



# Data Archiving Using Enhanced MAID

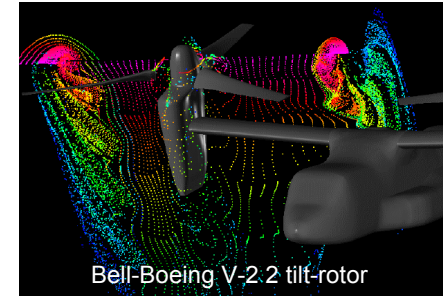
Aloke Guha  
CTO, COPAN Systems

23<sup>nd</sup> IEEE13th NASA Mass Storage Conference  
May 18, 2006

# Archive Data Needs

- **Retention and Disposition**
  - ◆ Corporate governance and regulatory guidelines
  - ◆ Highly variable
- **Performance**
  - ◆ Retrieval latency
  - ◆ Rate at which data can ingested into archive
- **Content Retrieval**
  - ◆ Ease of direct access
  - ◆ Retrieving data by various methods
- **Immutability: reference data**
- **Data Integrity**
- **Security**
  - ◆ Access Control
  - ◆ Audit Trails
- **Compliance: SEC-17a-4, SOX, DoD 5015.2-STD**
- **Cost Constraints**
  - ◆ Acquisition
  - ◆ Life-cycle

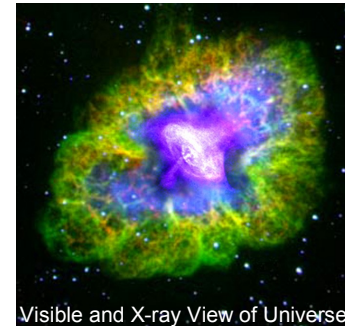
*CAD/CAM Design*



*Medical Imaging*



*Astronomy*

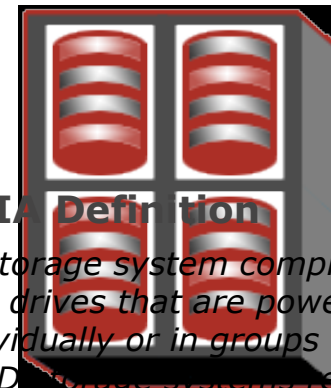
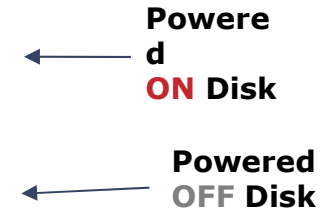


*Climate Modeling*



# MAID: Power Managed Disks

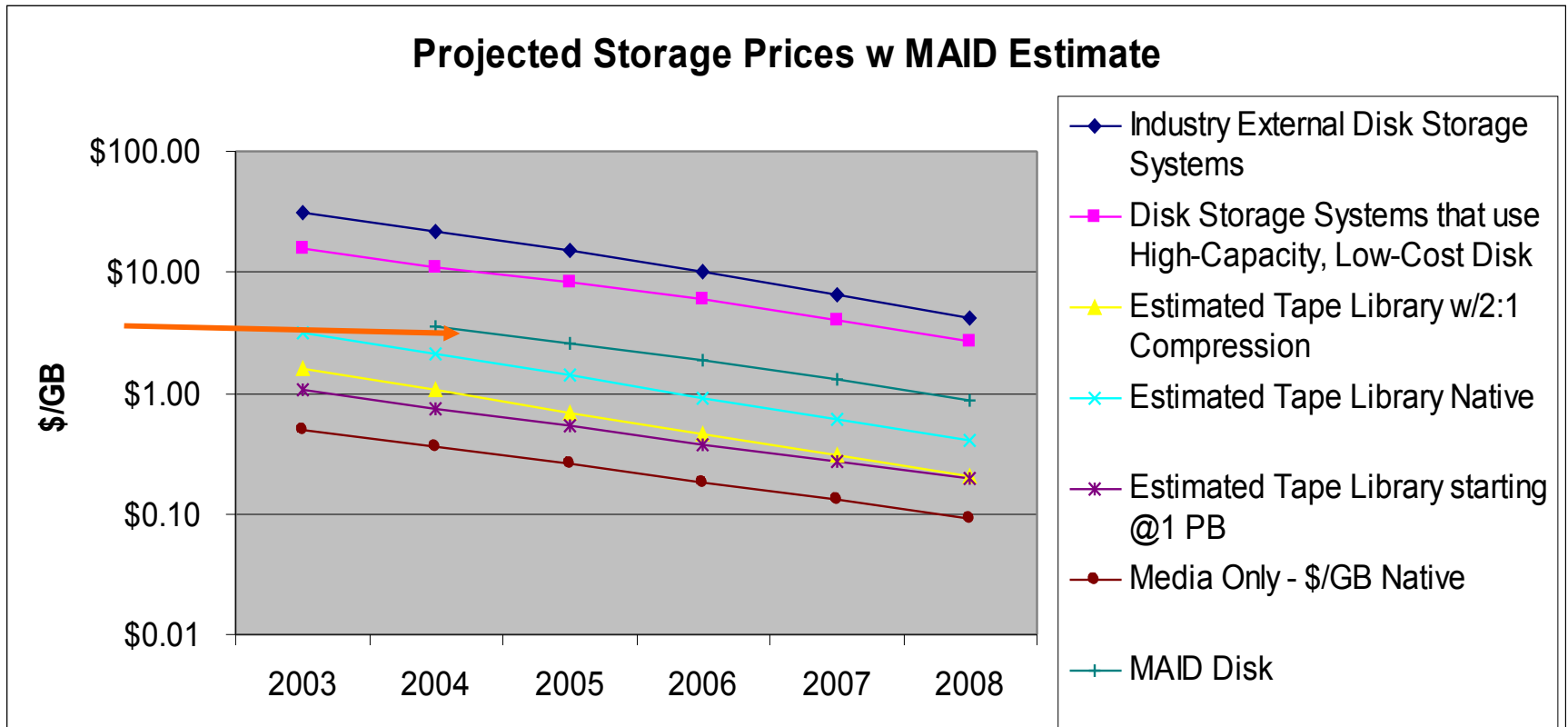
- MAID
  - ♦ Large number of power-managed disks
  - ♦ More than 50% drives powered off
  - ♦ Power-cycling by policy
  - ♦ Lower management and environmental costs and longer drive life
  
- *COPAN Systems' Enhanced MAID*
- Massively scalable storage solutions
  - Reliability and performance like disk
  - Scale and cost of tape
  - ♦ Three-Tier Architecture
    - Scales performance with capacity
  - ♦ **POWER MANAGED RAID® Software**
    - RAID protection for power-managed disks
    - Maximum of 25% drives spinning
    - 1/3 cost of traditional RAID systems
  - ♦ **DISK AEROBICS® Software**
    - Disk reliability and data integrity



## SNI Definition

"A storage system comprising an array of disk drives that are powered down individually or in groups when not required. MAID storage systems reduce the power consumed by a storage array."

# Cost Trends: Tape, MAID, Disk



**Disk Storage Systems Estimates Source: IDC, Worldwide Disk Storage Systems 2004-2008**

**Forecast and Analysis: Conservatism Persists, But Opportunities Abound, Aug. 2004**

**Tape Storage Price Estimates: Courtesy IBM**

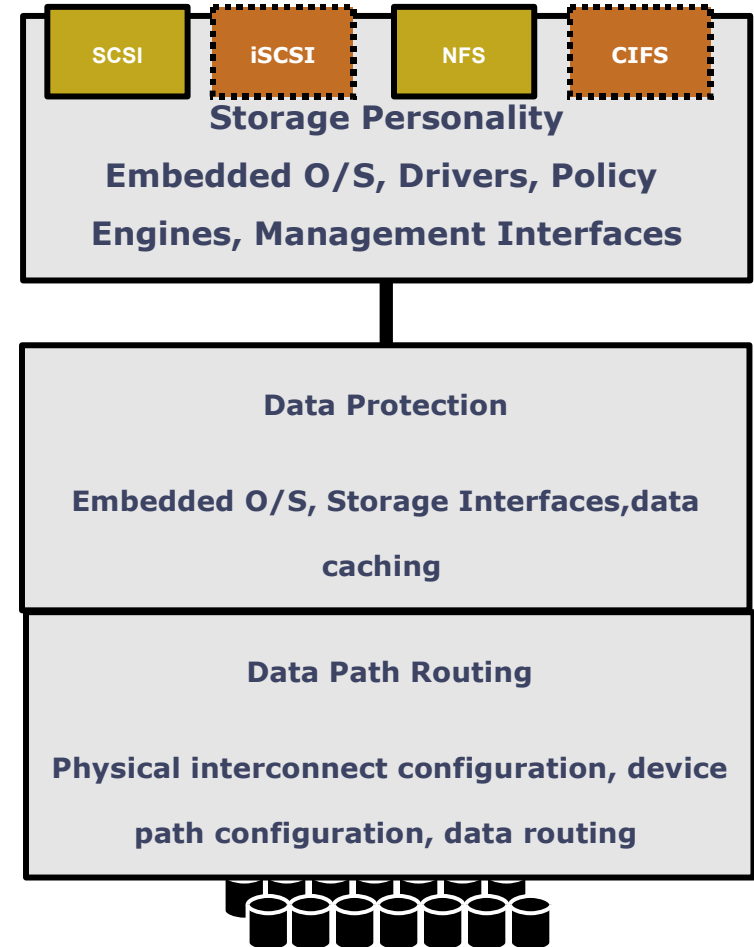
**MAID Estimate – INSIC Roadmap Workshop**

Source: INSCI Storage Roadmap, 2005

# 3-Tier Architecture

*Three levels of processing separate functionality, simplify management and allow performance to scale with capacity.*

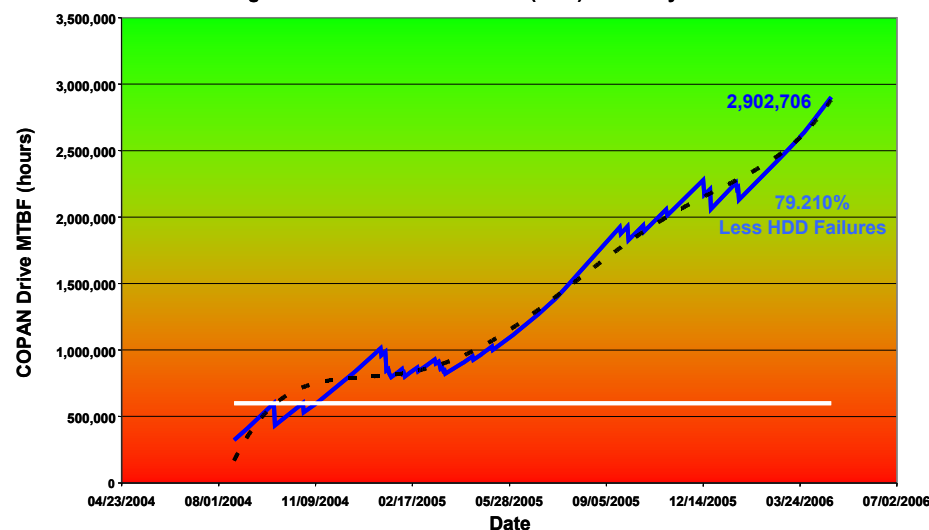
- **Layer 2 – Storage Personality**
  - ◆ Physical Domain (Rack Controller)
  - ◆ Storage Network Protocols
  - ◆ Logical Volume/Block Management
  - ◆ Performance and Load Balancing
  
- **Layer 1 – Data Protection**
  - ◆ Physical Domain (Shelf Controller)
  - ◆ RAID Support and Caching
  - ◆ Power Management
  - ◆ Device Management
  
- **Layer 0 – Data Path Routing**
  - ◆ Physical domain (Canister Controller)
  - ◆ Protocol Router
  - ◆ Management Microprocessor for environmentals and monitoring



# POWER MANAGED RAID® and DISK AEROBICS® Software



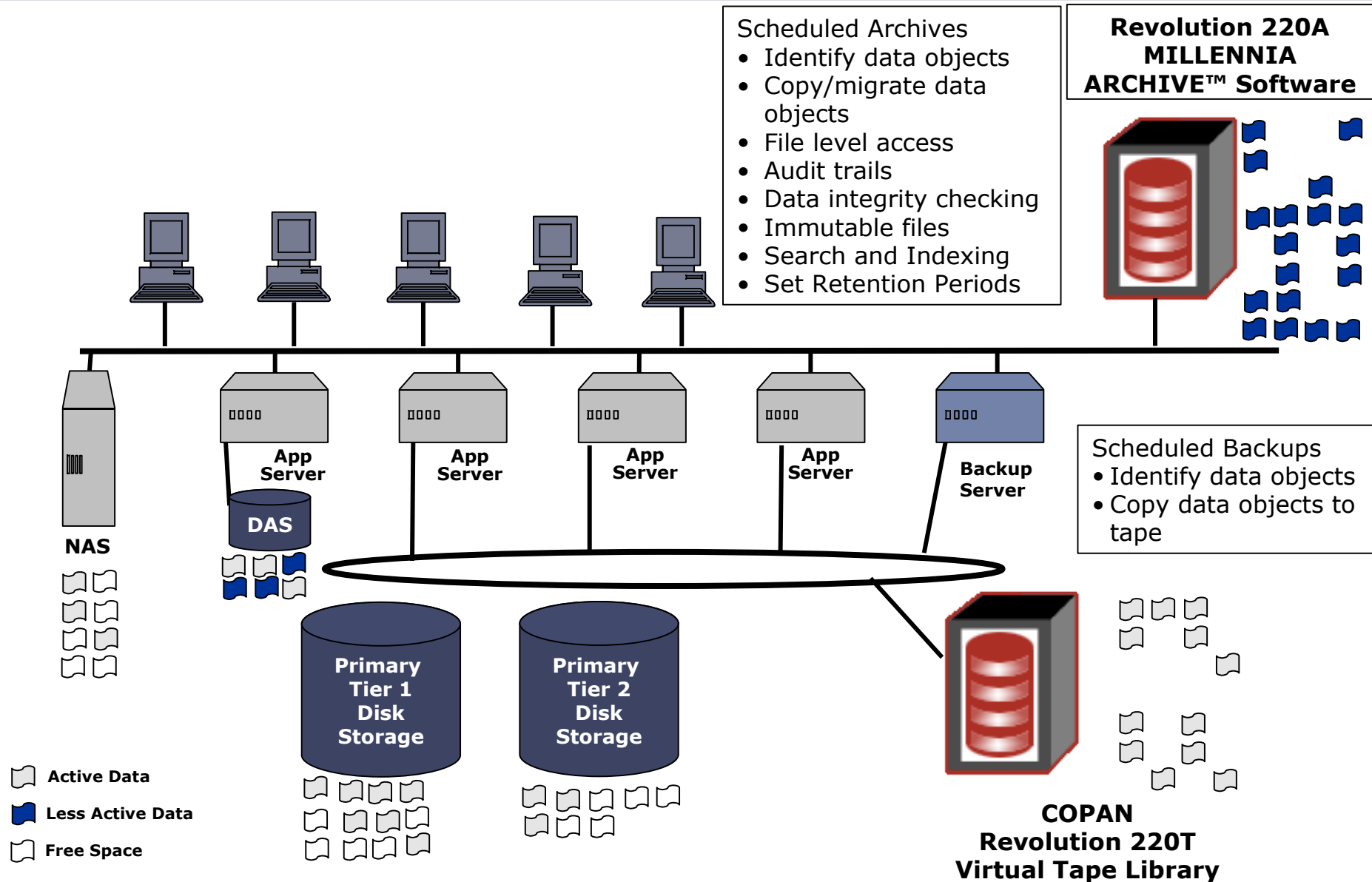
MAID Advantage in Terms of Hard Disk Drive (HDD) Reliability: Field HDD MTBF Growth



- Increase drive service life by > 4x
- Less than 79% failures than std SATA
- Explicitly manage drive health
- Proactively manage disk reliability
- Actively maintain data integrity
- Data reliability: 23X SATA, 6X FC

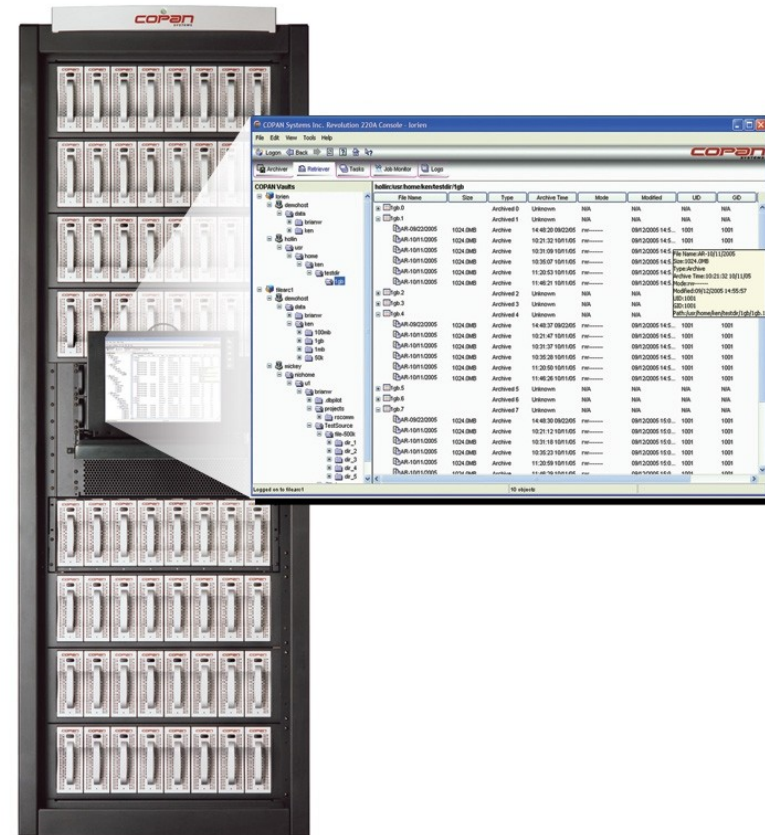
Feature	Benefit	Mechanism
POWER MANAGED RAID (Power Management)	Extends Drive Life	Drives spin only when necessary to meet application requirements
Periodic Exercise of Idle Drives	Assures Drive Health	Every idle drive powered-on and tested at least once every 30 days
Proactive monitoring and management of drives	Predictive Drive Maintenance	Monitors SMART parameters and environmental data
Disk Scrubbing	Assures Data Integrity	Background task identifies bad sectors on disk and copies data to new sector on drive
Data Migration	Data integrity and long-term data retention. Avoids long RAID rebuilds	Proactive failing of suspect drives – copies data to spare drive and “fails-out” suspect drive. Inserts new drive into RAID set.

# Uses of MAID Platform . . .



# Revolution 220A Unique Features

- **File Access Personality**
  - ◆ 1 billion files - 500 GB drives
  - ◆ 625 GB file size - 250 GB drives
  - ◆ 1.25 TB file size - 500 GB drives
  
- **MILLENNIA ARCHIVE™ software**
  - ◆ Automates data movement
  - ◆ Schedules archiving by date, time or event
  - ◆ Provides data immutability
  - ◆ Indexing and searching of files
  - ◆ User specified retention periods
  - ◆ Identifiable audit trails
  
- **Connectivity, Performance**
  - ◆ 4 1-GigE connections
  - ◆ File access latency: ms - 13 s

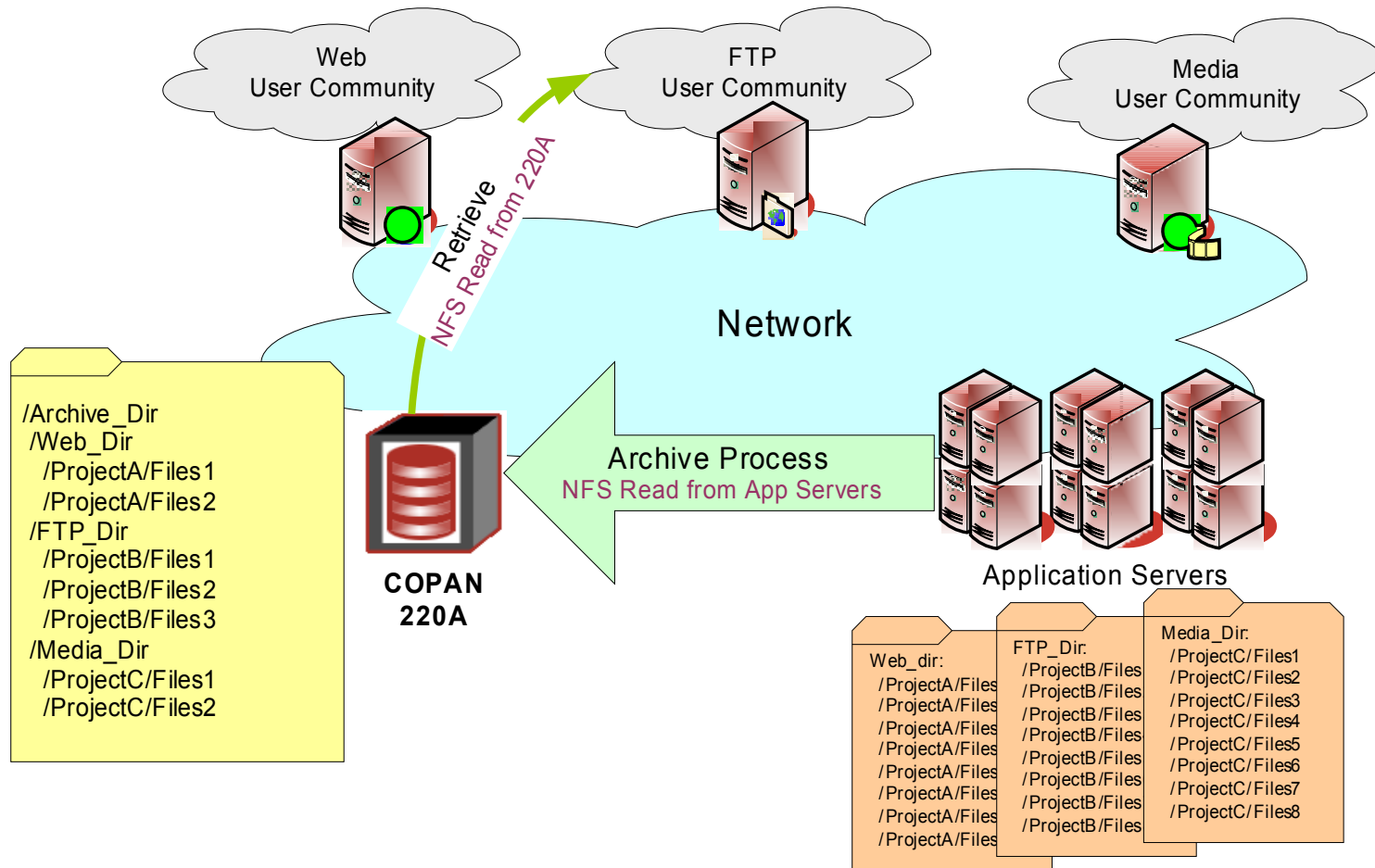


# Revolution 220A Archive Process

Archive\_Dir/Web\_Dir/...\*.\*

Archive\_Dir/FTP\_Dir/...\*.\*

Archive\_Dir/Media\_Dir/...\*.\*



The screenshot shows the COPAN Systems Inc. MILLENNIA ARCHIVE Console interface. The window title is "COPAN Systems Inc. MILLENNIA ARCHIVE Console - gerbil". The interface includes a menu bar (File, Edit, View, Tools, Help), a toolbar with Logon, Back, and other navigation icons, and a main toolbar with buttons for Archiver, Retriever, Tasks, Job Monitor, and Logs. The main area is divided into two panes. The left pane, titled "Mounted File Systems", shows a tree view of the file system structure. A blue arrow points to the "data on demohost" folder. The right pane, titled "demohost:/data/brianw/TestSource/file100k/dir\_10/subdir\_1", displays a table of files with columns for File Name, Size, Type, and Modifi. The table lists files from file\_100 to file\_109, each with a size of 50.0KB and a modification time of 09:23:2. The status bar at the bottom indicates "Logged on to gerbil" and "500 objects".

Mounted Directories and Files to be archived

File Name	Size	Type	Modifi
file_100	50.0KB	File	09:23:2
file_1000	50.0KB	File	09:23:2
file_101	50.0KB	File	09:23:2
file_102	50.0KB	File	09:23:2
file_103	50.0KB	File	09:23:2
file_104	50.0KB	File	09:23:2
file_105	50.0KB	File	09:23:2
file_106	50.0KB	File	09:23:2
file_107	50.0KB	File	09:23:2
file_108	50.0KB	File	09:23:2
file_109	50.0KB	File	09:23:2

# Retrieve

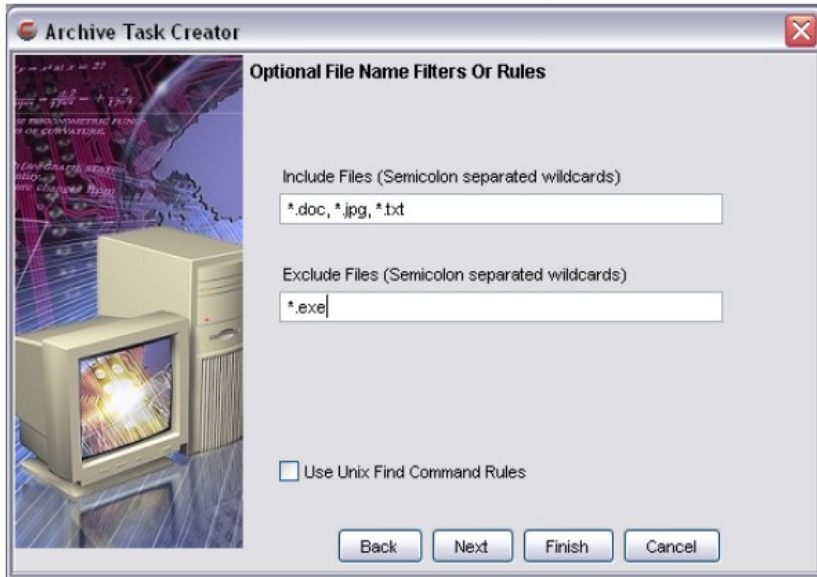
Directories and Files on COPAN available for retrieval

The screenshot shows the COPAN Systems Inc. MILLENNIA ARCHIVE Console interface. The window title is "COPAN Systems Inc. MILLENNIA ARCHIVE Console - gerbil". The interface includes a menu bar (File, Edit, View, Tools, Help), a toolbar with Logon, Back, and other navigation icons, and a main toolbar with buttons for Archiver, Retriever, Tasks, Job Monitor, and Logs. The "Retriever" button is highlighted with a pink box. The main area is divided into two panes. The left pane, titled "COPAN Vaults", shows a tree view of the file system structure. The right pane, titled "azog:/usr/share/aclocal/iimf", displays a table of files available for retrieval.

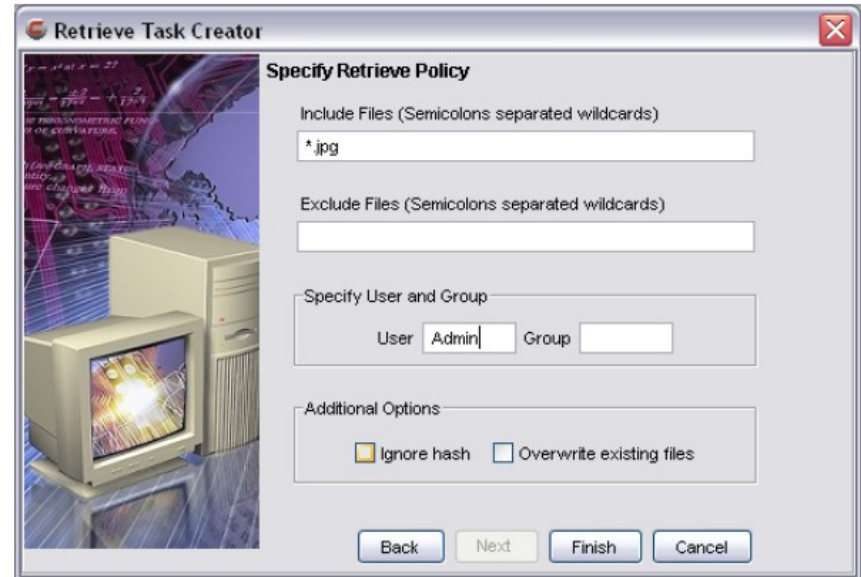
File Name	Size	Type	Archive T
check_canna.m4		Archived m4	N/A
AR-01/06/2006	417	Archive	15:28:26 01
check_freewnn.m4		Archived m4	N/A
check_gtk.m4		Archived m4	N/A
check_socket.m4		Archived m4	N/A
check_sun.m4		Archived m4	N/A
check_sys.m4		Archived m4	N/A
check_x.m4		Archived m4	N/A

Logged on to gerbil | 7 objects

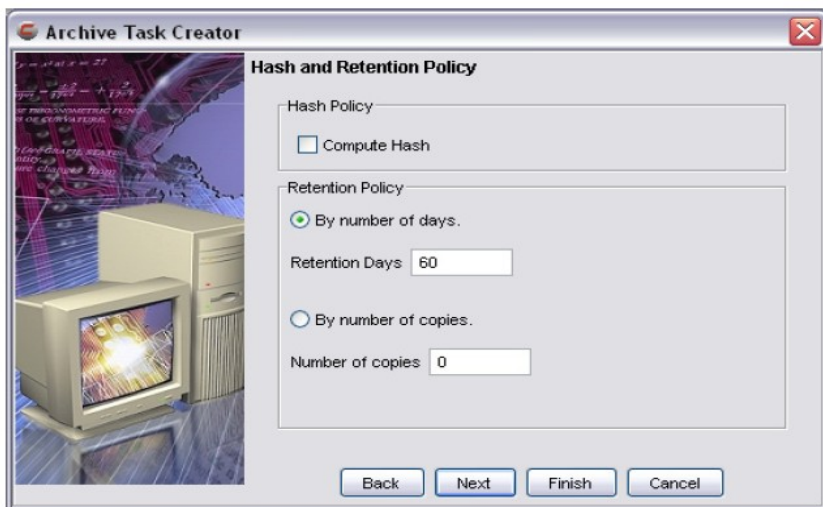
# E.g.: Archive and Retrieval Policies



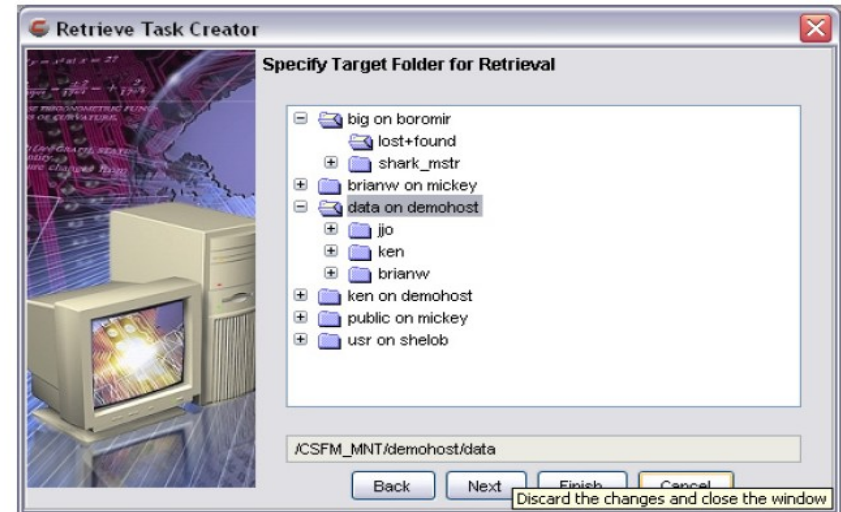
Archive setup with filtering options



Retrieve via GUI with policy options



Retention setup, versioning options



Retrieve via by path and file name

## ▪ Applications

- ♦ **Repurpose Digital Assets: media/content, engg. design**
- ♦ **Accessing Reference Information: medical**
- ♦ **Support Corporate Governance and Regulatory Compliance**
- ♦ **Document Management**
- ♦ **Data Collection and Analysis: scientific**
- ♦ **Eliminate the Need for Long-term Tape Retention**

## ▪ Actual Use Cases

- ♦ **Financial: historical stock trading ("tick") data, on disk vs tape, direct file access ⇒ from 2-4 days to < minute**
- ♦ **Media/Entertainment: archiving digital film material for future repurposing, ease of new content generation**
- ♦ **E-commerce: ticketing company keeps all transaction records online, major savings in storage cost, improved performance**

- **MAID storage ideal for long-term data and archiving**
  - ◆ Scale
  - ◆ Density
  - ◆ Reliability
  - ◆ Cost
  
- **File Archiver R220A designed for solving archiving needs**
  - ◆ Millennia Archive: archiving from existing file storage
  - ◆ Support for Retention and Immutability
  - ◆ Data integrity checking
  - ◆ Retrieve via NFS
  - ◆ Search by file and user-defined attributes
  - ◆ Security: secure views, access control, audit trails



**COPAN Systems provides intelligent storage solutions that unlock the value of long-term data.**

**Thank You**