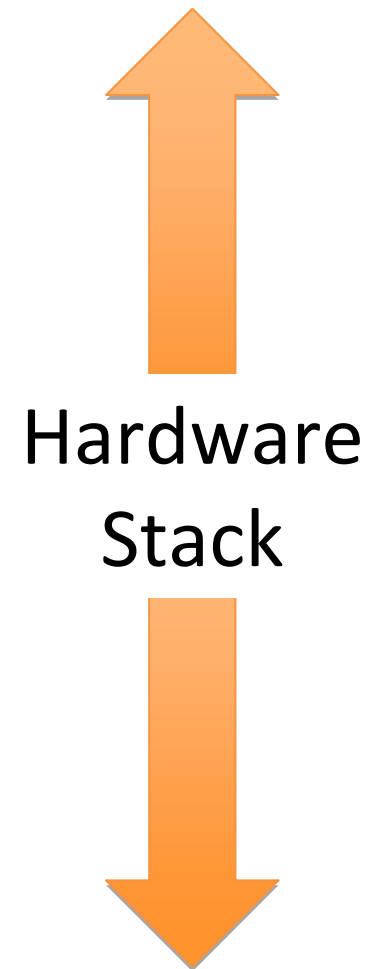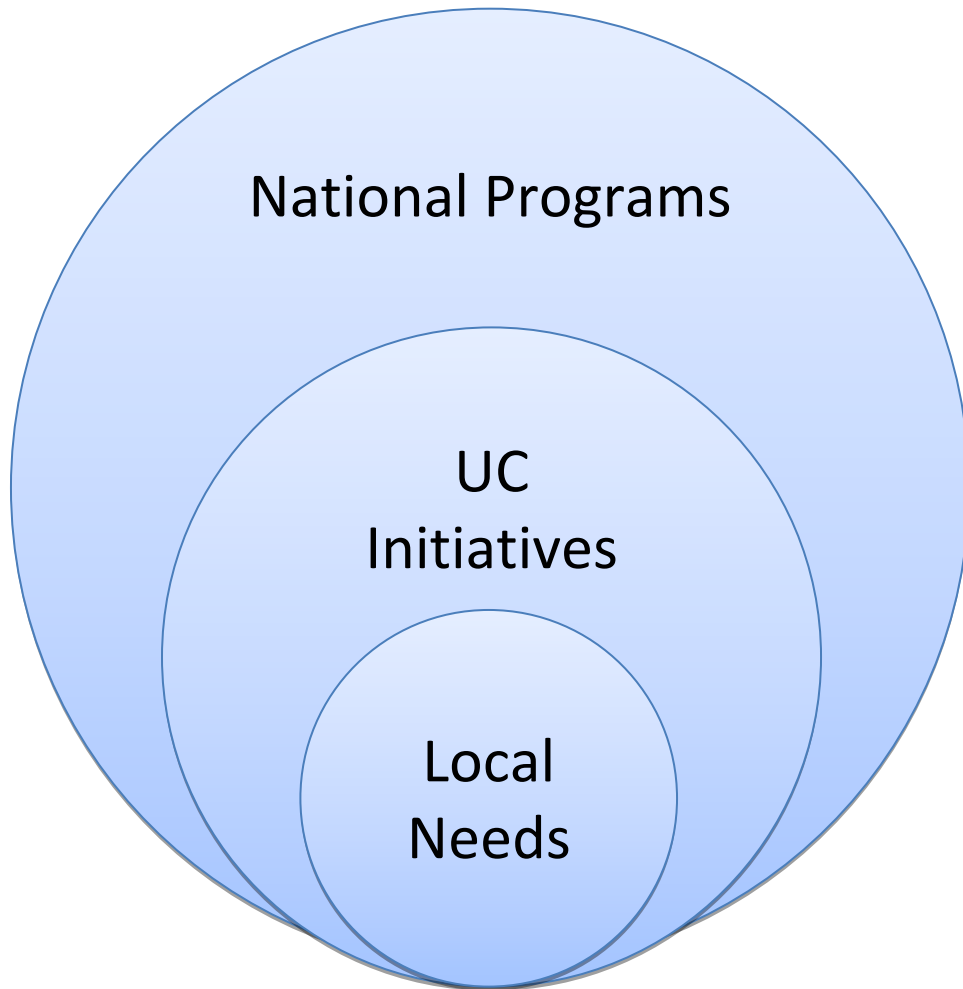# SDSC Storage
New Developments

David Minor

San Diego Supercomputer Center

UC San Diego
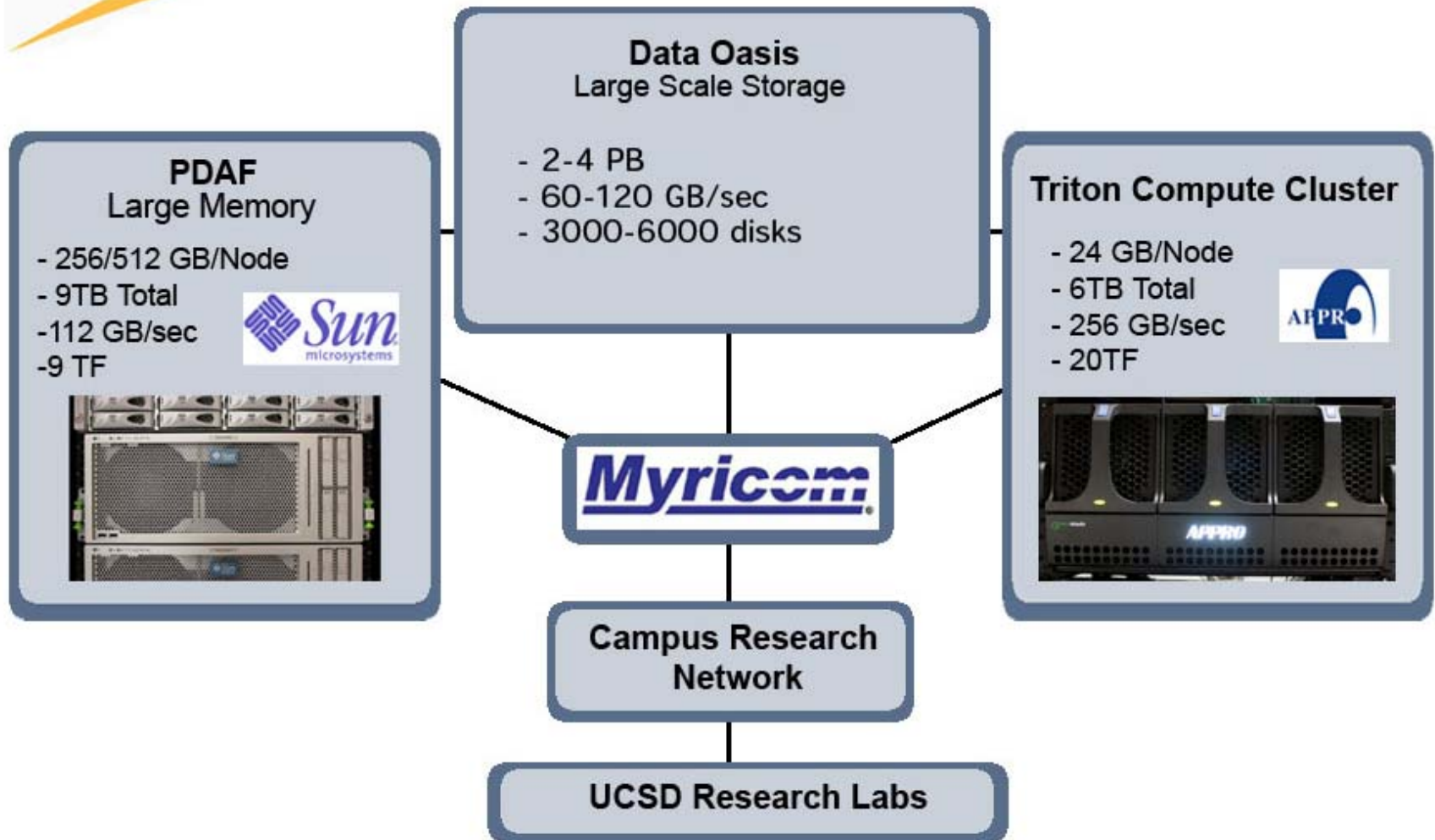
Next generation storage

- Based on feedback from users

- Reflects changing needs

- Opportunity to "start from scratch"

- Continues move from pure computation to data-driven science and research

National Programs

UC Initiatives

Local Needs

Hardware Stack

Data Oasis services overview

High Performance Computing Storage

- Purpose: Transient Storage to Support HPC
- Access Mechanisms: Lustre on HPC Systems, NFS/CIFS for data migrations

Traditional File Server Storage

- Purpose: Typical Project / User Storage Needs
- Access Mechanisms: NFS/CIFS

Archive

- Purpose: Long-Term Storage of Data that will be Infrequently Accessed
- Access Mechanisms: Lustre on HPC Systems, NFS/CIFS for data migrations, iRODS

Service layering

Provided Natively

- Large, scalable file system (HPC, Archive)
- High speed data movement in and out of the system (HPC)
- High speed data movement within the system (HPC)
- Data reliability via replication (Archive, FS)
- Implemented within one organizational entity/group. With single point of contact for users. (HPC, FS, Archive)

Could Layer Well with Software and Funding

- Key-value store and triple stores, e.g. for metadata, ontology, etc. management
- Collection-based (ie user metadata-driven) file management, with hooks to incorporate Data Object Identifiers (DOIs) or other persistent IDs as a stepping stone towards long-term preservation of digital data

Shared with Different Architectures

- Shared-nothing Distributed File System (HDFS)
- Database Management Systems (DBMS)

Data Oasis services SLAs

HPC Storage

- Availability
  - Best Efforts
  - 95% Uptime

- Reliability
  - Single-Copy
  - No Backups
  - NO GUARANTEES

- Performance (Today)
  - Single Node: 500MB/s
  - Aggregate: 2.5GB/s

- Access Mechanisms
  - Lustre (HPC Systems)
  - NFS/CIFS

- Future Options
  - Multiple Sites
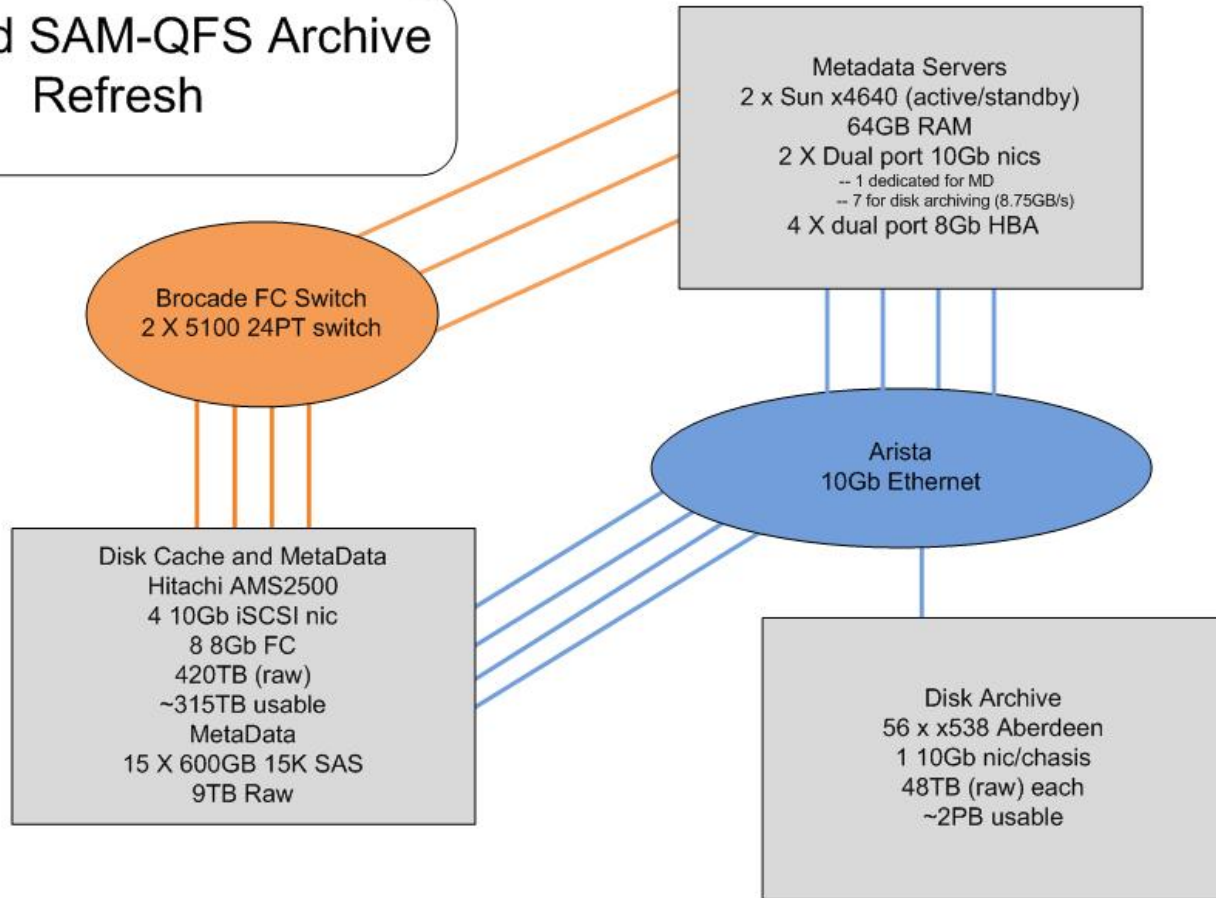 - Clustering

File Server Storage

- Availability
  - 24x7 On-Call Support
  - 99% Uptime

- Reliability
  - Replica Server
  - Snapshots
  - Historical Backups Optional

- Performance
  - Single Server: 300-500MB/s

- Access Mechanisms
  - NFS/CIFS
  - Webfarm

- Future Options
  - Multiple Sites
  - Clustering

Archive

- Availability
  - 24x7 On-Call Support
  - Maximum Uptime

- Reliability
  - Dual-Copy
  - No Historical Backups

- Performance
  - Cache: ~1 GB/s
  - Disk: 300-500 MB/s
  - Tape: 20-100 MB/s

- Access Mechanisms
  - NFS/CIFS
  - iRODS
  - ComVault
  - Web Client?

- Future Options
  - Multiple Site Replication via Chronopolis

Purposed SAM-QFS Archive Refresh

Metadata Servers
2 x Sun x4640 (active/standby)
64GB RAM
2 X Dual port 10Gb nics
-- 1 dedicated for MD
-- 7 for disk archiving (8.75GB/s)
4 X dual port 8Gb HBA

Brocade FC Switch
2 X 5100 24PT switch

Arista
10Gb Ethernet

Disk Cache and MetaData
Hitachi AMS2500
4 10Gb iSCSI nic
8 8Gb FC
420TB (raw)
~315TB usable
MetaData
15 X 600GB 15K SAS
9TB Raw

Disk Archive
56 x x538 Aberdeen
1 10Gb nic/chasis
48TB (raw) each
~2PB usable

Storage Summary:
315TB usable disk cache
Archive copies
1PB stored no compression 2 copies
1.2PB stored 20% compression 2 copies

Also in the pipeline…

- Simplifying SLAs and access mechanisms

- Monitoring national trends (e.g. NSF data requirements, LC initiatives)

- Working with UCSD cyberinfrastructure team

- Working with CDL/UC3 on Merritt

Thank you

David Minor

minor@sdsc.edu