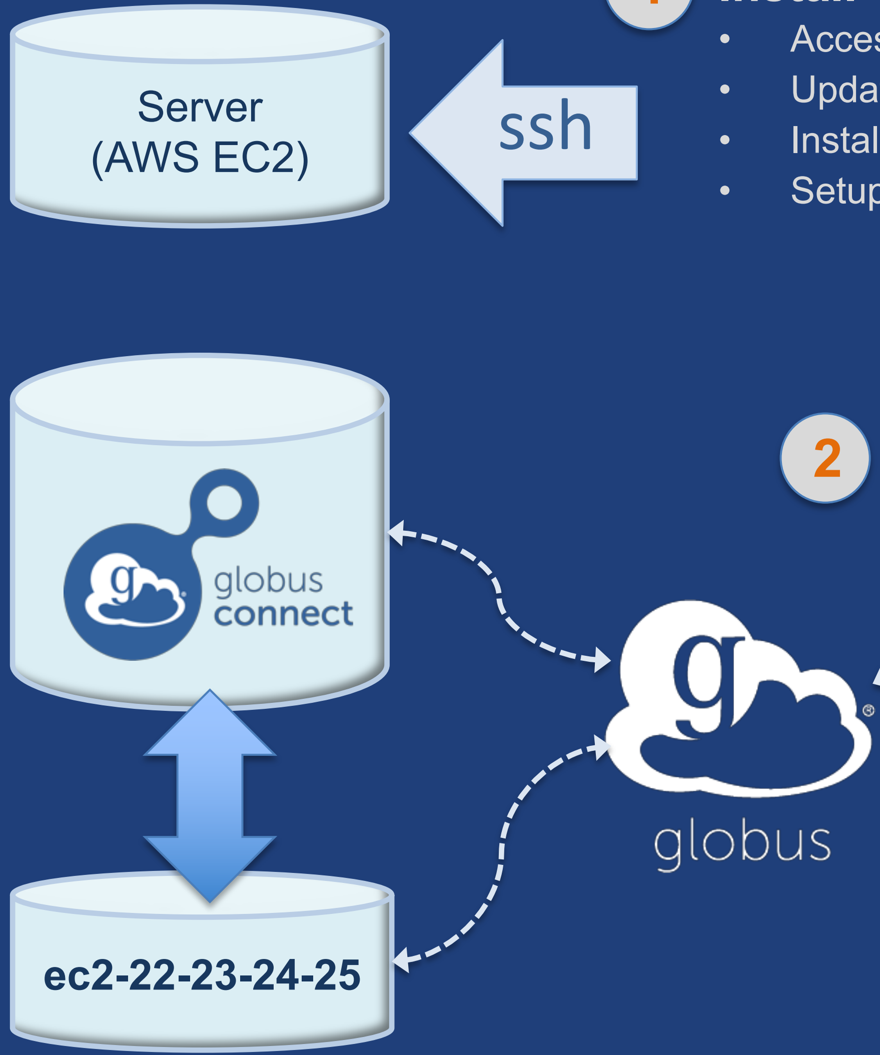
Globus Connect Server



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

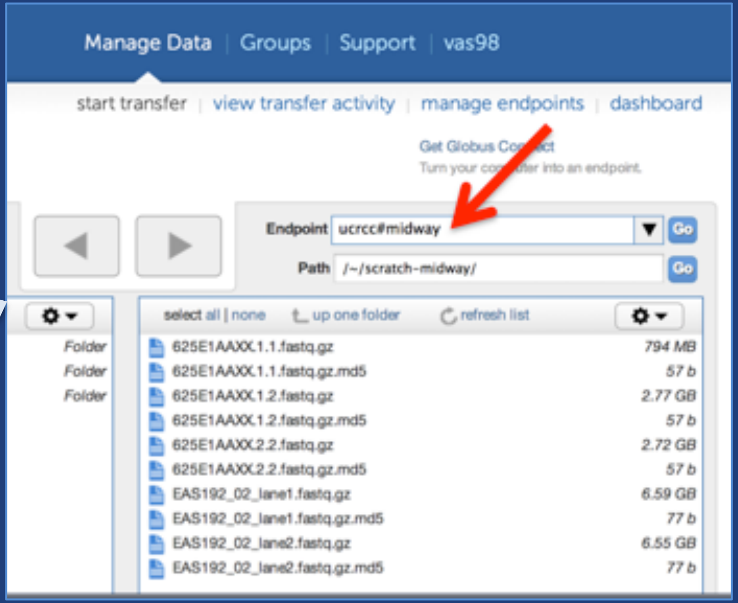
What we are going to do:



1 **Install Globus Connect Server**

- Access server as user “**campusadmin**”
- Update repo
- Install package
- Setup Globus Connect Server

2 **Log into Globus**


Manage Data | Groups | Support | vas98
start transfer | view transfer activity | manage endpoints | dashboard
Get Globus Co ct
Turn your co ter into an endpoint.
Go
Go
select all | none
up one folder
refresh list
Folder
Folder
Folder

Endpoint ucrcc#midway

Path /~/scratch-midway/

625E1AAXX.1.1.fastq.gz	794 MB
625E1AAXX.1.1.fastq.gz.md5	57 b
625E1AAXX.1.2.fastq.gz	2.77 GB
625E1AAXX.1.2.fastq.gz.md5	57 b
625E1AAXX.2.2.fastq.gz	2.72 GB
625E1AAXX.2.2.fastq.gz.md5	57 b
EAS192_02_lane1.fastq.gz	6.59 GB
EAS192_02_lane1.fastq.gz.md5	77 b
EAS192_02_lane2.fastq.gz	6.55 GB
EAS192_02_lane2.fastq.gz.md5	77 b

3 **Access the newly created endpoint (as user ‘researcher’)**

4 **Transfer a file**

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 –LOs
http://downloads.globus.org/toolkit/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 and 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. ESnet read-only) 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**
  - – Thanks to our superb security …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 web app to flag as managed endpoint**

* 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

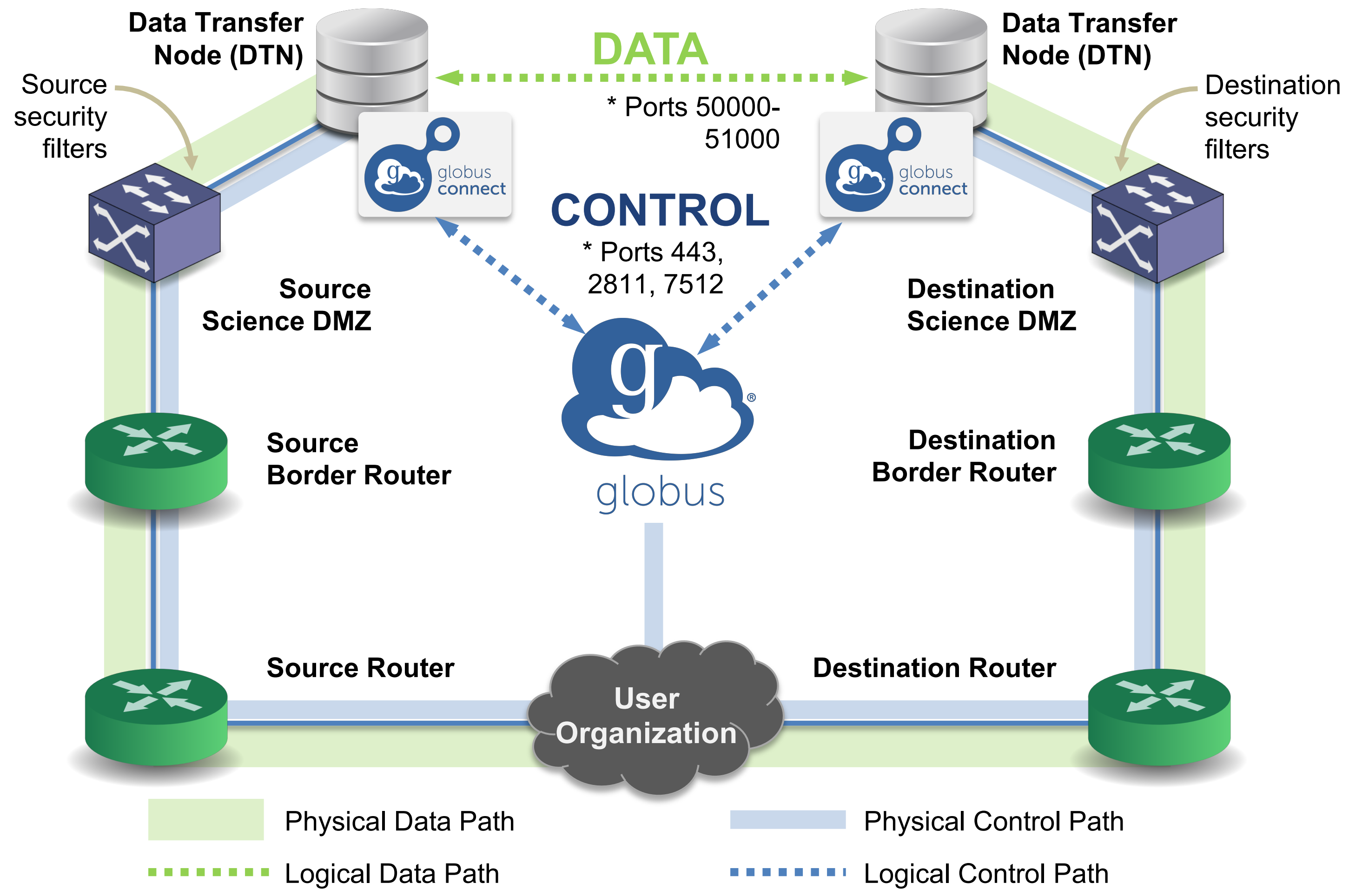
- **Restrict paths where users can create shared endpoints**
- **Use `SharingRestrictPaths` to customize**
  - Same syntax as `RestrictPaths`
- **e.g. Full access to home directory, deny hidden files:**
  - `SharingRestrictPaths = RW~,N~/.*`
- **e.g. Full access to public folder under home directory:**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

Network Paths - Illustrative



Data Transfer Node (DTN)

Source security filters

DATA

* Ports 50000-51000

Data Transfer Node (DTN)

Destination security filters

globus connect

globus connect

CONTROL

* Ports 443, 2811, 7512

Source Science DMZ

Destination Science DMZ

globus

Source Border Router

Destination Border Router

Source Router

User Organization

Destination Router

🟩 Physical Data Path

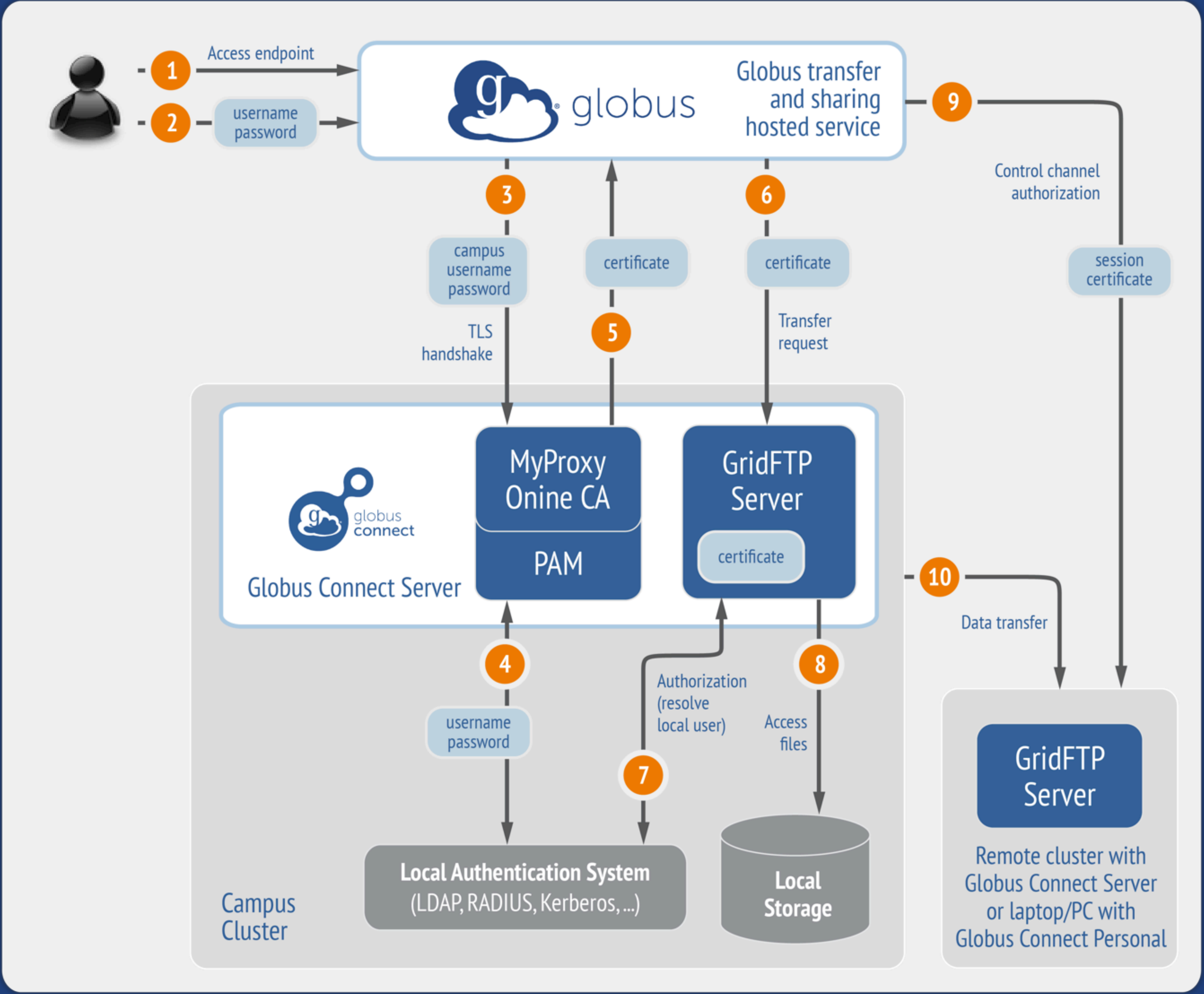
▬ Physical Control Path

┅ Logical Data Path

┅ Logical Control Path

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

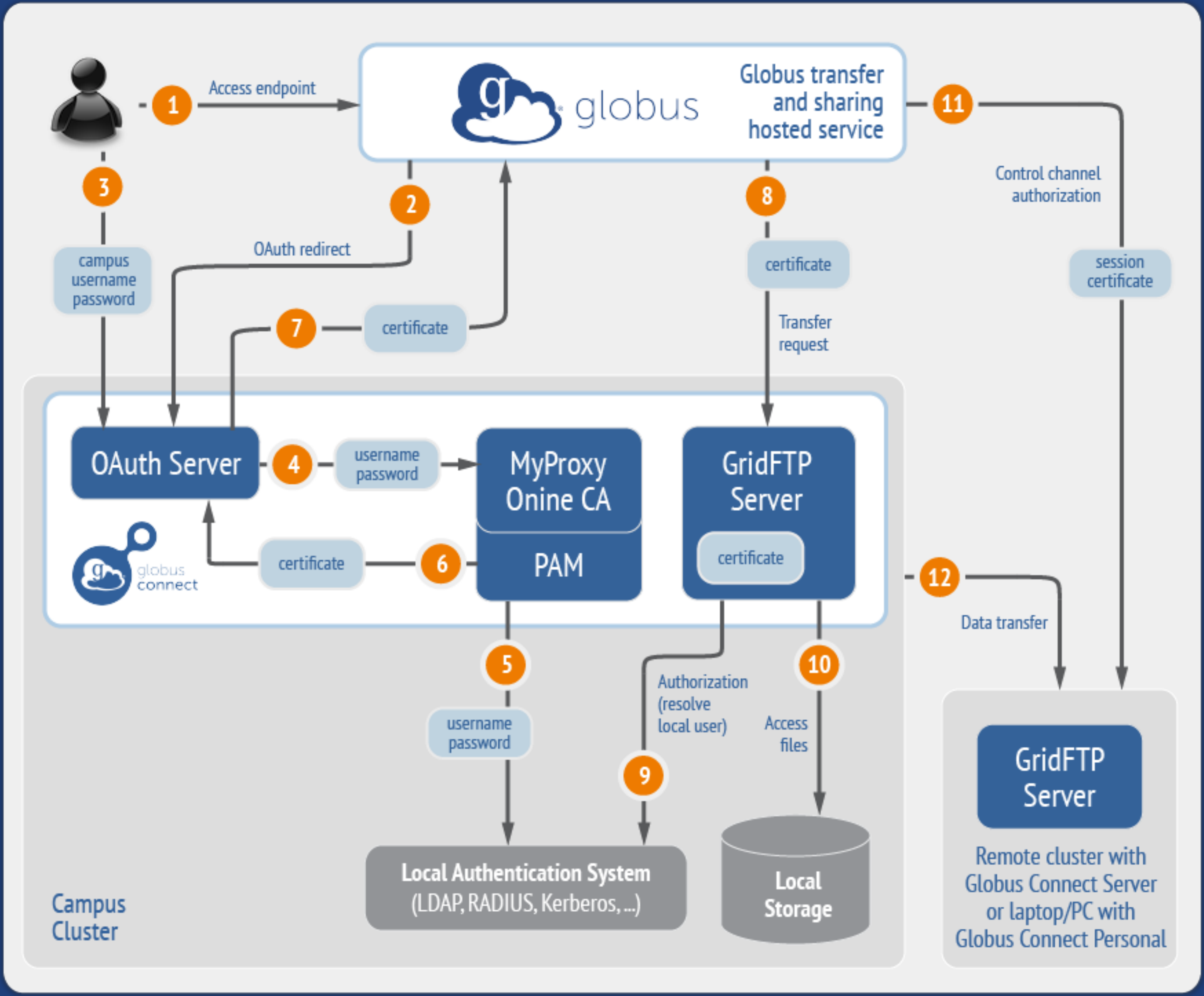
Endpoint activation using MyProxy



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:** <RandS>1</RandS>
- **Local account must match institutional ID (InCommon ID)**
  - Test by creating a local user with same name
- **In** `/etc/globus-connect-server.conf` **set:**

```
AuthorizationMethod = CILogon

CILogonIdentityProvider =
<institution_listed_in_CILogon_IdP_list>

   https://cilogon.org/include/idplist.xml
```

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

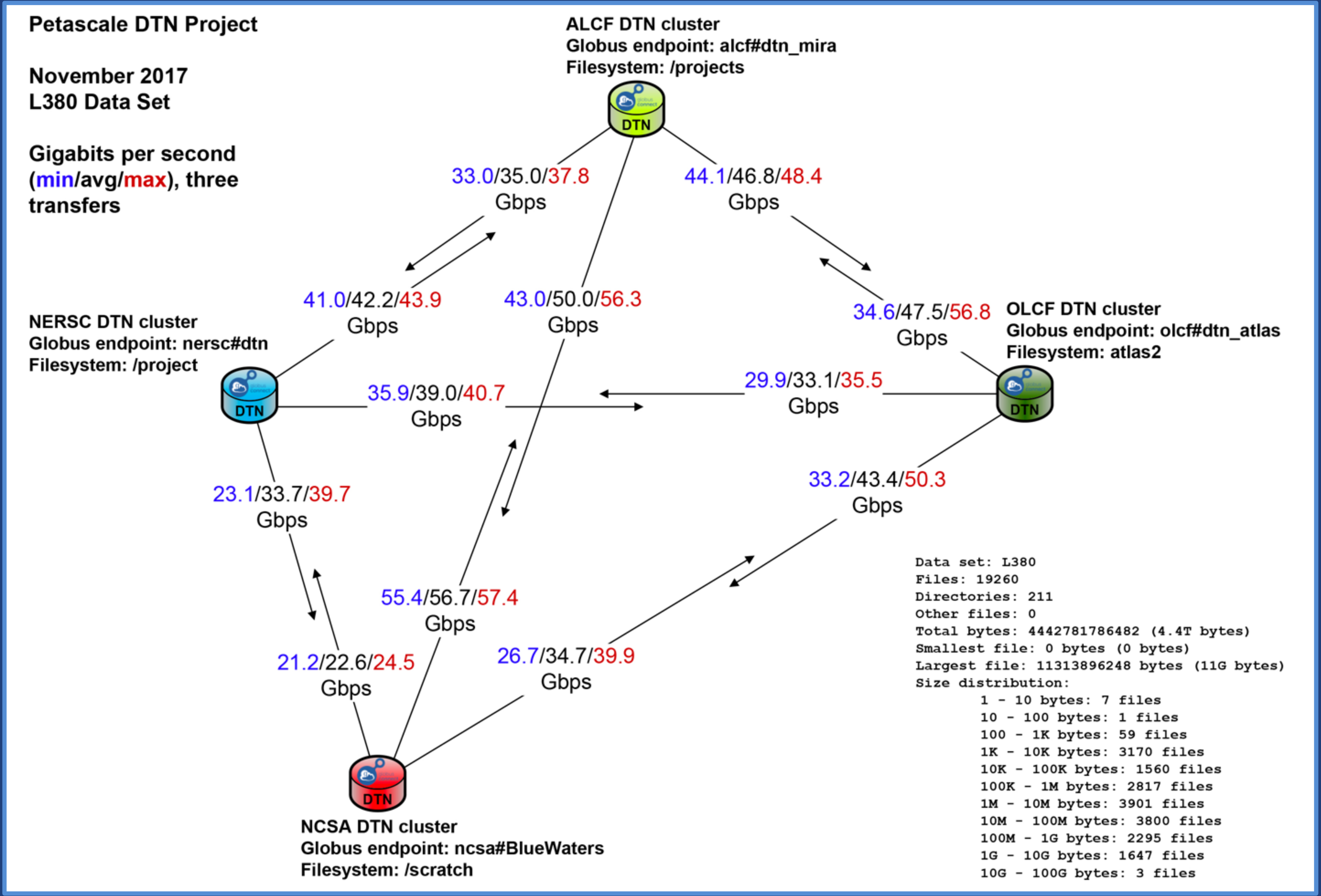
Balance: performance - reliability

- **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

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

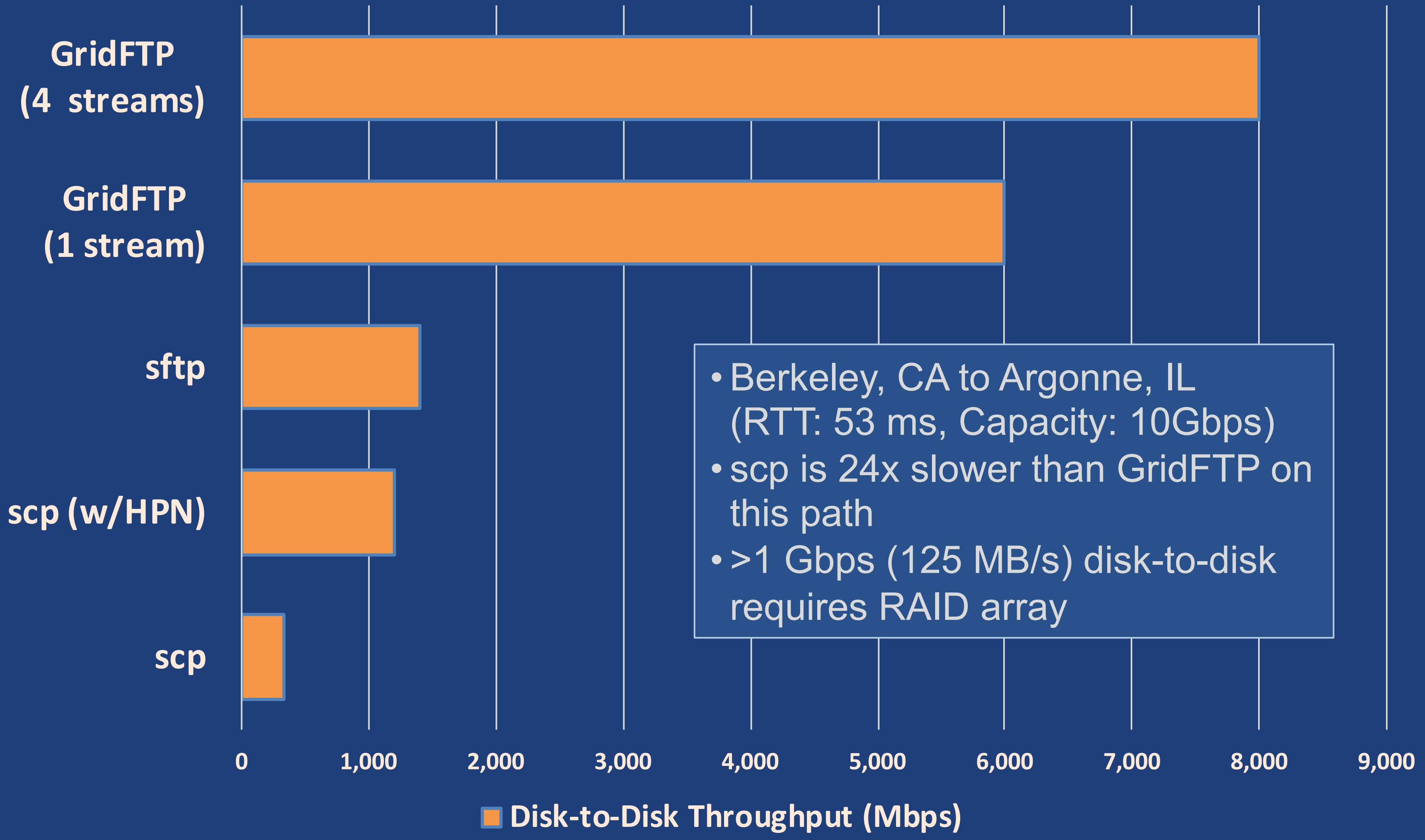
NERSC DTN cluster
Globus endpoint: nersc#dtn
Filesystem: /project

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

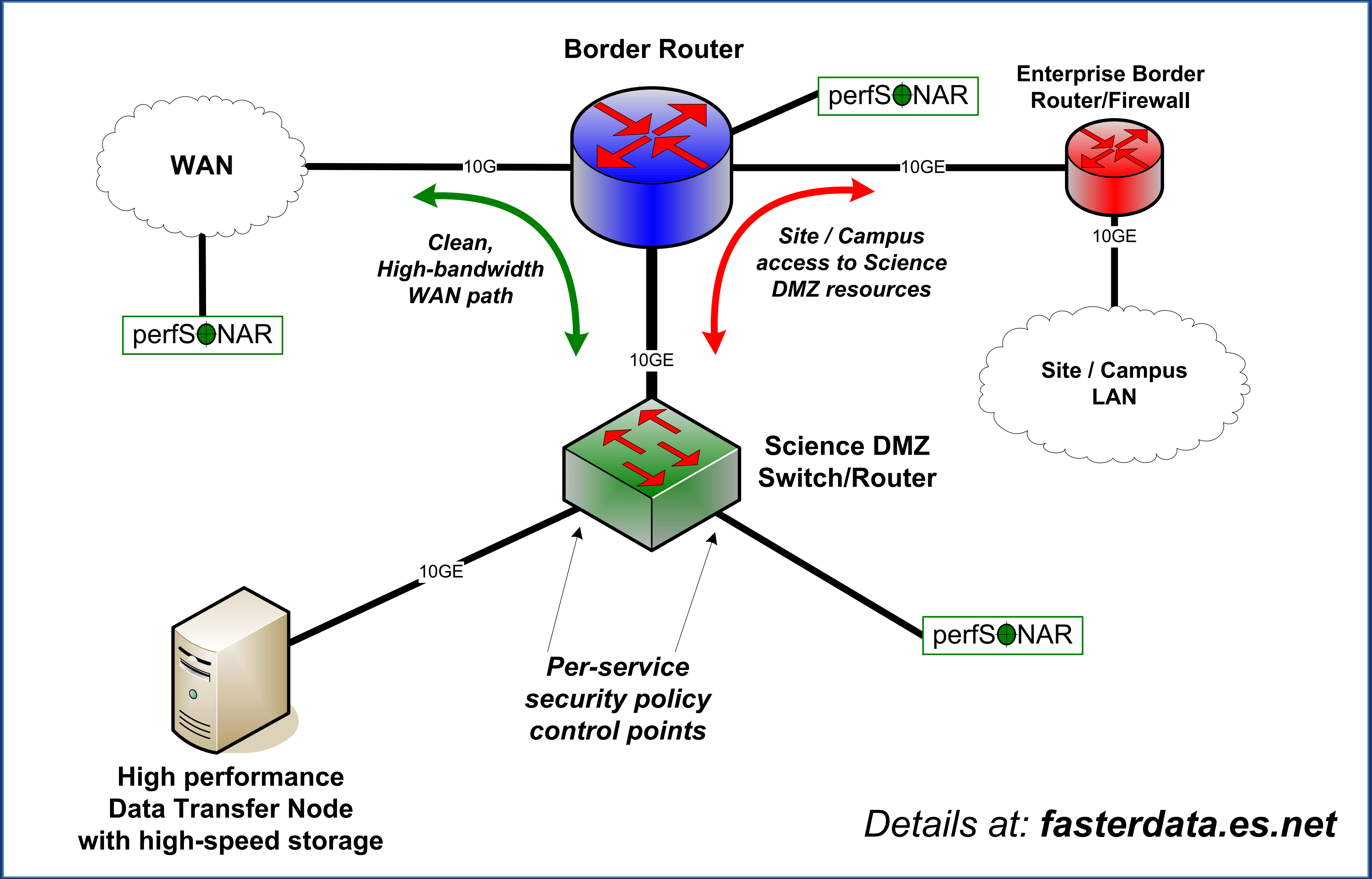


- Berkeley, CA to Argonne, IL (RTT: 53 ms, Capacity: 10Gbps)
- scp is 24x slower than GridFTP on this path
- >1 Gbps (125 MB/s) disk-to-disk requires RAID array

Source: ESnet (2016)

Deployment Scenarios

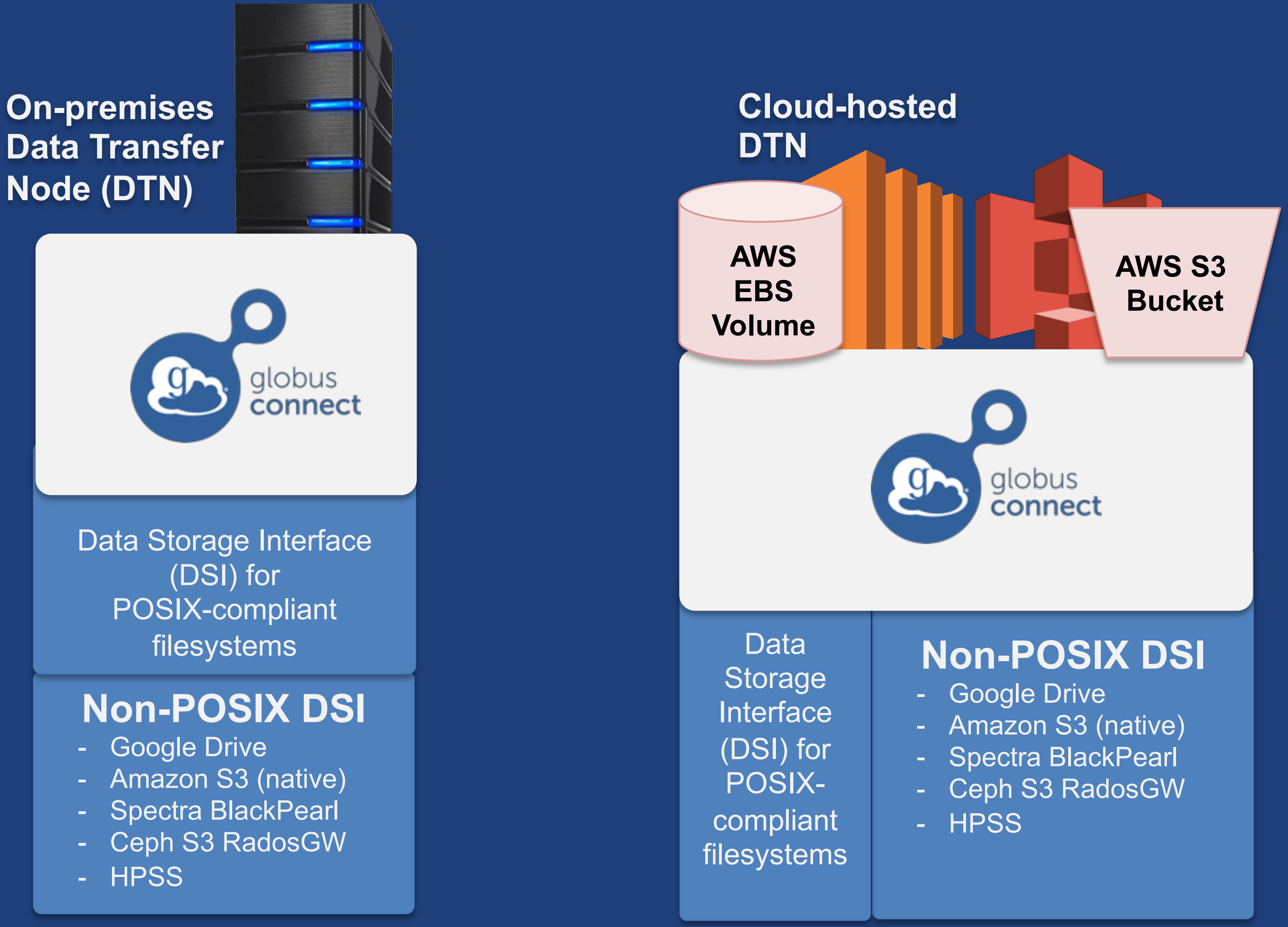
Best-practice deployment


Border Router
perfSONAR
Enterprise Border Router/Firewall
WAN
10G
10GE
Clean, High-bandwidth WAN path
Site / Campus access to Science DMZ resources
10GE
perfSONAR
10GE
Site / Campus LAN
Science DMZ Switch/Router
10GE
perfSONAR
Per-service security policy control points
High performance Data Transfer Node with high-speed storage

Details at: ***fasterdata.es.net***

The Data Transfer Node

On prem and cloud based endpoint hosting



On-premises Data Transfer Node (DTN)

globus connect

Data Storage Interface (DSI) for POSIX-compliant filesystems

Non-POSIX DSI

- Google Drive
- Amazon S3 (native)
- Spectra BlackPearl
- Ceph S3 RadosGW
- HPSS

Cloud-hosted DTN

AWS EBS Volume

AWS S3 Bucket

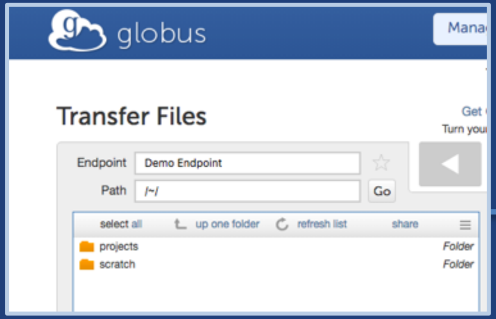
globus connect

Data Storage Interface (DSI) for POSIX-compliant filesystems

Non-POSIX DSI

- Google Drive
- Amazon S3 (native)
- Spectra BlackPearl
- Ceph S3 RadosGW
- HPSS

Common endpoint configuration

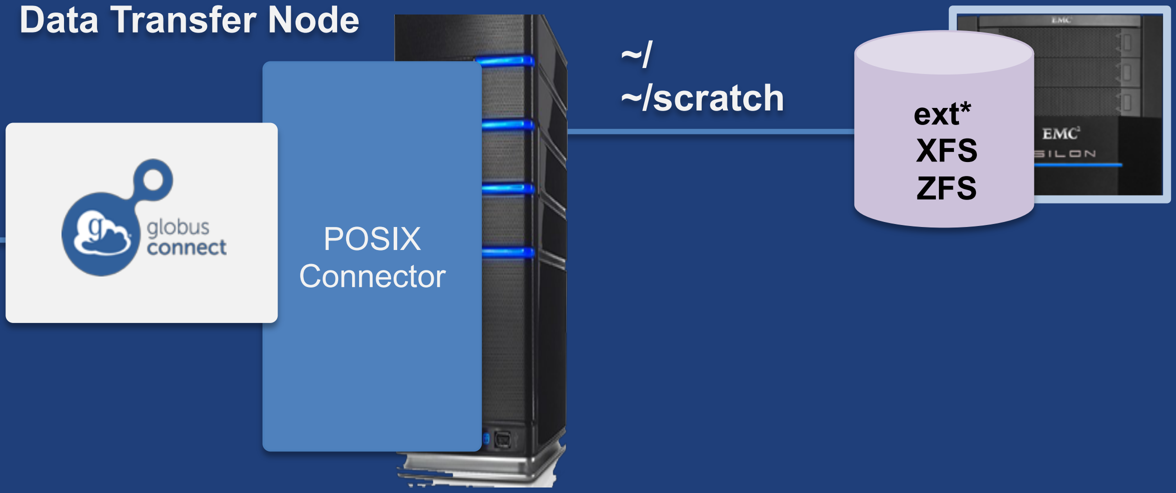


Transfer Files

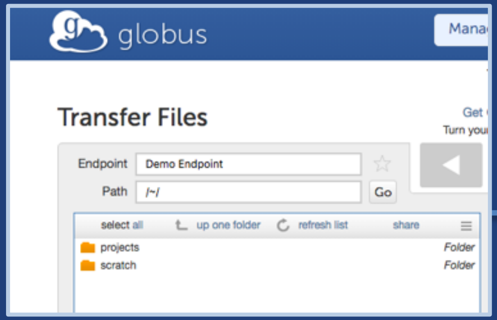
Endpoint: Demo Endpoint

Path: /~/

projects	Folder
scratch	Folder



Common endpoint configuration



globus

Mana

Transfer Files

Get

Turn you

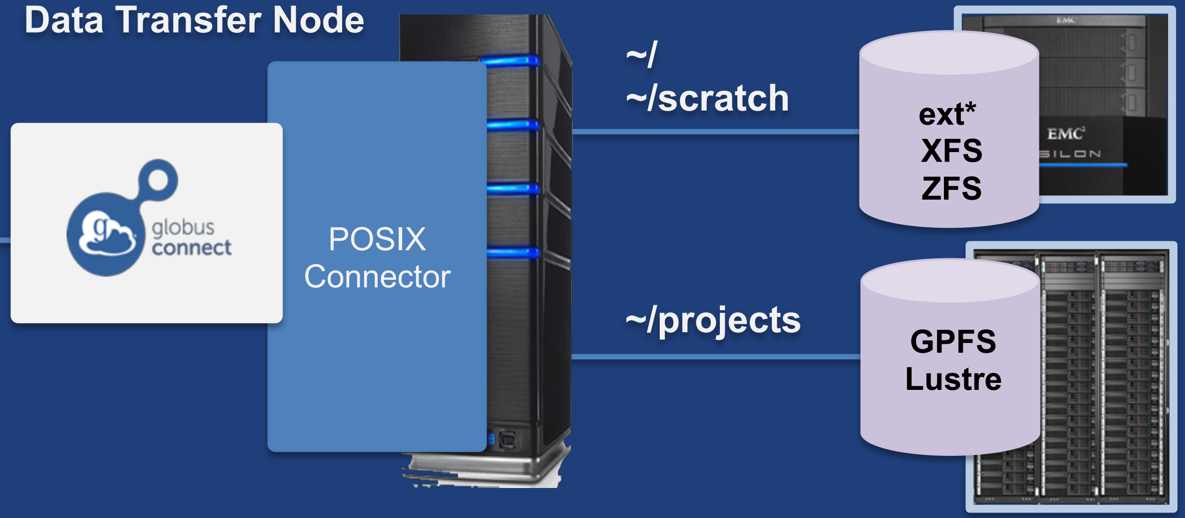
Endpoint: Demo Endpoint

Path: /~/

Go

select all | up one folder | refresh list | share

projects	Folder
scratch	Folder



Data Transfer Node

globus connect

POSIX Connector

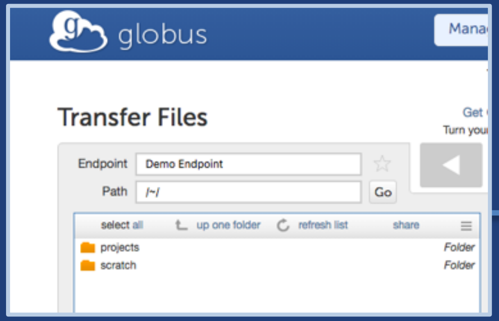
~/
~/scratch

ext*
XFS
ZFS

~/projects

GPFS
Lustre

Multi-endpoint configuration



globus

Mana

Transfer Files

Get

Turn you

Endpoint Demo Endpoint

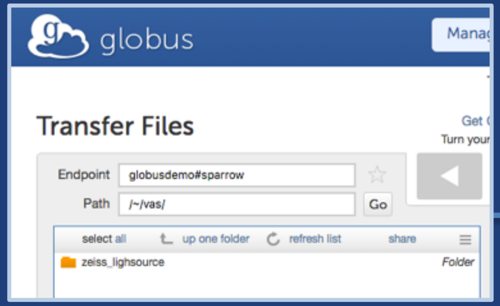
Path /~/

Go

select all | up one folder | refresh list | share

projects Folder

scratch Folder



globus

Manag

Transfer Files

Get

Turn your

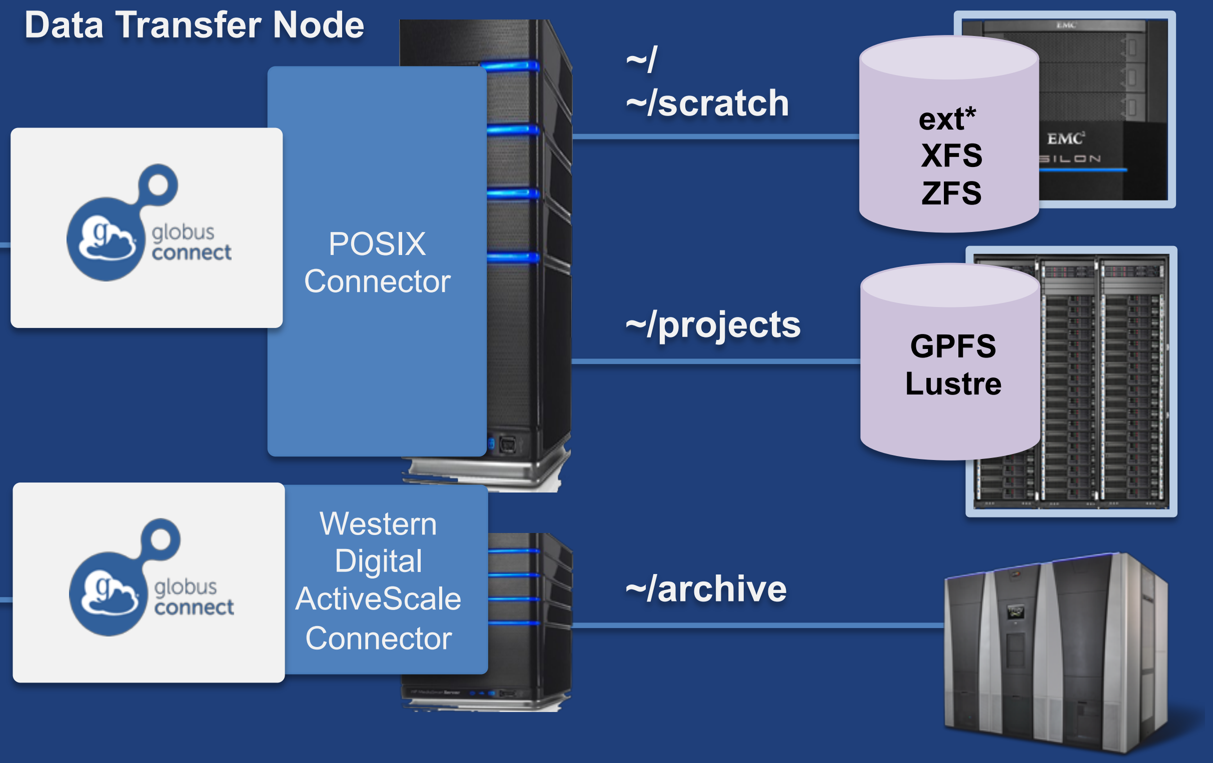
Endpoint globusdemo#sparrow

Path /~/vas/

Go

select all | up one folder | refresh list | share

zeiss_lighsource Folder



Data Transfer Node

globus connect

POSIX Connector

~/

~/scratch

ext*

XFS

ZFS

~/projects

GPFS

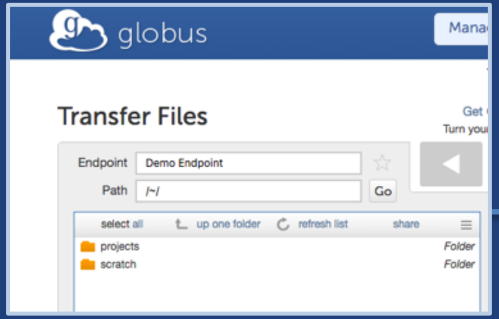
Lustre

globus connect

Western Digital ActiveScale Connector

~/archive

Multi-endpoint configuration



globus Mana

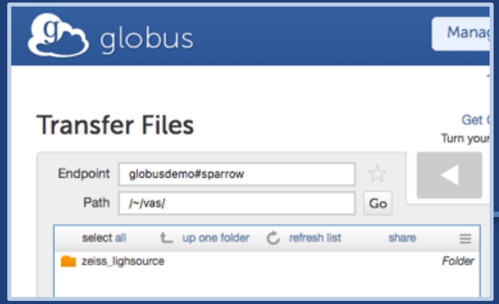
Transfer Files

Get
Turn you

Endpoint: Demo Endpoint
Path: /~/ Go

select all | up one folder | refresh list | share

projects	Folder
scratch	Folder



globus Manag

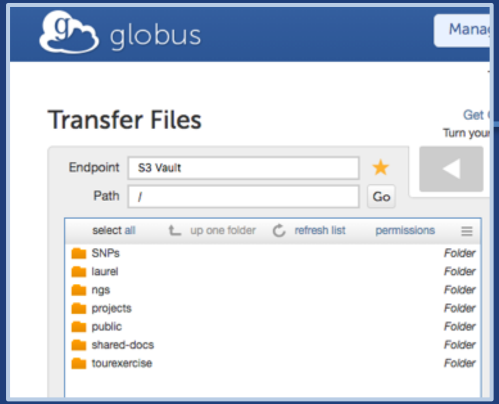
Transfer Files

Get
Turn your

Endpoint: globusdemo#sparrow
Path: /~/vas/ Go

select all | up one folder | refresh list | share

zeiss_lighsource	Folder



globus Mana

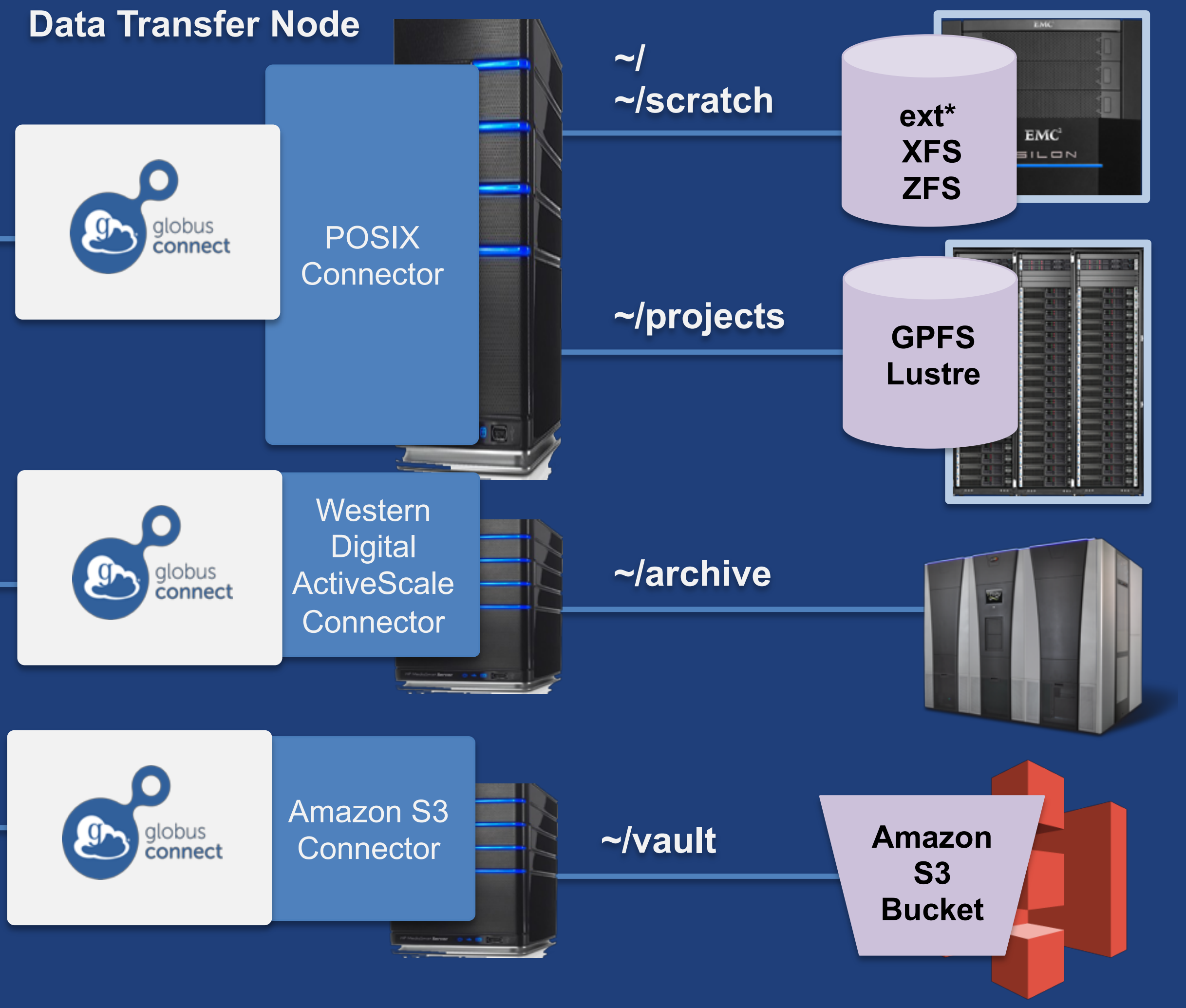
Transfer Files

Get
Turn you

Endpoint: S3 Vault
Path: / Go

select all | up one folder | refresh list | permissions

SNPs	Folder
laurel	Folder
ngs	Folder
projects	Folder
public	Folder
shared-docs	Folder
tourexercise	Folder



Western Digital ActiveScale

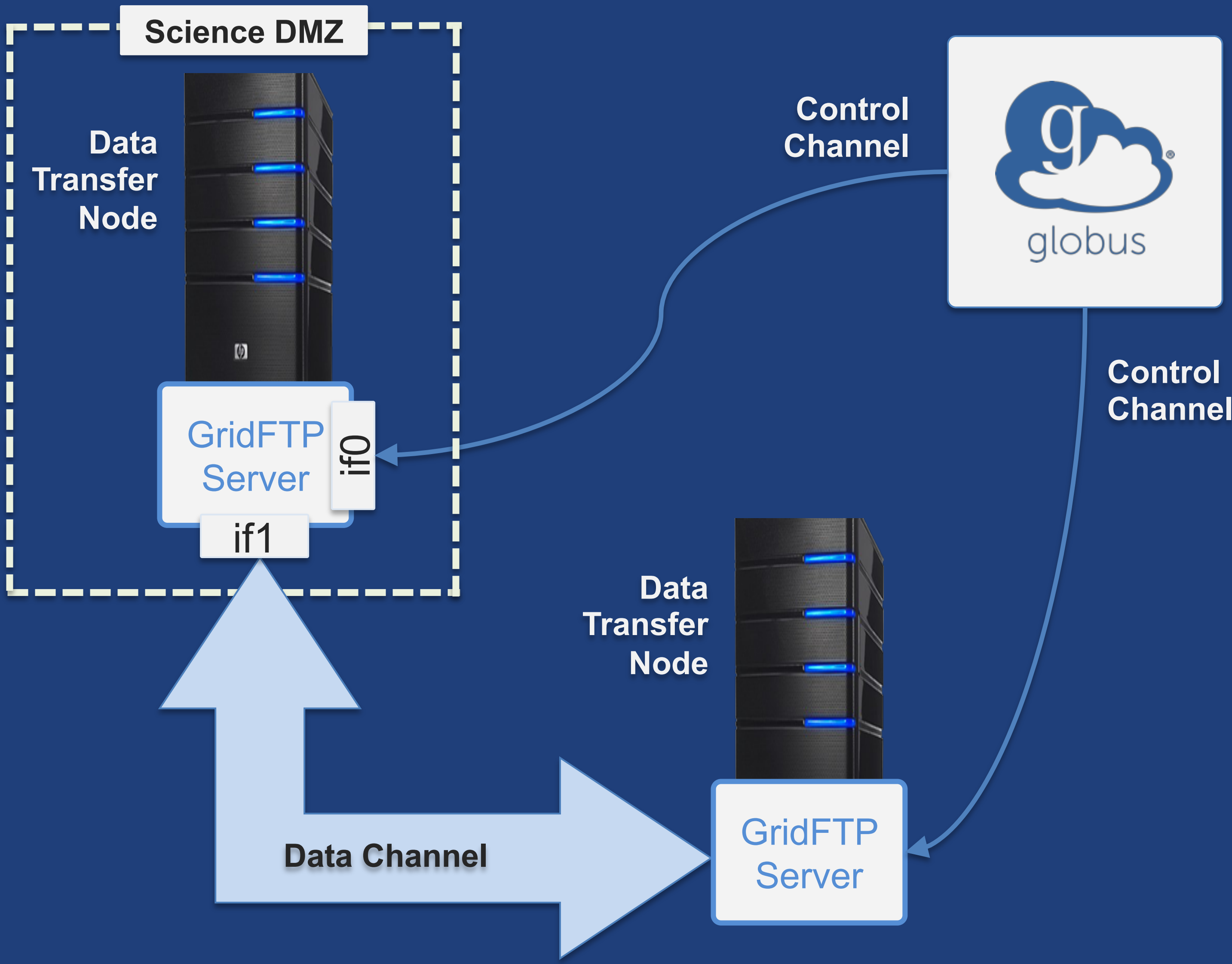
- Turnkey on-premise object storage
- Globus connector using S3 API
- Low TCO: Manufactures own drives
- Erasure coding
- Auto data integrity checks with self-healing
- Cloud-based systems management tools
- Data Forever: automatic migration to new tech

docs.globus.org/premium-storage-connectors/wd-activescale/

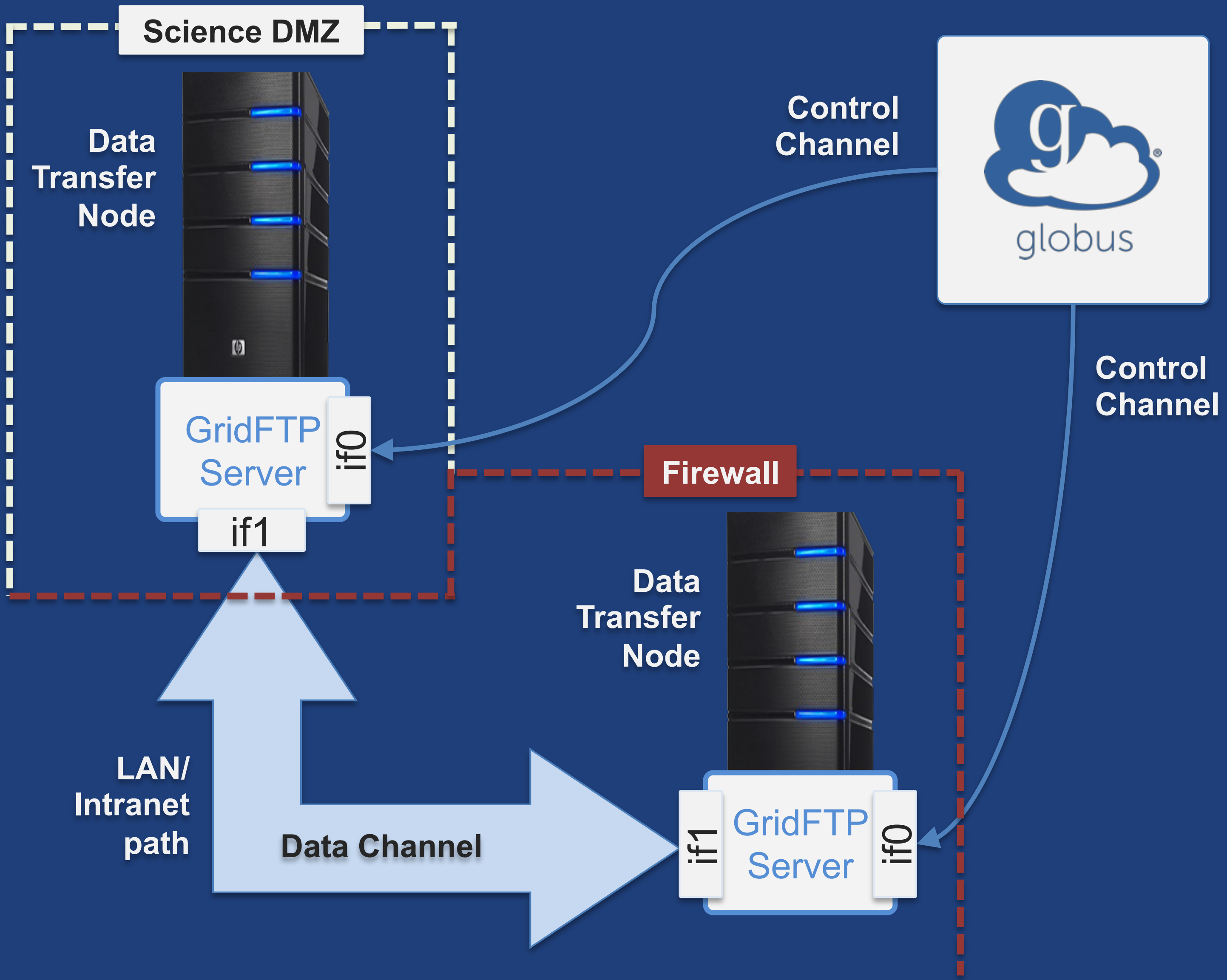
Network paths

- **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 – high speed data path



Other Deployment Options

Encryption

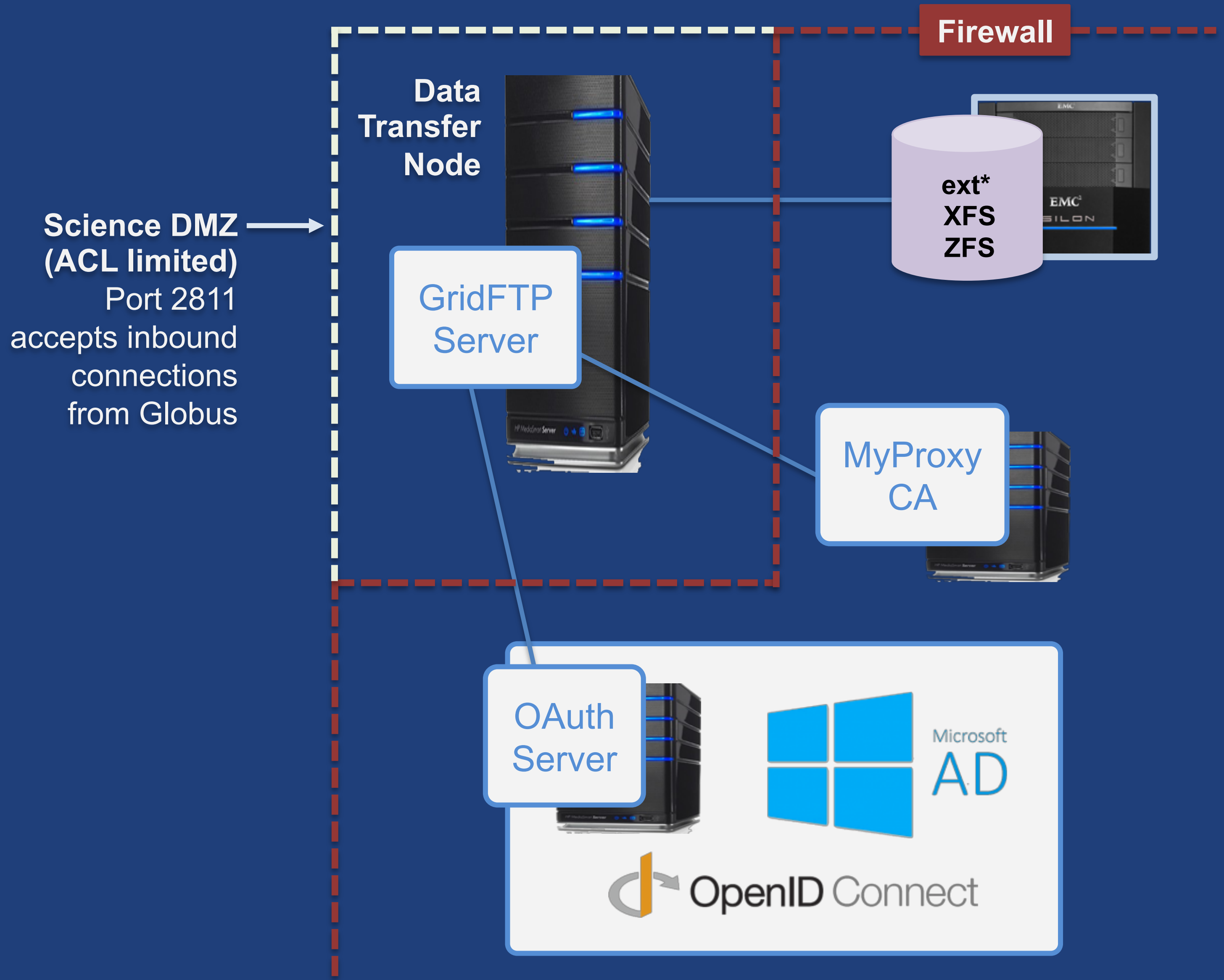
- **Requiring encryption on an endpoint**
  - User cannot override
  - Useful for “sensitive” data
- **Globus uses OpenSSL cipher stack as currently configured on your DTN**
- **FIPS 140-2 compliance: ensure use of FIPS capable OpenSSL libraries on DTN**

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

1. **Install Globus Connect Server on all servers**
2. **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**
3. **Copy Globus Connect Server configuration file to all servers**
4. **Run globus-connect-server-setup on the server running the –id component**
5. **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
```

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**