Storage Technology Requirements of the NCAR Mass Storage System

Gene Harano
National Center for Atmospheric Research (NCAR)
1850 Table Mesa Dr.
Boulder, CO 80303

Phone: +1-303-497-1203; FAX: +1-303-497-1848
email: snow@ucar.edu

Presented at the THIC meeting at the Embassy Suites Hotel, Denver, CO July 14, 1999

THIC Inc. The Premier Advanced Recording Technology Forum

Scientific Computing Division

Storage Technology Requirements of the NCAR Mass Storage System

Gene Harano
National Center for Atmospheric Research (NCAR)
1850 Table Mesa Dr.
Boulder, CO 80303

Phone: +1-303-497-1203; FAX: +1-303-497-1848
email: snow@ucar.edu

Presented at the THIC meeting at the Embassy Suites Hotel, Denver, CO July 14, 1999

Outline

- History
- Current Storage Technologies - MSS-III
- Statistics
- Archive Stewardship
- Handling Growth
- Requirements

History: TLIB

- The first computers were CDC 6600 and 7600, running an NCAR developed OS
- Mass Storage was 1/2” tape
- NCAR developed the TLIB tape staging software for the 7600

History: Ampex Tera-Bit Memory System

- The next Mass Store was the Ampex Tera-Bit Memory (TBM) System accessed from CDC 7600 and the Cray 1
- 2” video tape technology, each tape reel weighed 12 pounds and stored 5 GB, the mount/align time was 3-4 minutes and the data transfer rate was 1 MB/sec. Two copies were kept of all data.

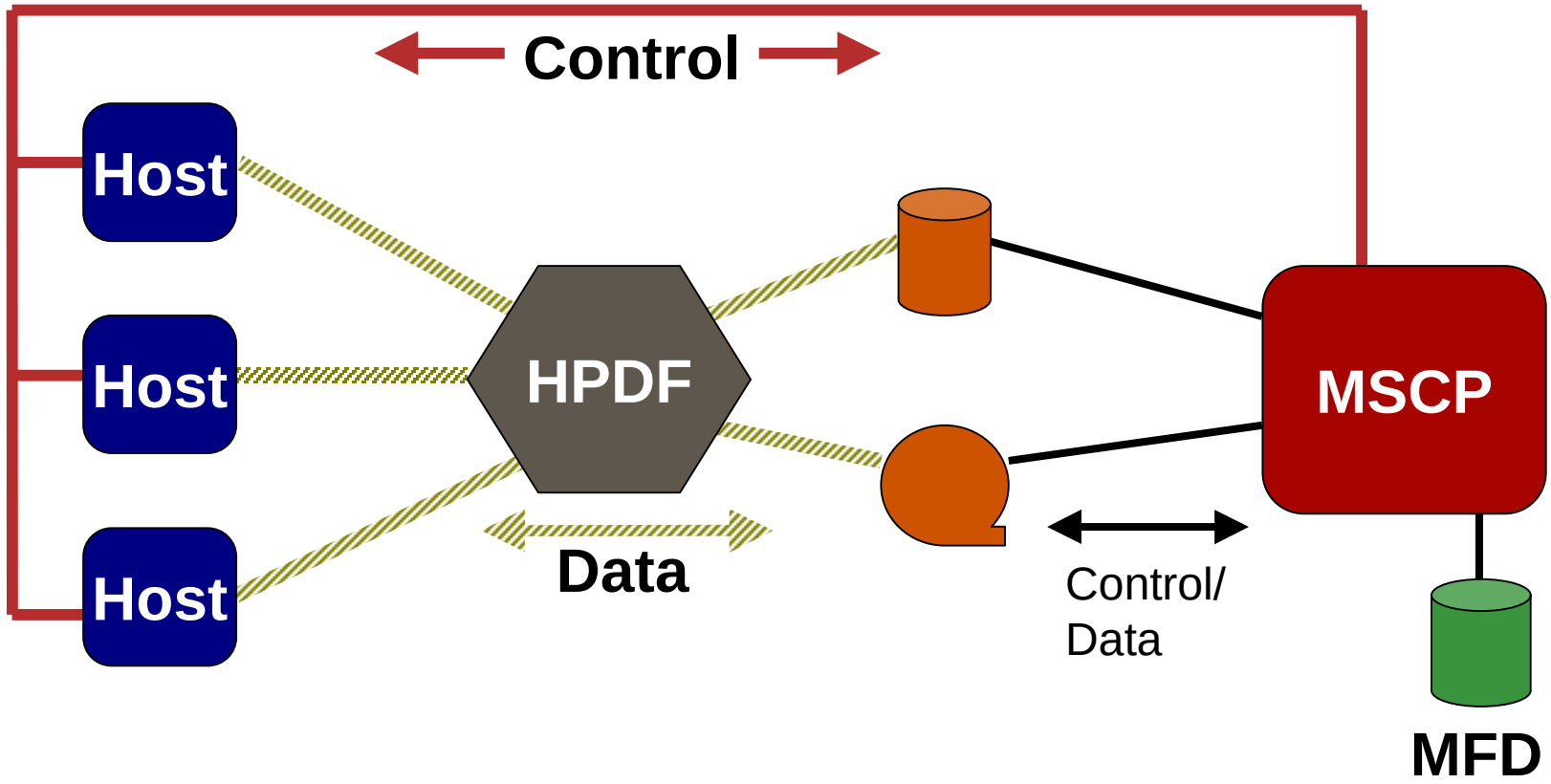
History: MSS III/MSS IV

- The next system deployed was the MSS III, designed by NCAR.
- MSS IV is the evolution of MSS III into a distributed, vendor/platform independent system.

MSS III

- Developed at NCAR starting in 1984
- Operational since 1986
- Based on the IEEE Mass Storage Model Version 2
- First to implement 3rd party transfers to optimize data movement

MSS III: Diagram



Scientific Computing Division

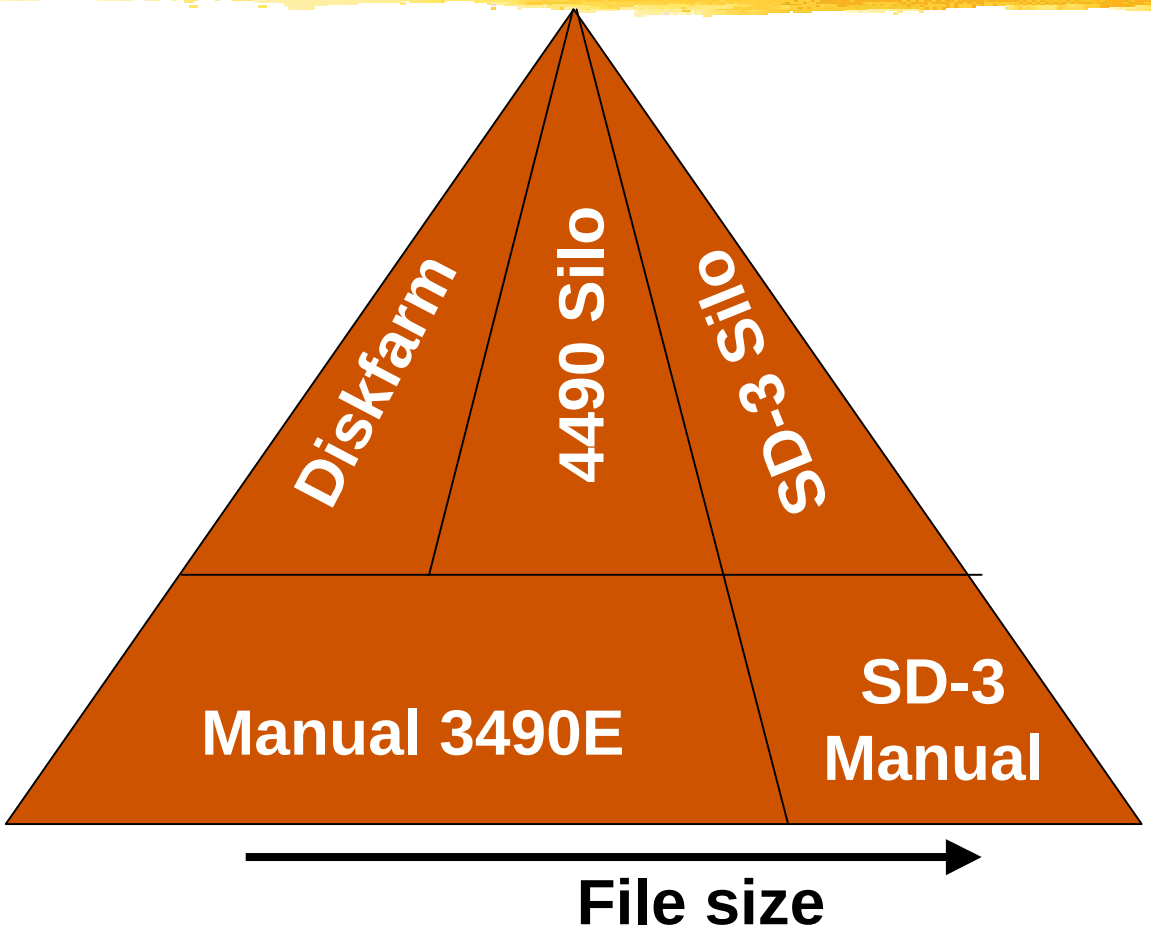
MSS III: Current Hardware Configuration

- IBM 9672-R41 OS390 V1.2 - Mass Storage Control Processor (MSCP)
- Client system based file movers
  - Unicos, IRIX, Solaris, AIX
- High Performance Data Fabric (HPDF)
- Metadata and distributed access services (DCS)

MSS III: Storage Devices

- 180 GB IBM 3390-3 Disk
- 16 STK 4490 tape drives on a STK 9310 Powderhorn Library
- 10 STK SD-3 tape drives on a STK 9310 Powderhorn Library
- 16 IBM 3490E manual mount drives
- 6 STK SD-3 manual mount drives

MSS III: Storage Hierarchy



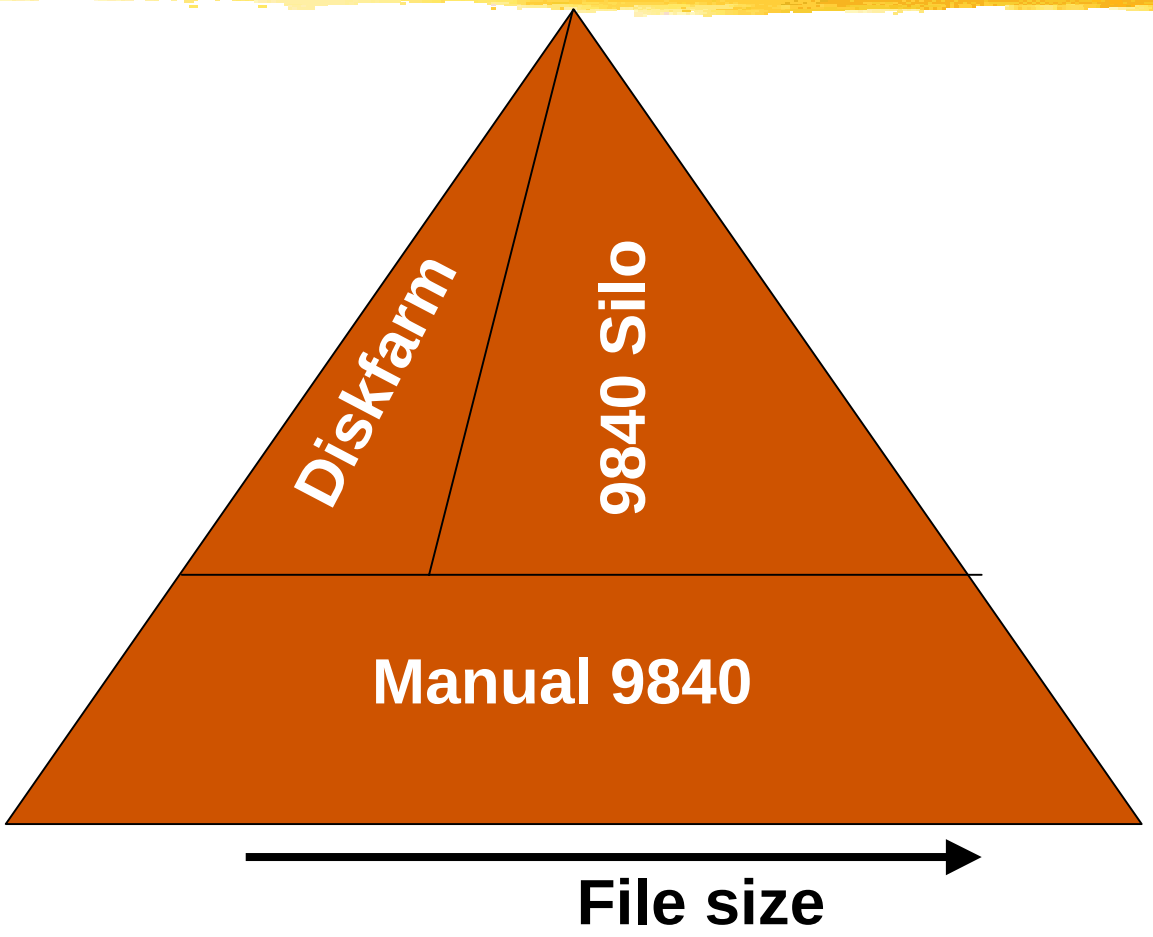
Primary

Secondary

MSS III: Storage Devices

- Introducing STK 9840 technology
- Replace 3490E and SD-3

MSS III: Storage Hierarchy



Primary

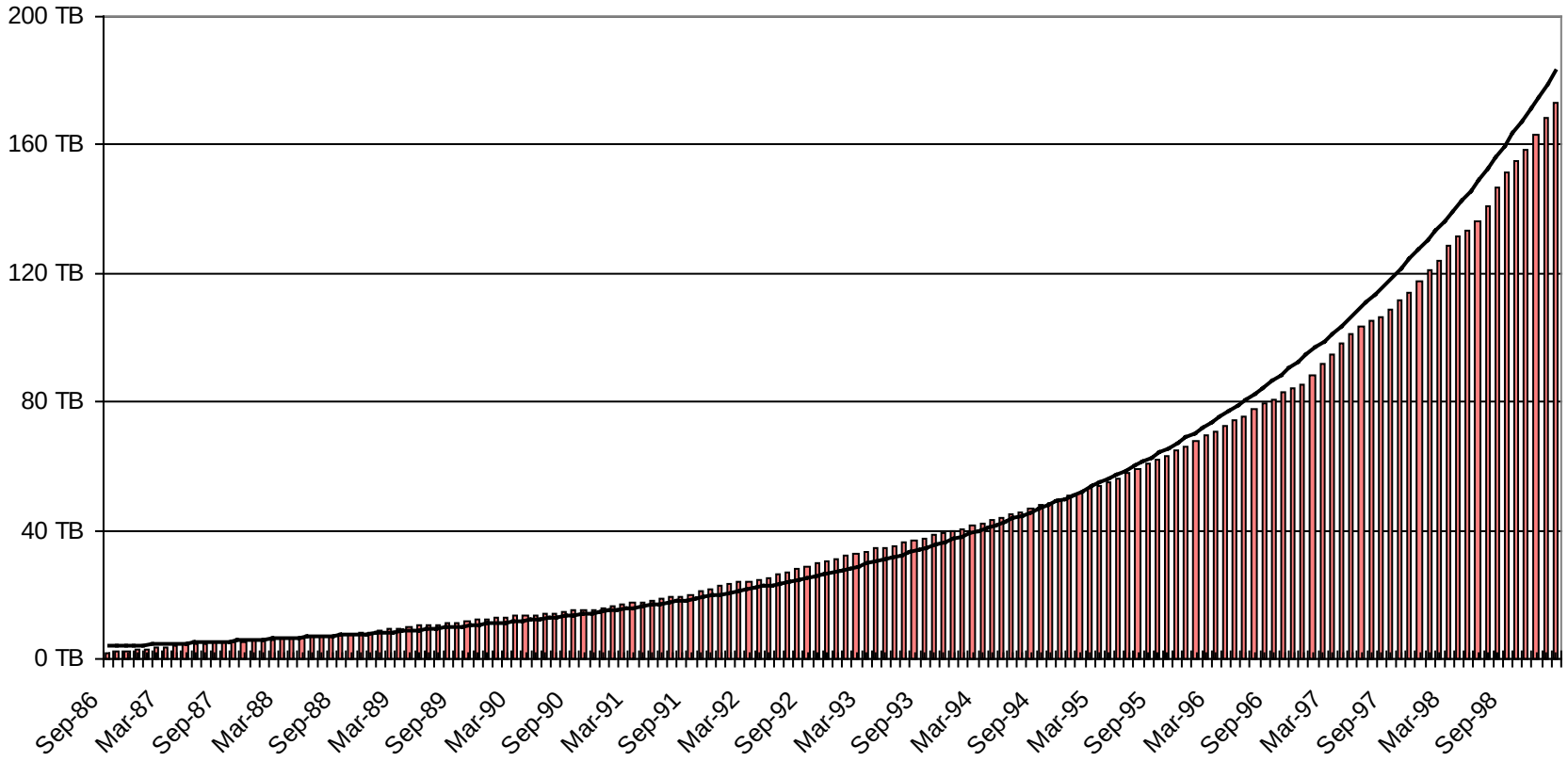
Secondary

Scientific Computing Division

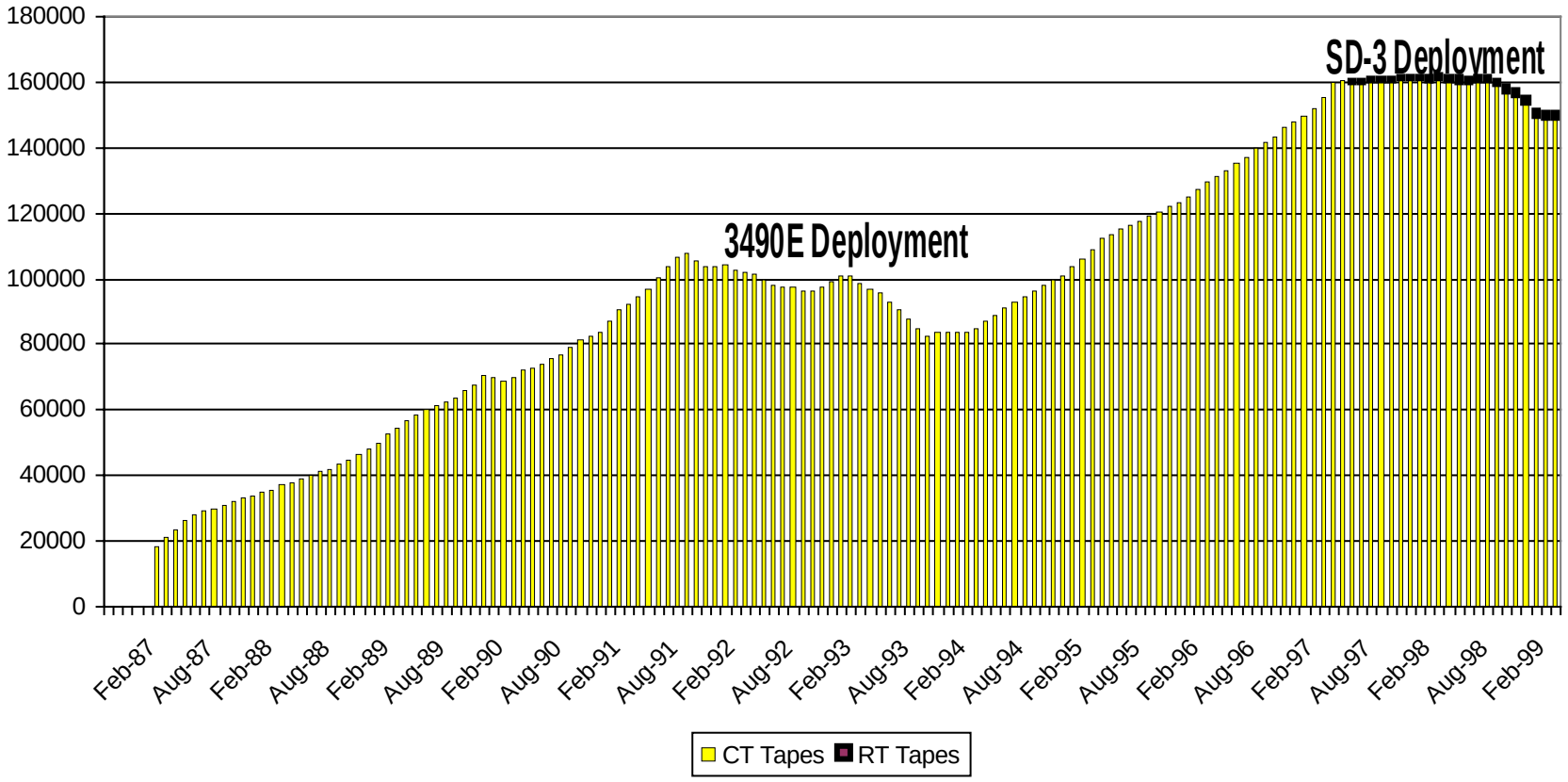
Statistics: July 1999

- 195 TB stored
- 162,000 tape cartridges
- 6,452,948 files
- 15,372,461 MFD entries
- 30 MB average file size
- 21 HPDF attached hosts

Statistics: TB Stored per Month



Statistics: Tape Count per Month



Archive Stewardship

- NCAR has several data archives
  - Geophysical
  - Oceanographic
  - Atmospheric
  - Weather Observational
  - Model Outputs
  - Special Projects Observations

Archive Stewardship: Example

- Historical weather observations have migrated across these types of media:
  - 1800s ship logs, land station logs
  - Punch cards, punch tape
  - 1/2” magtape
  - TBM
  - 3480/3490/SD-3 magtape
- This type of data can not be regenerated

Archive Stewardship: Issues

- How do you keep data accessible when you can not count on the life spans of:
  - Computer hardware/software vendors
  - Media access devices
  - Recording media
  - Documentation of the format of the archived data
- Migrate it!

Archive Stewardship: Migration

- Must migrate before:
  - Recording media disintegrates
  - Recording devices break or die of old age
  - Software used to drive devices and interpret data formats is lost or becomes unusable due to lack of computer hardware
  - Staff depart without leaving documentation

Archive Stewardship: 1st Migration

- In 1986 ~2.7 TB were offloaded from an Ampex Tera-Bit Memory (TBM) system to 3480 technology, taking ~3 months, while the system was accepting new data.
- Read-Only access to the TBM
- Both systems were run in parallel
- Completed only by scavenging parts from the other (decommissioned) TBM sites

Archive Stewardship: 1st Migration

- Lesson learned - Plan for obsolescence
  - What is the useful lifetime of the technology?
    - What is the expected shelf-life of the media?
    - Will a drive/system be available to read the media?
    - What is the useful lifetime of the data?
  - Start migration allowing sufficient time before the system becomes obsolete
  - Worst case - lose half the useful system lifetime

Archive Stewardship: 2nd Migration

- Between September 1991 and September 1993, 20 TBs were migrated from 3480 to 3490E technology while maintaining normal user access to the system and accepting new data.
- Read-Only access to 3480 media
- End-user transparent

Archive Stewardship: 3rd Migration

- STK SD-3 (Redwood) technology was integrated into the NCAR MSS in April 1997
- Sept. 1997 a concurrent migration of ~100 TBs was started from 3490E to SD-3
- The migration was stopped in Oct. 1997 because of problems encountered with the SD-3 drives. The MSS now creates 2 copies of the data.

Archive Stewardship: The Next Migration

- STK 9840 tape technology as replacement for 3490E and SD-3, scheduled for Spring 2000

Archive Stewardship: Maintaining Data Quality

- Reliability
  - Continue to make multiple copies on unproven technology
- Shelf life
  - Data Sniffing samples media quality and refreshes marginal data, purges suspect media.

Archive Stewardship: Data Availability

- Data Ooze - Maintain user access to the data while it is being migrated
  - Background process - Don't want to impact normal user access.
  - Takes longer to complete the migration unless additional resources are committed.

Archive Stewardship: Archive Growth

- Archive Growth is a major concern
- Physical floor space is a finite resource and close to being exhausted
- To accommodate archive growth
  - We will migrate to increase media density and data transfer rate while lowering $/MB
  - Growth reduction policies

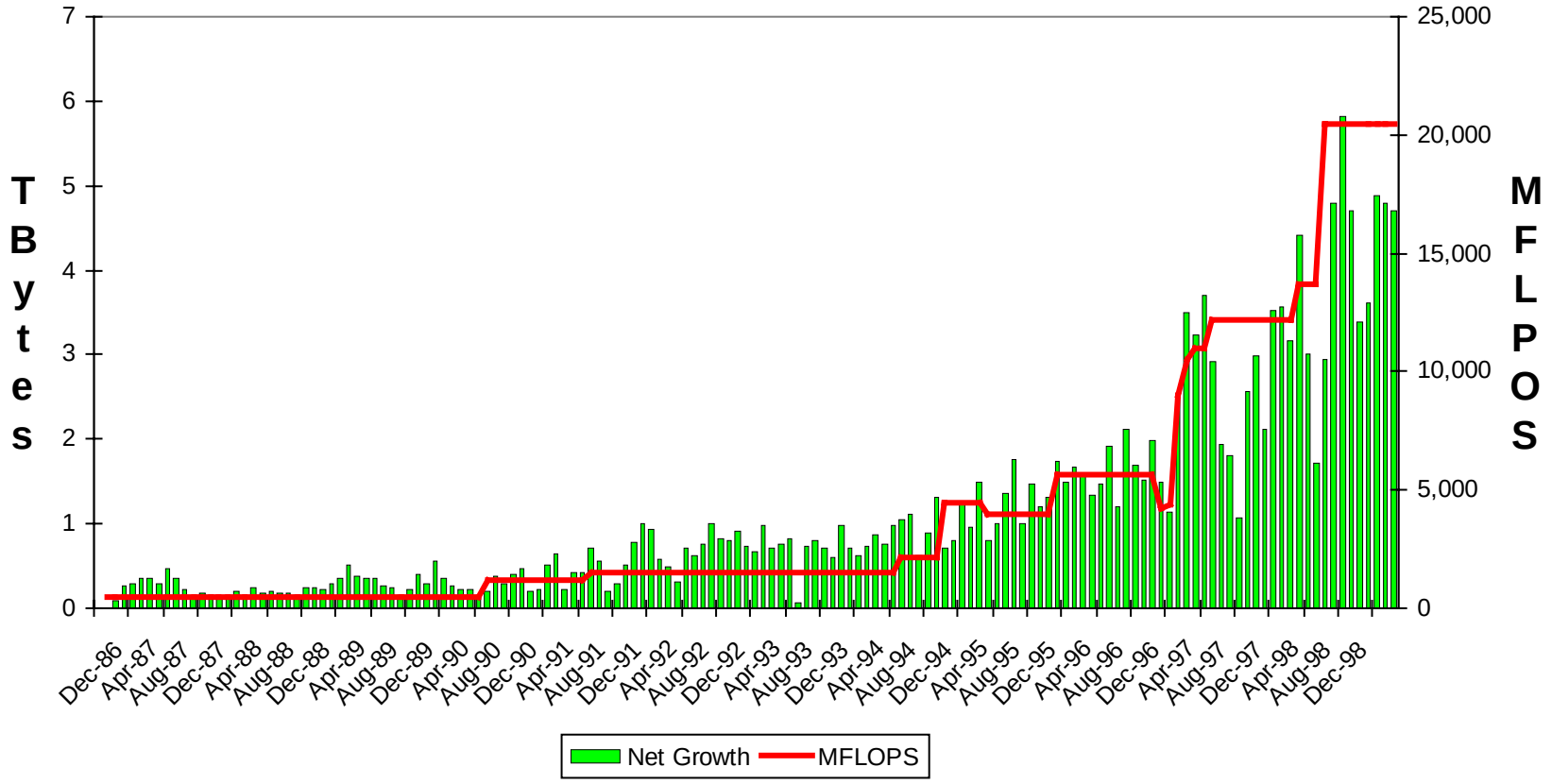
Archive Stewardship: Even possible?

- Perpetual migration
  - Can large archives be completely migrated before the end of the useful lifetime of the technology?
    - Multi-year process
    - Must have backward compatible drives
    - Migrating from multiple media technologies simultaneously

Handling Growth

- Forecast requirements based on past behavior
- Change future user behavior
  - User self-initiated changes
  - Economic (imposed) changes
- Pray media vendors provide necessary improvements to recording technologies

Handling Growth: Net Growth vs MFLOPS



Handling Growth: Projections

	eFY98	eFY99[1]	eFY05[1]
Gflops on NCAR's floor	~20	~40	~1,000
Total Storage (TB)	150	225	5,700
Total Files (10^6)	5.1	7.2	165
Net Growth (TB/month)	5	7	217
Data read/written (TB/month)	25	35	1,300
Data migrated(TB/month)	25	35	1,300

[1]Projected

25 TB/month ~= 10 MB/sec sustained

1,300 TB/month ~= 500 MB/sec sustained

Concerns

- Reliability
  - Corrected Bit Error Rate
  - Media shelf life
- Performance
  - Data Transfer rates
  - Search, load, rewind rates

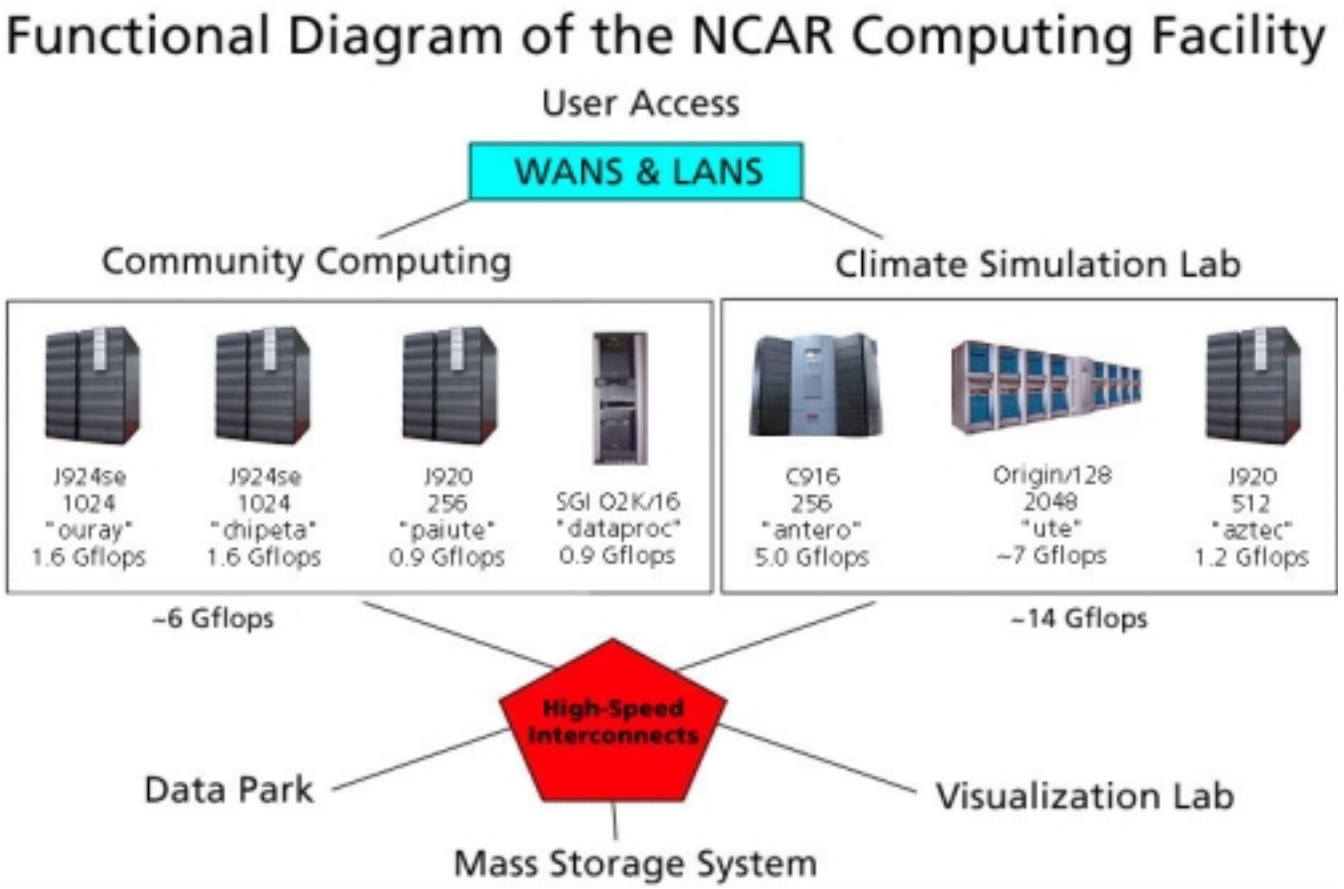
Concerns

- Capacity
- Interoperability
- Vendor Presence
  - Standards
  - Multiple Sources
- Data Sharing in a heterogeneous environment

Requirements: Reliability

Corrected	Transfer Rates		
Bit Error Rate	1 GB/sec	10 GBs/sec	100 GBs/sec
10^{-12}	2 min	12.5 sec	1.25 sec
10^{-13}	21 min	2 min	12.5 sec
10^{-14}	3.5 hrs	21min	2 min
10^{-15}	35 hrs	3.5 hrs	21 min
10^{-16}	14.5 days	35 hrs	3.5 hrs
10^{-17}	145 days	14.5 days	35 hrs
10^{-18}	4 years	145 days	14.5 days
10^{-19}	40 years	4 years	145 days
10^{-20}	400 years	40 years	4 years

Requirements: Data Sharing



Scientific Computing Division

Requirements: Data Sharing

- Global Shared File Systems
  - Storage Area Networks (SANs)
  - Network Attached Storage (NAS)
    - Disk
    - Tape
  - High Performance and Capacity
  - HSM functionality

Summary

Bigger, Faster, Cheaper!

More Info

- www.ucar.edu
- snow@ucar.edu