The globusworld TOUR

Edge.

Advanced Globus Admin

Vas Vasiliadis - vas@uchicago.edu

May 24, 2023

globus

THE UNIVERSITY OF CHICAGO

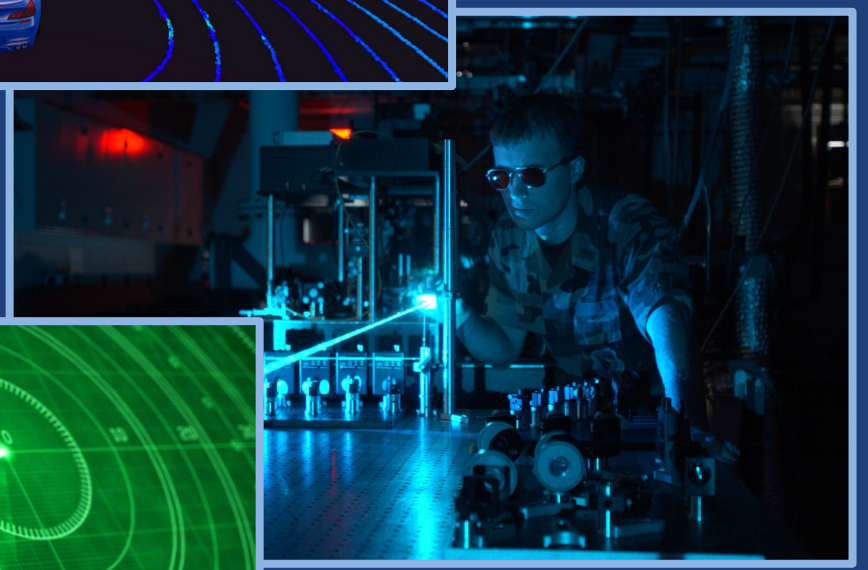
Argonne NATIONAL LABORATORY

Word of the Day

anacronym

(an-AK-ruh-nim)

noun: An acronym or abbreviation whose expansion is not widely known.



From: A.Word.A.Day with Anu Garg

Agenda

- Restricting collection access
- Multi-DTN deployments
- GCS troubleshooting
- Supporting non-POSIX storage systems
- Optimizing (or not!) file transfer performance
- Modifying the data channel interface
- Custom identity mapping example

? Migrating from GCS v4 to v5

Restrict collection access to filesystem

Setting path restrictions

- **E.g., Storage gateway will limit access to /home**
  - NB: No change to local permissions, only visibility via Globus
- **We specify the path restrictions in `paths.json`**
  - This file is in your admin user's home directory
- **Run: `storage-gateway create` command with the `--restrict-paths` option**
- **Create a new POSIX mapped collection**

Cheatsheet
bit.ly/gw-tut

Create a restricted storage gateway, collection

```
$ globus-connect-server storage-gateway create posix \
> "My Storage Gateway - Restricted" \
> --domain globusid.org \
> --authentication-timeout-mins 90 \
> --restrict-paths file:/home/adminN/paths.json

$ globus-connect-server collection create \
> 3926bf02-6bc3-11e7-a9c6-22000bf2d287 \
> / \
> "My Mapped Collection – Restricted"
```

Fully qualified filename containing rule(s) for restricting access to specific filesystem paths

Revisit your mapped collections

- **Your will need to authenticate on your new (restricted access) collection, and consent**
- **Note the access behavior differences between the two mapped collections**

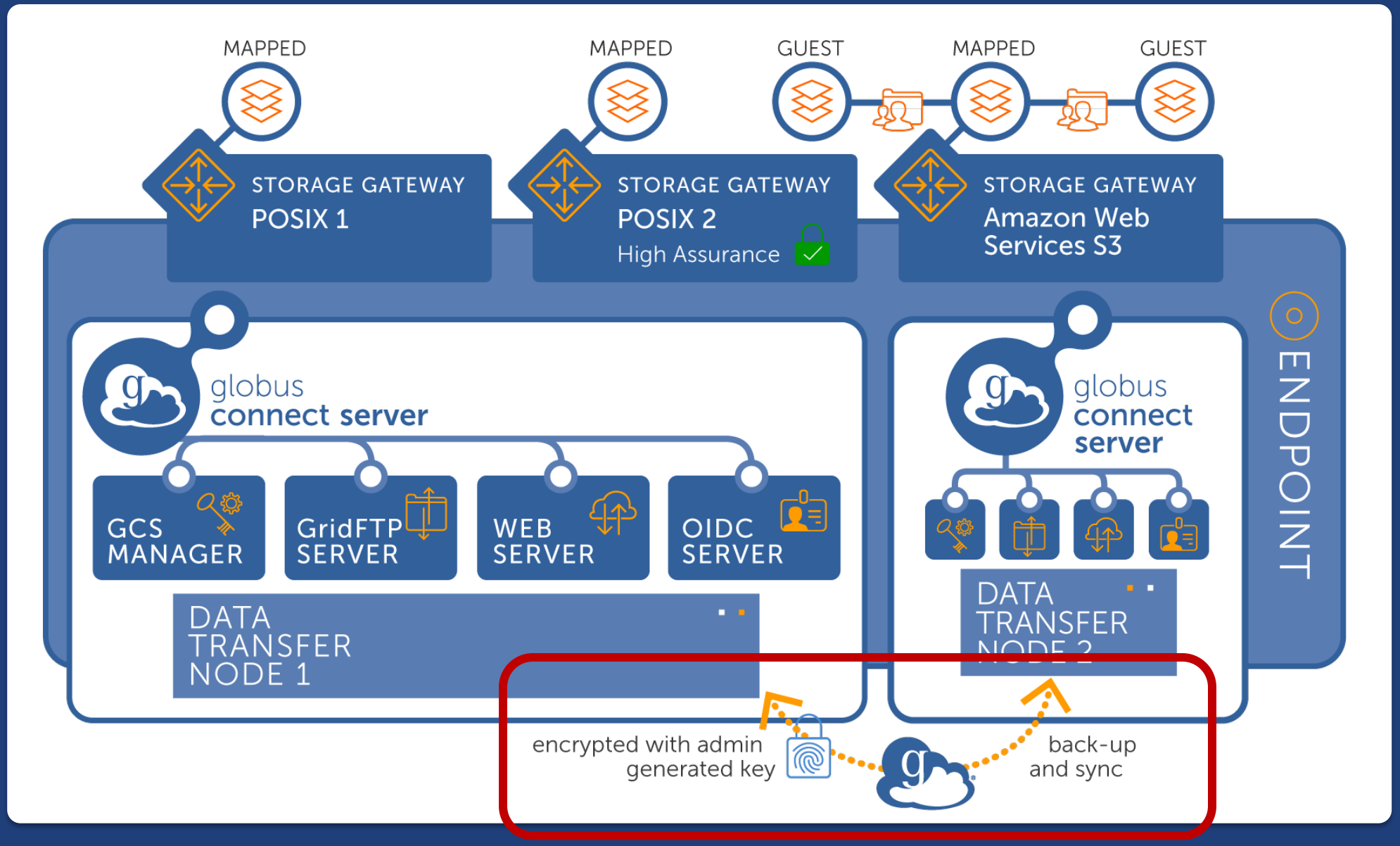
Let’s do some DTN gymnastics…

Adding DTNs to your endpoint

Multi-node DTN behavior

- **Transfer tasks sent to nodes in round-robin fashion**
- **Active nodes can receive transfer tasks**
- **Tasks on inactive node will pause until active again**
- **GCS manager assistant service**
  - Synchronizes configuration among nodes in the endpoint
  - Stores encrypted configuration values in Globus service

Recall: GCSv5 deployment key



Adding a node requires just two commands

```
$ globus-connect-server node setup --deployment-key THE_KEY
$ systemctl restart apache2
```

Copy the deployment key from the first node (DTN) to every other node

Node setup pulls configuration from Globus service

Check your DTN cluster status:
`globus-connect-server node list`

Migrating/refreshing DTNs

Migrating an endpoint to a new host (DTN)

- **An endpoints is a logical construct → replace host system without disrupting the endpoint**
  - Avoid replicating configuration data (esp. for guest collections!)
  - Maintain continuity for custom apps, automation scripts, etc., that use the endpoint UUID
- **1. Add new node to endpoint → 2. remove original node**
- **Again, deployment key is required**
  - Export node configuration with `node setup --export-node`
  - Import on new DTN using `node setup --import-node`

Troubleshooting Globus Connect Server

Before asking for help…

- `self-diagnostic` **can identify many issues**
  - Are services running? GCS manager/assistant, GridFTP server
- **Connectivity is a common cause**
  - Can Globus connect to the GCS Manager service?
  - Is the DTN control channel reachable?
  - Can the DTN establish data channel connection?

docs.globus.org/globus-connect-server/v5.4/troubleshooting-guide

…and we're always here for you: support@globus.org

Supporting non-POSIX systems

- **Update your GCS packages**
- **Add the appropriate storage gateway**
  - Non-POSIX systems require add-on connector subscription(s)
- **Gateway configuration options vary by connector**
  - e.g., specify bucket name(s) for AWS S3
- **Collection authentication options vary by connector**
  - e.g., provide user access key and secret key for AWS S3
  - Credentials must grant appropriate permissions
  - Mapped collection may not actually “map” to local user account

Accessing AWS S3

(and S3-compatible systems)

On performance…

You should have Great Expectations

Data set size	1 Minute	5 Minutes	20 Minutes	1 Hour
10PB	1,333.33 Tbps	266.67 Tbps	66.67 Tbps	22.22 Tbps
1PB	133.33 Tbps	26.67 Tbps	6.67 Tbps	2.22 Tbps
100TB	13.33 Tbps	2.67 Tbps	666.67 Gbps	222.22 Gbps
10TB	1.33 Tbps	266.67 Gbps	66.67 Gbps	[illegible]
1TB	133.33 Gbps	26.67 Gbps	6.67 Gbps	2.22 Gbps
100GB	13.33 Gbps	2.67 Gbps	666.67 Mbps	[illegible]
10GB	1.33 Gbps	266.67 Mbps	66.67 Mbps	22.22 Mbps
1GB	133.33 Mbps	26.67 Mbps	6.67 Mbps	2.22 Mbps
100MB	13.33 Mbps	2.67 Mbps	0.67 Mbps	0.22 Mbps

Time to transfer

ESnet EPOC target for all DOE labs
Requires at least a 10G connection

Globus transfer is fast …but it depends on…

- **Data Transfer Node (CPU, RAM, bus, NIC, …)**
- **Network (devices, path quality, latency, …)**
- **Storage (hardware, attach mode, …)**
- **Dataset make-up (file#, size, tree depth, …)**
  - Remember: LoSF == Great sadness
- **Things people do (one transfer per file …1M files)**
- …?

Performance is a pairs sport

- **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**

Globus network use parameters

- **May only be changed on managed endpoints**
- **Modify via the web app: Console → Endpoints tab**
- **Modify via Globus Connect Server CLI**
  - Run `globus-connect-server endpoint modify`
- **Strong recommendation: Do *not* change network use parameters before establishing baseline performance**

Modifying network use parameters

Configuring a “private” data channel

- **Default: data interface is set to the DTN’s public IP address (see** `data_interface` **in** `/etc/gridftp.d/globus-connect-server`**)**
- **Create** `/etc/gridftp.d/STORAGE_GATEWAY_ID`
- **Set** `data_interface PRIVATE_INTERFACE_IP_ADDRESS`
- **Replicate on every DTN (files in** `/etc/gridftp.d/` **are not sync'd between nodes by Globus)**

Customizing identity mapping

- **Recall: Globus identity** `userX@domain.edu` **maps to local user** `userX`
- **Customize via mapping expressions or external code**
- **Apply mapping expression(s) to storage gateway configuration**
- **Call external script; be aware of storage gateway type**
  - Can use static map files, database, etc.

Simple custom mapping example

Map
42032579@wassamottau.edu
to local user **vas**

Otherwise, default behavior
local user → domain username

Note: Requires the storage gateway to accept identities from *two* domains

```
{
  "DATA_TYPE":
"expression_identity_mapping#1.0.0",
  "mappings": [
    {
      "source": "{username}",
      "match": “42032579@wassamottau.edu",
      "output": "vas",
      "ignore_case": false,
      "literal": false
    },
    {
      "source": "{username}",
      "match": ".*@uchicago.edu",
      "output": "{0}",
      "ignore_case": false,
      "literal": false
    }
  ]
}
```

When you really need a clean slate…

- **Proper clean-up—*both on your system and in the Globus service*—is important!**
- **Execute these commands in the specified order:**
  - `globus-connect-server node cleanup` (on every DTN)
  - `globus-connect-server endpoint cleanup` (on last DTN)

Migrating GCSv4 to GCSv5

Goals

- **No user intervention should be required**
- **Recreate all host and guest endpoints**
- **Preserve all relevant configuration**
- **Preserve the UUIDs of the resource**
- **Minimize downtime**

Migration tools: Approach

- **Read v4 configuration to create migration plan**
- **Allow edits and changes by administrator to plan**
- **Apply migration plan to a vanilla install of v5**
- **Test v5 endpoint/collections and validate**
- **Finalize by assigning v4 UUID to the new endpoint**

Impact

- Downtime after final migration step (preserving UUID)
- Active transfers cancelled on final step; users notified
- Pause rules NOT preserved; must be recreated
- Custom applications using v4 host endpoint (with activation) must move to v5 collection (with consent) docs.globus.org/globus-connect-server/migrating-to-v5.4/application-migration

Resources

- GCSv5 Guides: docs.globus.org/globus-connect-server/
- Migration: docs.globus.org/globus-connect-server/migrating-to-v5.4/
- Globus support: support@globus.org