

Tape Data Storage in Practice

Minnesota Supercomputing Institute



UNIVERSITY OF MINNESOTA

GlobusWorld 2018
Jeffrey McDonald, PhD
Associate Director for Operations

© 2015 Regents of the University of Minnesota. All rights reserved.

Minnesota Supercomputing Institute
MSI



HPC Resources



- MSI Users
 - PI Accounts: 843
 - Users: > 4600
- Mesabi
 - Cores: > 18,000
 - Memory: 67 TB
 - Accelerators: 80 K40 gpGPUs
 - Peak: 675 TF
 - 320 Gbps to Storage
- New Technologies
 - FPGAs
 - Nvidia GPUs
 - Intel PHI
 - Storage (Intel NVME)

UNIVERSITY OF MINNESOTA

© 2015 Regents of the University of Minnesota. All rights reserved.

Minnesota Supercomputing Institute
MSI



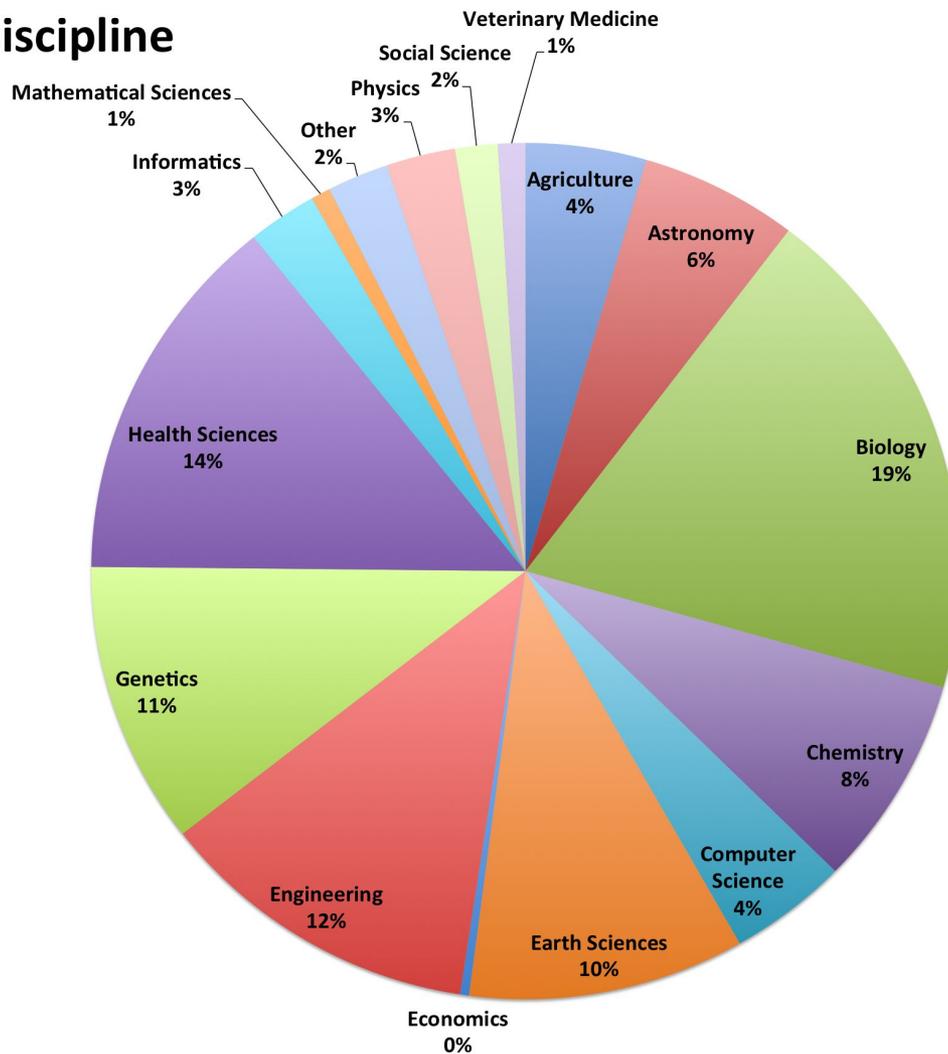
Storage

- High Performance Storage
 - 7.2 PB Usable
 - 48GB/s read/write
 - Available on HPC resources
- Tier-2 Storage
 - 3.1 PB Usable
 - Available via Amazon's S3 interface
 - Available anywhere in the world
- Archive Storage
 - > 3.5 PB tape-based storage
 - offline storage
- All available via Globus



Who Uses MSI?

Storage Allocated by Discipline



- System: Spectra T950, three frames.
- Blackpearl Enterprise model, LTO-7 model.
- Blackpearl has two dedicated LTO-7 drives.
- Tape Library Logically partition between TSM and Blackpearl.
- Dedicated 40 GbE network link to the blackpearl.



UNIVERSITY OF MINNESOTA



Minnesota Supercomputing Institute
MSI

Tape Storage as a Service

- Launched Summer of 2017.
- Three PIs purchased (four) LTO-7 tape buckets.
- Each unit is for a 5-year term, is by-tape, with LTO-7 (assume 12 TB of data storage) and include redundancy.
- Purchase rate for \$456.12 per unit
- Cost estimate based upon occupancy of a full frame.

Policies Considerations

- One 'bucket' per group.
- Tools to chunk data to adequate sizes
 - Duplicity
 - Tar
- Service model always; MSI never relinquishes the tape and we will not export.
- Globus provides data movement channel.

Tape Considerations

- LTO-7 error rates 1/200,000 tapes/1.25 Exabytes (5-9s of reliability), 1000x better than disk (LTO-consortium)
- MSI is conservative with storage, so we elected to make two copies of all user data
- Data can be deleted; space is not recovered unless bucket is re-initialized

Storage Domains				
Name	Tape Used	Tape Allocated	Pool Used	Pool Allocated
All Storage	879.0 MB	5.2 TB	0.0 Bytes	0.0 Bytes
BackupDomain	24.4 GB	10.4 TB	0.0 Bytes	0.0 Bytes
MSI ISO primary copy	10.1 TB	15.6 TB	0.0 Bytes	0.0 Bytes
MSI ISO redundant copy	10.1 TB	15.6 TB	0.0 Bytes	0.0 Bytes
Tape First Copy	29.0 TB	36.5 TB	0.0 Bytes	0.0 Bytes
Tape Second Copy	21.3 TB	26.1 TB	0.0 Bytes	0.0 Bytes

- Configuration uses 'redundant' copy configuration.
- Placement of data is controlled by the redundant copy configuration.
- Prevents 'mixing' of user's data
- Forces the redundant copy.

Breaking down the hardware costs

Item	Quantity	Item cost	Extended cost	% to Service	Term	Cost /Year	TB	Cost/TB/Year
Library Frame	1	25,000	25,000	100.00%	12	2,083		\$0.3245067
Robotics from initial purchase	1	0	0	100.00%	12	0		\$0.0000000
Robotics expansion	1	6,000	6,000	100.00%	12	500		\$0.0778816
Library Maintenance	1	10,000	10,000	100.00%	1	10,000		\$1.5576324
LTO-7 tape	1070	106.00	113,420	100.00%	5	22,684		\$3.5333333
LTO-7 drives	2	10,500.00	21,000	100.00%	7	3,000		\$0.4672897
BlackPearl server	1	100,000.00	100,000	100.00%	5	20,000		\$3.1152648
BlackPearl maintenance contract	1	4,800.00	4,800	100.00%	1	4,800		\$0.7476636
Slot activation	1070	65.00	69,550	100.00%	12	5,796	6,420	\$0.9027778
Total Hardware			349,770			68,863		\$10.7263499

Personnel Costing

- 10% of FTE for Technical Person.
- 10% of FTE for User Support Person.
- 5% FTE for management.

This adds a base of \$4.48 cost per TB per Year.

The total cost per TB (no utilities) is \$15.20 per TB per year (w/redundancy) or 456.12 per 5-years.

Power Considerations

- The archive and blackpearl use about 500W (averaged over duty cycle and share of T950). Independent of Capacity--Tapes don't use power.
- Compare to 6 PB spinning archive, 500-12 TB drives (no redundancy) 5W/drive = 2.5kW of power. Scales with capacity, no servers included here.
- 5 year lifetime savings: \$20k power + \$60k cooling. Saving grow with archive size (assuming at least 5 servers for spinning).

Globus Transfers

- Data set between 2017-06-30 and 2018-04-09.
- 131 Successful transfers.
- 28 Failed transfers (various reasons).
- 40 TB transferred.
- Rates as high as 95 MB/sec and average 30 MB/sec.
- Highly directional, one in 28 from archive.

Users/File Profiles

These are the users who purchased tapes in 2017 and their utilization. The primary concern for us was that the users aggregate their data to limit object counts on the Blackpearl system. MSI staff would spend a couple of hours with each user. Most transfers occur between our tier2 and tier3 storage for these users.

User Dept/Type	Tapes	Files (Size)	Average Size	Comments
Geneticist	2	3862 (16 TB)	4.13 GB	mostly large dna sequence files
Biologist	1	4 (150 GB)	36 GB	project archives
Biologist	1	56 (600 GB)	11 GB	large dna sequence files

Conclusions

- MSI should spend some money on Marketing!
- Globus is providing MSI with important tools toward an HSM-like framework for storage.
- Expanding the framework for some other appropriate use cases.
- Tape continues to offer a compelling use case for storage.

Backup Slides

UNIVERSITY OF MINNESOTA

© 2015 Regents of the University of Minnesota. All rights reserved.

Minnesota Supercomputing Institute
MSI