BJ Klingenberg – Senior Technical Staff Member
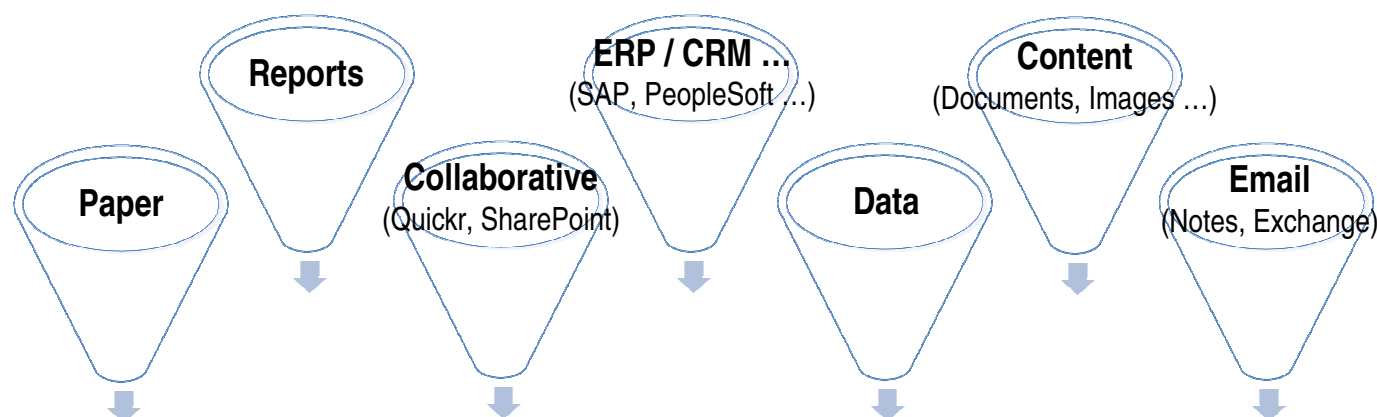
21 October 2009

IBM

IBM Information Archive

IBM Smart Archive Strategy

Reports

Paper

ERP / CRM ...
(SAP, PeopleSoft ...)

Collaborative
(Quickr, SharePoint)

Content
(Documents, Images ...)

Data

Email
(Notes, Exchange)

Optimized and Unified Assessment, Collection and Classification
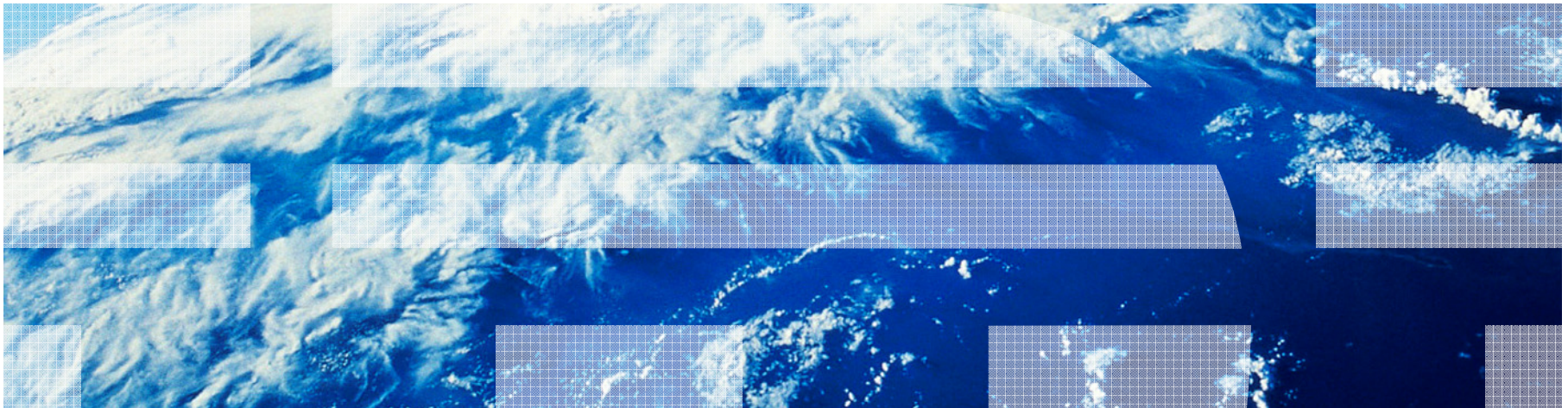
Value Added Services
- Optimization Services
- System Services
- Managed Services
- Reference Architecture
- Information Governance

Cloud Ready Archive Storage with Optional ECM

Flexible and Secure Infrastructure with Unified Retention and Protection

On Premise
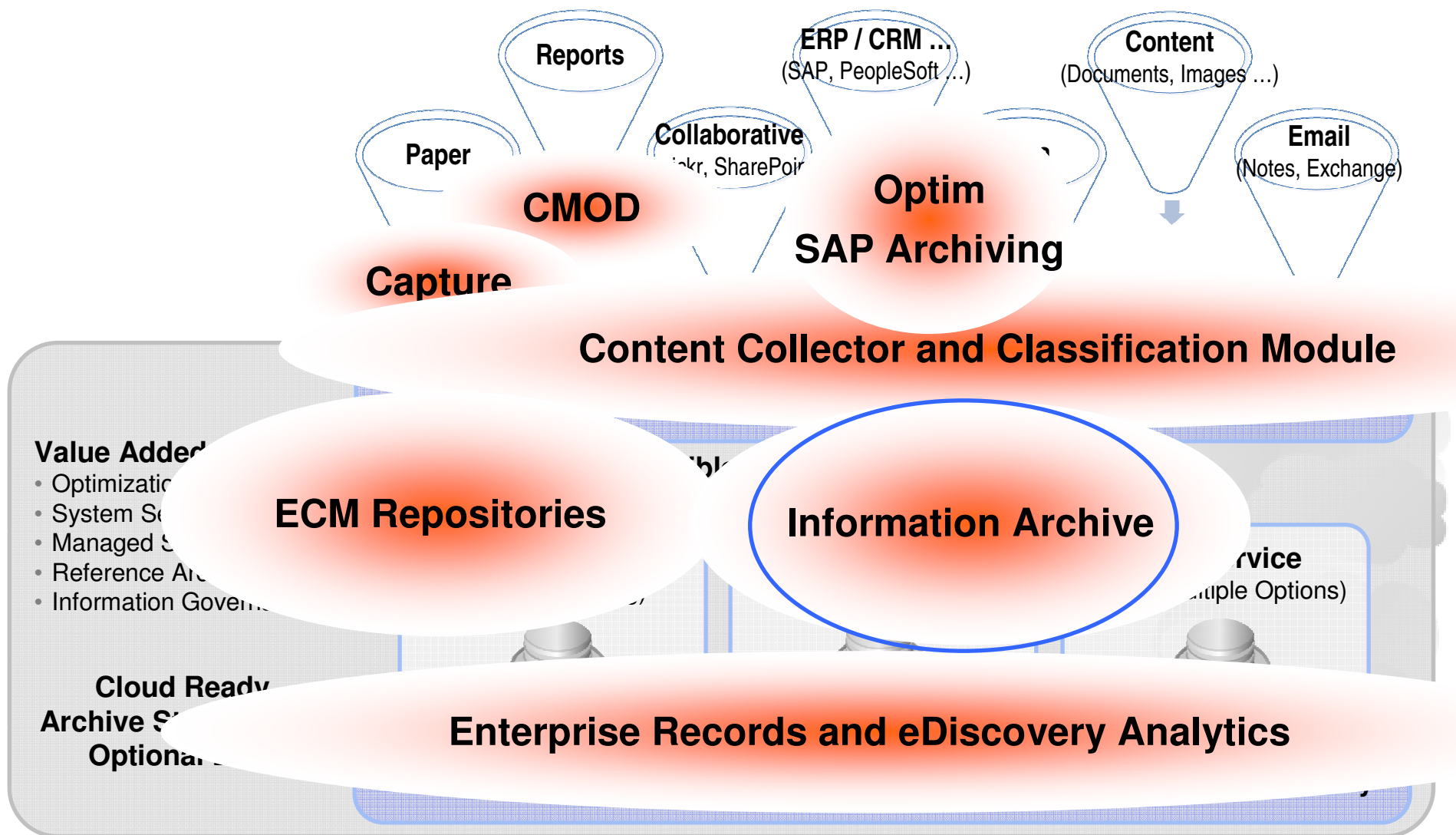(Custom Config)

Appliance
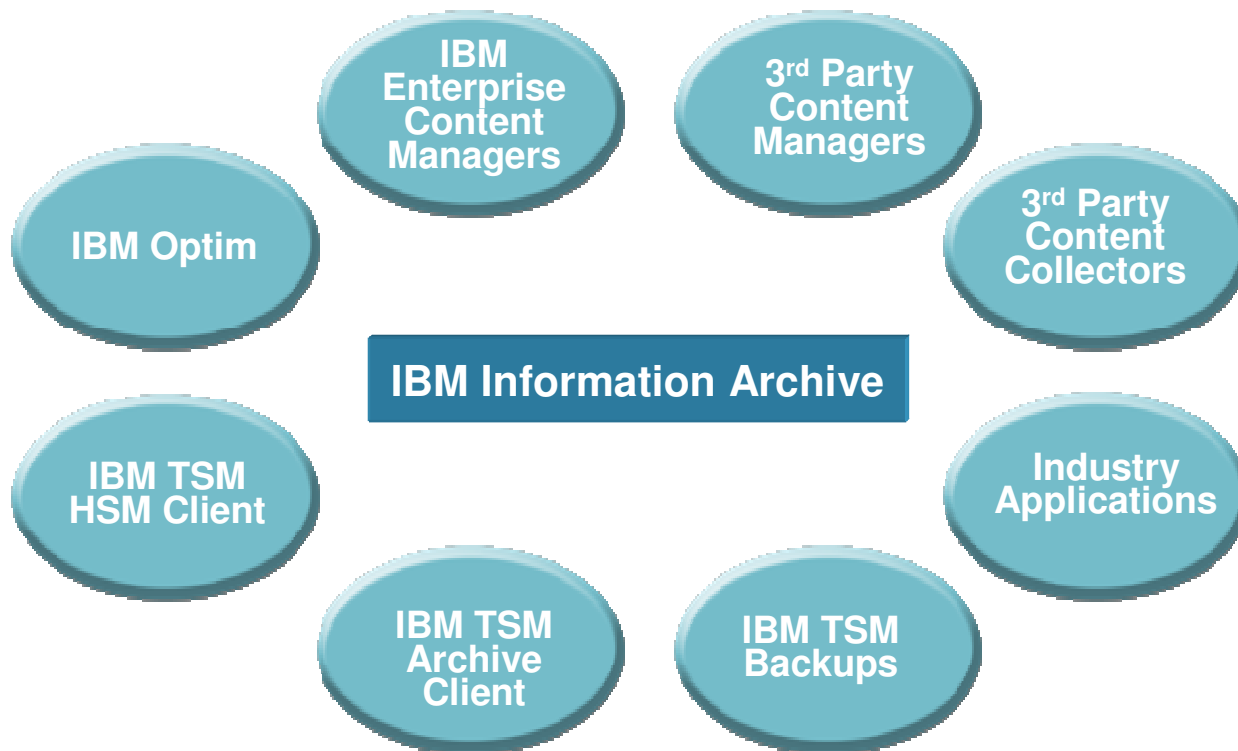(Pre-Config)

As A Service
(SaaS, Multiple Options)

Integrated Compliance, Records Management, Analytics and eDiscovery

IBM

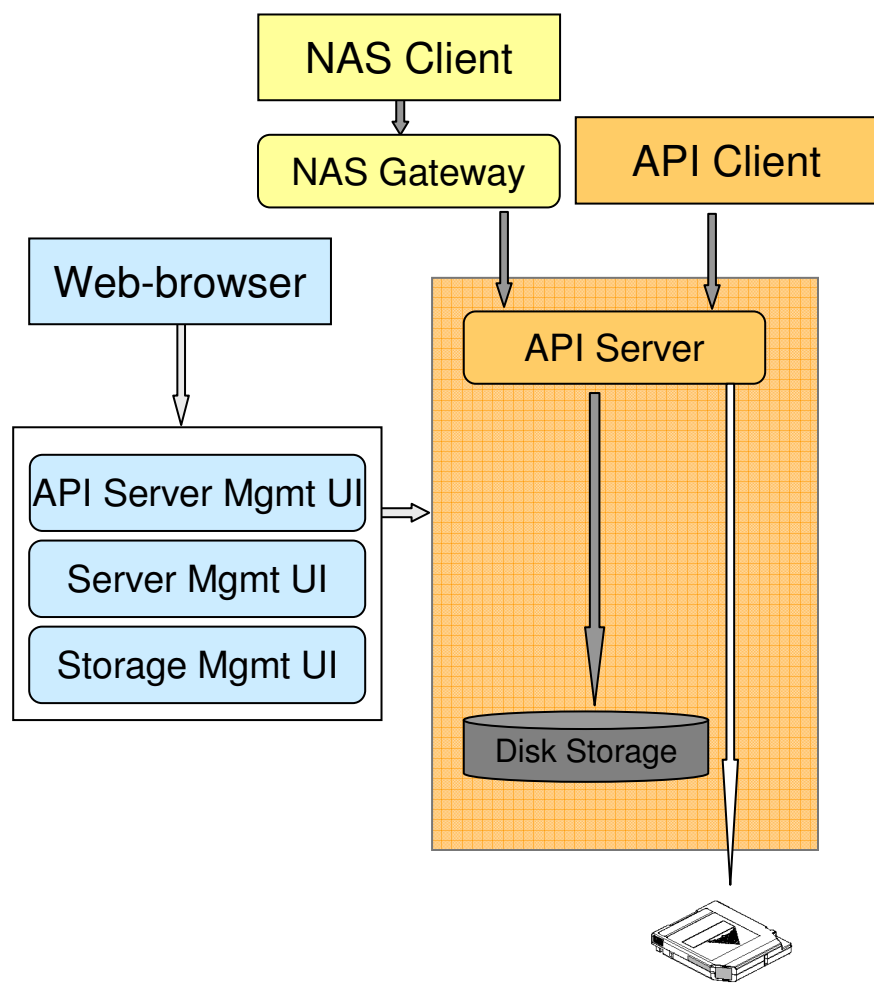# Existing Integrated IBM Archiving Solutions (Today)

Reports

ERP / CRM ...
(SAP, PeopleSoft ...)

Content
(Documents, Images ...)

Paper

Collaborative
...kr, SharePoi...

Email
(Notes, Exchange)

CMOD

**Optim
SAP Archiving**

Capture

Content Collector and Classification Module

Value Added...
- Optimizatio...
- System S...
- Managed S...
- Reference Ar...
- Information Gove...

ECM Repositories

Information Archive

...rvice
...ltiple Options)

Cloud Ready
Archive S...
Optiona...

Enterprise Records and eDiscovery Analytics

3   CMOD = Content Manager On Demand

IBM Information Archive (IA) Introduction

IBM Enterprise Content Managers

3rd Party Content Managers

IBM Optim

3rd Party Content Collectors

IBM Information Archive

IBM TSM HSM Client

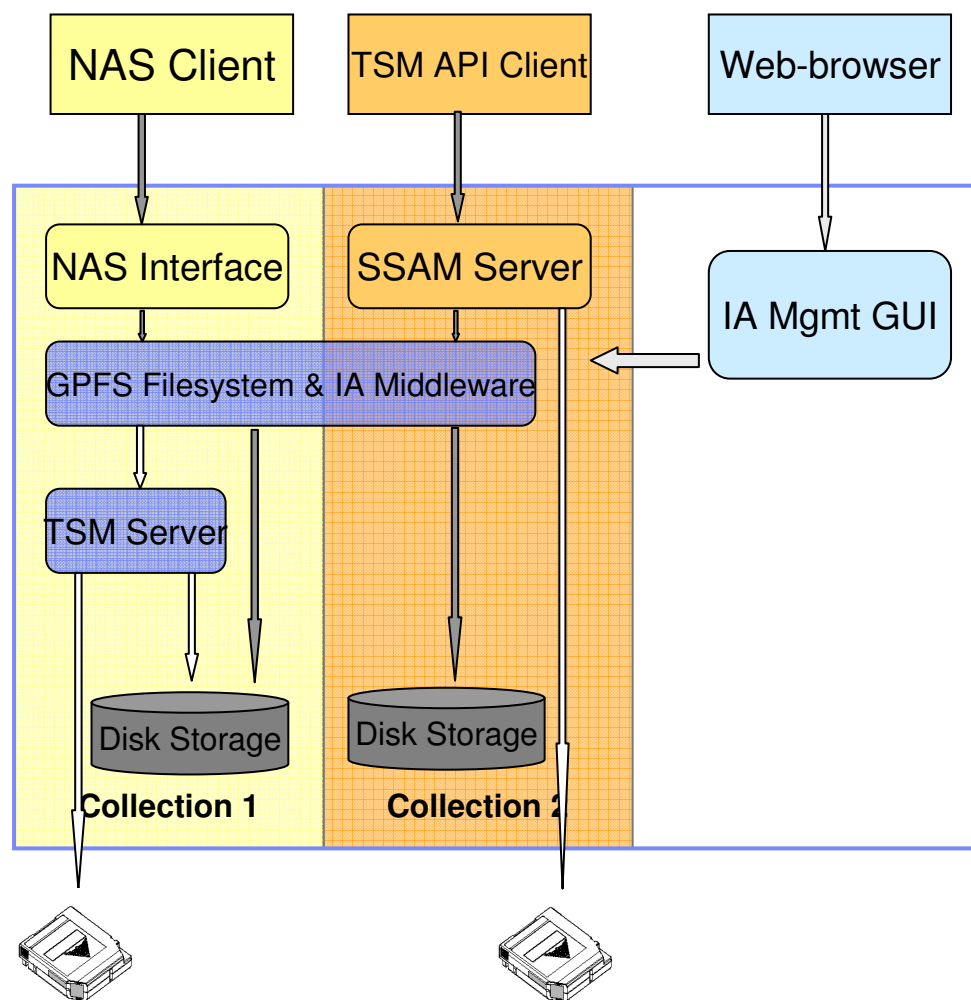Industry Applications

IBM TSM Archive Client

IBM TSM Backups

- Integrated appliance
- Replaces DR550
- More interfaces
- Better scalability and performance
- Accepts information from multiple sources
- Developing ISV ecosystem
- Close ties with IBM ECM

Information Archive Next Generation Architecture

Previous Generation Architecture

NAS Client

NAS Gateway

API Client

Web-browser

API Server

API Server Mgmt UI

Server Mgmt UI

Storage Mgmt UI

Disk Storage

Information Archive Architecture

NAS Client

TSM API Client

Web-browser

NAS Interface

SSAM Server

IA Mgmt GUI

GPFS Filesystem & IA Middleware

TSM Server

Disk Storage

Disk Storage

Collection 1

Collection 2

Information Archive Characteristics

Universal

- Common platform for archive of multiple types of data
- Variety of data interfaces (NAS, TSM)

Scalable

- Scale-out of processing and storage
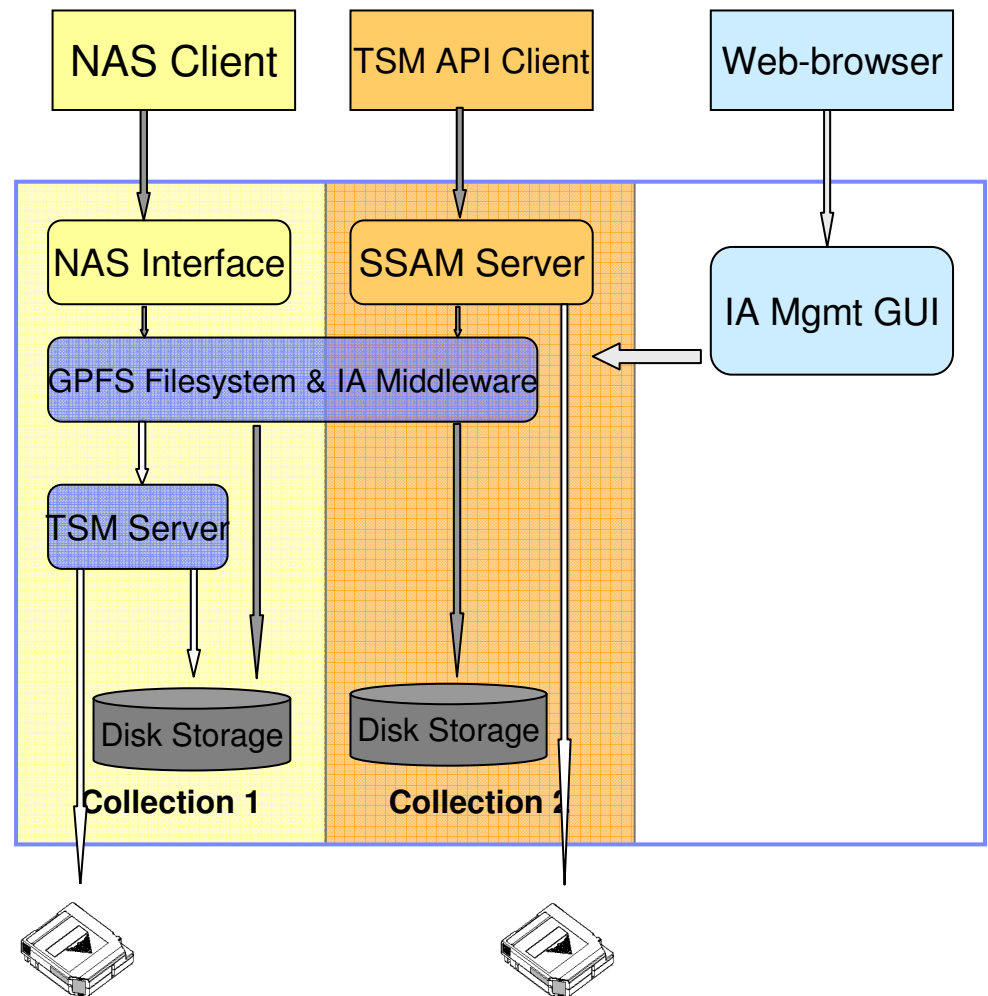- Tiered storage, including external tape

Adaptable

- Compliant and non-compliant archives
- Pluggable architecture for future data-specific function

Secure

- Fully protected, lockable compliant store
- No root access in full compliance mode through Enhanced Tamper Protection

Information Archive Architecture

NAS Client

TSM API Client

Web-browser

NAS Interface

SSAM Server

IA Mgmt GUI

GPFS Filesystem & IA Middleware

TSM Server

Disk Storage

Disk Storage

Collection 1

Collection 2

Information Archive Collections

- **Collection = Set of archived documents managed under the same policy domain**
- **Types of collections: NAS, SSAM**
- **Multiple collections per IA appliance (up to 3 in R1) with separation of data**

Interface Policy

- Interface protocol (NAS or TSM API)
- Object commit method (NAS only)
  - XML file, timeout, NetApp Snaplock™

Retention Policy

- Controlled internally or by external application
- Time-based (internal) and Event-based (external) retention
- Automatic (internal) or manual (external) deletion after expiration

Storage Policy

- Disk replication
- Tape storage tier
- Encryption (tape only)
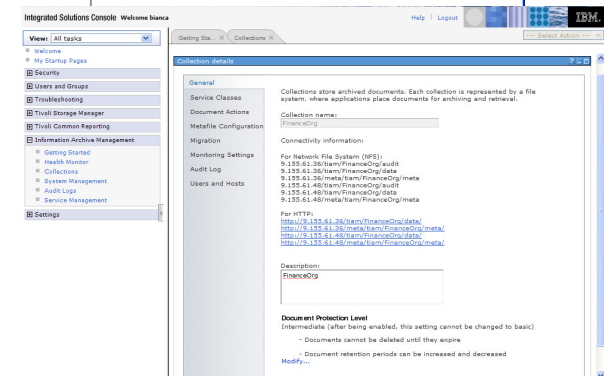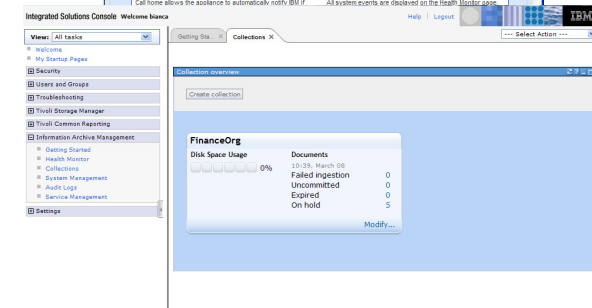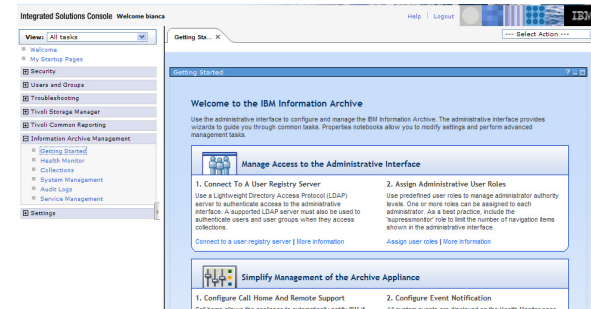- Deduplication
- Shredding (SSAM only in R1)

Compliance Policy

Mode	Delete before expire?	Retention Period	
		Shorten?	Lengthen?
Basic	Yes	Yes	Yes
Intermediate	No	Yes	Yes
Maximum*	No	No	Yes

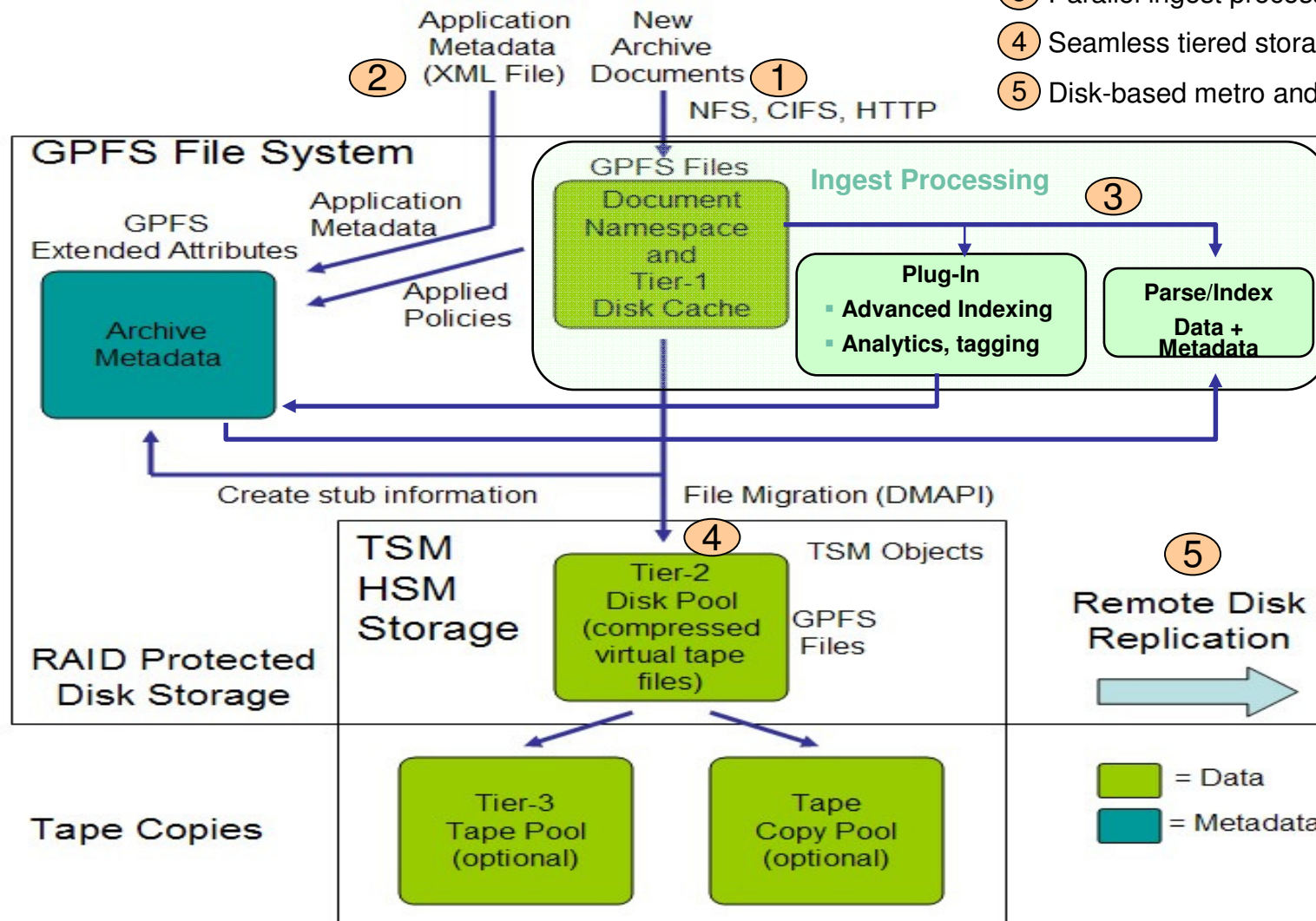*SSAM Collections are always Maximum compliance

Information Archive Integrated, Web-based Interface

- Initial configuration is integrated and guided
  - Start archiving data in 1 day or less

- Consolidated and integrated management in a single administrative interface
  - Security, administration, monitoring, troubleshooting, serviceability
  - Manage multiple collections from a single interface
  - Role-based administrative security

- Simple, efficient, consistent administrative experience
  - Wizards guide user on the creation of objects
  - Overviews provide high level situational awareness of system health
  - Consolidation of information for efficient administration

- Task-oriented, with drill down where needed for deeper configuration and trouble shooting

Information Archive NAS Collection Internal Architecture

1. Scale-out NAS interface with global namespace
2. XML-based application metadata (optional)
3. Parallel ingest processing, including indexing
4. Seamless tiered storage through TSM HSM
5. Disk-based metro and global mirroring

Application Metadata (XML File) 2

New Archive Documents 1

NFS, CIFS, HTTP

GPFS File System

GPFS Extended Attributes

Application Metadata

Applied Policies

Archive Metadata

GPFS Files

Ingest Processing

Document Namespace and Tier-1 Disk Cache

3

Plug-In
- Advanced Indexing
- Analytics, tagging

Parse/Index
Data + Metadata

Create stub information

File Migration (DMAPI)

TSM HSM Storage

RAID Protected Disk Storage

4

Tier-2 Disk Pool (compressed virtual tape files)

TSM Objects

GPFS Files

5

Remote Disk Replication

Tape Copies

Tier-3 Tape Pool (optional)

Tape Copy Pool (optional)

= Data
= Metadata

Document Archival

- TSM API or TSM Archive Client
  - Managed by SSAM server
  - SSAM retention policies and functions
  - 2 storage tiers
    - Disk – compression/deduplication
    - Tape (optional)

- NAS protocols
  - NFS, HTTP READ
    - Support time and event based retention and retention holds via:
      - Auto-protect with retention policies
      - NetApp SnapLock™ compatibility, or
      - Metadata files to control retention
  - Backend TSM server allows configuration with disk or disk + tape
  - 3 storage tiers
    - Tier 1 archive disk – high speed data transfer and ingestion (retention, …)
    - Tier 2 archive disk – compression/deduplication
    - Tape (optional)

Create collection

Welcome	General
General	The collection name provides a unique identifier for the collection. The name cannot be m... collection's file system is derived from the collection name.
Summary	

*Collection name:
DR550

Description:
DR550 Collection 1

File archive collections transfer files using standard network protocols such as Network File System (NFS). with System Storage DR550 clients.

○ Create a file archive collection
◉ Create a System Storage Archive Manager collection

< Back | Next > | Finish | Cancel

Security and Compliance Characteristics

- Role based security:

    - Security, Systems and Archive Administrators

    - Archive User, Service Engineer and Auditor Roles

- Audit logs provide compliant audit trail.

- Physical security: locking cabinet

- To achieve maximum compliance protection with IBM IA customers enable the *Enhanced Tamper Protection* feature
    - Removes root login capability from the IBM IA Cluster
    - Neither customer nor IBM has root login authority
    - Once enabled, cannot be disabled
    - Expected admin and support operations pre-programmed to remove need for root access
    - Best practice to enable during installation
    - We have a procedure for an unforeseen emergency requiring root access by delivering a signed - time-bounded patch to the customer.

Meta Files

- Use existing standards (NFS, XML) to extend the user's ability to manage files in the archive via NAS protocols
    - NetApp SnapLock^{TM} compatibility only provides time-based retention
    - The addition of metafiles enables support of time *and* event based retention, policy based retention and retention hold/release

- Customer/application can define user fields
    - Binding and non-binding (binding cannot be updated after document is committed)
    - User defined fields will be indexed in the future

- Event based retention, holds, etc. can be signaled via EVENT fields in meta files.

- User/application can use standard file protocols to update meta files.

Meta Files

- /IIA/col1/data
  - /IIA/col1/data/file1.doc

  Now is the time for all good men to come to the aid of their country.

- /IIA/col1/meta
  - /IIA/col1/meta/file1.doc

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<fields>
<_SYSTEM_md5Checksum>
da1e100dc9e7bebb810985e37875de38
</_SYSTEM_md5Checksum>

<USER_confidential> Yes
</USER_confidential>

<_EVENT_setRetention_> 20101231
</_EVENT_setRetention_>
</fields>
```
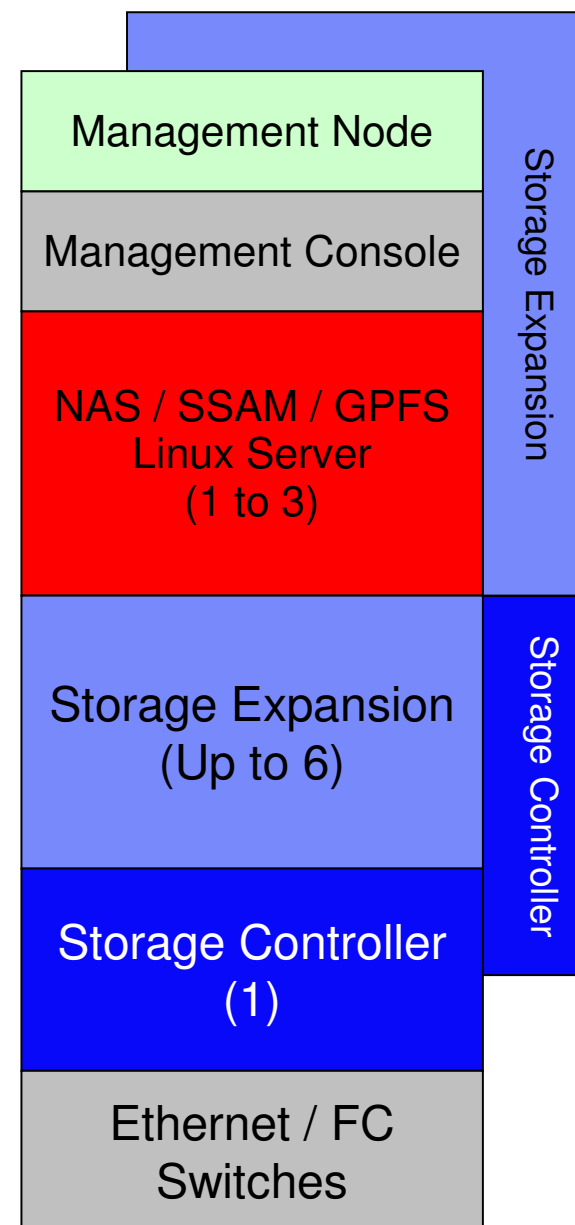
IBM Information Archive Key Feature Summary

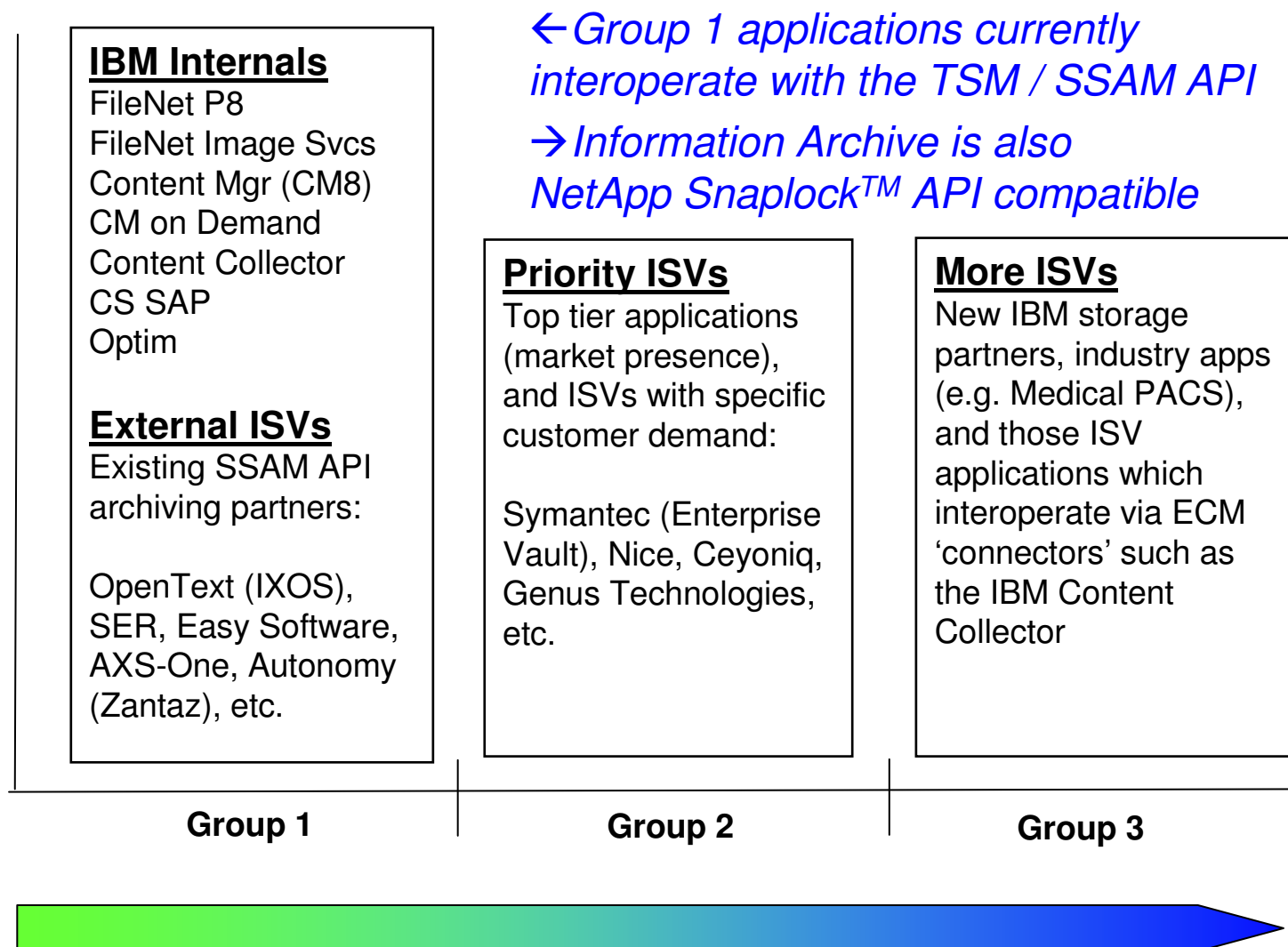| Feature	IBM Information Archive
Ease of use	Appears as a file-server, simple drag & drop files to IA.
Installation & Implementation	Information Archive is physically installed by a CE. Quick configuration via wizards. Dynamically add storage.
User Interface	Single, task oriented GUI to manage the entire solution (+ CLI)
Application Interfaces	NFS, CIFS*, HTTP*, FTP*, SSAM – all within IA
Scalability	Supports billions of objects – 1 billion per collection, up to 3 collections per IA
Retention Policies	Accepts retention policies from applications or creates them (Event-based or time-driven)
Compliance / Security	Each collection can be configured with different protection levels (basic, intermediate, maximum). New patent-pending tamper proof technology increases protection.
Index / Search	Full text indexing and search (both data and metadata) – R1.2.0

* Available at 2nd release

Inside IBM Information Archive (R1)

- Hardware
  - Dual-quadcore clustered Linux servers
  - Integrated (closed) appliance
  - Internal IP and SAN networks
  - SATA storage
    - RAID 6, 209 TB max usable
  - PBs of external tape
  - HW-based metro and global mirroring
  - 3 collections per appliance
- Software
  - System Storage Archive Manager
  - Tivoli Storage Manager – HSM
  - Scale Out File System – GPFS
  - Management Console
- Security
  - Role-based administration
  - LDAP or Active Directory for user management
  - No root access in compliant mode

Management Node

Management Console

NAS / SSAM / GPFS Linux Server (1 to 3)

Storage Expansion (Up to 6)

Storage Controller (1)

Ethernet / FC Switches

Storage Expansion

Storage Controller

IBM Information Archive ISV Roadmap

IBM Internals
FileNet P8
FileNet Image Svcs
Content Mgr (CM8)
CM on Demand
Content Collector
CS SAP
Optim

External ISVs
Existing SSAM API
archiving partners:

OpenText (IXOS),
SER, Easy Software,
AXS-One, Autonomy
(Zantaz), etc.

←Group 1 applications currently interoperate with the TSM / SSAM API

→Information Archive is also NetApp Snaplock™ API compatible

Priority ISVs
Top tier applications
(market presence),
and ISVs with specific
customer demand:

Symantec (Enterprise
Vault), Nice, Ceyoniq,
Genus Technologies,
etc.

More ISVs
New IBM storage
partners, industry apps
(e.g. Medical PACS),
and those ISV
applications which
interoperate via ECM
'connectors' such as
the IBM Content
Collector

**Group 1**   **Group 2**   **Group 3**

IBM Information Archive Summary

- Provides compliant archive storage as a key part of IBM's Smart Archive Strategy

- IBM's strategic information retention solution
    - **For all types of data (structured, unstructured)**
    - **For all business needs (compliance, long-term)**
    - **Affordable for mid-market and large enterprise**
    - **Universal, scalable, adaptable, secure**
    - **Replaces IBM System Storage DR550**

Smart Business System, Cloud-ready

Thank You

IBM
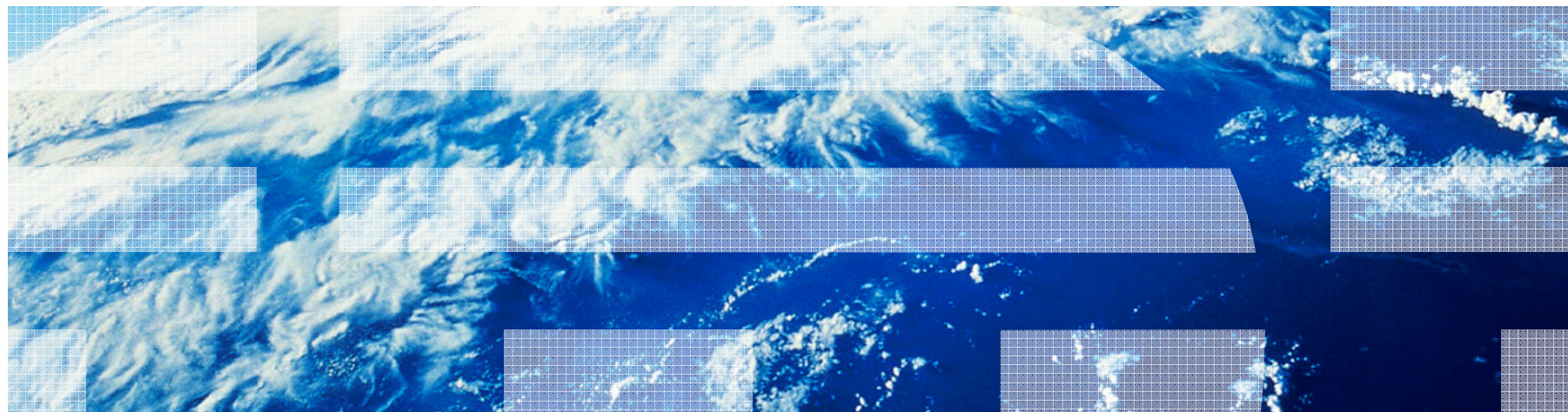
Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.