



GRAU

DATA STORAGE

Overcoming Obstacles to Petabyte Archives

Mike Holland

Grau Data Storage, Inc.

609 S. Taylor Ave., Unit 'E', Louisville CO 80027-3091

Phone: +1-303-664-0060 FAX: +1-303-664-1680

E-mail: Mike@GrauData.com

**Presented at the THIC Meeting at the National Center for
Atmospheric Research**

Boulder CO 80305-5602

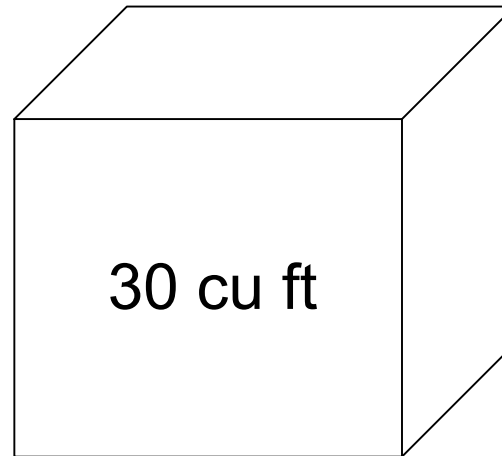
June 11-12, 2002

THIC Inc.

The Premier Advanced Recording Technology Forum



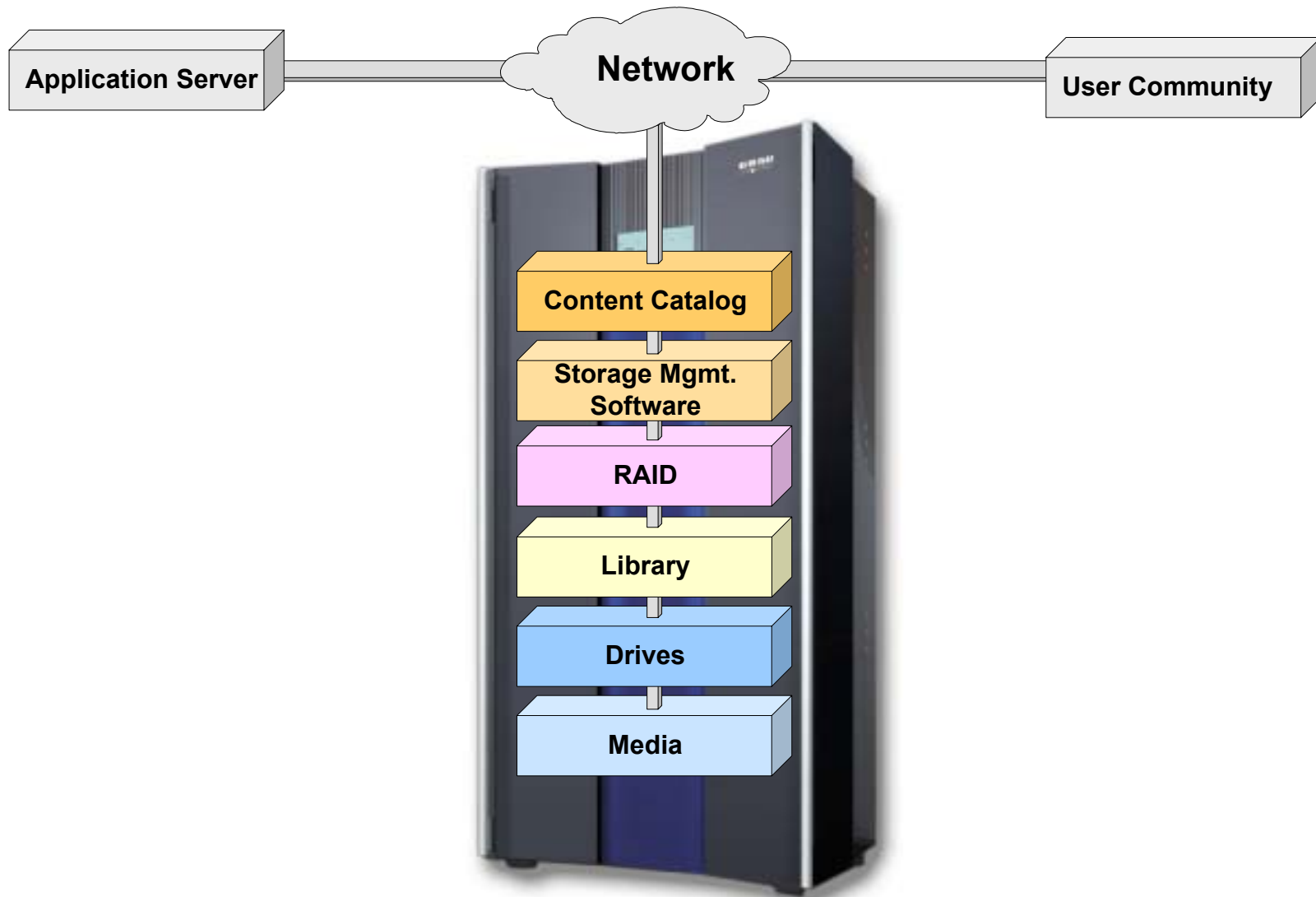
A Petabyte of Archive is Easy?



Our system utilizes SONY AIT technology, so 1 PB is only 10,000 tapes...or about enough tapes to fill a box 1M on a side

In the future, we can comment about the size of the cube...

Components of an Archive





The Challenges of a PB Archive have less to do with sheer Storage

- Support for Users
 - Common Access to all of the Data
 - Recall Characteristics
 - Data Acquisition Rates
 - Data Expiration Rates and Reorganization of Media
- Catalog Size (File System Entries)
- System Conversions
 - Absorption of Older Archives
 - Transcription to Newer Technologies
- Data Protection Schemes (Media Availability)
- System Availability (Fault Resilience)
 - Recovery



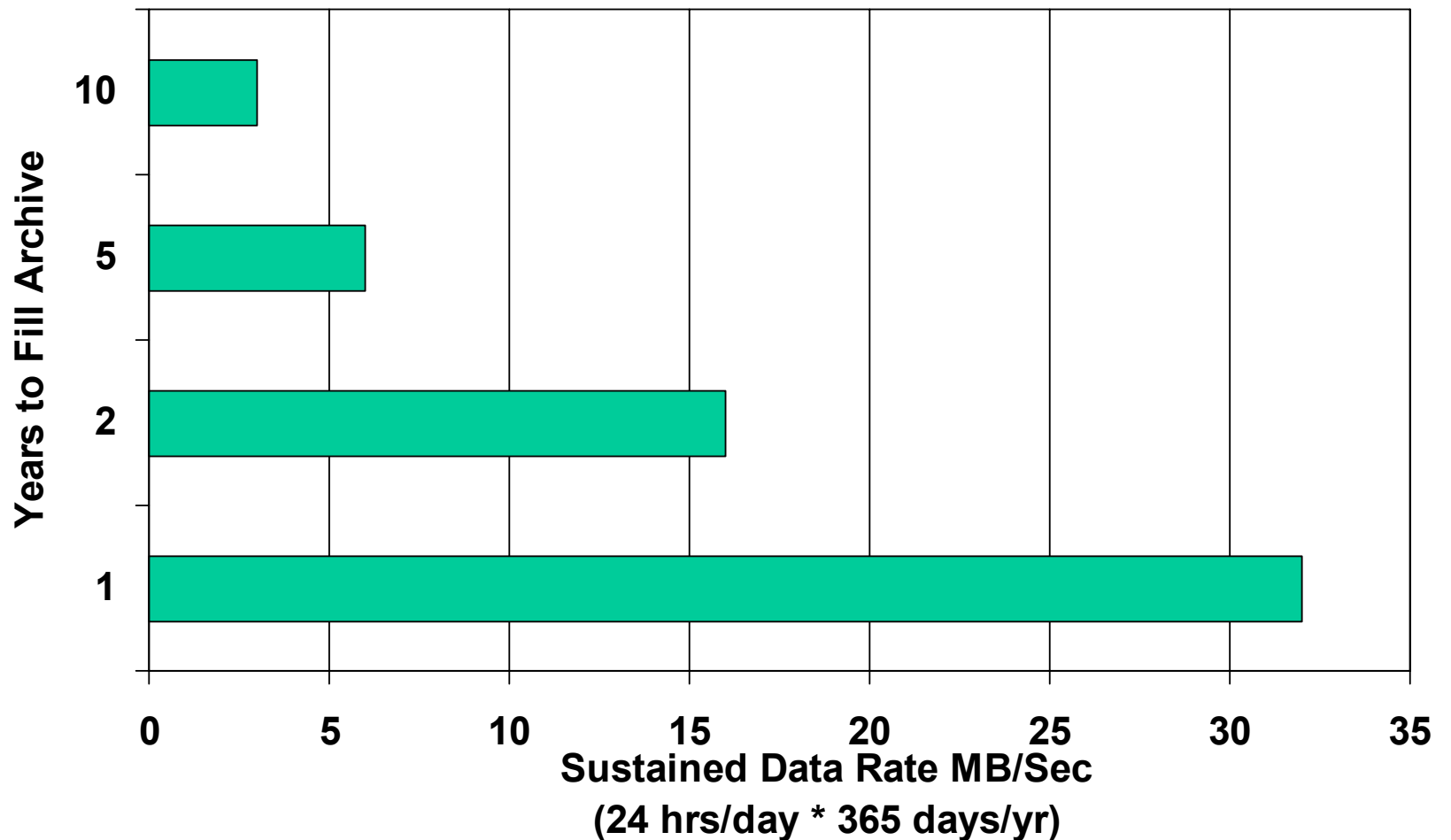
I/O Activity within Archive

- Number of Recall Users
 - Concurrency
 - Peak workloads
- Data Acquisition Rates
- Data Request Queue Time Expectations
- Data Access Patterns
 - Locality of reference for pre-stage data caching
 - Scheduled recalls staged for faster use

Actually, less concerned about I/O
than management within archive

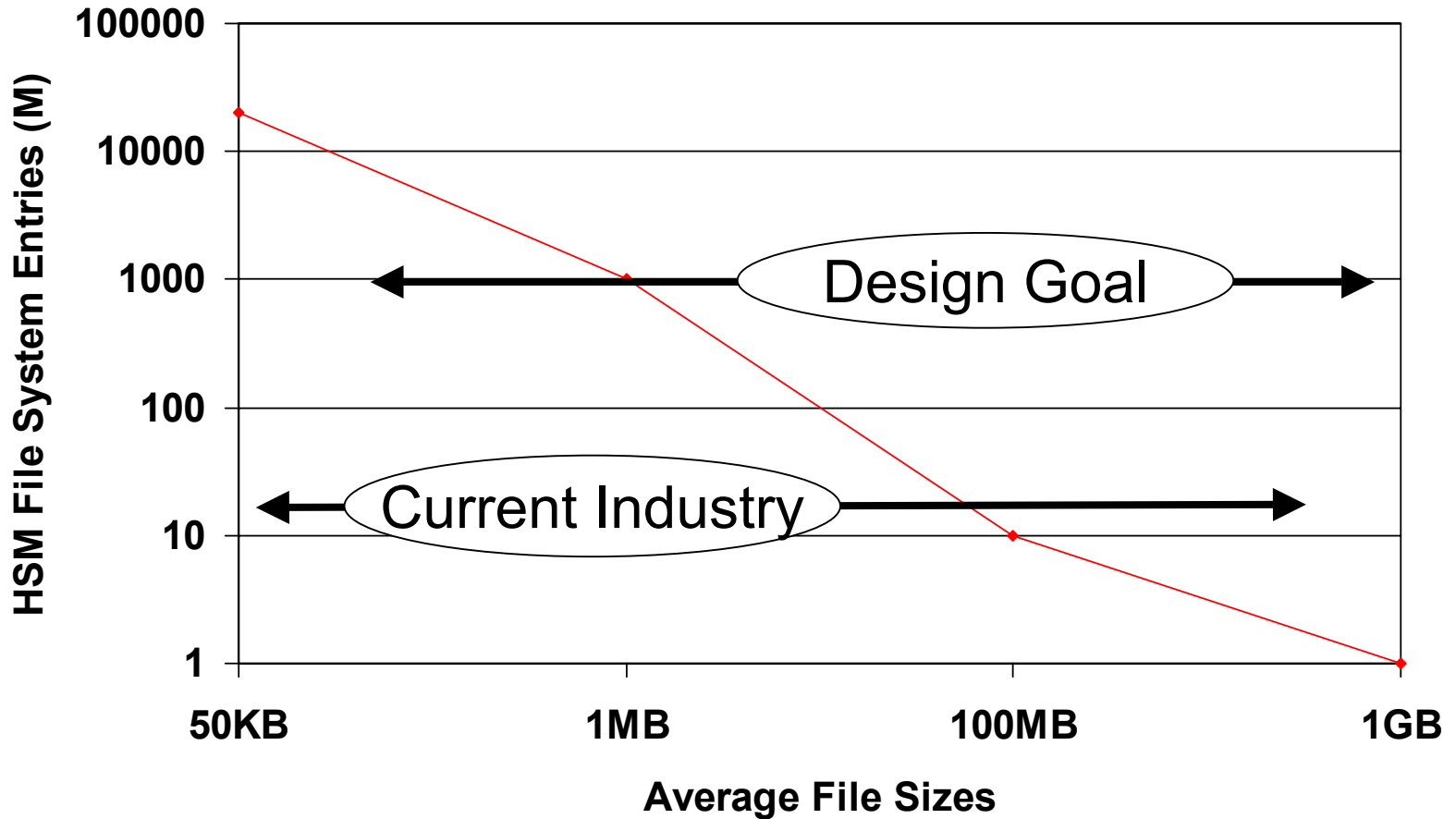


What Does a PetaByte Imply?

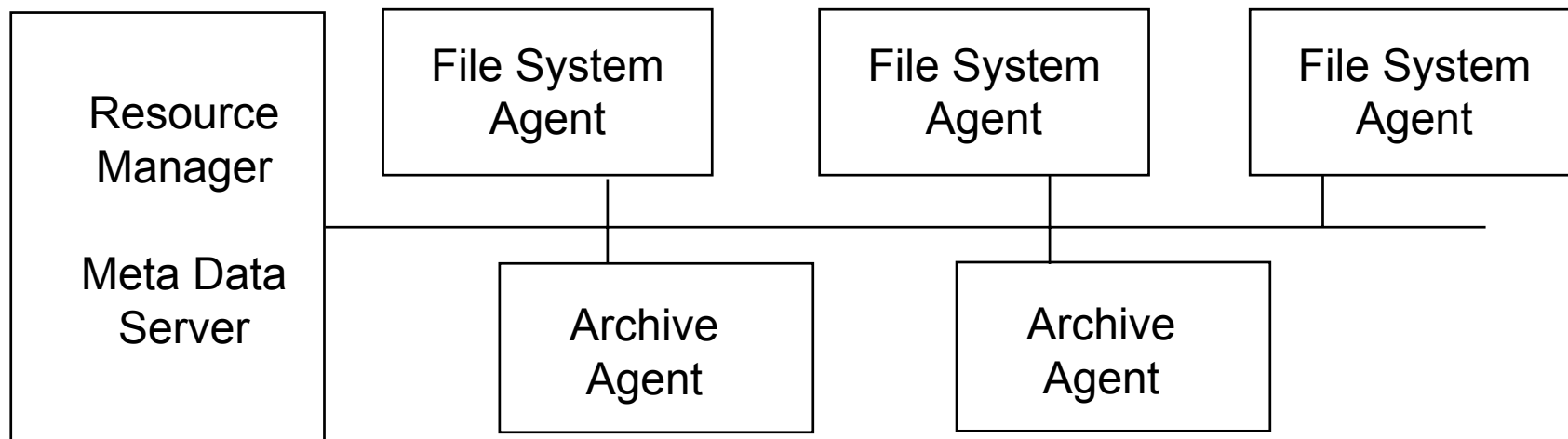




What Does a PetaByte Imply?



PB Archive Must Be Scalable Throughout

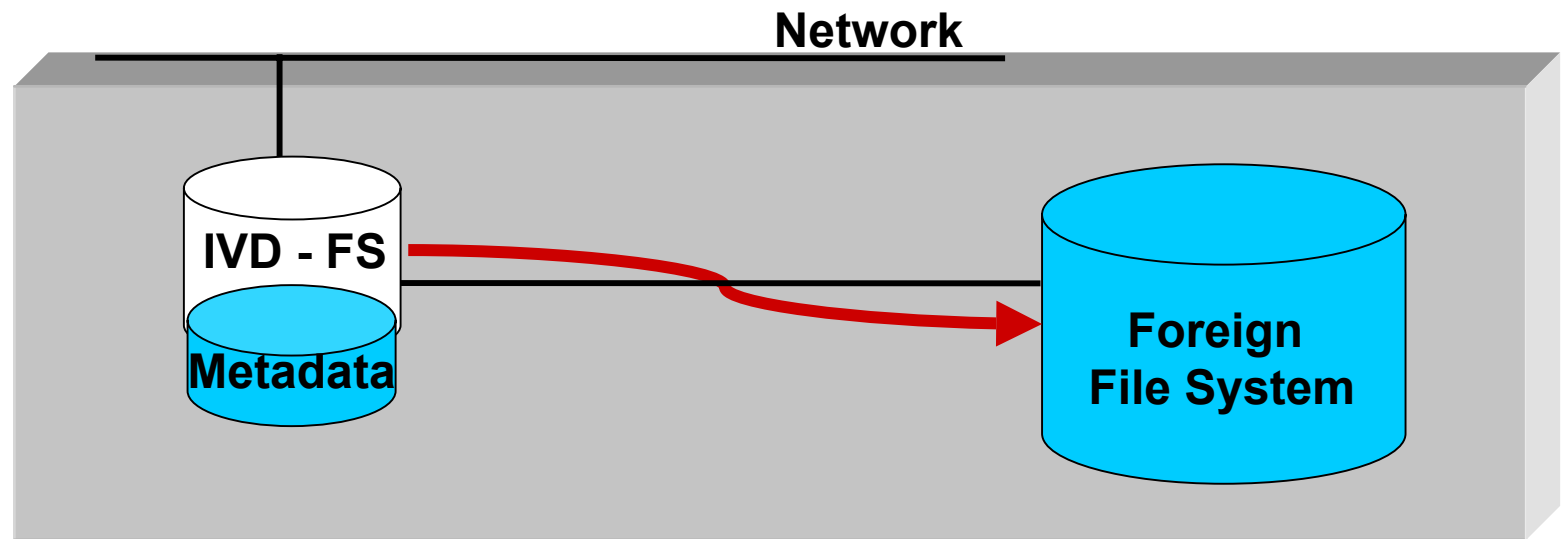


- # of files drives absolute requirement for multiple filesystems
- All agents are connected to the RM and MDS
- Front End Agents (File System Agent) are connected to Back End Agents (Archive Agent) through IP connections.
- Requirement for a stable archiving OS

So, It's Agreed!

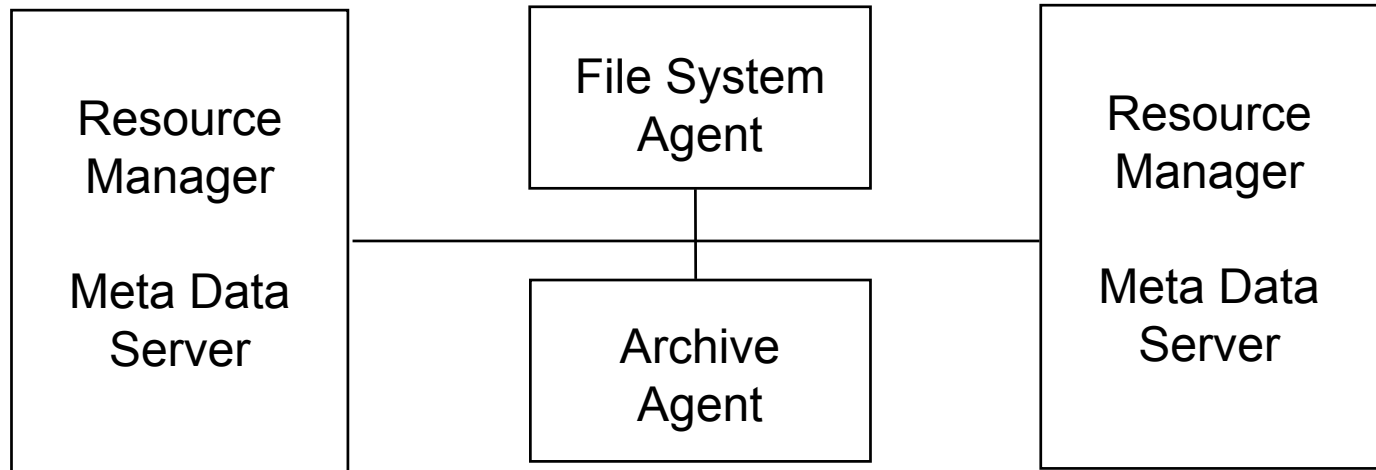
We Never Delete Any Files!

- Present a common view of data
- Add older systems to a current system
- Continually migrate data to newer storage media



- Capture 'old' catalog into new management system
- Migrate old data in background

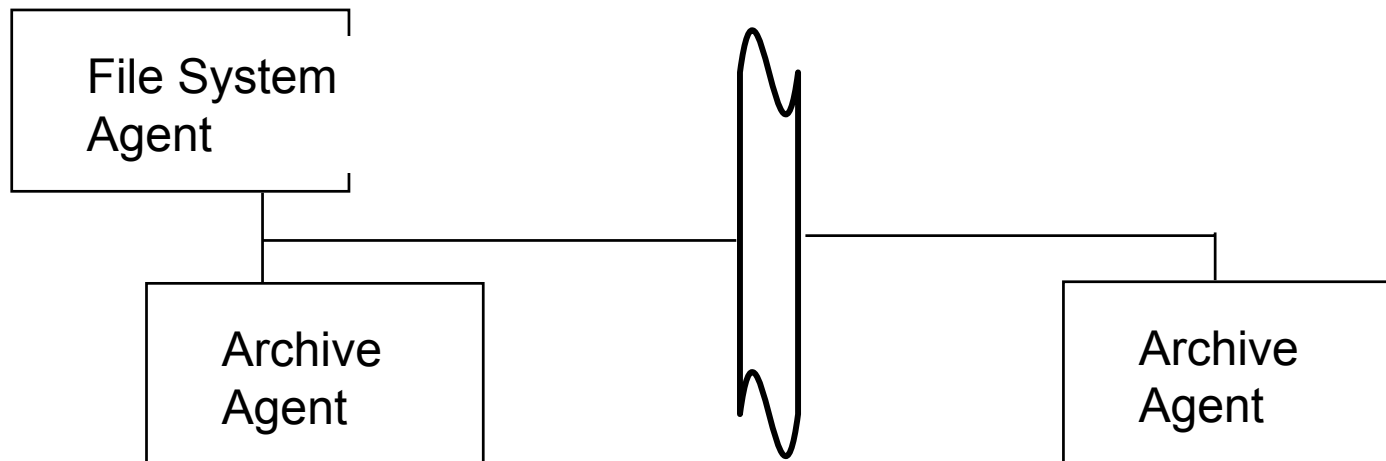
High Availability



- Multiple instances of Resource Manager and Meta Data Server may exist within in one system
- Two or more copies of control information are journaled onto separate disk subsystems

Data Protection Schemes

- To guarantee data security, IVD can store several copies of the data to different media pools, local or remote
- Pools may be located far away from each other but connected through TCP/IP





IVD - Policies

- Each FE Agent can manage several partitions
- A partition contains a file system and one or more tape pools
- Each partition has a wide set of migration and truncation policies:
 - Partition, directory or specific file level
 - File access time and file size based
 - Wildcard or file extension based
 - Exclude lists
 - Priority based job execution (migration, recall, reorg)
 - Extended via external filter script for user defined policies
 - Associative file grouping for migrate and recall as a group

▼ Policy Templates

Preconfigured policy sets for the most common use

- Premigration: Data are copied to the tape soon after they are stored to secondary media but stay on disk. If disk space is needed, older data are removed.
- Immediate Migration: Data are copied to secondary media soon after they are stored to IVD and removed from disk immediately.
- Triggered Migration: Certain filesystem events control data migration to secondary media.
- User is free to define policies of any style, using individual scripts

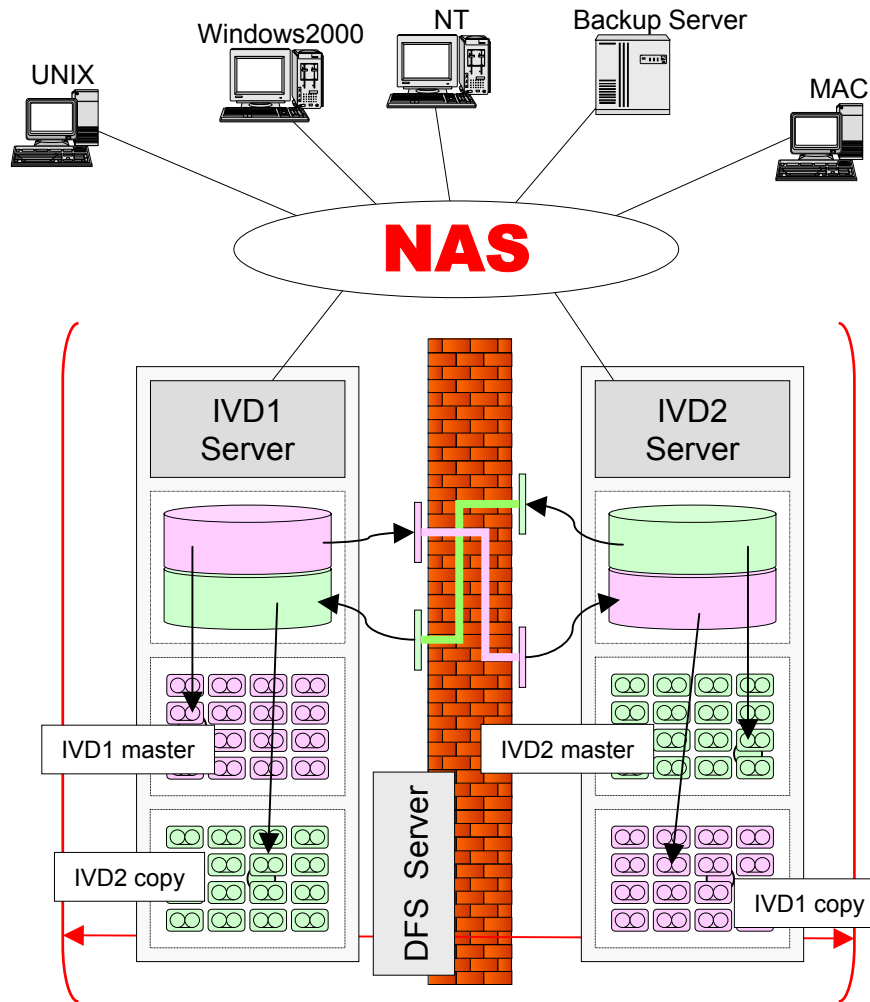


The Commercial Challenge

How do we fill up a 35TB archive, much less a Petabyte?

- Changes in the industry regarding management of data
 - More sophisticated data copy and migration
 - More attention to TCO of storage
 - Re-purposing data and data mining
- Regulated communities are seeing larger, online requirements including disaster recovery
- Data intensive video community ready to move beyond sneaker-net
- Closer relationship between backup and archiving

Archiving at it's Best: Today



Dual copy

- Asynchronous Replication of each partition in same unit

Remote copy

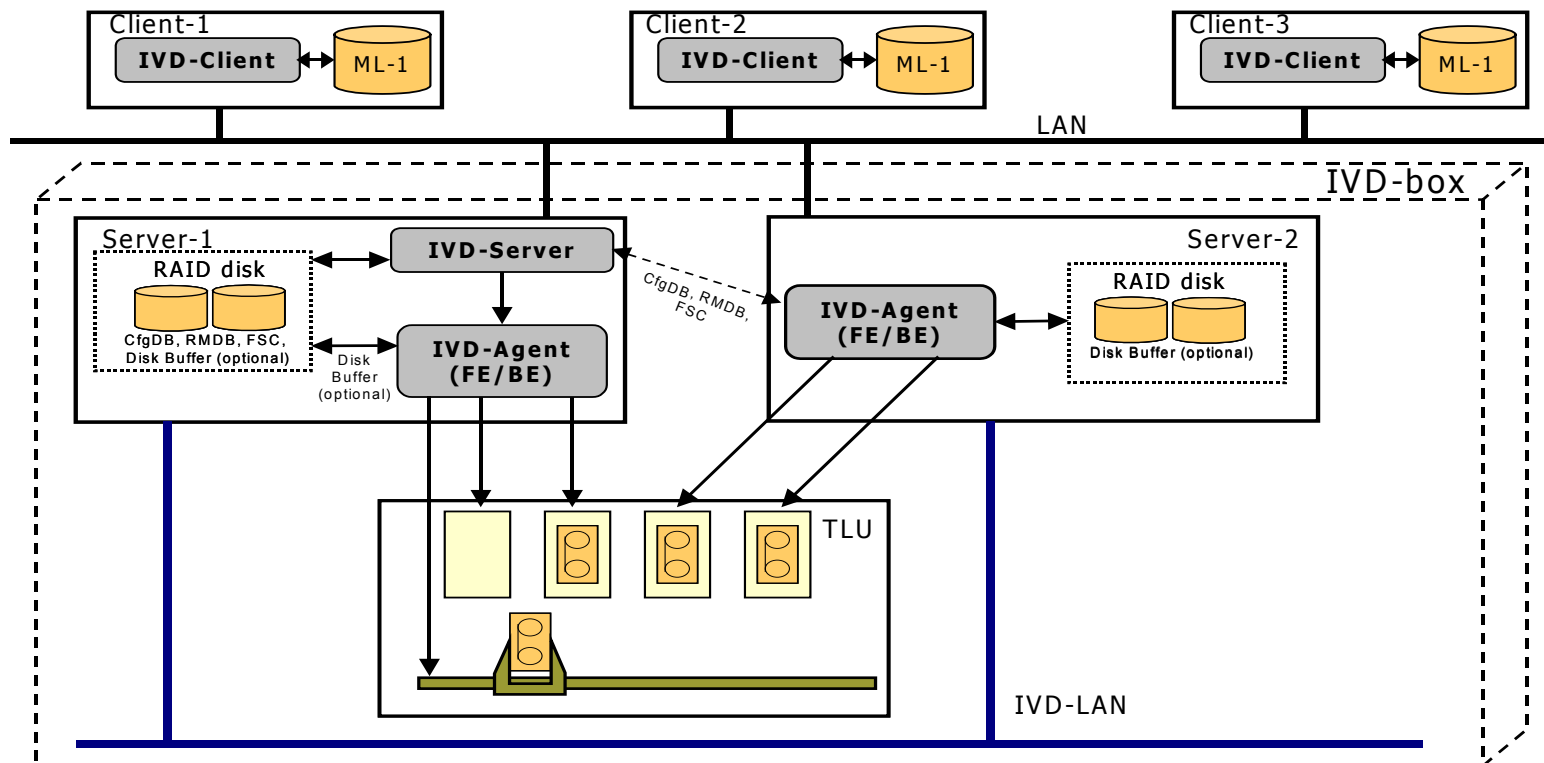
- Asynchronous replication to second unit
 - ◆ via the network
 - ◆ via dedicated network connection
- On campus, across town, across the country

High Availability

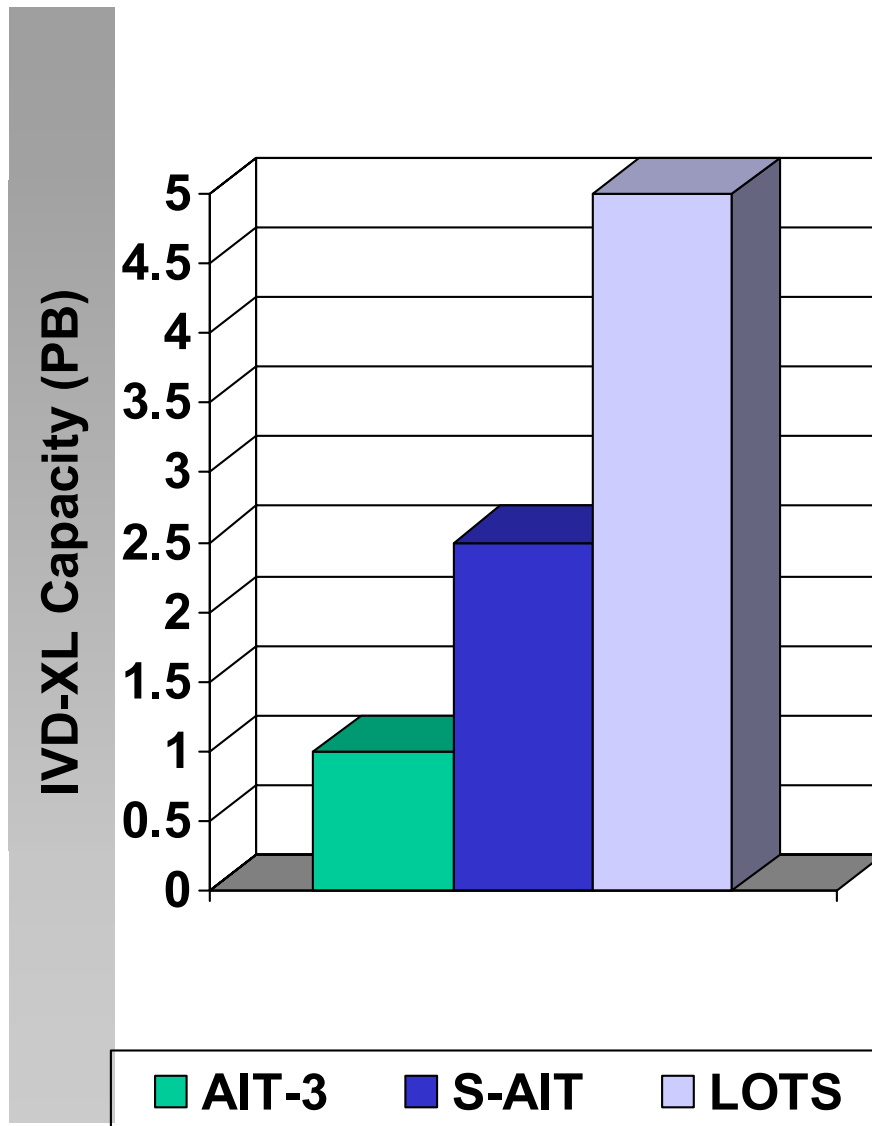
- DFS system
 - ◆ Load balancing
 - ◆ Disaster Recovery
 - ◆ High Availability
 - ◆ Single system view

Archiving later this Year

Multiple Servers within an IVD Complex plus multiple Data Movers to ingest data for the Archive



Grau System Capacities



- AIT-3 at 100GB/Cart
- S-AIT at 500GB/Cart
- LaserTAPE at 1000GB/Cart

- 576 MB/S with AIT-3
- 720 MB/s with S-AIT
- 960 MB/s with LaserTAPE



IVD-XL 1PB of Archive Capacity

