UMIACS
University of Maryland Institute for Advanced Computer

THIC Inc.
The Premier Advanced Recording Technology Forum

Digital Archiving and Long-Term Preservation An Early Experience with Grid and Digital Library Technology

Fritz McCall - University of Maryland
Robert Chadduck - National Archives and Record Administration
Collaborators: Joseph JaJa, Mike Smorul (UMD), and Reagan Moore (SDSC)
E-mail: fmccall@umiacs.umd.edu
2135 AV Williams Bldg, College Park MD 20742
Phone:+1-301-405-6757 FAX: +1-301-314-0568

Presented at the THIC Meeting at the Raytheon ITS Auditorium
1616 McCormick Drive, Upper Marlboro MD 20774-5301
October 26-27, 2004

Two Major Projects

- Persistent Digital Archives Project (SDSC/UMD/NARA)
  - Development of a technology framework for digital archiving and preservation, and a prototype persistent archive.
- Global Land Cover Facility
  - An Earth Science Digital Library with innovative information services and a large user base.

Persistent Digital Archive: Overall Architecture

- Digital object model that encapsulates content, structural, descriptive, and preservation metadata.
- Layered software architecture: data layer, information management, and access and presentation.
- Distributed architecture based on data grids (Storage Request Broker – SRB)

UMIACS
University of Maryland Institute for Advanced Computer Studies

Systems Infrastructure

Heterogeneous "grid bricks" at SDSC, UMD, and NARA, linked with the SRB.

Site	NARA	UMD	SDSC (2003)	SDSC (2004)
Name	DELL 2560	Dell 2550	Custom	Custom
CPU	2-1.4 Ghz	2-1.4 Ghz	1-1.7Ghz	1-2.8Ghz
CPU type	Pentium 3	Pentium 3	Celeron	Pentium 4
Memory	4 GB	4 GB	1 GB	1 GB
Disk controller	FastT200	FastT200, JetStor	3Ware	3Ware
Disk connect	2Gb FC DASD	2Gb FC SAN	IDE	IDE
Disk size	1.1 TB	4.1 TB	1.1 TB	5.25 TB
Disk drive	73 GB	73 GB ,160GB	160 GB	250 GB
Disk RPM	10000	10000, 7200	5400	5400
Throughput	166 MB/s	166 MB/s	110 MB/s	110 MB/s
Network	10/100/Gig-E	10/100/Gig-E	100/Gig-E	100/2-Gig-E
OS	Linux	Linux	Linux	Linux

11.5 Terabytes of Distributed Storage

Software Configuration

- The Storage Resource Broker ensures that data can be securely accessed from any site.
- The Grid Security Infrastructure supports uniform cross-site authentication through a Certificate Authority run by NARA.
- The Metadata Catalog (MCAT) federates heterogeneous catalogs running at each site. SDSC and NARA run Oracle. UMD runs Informix.

Some Observations

- Commodity grid bricks provided:
  - VERY low cost disk storage:
    - $4,000/TB in 2003 -- $2,000/TB in 2004.
  - Adequate performance for TCP/IP and data manipulation procedures (eg. Subsetting, aggregation, and metadata extraction)
- They required:
  - a detailed understanding of the low-level system properties.
  - higher operational vigilance. RAIDs were monitored daily in order to resolve all problems as soon as they occurred.

Back-Up Systems

- Collections written to the grid brick at UMD were archived into an archival storage system using the Tivoli Storage Manager as a Disaster Recovery Plan.
- TSM runs on an IBM P630 Server with one Terabyte of disk cache, an ADIC Scalar 10K tape library, and six LTO tape drives.

Lessons Learned

- We rescued a 1TB collection of images that should have been fine:
  - The media was standard and very reliable.
  - The tape drive and library were standard SCSI devices
  - Even the operating system was still supported
  - The problem was caused by the format of the disks!
- The complexity of any single system makes it difficult to gauge its ability to preserve data over the long-term – hence the need for heterogeneous and platform independent configurations.

UMIACS
University of Maryland Institute for Advanced Computer Studies

The Global Land Cover Facility

- Overview: The GLCF provides access to over 13TB of remotely sensed data and related derived products at http://glcf.umiacs.umd.edu.

- Mission: encourage the use of remotely sensed imagery to better understand the nature and causes of land cover change and its impact on the Earth’s environment.

UMIACS
University of Maryland Institute for Advanced Computer Studies

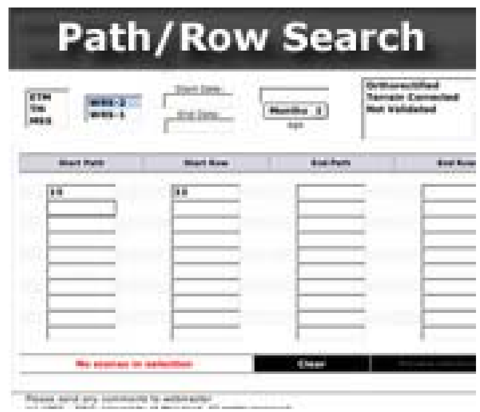
The GLCF Collection

	# of Scenes	Total MB Uncompressed
Landsat MSS	7,541	507,449
Landsat TM	8,355	3,402,795
Landsat ETM+	11,486	7,681,906
Landsat Data Subtotal		11,592,150
Landsat TM Mosaics	681	300,919
Landsat ETM+ Mosaics	762	109,332
MODIS 32-Day Composites	185	779,494
MODIS 16-Day CONUS NDVI	97	146,371
MODIS VCF	10	20,708
MODIS Data Subtotal	292	946,573
IKONOS	78	22,274
AVHRR Global Land Cover	15	7,450
AVHRR Continuous Fields Tree Cover	12	37,063
ASTER	~ 400	47,497
STRM	~14589	76,800
Total		13,140,057

Over 8TB/Month are downloaded by an international user community

Earth Science Data Interface

ESDI enables dynamic search, browsing, and downloading of products.



The whole library uses standard file formats and projection parameters and is available at no cost.

UMIACS
University of Maryland Institute for Advanced Computer Studies

System Configuration

- The main archive consists of:
  - Two Sun Distributed File System servers with over 26 TB of attached disk storage.
  - Twenty thin caching servers supporting http, ftp, and the SRB protocol.
- The catalog consists of:
  - Dual-processor Database servers with 4GB RAM and mirrored 15K rpm drives.
  - Dual-processor Application servers with 4GB of RAM and 10K rpm drives.

Software Support

- The Distributed File System supports replication, failover, and a uniform namespace.
- Database and Application servers organize and index the collection supporting:
  - Custom spatial data queries
  - Standard query interfaces through the ESRI Advanced Spatial Server or the freely available Mapserver package from the University of Minnesota.
- TSM backs up the entire collection.

GLCF Data Grid

- Links earth data centers at UMD, George Mason, and University of New Hampshire.
- Provide site independent namespace.
- Access data directly through high-level browsing and querying, not through redirection service.
- Basic spatial querying across different sites and data sets.
- Replication of data between sites.

Observations

- Traditional data transfer on removable media is complex:
  - A variety of media (eg. 8mm, DLT, SDLT, LTO, CD-ROM, DVD, USB/IEEE1394 Hard Drives)
  - A variety of formats (eg. Tar, ufsdump, dd, ntfs, ufs, ext2)
  - A variety of arrangements
- The cost of the various drives and trained operators is high.

Automated Ingestion

- Ingesting data is a key challenge for digital libraries and long-term preservation. We need efficient ingestion workflows that include:
  - Ways to agree on a transfer's details before it begins.
  - Means to assert that the transferred data is compliant as it is transferred, NOT after.
  - Tools to simplify the arrangement of data.
  - Mutual authentication and confidential communication without significant operational trust between the archive and a producer.

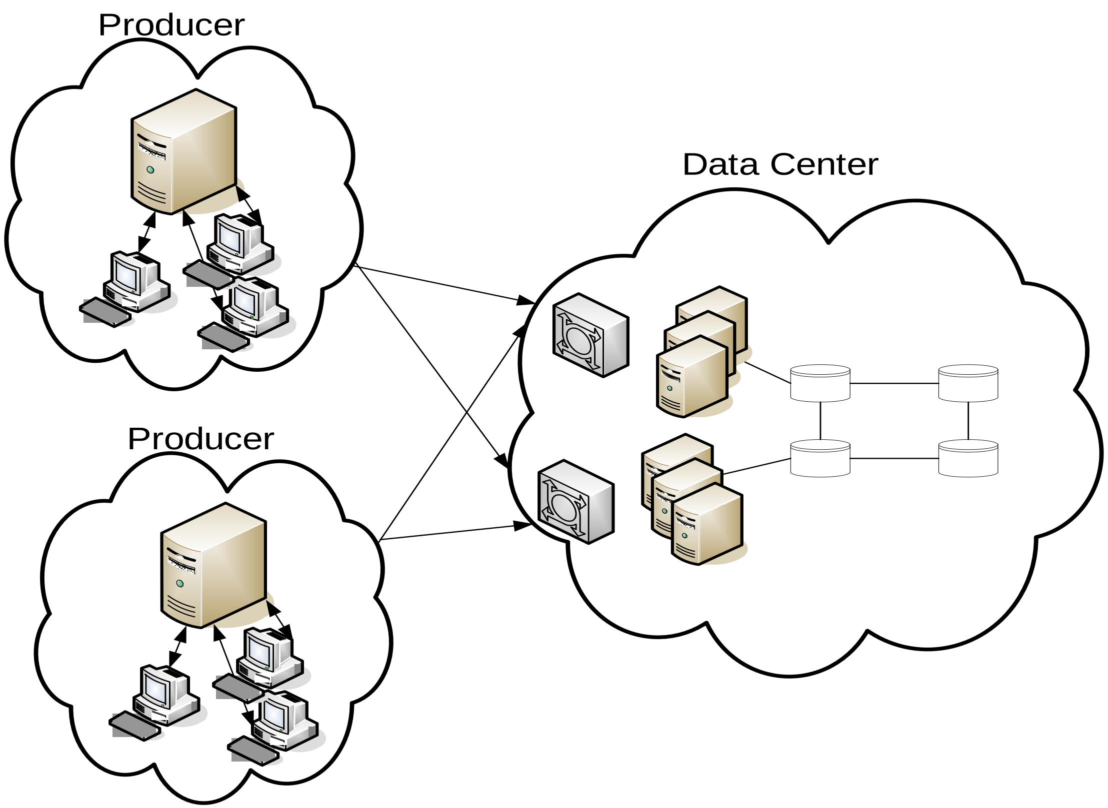
UMIACS
University of Maryland Institute for Advanced Computer Studies

A Prototype Ingestion System

- The UMD team implemented PAWN (The Producer -- Archive Workflow Network) to support archiving and long term preservation:
  - Scalable and platform independent; the entire system is built on web services and Java.
  - Marshal distributed resources at the producer site to encapsulate content with key metadata and transfer using the Metadata Encoding Standard (METS)
  - Simplify identity management, security, and trust between the producer and the archive using open standards (PKI, X.509, and GSI)

Main Components

- PAWN has three major components:
  - A management server, a thin client, a receiving server at the archive



UMIACS
University of Maryland Institute for Advanced Computer Studies

The PAWN Management Server

- The management server at the producer:
  - acts as a central point for the initial organization of the data
  - Identifies and tracks each bit-stream and their associated metadata as they are registered and transferred to the archive.
  - Records checksums/digital signatures, system metadata and other client supplied descriptive metadata
  - Provides the necessary security infrastructure to allow secure transfer of bit-streams between the producer and the archive.

UMIACS
University of Maryland Institute for Advanced Computer Studies

The PAWN Client

- Multiple thin clients run simultaneously at the producer site. They enable:
  - Automated bulk registration of bitstreams, checksums and their system and preservation metadata.
  - Assembly and transfer of valid Submission Information Packets (SIPs) to the archive
  - Automated harvesting of descriptive metadata as needed.

The PAWN Receiving Server

- Accepts SIPs from clients at the producer
- Verifies with the management server that all SIPs arrived intact
- Processes SIPs and initiates verification and validation processes on the received bit-streams.
- Coordinates user authentication with the management server at the producer site
- Negotiates authentication and authorization needed for the producer to securely transfer documents to the archive.

UMIACS
University of Maryland Institute for Advanced Computer Studies

Conclusion

- Aggregating and sharing storage resources across different domains is becoming a critical technology.
- Platform independent software infrastructure to support such activities is still lacking.
- Archiving and long term preservation are becoming major issues in a wide variety of organizations.