Architectural Assessment of Mass Storage Systems at GSFC

M. Halem, Code 930.0
J. Behnke, Code 633.1
P. Pease, Code 935.0
N. Palm, Code 931.0

NASA/Goddard Space Flight Center
Greenbelt, MD 20771

Architectural Assessment of

Mass Storage Systems at GSFC

NDADS: National Space Science Data Center
Data Archive and Distribution Service

GDAAC V.0: Earth Observing System Data Information System
Goddard Distributed Active Archive Center

M(DS)^2: NASA's Center for Computational Science
Mass Data Storage and Delivery System

by

Dr. M. Halem, J. Behnke, P. Pease and N. Palm

Space Data and Computing Division
NASA/Goddard Space Flight Center

Presentation for
Goddard Conference on
Mass Storage Systems &
Technologies
NASA/GSFC, Bldg. 8
September 24, 1992

OVERVIEW

- Background
- System Functionality
- Characteristics
- Data Sources
- Hardware/Software Systems
- Performance Assessments
- Conclusions

BACKGROUND OF MASS DATA STORAGE SYSTEMS

NDADS:

Prototype of the Hubble Space Telescope Data Archive and Distribution Service (HST-DADS) contracted to Loral AeroSys in 1989. Evolved as the Astrophysics and Space Physics archiving system for the National Space Science Data Center to maintain a mix of near and on-line data and manage a deeper data storage archive

GDAAC/V.0:

EOS prototype archive and distribution systems initiated in FY91 and planned for operational availability in FY94. One of nine geographically distributed discipline-oriented interoperable DAAC's

M(DS)^2:

A mass storage and delivery system serving more than 1400 users within the NASA Computational Science Center at Goddard that has to manage both the high-speed computer-generated simulation data, as well as space-borne observational data

SYSTEM REQUIREMENTS

	NDADS	GDAAC V.0	M(DS)^2
NEAR-ONLINE STORAGE /DEEP	2.6 TB/6 TB 16GB DASD	10TB/3TB 16GB/DASD	7TB/35TB 240 GB/DASD
SCALABLE UP TO	10 TB/50TB 100 GB DASD	18 TB 100 GB/DASD	225 TB/500 TB 3 TB/DASD
INGEST (RATE)	13 GB/DAY	30 GB/DAY	90 GB/DAY
DISTRIBUTE (RATE)	1050 MB/DAY - NET 700 MB/DAY - TAPE 100 PHOTOS/DAY	150 GB/DAY	100 GB/DAY
PEAK CONCURRENT USERS	146 240 CATALOG QUERIES/HR	100	MIN. 128 MIN. 32 SIMULTANEOUS FTP TRANSFERS

SYSTEM FUNCTIONAL CHARACTERISTICS

Data and Metadata Functions	NDADS	GDAAC	MDSDS
Network Access (Ethernet, FDDI, DecNet, UltraNet)	x	x	x
Security	x Barrier	x Barrier	x RACF/C2
Integrity and Quality Control	x	x	
Automated Data Migration and Compaction	Partial		x
User Ingest and Retrieval	Partial	Partial	x
Remote Ordering and Delivery Service	x	x	
Catalogues and Inventories	x	x	User
Browse - On-Line	x	x	
Interoperability		x	Partial
Database Queries and Subsetting	x	x	
Portable Software Operating Systems		x	x
Incorruptible Archive	x	Partial	Partial
Remote Back-Up/Safe Store	x		———
Data Compression	x	x	———
Redundancy (NSPOF)	Partial	Partial	x
Scalable Upgrade	x	x	x
Accounting & Monitoring	x	Partial	x

NDADS Data Sources:

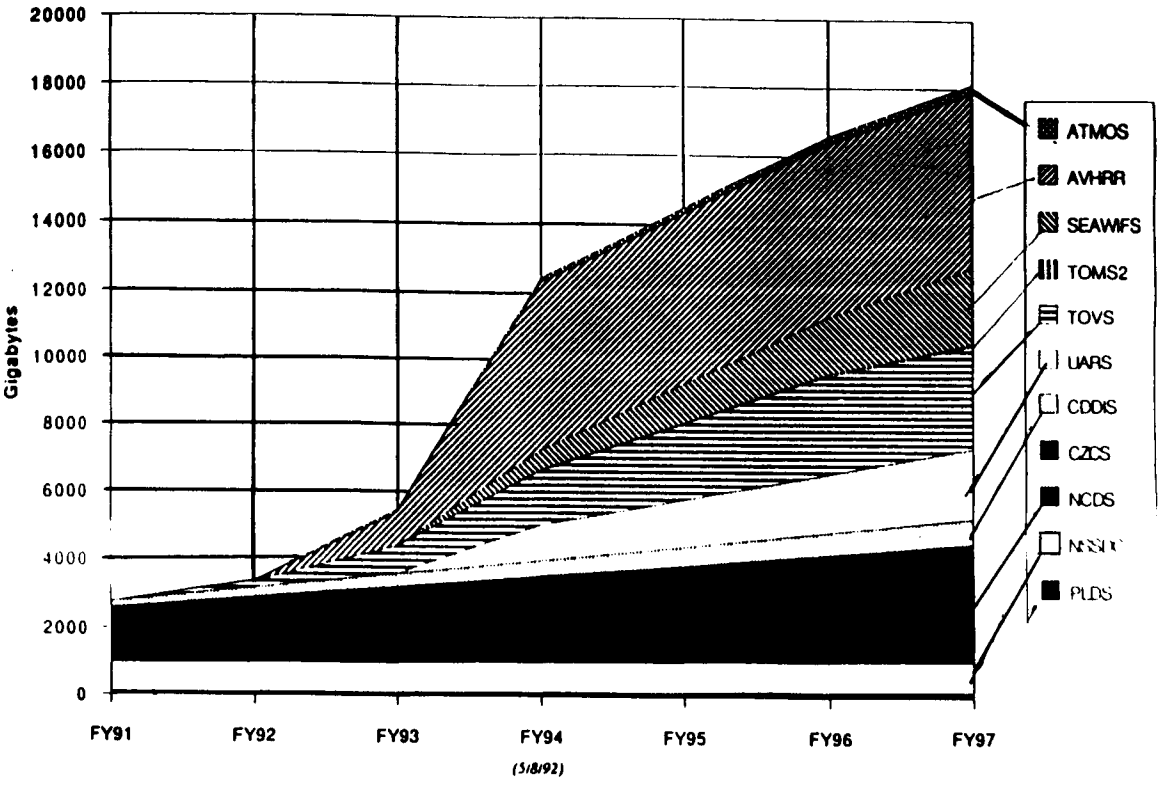
Astrophysics:

Wave Length	Project	DataTypes	Granules	Archived	Total Size
High Energy	EXOSAT	10	34K		150 GB
	HEAO-3	2	10K	100%	4 GB
	HEASARC	200	100K		20GB
	EINSTEIN	20	15K	30%	120 GB
	ASTRO-D	launch 1993	—	—	1650GB
	XTE	launch 1996	—	—	
	GRO	100	—	none	50GB/year
	ROSAT	60	> 100K	30%	100 GB
	VELA 5B	1	1K	100%	3GB
Ultra Violet	IUE	6/8	80K	100%	200 GB/750GB
Copernicus	EUVE	launch 1992	—	—	50 GB
Optical	HST	4			1GB/4GB
Infrared	COBE	6	100K	0	30 GB
	IRAS	6	150K	100%/10%	5GB+78GB
Radio	VLA	1	—	60%	1000 GB
Miscellaneous	ADC Catalogs	Various	1K	100%	1GB
TOTAL					2.41 + (1.75)

Space Physics:

Mission	Data Types	Granules	Archived	Total Size
SKYLAB	1	3500	100%	10 GB
ISTP-GOES	8			.4 GB
ISTP-IMP8	8			.4 GB
ISTP-GEOTAIL	8			1.0 GB
Atomic Physics	2	100	10%	1 GB
DE-1				100 GB
VOYAGER				2 GB
TOTAL				114 GB

GSFC V0 DAAC Data Volume Requirements (by Project)



Science Project Data Products

Project	Product Description
UARS	• Profiles of 15 trace species, temperature, and wind • Solar UV irradiance measurements (115 - 400 nm)
SeaWiFS	• Ocean pigment, chlorophyll a concentrations • 5 water leaving radiances, 3 aerosol radiances • Diffuse attenuation coefficient
Atlas / ATMOS	• Profiles of 30+ upper atmosphere trace species • Upper atmospheric temperature profiles
TOMS2	• Total ozone, effective tropospheric reflectivity • 6 backscattered UV radiances (313 - 340 nm)
AVHRR Pathfinder	• Binned 5 channel clear sky radiances • Daily cloud fraction, height, and reflectivity at 9 km and 1 degree spatial resolutions • Daily, weekly and seasonal surface reflectance NDVI at 9 km resolution • Daily Surface albedo at 9 km resolution • Aerosol optical thickness, longwave surface flux
TOVS Pathfinder	• Profiles of atmospheric temperature, humidity, and geopotential height • Precipitable water in 6 tropospheric layers, total ozone, and tropopause pressure • Surface air and skin temperatures, 3.7 micron bidirectional surface reflectance, and 50 GHz surface microwave emissivity • Cloud fraction, cloud top pressure, precipitation estimate, visible reflectance, outgoing longwave radiation, and longwave cloud forcing

M(DS)2 DATA SOURCES

PROJECTS: NIMBUS/TOMS
ISTP
IUE
GRO

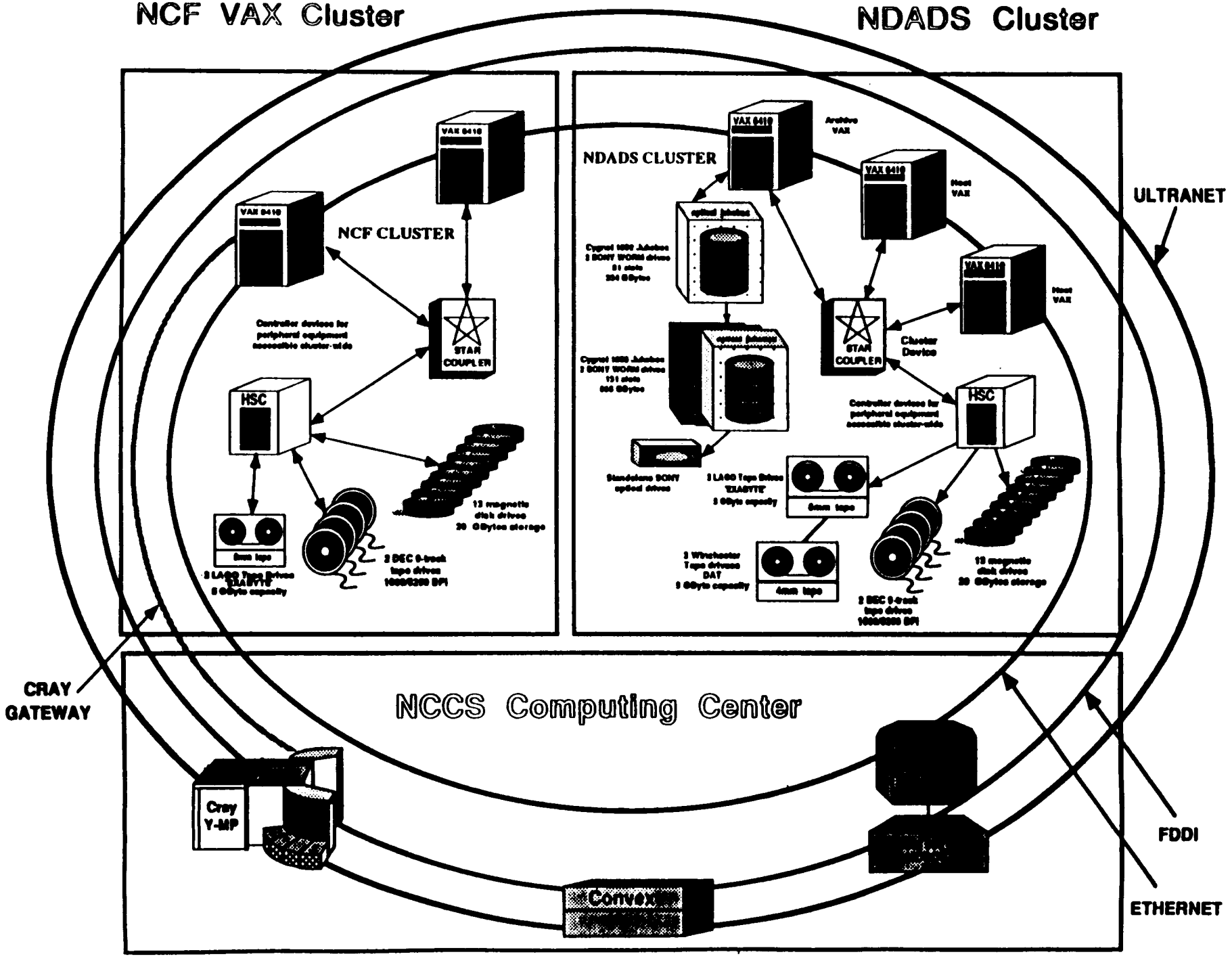
MODELING: DATA ASSIMILATION
COUPLED OCEAN/ATMOSPHERE/STRATOSPHERE
GEODYNAMICS
SPACE PHYSICS PLASMA MODELING

ANALYSIS: TOVS PATHFINDER
ALGORITHM DEVELOPMENT (TRMM, MODIS...)
HST IMAGE DEBLURRING

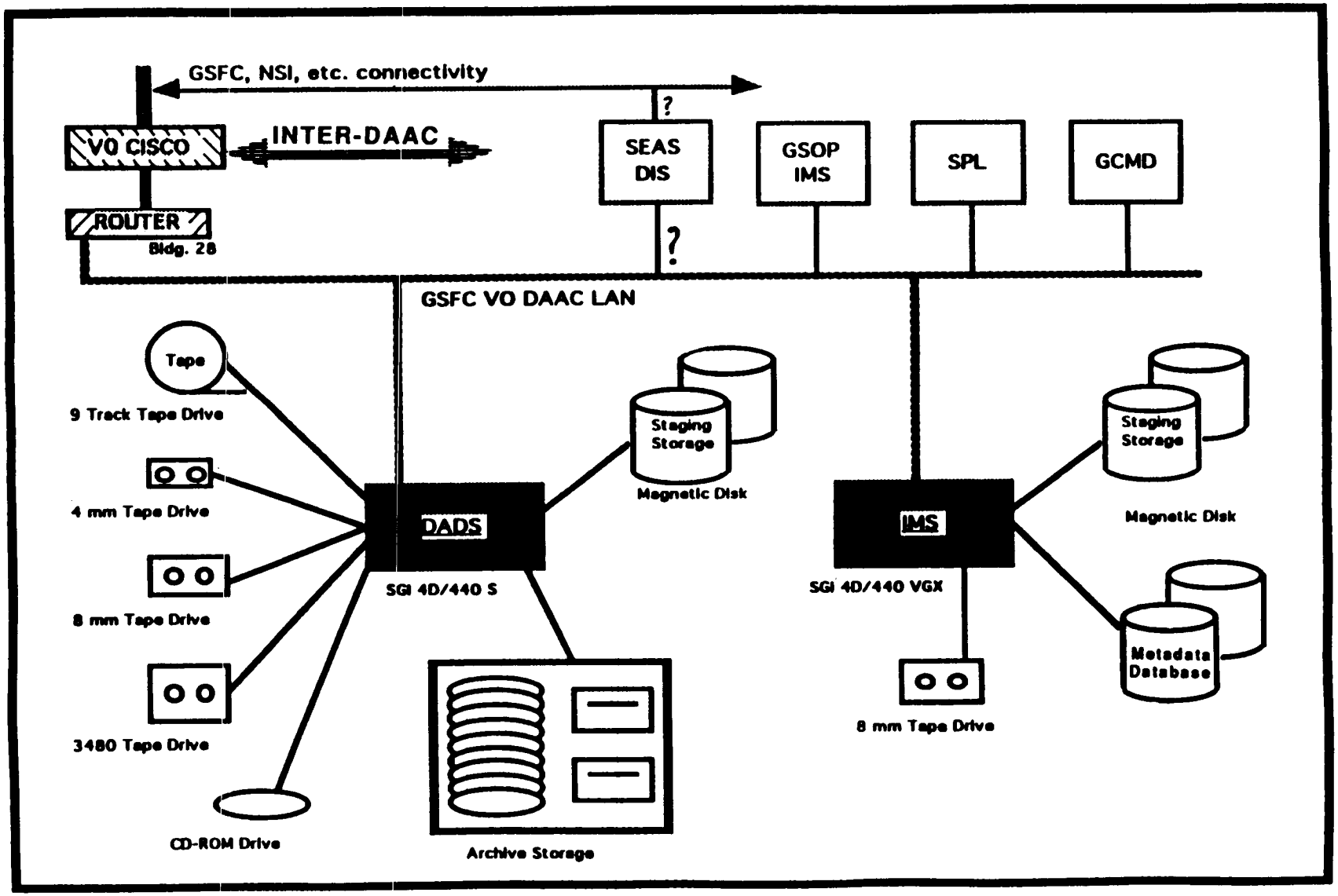
HPCC: EARTH AND SPACE SCIENCE TESTBEDS

Mass Storage Hardware Systems

NDADS	GDAAC	MDSDS
• 3 VAX 6410's (14GB/DASD)	• 2 SGI 4D/440 (16 GB DASD)	• IBM ES 9021/500 (56 ch, 128 MB • IBM 3980 - (240 GB/DASD) • Convex 3240 - (512 MB)
• 2 CYGNET WORM Jukeboxes with 4 SONY drives	• 2 CYGNET WORM Jukebox with 4 ATG 9001 drives (24 TB) • 1 Metrum ACS (8.7TB) w/ 4 drives	• 3 STK 440 (4.8TB) • 1 Dataware WORM Jukebox 3 34/850 (1.2 TB) --- • B-Test Helical E-Systems Tower (8.2 TB)
• 2 9-track Dec tape drives • 2 8mm Exabyte tape drives • 2 4mm Winchester tape drives • CD-ROM pre-mastering	•1 I/O power channel 80MB/s •5 6250 9 track tape •4 Exabyte 8 mm tape drives •2 3480 tape cartridge • CD-Rom pre-mastering • 2 4mm DAT tape drives	• 8 ESCON Channel • Ultranet • 40 Memorex 3480 compatible tape drives

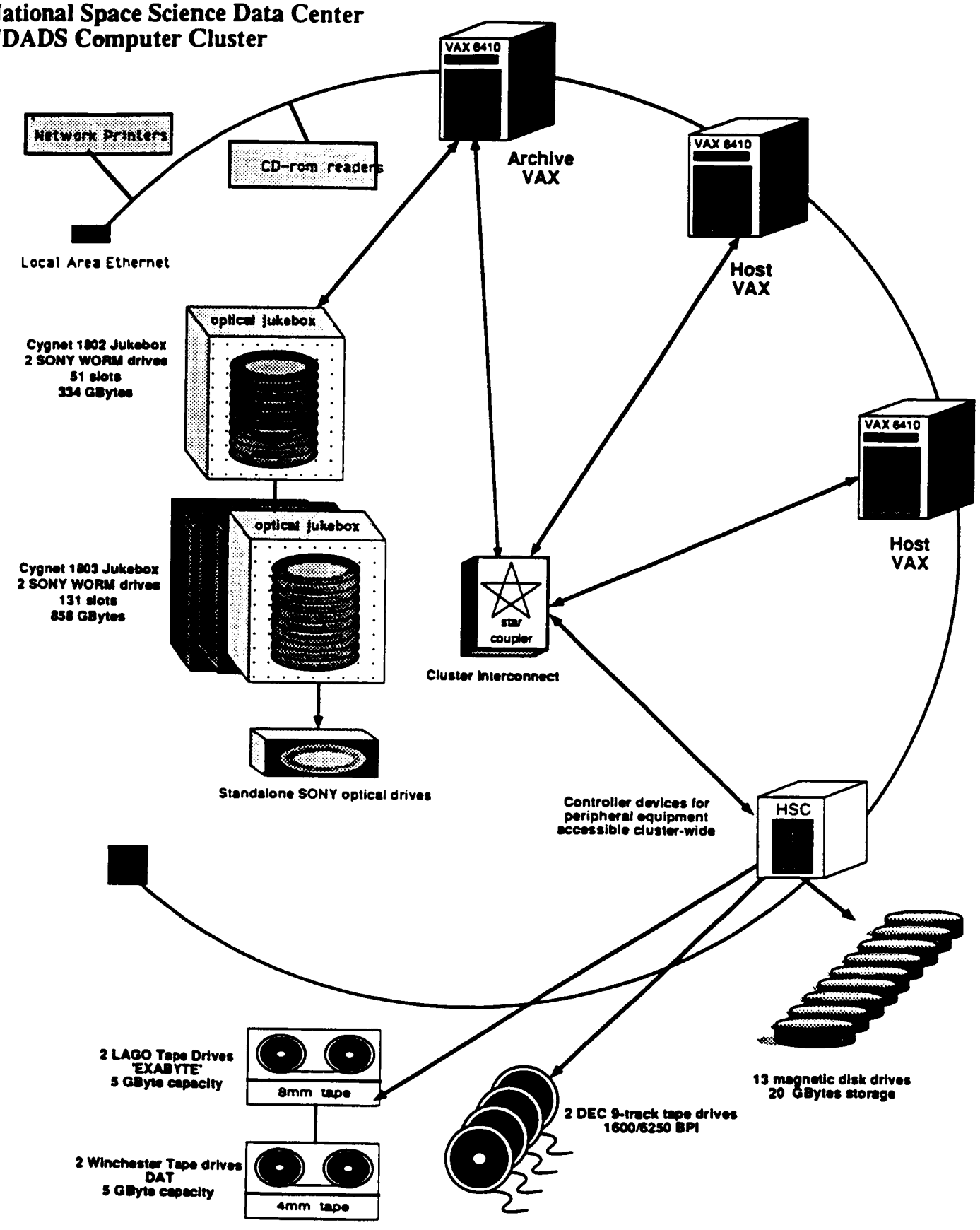

NCF VAX Cluster
NDADS Cluster
NCF CLUSTER
NDADS CLUSTER
STAR COUPLER
HSC
ULTRANET
CRAY GATEWAY
NCCS Computing Center
Cray Y-MP
FDDI
ETHERNET

Hardware Architecture Components



Goddard DAAC

National Space Science Data Center
NDADS Computer Cluster



Mass Storage Software Systems

Software System	NDADS	GDAAC	MDSDS
Client/Server Operating System	VAX VMS	IRIX	MVS, AIX
Networks Supported	DECnet TCP/IP SPRINTnet	TCP/IP	TCP/IP Ethernet BITNET NSFNET InterNet InterLink UltraNet
Database/Library Management System	FSTAGE/FSTORE SYBASE INGRES	Unitree ORACLE	HSM Unitree Oracle (Opt.)
Physical Storage Device Driver Software	JIMS SOAR	Unitree Drivers	Dataware STK Unitree Drivers
User Interface	NCDS ARMS* NSSDC ARCHIVE	NCDS/PLDS EOSDIS/IMS	USER CONTROL

* AUTOMATED RETRIEVAL MAIL SYSTEM

System Assessments

	NDADS	GDAAC	MDSDS
Strengths:	• Project customization • FTP Accessible • Distribution of archive media • Intelligent data access and optimization • Data compression • Metadata search/browse • Remote back-up	• Project customization • FTP Accessible • Interoperable • Metadata search/browse NCDS/PCDS Experience Intelligent Data Mgmt. SpatialTemp Data Fusion — • Metadata search/browse • IEEE Mass Storage Compliant • Open	• FTP Accessible worldwide • Archival/retrieval by user request • Intelligent hierarchical storage migration • Remote back-up option • IEEE Mass Storage Compliant
Weaknesses:	• Non-portable systems • In-house customization of software and hardware	• In-house customization of software • Embryonic HW/SW systems	• Embryonic HW/SW systems • Costs

PERFORMANCE CHARACTERISTICS

THESE ARE ALL PRELIMINARY ESTIMATES FOR THE NDADS M(DS)^2 AT THIS POINT!!

(DOES NOT INCLUDE PROJECT SPECIFICS)

NDADS

SONY Optical Disk Drives:	Read:	600 KB/sec
	Write:	300 KB/sec
Actual Rates:	Read:	250 KB/sec
	Write:	107.52 KB/sec (average)
		250 KB/sec (max)
Platter load speed:		11 seconds
Data Storage: Current:		120 GBytes
Growth:		240 GBytes/year
Storage Input:		5 GB/day (average), 17 GB/day (max)
Storage Output:		330 files/day, electronically
Inquiries on the archive:		24/day

M(DS)^2

DASD:	80 MB/S Throughput
Mass Storage:	18 MB/S Throughput

CONCLUSION

- Mass storage systems allow scientists to perform research previously impossible because of logistic burdens and maintain pace with rapid data growth arising from increasing computational power and observational resolution

- Mass storage hardware systems technology evolving faster than software available to integrate into system. IEEE mass storage standards model changing faster than vendors can keep up with; still need standards.

- Community needs to acquire much more performance test data reliability, stability and data access speeds across small to large mass storage systems

- Not yet clear whether many small distributed client-servers are more effective than fewer large-scale client servers

- Mass storage management systems need to become more robust and more stable