

Using Metadata to Enhance Globus Data Transfer Effectiveness

Martin Margo

Presented for 2015 Globus World Conference

Argonne, IL

April 14-15 2015

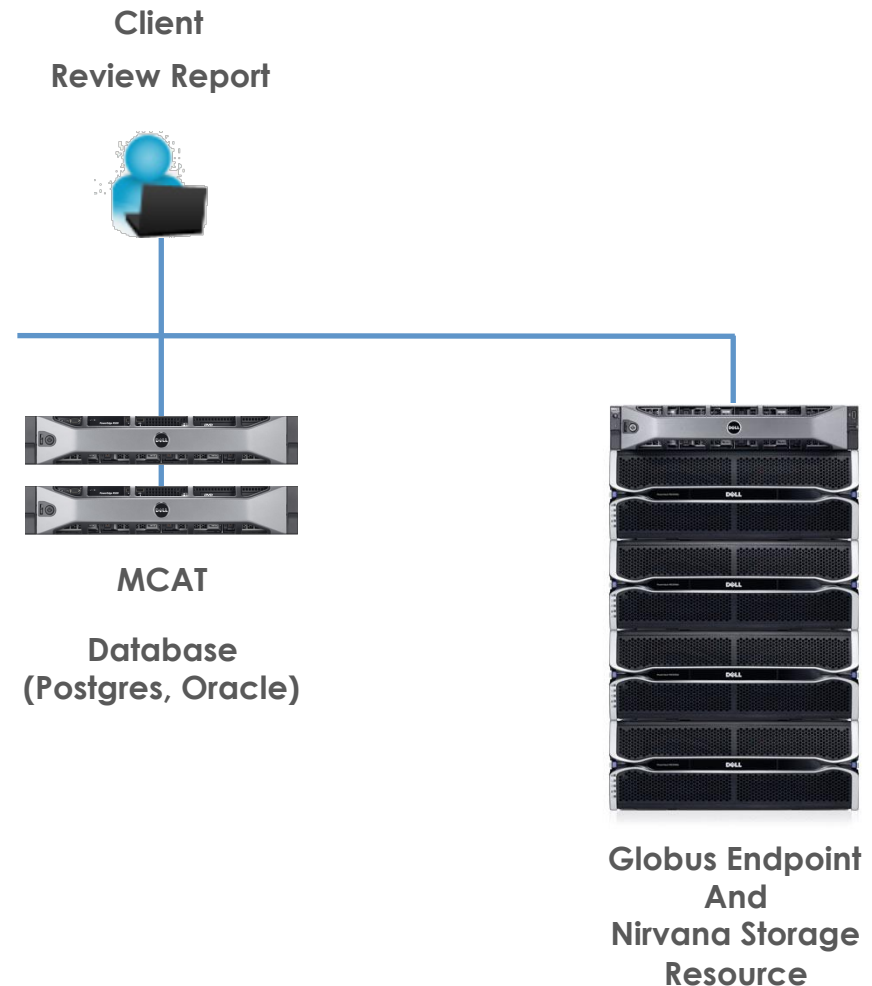
Challenge: User Data Accumulates in Globus Endpoints

1. Scientists are legitimately focusing on their day-to-day scientific work, not IT work
2. Globus is now very easy to use
3. Number of file transfer increases day by day
4. As such, number of copies left behind are increasing too
5. IT reacts when data transfer node endpoints are full



Nirvana as Reporting Engine

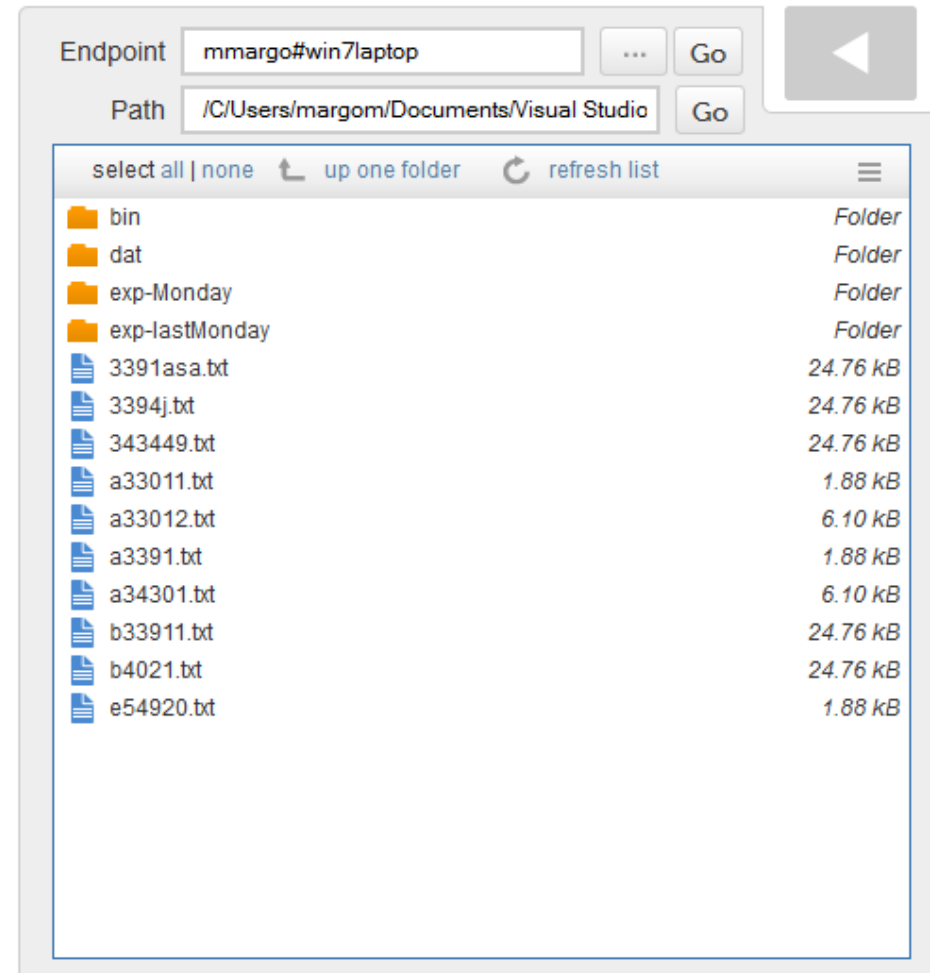
- Nirvana metadata catalog (MCAT) is hosted in RDBMS (Oracle or PostgreSQL)
- Use Nirvana Synchronization daemon to register Globus Endpoint to MCAT
- Daemon picks up POSIX metadata automatically
 - UID, GID, modified time, access time, filename, directory, etc.
- Admins can generate insightful report from MCAT



Using Nirvana SRB To Inventory The Data Junkyard

1. Install Nirvana SRB location daemon
2. Run Location agent on individual Globus Endpoint
3. Nirvana SRB synchronizes Globus Endpoint files to MCAT
4. Create reports to
 - Remedy duplicates
 - Clean up old files
 - Capture data trends (overall size, time distribution, hotspot in user community, etc.)

Transfer Files



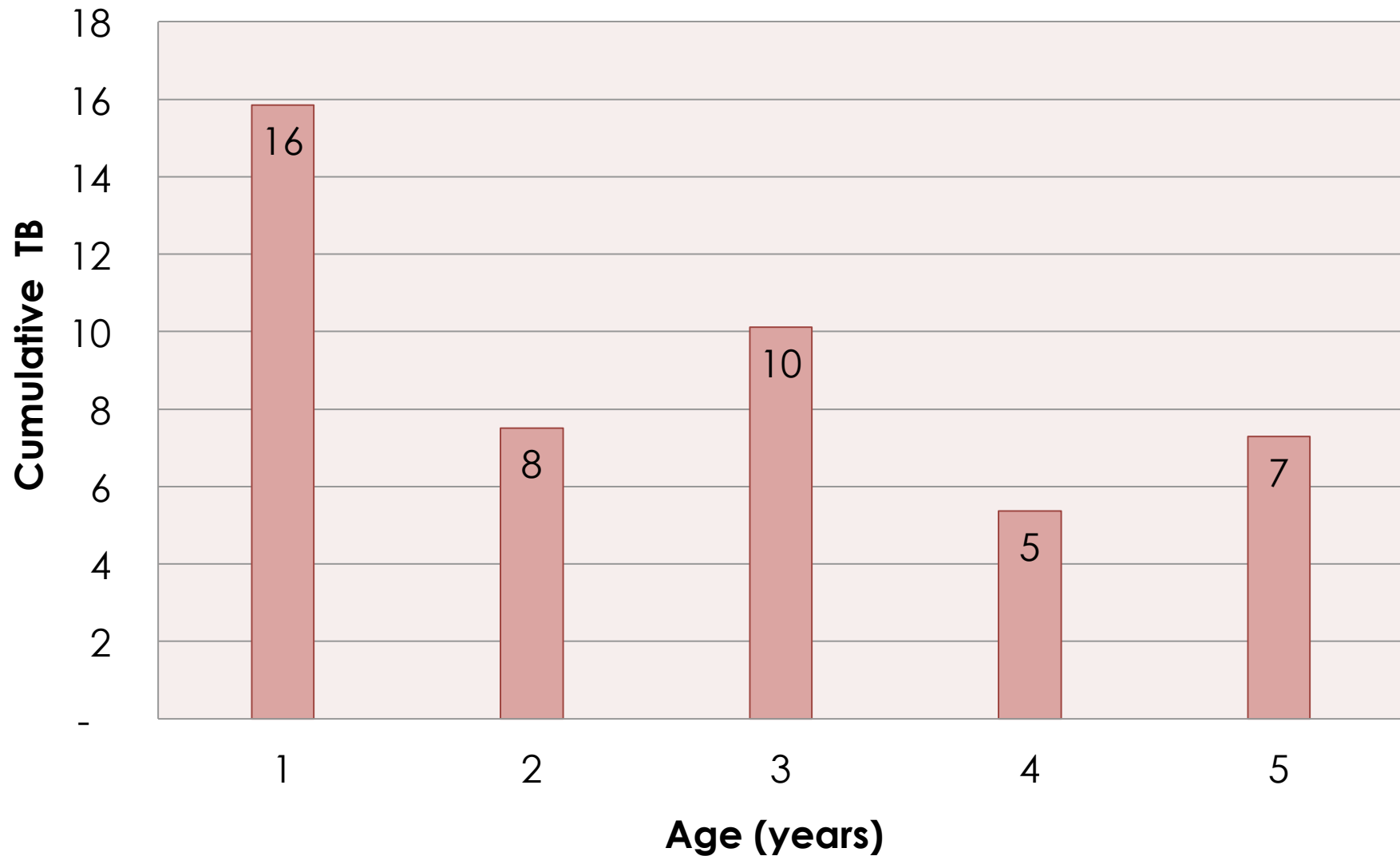
Sample Analysis: Top 10 Users By File Count

User	TB Total Use	Count	Avg Size
1	3.64	39,376,868	92 KB
2	0.13	4,145,415	31 KB
3	1.30	3,282,744	396 KB
4	1.81	1,883,409	961 KB
5	1.66	1,605,967	1 MB
6	0.80	1,516,296	520 KB
7	0.36	1,092,901	329 KB
8	2.21	902,696	2.4 MB
9	0.12	901,450	133 KB
10	0.06	767,108	78 KB

Analysis: File Type Inventory

Type Designator	Total TB	Count	Avg Size
EFIT	2.88	36x10 ⁶	80 KB
GAPProfile	1.53	24x10 ⁶	64 KB
CINE	10.68	5,400	197 MB
CIN	3.65	3,200	1.1 GB

Analysis: Cumulative TB By Last Access Age In Years



Analysis Capability Enables...

- Nirvana SRB gives us ability to identify top 10 users who left files in Globus endpoints
- Improve overall system responsiveness after clean-up
- Ability to set median age of files before administrator will delete files
- Sizing next generation data store



Thank You!



Nirvana