

Storage Services at the University of Michigan

Advanced Research Computing – Technology Services

Jeremy Hallum

April 2018

Advanced Research Computing - Technology Services: What is it?

ARC-TS provides infrastructure, support for research computing across all U-M academic units.

ARC-TS Storage Services

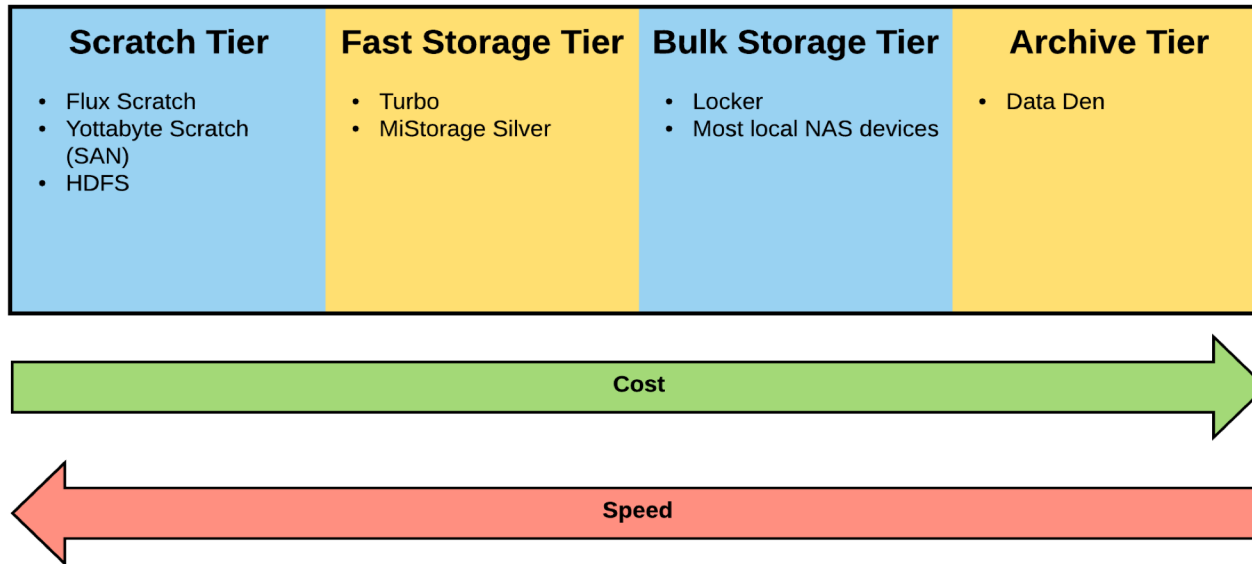
- Available to all units and departments @ UMICH.
- We provide both general purpose and specific-build systems.
- We have to build for all kinds of research.

Storage Strategy

- Build scalable, sensible storage solutions to meet specific demands.
- Encourage people to store their large research data sets with us, not on 'unsafe' devices.
- Enable the easy transport of data across platforms.

UM Storage Tiering

Storage Services at Michigan



'Scratch' Services

- Flux Scratch (Flux)
- HDFS (Flux-Hadoop/Cavium)
- Private Cloud SAN (Yottabyte Research Cloud)
- GPFS (ConFlux)
- All connected by high-performance networks to their associated devices.

Fast Tier Service

- Turbo
 - Flexible and yet high performance. (but also our most expensive)
 - Enterprise Class Isilon.
 - Designed for CIFS/NFS access to our major services for ‘nearly’ active data.
 - Multi-protocol access to data. (CIFS and NFS simultaneously)
 - Can host some types of restricted data, HIPAA. (but not CUI, yet)

Bulk Tier Service

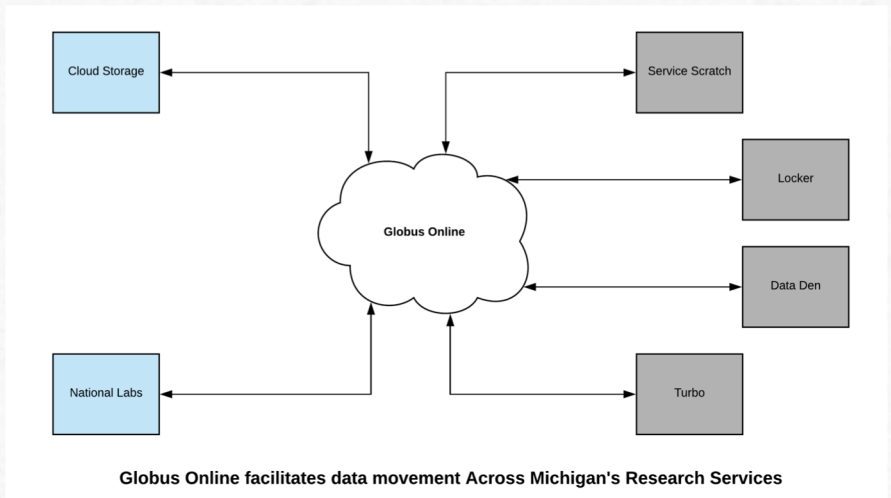
- Locker
 - Balanced for price vs performance vs scalability.
 - DDN SFA14K GPFS system.
 - SMB or NFS, but no multi-protocol.
 - Goal is to aim for CUI (as well as HIPAA)
 - Service is entering its final configuration, but we've already filled the service with 1 PB of data. 30%+ filled, and it's still a pilot.

Archive-Tier Service

- Data-Den
 - For archiving data for posterity (or because your grant agency said so)
 - Balance between data lifetime and low cost.
 - Goal: to replace all of those USB drives that hold data.
 - Service is still being acquired.
 - SMB or NFS, but no multi-protocol.
 - Goal is to aim for CUI (as well as HIPAA).
 - In development (still in procurement).

The glue

- Globus is the glue by which we plan to move data between services.
- It's our most recommended way to move data in bulk between services.
- Our Bioinformatics Core uses Globus Connect to share data with their clients from Turbo.



The Future

- Sensitive data in the cloud.
- Automatic migration across tiers. (It sounds good, but is it the future?)
- Automated and self-service provisioning across resources.
- More coherent campus networking.

Conclusion

- **Advanced Research Computing-Technology Services provides infrastructure, support and coordination for research computing across all U-M academic units.**
- **We have a long road to go, but we have made good progress in a scalable, sensible storage solutions for our researchers.**

Questions?

Jeremy Hallum | jhallum@umich.edu
Research Computing Manager

<http://arc-ts.umich.edu>