
Storage Models

“Disk-to-Tape Migration” – An Introduction

Michael Ernst

DESY

HEPiX Spring Meeting 2006

Tape is typically seen as ...

- Extremely boring!
 - An afterthought
 - Expensive
 - For backup only
 - Dead!
- “Migration” to be understood generically, i.e. not limited to tape



Challenges in Data Management

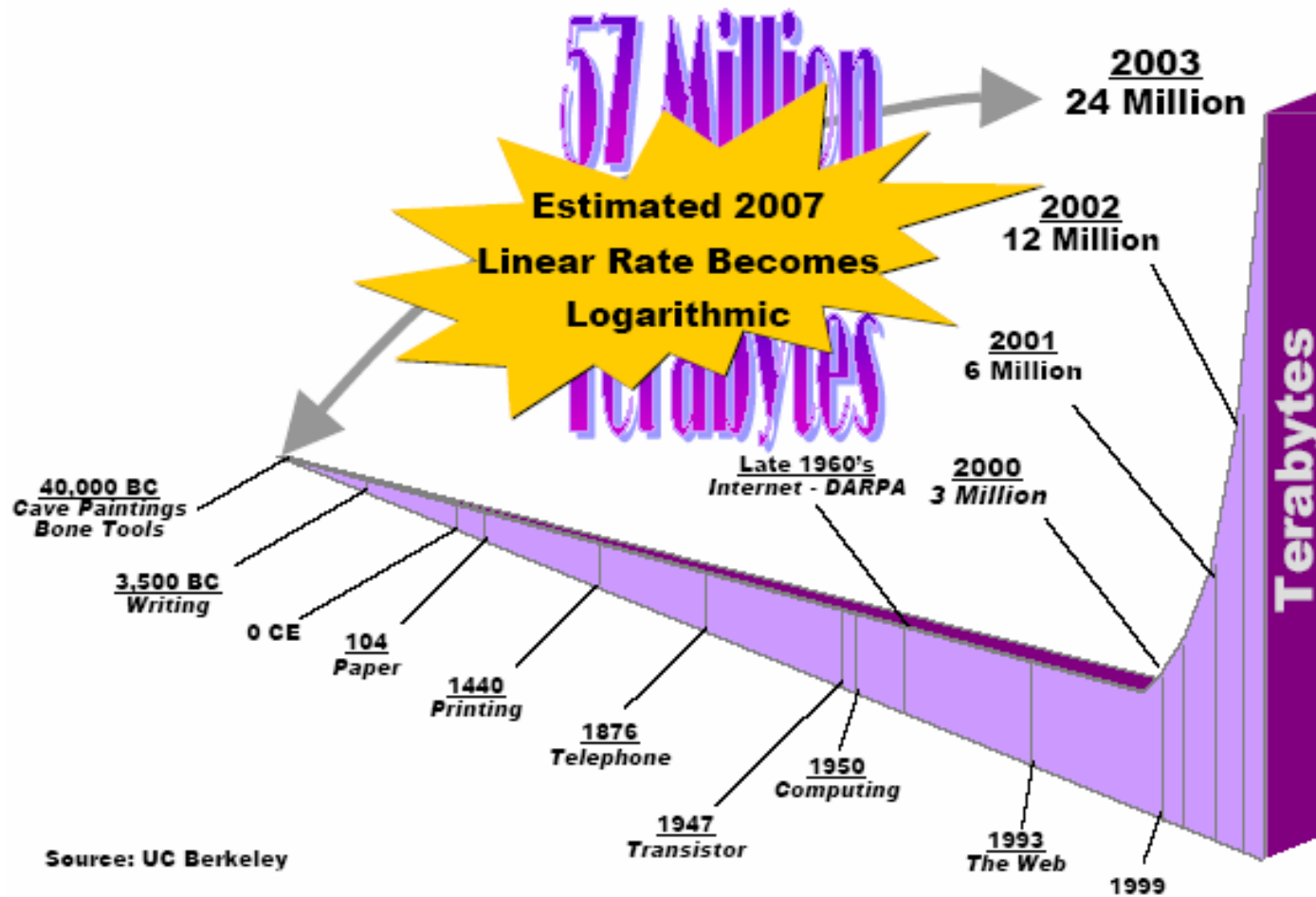
Observed by



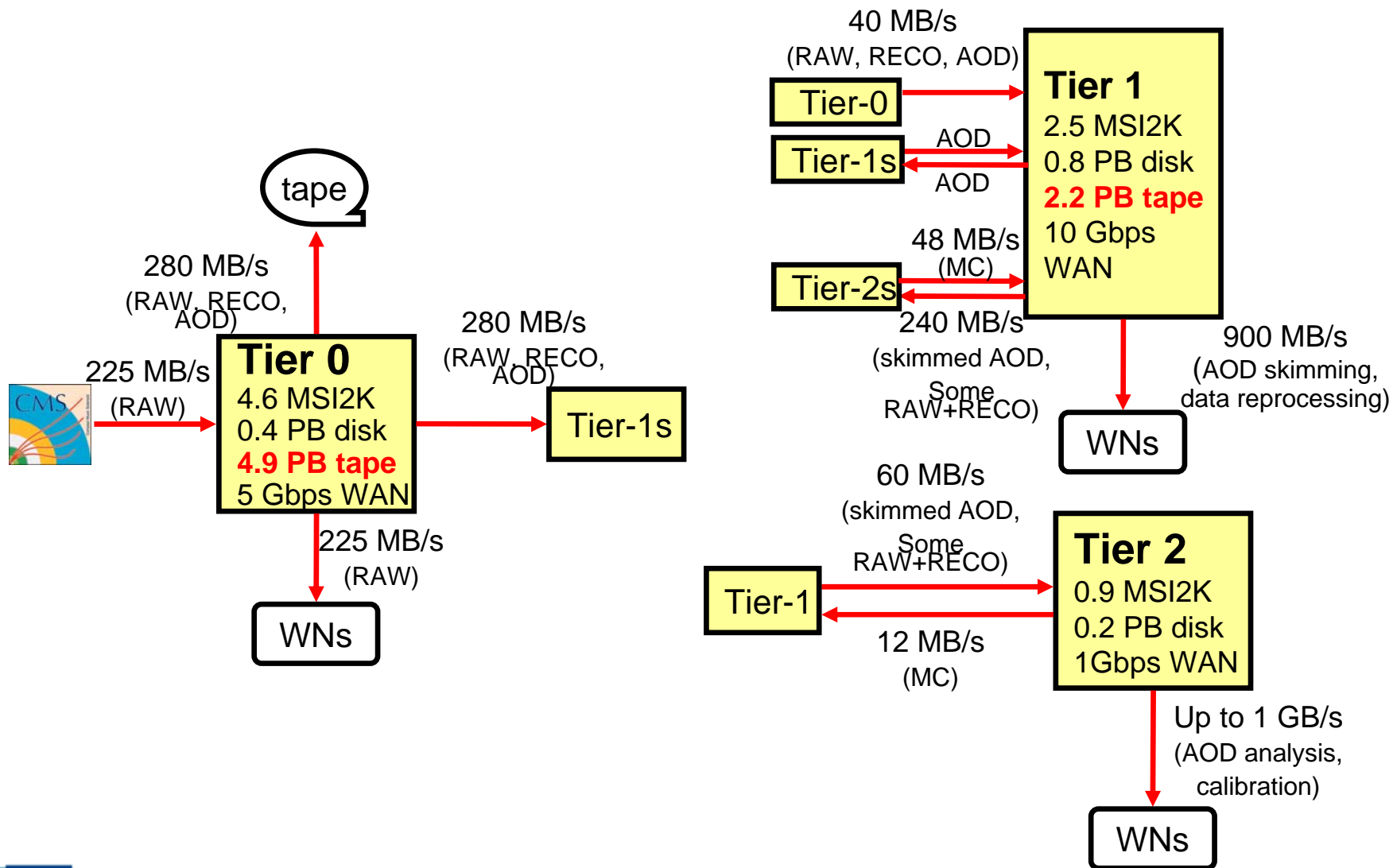
H. D. Simon

- 2000-2005: Data management issues becoming increasingly important.
- 2005-2010: Technology (primarily software) will not be able to meet needs unless it evolves.
- **Challenges**
 - ❑ Lack of standardization, performance, portability
 - ❑ Locating and accessing data becoming more and more difficult
 - ❑ Hierarchical storage management (many different types) will be important to reduce complexity
 - ❑ Scientific workflows copy data multiple times – how can technology help?
 - ❑ Increasing latency/bandwidth gap between memory and disk storage

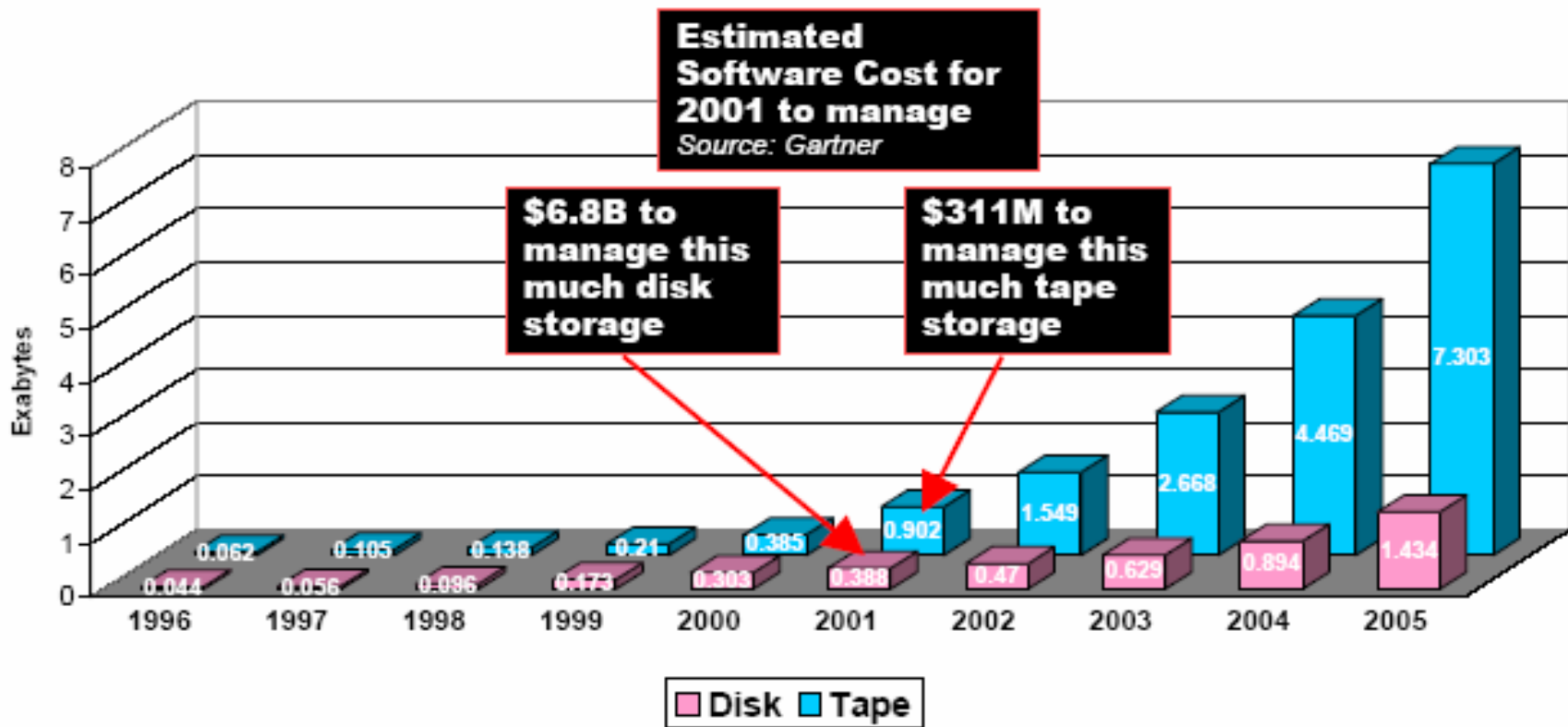
Information Growth



Example: CMS Resources and data flows in 2008



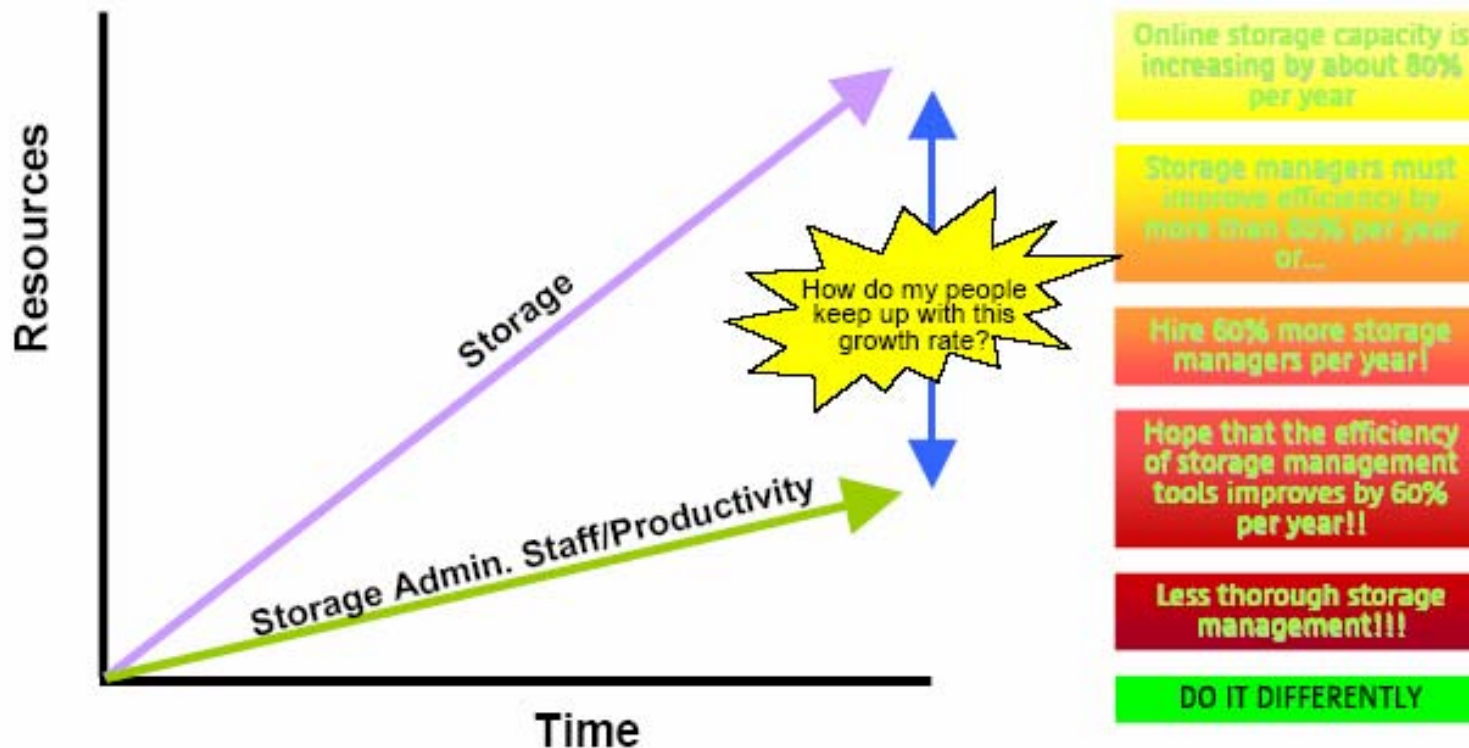
(Disk-) Storage – Cheap to buy, expensive to manage



Tape Source: Gartner Group

Disk Source: IDC

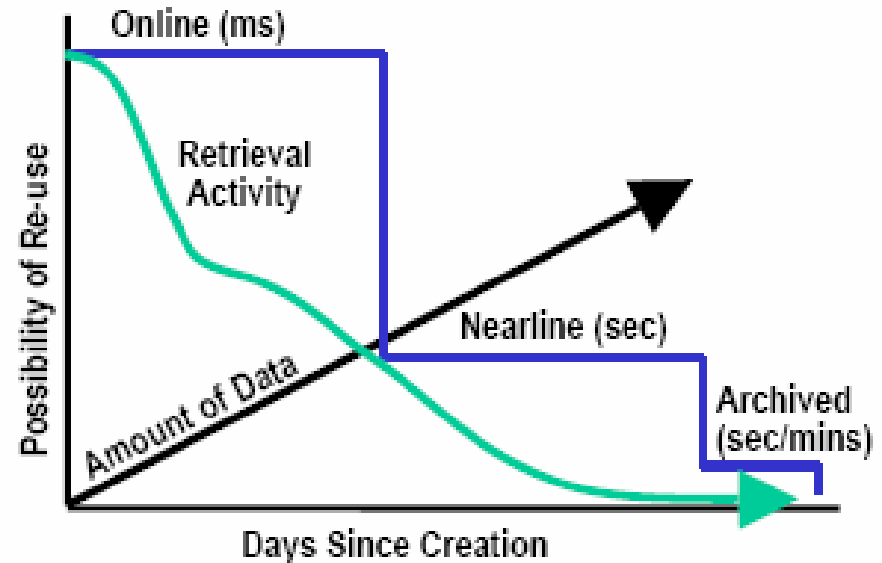
Storage Growth & Too Few People



By Sun Microsystems

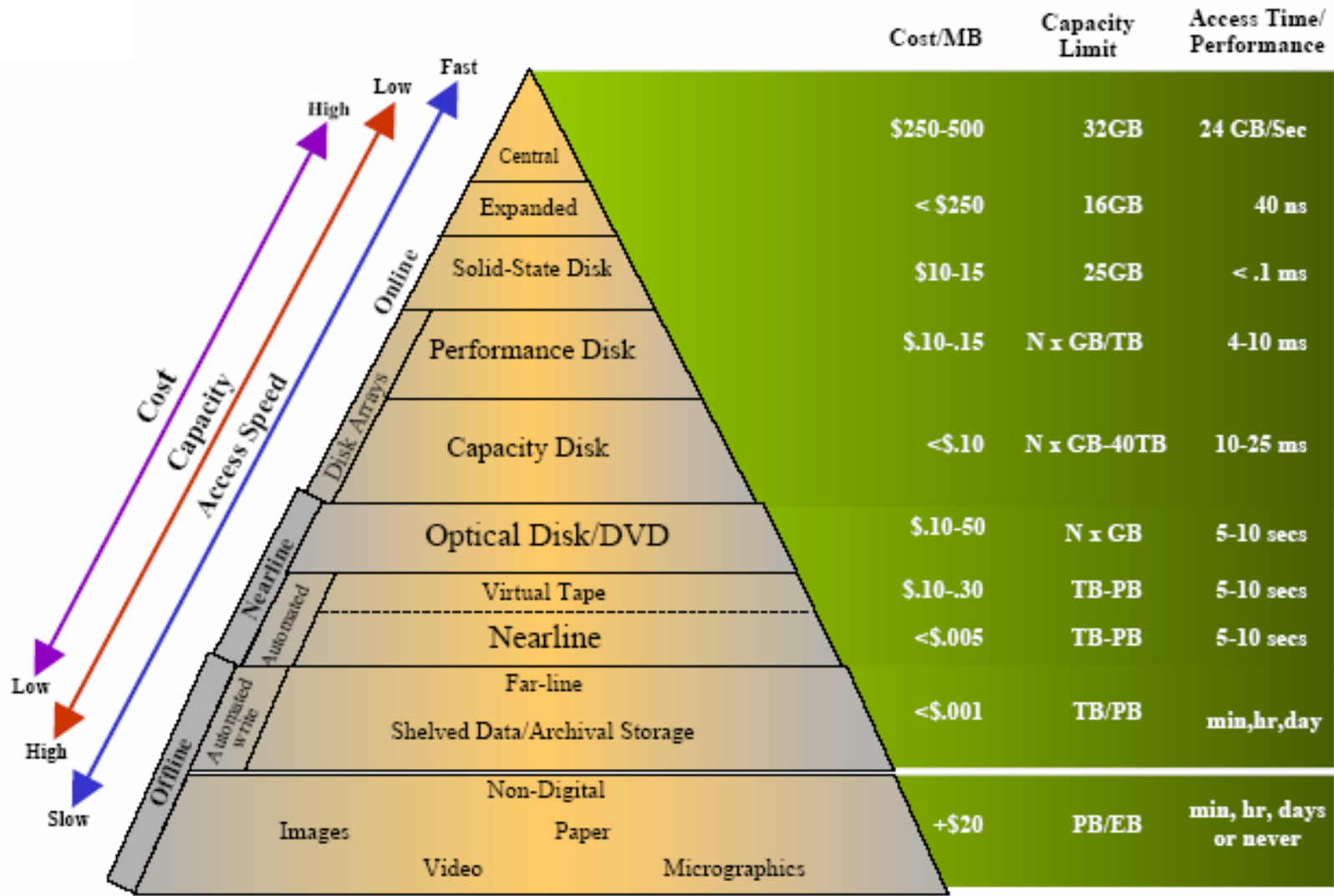
Can we save by managing data/storage differently?

- Data is created
- Most accesses occur within the first day or two
- After one week the data is rarely accessed
- is rarely accessed
- After 90 days the data is almost never accessed
- Different solutions required for different data



- Why do we mostly treat all data on disk the same, regardless of age?

Right Data, Right Place, Right Time, Right Cost



Agenda

- Presentations of three prominent Disk Pool Management Systems either integrated w/ (Tape-) Mass Storage Systems or providing Interfaces to support a Storage Hierarchy
 - Castor
 - dCache
 - HPSS
- ... on Tape Storage Backend for Disk Pool Managers & Stagers
 - TSM
- ... followed by short discussion at the end
- Trying to focus on application oriented issues
 - “Checklist” for compatibility

Criteria to be considered (1)

- **Functionality**
 - Access Protocols
 - Authentication / Authorization
 - Access to namespace service (e.g. URL, NFS, etc)
 - Policy Management
 - user-applied vs. centrally provided
 - pluggable vs. tightly integrated
 - Support for different Storage Spaces (durable, permanent ..)
 - Priorities
 - user definable, applicable to storage classes, queues
 - Support for Transactions (roll-back)

Criteria to be considered (2)

■ Scalability

- Maximum manageable Size (storage space)
- Maximum achievable data transfer rate (single vs. multiple streams, max. # of concurrent streams)
- ... of the Name Service Management Component
 - Maximum # of Files
 - Time required to create/delete a File
 - Latency associated w/ opening a File
- A Metric regarding the performance / unit

Criteria to be considered (3)

■ Flexibility

- ❑ Ability to support multiple/different MSSs and MSS Hardware (e.g. tape drives/robots, optical, inexpensive magnetic disk used in HSM configuration)
- ❑ Namespace administration
- ❑ Support for multiple / heterogeneous platforms (H/W, OS)
- ❑ Variety of supported storage technologies for repositories
- ❑ Configuration Management

IT Security

- Protection against
 - unauthorized access
 - local and WAN access protocols
 - break-ins on service ports
 - unintended alteration of data on storage pools (detection?)
 - corruption of data in transit (local and WAN protocols)
- Resiliency of namespace management component against
 - H/W and S/W failures
 - disaster (effort required to recover from entire loss of the catalog?)

System Operation & Support

■ System Operation

- Effort required to operate an instance (depending on size, application profile & heterogeneity ...)
- Admin Tools offered for efficient management

■ Support

- Continuity of User Support
- Quantify Development / Support Effort available for coming years

■ License Model (Terms and Cost)

- Relevant to HEP Sites
- other non-commercial Research Communities

An Ecosystem for Storage Management

- Platforms, software, institutions, applications, and people who solve HEP Computing Applications can be thought of collectively as an Ecosystem

