

AVAMAR



THE NEW WAY TO  
**BACKUP & RESTORE**

You are Drowning in Data  
Relax, Take a Deep Breath

Kevin Daly, CEO Avamar Technologies

THEN

NOW

# The Whyne of the Ancient (IT) Manager

Data, data, everywhere,  
And all the funds did shrink;  
Data, data, everywhere,  
Nor any time to think.

The very deep did rot: O Christ!  
That ever this should be!  
Yea, slimy things did crawl with legs  
Upon the slimy sea (of data).

(with apologies)

Samuel Taylor Coleridge

*The Rhyme of the Ancient Mariner*

# Topics

- Background
- Data Management Challenge
- Data Inflation
- Quantifying Costs
- Data Protection as a Proxy
- Where do we go from here?

# *Background*

# Definitions

- Data: digital content managed by IT systems
- Information: knowledge useful for achieving specific objectives
- Storage: a place to house data

*All three are measured in Bytes (MB, GB, TB, etc)  
... but they are very different entities*

# Storage Classes

Conceptually, we can segment storage into two classes:

- Transaction Class: storage required by IT applications (and people) to meet organizational objectives
- Content Class: storage required by IT systems to meet operation, financial, legal, performance, security and compliance objectives

<i>Transaction</i>	<i>Content</i>
Access	Capacity
Shared	Managed
Local	Distributed
Democratic	Autocratic
<i>Information Sparse</i>	<i>Information Rich</i>

# Content Class Storage Functions

**Backup  
& Restore**

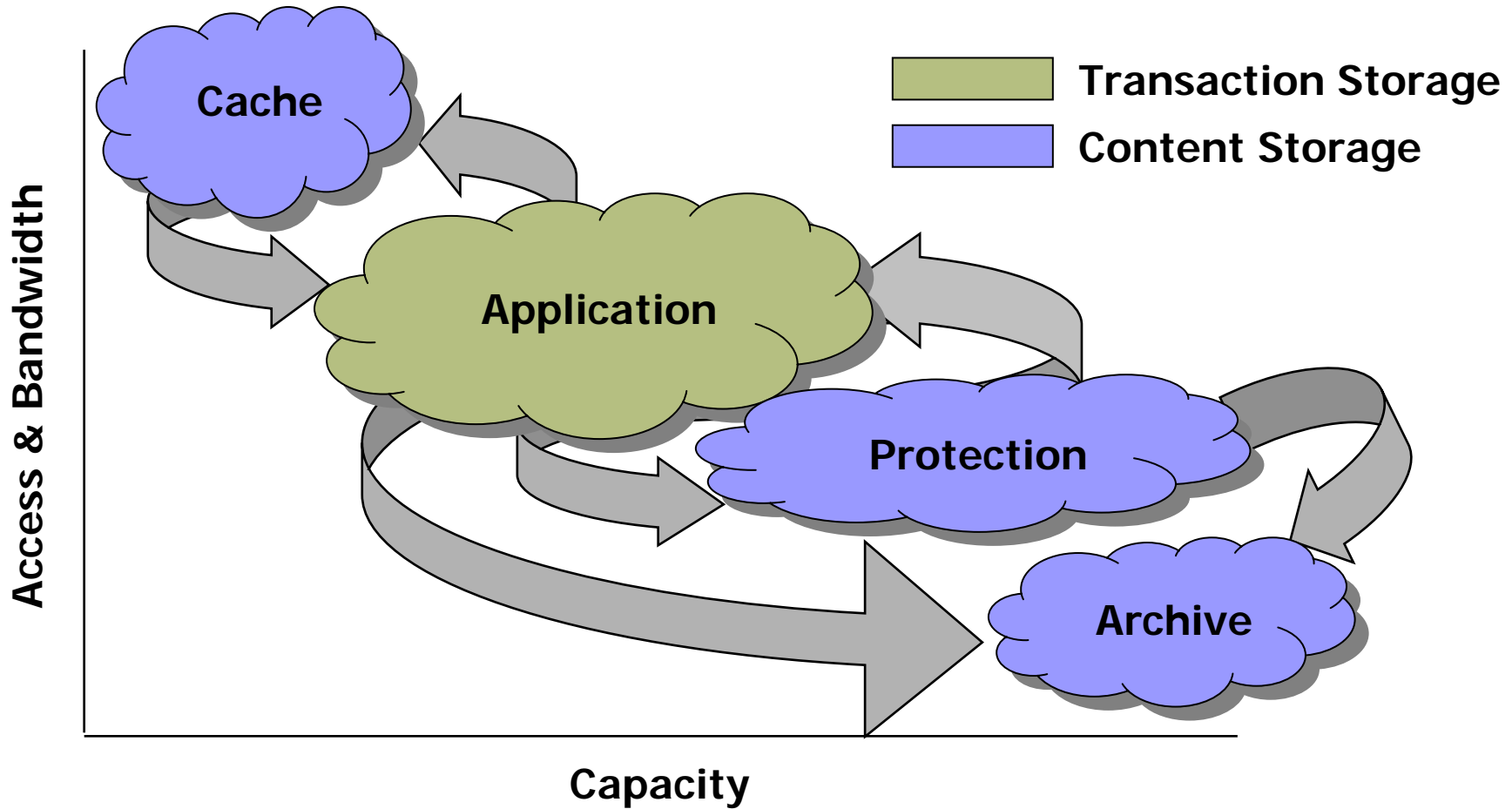
**Business  
Continuity**



**Disaster  
Recovery**

**Information  
Lifecycle Mgt**

# Storage Flows

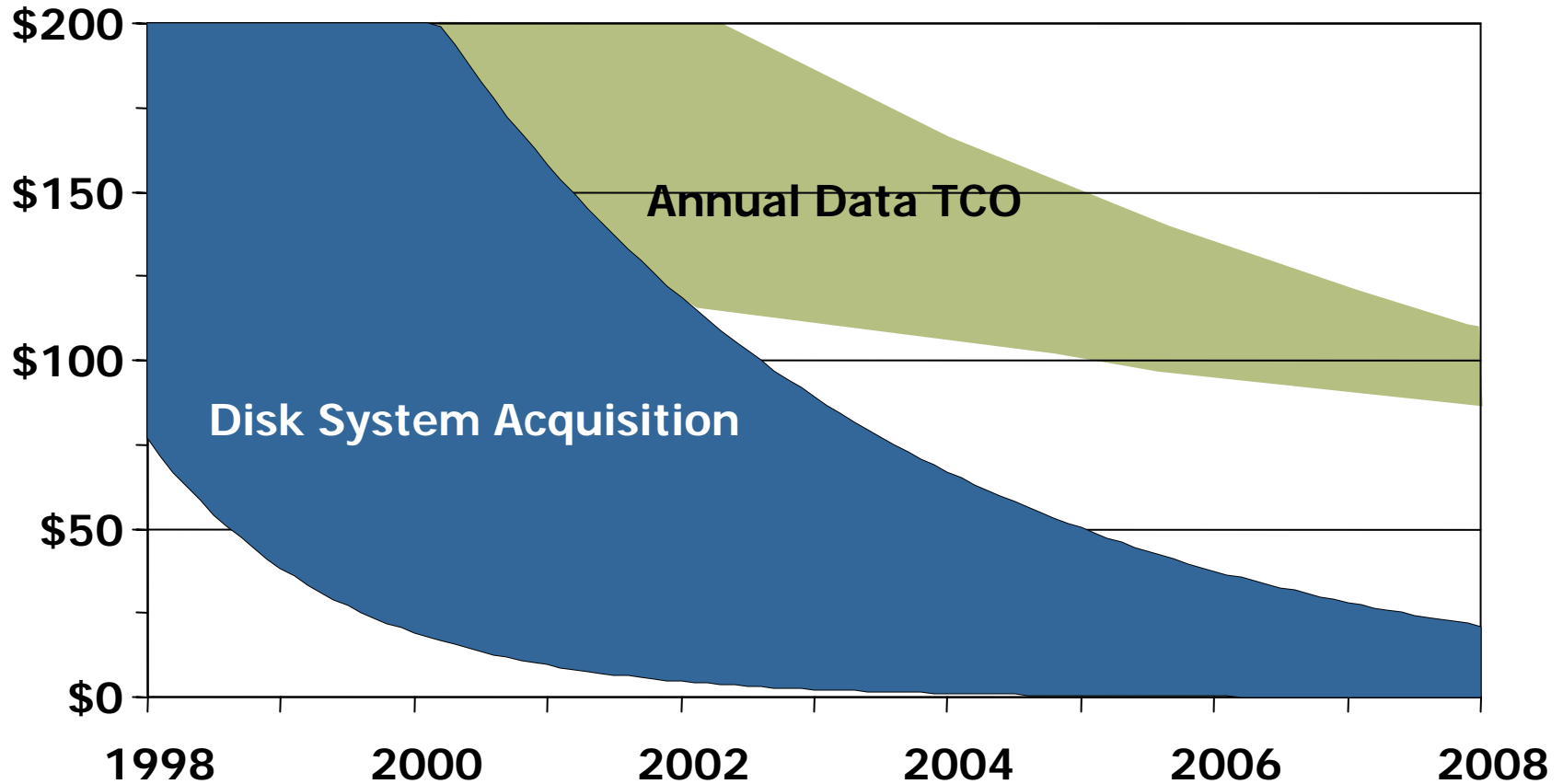


# *Data Management Challenge*

# Mission Impossible?

- Data is what we have – lots of it!
  - It's cheap (and getting much cheaper) to get
  - Its growing much faster than your budget
  - It's convenient for your users
  - But it's oh, so expensive to own
- Information is what we need – and it's buried among all that data
- Your job, should you agree to accept it (and you'd better), is to manage the information without breaking the bank

# Implications of Data Ownership



# The Data Balance Sheet

Data used to be just an asset;  
it is now both an asset and a liability

## ■ Assets:

- 92% of business information
- Productivity
- Competitiveness

## ■ Liabilities:

- Cost
- Discovery
- Compliance
- Privacy

*It's the information, stupid.*

The Mathematical Theory of Communication

Claude Shannon

1964

[Loosely interpreted]

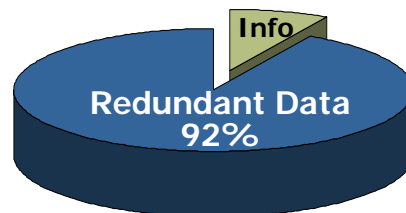
# Information Sources

- We create 6,000PB of new information annually<sup>§</sup>
  - 2,000PB (33%) is on hard disk
    - E-mail represents 400PB of it
- The telephone represents an information\_flow of 20,000PB annually – mostly analog, but rapidly becoming digital (VoIP)
- Internal data networks (LANs & SANs) represent an information flow of 400,000PB annually

<sup>§</sup> Source: SIMS, UC Berkeley, 2003

# Where Data Lives

- We have over 50,000PB of disk capacity
  - 12,500PB (25%) in networked environments
- We are consuming an additional 25,000PB of disk annually
- ... and most organizations have far more data (10x) on tape than they have on disk



***Why do we have so much data?***

# *Data Inflation*

# Data Inflation

- We prefer information-poor data representations
  - A picture may be worth 1,000 words, but it costs 1,000,000 words
- We make matters worse to make our lives easier
  - Caching, replication, etc.
- We keep data for various emotional, operational, and regulatory reasons
- As a result we often create thousands of times as much data as the underlying information requires – this is *data inflation*

# Data Inflation

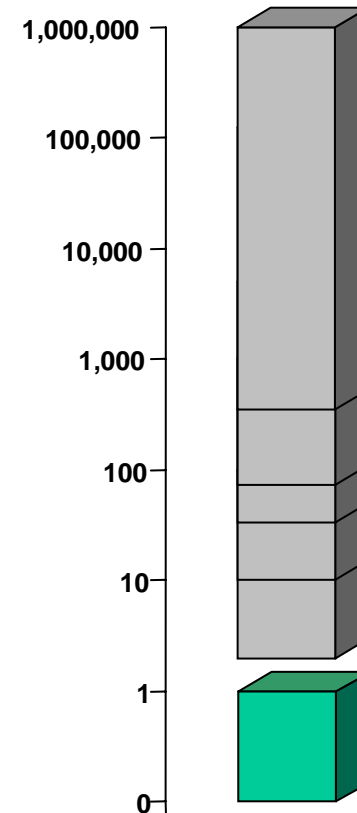
**Representations** – extra data helps communicate information to people

**Reliability** – extra data protects against hardware or system failures

**Recall** – extra data in caches and buffers speed access

**Replication** – extra data permits localization of data to people

**Revisions** – multiple versions create highly redundant data records



# Data Inflation

- Data inflation can be valuable for *Transaction* class storage – trade technical inefficiency for human and process efficiency
- Data inflation is seldom valuable for *Capacity* class storage – increases cost with little operational benefit

*A major storage management challenge is to preserve (or create) data inflation where it is valuable and to reduce (or eliminate) it where it is not*

# *Quantifying Costs*

# Key Cost Issues

*Keep that damn horse out of here!*

(Rough Translation)  
Laocoon & Cassandra  
Troy, 1200 BC

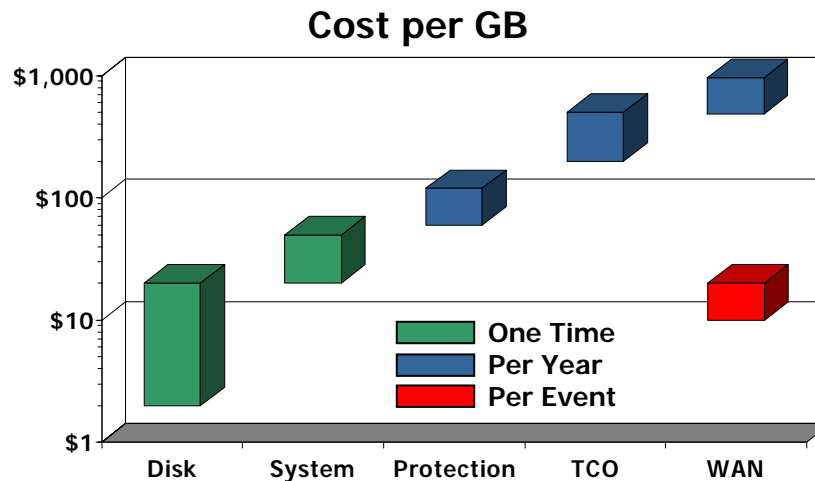
- Low acquisition costs often mask high ownership costs
- Efficiency is important – *brute force* solutions are too expensive
- Bandwidth is only cheap within the data center
- Effective operation requires discipline – not just management

*Catching a tiger is easier  
than letting one go.*

Chinese Proverb

# Cost of Data

*The value is in information, but the cost is in data*



- The median amount of data under management
  - Mid-Tier Enterprise: 1,000-2,000 GB
  - Large Enterprise: ~50,000 GB

# WAN Facts of Life

Channel	B/W [Mb/s]	Transfer [GB/Hr]	LAN Equivalent
<b>T1</b>	1.5	0.7	N/A
<b>T3</b>	45	20	100baseT
<b>OC3</b>	155	70	
<b>OC12</b>	622	280	GbE

***It costs between \$10/GB and \$20/GB [each time] to transfer data on a WAN***

# New Considerations

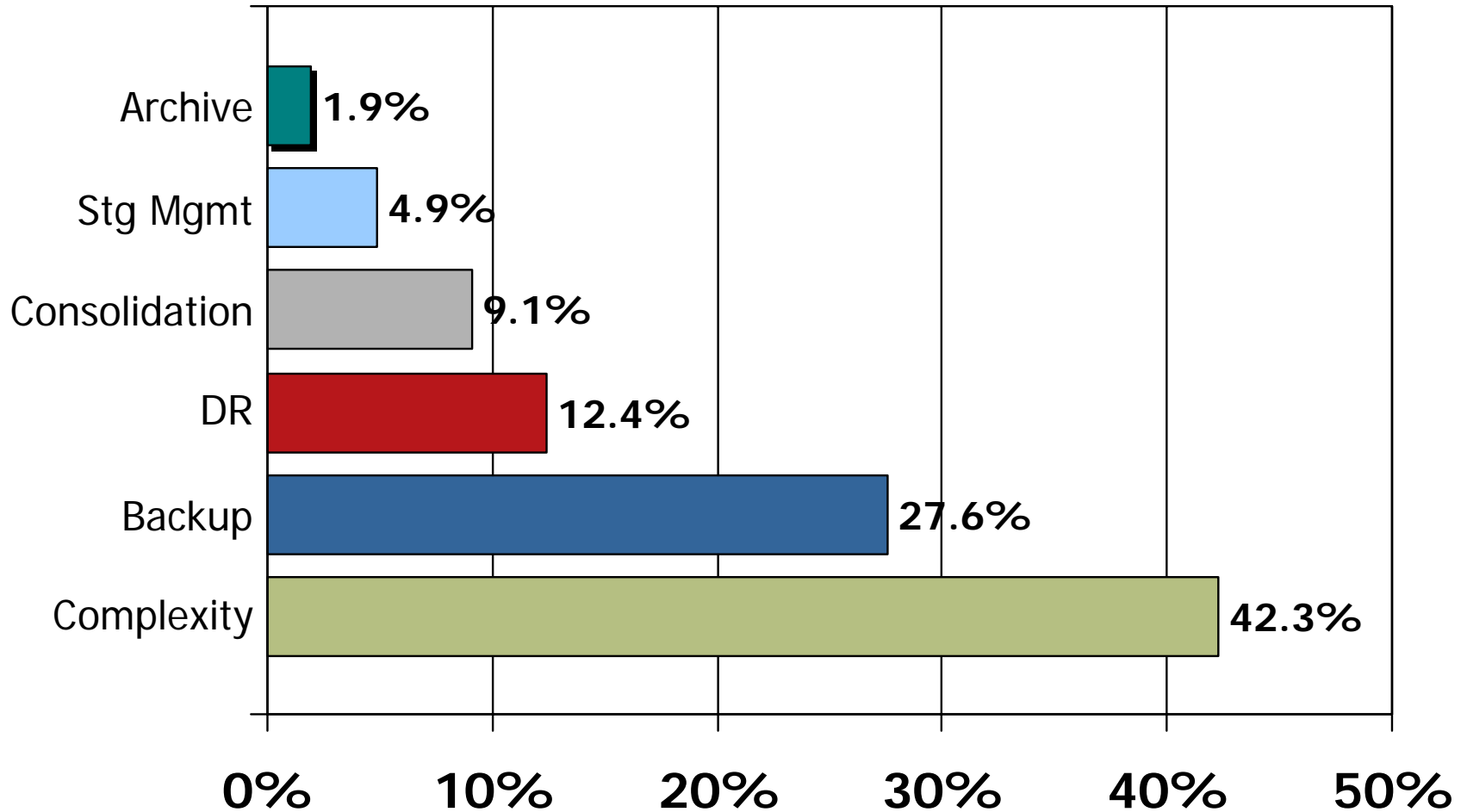
- Organizational dependence on data assets is just about complete
- Critical data has diffused from the data center
  - 30%-50% of critical data is *out there*
- Data loss events are no longer inconceivable
- WAN bandwidth costs are declining ... but they are still very high
- New service provider models are emerging
- ... and then there's compliance

# *Data Protection as a Proxy*

# Why Data Protection?

- It represents a large component of overall storage spending
- It is the nexus of storage and communications – both significant cost areas
- There are new – and evolving – requirements
- Everyone does it – in some way
- It is undergoing significant technology change

# Top Storage Spending Areas



Source: Strategic Research "Spending to Solve Top Storage Pain," 2004

# Disk-Based Data Protection

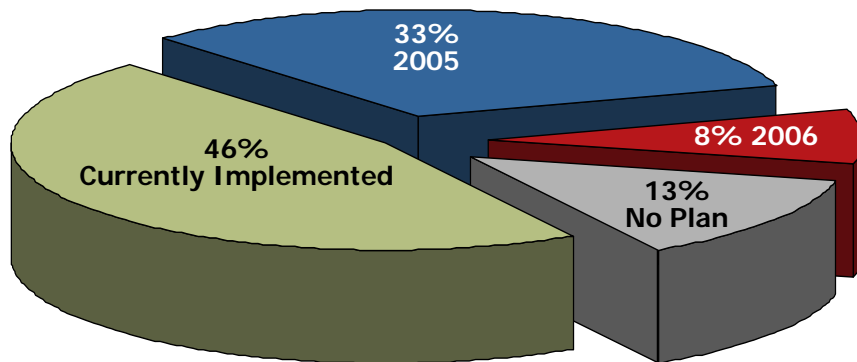
- Moves data protection into the mainstream of data management
- Rich set of technologies and processes
- Bridges large performance gaps
- Addresses new and emerging requirements
- Permits significant enhancements in data efficiency

# Disk-Based Data Protection

*Disk-to-disk backup  
is the future.  
Plan for it.*

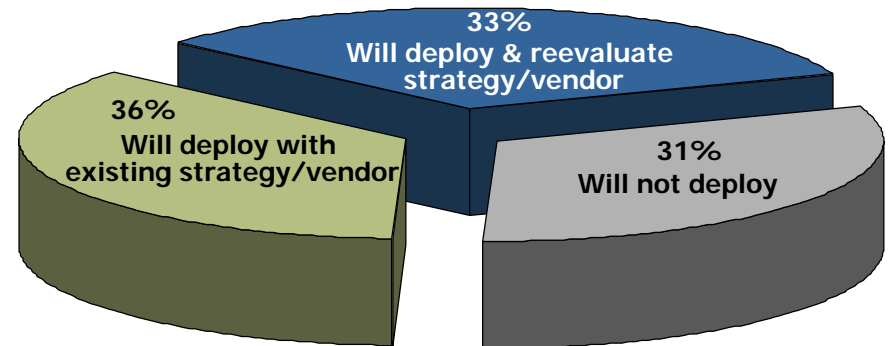
Carolyn DiCenzo  
Vice President, Gartner Research  
Data Center Conference, Dec 2004

## Plans for Disk-Based Backup



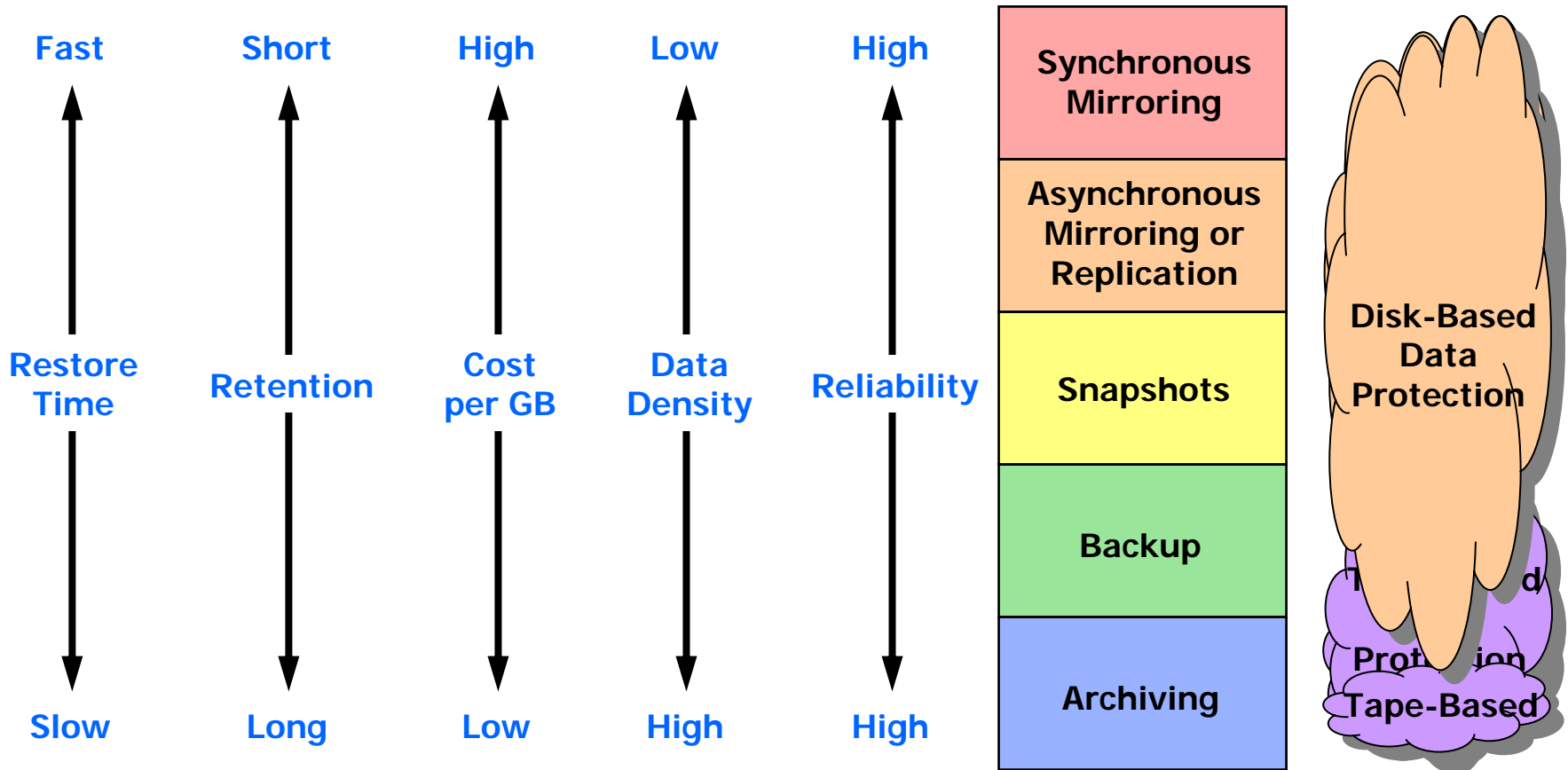
Source: InfoStor Backup/Recovery Survey, April 4, 2005

## Plans Over the Next 12 Months

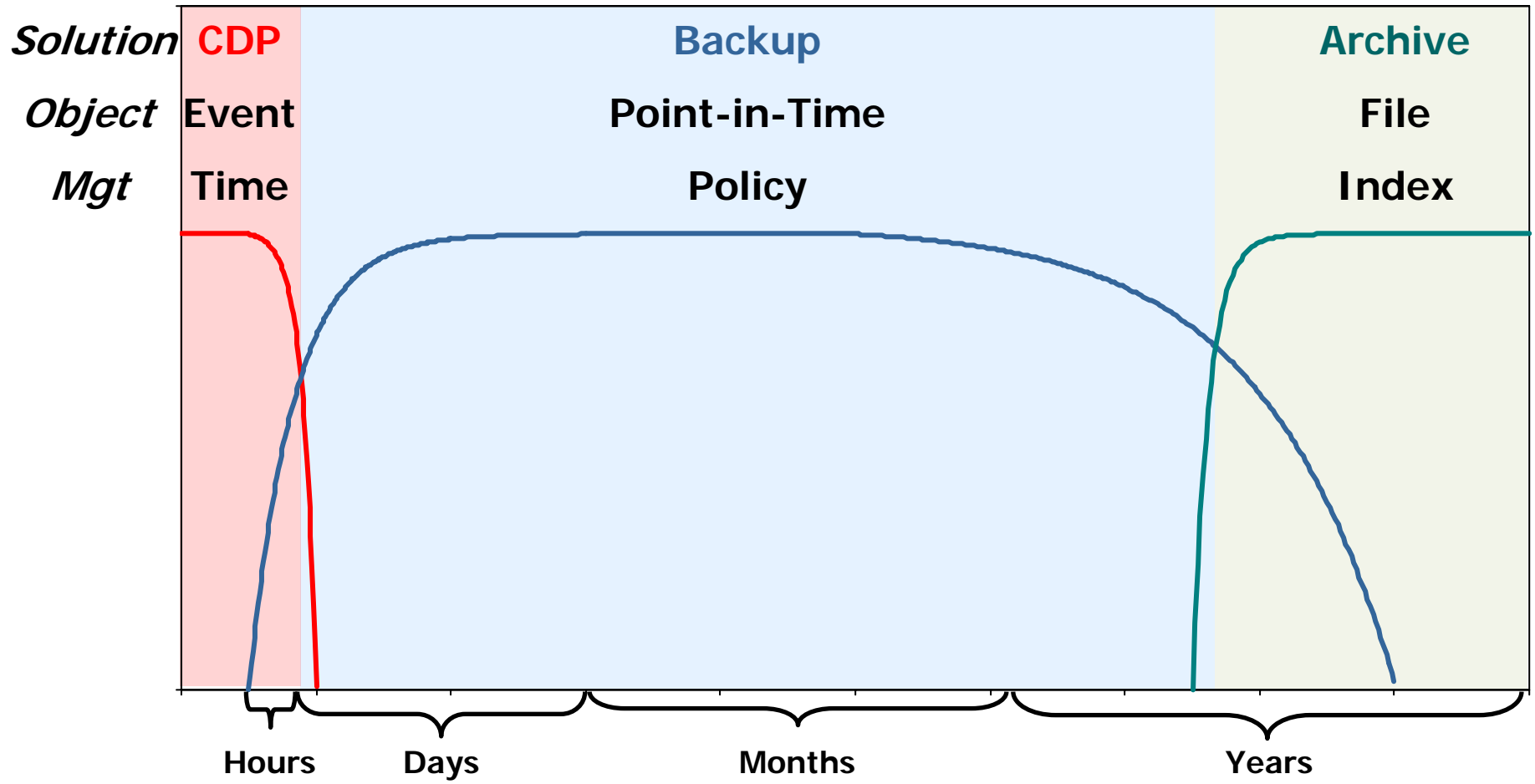


Source: Goldman Sachs IT Spending Survey, Nov 16, 2004

# Data Protection Spectrum



# Data Protection Solutions



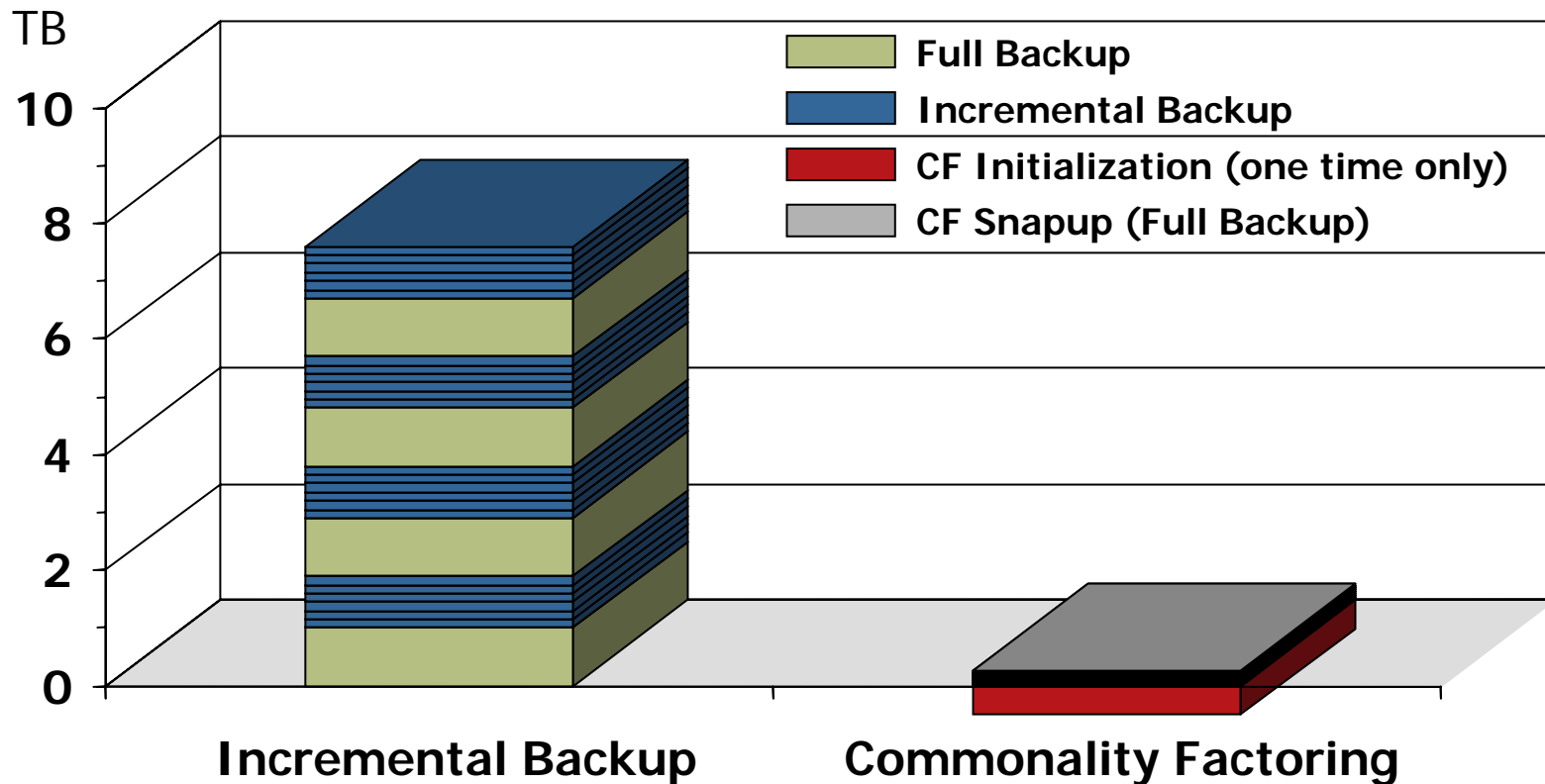
# New Data Protection Capabilities

***Disk-based data protection systems can significantly reduce data inflation in Content Storage***

Backup Process	Tape-Based	Disk-Based
Approach	<u>File Incremental</u>	<u>Commonality Factoring</u>
Elemental Unit	File	Sub-file Object
Strategy	Scope Limitation	Single Instance Store
Metric	File Metadata	Object Signature
Scope	Local	Global
Data Reduction	6x – 10x	100x – 300x

# Backup Data Transfer – First 30 Days

## Data Transfer to Protect 1TB of Data



# Data Inflation

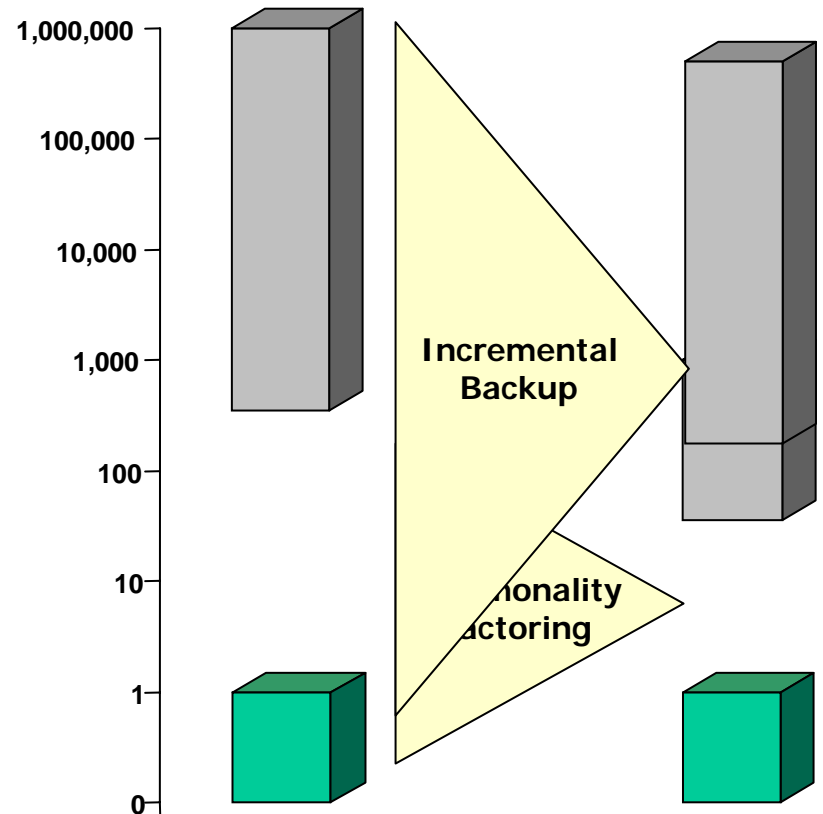
**Representations** – extra data helps communicate information to people

**Reliability** – extra data protects against hardware or system failures

**Recall** – extra data in caches and buffers speed access

**Replication** – extra data permits localization of data to people

**Revisions** – multiple versions create highly redundant data records

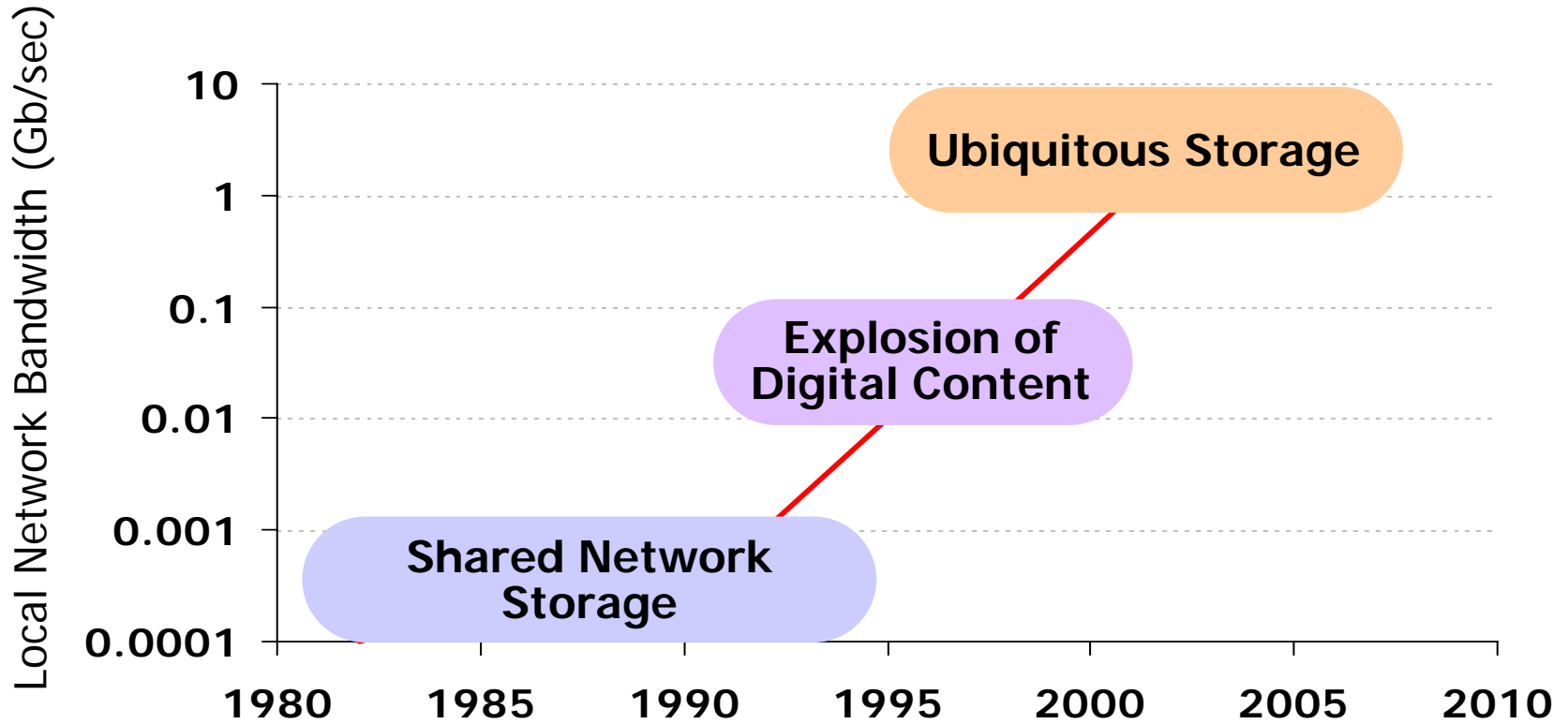


# *Where Do We Go From Here?*

# Current Trends

- iSCSI, sATA, SAS – now everyone can acquire more storage than they can possibly manage
- WAN bandwidth costs are decreasing – but they are still 100x LAN costs
- Compliance
  - (a) Everything can be digital
  - (b) If it is digital you are responsible for it
  - (c) If in doubt – see (a)

# Bandwidth Drives Storage



# What This Means for Storage

- Capital investment levels are increasing
  - 60% plan increased spending in 2005
- Disk system replacement of removable media will accelerate
  - Fixed content
  - Disk-based data protection
  - Compliance
- Storage complexity [e.g. tiered storage, ILM] is expensive and disempowering to end users
- You must contain Data inflation

# An Analogy: You Can Either ...

Address the issue  
at its source

**OR**

Try to handle the  
consequences



# Conclusions

- Storage will continue to increase as a percentage of overall IT spending
- Storage management growth will accelerate (Gartner predicts 12% growth to \$6.3B in 2005)
- Storage management functionality is becoming more strategic
  - Financial Return
  - Business Effectivity
  - Operational Improvements
  - Regulatory Compliance
- Information efficiency will trump both brute force and management elegance

*This is my stuff, that's your stuff, that'll be his stuff over there. That's all you need in life, a little place for your stuff ... wouldn't want somebody to come by and take some of your stuff. They always take the good stuff. They never bother with that crap you're saving ... Aww, no. NOW what ... you gotta pack an even SMALLER version of your stuff ... Only the stuff you know you're gonna need ... well, only the stuff you HOPE you're gonna need.*

*On Stuff*  
George Carlin