The Bumpy Road to Multi-Petabyte Archives

Byron V. Peters
EOSDIS Core System
Raytheon Company, 1616 McCormick Dr
Landover MD 20774-5301
email: bpeters@eos.east.hitc.com

Presented at the THIC Meeting at the Raytheon ITS Auditorium, 1616 McCormick Dr
Upper Marlboro MD 20774-5301
November 5-6, 2002
The Bumpy Road to Multi-Petabyte Archives

Byron V. Peters
EOSDIS Core System
Raytheon Company
Landover, Maryland
email: bpeters@eos.east.hitc.com
• The Earth Observing System (EOS)
  • Part of the U. S. Global Change Research Program (USGCRP)
  • 29 Instruments on 12 Earth Resource Satellites through 2015
    - Landsat 7
    - Terra (EOS AM-1)
    - Aqua (EOS PM-1)
• EOSDIS Core System (ECS)
  • Flight Operations
  • Science Processing
    – Ingest, process, distribute products via Distributed Active Archive Centers (DAACs)
The DAACs...

ECS-supplied DAACs: GSFC, LaRC, EDC and NSIDC
• “...Effectively adapt to change on an ongoing basis, and still achieve the stability required of a system that must support operational product generation and manage large volumes of EOS instrument data over the project lifetime”

• “...Gracefully grow and evolve, is flexible, and has reduced risk, complexity, and cost.”

• Storage will be “designed to allow incorporation of new standards in operating systems, data formats, software, and metadata formats in a way that minimizes adverse impact to the current users of the system”

• COTS to be used wherever possible

From the ECS Statement of Work
Unprecedented problem size for NASA

- Support many different data product types: ~1,300 different product types
- Archive large amounts of data: adding ~ 3 TB per day today, growing to 6 TB per day in 2003
- Data holdings are ~1.6 PB today, growing to ~ 3 PB by 12/03, ~20 PB by 2015
- Manage large geo-spatial databases: adding ~ 500 MB of metadata per day
- Distribute large amounts of data in several formats: ~ 3 TB per day

Complex system context

- Execute complex science algorithms provided by science community
- Interface with ~ 35 external systems
Initial Architecture

- ECS Custom Software
  - Object oriented, client-server
- Platforms
  - Use Sun systems for general purpose
  - Use HP systems for network and systems management
  - Use SGI systems for high-bandwidth I/O and applications
- Network-centric
  - Direct-attached storage
  - Heavy use of NFS and FTP for internal operation
  - HiPPI backend network for SGIs
- Archive Hardware
  - ADIC AML/2 hardware for browse (thumbnail) files
  - STK 9310 Powderhorn for main archive function
- Archive Software
  - ADIC AMASS
    - NOT hierarchical management - strictly an archive
Initial Generic Architecture Map

Staging Area

- SGI Science Processing and Storage
- SGI Ingest and Storage
- SGI Archive Management and Storage

Network

- Archive Control Connection (Network)
- Archive Data Connection (SCSI)

- Library Front End
- Tape Library

- Sun General Purpose (20)
- HP Network/System Management (5)

Object-oriented Client/Server software model
Data Storage Characteristics

- Most data are archived using the HDF data format
- Storage of the product data associated with a data type can be configured on a data type basis
- Support for online and near-line storage of product data
- Portions of a data collection may be allocated across multiple storage units (e.g., silos) to improve performance -- allocations may be changed over time to support new access patterns
- Data collections may be replicated to protect against data loss
Initial “Opportunities”

• Pre-Landsat 7 launch concerns:
  • Performance, Performance, Performance
  • Is ECS pushing the envelope too much???
    - 1-3 MB/s throughput to the SD-3s and needed at least 5 MB/s
      – Fixed through a new AMASS option - streaming I/O and using IRIX Extended Logical Volume manager (XLV) across eight RAID-3 devices
      – Results - 6-12 MB/s per drive on 8 drives simultaneously

• Post Terra launch, pre Aqua launch undertook initial drive swap
  • SD-3 Redwood pulled from the market and the design
  • STK 9940As replaced the Redwoods
Changes Along the Way...

- Movement to the “simpler is better” approach
  - Migration to Distributed File System (DFS) abandoned
  - Distributed Computing Environment (DCE) abandoned
    - Migration to native threads
  - HP’s abandoned; Replaced by Sun’s
- Moved to only use STK Powderhorn
  - One for one replacement of AML/2s
Where we are...

• ECS succeeded!
  • Performance routinely 6 times greater than baseline requirement
  • Synergy program funded real world uses of ECS data
    - Goal to provide a one-stop shopping “infomart” for remote sensing data

• Storage Area Networks added
  • Storage is more persistent - a data pool (direct user access to live data)
    - Better user response times through fewer retrievals from the archive
  • Adds a three layer storage management to ECS
    - Continued direct-attached storage where required
    - Expand SAN-based common areas
    - Continued use of the tape archive

• Browse moved to disk
• Using IBM/Tivoli SANergy to integrate both Solaris and IRIX systems on STK/LSI hardware
New Opportunities

• Technology Obsolescence
  • Migrating to new technology has become a major challenge

• Implementing large SANs in a Heterogeneous Environment
  • Sun Solaris, SGI IRIX, Linux platforms
  • RAID and tape target devices
  • Heterogeneity again pushing the envelope...

• Hardware Consolidation
  • Reducing the number of platforms to reduce maintenance costs and improve reliability

• System Integrity Monitor
  • Need an answer to “is my system working from end-to-end???”

• The Big Unanswered Question -- Implementing Data Management
Lessons to Ponder

• If you are going to build anything complicated, attempt to build a homogenous system
  • Single Operating System and hardware platform

• If a homogenous system not an option, minimize the choices especially in Operating Systems.

• Selecting a vendor that is not the industry leader will impact implementing the latest and greatest versions of COTS software

• Pick vendors that have a history of partnering
  • Doing anything non-standard often requires active vendor cooperation
  • SGI, STK and ADIC have been exemplary partners for ECS