Johns Hopkins University
April 11, 2019

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

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

What we hope to accomplish today

- **Agenda – can be dynamic**
  - Morning – higher level
  - Late Morning / Afternoon – more technical
- **Globus – always moving forward**
  - GCS 4.x to CGS 5.x
  - Old Web App to the current Web App
  - All in the name of High Assurance
- **Give you a taste of many different parts of Globus to encourage you to explore**

Agenda

09:00—10:30	Globus: Beyond File Transfer
10:30—10:45	snack break (refreshments provided)
10:45—11:30	Globus High Assurance for Protected Data
11:30—12:30	Automating Research Data Workflows
12:30—13:30	Lunch break (lunch provided)
13:30—14:30	Globus Endpoint Administration
14:30—15:30	Leveraging the Globus Platform
15:30—15:45	snack break (refreshments provided)
15:45—16:30	Best Practices for Data Sharing

Drives Down TCO for Large-Scale Research Collaboration

Platforms	Composable Infrastructure	Cloud	Primary
Ultrastar® Flash and HDD Platforms	**OpenFlex™** Software Composable Infrastructure	**ActiveScale™** Object Storage Geo-Distributed	**IntelliFlash™** Unified Block and File Storage NVMe, All-Flash, and Hybrid

Density   Capacity   Durability   Integrity   Performance   Manageability

ACTIVE SCALE™   +   globus
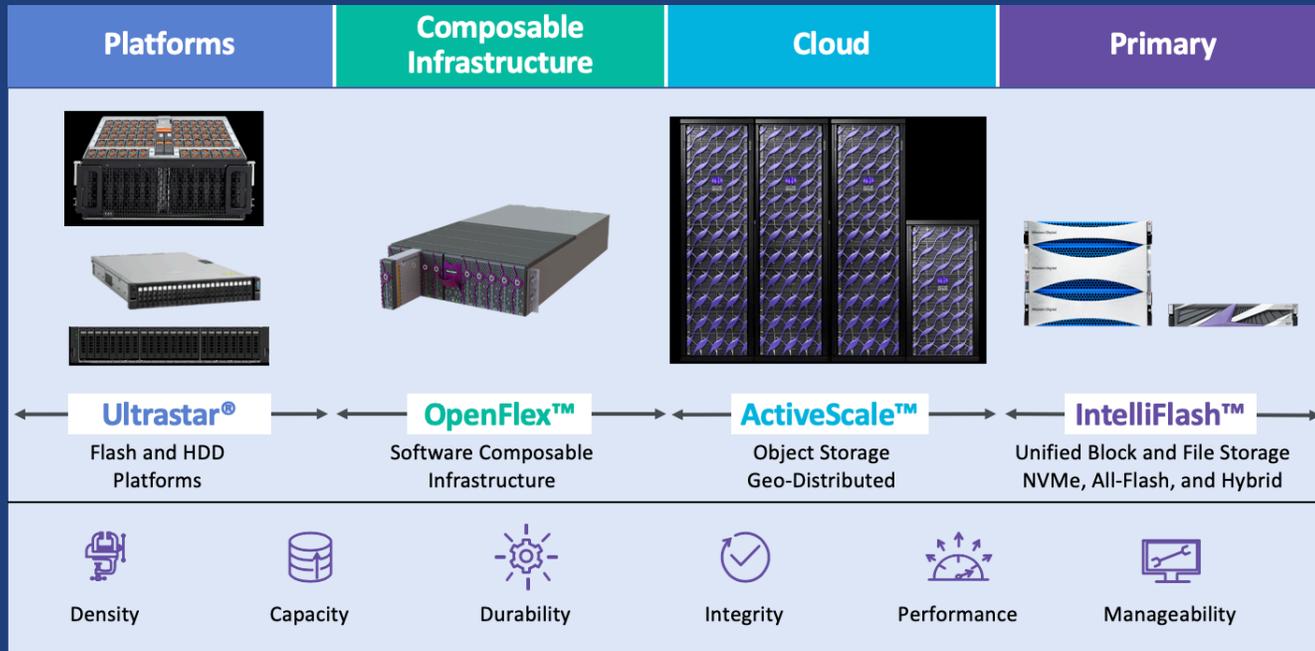
Western Digital Offers Expansive Portfolio of Platforms and Systems

Research data management today

Index?

How do we...
...move?
...share?
...discover?
...reproduce?

Globus delivers...

Fast and reliable ~~big~~ data transfer, sharing, publication, and discovery…

…directly from your own storage systems…

...via software-as-a-service using existing identities with the overarching goal of...

Unifying access to data across tiers
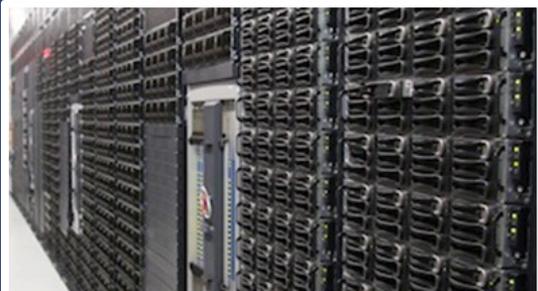
Research Computing HPC

National Resources
Welcome to **compute** canada
ACCESS SERVICES AND EXPERTS
DISCOVER COMPUTE CANADA

Personal Resources

Desktop Workstations

Mass Storage

Instruments

amazon web services | S3
red cloud
Windows Azure
Google Cloud Platform
Public Cloud

Storage Connectors - globus.org/connectors

Current

Planned

IBM Spectrum Scale

SPECTRA

box

amazon web services™ | S3

ceph

Google Cloud

Google Drive

globus connect

wasabi hot cloud storage

hadoop HDFS

Lustre®

HPSS

HGST ActiveScale
a Western Digital brand

Microsoft Azure

Share with collaborators/community

External campus storage

Project repositories, replication stores

Public repositories

Public / private cloud stores

Manage data from instruments

Next-Gen Sequencer
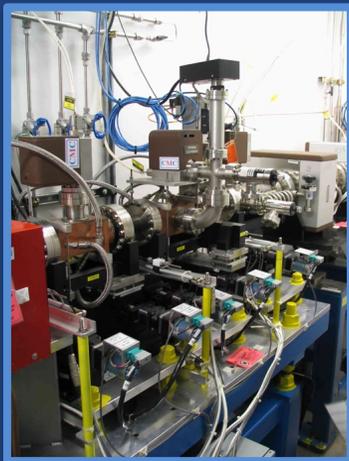
Advanced Light Source

Cryo-EM

MRI

Light Sheet Microscope

Analysis store
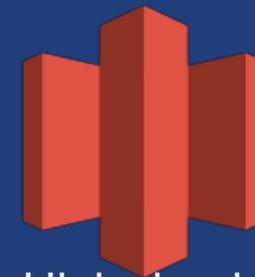
High-durability, low-cost store

Remote visualization

Personal system

Use(r)-appropriate interfaces
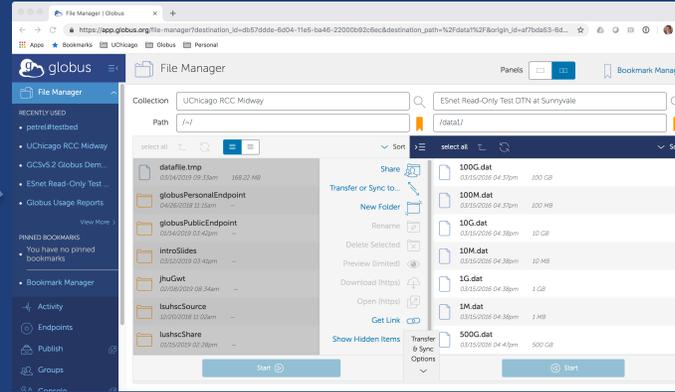
Web

CLI

```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

Options:
  -v, --verbose            Control level of output
  -h, --help               Show this message and exit.
  -F, --format [json|text] Output format for stdout. Defaults to text
  --map-http-status TEXT   Map HTTP statuses to any of these exit codes:
                           0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark       Manage Endpoint Bookmarks
  config         Modify, view, and manage your Globus CLI config.
```

Globus service

```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
…
```

Rest API

Globus SaaS / PaaS: Research data lifecycle

Instrument

Compute Facility

Globus transfers files reliably, securely

2

4 Globus controls access to shared files on existing storage; no need to move files to cloud storage!

Transfer

6 The Globus Command Line Interface, API sets, and Python SDK provide a platform…

Build

3

Share

1

Researcher initiates transfer request; or requested automatically by script, science gateway

Researcher selects files to share, selects user or group, and sets access permissions

7 … for building science gateways, portals and publication services.

Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus

5

8

Automating research workflows and ensuring those that need access to the data have it.

- **Use a Web browser or platform services**
- **Access any storage**
- **Use an existing identity**

Personal Computer

Conceptual architecture: Hybrid SaaS

Subscriber Control Domain

Source Endpoint

DATA Channel

No data relay or staging via Globus

Destination Endpoint

Subscriber owned and administered storage system

globus connect

globus connect

Globus "client" software

CONTROL Channel

powered by amazon web services

globus

Single, globally accessible multi-tenant service

Globus Control Domain

Conceptual architecture: Sharing

External User Control Domain

Subscriber Control Domain

DATA Channel

Managed Endpoint

globus connect

globus connect

CONTROL Channel

Subscriber managed filesystem permissions

powered by amazon web services

globus

Shared Endpoint

Globus Control Domain

Globus managed "overlay" permissions

...makes your storage system a Globus endpoint

Endpoints (Collections)

- **Storage abstraction**
  - All transfers happen between two endpoints
  - Globus Connect instantiates endpoints

- **Collection ~= Endpoint**

- **Test / Demo Endpoints**
  - Globus Tutorial Endpoint 1
  - Globus Tutorial Endpoint 2
  - ESnet Test Endpoints
    - Contain file samples of various sizes

- **Globus Connect Personal**
  - Now your laptop is an endpoint
  - https://www.globus.org/globus-connect-personal

Globus Connect Personal
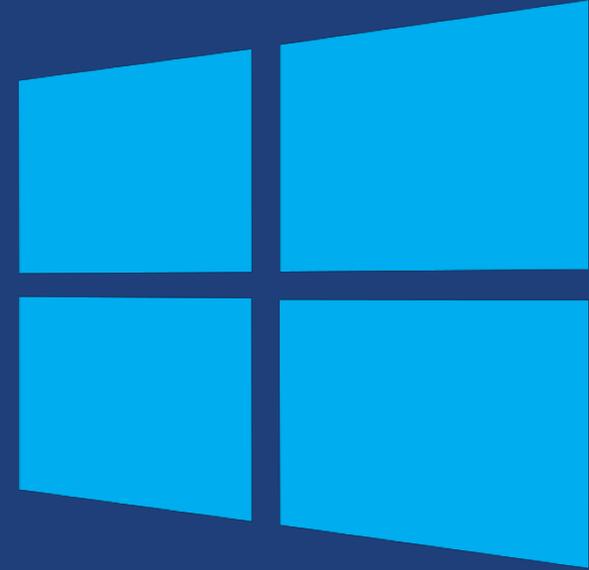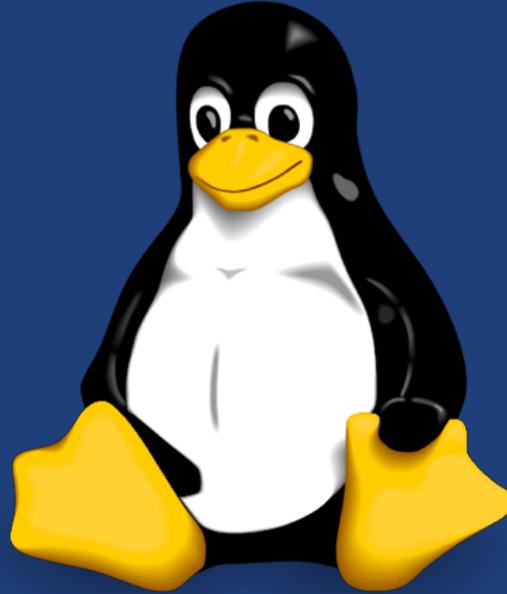
- **Installers do not require admin access**
- **Zero configuration; auto updating**
- **Handles NATs**
- **Installs in seconds – easy to delete - I'll prove it!**

The Globus Web App - Accounts

- **A Globus Account is**
  - A Primary Identity
  - Possible Linked Identities

- **Linking Identities**

- **Managing Identities**

- **Consents**

Demonstration
Identities
File Transfer
File Sharing

The Globus Web App - Hidden in Plain Sight

- **The Hamburger Menu**
  - Not always the same – based on the type of endpoint and the storage behind it
  - A great place to get the link to a share

- **Transfer Settings**
  - label – when using the activity monitor it's nice to see a recognizable name
  - sync - only transfer new or changed files
  - delete files on destination that do not exist on source
  - preserve source file modification times
  - verify file integrity after transfer
  - encrypt transfer

- **Search**
  - The magic of the magnifying glass
  - Search for: Endpoints / Users / Groups

Activity Monitoring

- **Recent / History / Filter**

- **Drilling Down**
  - File transfer statistics
  - Overview
  - Event Log
  - Cancelling an active task

Endpoint Sharing and Roles

- **Sharing**
  - Select the directory and create the "share"
  - A "share" is another type of endpoint
  - Share with: Users / Groups / All Globus Users

- **Roles**
  - Giving others (or groups of others) control or monitoring rights for your endpoints

Groups

- **What can they be used for?**
  - Sharing: Access permissions for more than one person
  - Roles: Endpoint management and monitoring

- **Groups**
  - Creating groups and setting the visibility
  - Members (invitations), Subgroups, Settings
  - Settings
    - Policies / Membership Fields / Terms & Conditions
  - Roles
    - Giving others authority over your groups

Bookmarks

- **Just like browser bookmarks – frequently used, or maybe not used frequently enough!**

- **Creating a bookmark**

- **Using a bookmark**

- **Sorting and Filtering**

- **Editing and Deleting**

Globus Command Line Interface

```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

Options:
  -v, --verbose           Control level of output
  -h, --help              Show this message and exit.
  -F, --format [json|text]  Output format for stdout. Defaults to text
  --map-http-status TEXT  Map HTTP statuses to any of these exit codes:
                          0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark         Manage Endpoint Bookmarks
  config           Modify, view, and manage your Globus CLI config.
  delete           Submit a Delete Task
  endpoint         Manage Globus Endpoint definitions
  get-identities   Lookup Globus Auth Identities
  list-commands    List all CLI Commands
  login            Login to Globus to get credentials for the Globus CLI
  logout           Logout of the Globus CLI
  ls               List Endpoint directory contents
  mkdir            Make a directory on an Endpoint
  rename           Rename a file or directory on an Endpoint
  task             Manage asynchronous Tasks
  transfer         Submit a Transfer Task
  version          Show the version and exit
  whoami           Show the currently logged-in identity.
```

Open source, uses Python SDK

docs.globus.org/cli

github.com/globus/ globus-cli

The Globus CLI

- **Installation**
  - docs.globus.org/cli/installation
  - Prerequisites

- **Logging On (remember the consents?)**
  - globus login / logout

- **Getting help / list of commands**
  - globus –help
  - globus list-commands

- **Doing something**
  - It all about the UUIDs
  - Don't forget the file paths!

The Globus CLI – Let's do a few things...

- **Find endpoints**
  - globus endpoint search Midway
  - globus endpoint search ESNet
  - globus endpoint search --filter-scope=recently-used

- **Find endpoint contents**
  - globus ls af7bda53-6d04-11e5-ba46-22000b92c6ec

- **Transfer a file**
  - From ESnet Read-Only Test DTN at CERN to Midway
  - Note the specific paths
  - globus transfer d8eb36b6-6d04-11e5-ba46-22000b92c6ec:/~/data1/1M.dat af7bda53-6d04-11e5-ba46-22000b92c6ec:/~/1M.dat

- **Transfer a directory**
  - From Globus Tutorial Endpoint 2 to Midway (create directory and contents)
  - globus transfer --recursive ddb59af0-6d04-11e5-ba46-22000b92c6ec:/~/sync-demo af7bda53-6d04-11e5-ba46-22000b92c6ec:/~/syncDemo

- **https://docs.globus.org/cli/examples/**

A bit of Globus history

Globus sustainability model

- **Standard Subscription**
  - Shared endpoints
  - Data publication
  - Management console
  - Usage reporting
  - Priority support
  - Application integration
  - HTTPS support (coming soon)
- **Branded Web Site**
- **Premium Storage Connectors**
- **Alternate Identity Provider (InCommon is standard)**
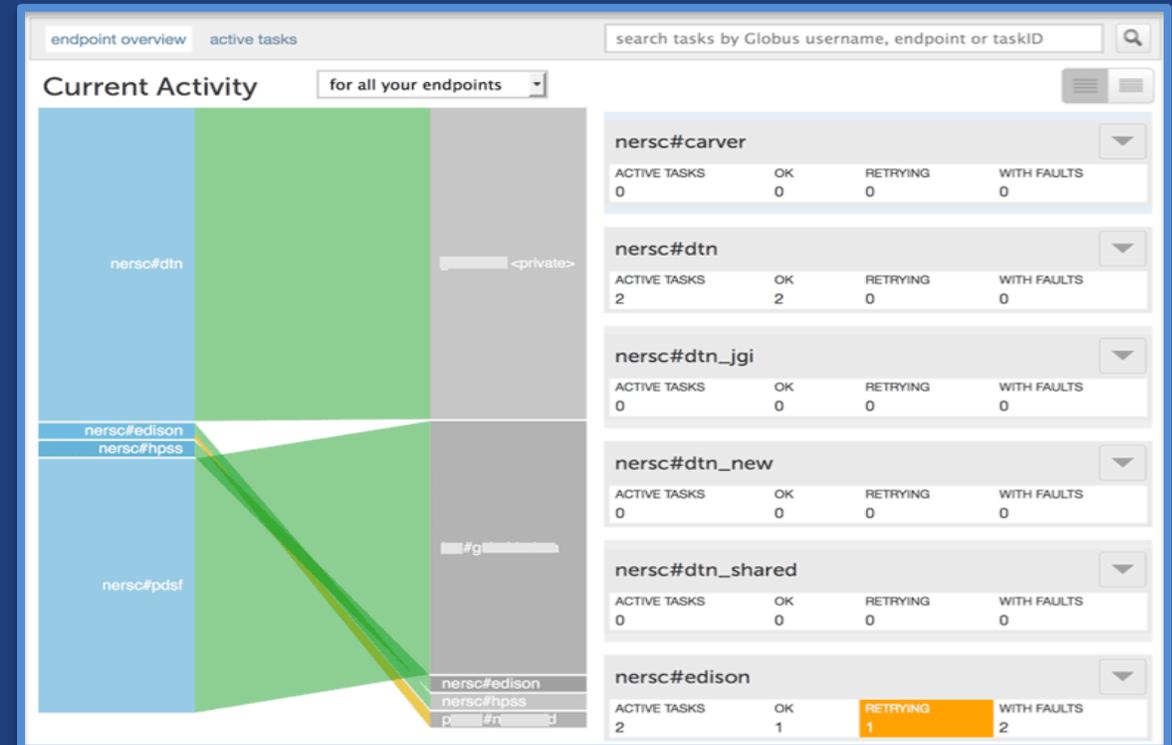
The path to sustainability

Globus by the numbers

1,042 most shared endpoints at a single institution

575 PB transferred

70 billion files processed

1,700 active GCS endpoints

100+ subscribers

100,000 users

3 months longest running transfer

18,000 active GCP endpoints

500+ identity providers

1 PB largest single transfer to date

8,000 active shared endpoints

99.9% availability

Globus on your Campus

- **Webinars**
- **Programs**
  - Helping you evangelize Globus within your institution.
- **Professional Services**
- **Globus World Tour**
  - Taking the show on the road.