Cost Effective Archive with Hierarchical Storage Management (HSM)

Dan Albright

Product Manager

Oracle Corporation

SNIA
SNW
COMPUTERWORLD

October 11-14, 2010
The Gaylord Texan
Dallas, Texas

Wikipedia def:

Hierarchical Storage Management (HSM) is a data storage technique which automatically moves data between high-cost and low-cost storage media.

In effect, HSM turns the fast disk drives into caches for the slower mass storage devices.

Agenda

- Major Challenges

- Archive Facts

- Cost Analysis

Major Challenge$

- Managing data across storage tiers

- Access is complicated, data location is not known, compromising value of data

- Disk system CAPEX and OPEX place significant pressure on IT budgets

- Storage capacity scales, people don't

Archive Facts

Today's Storage Costs - OPEX

- Operating Expense
  - Energy
  - Human resources
  - Floor space

Average Electrical Power usage inside Data Centers

IT equipment as a percentage of IT electrical consumption	
Servers	48%
Storage (disk)	30%
Storage (tape)	6%
Networks	12%
Other	4%

Source: Horison, Inc. and estimates from various industry

Today's Storage Costs - CAPEX

- Capital Expense / Capital Equipment

Subsystem category	ASP (Capex) Range $/GB	Notes
Solid State Disk (DRAM) FC, SCSI (Flash SSD/HDD)	$500-$1000 <$100 (current ASP est.)	ASP range based on capacity and high-availability features
Enterprise disk FICON, ESCON, SCSI, FC, SAS	$12-22	Includes controller, cache & drives. Add-on storage modules with no controller and cache are lower
Midrange disk SCSI, FC, SAS, Virtual Tape, NAS	$2-10	Includes controller, cache & drives. Add-on storage modules with no controller and cache are lower
Optical disc library (Blu-laser) SCSI/WORM	$5-$20	Includes drives, media and library
Economy disk SATA, MAID, JBOD	$1.50-$10	Price range varies widely based on capacity
Automated tape library	$.20-$2.00 (non mainframe) <$.20 (mainframe, varies by configuration and cartridge to drive ratio)	Includes tape drives, media and library and uses a 2:1 compression to determine library capacity

Source: Horison Information Strategies (April 2010 update)

Questions to Ask When Archiving Data
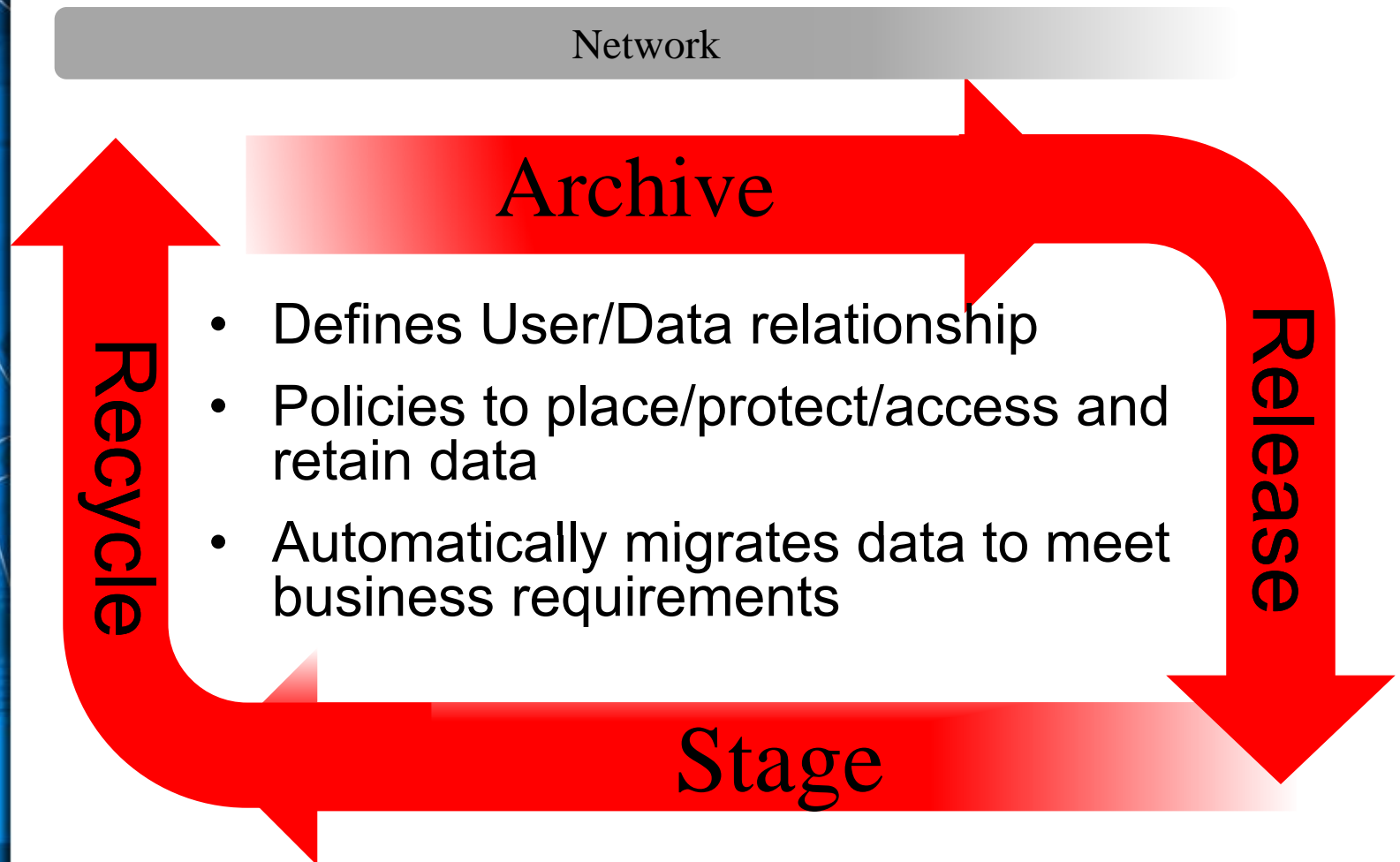
File Servers

Email

Data

- Can data be classified?
- How long is the data retention window?
- What regulation is data governed?
- To what degree does the content need to be protected?
- How will the data be accessed over its lifetime?
- Who needs access to the data?
- How is data searched?
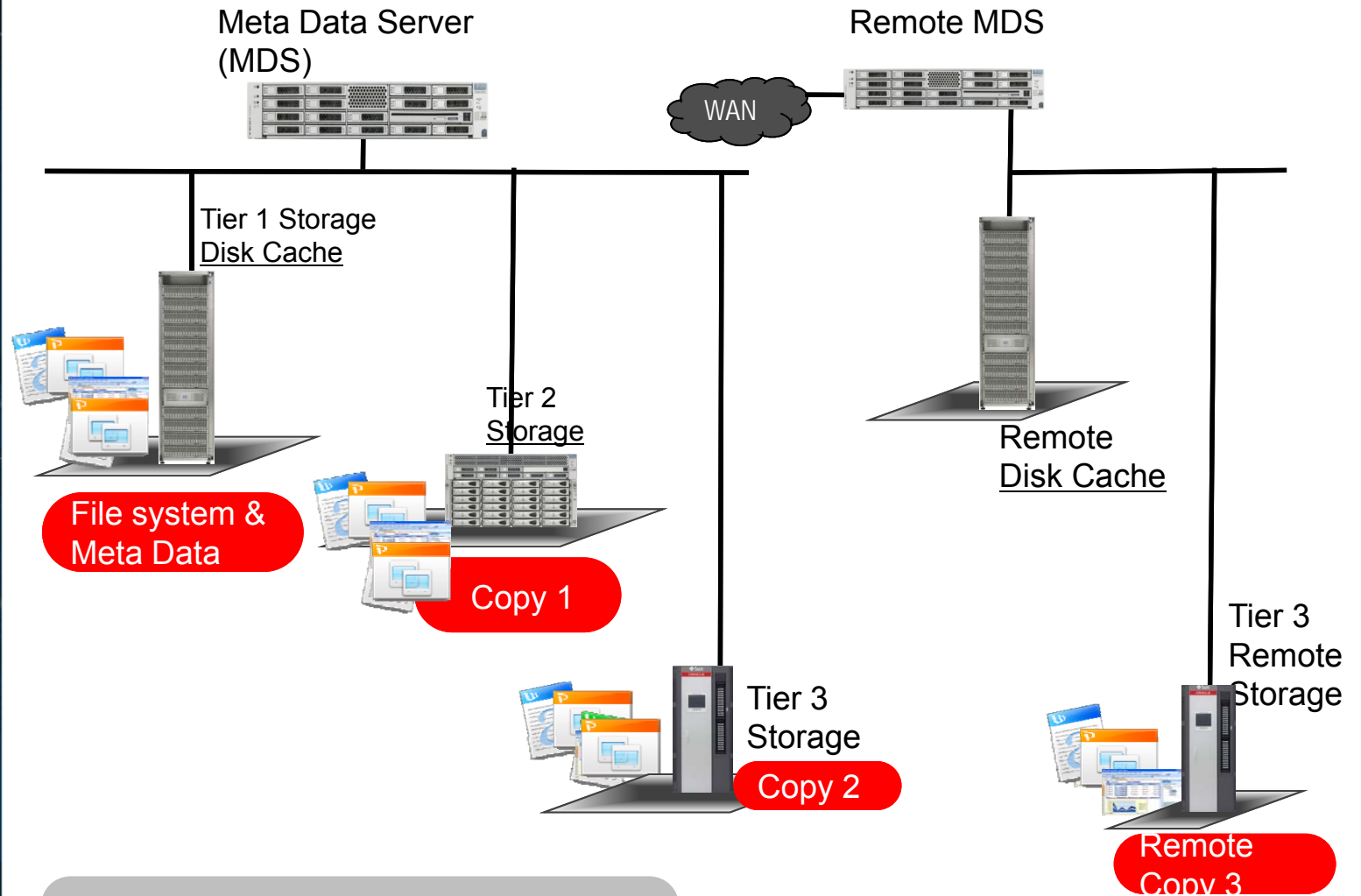- Who is responsible for managing the data and storage?
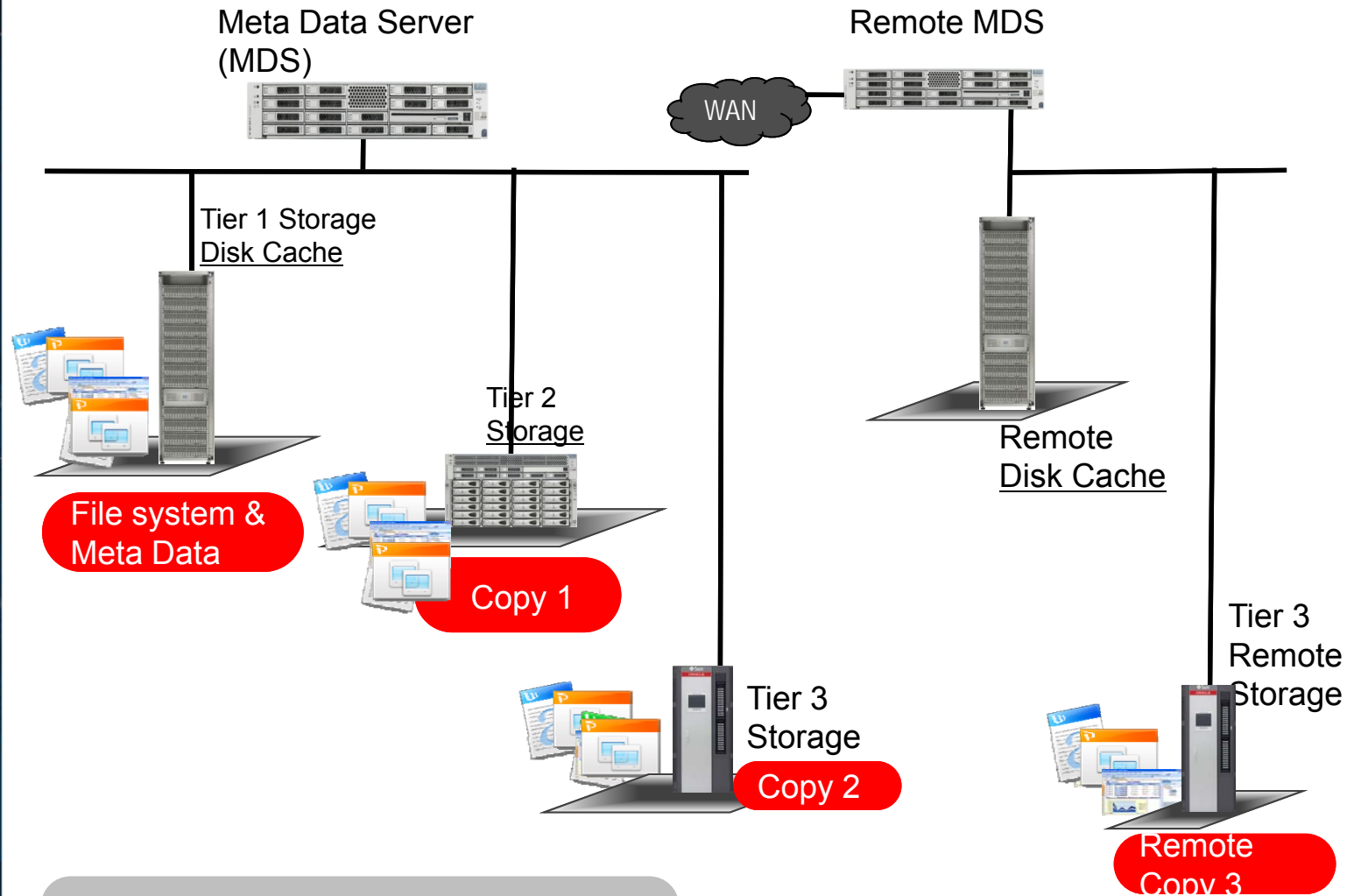
HSM Archive Basics
What is needed– 4 Core Processes

Network

Archive

Recycle

Release

Stage

- Defines User/Data relationship
- Policies to place/protect/access and retain data
- Automatically migrates data to meet business requirements

Process: Initial Archive

Meta Data Server (MDS)

Remote MDS

WAN

Tier 1 Storage
Disk Cache

File system &
Meta Data

Tier 2
Storage

Copy 1

Remote
Disk Cache

Tier 3
Storage

Copy 2

Tier 3
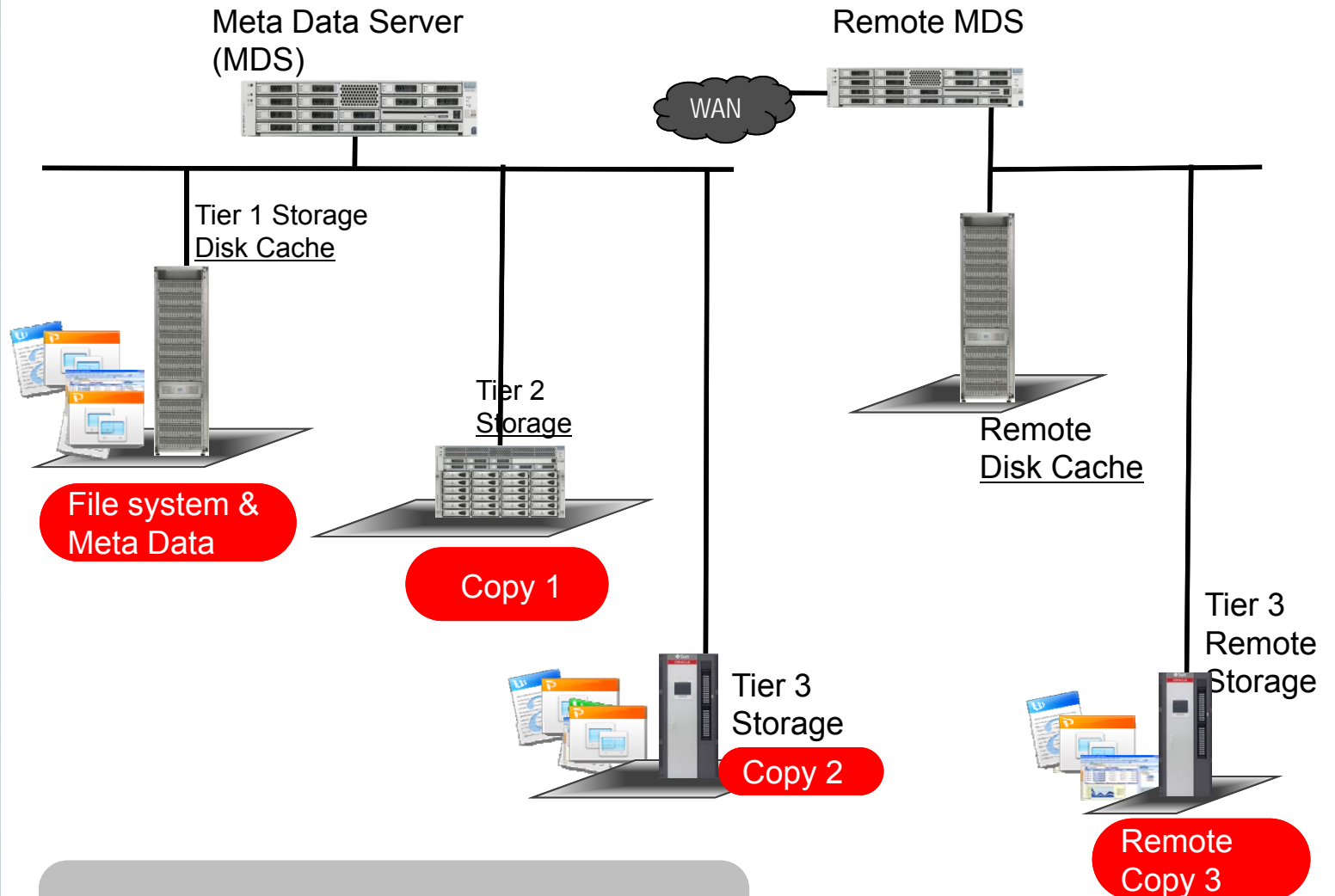Remote
Storage

Remote
Copy 3

−Archive to Tier 2 immediately
−Archive to Tier 3 after 1 day
−Archive Remote Immediately

RMAN

Process: Releasing Data to Another Storage Tier

Meta Data Server
(MDS)

Remote MDS

WAN

Tier 1 Storage
Disk Cache

Tier 2
Storage

Remote
Disk Cache

**File system &
Meta Data**

Copy 1

Tier 3
Storage

Tier 3
Remote
Storage

Copy 2

**Remote
Copy 3**

- Release from tier 1 after 6 months
- Release from tier 2 after 1 year

Process: Staging to Access/ Reuse Data

Meta Data Server (MDS)

Remote MDS

WAN

Tier 1 Storage
Disk Cache

File system & Meta Data

Tier 2 Storage

Copy 1

Remote Disk Cache

Tier 3 Storage

Copy 2

Tier 3 Remote Storage

Remote Copy 3

− Stage from Tier 3 when accessed
− Pre-stage based on file association

RMAN

Process: Recycle to Consolidate/Reclaim

Meta Data Server
(MDS)

Remote MDS

WAN

RMAN

Tier 1 Storage
Disk Cache

Tier 2
Storage

File system &
Meta Data

Copy 1

Remote
Disk Cache

Tier 3
Storage

Tier 3
Remote
Storage

Copy 2
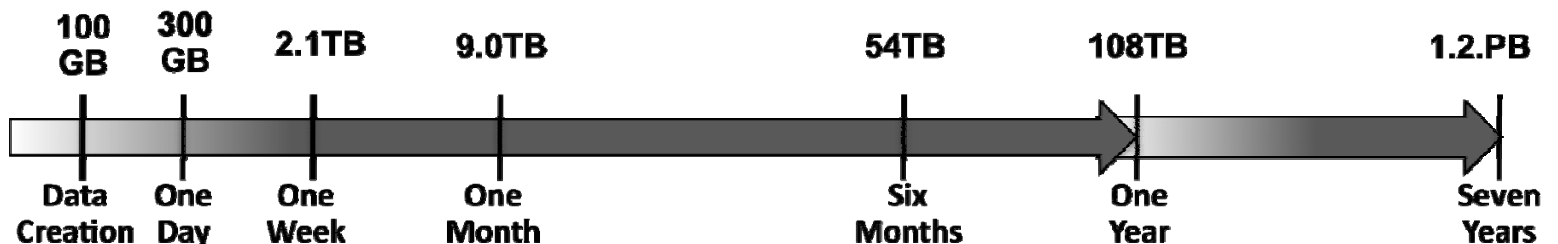
Remote
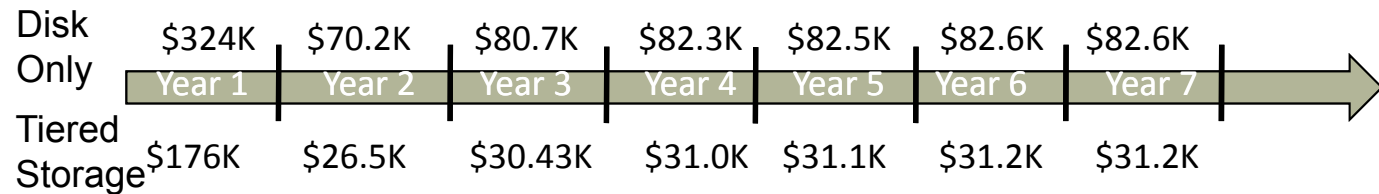Copy 3
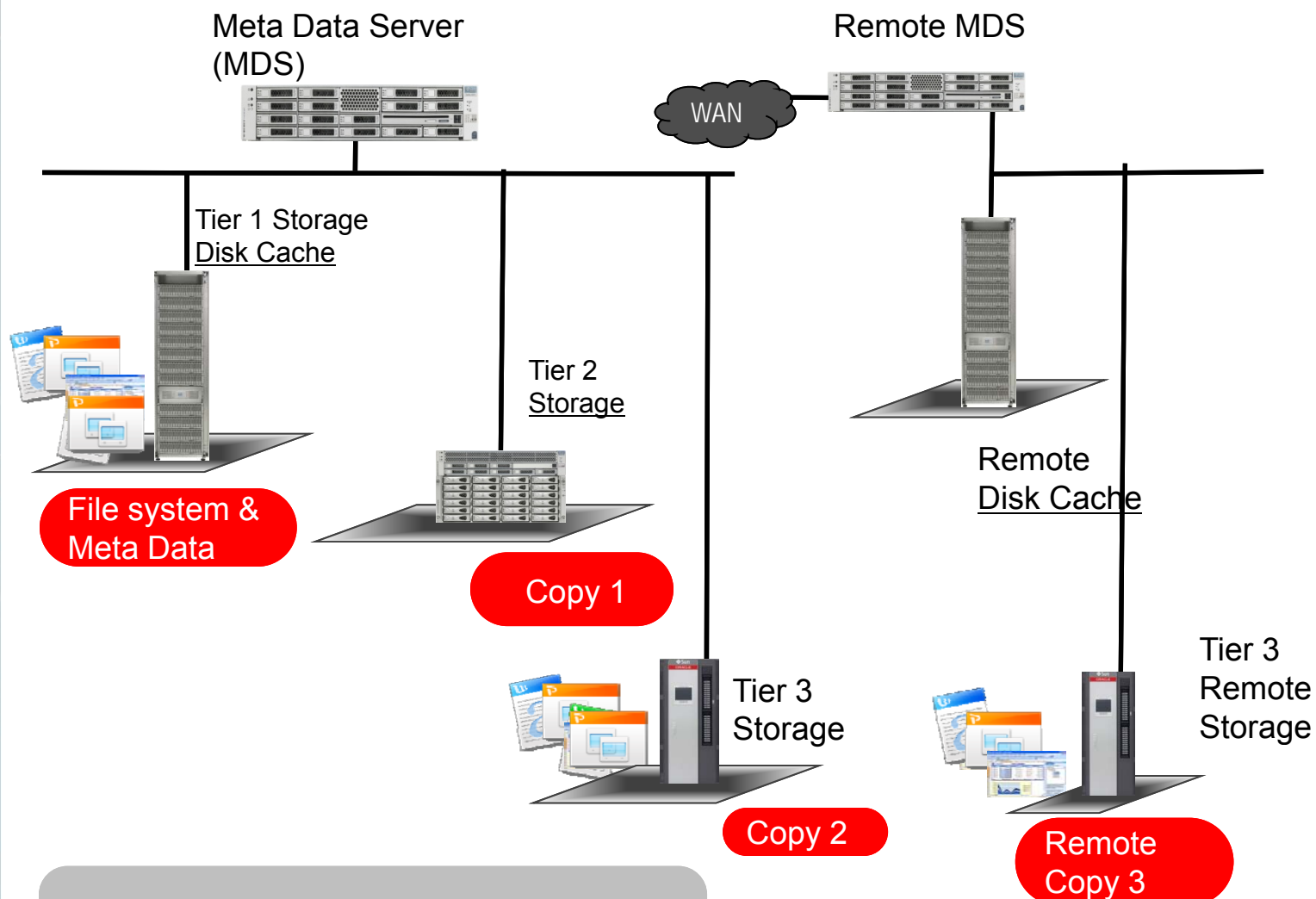
‒ Reclaim space on archive media

Cost Analysis

Example: 7 Year Storage Archive

- 2 Systems Administrators (est. burden $125K each)
- 7 year retention
- 100 GB per day
- 3 copies made
  - Copy 1 for real time use
  - Copy 2 for local redundancy
  - Copy 3 for Disaster Recover and Resume Operations
- Annual Data Growth = 15%
- IRR = 11%

100 GB	300 GB	2.1TB	9.0TB		54TB	108TB	1.2.PB
Data Creation	One Day	One Week	One Month		Six Months	One Year	Seven Years

Example: 7 Year Storage Archive CAPEX and OPEX

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
Disk Only	$324K	$70.2K	$80.7K	$82.3K	$82.5K	$82.6K	$82.6K
Tiered Storage	$176K	$26.5K	$30.43K	$31.0K	$31.1K	$31.2K	$31.2K

	CAPEX	OPEX	Total
Without SAM QFS	$595,034	$1,178,049	$1,773,083
With SAM QFS	$273,187	$ 117,804	$ 309,991
Total Savings	$321,854	$1,060,245	$1,463,092

Summary

An Effective Tiered Archive

- Simplifies tiered storage management

- Optimizes the value of your data

- Reduces CAPEX and OPEX in your IT operations

Cost Effective Archive with Hierarchical Storage Management (HSM)

Dan Albright

Product Manager

Oracle Corporation

Thank You!

October 11-14, 2010
The Gaylord Texan
Dallas, Texas