SNIA

STORAGE NETWORKING INDUSTRY ASSOCIATION

EDUCATION

ILM: Tiered Storage & The Need For Classification

Nik Simpson (Scentric Inc.)

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.

About SNIA and the DMF

About the Storage Networking Industry Association (SNIA)

- SNIA's primary goal is to ensure that storage networks become complete and trusted solutions across the IT community
- For additional information about SNIA see www.snia.org
- SNIA's "Dictionary of Storage Networking Terminology" is online at www.snia.org/dictionary

About the Data Management Forum (DMF)

- Founded in 2004, the Data Management Forum is a sub-group of SNIA specializing in data management and protection throughout the lifecycle of information.
- More information about the DMF including resources on data and information lifecycle management can be found at www.snia-dmf.org

Agenda

- Tiered storage
  - What is it?
  - Why would you want it?
  - What's missing?
- Data classification
  - Why do it?
  - Using data classification to create & assign SLOs
  - Combining data classification with tiered storage
- Industry directions

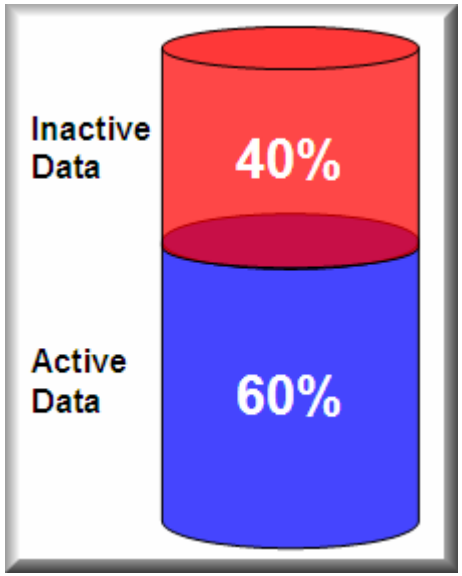
Abstract

Tiered Storage & The Need For Classification

This session will appeal to Storage practitioners looking to understand the benefits of information and data classification as part of any tiered storage implementation.

The session will examine the need for classification and how classification can be applied to help tier data in a tiered storage environment. The session will also look at the emerging solutions and technologies aimed at automating aspects of the classification process.

What's In Your Storage?



Typically, 40%* or more enterprise data is inactive, but…

- It consumes expensive storage capacity
- It gets managed, backed up, replicated, ...
- It has significant legal & compliance risks
- It all has to be recovered in a DR scenario

* Source – Strategic Research Corp.

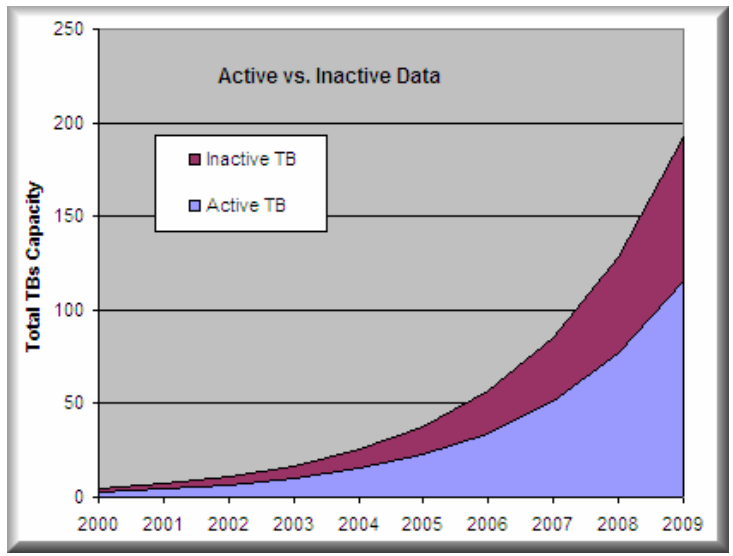
5TB of total data in Y2000 translates to

- 115 TB of active data
- 77 TB of inactive data

by 2010 ...

(assumes 50% CAGR*)

* CAGR = Compound Aggregate Growth Rate



“Tiers” of Storage

- “Tiers” of Storage = service classifications for storage
- ILM is much more than just tiered storage

- Tiering is relative to your environment
  - It’s about quality of service attributes; speed, reliability,...
  - Tiers can be specialized, i.e. Compliance archive
  - Tiers can be non-disk, i.e. CDR, DVD-RW, Tape etc

Tier-1	Tier-2	Tier-n	...	Specialized
Highest performance Most reliable Most scalable Most expensive	Medium performance OK reliability Limited scalability Less expensive	Hi-capacity Low performance Very low $/TB		Offsite tape archive Write-once compliance archive D2D-Backup CDP...

Storage Technology & Tiers

SNIA
STORAGE NETWORKING INDUSTRY ASSOCIATION
EDUCATION

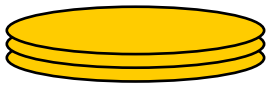
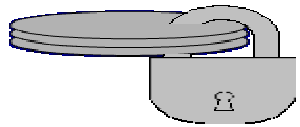
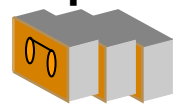
Attribute Name	FC/SAS	SATA	MAID*	Tape**	Optical
Initial Access Time (MTFB)	ms	ms	ms-secs	secs-hours or more; Depends on location of cartridge	secs
Storage Density	Moderate	Moderate - High	High	High	Low
Reliability	High	Moderate	Moderate-High; Long Service Life	Low-Moderate; Long Service Life	Moderate-High; Long Service Life
Data Redundancy	RAID	RAID	RAID	Media Duplication	Media Duplication
Power/Cooling	High	High	Low-Moderate	Very Low - Low Depends on number of drives	Low-Moderate Depends on number of drives
Storage Cost	High	Moderate	Low	Very Low - Low Depends on number of drives.	Moderate Depends on number of drives.

* Lower power duty cycle results in lower power-cooling needs, longer effective service life with fewer technology refreshes versus std SATA drives. Ref. Anderson et al,"More than an interface - SCSI vs ATA," 2nd FAST Conference, March 2003

** Lowest for large ratio of cartridges to drives

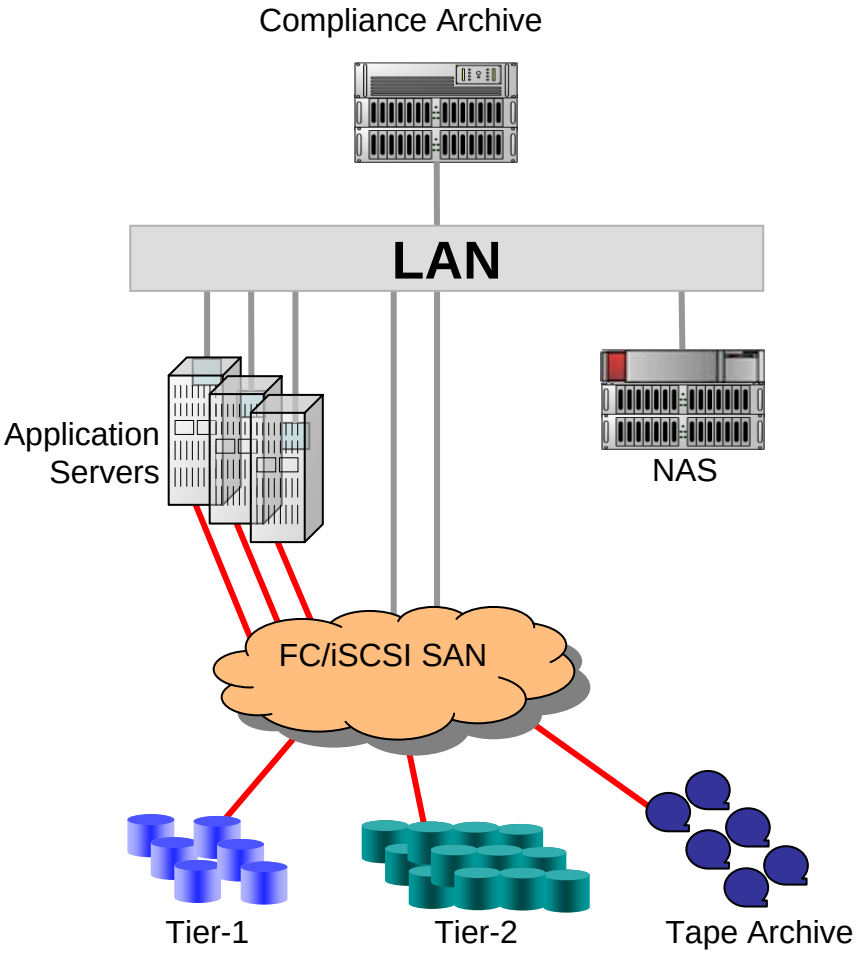
The business case looks simple...

SNIA
Storage Networking Industry Association
EDUCATION

	Tier 1 	Tier 2 	Compliance 	Tape vault 
Characteristics	Hi-performance, hi-availability	Medium performance, high storage density	Write-once, Secure, Immutable	Low performance, removable
Usage	Production data	Reference data	Compliance data	Offsite DR
Relative Cost (CAPEX/OPEX*)	10	2	3	1
%age capacity	20%	35%	30%	15%
Weighted cost	3.75 (62.5% reduction vs. all Tier 1)			

* CAPEX/OPEX = Capital & Operational Expenditure

What it might look like...



So that was easy!

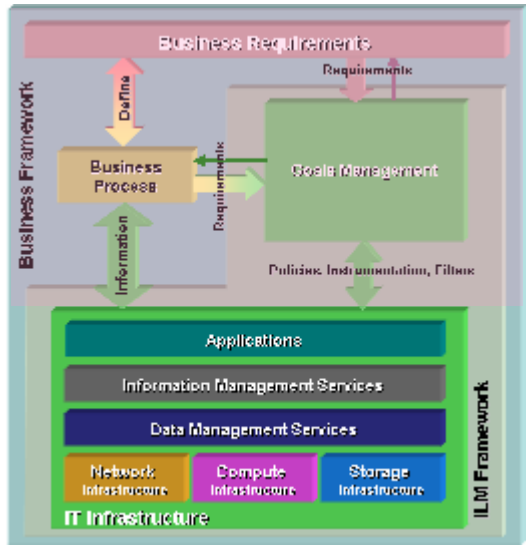
But what about...

- What data/application for which tier?
- When and under what conditions to move/copy data from tier to tier?
- How should data be moved, block, file...?
- What to keep on the compliance archive?
- ...

That's why we need Data Classification

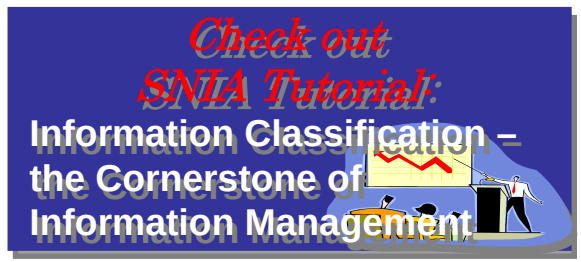
Why Classify Data

- Align data requirements and storage services
- Data Classification enables
  - More efficient storage utilization
  - Consolidated Operational Recovery practices
  - Consolidated Disaster Recovery practices
  - Archiving to meet compliance & other needs
- Build a business case
  - Improved alignment of IT with business priorities
  - Reduced hardware costs
  - Improved utilization & management
  - Reduced footprint
  - Improved environmental resource utilization
- Practiced by many IT thought leaders today
  - In various phases of definition & implementation
  - Cornerstone for ILM



Using Information Classification to Derive Data Classifications

- Process & policies to define service requirements for data
  - Allows IT to create multiple service level offerings
  - Allows Lines Of Business (LOB) to select services based on value of data
  - May use software to enable some of the process
- Represent corporate requirements:
  - Security officer: Secret, confidential, proprietary, …
  - Records Manager: retention time, …
  - Compliance officer (HIPAA, SOX, …): authorization, retention, …
- Represent LOB requirements:
  - Application performance, availability, recoverability, …
  - Staff response time, asset reporting, …



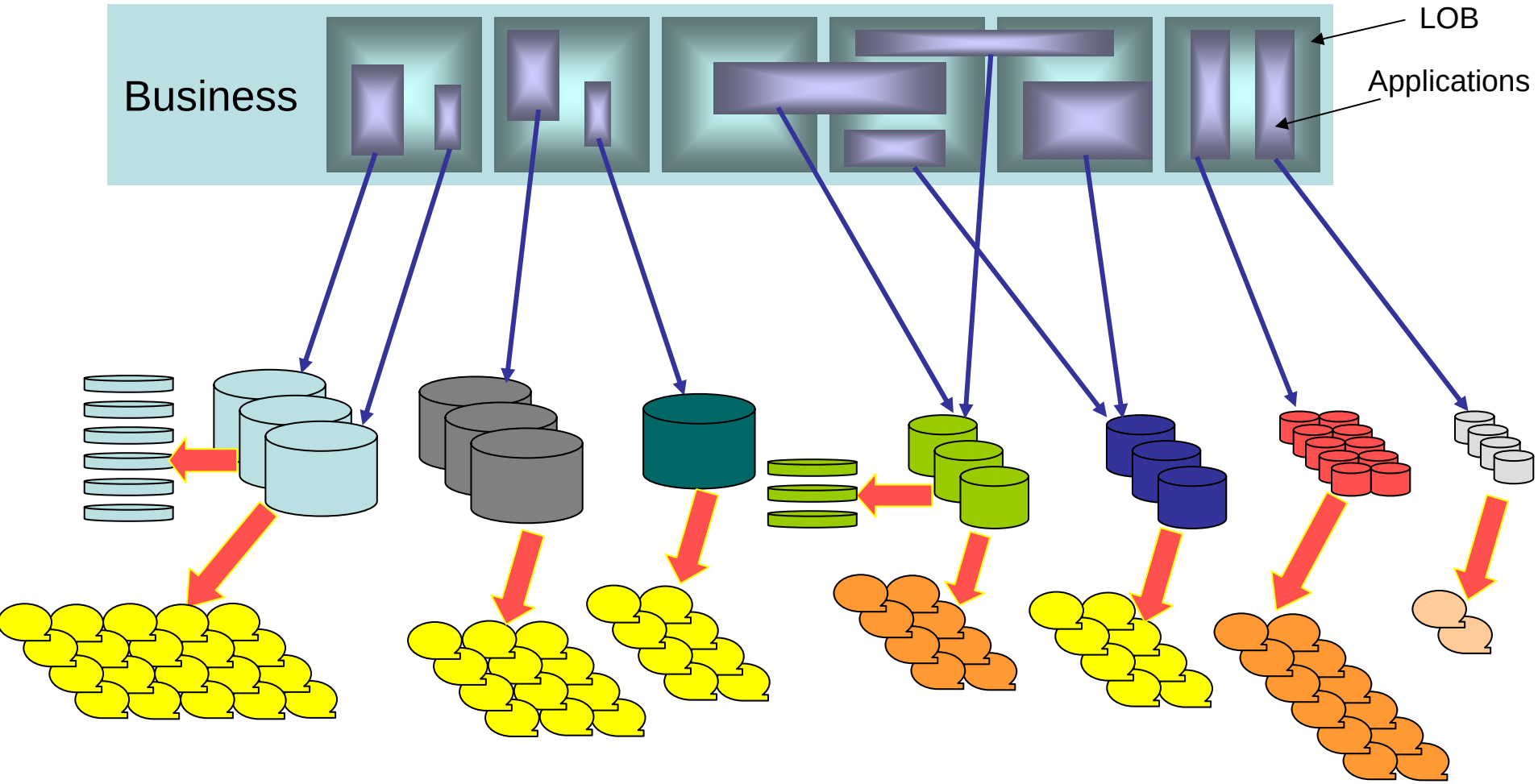
- This process enables the IT Organization to:
  - Match data with appropriate resources based on its lifecycle
  - Map data requirements to appropriate storage devices & services

How is Data Classified?

Data Policies, Service Level Objectives (SLOs) & Data Classification are tightly coupled

- Classify by application
  - All data assigned same classification
  - Simple; good start; a first approximation
  - Net effect: ranking of applications to services
- Classify by metadata
  - Time last accessed, owner, etc
  - E.g., Hierarchical Storage Management (HSM) for a file server
- Classify by content
  - Content-driven alignment of data and SLOs

Pre-Classification: Many data silos



Inefficient, unsustainable, and expensive to manage

Use Resource Classification to Identify Tiers of Service - SLOs

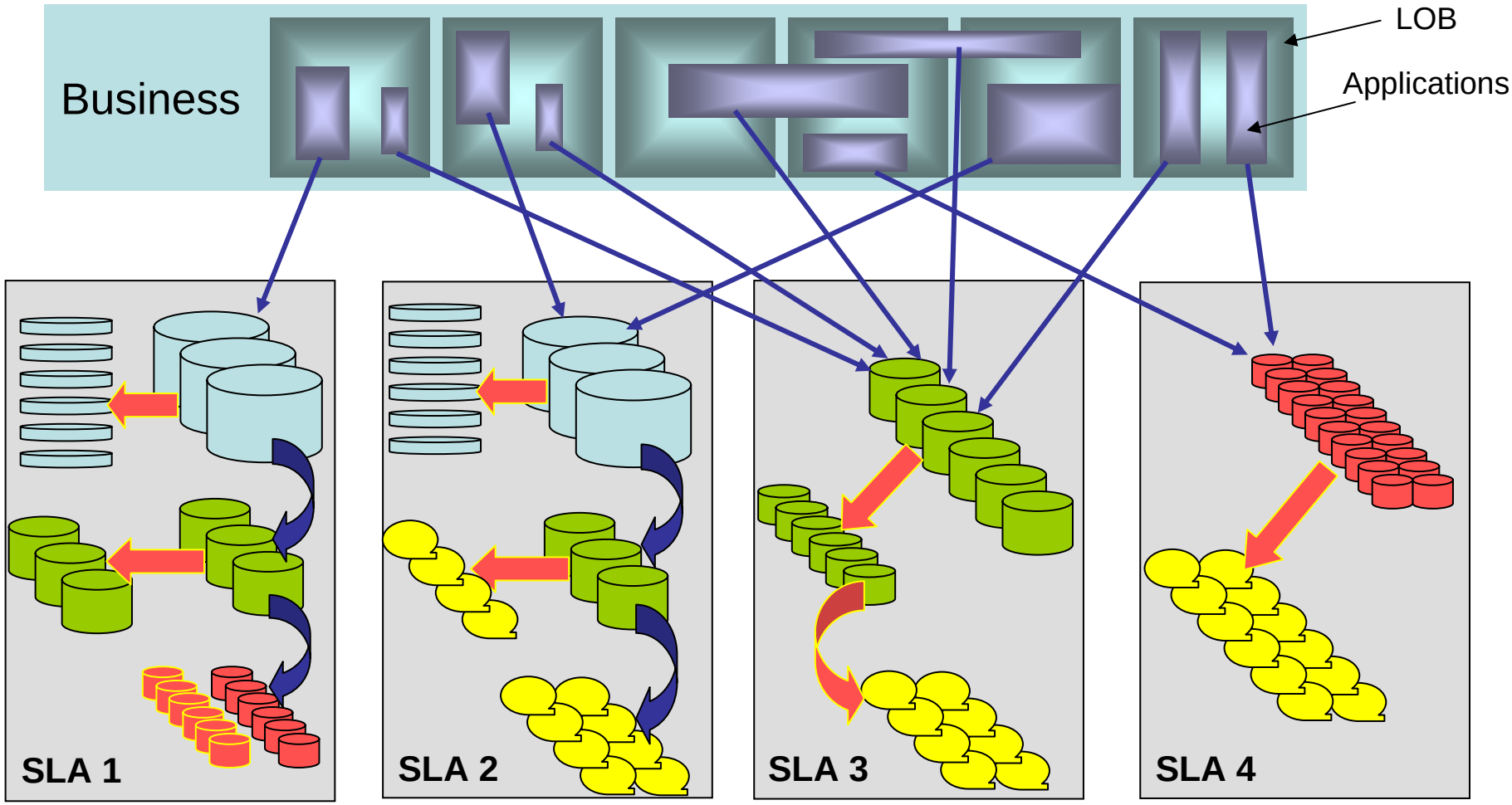
- Define Service Level Objective framework
  - Class of infrastructure for performance & resiliency
  - Availability requirements (99.xxx%)
  - Data Protection & Recovery classes (RTO-RPO mins to days)
  - Archival classes (online, tape, off-site, …)
  - Compliance classes (HIPAA, SOX, …)
  - Confidentiality (in the host, in the network, on storage, at rest…)
  - Others …
- Focus on what level of service is required for data
  - Not on how it is delivered
  - Technology changes, but service levels may not
- Only create SLOs that are important to your business

Acme Service Catalog

SNIA
STORAGE NETWORKING INDUSTRY ASSOCIATION
EDUCATION

	Mission Critical	Business Critical	Business Important	Development
Requirement				
Availability	99.99%	99.9%	99%	97%
Threshold based Automated provisioning	Up to 20% of current file System allocation within 1 business Day	Up to 20% of current file system allocation within 2 business days	Up to 10% or current file system allocation within 4 business days	Up to 20%of file system allocation within 1 business week scratch based allocation
RTO	15 minutes	1 hour	8 hours	24 hours
RPO	1 hour	12 hours	48 hours	96 hours
Restore Requests	100 requests/ week	100 requests / week	50 requests / week	50 requests / week
Backup success rates	97%	95%	90%	90%
Archive Policy	No access in 90 days	No access in 30 days	No access in 90 days	No access in 180 days
Archive access time	Seconds	Seconds	Up to 4 Hours	24-48 hours
Regulations	SOX	HIPAA	None	None
Forecasting	Monthly	Quarterly	Yearly	Yearly
Incident classification and notification	Sev. 1 < 15 minutes Sev. 2 < 30 minutes Sev. 3 < 1 day Sev. 4 < 1 day	Sev. 1 < 25 minutes Sev. 2 < 40 minutes Sev. 3 < 1 day Sev. 4 < 1 day	Sev. 1 < 25 minutes Sev. 2 < 40 minutes Sev. 3 < 1 day Sev. 4 < 1 day	Sev. 1 < 25 minutes Sev. 2 < 40 minutes Sev. 3 < 1 day Sev .4 < 1 day
Asset reporting to support chargeback	Weekly	Bi-weekly	Monthly	Monthly
Cost	$$$$	$$$	$$	$

With Resource Classification: Standard Configurations



Simplified Management, more efficient, scalable

Use Service Criteria Relevant to Your Business (1)

SNIA
Storage Networking Industry Association
EDUCATION

	Mission Critical	Business Critical	Business Important	Development
Requirement				
Availability	**99.99%**	**99.9%**	**99%**	**97%**
Unsched. Downtime	< 1 hour/year	< 10 hours/year	< 100 hours/year	No commitment
Planned Downtime	< 1 hour/month	< 2 hour/month	< 8 hours/month	Intermittent
Advanced Notice	2 weeks	1 week	2 days	None
Dependencies	Requires SLA1 for all	SLA 1 or 2 for storage		

Accessibility	**High performance**	**Standard**	**Economy**	**Offline**
Throughput	Performance-optimized configuration	Default configuration	Cost-optimized configuration	N/A: must import from offline
Compute Style (choice)	Read-intensive, write-intensive, or mixed	Read-intensive, write-intensive, or mixed	Mixed	N/A: must import from offline
Threshold based Automated provisioning	Up to 20% of current file System allocation within 1 business Day	Up to 20% of current file system allocation within 2 business days	Up to 10% of current file system allocation within 4 business days	Up to 10% of current offline allocation within 1 business Day
Data Access (choice)	Block	Block or File	File	N/A
Migrate to Economy	No access in *n* days	No access in *n* days	N/A	N/A
Migrate Offline	No access in *n* days	No access in *n* days	No access in *n* days	N/A
Offline import response time	< 30 minutes 100 requests/ week	< 45 minutes 100 requests / week	2 hours 50 requests / week	N/A

Use Service Criteria Relevant to Your Business (2)

SNIA Storage Networking Industry Association
EDUCATION

Requirement	Mission Critical	Business Critical	Business Important	Development
Site DR	**Immediate**	**Fast**	**Standard**	**Non-critical**
RTO	15 minutes	1 hour	8 hours	24 hours
RPO	1 hour	12 hours	48 hours	96 hours
Recoverable images	Point-in-time only	Two	13 weekly, 7 daily	4 weekly, 7 daily
Retention	N/A – Point in time only	24 hours	Daily: 30 days; Weekly: 52 weeks	Daily: 7 days; Weekly: 13 weeks
Protection success rate	99.999%	97%	90%	90%

Operational Recovery	Immediate	Fast	Standard	Non-critical
RTO	15 minutes	1 hour	8 hours	24 hours
RPO	4 hour	8 hours	24 hours	24 hours
Granularity	Application only	Application or file	Application or file	Application or file
Recoverable images	12	2	One onsite; others offsite	One onsite; others offsite
Retention	48 hours	48 hours	Daily: 30 days; Weekly: 52 weeks	Daily: 7 days; Weekly: 13 weeks
Protection success rate	99%	95%	90%	90%

Use Service Criteria Relevant to Your Business (3)

SNIA
Storage Networking Industry Association
EDUCATION

Requirement	Mission Critical	Business Critical	Business Important	Development
Helpdesk	**Platinum**	**Gold**	**Silver**	**Bronze**
Restore from backup response time	< 30 minutes 100 requests/ week	< 45 minutes 100 requests / week	2 hours 50 requests / week	4 hours 50 requests / week
Forecasting	Monthly	Quarterly	Yearly	Yearly
Incident classification and notification	Sev. 1 < 15 minutes Sev. 2 < 30 minutes Sev. 3 < 1 day Sev. 4 < 1 day	Sev. 1 < 25 minutes Sev. 2 < 40 minutes Sev. 3 < 1 day Sev. 4 < 1 day	Sev. 1 < 25 minutes Sev. 2 < 40 minutes Sev. 3 < 1 day Sev. 4 < 1 day	Sev. 1 < 25 minutes Sev. 2 < 40 minutes Sev. 3 < 1 day Sev .4 < 1 day
Asset reporting to support chargeback	Weekly	Bi-weekly	Monthly	Monthly

Security	**Public**	**Internal**	**Confidential**	**Secret**
Classification Level				

Compliance	**HIPAA**	**SEC 17a4**	**Sarbanes-Oxley**	**None**
Compliance categories				

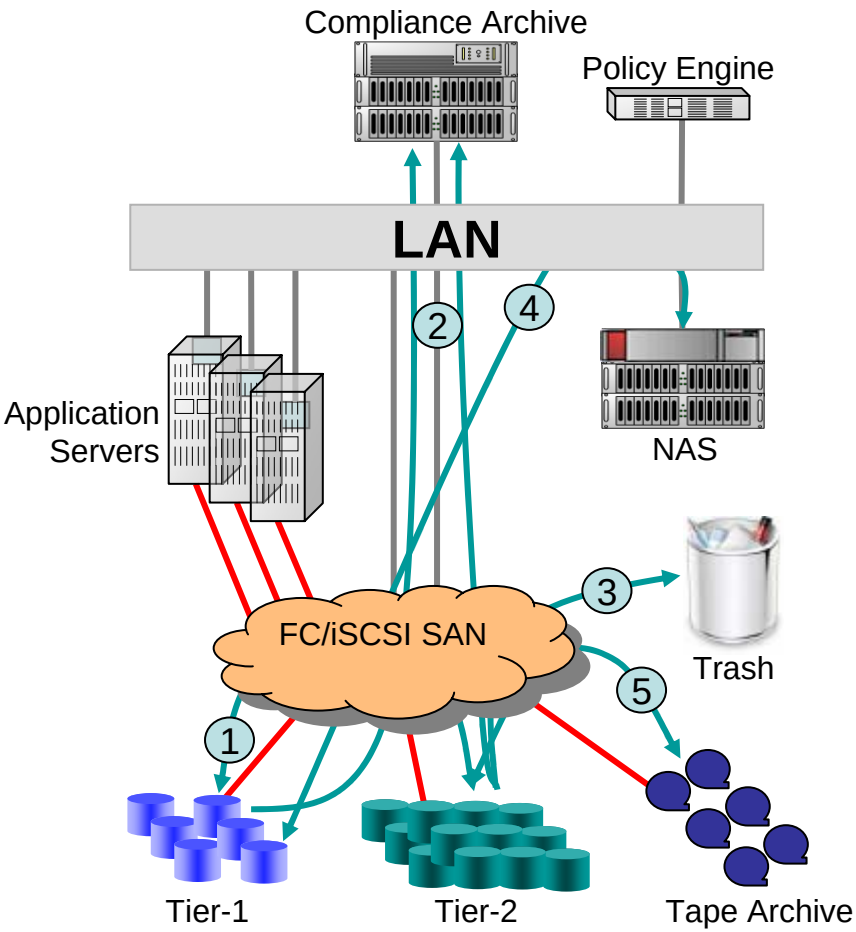
Implementation Considerations

Three primary data types

- Unstructured – Files in file systems
- Semi-structured – Email messages in Email systems
- Structured – Database records

Each type presents different challenges

Unstructured Data (Files)



It's more than traditional HSM.

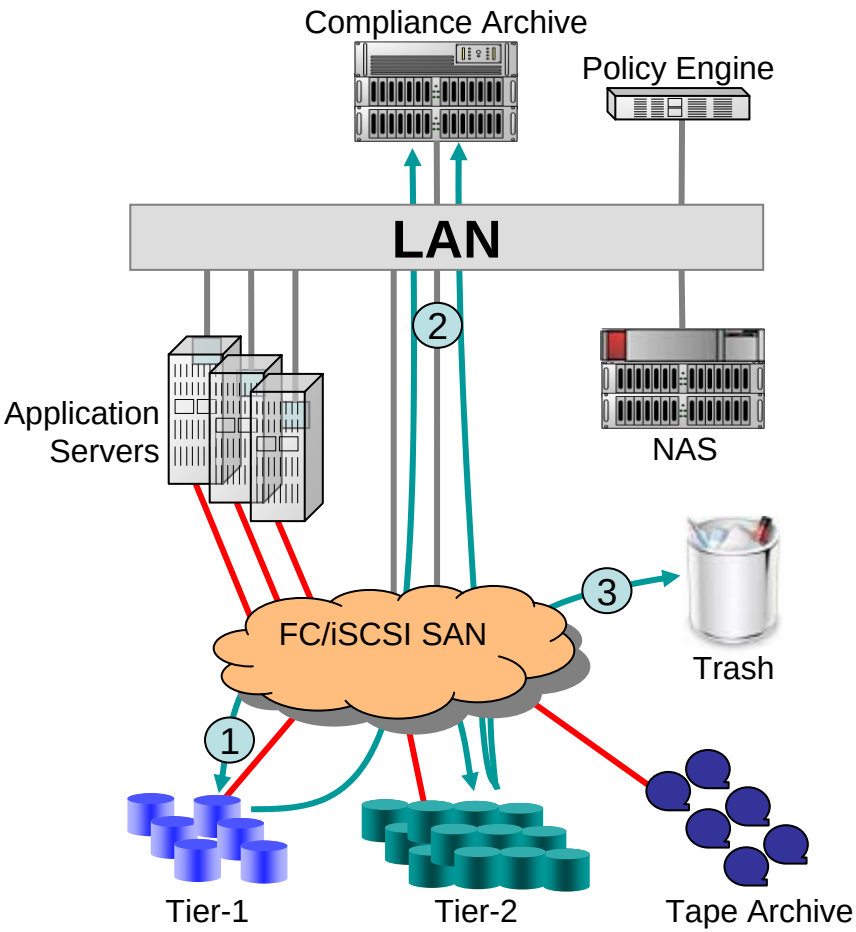
Selection criteria include

- Age/owner type metadata
- File content

Movement can be

1. Tier – tier (bi-directional)
2. To compliance archive
3. To "trash"
4. From SAN to NAS & vice-versa
5. To tape archive for offsite vaulting

Semi-Structured Data (Email)



Selection criteria include

- To/From type metadata
- Age & access metadata
- Content

Actions might include

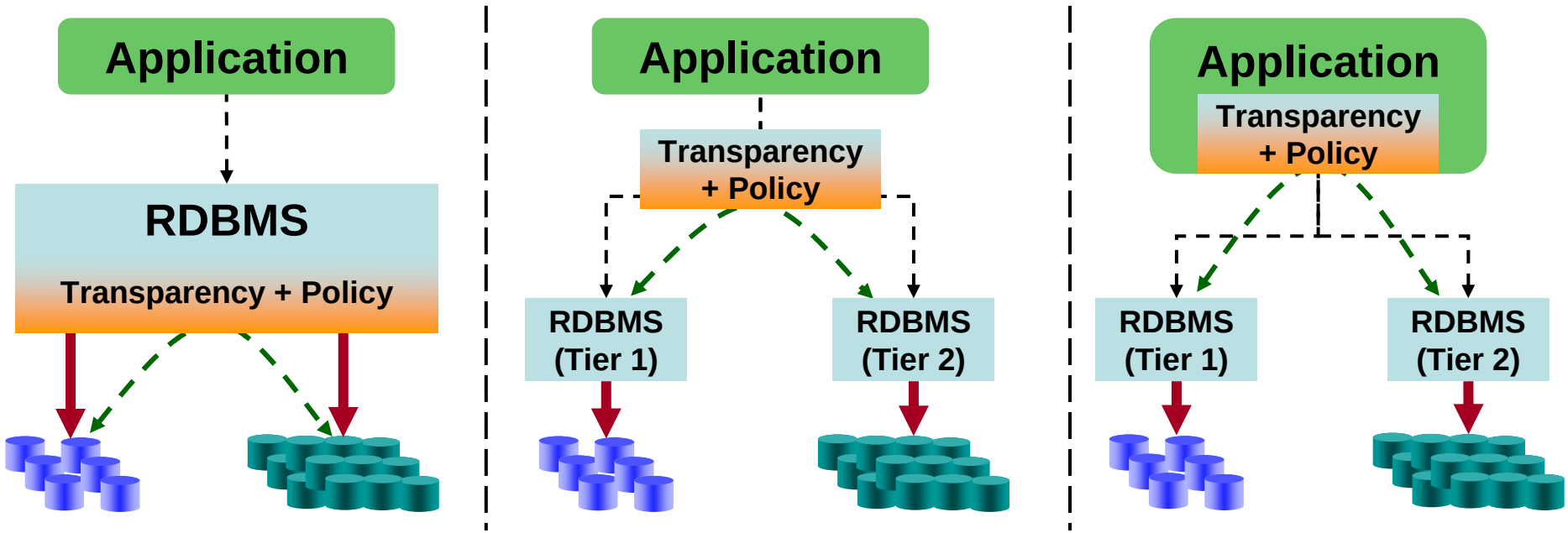
1. Create single instance store & attachment migration
2. Send email to compliance archive
3. Get rid of trash

Structured Data (RDBMS)

Represents significant challenges because:

- Content & organization is a function of the application
- Each application is different
- Must maintain transactional integrity

Three basic approaches….



A Look To The Future...

New data classification & management systems will be:

- Comprehensive
  - Able to catalog & manage all types of data
  - Enable “business view” of data management requirements
  - Able to scale to enterprise wide deployments
- Application integrated and business-driven
  - Files (unstructured data) - beyond traditional HSM
  - Databases (structured data) - framework for higher level application integration
  - Email & Messaging (semi-structured data) - variety of management capabilities for email and attachments
- Policy-based
  - Support many policies to address common data management problems
  - Support wide variety of data movement options at file & block level
  - Data movement triggered by time, content, event, …

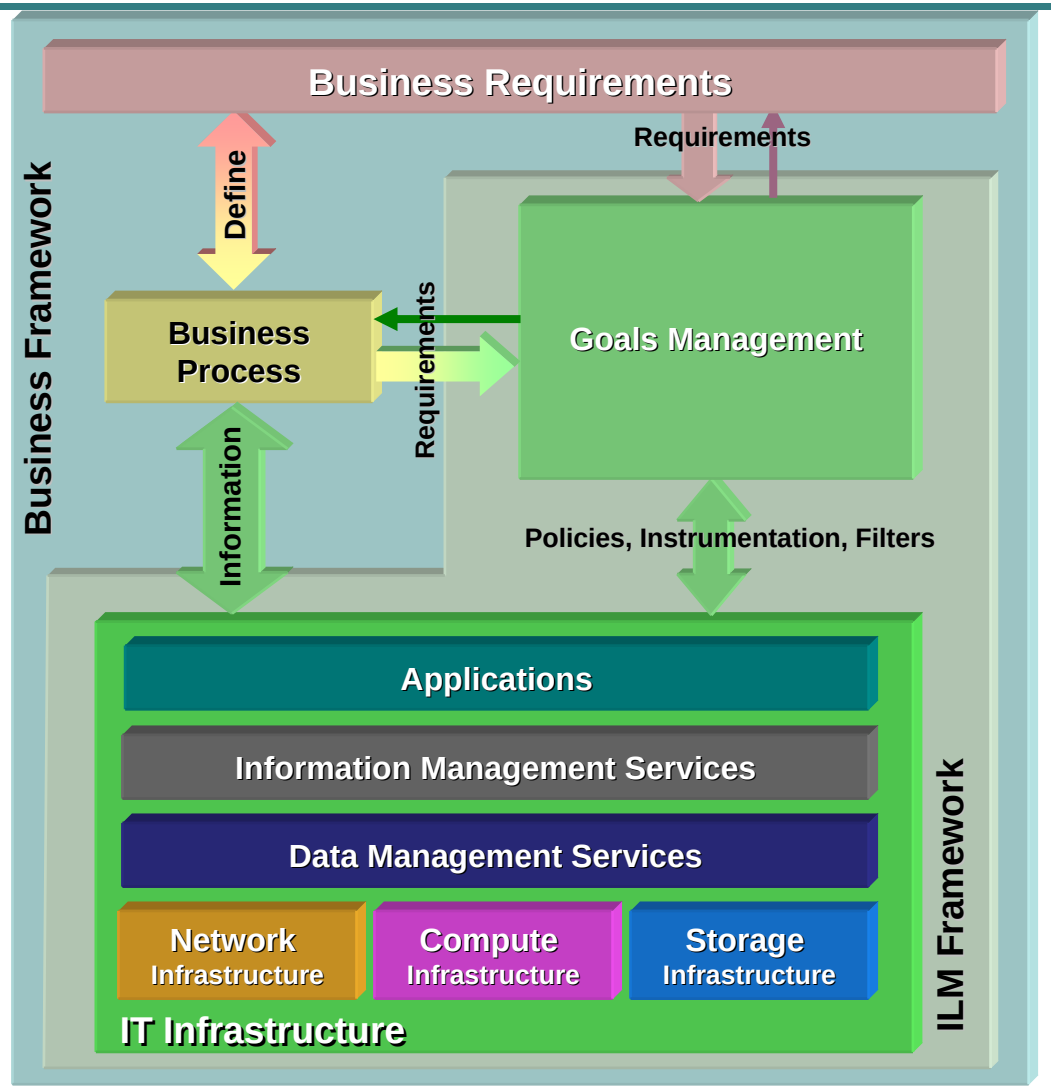
DMF's ILM Framework for the Datacenter

SNIA
Storage Networking Industry Association
EDUCATION

For more information on SNIA's Data Management Forum (DMF) visit the DMF website at

http://www.snia-dmf.org

Also see the *Data Management Solutions Center* in the Show Exhibit area



Please send any comments on this tutorial to SNIA: track-datamgmt@snia.org

Many thanks to the following individuals for their contributions to this tutorial:

Nik Simpson
Dan Tanner
Edgar St.Pierre
Alex Adamopoulos