

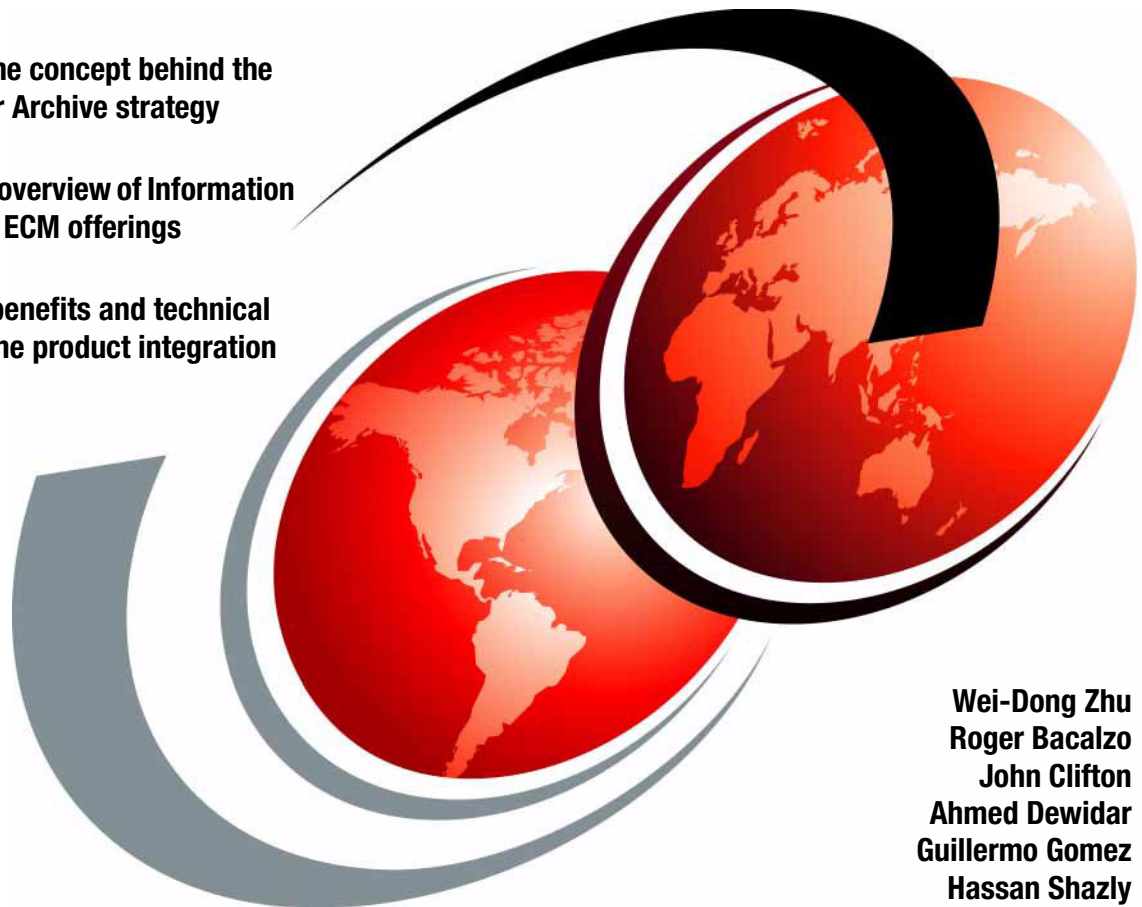


IBM Enterprise Content Management and IBM Information Archive Providing the Complete Solution

Discusses the concept behind the
IBM Smarter Archive strategy

Provides an overview of Information
Archive and ECM offerings

Details the benefits and technical
aspects of the product integration



Wei-Dong Zhu
Roger Bacalzo
John Clifton
Ahmed Dewidar
Guillermo Gomez
Hassan Shazly



International Technical Support Organization

**IBM Enterprise Content Management and
Information Archive: Providing the Complete
Solution**

January 2011

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (January 2011)

This edition applies to Version 5, Release 0 of IBM FileNet P8 Platform products, and Version 1, Release 2, of IBM Information Archive (program number 5608-IAF).

This document created or updated on January 19, 2011.

© Copyright International Business Machines Corporation 2011. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Notices	vii
Trademarks	viii
Preface	ix
The team who wrote this paper	x
Now you can become a published author, too!	xi
Comments welcome	xii
Stay connected to IBM Redbooks	xii
Chapter 1. End-to-end Smart Archive solutions with the IBM Smart Archive strategy	1
1.1 Executive summary	2
1.2 The history of archiving	2
1.2.1 Reasons behind archiving	2
1.2.2 Archiving technologies	3
1.3 Archiving compared to backup	4
1.3.1 Backup	4
1.3.2 Archiving	4
1.4 The IBM Smart Archive strategy	5
1.5 IBM solutions for the Smart Archive strategy	8
1.6 The Information Archive device and ECM	9
1.6.1 End-to-end Smart Archive solutions	10
1.7 Summary	11
Chapter 2. Introducing the IBM Information Archive device	13
2.1 Introduction	14
2.2 Archiving and retrieving information	15
2.3 Collections	15
2.4 Information protection	16
2.5 Retention policies	17
2.6 Data encryption	18
2.7 Disaster recovery protection	18
2.8 Summary	19
Chapter 3. Introducing IBM Enterprise Content Management	21
3.1 Introduction	22
3.2 Essential ECM	22
3.2.1 IBM FileNet Content Manager	23
3.2.2 IBM Content Manager	24

3.2.3 IBM Content Manager OnDemand	24
3.2.4 IBM Datacap	25
3.2.5 IBM FileNet Capture	25
3.3 Advanced case management	26
3.4 Information life cycle governance	27
3.4.1 Content Analytics for Assessment	27
3.4.2 Content Collection and Archiving	28
3.4.3 Advanced Classification	29
3.4.4 Records Management	30
3.4.5 eDiscovery Search and Analytics	30
3.5 Trusted Content Analytics	31
3.6 Summary	32
Chapter 4. Integrating IBM FileNet P8 with the Information Archive device	37
4.1 Introduction	38
4.2 IBM FileNet P8 Platform overview	38
4.3 Integration architecture	39
4.3.1 Deployment model	40
4.3.2 Data model	42
4.3.3 Retention	44
4.3.4 Hierarchical storage	47
4.4 Integration implementation example	48
4.4.1 Configure the Information Archive device	48
4.4.2 Install the Tivoli Storage Manager client	52
4.4.3 Configure the Content Engine	54
4.4.4 Test the integration	63
4.5 Hints and tips	68
4.5.1 Best practices	68
4.5.2 FileNet P8 configuration troubleshooting	69
4.5.3 Logs	69
4.6 Summary	70
Chapter 5. Integrating IBM Content Manager with the Information Archive device	71
5.1 Introduction	72
5.2 IBM Content Manager overview	72
5.2.1 IBM Content Manager architecture	73
5.2.2 Benefits of using an Information Archive device with IBM Content Manager	76
5.3 Integration architecture	78
5.3.1 Deployment model	78
5.3.2 Data model	80

5.3.3	Retention	81
5.3.4	Hierarchical storage	85
5.4	Integration implementation	86
5.4.1	Configure Tivoli Storage Manager	86
5.4.2	Configure the Resource Manager	90
5.4.3	Test the integration	98
5.5	Hints and tips.	101
5.5.1	Configuring storage system	101
5.5.2	Staging folder	102
5.5.3	Logs	102
5.6	Summary	102
Chapter 6. Integrating IBM Content Manager OnDemand with the Information Archive device		103
6.1	Introduction	104
6.2	IBM Content Manager OnDemand overview.	104
6.2.1	Cache storage manager	106
6.2.2	Archive storage manager	106
6.3	Integration architecture	107
6.3.1	Supported configurations	108
6.3.2	Content Manager OnDemand data stored in the Information Archive device	109
6.3.3	Content Manager OnDemand data migration to the Information Archive device	109
6.3.4	Retention: Coordinating Tivoli Storage Manager and Content Manager OnDemand policies	110
6.3.5	Protecting data with the data retention protection protocol	111
6.4	Integration implementation	111
6.4.1	Integrate Content Manager OnDemand with the Information Archive device	112
6.4.2	Modify the configuration files.	115
6.4.3	Customize the setup	119
6.4.4	Test the integration	126
6.5	Hints and tips.	128
6.5.1	Coordinating Tivoli Storage Manager and OnDemand policies	128
6.5.2	Ensuring that the Information Archive device is properly installed	128
6.6	Summary	128
Chapter 7. IBM Information Archive for Email, Files, and eDiscovery		131
7.1	Introduction	132
7.2	The solution bundle	133
7.3	Solution configuration	135
7.4	Use cases	137

7.4.1 Use case 1: Mailbox archiving, file archiving, and eDiscovery for compliance	137
7.4.2 Use case 2: Archiving for storage and business optimization	138
7.5 Summary	138
Related publications	141
IBM Redbooks publications	141
Online resources	141
How to get Redbooks publications	142
Help from IBM	142

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of 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.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.


COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

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

AIX®	OS/400®	System x®
DB2®	pSeries®	System z®
developerWorks®	Quickr™	Tivoli®
FileNet®	Redbooks®	WebSphere®
i5/OS®	Redpaper™	z/OS®
IBM®	Redbooks (logo)  ®	zSeries®
Lotus®	System i®	
OmniFind®	System Storage®	

The following terms are trademarks of other companies:

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Preface

The need to archive information is on the rise, driven by content and data growth, regulatory compliance, legal discovery, and data protection requirements. The IBM® Smart Archive strategy is a comprehensive, unified, and integrated archive strategy that combines IBM software, systems, and service capabilities that are designed to help organizations extract value and to gain new intelligence from information by collecting, organizing, analyzing, and using that information.

IBM Enterprise Content Management (ECM) products and offerings combined with the IBM Information Archive device provides the type of end-to-end Smart Archive solution that is a critical component of the IBM Smart Archive strategy.

This IBM Redpaper™ publication focuses on the benefit and technical details of the integration of ECM products and offering with the Information Archive device. We explain the need and concept behind the IBM Smart Archive strategy, provide an overview of the Information Archive device and ECM products and offerings, and discuss how integrating them can benefit an organization. The technical details that we provide include integrating the Information Archive device with the following ECM products and offerings:

- ▶ IBM FileNet® P8
- ▶ IBM Content Manager
- ▶ IBM Content Manager OnDemand

The Information Archive for Email, Files, and eDiscovery solution comes with the preintegrated and preconfigured Information Archive device and the preinstalled ECM software.

The paper introduces technical sales people and IT specialists to the IBM Smart Archive strategy and the integration of Information Archive and ECM products and offerings. At the same time, it provides IT specialists specific guidance about performing the integrations.

The team who wrote this paper

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization (ITSO), Rochester Center.

Wei-Dong Zhu (Jackie) is an Enterprise Content Management Project Leader with ITSO. Jackie joined IBM in 1996 and has more than 10 years of software development experience in accounting, image workflow processing, and digital media distribution. She is a Certified Solution Designer for IBM Content Manager and has managed and led the production of many Enterprise Content Management IBM Redbooks® publications. Jackie holds a Master of Science degree in Computer Science from the University of the Southern California.

Roger Bacalzo is a Senior Software Engineer in the IBM Enterprise Content Management group based in Seattle, Washington, U.S. Roger holds a Masters of Science degree from the University of Colorado. He has over 20 years experience in enterprise software development with the past 5 years in the storage development group of IBM FileNet Content Manager. His areas of expertise in the IBM FileNet Content Manager product include content storage, content federation services (CFS), and content search. Roger is published in these areas on IBM Redbooks and developerWorks®.

John Clifton is the Technical Strategist for IBM Enterprise Content Management in the U.K. and Ireland. John has more than 15 years of experience in enterprise content management and the last 10 years working for IBM. John supports large Financial Services organizations. He acts as an advisor to customers on complex projects, where multiple products are involved or where product choice is not clear. John also serves as a representative of the technical community for the U.K. and Ireland ECM management team. John is a Chartered Engineer and Chartered IT Professional with a broad knowledge of and experience in IT, including software development, consulting, and project management.

Ahmed Dewidar is an IBM Enterprise Content Management Technical Sales Specialist in Middle East and North Africa. Ahmed has over 5 years experience with ECM technical sales in the region. Due to his knowledge in ECM, Ahmed attained his Level 1 Certification for IT Specialist. Ahmed is also IBM Certified Solution Designer for IBM Content Manager v8.3 and IBM Certified Designer for FileNet Content Manager 4.5. Due to his engagements with customers, Ahmed is focused primarily on the essential content and advanced case management waves portfolio, out of the four waves of IBM ECM innovations.

Guillermo Gomez is a Product Manager in the Storage-Archiving group and former Software Engineer in the IBM System i® with more than 13 years of experience. He is based in Guadalajara, Mexico. Guillermo joined IBM in 1996 in the Software Development area. As a part of the IBM Smart Archive strategy,

Guillermo leads the integration phase through testing phase of the IBM Information Archive for Email, Files, and eDiscovery solution.

Hassan Shazly (AI) is a Senior Software Engineer and Manager in the U.S. He has 35 years of experience in Information Systems. He holds a Ph.D. in Remote Sensing and Image Processing from the University of South Carolina. AI joined IBM in 1996. His current areas of expertise include Content Manager and Content Manager OnDemand for z/OS®. He has written multiple articles and presented at several conferences on topics including ECM, client and server technology, image processing, and systems performance.

Thanks also to the following people for their contributions to this project:

Servando Varela
Aliye Ozcan Ergulen
Ben Boltz
Patrick Chesnot
Bob Kreuch
David Lebuttsch
Ruth Hildebrand-Lund
IBM Software Group

Frank Boerner
Roland Hoppe
Rene Wuellenweber
IBM Global Technical Services

Manuel Avalos Vega
IBM Systems and Technology Group

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks publications in one of the following ways:

- ▶ Use the online **Contact us** review Redbooks form found at:

ibm.com/redbooks

- ▶ Send your comments in an e-mail to:

redbooks@us.ibm.com

- ▶ Mail your comments to:

IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- ▶ Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- ▶ Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- ▶ Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- ▶ Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>

- ▶ Stay current on recent Redbooks publications with RSS Feeds:

<http://www.redbooks.ibm.com/rss.html>



End-to-end Smart Archive solutions with the IBM Smart Archive strategy

The IBM Smart Archive strategy is a comprehensive, unified, integrated, and information-aware archive strategy. Many IBM solutions, services, and offerings are available that derive from the Smart Archive strategy. The end-to-end Smart Archive solutions that are created by combining the IBM Enterprise Content Management portfolio of software products and offerings with the IBM Information Archive compliance storage hardware offer the most value from the hardware, software, and associated service elements.

This chapter introduces the IBM Smart Archive strategy and the end-to-end Smart Archive solutions that use Enterprise Content Management (ECM) and the Information Archive device. It includes the following topics:

- ▶ Executive summary
- ▶ The history of archiving
- ▶ Archiving compared to backup
- ▶ The IBM Smart Archive strategy
- ▶ IBM solutions for the Smart Archive strategy
- ▶ The Information Archive device and ECM
- ▶ Summary

1.1 Executive summary

The need to archive information is on the rise, driven by content and data growth, regulatory compliance, legal discovery, and data protection requirements. With these ever-increasing requirements for long-term retention, organizations are looking for an integrated approach to manage and use information over its lifetime while driving down infrastructure costs and complexity.

The IBM Smart Archive strategy is a comprehensive, unified, and integrated archive strategy. The Smart Archive strategy is not a single product. Instead, it is a comprehensive approach and strategy to archiving. The Smart Archive strategy combines optimization of the archive infrastructure with information awareness, which enables companies to intelligently and systematically know the information that needs to be collected, classified, and archived and provides the capabilities for doing so in a unified manner. These capabilities deliver a total solution with a modular approach, facilitating archiving across multiple content types, repositories, content generators and retention requirements and help companies to maximize the value of their hardware, software, and services.

The Information Archive device provides the hardware along with the archive and retention software for securely storing retention-managed information for long terms. ECM provides a complete set of products and offerings for collecting, assessing, classifying, analyzing, archiving, and managing information over its life cycle. The ECM portfolio of products and offerings along with the Information Archive device provide an end-to-end Smart Archive solution as part of the IBM Smart Archive strategy. IBM Information Archive for Email, Files, and eDiscovery is one such solution. The Information Archive for Email, Files, and eDiscovery solution comes with the preintegrated and preconfigured Information Archive device and the preinstalled ECM software.

1.2 The history of archiving

An *archive* is a place where public or corporate records are kept. *Archiving* is the act of preserving information, typically for long periods of time, to meet multiple requirements.

1.2.1 Reasons behind archiving

Archiving has been going on for many years, particularly in the physical world. As the volume of electronic information grows ever more rapidly, the need for a strategic approach to archiving becomes imperative.

Reasons for archiving information include:

- ▶ Business operational requirement: Certain information is required to remain available over a long period of time. For example, the details of a life insurance policy is required to be kept for a specific period of time, although request for this information might be none or few.
- ▶ Corporate governance requirement: An organization might need to retain certain information to meet the governance obligations.
- ▶ Regulatory compliance: An organization is subject to multiple regulatory rules which require the preservation and production of records.
- ▶ Legal discovery (also known as *eDiscovery* in the U.S., *eDisclosure*, in the U.K., and *litigation and response* in general): Potential litigious situations must be addressed adequately through the ability to find and produce documentary evidence.
- ▶ Corporate intelligence: Long-term information, especially those with unstructured content, can be analyzed to discover actionable insight that can help to make better use of systems, improve business processes and operations, or bring new ways of doing business.

1.2.2 Archiving technologies

From the early days of tape, numerous technologies have at one time or another been applied to the problem of preserving information for long periods after it is no longer required for day-to-day operations. Such technologies include:

- ▶ Tape
- ▶ Floppy disk: 8-in, 5.25-in, and 2.5-in
- ▶ CD-ROM
- ▶ Magneto Optical Disk
- ▶ DVD

The storage technology is changing. Ever larger, smaller magnetic disks are the order of the day.

The IBM Information Archive device exploits the latest SATA disk technology to good effect. Combined with a tiered approach to storage management and the ability to mix disk and tape technologies, it provides a unique and unbeatable proposition when it comes to total cost of ownership.

1.3 Archiving compared to backup

It is important to understand the differences between archiving and backup. Although meeting the requirements of both might involve common technology, the business reasons for archiving and backup are distinct and need to be well understood.

1.3.1 Backup

Backup is required in an organization to preserve operational data and to allow the data to be restored in the event of a system failure. Typically, backups are retained and recycled over short periods of time. Multiple versions of a backup are usually kept. The data that is backed up is usually a copy or delta from its original data source in its original location. The backup strategy that any organization adopts is determined by the business operational needs and IT infrastructure of the business.

1.3.2 Archiving

Archiving, by contrast, entails the long-term retention of information, typically after it has completed its active phase of the life cycle. Archiving normally involves moving the data from its operational location to one that has the following characteristics:

- ▶ Less expensive
- ▶ Less performant
- ▶ More secure
- ▶ Compliant with regulatory requirements
- ▶ Use and reuse of the archived information

Retention and disposition

A key aspect of the archiving of information is how long the information will be retained. Retention periods can be determined according to the factors affecting the different types of data to be archived, including operational requirements, governance, and compliance and legal requirements.

Information *retention* is important to meet organization's objective and obligations. Equally important to retention is information *disposition*. Data that should be kept by an organization can be considered and exploited as an asset. Conversely, data that should not be kept in the organization becomes a liability if it is not properly disposed. When planning for archiving, organizations need to address the requirement for both retention and disposal.

Archived information should also be eDiscovery ready, meaning that it should be properly indexed, dynamically available for litigation searches, and be able to accommodate hold requests. When a hold is put on a piece of archived information, it should not be deleted even when the retention period is expired.

1.4 The IBM Smart Archive strategy

Information is growing by volume, variety, and velocity:

- ▶ Experts estimate that petabytes of new information is generated daily.
- ▶ The majority of new data growth is in unstructured content.
- ▶ The amount of information and the speed of its growth has long surpassed storage capacity and is spiraling out of control.

Organizations can no longer keep information forever. Rethinking what information to store and how information infrastructures work is a must.

Many existing stand-alone archiving solutions do not address today's challenges regarding volume, variety, and velocity information growth. In addition, the existing solutions typically cover only one information type, such as email. However, an organization's archiving solution should cover all information that needs to be archived. Furthermore, performing search and eDiscovery on existing solutions can be challenging because the existing solutions were not created to accommodate this type of request and need.

Many new archiving solutions in the market today are for a specific hardware or software only. To integrate multiple hardware and software components, often from multiple vendors, without much best practice guidance, the integration can be a challenge.

Another issue that organizations are facing today is that there is no single archiving infrastructure that has the flexibility to meet multiple requirements, resulting in the need to deploy multiple point archiving storage solutions for different requirements. Organizations often must choose one storage environment to meet one regulation or to meet one requirement and then must choose another storage deployment to meet other requirements. Organizations end up with a collection of separate stand-alone infrastructures or storage platforms with no commonality between them.

All of these challenges and issues add cost, risk, and complexity to an archiving solution.

A successful archiving strategy can help organizations:

- ▶ Understand what information is housed enterprise-wide to make it easier to know, manage, trust, and use.
- ▶ Ensure that important information is properly retained and protected to support compliance efforts and to help reduce risk.
- ▶ Reduce costs and complexity while increasing the flexibility of the information infrastructure.

The Smart Archive strategy helps organizations understand the choices in archiving solutions with multiple modular yet integrated solutions that provide choice of management and delivery models based on a common information lifespan and policies. It enables organization to decide how to manage and access the archived information without sacrificing the economies of scale and cost benefits of the organization's information infrastructure options. Using information-aware devices, organizations can maintain consistency in retention and disposition policies, deduplication, migration, and long-term preservation of information.

The IBM Smart Archive strategy consists of the following elements:

- ▶ Optimized and unified ingest (assessment, collection, and classification)
- ▶ Flexible and secure infrastructure with unified retention and protection
- ▶ Integrated compliance, records management, analytics, and eDiscovery

Figure 1-1 summarizes the IBM Smart Archive strategy.

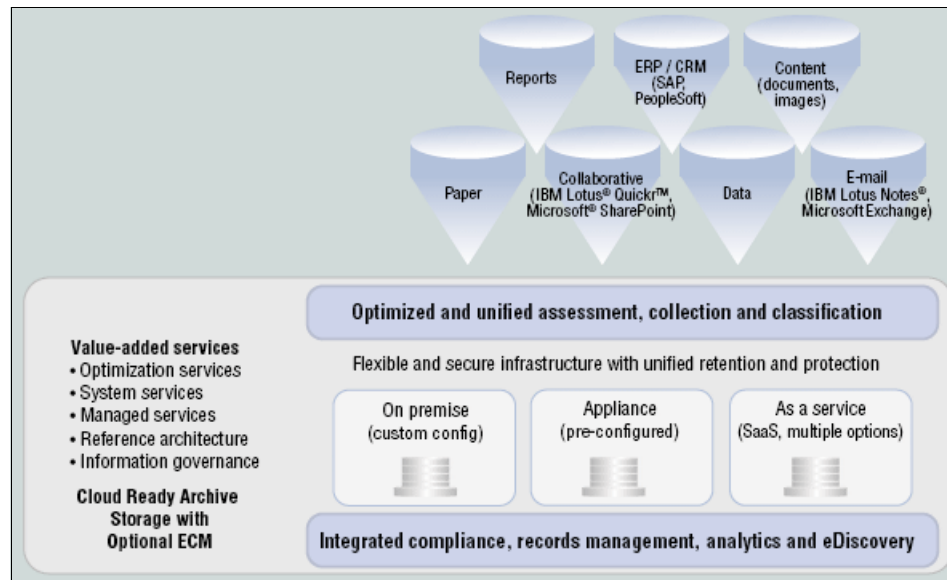


Figure 1-1 The IBM Smart Archive strategy

Optimized and unified content ingest includes the ability to handle a broad spectrum of information types, such as paper, reports, collaborative content, enterprise application data, structured data, documents, email, or images. It enables a deeper understanding of what information to archive through discovery- and analytics-based assessment technologies. It helps eliminate point solution complexity and cost by unifying data and content archiving through common collection (ingest) and classification technologies.

Flexible and secure infrastructure with unified retention and protection supports multiple deliverable models to make cost-optimized retention possible. It is essential to build an infrastructure that allows organizations to easily set and enforce policies, protect the integrity of business records with enhanced security, and support business continuity and resiliency while optimizing and streamlining archiving and storage management. Three models are available:

- ▶ The *on-premise delivery model* is where the archive is deployed on the customer premise.
- ▶ The *appliance delivery model* is an appliance that is preconfigured with the necessary software.
- ▶ The *service model* is where potentially organizations can use IBM Business Partners or IBM Global Services to deliver a flexible and secure infrastructure.

Integrated compliance, records management, analytics, and eDiscovery reduces risk, enables organizations to respond more quickly to legal inquiries, to establish trust, and to use information. It helps to ensure compliance by managing all records throughout their lifespan, regardless of type, media, or storage location, thus helping improve record authenticity, reliability, integrity, and usability. Organizations can also collect, hold, manage, assess, and intelligently cull collected electronic information to lower eDiscovery response costs and to form early case strategies.

The Smart Archive strategy is a comprehensive strategy that helps organizations visualize how to take advantage of IBM's capabilities to design specific strategic archiving solutions for their unique requirements. The advantage of having a comprehensive strategy is for organizations to start with specific intermediate archiving goals in mind (such as email archiving) and to expand into other archiving areas intelligently that grows in parallel with business archiving needs. Organizations do not have to do all at once. You can start at a point you need, and expand at your own pace.

1.5 IBM solutions for the Smart Archive strategy

IBM has combined some of the leading software, hardware systems, and services to create flexible, information-aware solutions for archiving that can start delivering value almost immediately. Organizations can speed time to value and simplify their environment through multiple modular yet integrated solutions, including a choice of management and delivery models based on common information lifespan and policies.

IBM offers specific solutions that address the following needs:

- ▶ Optimized and unified ingest capabilities
 - IBM Content Analytics for Assessment
 - IBM Content Collector family of offerings
 - IBM Classification Module software
 - Datacap
- ▶ Flexible and secure infrastructure
 - IBM Enterprise Content Management content repositories
 - IBM Information Archive storage repositories
 - IBM and partner archive services
- ▶ Integrated compliance, records management, analytics and eDiscovery
 - IBM Content Analytics for Assessment
 - IBM Content Collector family of offerings
 - IBM Classification Module software

- IBM Enterprise Records software
- IBM Records Manager software
- IBM eDiscovery Manager and eDiscovery Analyzer software

To learn more about the Smart Archive strategy and its associated solutions, see:

<http://ibm.com/software/data/smart-archive>

1.6 The Information Archive device and ECM

The Information Archive device is the next-generation information retention solution that is designed as a universal archiving repository for all types of content for midsize to enterprise clients. It provides the hardware along with the archiving and retention software to securely store retention-managed information for long terms. For an overview of the Information Archive device, see Chapter 2, “Introducing the IBM Information Archive device” on page 13.

ECM provides a complete set of products and offerings for collecting, assessing, classifying, analyzing, archiving, and managing information over its life cycle. It also provides a set of products and offering for eDiscovery specific capabilities. For an introduction to all of the ECM products and offerings, see Chapter 3, “Introducing IBM Enterprise Content Management” on page 21.

Combining the Information Archive device with ECM, organizations get the hardware appliance that provides the superior archiving infrastructure and the software products and solutions that provide comprehensive enterprise content management capabilities. Working together, they offer the end-to-end archive and enterprise content management solutions as part of the Smart Archive strategy.

Combining ECM products and offerings with the Information Archive device provides the following features and benefits:

- ▶ Ability for organizations to manage their information growth (volume, variety, and velocity) challenge.
- ▶ Archiving and retention management of a variety of information types.
- ▶ Comprehensive, enterprise-level content management capabilities.
- ▶ Advanced classification, content analytics, records management, search, and eDiscovery capabilities.
- ▶ Smooth integration of hardware (the Information Archive device) and software (ECM) from one vendor.
- ▶ A single archiving infrastructure that has the flexibility to meet multiple requirements.

1.6.1 End-to-end Smart Archive solutions

Combining ECM products and offerings with the Information Archive device enables organizations to create end-to-end Smart Archive solutions that maximize the values of software, hardware, and associated services. For example, a solution can provide the following functions with the associated product offerings:

- ▶ Content management using IBM FileNet Content Manager
- ▶ Content collection and archiving using IBM Content Collector for Email, IBM Content Collector for File Systems, and IBM Content Collector for SharePoint
- ▶ SAP archiving and document enabling using IBM Content Collector for SAP
- ▶ Additional enterprise content management functions:
 - Content analytics using IBM Content Analytics.
 - Advanced content classification using IBM Content Classification.
 - Records management using IBM Enterprise Records.
- ▶ Physically storing content in IBM Information Archive for compliance with regulations requiring WORM or NENR storage
- ▶ Managing, reusing, and searching and discovering of the archived content by using IBM eDiscovery Manager and IBM eDiscovery Analyzer along with the IBM Information Archive device and IBM FileNet Content Manager

For information about content management product and offerings, see 3.2, “Essential ECM” on page 22.

Content collection and archiving, content analytics, advanced content classification, records management, search, and eDiscovery product and offerings are all related to IBM Information Lifecycle Governance (ILG), a comprehensive compliance platform with flexible, integrated solutions that helps companies gain control and manage the life cycle of their information. For more information about ILG related product and offerings and their capabilities, see 3.4, “Information life cycle governance” on page 27.

Currently, as part of the Smart Archive strategy, one solution IBM offers is IBM Information Archive for Email, Files, and eDiscovery. This solution is preintegrated, preconfigured, and preinstalled bundle of software, hardware, and implementation services. The solution provides built-in mailbox management and journaling, file systems archiving, and eDiscovery capabilities. For more information about this solution, see Chapter 7, “IBM Information Archive for Email, Files, and eDiscovery” on page 131.

1.7 Summary

The IBM Smart Archive strategy is a comprehensive, unified, and integrated archive strategy. There are many products, offerings, and services available to materialize the strategy into deliverable solutions. Combining the Information Archive device with ECM product and offerings is part of the Smart Archive strategy that delivers many of the end-to-end Smart Archive solutions.

This paper focuses on the integration between the Information Archive device and ECM products. The remainder of the chapters in this paper are organized as follows:

- ▶ Chapter 2, “Introducing the IBM Information Archive device” on page 13 introduces the IBM Information Archive device.
- ▶ Chapter 3, “Introducing IBM Enterprise Content Management” on page 21 provides an overview of the products and offerings from IBM Enterprise Content Management.
- ▶ Chapter 4, “Integrating IBM FileNet P8 with the Information Archive device” on page 37 explains the integration of IBM FileNet P8 with the Information Archive device.
- ▶ Chapter 5, “Integrating IBM Content Manager with the Information Archive device” on page 71 explains the integration of IBM Content Manager with the Information Archive device.
- ▶ Chapter 6, “Integrating IBM Content Manager OnDemand with the Information Archive device” on page 103 explains the integration of IBM Content Manager OnDemand with the Information Archive device.
- ▶ Chapter 7, “IBM Information Archive for Email, Files, and eDiscovery” on page 131 discusses this preintegrated, preconfigured, and preinstalled solution.



Introducing the IBM Information Archive device

The IBM Information Archive device is the next-generation information-retention solution. It is designed as a universal archiving repository for all types of content to help midsize and enterprise clients reduce cost, manage risk, and address an organization's total information retention needs. This chapter introduces the IBM Information Archive device and provides a summary of its features and benefits.

This chapter includes the following topics:

- ▶ Introduction
- ▶ Archiving and retrieving information
- ▶ Collections
- ▶ Information protection
- ▶ Retention policies
- ▶ Data encryption
- ▶ Disaster recovery protection
- ▶ Summary

2.1 Introduction

The IBM Information Archive device is a universal storage repository for all types of content, structured and unstructured, compliant or non-compliant. It is a scalable appliance with a raw capacity that can scale up to 608 TB. It provides extensive capabilities, such as documents deduplication and compression, to maximize its capacity.

The Information Archive device is also a one box appliance (Figure 2-1) that is designed to meet the requirements for long-term managed retention of electronic stored information. It is a simple answer to the complex challenge of how to address the growing burden of information retention.



Figure 2-1 The IBM Information Archive device

The Information Archive device offers the following features and benefits:

- ▶ Built-in data deduplication and compression features that reduce disk capacity requirements.
- ▶ Support of transparent migration of information from disk to tape over the information life cycle through policy-driven, hierarchical storage management capabilities that eliminate the cost of services to migrate archive data to new hardware.

- ▶ Easy movement of data between storage tiers that reduces the total cost of ownership when moving long-term, less accessed information to lower-cost storage tiers.
- ▶ Support for up to three separate information repositories (collections) in a single system and three information protection levels per collection that provides flexibility to meet various information retention needs.
- ▶ Preintegrated and preconfigured that makes scaling easy, that saves time for set up, and that enables adding and removing storage as needed.
- ▶ A compliant storage repository with non-erasable, non-rewritable (NENR), Write Once Read Many (WORM) data storage protection, and data encryption options.
- ▶ Single user interface, which enables easy usage and administration.

The Information Archive device is fully integrated archive appliance that addresses the complete information retention needs of midsize and enterprise organizations. The Information Archive device offers intelligent high availability and disaster protection functions, and it supports applications such as IBM Enterprise Content Management (ECM) products with policy harmony.

2.2 Archiving and retrieving information

With the Information Archive device, you can archive and retrieve documents from a wide range of software applications using industry-standard interfaces such as Network File System (NFS). The Information Archive device supports a variant of the IBM System Storage® Archive Manager interface.

With the Information Archive device, users can archive information directly to protected storage from their workstations or file servers, and they can also retrieve information directly from storage. Users can archive information using the System Storage Archive Manager client or using NFS and can retrieve information using System Storage Archive Manager, NFS, or HTTP.

2.3 Collections

To help customers manage information with different retention requirements, the Information Archive device offers up to three collection options within a single system. A *collection* is a logical container that is used to store documents and to define retention and access policies. Storing documents within collections provides the flexibility to assign different levels of protection, access points, and retention policies for different kinds of information. It also allows for information

movement between the Information Archive device's two storage tiers, according to the collection's classification.

A collection can be either NFS based or System Storage Archive Manager based. Applications archive data using the NFS or System Storage Archive Manager client.

NFS collection

An NAS collection provides standard NFS interfaces, such as NFS and HTTP Read. It allows for different levels of information protection and supports multiple methods for setting retention by policy and application. In addition, it supports automatic deletion of data following data expiration.

System Storage Archive Manager collection

A System Storage Archive Manager collection provides a secure retention managed environment. It uses the Tivoli® Storage Manager API or Tivoli Storage Manager Archive Client for archiving. It supports various retention management policies and provides a fully supported integration point to the ECM repositories.

2.4 Information protection

The Information Archive device provides three information protection levels for up to three archive collections within a single system. This 3x3 combination provides the flexibility of protection to meet various information retention requirements. With different collections, you can assign customized levels of protection and retention policies for different types of data based on information retention requirements.

The Information Archive device offers the following protection levels:

- ▶ Basic
- ▶ Intermediate
- ▶ Maximum

Basic level of protection

This level of protection provides maximum flexibility for managing an organization's information retention. Documents can be deleted before they expire. Retention periods can be increased and decreased. Documents with an extended retention period due to a retention hold can be deleted. The document protection level can be changed to another level at any time.

Intermediate level of protection

This level of protection provides an intermediate level of flexibility for information retention. With this protection level, the retention periods can be increased and decreased, and documents cannot be deleted until the retention period expires. Documents with an extended retention period due to a retention hold cannot be deleted. The document protection level can be changed to maximum at any time, but it cannot be lowered to the basic level. The collection can never be deleted.

Maximum level of protection

This level of protection provides the least flexibility for information retention but offers maximum protection to meet strict business, legal, or regulatory retention needs. Similar to the intermediate protection level, documents cannot be deleted until the end of their retention period. Those documents with an extended retention period due to a retention hold cannot be deleted. Unlike the intermediate level, document retention periods can only be increased and not decreased. After enabling the maximum protection, you cannot change the protection level to any other level. Similar to the intermediate protection level, the collection with a maximum level of protection can never be deleted.

2.5 Retention policies

The Information Archive device prevents documents from being physically deleted, whether intentionally or by mistake, at the storage level.

To meet various retention demand, the Information Archive device can accept and execute policies that are created by other applications. For example, when integrating with ECM products such as IBM Content Manager or IBM FileNet Content Manager, the retention policies set in these applications can take over and control how the documents are to be retained in the Information Archive system.

The Information Archive device also provides the following retention policies that are customizable based on specific retention requirements:

- ▶ Time-based and event-based information retention policies to protect information from intentional or accidental deletion or modifications.
- ▶ Retention hold and deletion hold policies to set an indefinite period of retention for a file, such as during legal discovery.

Time-based retention policy

For a time-based retention policy, data is retained for a predetermined, fixed length of time (for example, keep all emails for 6 years). During the specified period, emails are not deleted.

Event-based retention policy

For an event-based retention policy, the data must be retained until a certain event occurs (for example, keep all customer account records until the account is closed). In this scenario, the account records are kept indefinitely and cannot be deleted until after the account is closed.

Combination retention policy

For a combination of a time-based and an event-based policy, the data must be retained for a predetermined, fixed length of time after a certain event occurs (for example, keep all customer account records for 7 years after the account is closed). In this scenario, the account records are kept indefinitely. If the account is closed, during the 7 years after which the account is closed, the account records cannot be deleted.

Retention hold and deletion hold

If there is a retention hold or deletion hold on a document due to, for example, a legal discovery, the document cannot be deleted and must be kept indefinitely, regardless of the retention policy that is associated with the document. When the hold is removed, then the retention policy becomes effective again.

2.6 Data encryption

The Information Archive device has encryption capabilities that enable companies to increase security for data storage and remote data transmission. The Information Archive device encryption provides enhanced security for businesses using 128-bit AES or 56-bit DES encryption technology. It can manage encryption keys offered with IBM System Storage TS1120 and LTO-4 tape drives.

2.7 Disaster recovery protection

The Information Archive device helps to protect data during disasters. The Information Archive device supports Enhanced Remote Mirroring, which enables asynchronous copies of the information (with the support of consistency group) to be maintained on two separate Information Archive devices. When one site is

down, the second Information Archive device from another site still has the data for business operations.

The Information Archive device also provides the capability to use advanced LTO-3 or LTO-4 WORM tape libraries to provide efficient and cost-effective copies of the information, and enables the Information Archive database to support off-site data storage and recovery in the event of a failure or the need to relocate to an alternate facility.

2.8 Summary

Organizations are challenged with managing ever increasing volumes of information, while delivering a flexible and cost effective storage infrastructure. The IBM Information Archive device helps to meet these challenges with a fully integrated IBM hardware and software solution that can be configured to meet multiple diverse storage requirements and that can be integrated with donor applications using open interfaces.

This paper focuses on the integration between the Information Archive device and ECM products. The remainder of the chapters are organized as follows:

- ▶ Chapter 3, “Introducing IBM Enterprise Content Management” on page 21 provides an overview of the products and offerings from IBM Enterprise Content Management.
- ▶ Chapter 4, “Integrating IBM FileNet P8 with the Information Archive device” on page 37 explains the integration of IBM FileNet P8 with the Information Archive device.
- ▶ Chapter 5, “Integrating IBM Content Manager with the Information Archive device” on page 71 explains the integration of IBM Content Manager with the Information Archive device.
- ▶ Chapter 6, “Integrating IBM Content Manager OnDemand with the Information Archive device” on page 103 explains the integration of IBM Content Manager OnDemand with the Information Archive device.
- ▶ Chapter 7, “IBM Information Archive for Email, Files, and eDiscovery” on page 131 discusses this preintegrated, preconfigured, and preinstalled solution.



Introducing IBM Enterprise Content Management

IBM Enterprise Content Management (ECM) provides a comprehensive set of capabilities for managing structured and unstructured information. This chapter introduces the ECM solutions in four product areas and introduces some of the products in the portfolio. It includes the following topics:

- ▶ Introduction
- ▶ Essential ECM
- ▶ Advanced case management
- ▶ Information life cycle governance
- ▶ Trusted Content Analytics
- ▶ Summary

3.1 Introduction

ECM solutions help companies and government agencies reduce costs and make quick, smart, and cost-effective decisions. The ECM solutions are organized in the following areas:

- ▶ **Essential ECM:** Manages essential content anywhere. Production imaging and capture, enterprise report management, office document management (ODM), and standardization and consolidation are content management foundations for reducing costs and improving efficiencies.
- ▶ **Advanced case management:** Takes workflow and business process management to the next level. It brings people, processes, and information together to drive better case outcomes.
- ▶ **Information life cycle governance:** Expands the value of your repositories. It can reduce costs and risks while improving your compliance posture.
- ▶ **Trusted content analytics:** Helps you know, trust, and use your content to derive unexpected content insights.

3.2 Essential ECM

Most companies operate with strict content management requirements to help them control content and automate content related processes. As unstructured content grows exponentially, you need content management to capture, store, manage, integrate and deliver all forms of content across your enterprise. You need the right information at the right time to make the right decision.

IBM ECM content management products offer essential capabilities for production imaging and capture, enterprise report management, office document management, and content consolidation and standardization.

Essential ECM include the following offerings:

- ▶ **IBM FileNet Content Manager:** Essential for production imaging, office document management, and content consolidation and standardization. It is the core of the FileNet P8 Platform and is also offered in a Starter Pack.
- ▶ **IBM Content Manager:** Essential for production imaging on multiple platforms, office document management, and content integration. It is offered for z/OS and in a Starter Pack.
- ▶ **IBM Content Manager OnDemand:** Adds enterprise report management to your content management system. It integrates with FileNet P8 Platform.

- ▶ IBM Datacap: Provides premium data and document capture. Datacap was acquired by IBM in 2010.
- ▶ IBM FileNet Capture: Provide imaging capture to get critical content into FileNet P8 repositories.

IBM FileNet Content Manager, IBM Content Manager, and IBM Content Manager OnDemand can all integrate directly with the IBM Information Archive device to create end-to-end Smart Archive solutions. For technical integration detail, see the following chapters:

- ▶ Chapter 4, “Integrating IBM FileNet P8 with the Information Archive device” on page 37
- ▶ Chapter 5, “Integrating IBM Content Manager with the Information Archive device” on page 71
- ▶ Chapter 6, “Integrating IBM Content Manager OnDemand with the Information Archive device” on page 103

3.2.1 IBM FileNet Content Manager

IBM FileNet Content Manager is the core content management solution for the FileNet P8 Platform.

FileNet Content Manager combines document management with ready-to-use workflow and process capabilities to automate and drive your content-related tasks and activities.

FileNet Content Manager offers the following features and benefits:

- ▶ Provide the core engines for content, security and storage management for FileNet P8 Platform.
- ▶ Provide a foundation for office document management.
- ▶ Enable companies to manage content anywhere by using the Content Federation Services feature. Companies can choose to manage content in their current repositories, thus avoiding or postponing costly content migrations yet still being able to standardized on FileNet P8 Platform.
- ▶ Support active content which means the actions associated with content can trigger predefined actions and enable automation.
- ▶ Provide core capabilities such as document versioning, content security, and life cycle management for enterprise compliance management.
- ▶ Provide WorkPlace XT, an ready-to-use user interface designed for the business user. Also support portals, integration with collaboration tools, integration with Microsoft® Office 2007 and more.

3.2.2 IBM Content Manager

IBM Content Manager manages all types of digitized content across multiple platforms, databases and applications. It provides imaging, digital asset management, web content management and content integration. Built on a multi-tier distributed architecture, it provides the scalability to grow from a single department to a geographically dispersed enterprise. The products and solutions help large and smaller organizations meet regulatory compliance requirements, collaborate with greater productivity and improve customer service.

Content Manager products are:

- ▶ Content Manager Enterprise Edition: Store, manage, and integrate content on Linux®, UNIX®, Windows® and Linux on System z®.
- ▶ Content Manager Collaboration Edition: Collaborate on business content with Lotus® Quickr™ while using Content Manager Enterprise Edition.
- ▶ Content Manager Starter Pack: Provide essential content management functions.
- ▶ Content Manager for z/OS: Store, manage, and integrate high volumes of content on z/OS.
- ▶ Content Manager for IBM i: Integrate content and advanced workflow for the enterprise on i5/OS® or OS/400®.
- ▶ IBM Web Interface for Content Manager: Use Web 2.0 and AJAX technologies and open standards with the readily-available, easy-to-use, highly interactive Web Client.

3.2.3 IBM Content Manager OnDemand

Enterprise report management software manages formatted computer output and reports. IBM Content Manager OnDemand is high-performance middleware for enterprise report management.

Content Manager OnDemand helps companies transforming costly high-volume print output to electronic information capture and presentation in support of customer service.

Content Manager OnDemand products for enterprise report management are:

- ▶ Content Manager OnDemand for Multiplatforms: Enterprise report management software for AIX®, HP-UX, Linux, Linux on zSeries®, Solaris, and Windows.
- ▶ Content Manager OnDemand for z/OS: Enterprise report management software using z/OS.

- ▶ Content Manager OnDemand for i: Enterprise report management software architected for IBM i.
- ▶ IBM Web Interface for Content Manager OnDemand: A web client that employs open standards and supports Web 2.0 and AJAX technologies.

Both OnDemand for Multiplatforms and OnDemand for z/OS integrate with FileNet P8 repositories.

3.2.4 IBM Datacap

Datacap is one of IBM's newest acquisitions, and completes the end to end document management solution. Datacap's document capture and integration products help green initiatives by reducing paper requirements. Datacap captures, manages, and automates business information and then integrates it tightly with the p8 platform, creating streamlined enterprise solutions.

The Datacap product line consists of three main products:

- ▶ IBM Datacap Taskmaster Capture
- ▶ IBM Datacap Rulerunner
- ▶ IBM Datacap FastDoc Capture

Taskmaster and FastDoc Capture are document ingestion products. FastDoc Capture is a stand-alone product that integrates with file systems and Microsoft SharePoint. It can also be operated offline. Taskmaster is an SOA capture and automation solution with both web and thick clients. The Rulerunner service is a core differentiator for Datacap products and is embedded in Taskmaster and FastDoc. It can also run decoupled and can run as a web service. Easily extensible, it drives business rules and actions which make Datacap solutions so flexible and powerful.

Building on these core products, Datacap also offers a number of industry specific and technology specific solutions.

For more information about Datacap, see:

<http://www.ibm.com/software/data/content-management/datacap>

3.2.5 IBM FileNet Capture

IBM FileNet provides for ingesting paper and image-based content with another set of existing products:

- ▶ IBM FileNet Capture Desktop
- ▶ IBM FileNet Capture Professional

Similar to Datacap solutions, IBM FileNet Capture products automate control and classification during the capture process. This enhances IBM FileNet P8 compliance by increasing accuracy and lowering the risk of lost or inaccessible information. IBM FileNet Capture functionality is mostly a subset of Datacap's offerings, with three key differences:

- ▶ IBM FileNet Capture has internationalization and localization supported, particularly Asian language support for OCR in version 5.2.1.
- ▶ IBM FileNet Capture does not include advanced data extraction capabilities. The OCR in FileNet Capture is limited to simple document properties and full text.
- ▶ IBM FileNet Capture does not include many advanced features contained in Datacap capture products including web-based scanning and indexing, handwriting recognition, complex data validation, and data export.

IBM FileNet Capture Desktop and Professional scan, index, and convert content and stores it in IBM FileNet P8.

3.3 Advanced case management

The advanced case management strategy from IBM unifies information, processes and people to provide a 360-degree view of the case. In addition to content and process management, it relies on advanced analytics, business rules, collaboration and social software to help drive more successful, optimized case outcomes. Moreover, advanced case management solutions help capture industry best practices in frameworks and templates to empower business users and accelerate return on investment.

The advanced case management strategy and solutions are designed to deliver key case management improvements: insight, responsiveness, flexibility, better customer service and regulatory compliance.

The advanced case management product is IBM Case Manager.

Case Manager integrates content, process and people (capabilities) to deliver the next-generation of case management applications. It provides a foundation for capturing organizational best practices in templates and solutions to shorten the time-to-value for companies. It also provides the business-oriented tools for the development and deployment of these applications to assure the business users participation and ownership of the case application.

For more information about advanced case management strategy and the Case Manager product, see:

<http://www.ibm.com/software/data/advanced-case-management>

3.4 Information life cycle governance

Information is growing at an extremely fast pace in a great variety. Information must be managed properly, over its life span, from creation to destruction. Through its life span, it must be stored, secured and managed in the appropriate manner. Information must be available to the right people, in the right context, and at the right time, enabling better decisions to be made faster.

IBM Information Lifecycle Governance (ILG) is a comprehensive compliance platform with flexible, integrated solutions that help companies gain control and manage the life cycle of information. Information Lifecycle Governance uses the following solutions:

- ▶ Content Analytics for Assessment: Take charge of content. Assess and decide what information to manage, trust, and use, what information to preserve, and what to decommission or remove.
- ▶ Content Collection and Archiving: Manage the explosion of information volumes and types by collecting and archiving the right information into enterprise-level content repositories.
- ▶ Advanced Classification: Reduce the burden on users by accurately automating classification of content in its proper context.
- ▶ Records Management: Enforce retention and disposition policies. Proactively improve defensibility, lower risks and costs, and confidently dispose of information.
- ▶ eDiscovery Search and Analytics: Quickly and cost-effectively respond to eDiscovery, audit and internal investigation requests.

3.4.1 Content Analytics for Assessment

For many companies, information exist on file systems and shares, local mail files and databases. This information is typically created locally and shared selectively, through for example, email distribution. The proliferation of collaboration tools and basic content services has encouraged people to create more content.

Content Analytics for Assessment helps companies to discover the information that is created and make decisions about how that information should be

managed. It helps companies gain control over structured and unstructured content by:

- ▶ Eliminating unneeded content and systems.
- ▶ Preserving valued content.
- ▶ Deciding which content to manage, archive, trust, and use.

Content Analytics for Assessment is a solution that is a bundled offering of IBM ECM software working together:

- ▶ IBM Content Analytics: A robust software with user-friendly interface for exploratory analytics of unstructured content.
- ▶ IBM Classification Module: Automates classification of unstructured content by analyzing full text of documents and emails.
- ▶ IBM Content Collector for File Systems: Collects and archive content from file systems.

Content Analytics for Assessment solution helps companies in identify relevant content that needs to be moved into their trusted and managed repositories such as Content Manager Enterprise Edition and FileNet Content Manager.

3.4.2 Content Collection and Archiving

Traditional methods for archiving are typically based upon disparate and isolated approaches dependent upon data type. There is no single view of archived information across the organizations.

The goal of Content Collection and Archiving is to provide optimized and unified ingestion to:

- ▶ Enable a deeper understanding of what information to archive through discovery and analytics-based assessment technologies.
- ▶ Eliminate point solution complexity and cost by unifying data and content archiving through common collection (ingestion) and classification technologies.

IBM Content Collector is the key software that provides the Content Collection and Archiving solutions. Content Collector collects, enhances and manages email, file systems, SharePoint and more. In addition, Content Collector provides the interface and functions that enable companies to connect and collect information from various source systems and applications and archive to a central repository, such as IBM Content Manager and IBM FileNet Content Manager.

Content Collector includes the following product offerings:

- ▶ IBM Content Collector for Email: Collect and archive emails.
- ▶ IBM Content Collector for File Systems: Controls document and content on network file shares.
- ▶ IBM Content Collector for Microsoft SharePoint: Collects SharePoint documents into a centrally managed archive.
- ▶ IBM Content Collector for SAP: Provides SAP data and document archiving, SAP content-enabling, and complementary process management for SAP environments.
- ▶ IBM Content Collector Discovery Analytics: Complete collection, archiving, eDiscovery and analytics capabilities in one solution.

3.4.3 Advanced Classification

With explosion of information and variety of information created, automate content classification is a necessity. With Content Collection and Archiving solutions, collecting and archiving everything is costly, time-consuming, and might not be necessary. One way to reduce the cost is using automated classification.

Advanced Classification reduces burdens on users and improves ability to classify unstructured content automatically. It also changes content collection and archiving by reducing the hardware requirement (by not archiving everything) and increasing operation efficiency (by not asking users to make decision, item-by-item about the content).

IBM Classification Module software is the key element in Advanced Classification solutions. Classification Module helps classifying unstructured content. It is capable of learning the content categories from a set of training documents and then proceed to automatically archive large volume of content. It can also propose taxonomies, provide suggestions for categories, and improve its classification ability through learning.

Classification Module combines rules with context based analysis for consistent, accurate, adaptive results. It can be embedded within Content Collector, so that additional information about the content can be obtained (from Classification Module) before Content Collector making decisions about what to do with the content after it is being collected.

3.4.4 Records Management

Records Management solutions help users address evolving governance mandates to meet regulatory, legal, and fiduciary requirements. Proactive adherence to information retention policies and procedures is a critical facet of companies' compliance strategies. The Records Management products help companies enforce centralized policy management for file plans, retention schedules, legal preservation holds, and auditing.

The Records Management products are designed to help users streamline processes and automate manual user and administrator activities throughout the record life cycle. Easily deployed and U.S. DoD 5015.2-certified, Records Management solutions deliver a rapid and compelling return on investment while reducing the risks and costs in managing electronic and physical record assets.

Records Management products are:

- ▶ IBM Enterprise Records (formerly called FileNet Records Manager): Securely captures, declares, classifies, stores and disposes of electronic and physical records.
- ▶ IBM Records Manager: A records management engine for electronic records management in applications.

3.4.5 eDiscovery Search and Analytics

Electronic discovery, or eDiscovery, is the process of identifying, locating, securing and producing electronically stored information (ESI) and materials, typically in response to litigation. It is a critical part of information life cycle governance solutions.

eDiscovery Search and Analytics solutions help reduce the cost from eDiscovery, balance in-house implementation with outsourced components, and proactively manage electronically stored information. eDiscovery Search and Analytics solutions manage content within ongoing business processes and helps increase defensibility - in a security-rich, consistent, auditable manner. It enables a disciplined approach to reducing costs and risk while improving ongoing operational benefits. eDiscovery Search and Analytics is a key family of solutions, along with Content Analytics for Assessment, Content Collection and Archiving, Advanced Classification, and Records Management, to help companies achieve information life cycle governance.

eDiscovery Search and Analytics solution includes the following products:

- ▶ IBM eDiscovery Manager: Provide IT users the ability to respond easily to discovery requests. It also provides comprehensive audit tracking to support authenticity and prove chain of custody.
- ▶ IBM eDiscovery Analyzer: Provide the detailed analysis and review capabilities that enable early case assessment and provides the opportunity to reduce case costs.
- ▶ IBM Content Collector: Enable collecting information from various source systems and applications (including email systems, file systems, Microsoft Sharepoint) and archive to a central repository.
- ▶ IBM Classification Module: Automatically and intelligently classify unstructured content to help organizations lower risk and expense.

3.5 Trusted Content Analytics

Enterprise Content Management software includes offerings focused on trusted content analytics to help companies know, trust, use, and exploit all of their enterprise content, not just what is in their repositories. Trusted Content Analytics analyze unstructured content to unlock critical business insights. It provides new ways to reduce costs, deliver business innovation, and create competitive advantage.

Content Analytics and Search products and solutions include:

- ▶ IBM Content Analytics: Discover and use content to derive unexpected business insights, take action, and share with other users and systems.
- ▶ IBM Content Analytics for Assessment: Know your content. Analyze what you have. Decommission the unnecessary. Preserve what matters.
- ▶ IBM Classification Module: Automatically and intelligently classify unstructured content to help organizations lower risk and expense.
- ▶ Master Content for Master Data Management Server: Connect to enterprise content. Create and manage 360 degree trusted content views to enrich master data initiatives.
- ▶ OmniFind® Enterprise Edition: Drive users to the content that matters through knowledge driven search.

3.6 Summary

IBM Enterprise Content Management helps companies make better decisions faster through four product areas:

- ▶ **Essential ECM:** Essential Enterprise Content Management infrastructure software for capture, imaging, office document management, and enterprise report management.
- ▶ **Advanced case management:** Software powered by collaboration, content, process, analytics, and business rules to provide an advanced 360-degree case view.
- ▶ **Information life cycle governance:** Integrated software for content analytics for assessment, collection and archiving, classification, records management, and ediscovery.
- ▶ **Trusted content analytics:** Software to help you find, know, use, and exploit all of your enterprise content.

Table 3-1, Table 3-2, Table 3-3, and Table 3-4 summarize the IBM Enterprise Content Management offerings for each area.

Table 3-1 Essential ECM products and offerings

Product name	Description
IBM FileNet Content Manager	Essential for production imaging, office document management, and content consolidation and standardization. It is the core of the FileNet P8 Platform and is also offered in a Starter Pack.
IBM Content Manager	Essential for production imaging on multiple platforms, office document management, and content integration. It is offered for z/OS and in a Starter Pack.
IBM Content Manager OnDemand	Adds enterprise report management to your content management system. It integrates with FileNet P8 Platform.
IBM IBM Datacap	Provides premium data and document capture. Datacap was acquired by IBM in 2010.
IBM FileNet Capture	Provides imaging capture to get critical content into FileNet P8 repositories.

Table 3-2 Advanced case management product

Product name	Description
IBM Case Manager	Integrates content, process and people (capabilities) to deliver the next-generation of case management applications.

Table 3-3 Information life cycle governance products and offerings

Product name	Description
Content Analytics for Assessment	
IBM Content Analytics	A robust software with user-friendly interface for exploratory analytics of unstructured content.
IBM Classification Module	Automates classification of unstructured content by analyzing full text of documents and emails.
IBM Content Collector for File Systems	Collects and archive content from file systems.
Content Collection and Archiving	
IBM Content Collector for Email	Collect and archive emails.
IBM Content Collector for File Systems	Controls document and content on network file shares.
IBM Content Collector for Microsoft SharePoint	Collects SharePoint documents into a centrally managed archive.
IBM Content Collector for SAP	Provides SAP data and document archiving, SAP content-enabling, and complementary process management for SAP environments.
IBM Content Collector Discovery Analytics	Complete collection, archiving, eDiscovery and analytics capabilities in one solution.
Advanced Classification	
IBM Classification Module	Automates classification of unstructured content by analyzing full text of documents and emails.
Records Management	
IBM Enterprise Records (formerly called FileNet Records Manager)	Securely captures, declares, classifies, stores and disposes of electronic and physical records.

Product name	Description
IBM Records Manager	A records management engine for electronic records management in applications.
eDiscovery Search and Analytics	
IBM eDiscovery Manager	Provide IT users the ability to respond easily to discovery requests. It also provides comprehensive audit tracking to support authenticity and prove chain of custody.
IBM eDiscovery Analyzer	Provide the detailed analysis and review capabilities that enable early case assessment and provides the opportunity to reduce case costs.
IBM Content Collector	Enable collecting information from various source systems and applications (including email systems, file systems, Microsoft Sharepoint) and archive to a central repository.
IBM Classification Module	Automatically and intelligently classify unstructured content to help organizations lower risk and expense.

Table 3-4 Trusted Content Analytics

Product name	Description
IBM Content Analytics	Discover and use content to derive unexpected business insights, take action, and share with other users and systems.
IBM Content Analytics for Assessment	Know your content. Analyze what you have. Decommission the unnecessary. Preserve what matters.
IBM Classification Module	Automatically and intelligently classify unstructured content to help organizations lower risk and expense.
Master Content for Master Data Management Server	Connect to enterprise content. Create and manage 360 degree trusted content views to enrich master data initiatives.
OmniFind Enterprise Edition	Drive users to the content that matters through knowledge driven search.

For more information about IBM Enterprise Content Management products, see:
<http://www.ibm.com/software/data/content-management>

This paper focuses on the integration between the Information Archive device and Enterprise Content Management products. The remainder of the chapters are organized as follows:

- ▶ Chapter 4, “Integrating IBM FileNet P8 with the Information Archive device” on page 37 explains the integration of IBM FileNet P8 with the Information Archive device.
- ▶ Chapter 5, “Integrating IBM Content Manager with the Information Archive device” on page 71 explains the integration of IBM Content Manager with the Information Archive device.
- ▶ Chapter 6, “Integrating IBM Content Manager OnDemand with the Information Archive device” on page 103 explains the integration of IBM Content Manager OnDemand with the Information Archive device.
- ▶ Chapter 7, “IBM Information Archive for Email, Files, and eDiscovery” on page 131 discusses this preintegrated, preconfigured, and preinstalled solution.



Integrating IBM FileNet P8 with the Information Archive device

In this chapter, we discuss the integration between the IBM Information Archive device and IBM FileNet P8. We describe the required configuration to enable FileNet P8 to use the Information Archive device as an archive storage system with hierarchical storage devices and retention capabilities.

This chapter includes the following topics:

- ▶ Introduction
- ▶ IBM FileNet P8 Platform overview
- ▶ Integration architecture
- ▶ Integration implementation example
- ▶ Hints and tips
- ▶ Summary

4.1 Introduction

Today's organizations face the challenge of archiving information for long-term retention. For some organizations, IBM FileNet is already part of their overall enterprise content management strategy. For others, there is no such strategy or solution.

From a software perspective, you can use IBM Content Collector to collect email, files, and application data and store them in an IBM FileNet content repository. This solution can also be used in combination with other products, such as IBM Classification Module, IBM Content Analytics, IBM Enterprise Records, IBM eDiscovery Manager, and IBM eDiscovery Analyzer, to provide advanced content classification, analysis, records management, and eDiscovery capabilities.

From a hardware perspective, for long-term, retention managed storage, IBM FileNet integrates with the Information Archive device to provide an end-to-end archiving infrastructure. In addition, organizations with an existing enterprise content management repository, such as IBM FileNet, can now require a compliance storage solution to house their current content collections.

4.2 IBM FileNet P8 Platform overview

IBM FileNet P8 Platform (FileNet P8) is a collection of tightly integrated components that are bundled together under a common platform. The broad functionality provided by these integrated components constitutes an enterprise content and process management platform. Key elements of this platform are a metadata repository, a process management repository, a readily-available user interface for accessing content and process elements, and a storage framework that can support a wide range of storage devices and platforms.

To provide these services, the FileNet P8 relies on the following core components:

- ▶ Content Engine
- ▶ Process Engine
- ▶ Workplace XT

All other add-on products are built on the foundation that these components provide.

The *Content Engine* is the engine that is responsible for storing and retrieving all content within a FileNet P8 system. It provides a series of services for creating, retrieving, updating, deleting, and securing content within the environment. In

addition, it provides interfaces for handling event-based actions, documenting life cycle, and integrating with various storage systems, such as the Information Archive device.

The remainder of this chapter focuses on the Content Engine component of FileNet P8 and how it integrates with the Information Archive device.

4.3 Integration architecture

The Information Archive device is the latest generation of information retention solutions. It is designed as a universal storage repository for all types of content (structured or unstructured) to help organizations of any size address complete information retention needs.

The flow of a document from FileNet P8 to the Information Archive device (shown in Figure 4-1) is as follows:

1. Documents are created by FileNet P8 client applications.
2. Documents submitted to the FileNet P8 system are cataloged by the Content Engine (Content Engine Server).
3. The Content Engine persists the document into a storage system, such as the Information Archive device, which comes bundled with IBM Tivoli Storage Manager and IBM System Storage Archive Manager software.

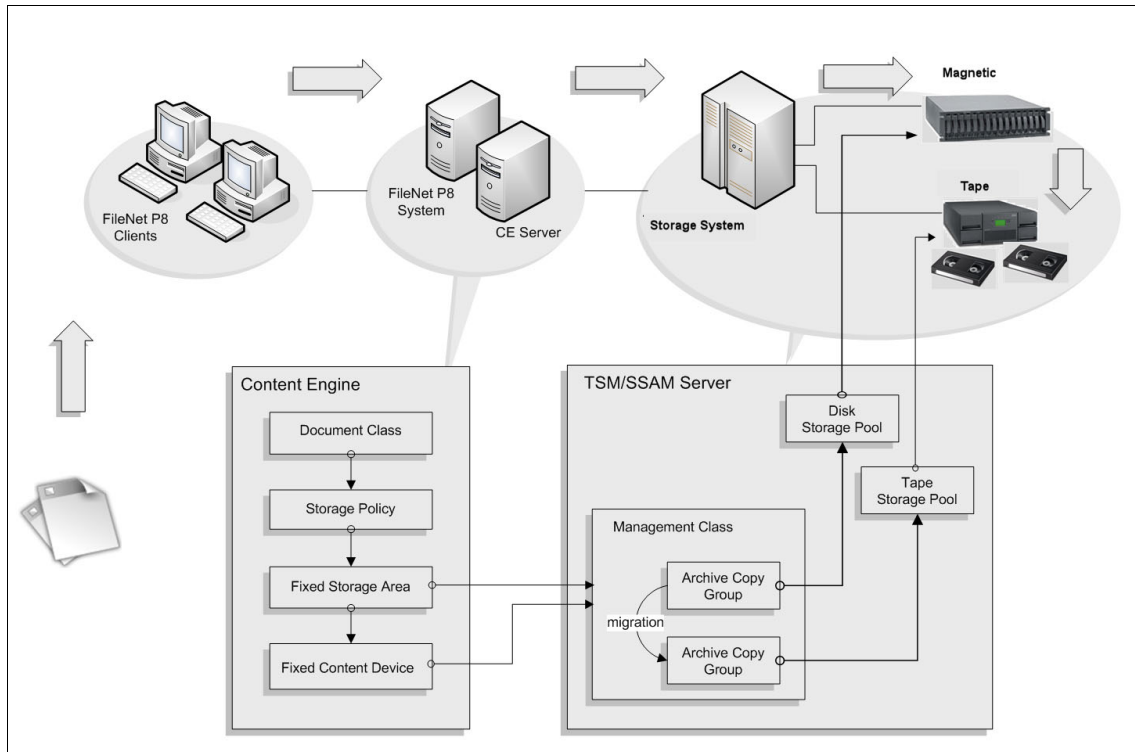


Figure 4-1 FileNet P8: The Information Archive device integration

The remainder of this section describes:

- ▶ A deployment model for FileNet P8 and the Information Archive device components
- ▶ A data model of FileNet P8 and Information Archive objects
- ▶ Retention
- ▶ Hierarchical storage

4.3.1 Deployment model

FileNet P8 uses the Tivoli Storage Manager Archive Client API to communicate with the Tivoli Storage Manager server component of the Information Archive device. Thus, it requires that the Tivoli Storage Manager client software is installed on every system that hosts a FileNet P8 Content Engine.

Figure 4-2 depicts a physical deployment of FileNet P8 and the Information Archive device with two Content Engine servers on different FileNet P8 host systems storing objects on the same Information Archive server.

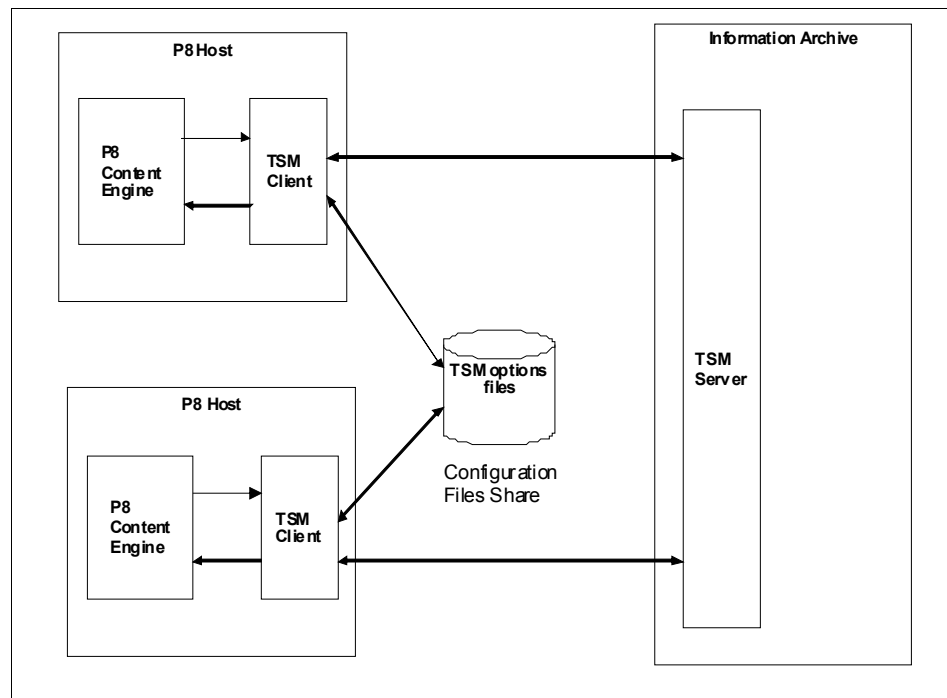


Figure 4-2 FileNet P8: The Information Archive device deployment model

The deployment model consists of the following components:

- ▶ FileNet P8 host

The FileNet P8 host is the physical server on which the FileNet P8 Content Engine runs, for example an IBM pSeries® server running AIX. As depicted in Figure 4-2, a Tivoli Storage Manager client must be installed on each FileNet P8 host system.

In addition, the application server on which Content Engine is deployed must be modified to reference the Tivoli Storage Manager API native libraries that are packaged with the Content Engine software. We discuss Tivoli Storage Manager client installation further in 4.4.2, “Install the Tivoli Storage Manager client” on page 52.

- ▶ Tivoli Storage Manager options files

The Tivoli Storage Manager options files hold the connection parameters that are used by the Tivoli Storage Manager client to communicate with the

Information Archive device. The Tivoli Storage Manager options files are created and managed by FileNet P8 in a shared directory that is accessible by all the Content Engine instances in a FileNet P8 domain. Thus, the operating system user who starts the Content Engine application server, such as WebSphere®, must have read/write access to this directory.

This directory location is a property of the Tivoli Storage Manager fixed content device (FCD) object of FileNet P8. We discuss the Tivoli Storage Manager options files in more detail in “Tivoli Storage Manager options files” on page 53.

► The Information Archive device

The Information Archive device has a Tivoli Storage Manager server component that manages the storage and retrieval of objects that are stored on the Information Archive device for application clients, such as FileNet P8, that use the Tivoli Storage Manager client API.

4.3.2 Data model

Figure 4-3 depicts the cardinality relationships between FileNet P8 objects and the Information Archive device objects.

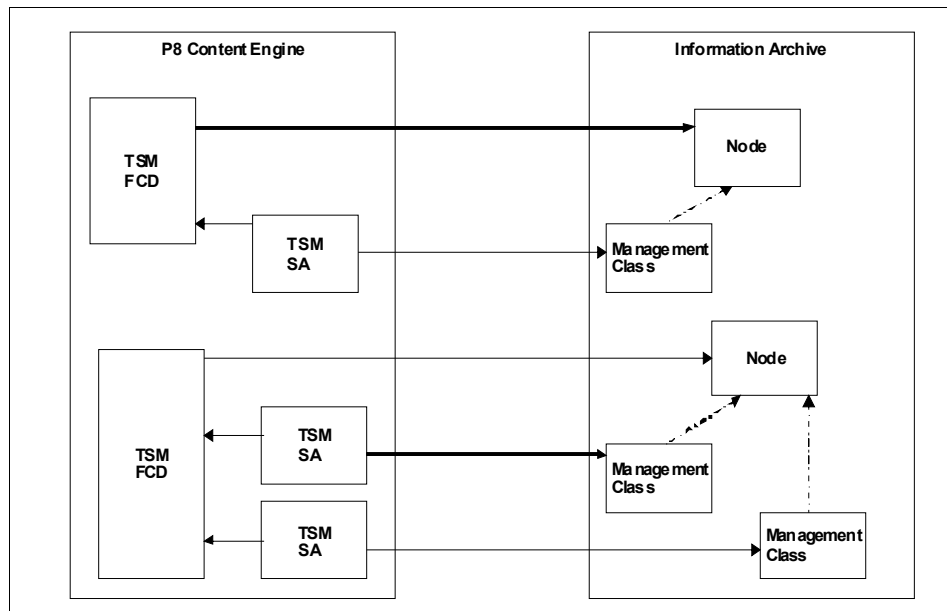


Figure 4-3 FileNet P8: The Information Archive device data model cardinality relationships

Multiple Tivoli Storage Manager FCDs can be defined in a FileNet P8 Content Engine. Each FCD can have one or more Tivoli Storage Manager storage areas associated with it. Documents created by FileNet P8 applications are stored in a specific storage area. When documents are stored in a storage area that is associated with a Tivoli Storage Manager FCD, they are transferred to the Information Archive device for persistent storage and are assigned a retention that is governed by the Information Archive device management class that is associated with the Tivoli Storage Manager storage area.

When multiple Tivoli Storage Manager FCDs are defined, each FCD must reference a distinct node on the Information Archive device. Management classes on the Information Archive device are associated with a node through a set of other Tivoli Storage Manager objects that are defined on the Information Archive device, as shown in Figure 4-4.

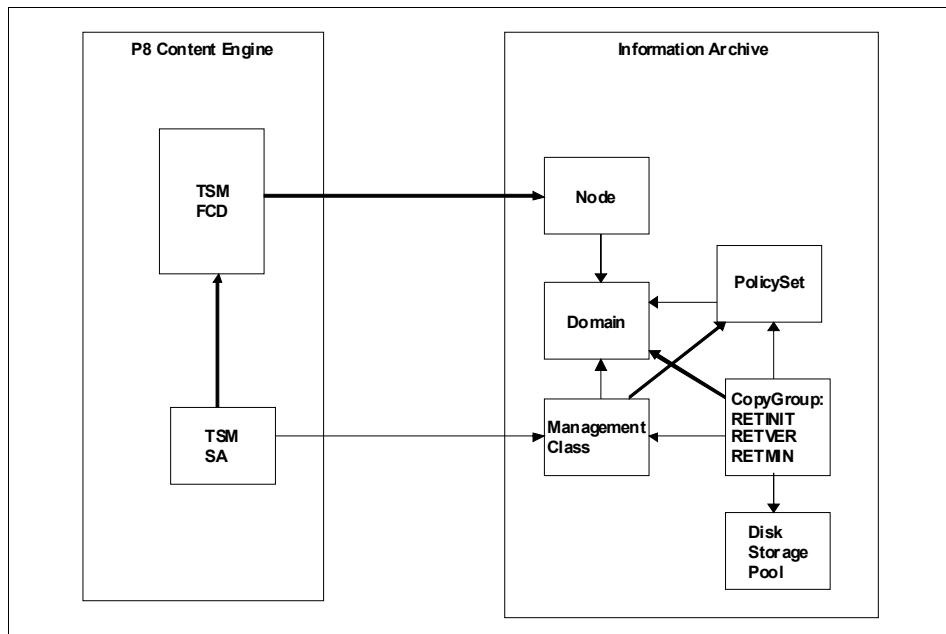


Figure 4-4 FileNet P8: The Information Archive device data model object relationships

We describe each of these Tivoli Storage Manager objects in more detail in “Create storage and policy configurations” on page 48.

When FileNet P8 documents are transferred to the Information Archive device for persistent storage, they are stored in the disk storage pool that is associated with the node. From the disk storage pool, documents can be migrated further to other storage pools as described in 4.3.4, “Hierarchical storage” on page 47. This migration to other storage pools is transparent to FileNet P8.

Although the Tivoli Storage Manager storage area uses an Information Archive device management class to set the retention of FileNet P8 documents, the retention parameters on the Information Archive device actually are set on a different Tivoli Storage Manager object, the *copygroup*, that references the management class.

We discuss retention further in the next section.

4.3.3 Retention

Retention relates to when a document can be deleted. Documents cannot be deleted until the retention period that is associated with the document has passed.

When using FileNet P8 with the Information Archive device, you can enforce retention using either the Information Archive device or FileNet P8.

Retention enforced by the Information Archive device

The Content Engine supports *static* retention, which is applied when a document is checked in and which prevents deletion until after the retention period expires. When connected to the Information Archive device, the Content Engine can use either the *chronological* or *event-based retention* scheme that is used by the Information Archive device to enforce this retention.

The Content Engine uses the behavior of the Tivoli Storage Manager management class to control when a document can be deleted. Although either scheme can be used, the preferred practice is to use event-based retention, because it does not require that a hold be placed on a document when it is created. This method results in less load on the Tivoli Storage Manager server and better performance than chronological retention.

The Content Engine always remains in control of the document deletion process. So, even though the Information Archive device supports the automatic deletion of objects after their retention period has expired, the Content Engine will not allow automatic deletion until after the retention period has expired *and* until a user specifically deletes the associated document from the Content Engine. When both conditions are met, the Content Engine relinquishes control of the Tivoli Storage Manager object, allowing the Information Archive device to delete it automatically.

Note that deletion on the Information Archive device typically happens as a background process. Thus, requests by the Content Engine to delete an object on the Information Archive device might not occur immediately. You can set up automatic expiration processing on the Information Archive device using the

EXPINTERVAL server option. The Information Archive device inventory expiration process removes files based on the archive copy groups of the management classes to which the files are bound.

To control retention while suppressing automatic deletion, Content Engine uses the specific parameters of the retention schemes that are available in the Information Archive device, as shown in Table 4-1.

Table 4-1 Retention parameters

Parameter and purpose	Value for chronological retention	Value for event-based retention
RETINIT determines when to initiate the retention period defined in the RETVER attribute	CREATION (start counting when the object is archived)	EVENT (start counting when the event is triggered)
RETVER is the number of days to retain the archived object after retention is initiated	0 to 30,000 days, or NOLIMIT	0 to 30,000 days
RETMIN is the minimum number of days to retain the archived object	Not applicable	n, where n represents a number of days

Chronological retention

A chronological management class uses the RETVER parameter to protect a Tivoli Storage Manager object. The RETVER parameter defines the number of days that the object should be retained, starting from the date that the object is stored on the Information Archive device.

Tivoli Storage Manager automatically deletes objects when their retention period (RETVER) on the Information Archive device has expired. To enable the Content Engine to retain control of the objects with chronological retention, the Content Engine places a hold on the object after creation. Later, when the document is finally eligible for deletion, after the RETVER has expired and after a user deletes the associated document, the Content Engine releases the hold and allows the Tivoli Storage Manager automatic deletion process to delete the object.

Placing a hold on an object uses the IBM System Storage Archive Manager component of the Information Archive device. This component is the same as Tivoli Storage Manager with the archiveretentionprotection attribute set to on.

When you create a Tivoli Storage Manager management class to support chronological retention, use the following parameters:

```
RETINIT=CREATION  
RETVER=<actual retention period>
```

Note: Chronological retention cannot be used for a zero duration retention. If you need to define a zero duration retention, use event-based retention, which we describe in the next section.

Event-based retention

An event-based retention (EBR) management class has two retention settings that combine to protect a Tivoli Storage Manager object:

- ▶ The RETMIN parameter defines the minimum retention period for the object from creation time.
- ▶ The RETVER parameter defines the retention from the time of activation.

As long as the EBR activate event is not issued, the object is protected from deletion, no matter what the value is for RETMIN. When the EBR activate event is issued, the retention period that is defined in RETVER is applied from that moment. The retention period set in RETVER can overlap with RETMIN. For example, if RETVER=0 and the activate event is issued during the RETMIN period, the object continues to be protected until RETMIN expires.

The Content Engine uses the behavior of the EBR management class to control when the Tivoli Storage Manager object can be deleted. If in Tivoli Storage Manager you set an EBR class with RETMIN=365 and REVER=0, then you can create in the Content Engine a Tivoli Storage Manager fixed storage area that uses the EBR management class. When a document is checked into this fixed storage area, the Content Engine calculates the delete date (today + 365 days) and writes it to the ContentRetentionDate attribute of the document. When a user wants to delete the document more than one year later, the Content Engine delete call sends an activate event to the Tivoli Storage Manager server. Because RETVER=0, the retention expires immediately, which allows the object to be deleted automatically by the Tivoli Storage Manager background task.

When you create a Tivoli Storage Manager management class to support event-based retention, use the following parameters:

```
RETINIT=EVENT  
RETVER=0  
RETMIN=<actual retention period>
```

Retention enforced by IBM Enterprise Records

IBM Enterprise Records is a FileNet P8 application that assumes all retention management responsibility for documents under its control. Hence, when using the Information Archive device as a repository for such documents, retention must be disabled on the Information Archive device.

You can disable retention on the Information Archive device by defining a management class on the Information Archive device using the following parameters:

```
RETINIT=EVENT  
RETVER=0  
RETMIN=0
```

Then, associate the Content Engine storage area for documents that are controlled by IBM Enterprise Records to use the Information Archive management class with no retention.

4.3.4 Hierarchical storage

The Information Archive device supports hierarchical storage capabilities to allow you to create policies so that data is stored on the type of media that best meets data longevity, access speed, and cost needs. For example, with the Information Archive device, you can attach tape devices to archive data that is accessed infrequently. Migration automates moving data from one type of media to another as media needs change and as new types of media become available in the market.

When using hierarchical storage with FileNet P8, keep in mind the following considerations:

- ▶ Always use a disk storage pool as the first storage pool for incoming documents from FileNet P8. For example, FileNet P8 documents should not be written directly from the Content Engine to a tape storage pool. This method can lead to performance problems when transferring documents from FileNet P8 to the Information Archive device.

Instead, set up additional storage pools and migration policies from the incoming disk storage pool for these other media types, such as tape.

- ▶ Use the Content Engine content cache to improve retrieval performance from tape.
 - The performance improvement applies to the initial retrieval of the content (especially large content in a Content Engine cluster environment) and to subsequent retrievals (content will be cached on fast storage).

- The Content Cache has its own inherent timeout mechanism. If a retrieval from tape takes longer than this internal timeout, the content cache returns a message to try the retrieval again later. In the meantime, the content cache continues to try to retrieve the tape content so that the content is available in its cache area the next time a user attempts to retrieve it.
- ▶ When using tape, adjust the MOUNTWAIT and MOUNTRETENTION of the device class for the tape library and IDLETIMEOUT for the system as discussed in “Modify the Information Archive device server parameters” on page 50. Adjusting these parameters prevents FileNet P8 users from waiting for an unreasonable length of time when retrieving documents stored on tape.

4.4 Integration implementation example

The implementation example in this section describes how to set up a storage configuration on an Information Archive device that is integrated with FileNet P8. The high-level steps are as follows:

1. Configure the Information Archive device:
 - a. Create storage and policy configurations.
 - b. Modify the Information Archive device server parameters.
2. Install the Tivoli Storage Manager client.
3. Configure the Content Engine:
 - IBM Tivoli Storage Manager FCD
 - Create a fixed storage area
 - Configure the content cache
4. Test the integration.

4.4.1 Configure the Information Archive device

To show how to configure information archive, in this section we provide example commands that can be executed from a Tivoli Storage Manager command line client to create and update the necessary Tivoli Storage Manager server objects required for our example FileNet P8 system.

Create storage and policy configurations

To create storage and policy configuration, perform these steps:

1. From the Information Archive device, run the Tivoli Storage Manager administration command line client:

```
dsmadm -server =collection name
```

2. Create a storage pool by entering the following command:

```
DEFINE STGPOOL IA_P8_POOL DISK DESCRIPTION="IA P8 Storage Pool"
```

Note: The storage pool that is used by FileNet P8 must be a disk storage pool. You can define other storage pool types, such as a tape pool, behind this disk storage pool in the Information Archive device using the NEXTSTGPOOL option to the DEFINE STGPOOL command.

3. Assign a storage volume of 100 MB (or other desired value) to the storage pool:

```
DEFINE VOLUME IA_P8_POOL P8DISK FORMATSIZE=100 WAIT=YES
```

4. Create a policy domain and a policy:

```
DEFINE DOMAIN IA_P8_DOMAIN DESCRIPTION="IA P8 Policy Domain"  
DEFINE POLICYSET IA_P8_DOMAIN IA_P8_POLICY DESCRIPTION="IA P8 Policy Set"
```

5. Create a management class and archive copy group with 30 day retention:

```
DEFINE MGMTCLASS IA_P8_DOMAIN IA_P8_POLICY IA_30_DAY_RET  
DESCRIPTION="IA Management Class with 30 day retention."
```

```
DEFINE COPYGROUP IA_P8_DOMAIN IA_P8_POLICY IA_30_DAY_RET  
DESTINATION=IA_P8_POOL TYPE=ARCHIVE RETINIT=EVENT RETVER=0 RETMIN=30
```

6. Create a management class and archive copy group with 90 day retention:

```
DEFINE MGMTCLASS IA_P8_DOMAIN IA_P8_POLICY IA_90_DAY_RET  
DESCRIPTION="IA Management Class with 90 day retention."
```

```
DEFINE COPYGROUP IA_P8_DOMAIN IA_P8_POLICY IA_90_DAY_RET  
DESTINATION=IA_P8_POOL TYPE=ARCHIVE RETINIT=EVENT RETVER=0 RETMIN=90
```

7. Create a management class and archive copy group with no retention:

```
DEFINE MGMTCLASS IA_P8_DOMAIN IA_P8_POLICY IA_NO_RET DESCRIPTION="IA  
Management Class with no retention."
```

```
DEFINE COPYGROUP IA_P8_DOMAIN IA_P8_POLICY IA_NO_RET  
DESTINATION=IA_P8_POOL TYPE=ARCHIVE RETINIT=EVENT RETVER=0 RETMIN=0
```

8. Set the IA_NO_RET management class as default:

```
ASSIGN DEFMGMTCLASS IA_P8_DOMAIN IA_P8_POLICY IA_NO_RET
```

9. Validate the policy set and activate the policy set:

```
VALIDATE POLICYSET IA_P8_DOMAIN IA_P8_POLICY  
ACTIVATE POLICYSET IA_P8_DOMAIN IA_P8_POLICY
```

10. Register a node to allow FileNet P8 to connect to the Information Archive device:

```
REGISTER NODE p8 p8password DOMAIN=IA_P8_DOMAIN
```

Modify the Information Archive device server parameters

Validate or update the following properties on the Information Archive device when using the Information Archive device with FileNet P8:

► **Archive Retention Protection**

The IBM System Storage Archive Manager software component of the Information Archive device automatically sets this attribute to on. You can verify that the setting is correct using the **query status** command and noting the value for the Archive Retention Protection property. This value corresponds to the Archive Protection Flag property on the FileNet P8 Tivoli Storage Manager FCD.

► **Node**

The node name that you register to allow FileNet P8 to connect to the Information Archive device must have the following capabilities:

- Read/write access to the appropriate Information Archive storage devices.
- Ability to make multiple concurrent connections to the Information Archive device.
- The password expiration period for the client node can be set to 0 (never expires), which allows the system administrator to change the password on both the Information Archive device and the Tivoli Storage Manager FCD at their convenience (the Content Engine cannot access the Information Archive device if the client node password expires).

For example:

```
update node p8 passexp=0
```

- If using hierarchical storage where some FileNet P8 content might be migrated to tape, increase the maximum number of client mount points per node so that multiple Content Engine servers can access the tape device using the same node.

Also, allow mount points to be released during a session so that they can be made available to other sessions that need them as quickly as possible.

For example:

```
update node p8 MAXNUMMP=999 KEEPMP=No
```

► **MAXSESSIONS**

Adjust the maximum number of simultaneous client sessions to the Information Archive device, if required. FileNet P8 uses the FCP Pool Max In

Use property of the Tivoli Storage Manager FCD to control the number of concurrent connections to the Information Archive device for each Content Engine server. Thus, the MAXSESSIONS value on the Information Archive device should be greater than the FCP Pool Max In Use property multiplied by the number of Content Engine servers.

For example:

```
setopt maxsessions 2048
```

► TXNGROUPMAX

This parameter specifies the number of files transferred as a group between Tivoli Storage Manager commit points. Content Engine commits all the content elements for a document in a single Tivoli Storage Manager transaction. Thus, if your documents can contain a large number of content elements, then consider increasing this parameter on the Information Archive device. The typical default for this parameter is 256 with a maximum allowed value of 65000.

For example:

```
setopt txngroupmax 4096
```

► MOUNTWAIT

If using tape, adjust this parameter on the device class to reduce the time the Information Archive device waits for a volume to be mounted in the tape drive. This parameter affects how long a FileNet P8 user waits for content that has migrated to tape.

For example, assuming you have an existing device class called P8TAPES, this command changes the wait time to 1 minute:

```
update devclass P8TAPES mountwait=1
```

► MOUNTRETENTION

If using tape, adjust this parameter on the device class to keep an idle tape mounted for a period of time. This adjustment can improve access to FileNet P8 content that is stored on the same tape.

For example, assuming that you have an existing device class called P8TAPES, the following command changes the time that the tape remains mounted while idle to 5 minutes:

```
update devclass P8TAPES mountretention=5
```

► IDLETIMEOUT

Adjust this system parameter to reduce the time that FileNet P8 waits for a response from the Information Archive device. If retrieving content from tape, the smaller of IDLETIMEOUT or MOUNTWAIT controls how long a FileNet P8

user waits before receiving a message to try again later to retrieve the content.

For example, the following command sets the idle timeout to 1 minute:

```
setopt idletimeout 1
```

4.4.2 Install the Tivoli Storage Manager client

The Tivoli Storage Manager client must be installed on each system that hosts a FileNet P8 Content Engine. To install and configure the Tivoli Storage Manager client, follow the instructions in the information center:

<http://publib.boulder.ibm.com/infocenter/p8docs/v4r5m1/topic/com.ibm.p8.install.doc/p8pin230.htm>

In addition, the application server on which the Content Engine is deployed must be modified to reference the Tivoli Storage Manager API native libraries that are laid down when you run the Content Engine software installer. These instructions are also in the information center.

For example, the Content Engine installer places the Tivoli Storage Manager API native libraries in the following directory for a typical Windows system installation:

```
C:\Program Files\IBM\FileNet\ContentEngine\tsm100
```

The Tivoli Storage Manager API native libraries are separate from the Tivoli Storage Manager client. They are bundled as part of the Content Engine software, whereas you must download the Tivoli Storage Manager client from an IBM support site, such as the following website for a Tivoli Storage Manager 6.1 client:

<http://www.ibm.com/support/docview.wss?uid=swg24022622>

If you use an application server specific deployment mechanism, such as IBM WebSphere Application Server Network Deployment, to deploy Content Engine to a cluster, then make sure you complete the following tasks:

1. Copy the Tivoli Storage Manager API native library folder to each system where you run a Content Engine server in the cluster.
2. Ensure that the directory path for the Tivoli Storage Manager API native library folder is identical on each system. For example, always copy it to the following folder if that is the folder where the Content Engine installer initially installed it:

```
C:\Program Files\IBM\FileNet\ContentEngine\tsm100
```

3. Install the Tivoli Storage Manager client on each of these same systems.

Tivoli Storage Manager options files

The Tivoli Storage Manager client uses Tivoli Storage Manager options files to manage the connection to the Information Archive device. FileNet P8 creates and maintains these options files automatically in the directory specified by the Configuration Files Share property of the Tivoli Storage Manager FCD.

The directory that is specified by the Configuration Files Share property must be located on a network share that is accessible by all Content Engine servers in the FileNet P8 domain, using the same file path. This directory must also have read/write access for the operating system user who runs the Content Engine application server.

Windows systems

When the Content Engine is installed on Windows systems, the Tivoli Storage Manager client on each system for each Content Engine server uses the Tivoli Storage Manager options file located in the Configuration Files Share directory.

The Content Engine and the Windows system Tivoli Storage Manager client use the following options files:

► *FCDID.opt*

This file contains all the configuration parameters that are used to connect to the Information Archive device. *FCDID* is the ID of the Tivoli Storage Manager FCD. For example:

19528730-E2CE-48AE-994B-593B8A46C128.opt

When changes are made to the Tivoli Storage Manager FCD, a backup of this options file is created that contains the previous values. The backup file uses the naming convention *FCDID.opt.xxxx*, where *xxxx* is the hexadecimal representation of the current time. For example:

19528730-E2CE-48AE-994B-593B8A46C128.opt.12ba85165e0

► *IBMFileNetP8.opt*

This file is typically empty but is required to be present for the Tivoli Storage Manager client. The Content Engine does not update this file, so no backup copies are made when the Tivoli Storage Manager FCD is updated.

UNIX systems

When the Content Engine is installed on a UNIX system, the Tivoli Storage Manager client on each UNIX system uses a local copy of the Tivoli Storage Manager options files under the Tivoli Storage Manager client installation directory. In this case, the Configuration Files Share directory holds a merged copy of all the Tivoli Storage Manager options files that are located on each UNIX system. This merged copy is used to keep all the local Tivoli Storage Manager

options files in sync with parameter changes made through the Tivoli Storage Manager FCD.

The location of the local Tivoli Storage Manager options file on each system is specified by the `DSMI Directory` property on the Tivoli Storage Manager FCD and must be the same on all UNIX hosts. For example:

```
/usr/tivoli/tsm/client/ba/bin
```

The following options files are used by the Content Engine and the UNIX system Tivoli Storage Manager client:

- ▶ `dsm.sys`

This is the local file on each UNIX system that contains all the configuration parameters that are used to connect to the Information Archive device. It is located under the Tivoli Storage Manager client installation directory. For example:

```
/usr/tivoli/tsm/client/ba/bin/dsm.sys
```

- ▶ `FCDID.opt`

This file contains the server name for the Tivoli Storage Manager connection referenced in the local `dsm.sys` file. It is located in the `Configuration Files Share` directory where `FCDID` is the ID of the Tivoli Storage Manager FCD. For example:

```
19528730-E2CE-48AE-994B-593B8A46C128.opt
```

This file is created when the Tivoli Storage Manager FCD is created. It is not updated, so no backups are maintained for it.

- ▶ `IBMFileNetP8.dsm.sys`

This is the master `dsm.sys` file that holds the merged copy of all local `dsm.sys` files for each UNIX system. It is located in the `Configuration Files Share` directory.

When changes are made to the Tivoli Storage Manager FCD, a backup of this file is created containing the previous values. The backup file uses the naming convention `IBMFileNetP8.dsm.sys.xxxx`, where `xxxx` is the hexadecimal representation of the current time. For example:

```
IBMFileNetP8.dsm.sys.120aab0ef2b
```

4.4.3 Configure the Content Engine

In this section, we discuss the FileNet P8 objects that must be configured when integrating with the Information Archive device. We configure FileNet P8 in this section using FileNet Enterprise Manager.

IBM Tivoli Storage Manager FCD

FileNet P8 uses FCD objects to communicate with external repositories such as the Information Archive device. For the Information Archive device integration, create an FCD of type *IBM Tivoli Storage Manager (FCD)* by running the New Fixed Content Device wizard on FileNet Enterprise Manager.

You can have multiple Tivoli Storage Manager FCDs configured for the same Information Archive server. However, each FCD must use a different Node Name to establish a client session with the Information Archive device.

You can find the full list of configurable properties for the Tivoli Storage Manager FCD at:

http://publib.boulder.ibm.com/infocenter/p8docs/v4r5m1/topic/com.ibm.p8.doc/ce_help/aboutem/aem_dom_props_fcd_general.htm#tivoli

We discuss a few of these properties further in this section to provide insight into how they affect the integration with the Information Archive device:

- ▶ FCP Pool Max in Use

Controls the maximum number of concurrent connections to the Information Archive device for a single Tivoli Storage Manager FCD in a single Content Engine server. The Information Archive device uses its MAXSESSIONS property to control the number of sessions across all client nodes. Hence, the number of sessions originating from all Content Engine servers must be under this MAXSESSIONS limit.

To calculate the maximum number of sessions that might originate from FileNet P8, use the following formula:

(Number of Content Engine servers) x (FCP Pool Max in Use for each Tivoli Storage Manager FCD added together)

- ▶ Configuration Files Share

Specifies the location for the generated Tivoli Storage Manager options files. This share must be accessible to all Content Engine instances of the FileNet P8 domain (using the same path), and must be backed up on a regular basis.

- ▶ IP address

The IP address or computer name of the Tivoli Storage Manager server. This property corresponds to the TCPSERVERADDRESS option in the Tivoli Storage Manager options file.

- ▶ Archive Protection Flag

The archive protection flag setting of the FCD must match the Archive Retention Protection value defined on the Tivoli Storage Manager server (set to true if protection is on, false if it is off). This property corresponds to the

ENABLEARCHIVERETENTIONPROTECTION option in the Tivoli Storage Manager options file. When integrating with the Information Archive device, this value should always be true.

► Node Name

Tivoli Storage Manager client node to log on and access the Information Archive server. Each Tivoli Storage Manager FCD that is connected to the same Information Archive server must use a different node name.

► Filespace Name

A logical space on the Information Archive device used to contain the data for this node. The file space name must begin with a forward slash (/). Use the file space to organize and manage the data on the Information Archive device for this node.

It is not required that file space names be unique across Tivoli Storage Manager FCDs. However, if you want to associate a file space to a particular Tivoli Storage Manager node, provide a different file space name for each Tivoli Storage Manager FCD.

► DSMI Directory

This value is set only on UNIX systems and must point to the Tivoli Storage Manager client installation on the local computer that contains the `dsm.sys` file. For example:

```
/opt/tivoli/tsm/client/ba/bin
```

► Concurrent Tape Readers

Controls how many users can be waiting for tape to be mounted. When the value is exceeded, any user request to retrieve content from tape will be rejected by the Content Engine.

Create a fixed storage area

The fixed storage area in the Content Engine is the area where content is uploaded by FileNet P8 applications and where content resides until it is transferred to the Information Archive device.

The storage area is also where retention is specified for the Information Archive device by associating the storage area to an Information Archive management class. FileNet P8 documents stored in this storage area are assigned a retention value that corresponds to the retention settings of the Information Archive management class.

Create a storage area with no retention

To create a storage area with no retention, choose a management class that you created on the Information Archive device with zero retention. FileNet P8 documents stored in this storage area have no retention enforcement by the Information Archive device. Follow these steps:

1. On FileNet Enterprise Manager, run the Create a Storage Area wizard and give a name to the storage area, for example ia_0day_sa as shown in Figure 4-5.

Create a Storage Area

Name and Describe the Storage Area
Enter the Storage Area name and description. The name identifies the Storage Area and must be unique.

Name: ia_0day_sa

Description: ia_0day_sa

Existing Names: Default Database Storage Area

< Back Next > Cancel Help

Figure 4-5 Storage area with no retention

- Specify the staging area path and choose the management class that you created in “Create storage and policy configurations” on page 48 that has no retention, as shown in Figure 4-6.

Staging area path: If you have multiple Content Engine servers, the staging area path must be on a network share that can be written to by all Content Engine servers using the same file path.

Create a Storage Area

Fixed Storage Area
Select the Fixed Content Device for the Fixed Storage Area.

Fixed Content Device: IA_FCD

Staging Area Path: C:\p8_ia\ia_0day_sa

Management Class:

Management Class	Retention Scheme	RETMIN	RETVER	CE Retention
IA_30_DAY_RET	Event Based	30 day(s)	0 day(s)	30 day(s)
IA_90_DAY_RET	Event Based	90 day(s)	0 day(s)	90 day(s)
IA_NO_RET	Event Based	0 day(s)	0 day(s)	0 day(s)

< Back Next > Cancel Help

Figure 4-6 Management class with no retention

3. Accept the default values on the next few screens, and click **Finish** to complete the wizard (Figure 4-7).

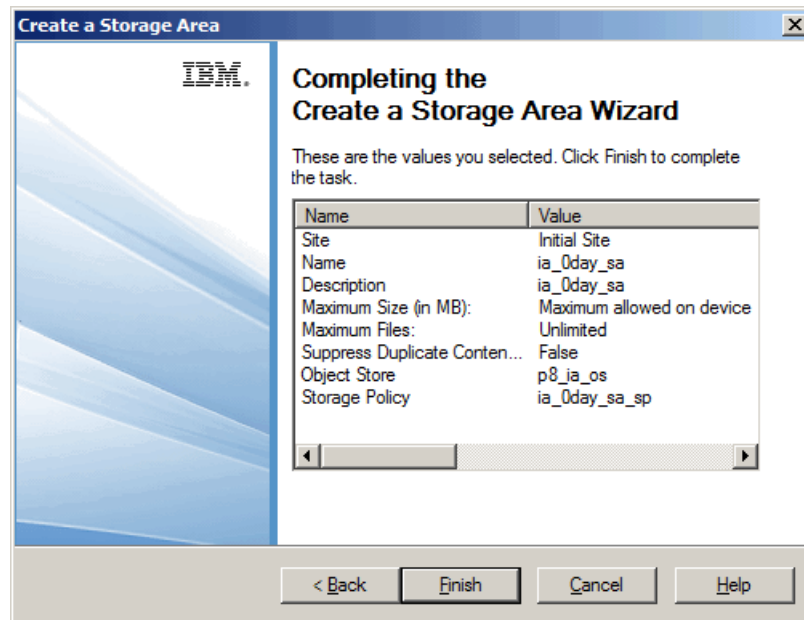


Figure 4-7 Storage area summary

Create a storage area with a 30 day retention

To create a storage area with a 30 day retention, choose a management class that you created on the Information Archive device with 30 day retention when running the Create a Storage Area wizard, as shown in Figure 4-8. FileNet P8 documents stored in this storage area cannot be deleted for at least 30 days.

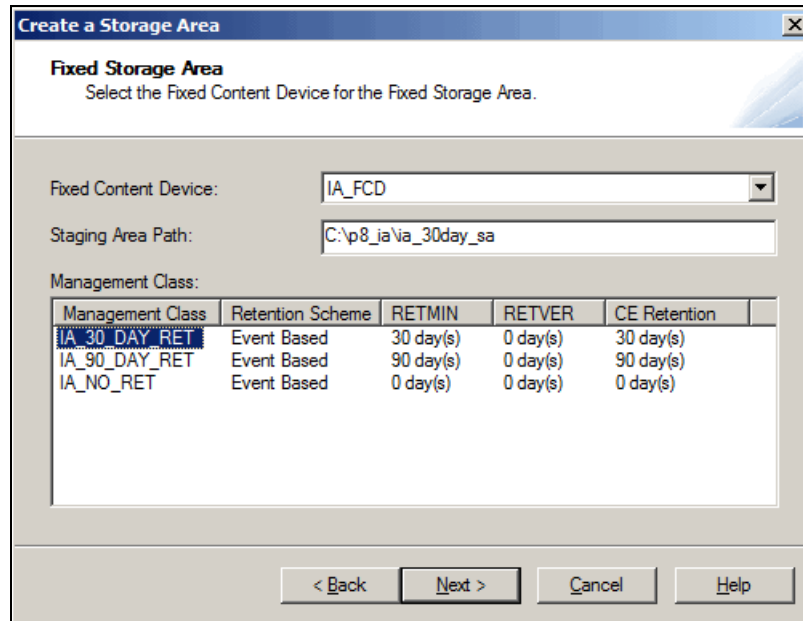


Figure 4-8 Management class with 30 day retention

Configure the content cache

If using hierarchical storage on the Information Archive device, where FileNet P8 content is migrated to a slow device such as tape, you must configure the FileNet P8 content cache.

Follow the instructions in the information center to create a content cache:

http://publib.boulder.ibm.com/infocenter/p8docs/v4r5m1/topic/com.ibm.p8.doc/ce_help/contentstores/cs_create_content_cache_store.htm

Figure 4-9 shows the summary page of the Create a Content Cache Area wizard that we used for this example.

Using multiple Content Engine servers: If you have multiple Content Engine servers, the Shared Folder for the content cache must be on a network share that can be written to by all Content Engine servers using the same file path. Thus, the operating system user who starts the Content Engine application server must have read/write access to the network share.

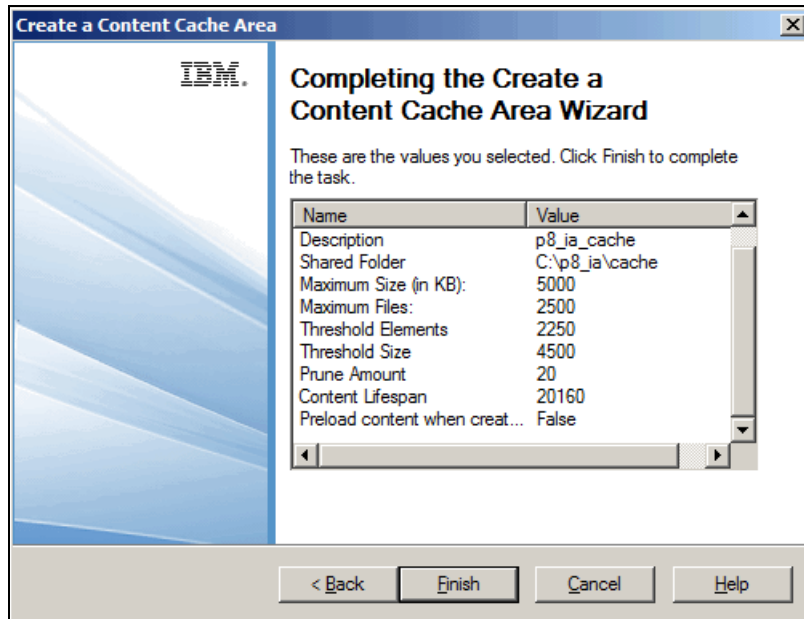


Figure 4-9 Content cache wizard summary

After creating a content cache area, assign it to a site as described in the information center:

http://publib.boulder.ibm.com/infocenter/p8docs/v4r5m1/topic/com.ibm.p8.doc/ce_help/contentstores/cs_assign_a_content_cache_area.htm

Figure 4-10 shows that the content cache that we created, p8_ia_cache, is assigned to the Initial Site. Make sure this name displays in the Selected Content Cache Area field.

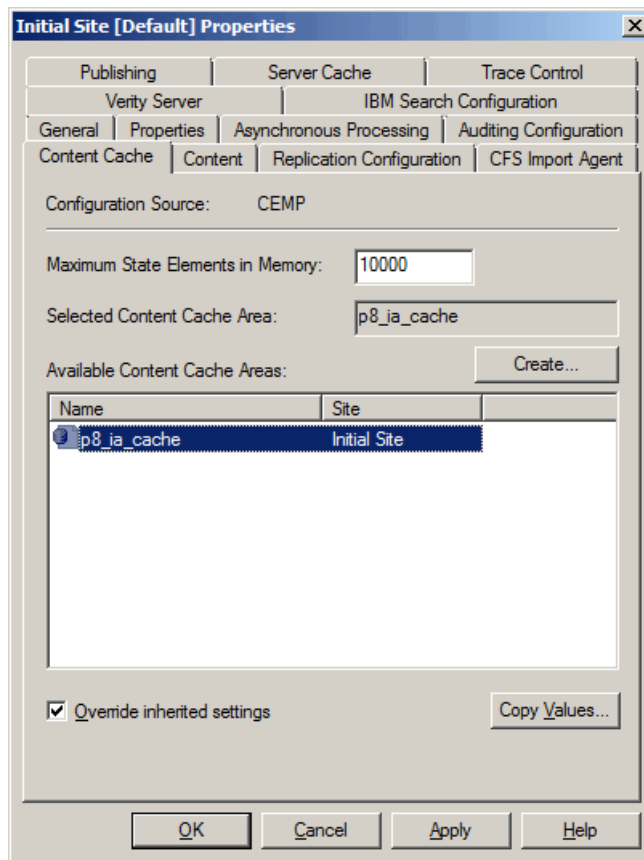


Figure 4-10 Assign content cache area to a site

Enable the content cache for all fixed storage areas that are associated with the Information Archive server. On FileNet Enterprise Manager:

1. Select the storage area Properties option.
2. On the Configuration tab and select the Allowed option in the Content Caching section, as shown in Figure 4-11.

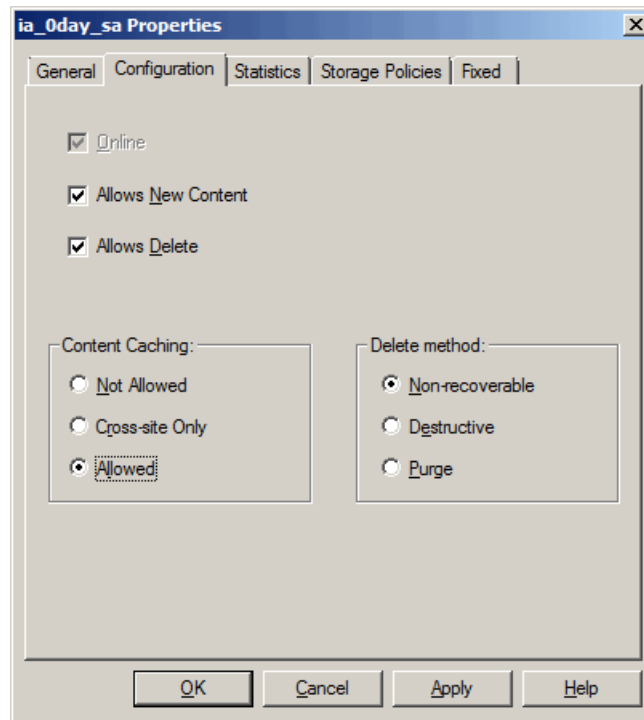


Figure 4-11 Content Caching Allowed option

4.4.4 Test the integration

Always test the integration to make sure that everything is set up correctly. To test, we add a document to each of the fixed storage areas that we created. These documents are stored on the Information Archive device.

We add the documents in this example using a storage policy that is explicitly selected to be the same as the storage area that we created earlier for the Information Archive device. A better practice is to create a FileNet P8 document class and specify the Information Archive storage policy as the default storage policy for this new class. Then, when instances of this new document class are created, they are stored in the Information Archive storage area by default.

Add a document

Use the following steps to add a document using FileNet Enterprise Manager:

1. Select your **object store** → **Root Folder** on the FileNet Enterprise Manager navigation pane.
2. Choose **Action** → **New Document**.
3. Follow the instructions in the Create New Document Wizard to create a document with content, as shown in Figure 4-12.

Create New Document Wizard

Create a document object
Select whether the object has content.

Document Title:

With content Create an object with one or more content elements.

Without content Create an object with no content element.

Save As Reservation

Create Minor Version

Create As Compound Document

< Back Next > Cancel Help

Figure 4-12 Create new document wizard

4. Select the storage policy for the document that corresponds to the storage area created in “Create a storage area with no retention” on page 57 and complete the wizard by clicking **Finish**, as shown in Figure 4-13.

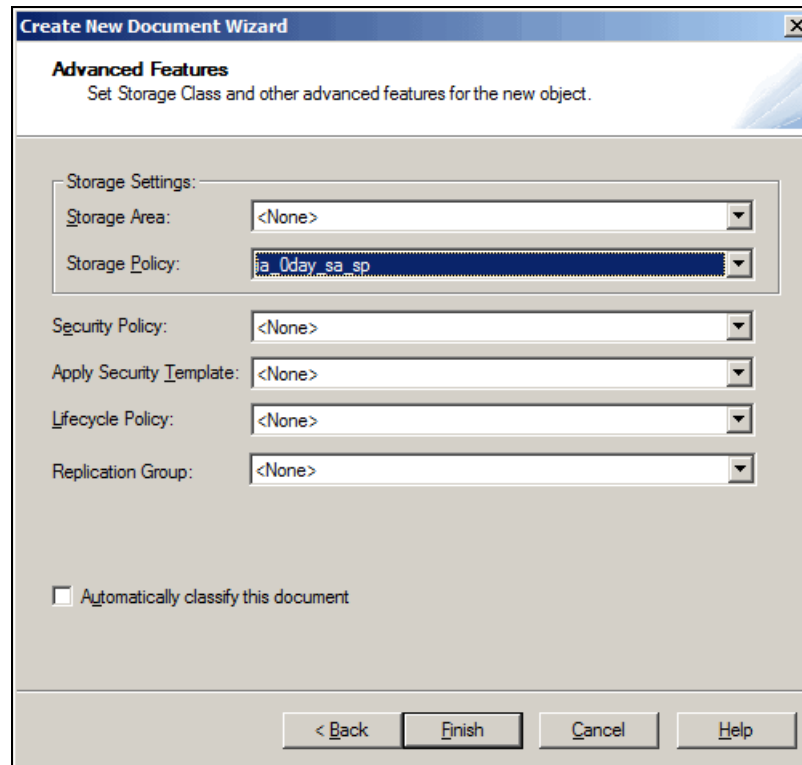


Figure 4-13 Assign document to storage policy with no retention

5. Create another document, repeating steps 1-4. On the final step, select the storage policy that corresponds to the storage area that you created in “Create a storage area with a 30 day retention” on page 60 and complete the wizard by clicking **Finish**. See Figure 4-14.

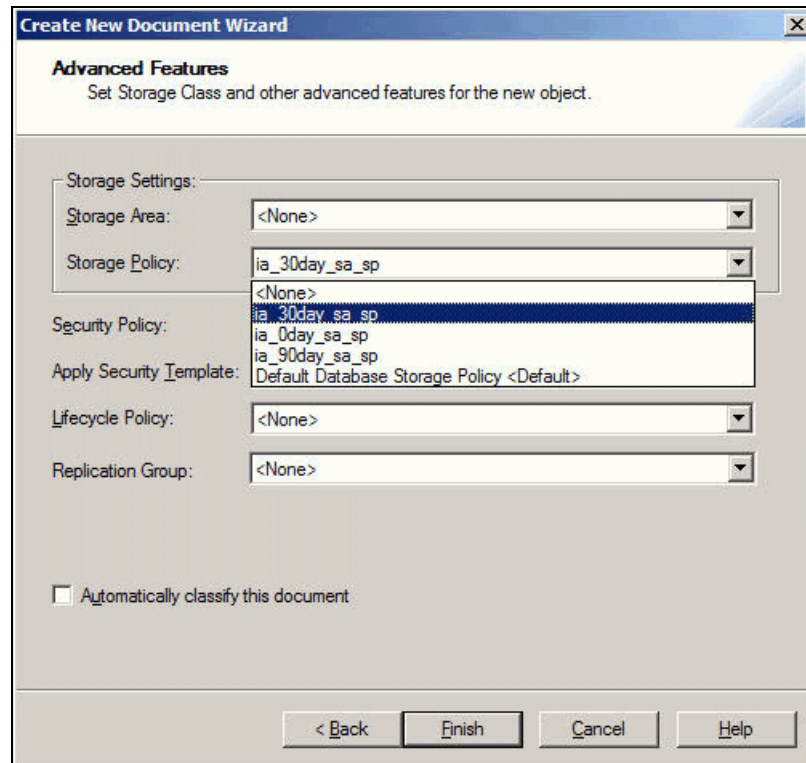


Figure 4-14 Assign document to storage policy with 30 day retention

Verify document storage location

To verify that the document that we just added is stored on the Information Archive device, use the Tivoli Storage Manager administration command line client:

```
dsmadm
```


Run the command shown in Figure 4-15 that displays the number of files in the Information Archive device that belong to a given node. The result shows that node p8, the node used by our Tivoli Storage Manager FCD, has two files associated with it. This number corresponds to the two documents that we just added in the Content Engine using FileNet Enterprise Manager.

```
tsm: SSAM1>SELECT node_name, SUM(num_files) as "Number of files"
FROM occupancy GROUP BY node_name
```

NODE_NAME	Number of files
P8	2

Figure 4-15 Command line to display number of files for node after documents created

Now delete the document that we created with no retention from the Content Engine using FileNet Enterprise Manager (see Figure 4-16) and validate that it is deleted from the Information Archive device by confirming the new value for Number of files.

Note that we run the `expire inventory` command first to force the Information Archive device to remove documents that can be deleted rather than waiting for the Information Archive device's automated expiration process to run.

```
tsm: SSAM1>expire inventory
ANS8003I Process number 131 started.
```

```
tsm: SSAM1>SELECT node_name, SUM(num_files) as "Number of files"
FROM occupancy GROUP BY node_name
```

NODE_NAME	Number of files
P8	1

Figure 4-16 Expire inventory

4.5 Hints and tips

This section provides best practices, hints, and tips for your integration planning and implementation.

4.5.1 Best practices

Consider the following best practices when integrating FileNet P8 and an Information Archive device:

- ▶ Always use a disk storage pool as opposed to a tape storage pool as the primary storage pool that is associated with a management class used by FileNet P8.
- ▶ Do not modify the retention settings of the Information Archive management class after you have started using it with FileNet P8. The retention date for documents already created in FileNet P8 does not change.
- ▶ When using IBM Enterprise Records with the Information Archive device, use a management class with zero retention, using the event based retention mechanism, so that IBM Enterprise Records can control all retention policies and dispositions.
- ▶ Use the event based retention scheme on the Information Archive device rather than a chronological retention scheme when defining management classes to use with FileNet P8. This method results in less load on the Tivoli Storage Manager server and better performance than chronological retention.
- ▶ Ensure that the directory specified by the Configuration Files Share property on the Tivoli Storage Manager FCD is on a network share that allows all Content Engine servers in a cluster to write to the directory using the same file system path.
- ▶ If you have multiple Tivoli Storage Manager FCDs for the same Information Archive server, use a different node name for each FCD.
- ▶ Make sure that you include the Configuration Files Share directory and the DSMI directories on UNIX in your backup planning for FileNet P8.
- ▶ Create a FileNet P8 document class with a default storage policy set for your Information Archive storage area to use for all documents that you want to store on the Information Archive device.
- ▶ To improve retrieval of large documents stored on the Information Archive device, use the Content Engine content cache. When a large document is requested, the content cache begins populating the file in its cache using a single Tivoli Storage Manager session. Multiple round trips from a FileNet P8 client to the Content Engine server for chunks of this document's content only access this cache file. If the content cache is not used, each Content Engine

server can establish separate Tivoli Storage Manager sessions to the Information Archive device to retrieve chunks of this document's content.

- ▶ To improve the retrieval of large numbers of documents at a time, increase the value of the `FCP Pool Max in Use` property on the Tivoli Storage Manager FCD. This value increases the number of concurrent requests that can be satisfied for a FileNet P8 client. You might also have to increase the `MAXSESSIONS` option on the Information Archive device as described in "Modify the Information Archive device server parameters" on page 50.

4.5.2 FileNet P8 configuration troubleshooting

If you cannot create Tivoli Storage Manager FCD, there is either a configuration problem or a communication problem to the Information Archive device.

Check the FileNet P8 error log for the following clues:

- ▶ Check whether the Content Engine application server uses a 32-bit or 64-bit JVM. If it uses a 32-bit JVM, make sure that you have installed a 32-bit Tivoli Storage Manager client. Likewise, if it uses a 64-bit JVM, make sure that you have installed a 64-bit Tivoli Storage Manager client.

Note that a 32-bit JVM can run on a 64-bit operating system. So do not assume Content Engine uses a 64-bit JVM just because it is running on a 64-bit operating system.

- ▶ Verify that the operating system user account that starts the Content Engine application server, such as WebSphere, can write to the directory that is specified by the `Configuration Files Share` property on the Tivoli Storage Manager FCD.

4.5.3 Logs

This section describes how to set up logging for FileNet P8 and the Tivoli Storage Manager client.

FileNet P8 logs

Use FileNet Enterprise Manager to enable trace logging on Content Engine using the instructions in the information center:

http://publib.boulder.ibm.com/infocenter/p8docs/v4r5m1/topic/com.ibm.p8.doc/ce_help/trace/logging/tl_enable_and_disable_trace_logging.htm

In the Trace Control window of FileNet Enterprise Manager, enable detail trace for the Content Storage and Fixed Content Provider subsystems.

Tivoli Storage Manager client logs

Enable Tivoli Storage Manager client tracing by manually editing the Tivoli Storage Manager options file that is located in the Configuration Files Share directory specified in the Tivoli Storage Manager FCD. The *FCDID.opt* file is used to control Tivoli Storage Manager client tracing for both Windows and UNIX Content Engine environments. See “Tivoli Storage Manager options files” on page 53 for additional information.

Add the following lines at the end of the file:

```
TRACEFILE /opt/ibm/tsmclient.log
TRACEFLAGS api api_detail comm commdetail verbinf verbdetail timestamp
```

You can adjust the path to the tracefile, as appropriate.

Make sure that you write the TRACEFILE to a location specific to each Content Engine server host. For example, do not put it in a shared network folder where multiple Content Engine servers can write to the same trace file.

Restart all Content Engine servers in the cluster to enable tracing.

To disable Tivoli Storage Manager tracing, remove the trace lines added to the FCDID.opt file. Then restart all Content Engine servers in the cluster.

4.6 Summary

This chapter provided a description of IBM FileNet P8 integration with the Information Archive device as an archive storage system with hierarchical storage devices and retention capabilities. We explained the integration architecture of FileNet P8 and the Information Archive device including:

- ▶ Deployment model
- ▶ Data model
- ▶ Retention
- ▶ Hierarchical storage

We also covered the steps that are required to set up FileNet P8 and the Information Archive device to enable this integration.

Finally, we provided some basic best practices and tips to help with installation, configuration, and troubleshooting the integration.



Integrating IBM Content Manager with the Information Archive device

In this chapter, we discuss the integration between IBM Content Manager and the IBM Information Archive device. We describe the required configuration to enable IBM Content Manager to work with the Information Archive device.

This chapter includes the following topics:

- ▶ Introduction
- ▶ IBM Content Manager overview
- ▶ Integration architecture
- ▶ Integration implementation
- ▶ Hints and tips
- ▶ Summary

5.1 Introduction

Today's organizations face the challenge of archiving information for long-term retention. For some organizations, IBM Content Manager is already part of their overall enterprise content management strategy. For others, there is no such strategy or solution.

From a software perspective, you can use IBM Content Collector to collect email, files, and application data and store them in an IBM Content Manager content repository. You can also use this solution in combination with other products, such as IBM Classification Module, IBM Content Analytics, IBM Enterprise Records, IBM eDiscovery Manager, and IBM eDiscovery Analyzer, to provide advanced content classification, analysis, records management, and eDiscovery capabilities.

From a hardware perspective, for long-term, retention managed storage, IBM Content Manager integrates with the IBM Information Archive device to provide an end-to-end archiving infrastructure. In addition, organizations with an existing enterprise content management repository, such as IBM Content Manager, can now require a compliance storage solution to house their current content collections.

5.2 IBM Content Manager overview

IBM Content Manager is a highly scalable and secure repository that allows document creation, indexing, securing, and life cycle management, that delivers to authorized users, and that integrates content with other applications. IBM Content Manager manages a variety of digital content, such as images, word processor, spreadsheets, multimedia, and email messages across multiple platforms, databases, and applications.

IBM Content Manager can handle a large volume of documents within the same system. It provides a powerful security scheme to assign different privileges to users. IBM Content Manager provides a set of APIs to allow integration of enterprise applications with IBM Content Manager repository. It also provides a document routing engine to automate content-centric workflows.

IBM Content Manager provides organizations with the following benefits:

- ▶ A flexible data model for document metadata and relationships to meet the organization's needs.
- ▶ A flexible, yet powerful authorization model to maintain documents secrecy and confidentiality.

- ▶ Flexible storage management techniques to use a variety of storage devices and to meet retention policies.
- ▶ Flexible APIs to integrate with other applications.
- ▶ The ability to extend its capabilities, such as incorporating with records management to meet the organization's strategies.

5.2.1 IBM Content Manager architecture

If you go into a library to look for a book, the first thing you do is browse through the library's catalog (either online or using the index cards). In the library's catalog, you can search for a book by its category or subject, the book title, or the author's name. The catalog can also tell you the location of the book on the shelf. You can then go to the shelf and get the book.

IBM Content Manager architecture is similar to the library catalog example that we mention here. IBM Content Manager consists of the following major components:

- ▶ The Library Server, which represents a library's catalog
- ▶ The Resource Manager, which represents the shelves in the library that hold the books

A typical IBM Content Manager system contains only one Library Server and one or more Resource Managers. The Library Server and Resource Manager are loosely coupled logical components that can be installed on the same system or on different systems.

Figure 5-1 shows the IBM Content Manager system architecture.

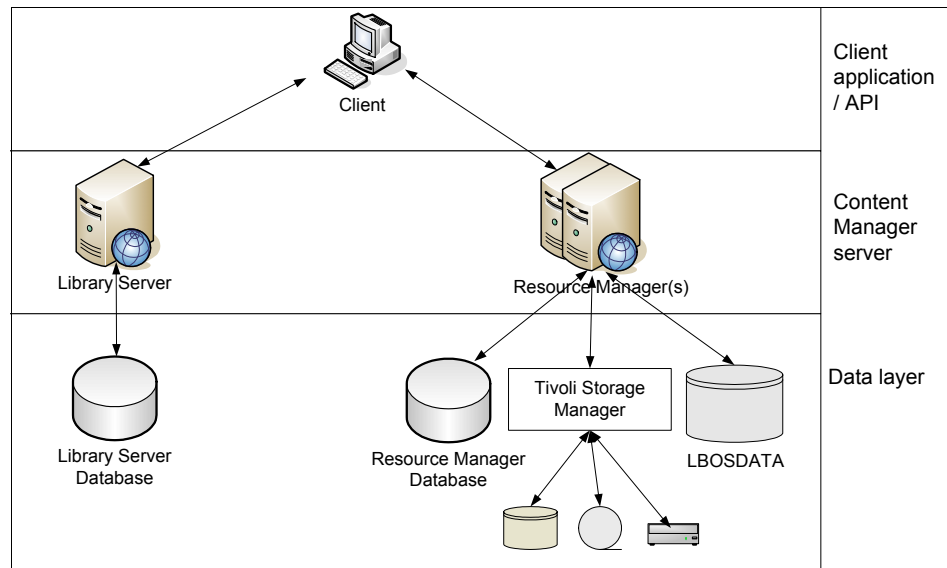


Figure 5-1 IBM Content Manager system architecture

Library Server

The Library Server is the “mind” of a Content Manager system. It is a centralized component containing system’s definitions for document types, security, workflow, and configurations. It also stores document metadata. The Library Server consists mainly of a database that includes all this information. Clients communicate with the Library Server by calling stored procedures on the Library Server database.

When a user connects to the system, the system performs an authentication operation using the Library Server’s defined users. When the user logs in successfully, the user performs a search operation for a document, also referred as an *item*. A set of tokenized search results is returned to the user. Depending on the user’s authorization, the user can perform a variety of actions on the documents. To access the document content, the token is passed to the Resource Manager to retrieve the document from the repository.

Resource Manager

The Resource Manager is the “muscle” of the system. It consists of the following major components:

- ▶ Resource Manager database: A database that is used internally to manage stored files.
- ▶ Resource Manager application: A J2EE application that is responsible for delivering (or retrieving) documents to users and for storing documents that are imported (or scanned) by users.
- ▶ Resource Manager services: Services that perform operations, such as data replication, data purge, and data migration, from one storage device to another.

Documents are stored in the Resource Manager in the following ways:

- ▶ LBOSDATA (file system structure): Enables storing documents on local disks, SAN storage, NAS storage, or DASD storage.
- ▶ Tivoli Storage Manager: Enables storing documents on other storage devices, including the Information Archive device. Through this mechanism, you can also define document retention policies. In addition, it provides documents Hierarchical Storage Management, which allows moving documents from one storage device to another until the document's final disposition.

When a Resource Manager application receives a token that is generated by the Library Server, it decrypts the token. Then, it searches for the entry in the Resource Manager database for the actual location of the document. Finally, it retrieves the document and returns it to the user.

For more information about IBM Content Manager, refer to:

<http://www.ibm.com/software/data/cm/cmgr/mp/edition-enterprise.html>

For IBM Content Manager product documentation, refer to:

<http://www.ibm.com/support/docview.wss?rs=86&uid=swg27015910>

For the IBM Content Manager help center, visit the information center at:

<http://publib.boulder.ibm.com/infocenter/cmgmt/v8r4m0/index.jsp>

To implement and configure IBM Content Manager, refer to *Content Manager Implementation and Migration Cookbook*, SG24-7051.

5.2.2 Benefits of using an Information Archive device with IBM Content Manager

The Information Archive device adds the following major enhancements to IBM Content Manager:

- ▶ Compliance and retention management
- ▶ Storage management

Compliance and retention management

Enterprises worldwide are facing challenges to comply internal regulations with industry and country regulations. This challenge includes the challenge of preserving documents and emails from being deleted or updated for any reason.

Integrating IBM Content Manager and the Information Archive device provides a solid retention framework to protect a company's documents and to mitigate risks.

The Information Archive device prevents documents from being physically deleted (intentionally or by mistake) at the storage level, assuring that a document is never deleted before its retention period expires. For each document type, a different retention policy can be applied.

The Information Archive device supports retention holds. Placing a hold on a document is required when the document is or might be used as an evidence in a legal or discovery case. To prevent potential deletion of the document through the normal disposition cycle, you can apply the necessary legal hold, and the document is preserved until the removal of the hold.

An example of an end-to-end archiving solution that integrated the Information Archive device and IBM Content Manager is IBM Information Archive for Email, Files, and eDiscovery. This solution comes with software bundle preinstalled and preconfigured on the hardware, which helps to simplify and accelerate the implementation of an end-to-end archiving and eDiscovery solution from weeks to days.

An important aspect of this bundle is responding to audit and litigation issues. Preserving documents and finding documents are two sides of the same coin; they are both inseparable and are key for any organization. This solution is the perfect solution to preserve documents and then to respond to discovery requests promptly.

Storage management

The Information Archive device is a scalable storage appliance whose capacity can surpass 600 TB of storage capacity. In addition, it provides extensive

capability for document deduplication and compression to maximize the use of the initial capacity.

Also, the Information Archive device adds to IBM Content Management tiered storage management capabilities. The following example helps to explain the concept of tiered storage management.

Your company wants to store electronic contracts for a period of 30 years after the contract is completed.

For the first 2 years immediately after the contract completion date, the electronic contracts are accessed frequently as reference for other related new contracts. At this early stage of the contracts life cycle, the contracts are stored on a fast storage to serve the frequent access requests.

For the next 5 years, the electronic contracts are no longer accessed as often. They are rarely retrieved to be used as reference for new projects. Thus, they are stored on a medium speed online storage, such as SATA disks.

Finally, for the remaining 23 years, these contracts are almost never accessed, except for special occasions such as in the case of litigation. For this last stage, the contracts are kept on tapes for long-term archiving.

IBM Content Manager integrates with Tivoli Storage Manager to fulfill the tiered storage management requirements.

The IBM Information Archive device (its predecessors being DR550 and DR450) are the perfect solution to be combined with fast storage devices because they provide a variety of storage media types, such as slow disks and tapes. This complete solution mitigates risks of compliance and discovery while reducing storage costs by compressing data and eliminating duplicates and by enhancing overall system performance, including content management repository, messaging systems, file shares, and so forth.

For more information about the Information Archive device, refer to:

<http://www.ibm.com/systems/storage/disk/archive/>

Also, consult the information center at:

<http://publib.boulder.ibm.com/infocenter/tivihelp/v37r1/index.jsp>

For more information about the Information Archive architecture and deployment, refer to *IBM Information Archive Architecture and Deployment*, SG24-7843.

5.3 Integration architecture

The Information Archive device is the latest generation of information retention solutions. It is designed as a universal storage repository for all types of content (structured or unstructured) to help organizations of any size address complete information retention needs.

Tivoli Storage Manager is mainly used for backup and restore solutions. However, it plays an important role with IBM Content Manager. Tivoli Storage Manager is a mediator to read or write to more than 700 different types of storage devices, including the Information Archive device. It has the following components:

- ▶ Tivoli Storage Manager server

The Tivoli Storage Manager server component provides backup, archive, and space management services to the clients over the objects that are stored in the final archive repository, for example the Information Archive device. The Tivoli Storage Manager server component must be installed on the Information Archive device.

- ▶ Tivoli Storage Manager client

The Tivoli Storage Manager client must be installed on the Resource Manager system. The Resource Manager uses Tivoli Storage Manager backup and restore APIs to fully control documents and to perform operations such as document archiving, deletion, or moving of documents from one storage device to another.

The remainder of this section describes:

- ▶ A deployment model for the Content Manager and Information Archive components
- ▶ A data model for Content Manager and Information Archive objects
- ▶ Retention
- ▶ Hierarchical storage

5.3.1 Deployment model

IBM Content Manager uses the Tivoli Storage Manager Archive Client API to communicate with the Information Archive device through the Tivoli Storage Manager server component interface. Thus, it requires that the Tivoli Storage Manager client software is installed in the system that hosts the Resource Manager server.

Figure 5-2 illustrates Tivoli Storage Manager deployment in an IBM Content Manager environment.

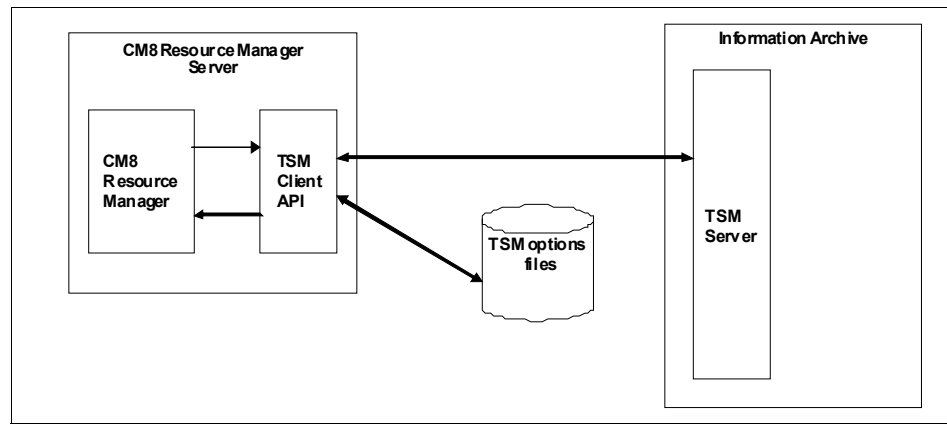


Figure 5-2 Tivoli Storage Manager deployment in IBM Content Manager environment

The deployment model consists of the following components:

- ▶ **CM8 Resource Manager server**
IBM Content Manager (CM8) Resource Manager server is the physical server on which the Resource Manager runs. If there are multiple Resource Manager servers installed in a system, a Tivoli Storage Manager client must be installed on each server.
- ▶ **Tivoli Storage Manager options file**
The Tivoli Storage Manager options file hold the connection parameters that are used by the Tivoli Storage Manager client to communicate with the Information Archive device. We discuss how to create the Tivoli Storage Manager options file in “Create the Tivoli Storage Manager options file (Tivoli Storage Manager client)” on page 88.
- ▶ **Tivoli Storage Manager server**
The Tivoli Storage Manager server provides backup, archive, and space management services to the clients. It enables IBM Content Manager to store data on the Information Archive device. The Tivoli Storage Manager server can be installed in the same host system where the Tivoli Storage Manager client resides or the Tivoli Storage Manager server component that is included in the Information Archive device can be used.
- ▶ **The Information Archive device**
The Information Archive device has a Tivoli Storage Manager server component that manages the storage and retrieval of objects that are stored

on the Information Archive device for application clients that use the Tivoli Storage Manager client API.

In the remaining sections of the chapter, we discuss how to configure Content Manager to use the Information Archive device.

5.3.2 Data model

Figure 5-3 depicts the data model relationships between the Content Manager and the Information Archive device objects.

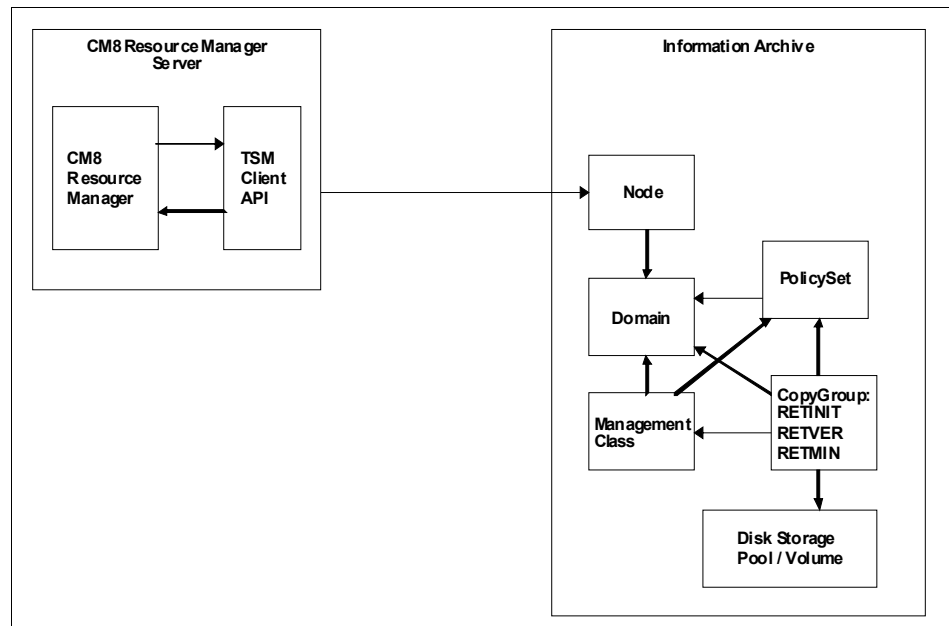


Figure 5-3 IBM Content Manager and Information Archive data model object relationship

We describe each of these Tivoli Storage Manager objects in more detail in “Create storage and policy configurations (Tivoli Storage Manager server)” on page 87. When Content Manager documents are transferred to the Information Archive device for persistent storage, they are stored in the disk storage pool that is associated with the node.

From the disk storage pool, documents can be migrated further to other storage pools, as described in 5.3.4, “Hierarchical storage” on page 85. This migration to other storage pools is transparent to Content Manager.

5.3.3 Retention

Retention relates to when a document can be deleted. Retention is the length of time a file is stored in the archive. Documents cannot be deleted until the retention period associated with the document has passed. When using Content Manager with the Information Archive device, retention can be enforced using one of the following mechanisms:

- ▶ The Information Archive device
- ▶ Content Manager, Tivoli Storage Manager

Retention enforced by the Information Archive device

Content Manager supports *static* retention, which is applied when a document is checked in and which prevents deletion until after the retention period expires. When connected to the Information Archive device, Content Manager can use either the *chronological* or *event-based retention* scheme that is used by the Information Archive device to enforce this retention.

Content Manager uses the behavior of the Tivoli Storage Manager management class to control when a document can be deleted. Although either scheme can be used, the preferred practice is to use event-based retention, because it does not require that a hold be placed on a document when it is created. This hold results in less load on the Tivoli Storage Manager server and better performance than chronological retention.

Content Manager always remains in control of the document deletion process. Thus, although the Information Archive device supports the automatic deletion of objects after their retention period has expired, Content Manager will not allow automatic deletion until after the retention period has expired *and* until a user specifically deletes the associated document from the Content Manager. When both conditions are met, Content Manager relinquishes control of the Tivoli Storage Manager object, allowing the Information Archive device to delete it automatically.

Note that deletion on the Information Archive device typically happens as a background process. Thus, requests by Content Manager to delete an object on the Information Archive device might not occur immediately. You can set up automatic expiration processing on the Information Archive device using the EXPINTERVAL server option. The Information Archive device inventory expiration process removes files based on the archive copy groups of the management classes to which the files are bound.

To control retention while suppressing automatic deletion, Content Manager uses the following specific parameters of the retention schemes that are available in the Information Archive device as shown in Table 5-1.

Table 5-1 Retention parameters

Parameter and Purpose	Value for Chronological Retention	Value for Event-Based Retention
RETINIT determines when to initiate the retention period defined in the RETVER attribute	CREATION (start counting when the object is archived)	EVENT (start counting when the event is triggered)
RETVER number of days to retain the archived object after retention is initiated	0 to 30,000 days, or NOLIMIT	0 to 30,000 days
RETMIN minimum number of days to retain the archived object	Not applicable	n, where n represents a number of days

Chronological retention

Using this method of retention prevents a file from being deleted until it has been retained for a defined time. A chronological management class uses the RETVER parameter to protect a Tivoli Storage Manager object. The RETVER parameter defines the number of days that the object should be retained, starting from the date that the object is stored on the Information Archive device.

Tivoli Storage Manager automatically deletes objects when their retention period (RETVER) on the Information Archive device has expired. To enable Content Manager to retain control of the objects with chronological retention, the Content Manager places a hold on the object after creation. Later, when the document is finally eligible for deletion, after the RETVER has expired and a user deletes the associated document, the Content Manager releases the hold and allows the Tivoli Storage Manager automatic deletion process to delete the object.

Placing a hold on an object uses the IBM System Storage Archive Manager component of the Information Archive device. This component is the same as the IBM Tivoli Storage Manager with the `archiveretentionprotection` attribute set to on.

When you create a Tivoli Storage Manager management class to support chronological retention, use the following parameters:

```
RETINIT=CREATION
RETVER=<actual retention period>
```


Note: Chronological retention cannot be used for a zero duration retention. If you need to define a zero duration retention, use the event-based retention that we describe in the next section.

Event-based retention

All management classes with an archive copy group must specify a retention period, for example the number of days that an archived object will be stored on the server before being deleted. An event-based retention policy provides the option of beginning the retention period either at the time the object is archived or at a later date when an activation event is sent to the server for that object.

Using this method of retention prevents documents from being deleted until an event is signaled. The event is signaled when the retention time (RETVER) is activated through an API retention event. The System Storage Archive Manager collection then assigns an expiration date for this object.

An event-based retention (EBR) management class has two retention settings that combine to protect a Tivoli Storage Manager object:

- ▶ The RETMIN parameter defines the minimum retention period for the object from creation time.
- ▶ The RETVER parameter defines the retention from the time of activation.

As long as the EBR activate event is not issued, the object is protected from deletion, no matter what the value is for RETMIN. When the EBR activate event is issued, the retention period that is defined in RETVER is applied from that moment. The retention period set in RETVER can overlap with RETMIN. For example, if RETVER=0 and the activate event is issued during the RETMIN period, the object continues to be protected until RETMIN expires.

Content Manager uses the behavior of the EBR management class to control when the Tivoli Storage Manager object can be deleted. If you set in Tivoli Storage Manager an EBR management class with RETMIN=365 and RETVER=0, Content Manager calculates the delete date (today + 365 days) and writes it to the ContentRetentionDate attribute of the document. When a user wants to delete the document more than one year later, the Content Manager delete call sends an activate event to the Tivoli Storage Manager server, and because RETVER=0, the retention expires immediately, which allows the object to be deleted automatically by the Tivoli Storage Manager background task.

When you create a Tivoli Storage Manager management class to support event-based retention, use the following parameters:

```
RETINIT=EVENT  
RETVR=0  
RETMIN=<actual retention period>
```

Deletion hold

You can use a *deletion hold* (also called a *retention hold* or *legal hold*) to retain a file longer than the minimum retention period that it was originally assigned. For example, a deletion hold can be used if a file needs to be saved for the duration of a legal or company-required audit.

A deletion hold can be applied to an archived file at any time during its retention period. The file is retained until a deletion release is applied. If a deletion release is not applied, the object is retained indefinitely. A deletion hold can be placed on a file with either a chronological archive retention or an event-based retention policy.

You can hold a file by using an archiving client to issue a retention event with the Hold parameter. When you no longer need to hold the document, use an archiving client to issue a retention event with the Release parameter. A file in a deletion hold cannot be deleted until you release it.

Retention enforced by IBM Content Manager, Tivoli Storage Manager

The definition of the Tivoli Storage Manager storage system is slightly different when retention protection is enabled. There are two supported retention modes:

- ▶ Standard
- ▶ Aggregate

The aggregate retention option improves performance when most of the files being stored are small ones. With standard retention, migration does not start until a specified amount of data is ready to be migrated. Configuration is the same for both retention modes.

Tip: Selecting the correct retention mode can improve migrator performance. You can change between retention modes at any time by changing a few configuration settings.

We discuss how to configure the Tivoli Storage Manager storage system in more detail in 5.4.1, “Configure Tivoli Storage Manager” on page 86.

5.3.4 Hierarchical storage

A *migration policy* is a user-defined schedule for moving objects from one storage class to the next. It describes the retention and class transition characteristics for a group of objects in a storage hierarchy. Creating a migration policy and defining the migrator schedule automates the migration of objects so you do not have to manually monitor migration. (Tivoli Storage Manager calls its migration policies *management classes*.)

Each migration policy belongs to a collection of *objects* and contains the *rules* for migrating the objects in that collection. When you create your migration policy, you decide how long to store a collection in a storage system. An object is set to migrate according to the retention period that is specified in its migration policy. However, if the volume becomes full, the object might migrate early because of threshold migration.

The *threshold cycle* is the amount of time in hours and minutes that must elapse before the system checks the capacity of volumes. You specify the threshold cycle in the Threshold field on the Cycles page of the Resource Manager Configuration window. Threshold migration works just like normal migration, except that it occurs when the volume space exceeds the threshold.

The Information Archive device supports hierarchical storage capabilities to allow you to create policies so data is stored on the type of media that best meets data longevity, access speed, and cost needs. For example, with the Information Archive device, you can attach tape devices to archive data that is accessed infrequently.

Migration automates moving data from one type of media to another as media needs change and as new types of media become available in the market.

When using hierarchical storage with Content Manager, keep in mind the following important considerations:

- ▶ Always use a disk storage pool as the first storage pool for incoming documents from Content Manager. For example, Content Manager documents should not be written directly to a tape storage pool. This method can lead to performance problems when transferring documents from Content Manager to the Information Archive device.

Instead, set up additional storage pools and migration policies from the incoming disk storage pool for these other media types such as tape.

You can arrange storage pools in a storage hierarchies, which consist of at least one primary storage pool to which a client node backs up, archives, or migrates data. Typically, data is stored initially in a disk storage pool for fast

client restores and then moved to a tape-based storage pool, which is slower to access but which has greater capacity.

- ▶ When using tape, adjust the MOUNTWAIT and MOUNTRETENTION parameters of the device class for the tape library and the IDLETIMEOUT parameters for the system. Adjusting these parameters prevents Content Manager users from waiting for an unreasonable length of time when retrieving documents stored on tape.

5.4 Integration implementation

The implementation example in this section describes how to set up a storage configuration on an Information Archive device that is integrated with IBM Content Manager. The high-level steps are as follows:

1. Configure Tivoli Storage Manager
 - a. Create storage and policy configurations (Tivoli Storage Manager server).
 - b. Create the Tivoli Storage Manager options file (Tivoli Storage Manager client).
2. Configure the Resource Manager:
 - a. Create a collection in the Resource Manager to use the Information Archive device as a permanent storage device.
 - b. Assign the collection to the desired item type.
3. Test the integration.

We cover the details of each steps in the following sections.

5.4.1 Configure Tivoli Storage Manager

In this section, we show how to configure Tivoli Storage Manager to enable IBM Content Manager to store data on the Information Archive device. This configuration requires the following high-level steps:

1. Create storage and policy configuration:
 - a. Create a storage pool.
 - b. Assign a volume to this storage pool.
 - c. Create a policy domain.
 - d. Create a policy set.
 - e. Create a management class.

- f. Create a backup copy group.
 - g. Set the created management class as default for policy set.
 - h. Validate and activate the policy set.
 - i. Register the node to allow communication between IBM Content Manager server (that is, the Tivoli Storage Manager client on the Resource Manager system) and the Tivoli Storage Manager server.
2. Create the Tivoli Storage Manager options file.

Create storage and policy configurations (Tivoli Storage Manager server)

To create storage and policy configuration, perform these steps:

1. From the Tivoli Storage Manager server, open the Tivoli Storage Manager Console and log in.
2. Start the Command dialog box by expanding the tree on the left to **Tivoli Storage Manager** → **<server_name>** → **TSM Server** → **Command Line**. See Figure 5-4.

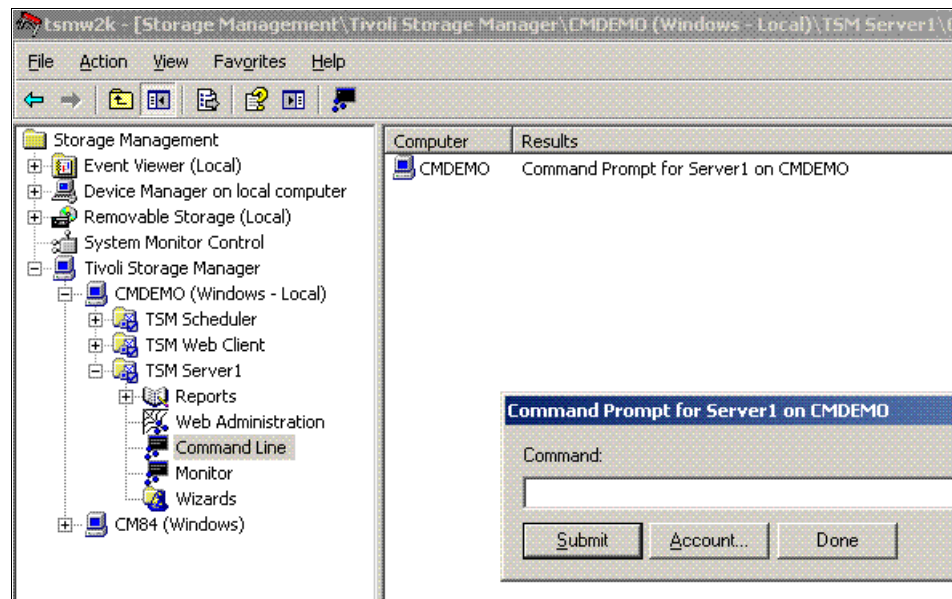


Figure 5-4 Tivoli Storage Manager Management Console and the Command Line

3. Create a *storage pool* by entering the following command:

```
DEFINE STGPOOL IA_POOL DISK DESCRIPTION="IA Storage Pool"
```

4. Assign a *storage volume* of 100 MB (or other desired value) to the storage pool:

```
DEFINE VOLUME IA_POOL DISK FORMATSIZE=100 WAIT=YES
```

5. Create a *policy domain* and a *policy*:

```
DEFINE DOMAIN IA_DOMAIN DESCRIPTION="IA Policy Domain"  
DEFINE POLICYSET IA_DOMAIN IA_POLICY DESCRIPTION="IA Policy Set"
```

6. Create a *management class*:

```
DEFINE MGMTCLASS IA_DOMAIN IA_POLICY IA_DISK DESCRIPTION="IA  
Management Class"
```

7. Create a *backup copy group* for the created management class and set the destination as the Information Archive device defined storage pool:

```
DEFINE COPYGROUP IA_DOMAIN IA_POLICY IA_DISK DESTINATION=IA_POOL  
VEREXISTS=1 VERDELETED=0
```

8. Set IA_DISK management class as default:

```
ASSIGN DEFMGMTCLASS IA_DOMAIN IA_POLICY IA_DISK
```

9. Validate the policy set and activate the policy set using the following command:

```
VALIDATE POLICYSET IA_DOMAIN IA_POLICY  
ACTIVATE POLICYSET IA_DOMAIN IA_POLICY
```

10. Register a node to allow the Resource Manager to connect to Tivoli Storage Manager server:

```
REGISTER NODE cmdemo password DOMAIN=IA_DOMAIN BACKDELETE=YES  
FORCEPWRESET=NO
```

11. Close the Tivoli Storage Manager management console.

Create the Tivoli Storage Manager options file (Tivoli Storage Manager client)

To create the Tivoli Storage Manager options file, perform the following operations on the Tivoli Storage Manager client system, which is the same system as the Resource Manager:

1. Open the TSM Client Configuration Wizard by clicking **Start** → **Programs** → **Tivoli Storage Manager** → **Backup-Archive GUI**.
2. Click **Next**.
3. Choose **Create a new options file** and click **Next**.
4. Enter cmdemo in the Node Name field, and click **Next**.
5. For communication method, choose **TCP/IP** and click **Next**.

6. Enter the Tivoli Storage Manager server's IP address, enter 1500 as the port number, and click **Next**.
7. Leave the defaults for the next windows.
8. Click **Apply** to confirm the configuration.
9. When prompted for a user name and password, set the user name to ia_node and set the appropriate password.
10. Click **Finish**.
11. Enable archiving and retention on the Information Archive server:
 - a. Open the Tivoli Storage Manager options file using a note pad or another text editor:
 C:\Program Files\tivoli\tsm\baclient\dsm.opt
 - b. At the end of the file, add the following line:
 ENABLEARCHIVERETENTIONPROTECTION YES

Figure 5-5 shows the added line at the end of the file.

NODENAME	CMDEMO		
TCPSEVERADDRESS		192.168.2.129	
TCPPOINT	1500		
ENABLEARCHIVERETENTIONPROTECTION			YES

Figure 5-5 Tivoli Storage Manager options file sample

This concludes the configuration on the Tivoli Storage Manager server and client sides. Now, you can make the necessary configuration changes on the IBM Content Manager side.

5.4.2 Configure the Resource Manager

In this section, we show the steps to configure the IBM Content Manager Resource Manager to store documents on the Information Archive device. This configuration requires the following high-level steps:

1. Create a collection in the Resource Manager to use the Information Archive device as a permanent storage device:
 - a. Create a server.
 - b. Enable the Tivoli Storage Manager Class.
 - c. Create a Storage Class.
 - d. Create a Storage Group.
 - e. Create a Tivoli Storage Manager Storage System.
 - f. Create a Migration Policy.
 - g. Create a Workstation Collection.
2. Assign the collection to the desired item type.

Create a collection in the Resource Manager to use the Information Archive device as a permanent storage device

To configure the Resource Manager to store documents on the Information Archive device:

1. Open **IBM Content Manager System Administration Client** by selecting **Start → Programs → IBM DB2® Content Manager Enterprise Edition → System Administration Client**.
2. Log in using the default `icmadmin` user ID (or another ID that you use for administrator) and the associated password.
3. Expand **Resource Manager** → **rmdb**.
4. Create a new server definition:
 - a. Right-click **Server Definitions** and select **New**.
 - b. Enter the following information when prompted:
 - Name: `cmdemo`
 - Server Type: Select **Tivoli Storage Manager**
 - Hostname: Enter the host name or IP address
 - User ID: `cmdemo`
 - Password: `password`
 - Port Number: `1500`
 - Client Options File: `C:\Program Files\tivoli\tsm\baclient\dsm.opt`

- c. Leave default values for the remainder of the fields. See Figure 5-6.
- d. Click **OK**.

The screenshot shows a dialog box titled "Server Definition Properties - cmdemo". It contains the following fields and values:

- Name: cmdemo
- Server type: Tivoli Storage Manager
- Hostname: cmdemo
- Platform: Windows
- User ID: cmdemo
- Password: *****
- Protocol: ftp
- Port number: 1500
- Schema: (empty)
- Path: (empty)
- Client options file: files\Tivoli\TSM\baclient\dsm.opt
- Buffer size: 131072
- File space name: C:
- File space information: (empty)

At the bottom of the dialog are four buttons: OK, Cancel, Apply, and Help.

Figure 5-6 Add a new server on the Resource Manager

5. Update the ICMADDM device manager properties:
 - a. Click **Device Managers**, right-click **ICMADDM**, and select **Properties**.
 - b. Enter the following parameter values (see Figure 5-7):
 - Parameters: mode=retention_aggregate
 - Class: TSM
 - Device Manager: Select **Enable**
 - c. Click **OK**.

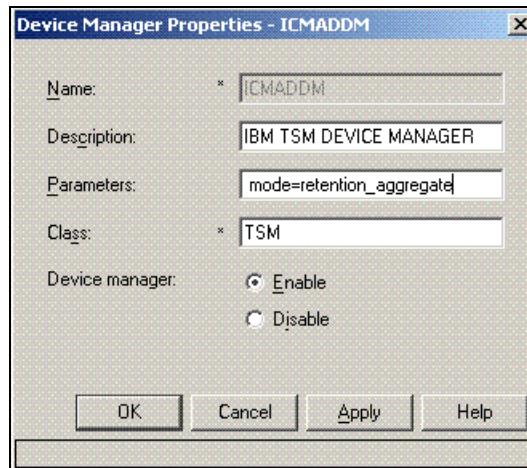


Figure 5-7 Configure the storage device properties

6. Create a new storage class:
 - a. Right-click **Storage Classes** and select **New**.
 - b. Enter the following parameters (see Figure 5-8):
 - Name: IA_CLASS
 - Destination: Select **Local Destination**
 - Device Manager: Select **ICMADDM**
 - c. Click **OK**.

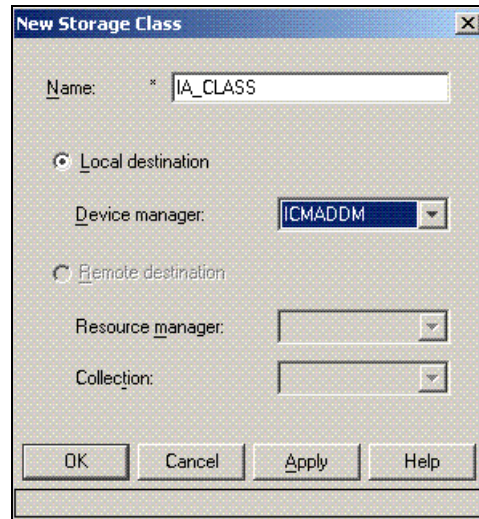


Figure 5-8 Configure a new storage class

7. Create a new storage group:
 - a. Right-click **Storage Groups** and select **New**.
 - b. Enter the following parameters (see Figure 5-9):
 - Name: IA_SG
 - c. Click **OK**.

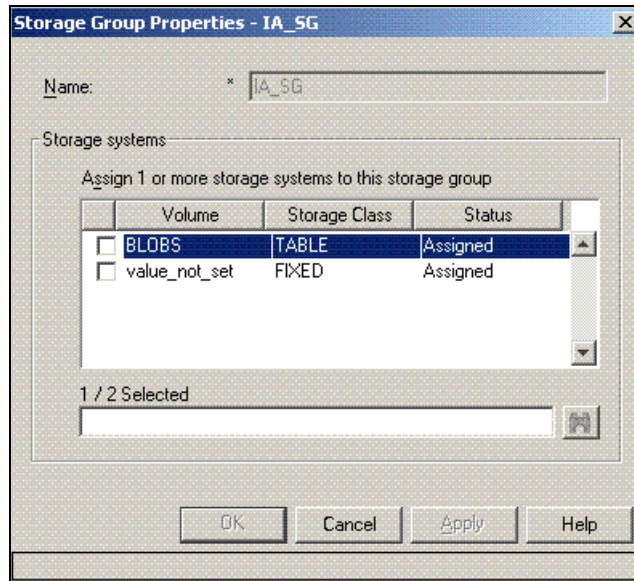


Figure 5-9 Configure storage groups

8. Create a new Tivoli Storage Manager volume:
 - a. Expand **Storage Systems**.
 - b. Right-click **Tivoli Storage Manager Volume** and select **New**.
 - c. Enter the following parameters (see Figure 5-10):
 - TSM Management Class: **IA_DISK**
 - Server Name: Select **cmdemo**
 - Storage Class: Select **IA_CLASS**
 - Assignment: Select **Assigned** and choose **IA_SG**.
 - d. Click **OK**.

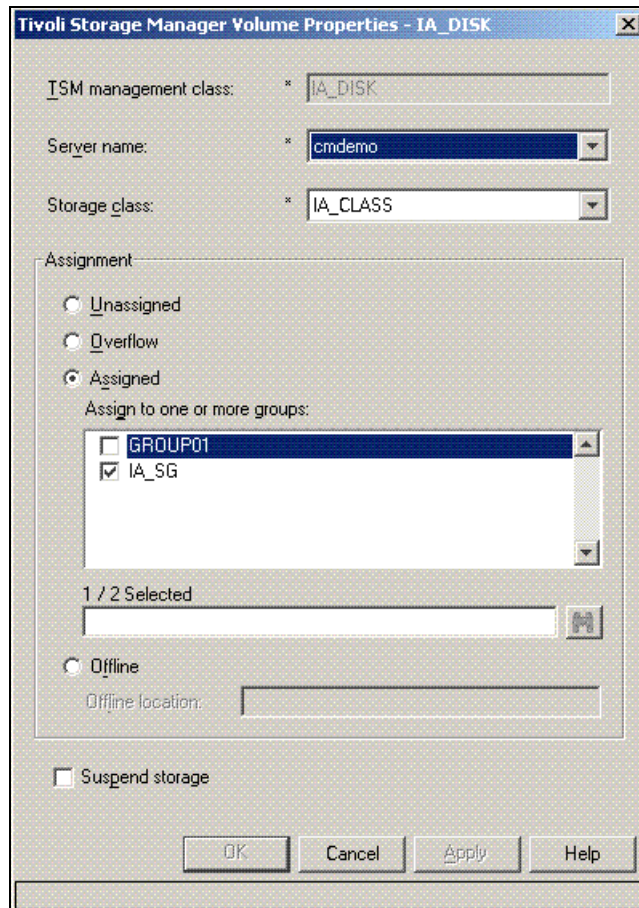


Figure 5-10 Create a new storage system

9. Create a new migration policy (see Figure 5-11):
 - a. Right-click **Migration Policies** and select **New**.
 - b. Enter the following parameters:
 - Name: IA_MIG_POL
 - c. Click **Add** and enter the following parameters:
 - Storage Class: Select **IA_CLASS**
 - Retention Period: Select **Forever**
 - d. Click **OK** to save the new storage class in the migration policy.
 - e. Click **OK** to save the migration policy.

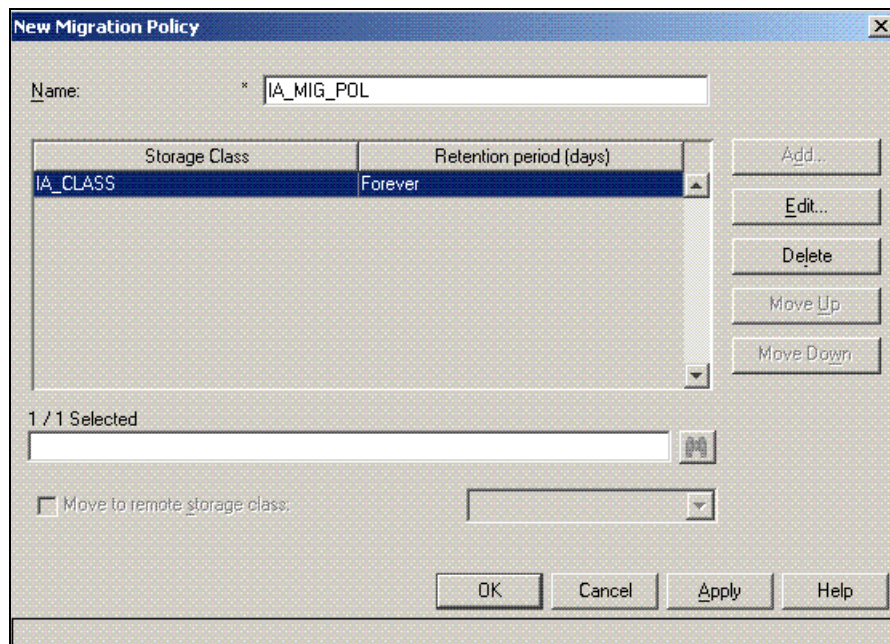


Figure 5-11 Create a new migration policy

10. Create a new workstation collection:
 - a. Right-click **Workstation Collections** and select **New**.
 - b. Enter the following parameters (see Figure 5-12):
 - Name: IA_COLL
 - Migration Policy: Select **IA_MIG_POL**
 - Storage Group: Select **IA_SG**
 - c. Click **OK**.

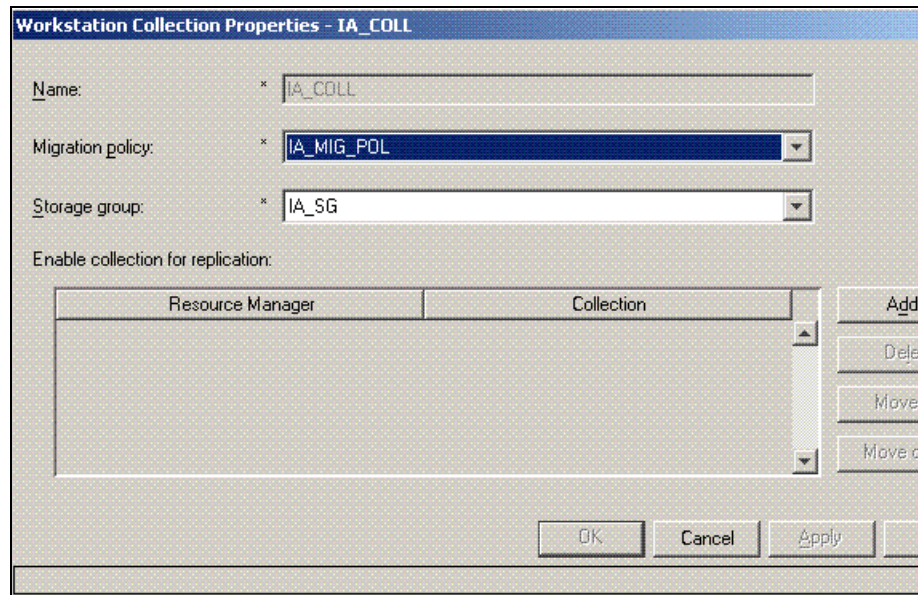


Figure 5-12 Create a new workstation collection

So far, you have created a definition on the Resource Manager to use the Information Archive device as a permanent storage device. Now, it is time to set the storage for an item type to the storage the collection that we have just defined.

Assign the collection to the desired item type

To set the item type of the documents that you want to store on the Information Archive device to the collection that you created previously, follow these steps:

1. Click the Document Management tab.
2. For each of the fields, complete the following steps (as shown in Figure 5-13):
 - a. Click the item type.
 - b. Click **Edit**.
 - c. For collection, select **IA_COLL**.
 - d. Click **OK**.

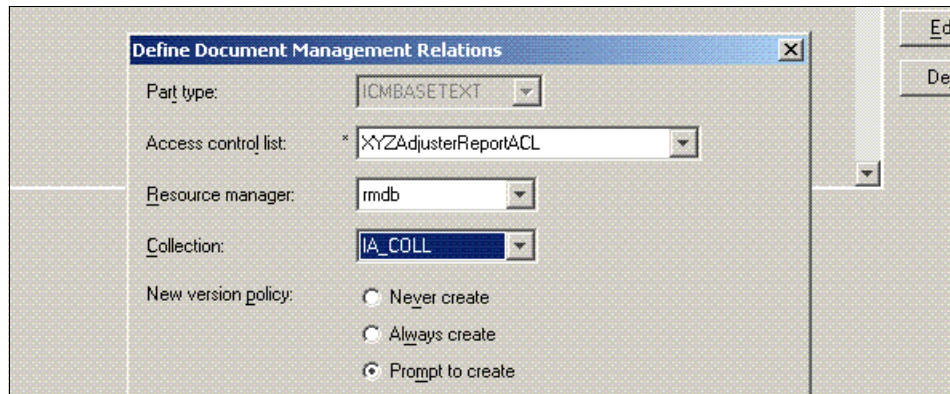


Figure 5-13 Assign the collection to the item type

Now, you can test the integration.

5.4.3 Test the integration

Always test the integration to make sure everything is set up correctly. To test, add documents to the item type that was assigned to the newly created collection. These documents should be stored on the Information Archive device.

Add a document

Use the following steps to add a document:

1. Open Tivoli Storage Manager Administrative Command Line Client by clicking **Start** → **Run** → **Type dsmdmc**.
2. Execute the following query before adding any document:

```
SELECT node_name, SUM(num_files) FROM occupancy GROUP BY node_name
```

Save the result number for later comparison.

3. Open IBM Content Manager Client for Windows from **Start** → **Programs** → **IBM DB2 Content Manager Enterprise Edition** → **Client for Windows**.
4. Log in using the default `icmadmin` user ID (or another ID that you use for administrator) and the associated password.
5. Import a document (Figure 5-14):
 - a. In the Welcome window, select **Import**.
 - b. Click **Add Files to Import**.
 - c. Select any text file.
 - d. In File Type, choose **Text Document**.
 - e. In Item Type, choose the item type you set up earlier.
 - f. Populate the related metadata information.
 - g. Click **Import**.
 - h. Wait until progress window closes.

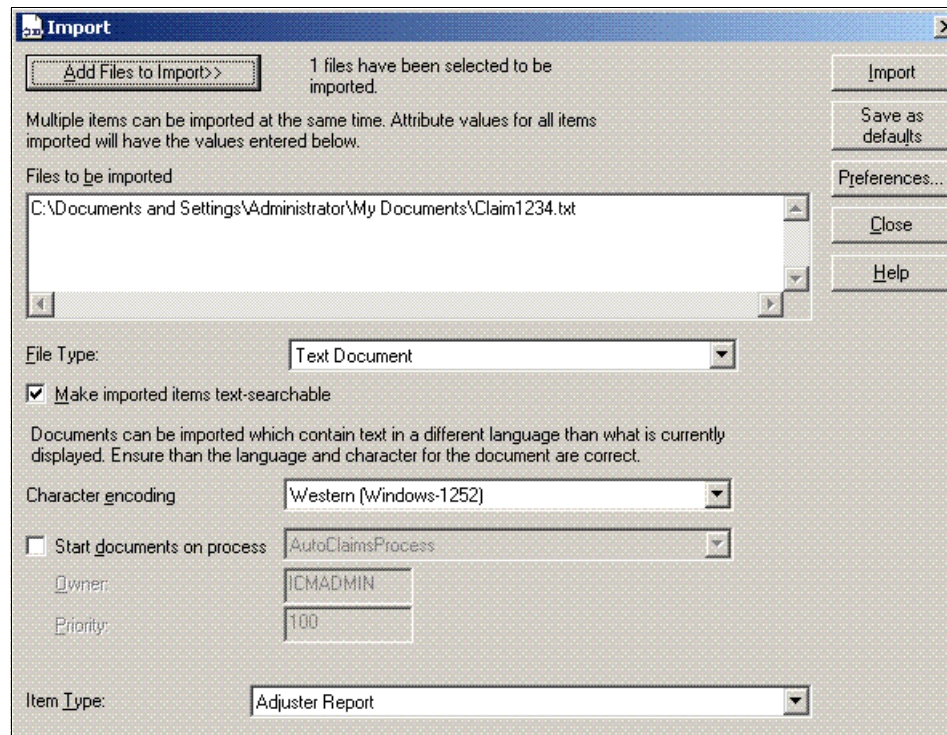


Figure 5-14 Import document using Client for Windows

Search for a document

To search for the document that you just imported (Figure 5-15 on page 100):

1. Press **Ctrl+B**.
2. Choose the item type in context.

3. Enter the required metadata to locate the imported document.
4. Click **OK**.
5. Double-click the document to view it.

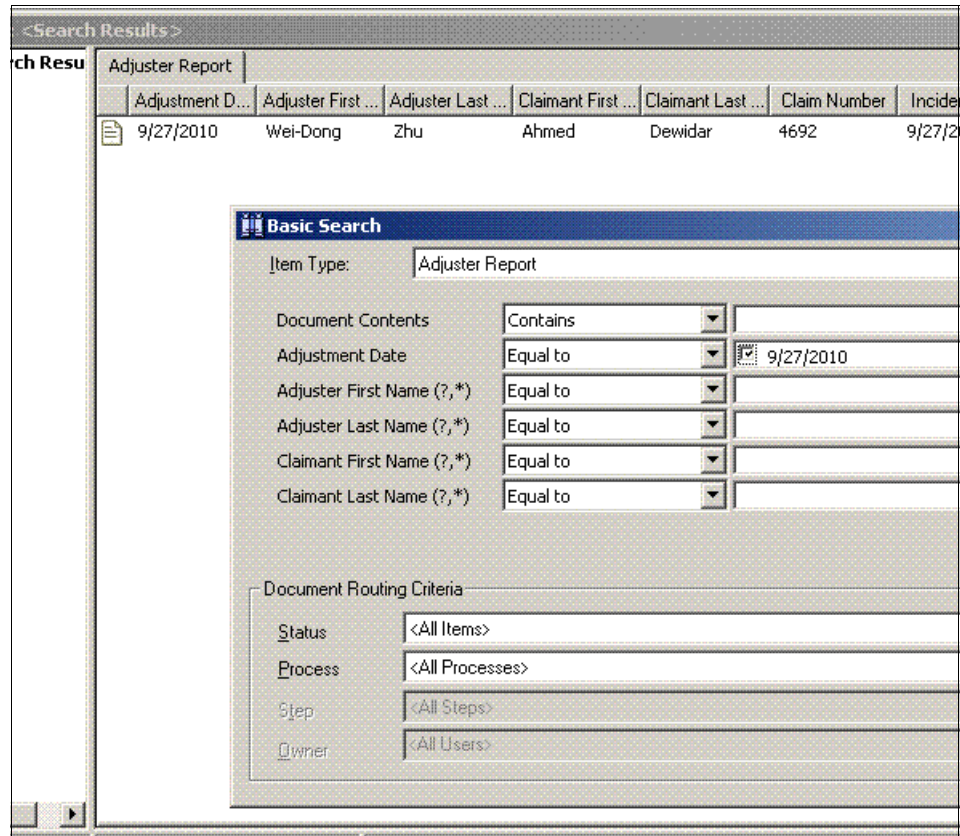


Figure 5-15 Search for a document using Client for Windows

Verify the document storage location

You also need to verify that the document that you just imported is stored on the Information Archive device:

1. Open the Tivoli Storage Manager Administrative Command Line Client by clicking **Start** → **Run** → **Type dsmdmcc**.

2. Execute the following query:

```
SELECT node_name, SUM(num_files) FROM occupancy GROUP BY node_name
```

Compare this number with the number that you obtained when you executed the query before adding the document.

3. As indicated in Figure 5-16, check to make sure that the number is increased by one, which indicates that the document is stored successfully by IBM Content Manager on the Information Archive device.

```
tsm: CMDEMO_SERUER1>select node_name, SUM(num_files) as "Number of Files"  
ccupancy GROUP BY node_name  
  
NODE_NAME          Number of Files  
-----          -  
CMDEMO              1  
  
tsm: CMDEMO_SERUER1>
```

Figure 5-16 Verifying from Tivoli Storage Manager that document is archived

This concludes our scenario. We have successfully tested the integration between IBM Content Manager and the Information Archive device by adding a document to the system, which is already configured to store documents on the Information Archive device, and then by retrieving the document.

5.5 Hints and tips

This section provides some hints and tips to assist your integration task in the area of storage system configuration, staging folder, and logs.

5.5.1 Configuring storage system

You might experience errors while configuring Tivoli Storage Manager storage system on the Resource Manager. These errors are mainly due to the lack of communication between the Resource Manager and the Tivoli Storage Manager server.

We suggest the following procedure to resolve the issue:

1. Verify that the Tivoli Storage Manager client APIs are installed on the Resource Manager system. Make sure that the following path exists:
C:\Program Files\tivoli\tsm\baclient\dsm.opt
2. Ensure that DSMI_DIR and DSMI_LOG_PATH are properly set:
 - a. Enter the following URL in a web browser on the Resource Manager system:
<https://localhost:9443/icmrm/admin>
 - b. Log in with the radmin user ID and the corresponding password.
 - c. Ensure that the values for DSMI_DIR and DSMI_CONFIG are set properly.

5.5.2 Staging folder

To quickly identify if there is a problem in retrieving documents from the Information Archive device, you can check whether the corresponding document is copied to the staging folder. The staging folder is a designated folder that is specified during the Content Manager installation (the default is `C:\staging`) to cache documents coming from Tivoli Storage Manager-managed storage devices.

If the document is copied to the staging folder, the problem exists in the client side; otherwise, the error is on the Resource Manager side.

5.5.3 Logs

The Resource Manager provides multiple services, including data replication, migration, and other services. The Resource Manager records errors into multiple 2 MB-capped (default) text log files called `icmmr.logfile`.

The Information Archive device is concerned about the Resource Manager Migrator service. To configure the log for the migrator service, you can change the configuration specified in the `icmmr_migrator_logging.xml` file. Log files are a good place to easily identify potential problems.

5.6 Summary

In this chapter, we provided an overview of IBM Content Manager and its integration with the Information Archive device as an archive storage system with hierarchical storage devices and retention capabilities. We described the benefits of using the Information Archive device with IBM Content Manager and explained the following integration architecture:

- ▶ Deployment model
- ▶ Data model
- ▶ Retention
- ▶ Hierarchical storage

We also covered the steps that are required to set up IBM Content Manager and the Information Archive device to enable this integration.

Finally, we provided hints and tips to help with configuration and troubleshooting.



Integrating IBM Content Manager OnDemand with the Information Archive device

In this chapter, we discuss the integration between IBM Content Manager OnDemand and the IBM Information Archive device as an archive storage system. We begin with a general discussion of Content Manager OnDemand from an archive management perspective. Content Manager OnDemand uses the Tivoli Storage Manager APIs to communicate with the Information Archive device. We describe the required configuration to set up Content Manager OnDemand and the Tivoli Storage Manager APIs to enable the usage of the Information Archive device as a storage archive.

This chapter includes the following topics:

- ▶ Introduction
- ▶ IBM Content Manager OnDemand overview
- ▶ Integration architecture
- ▶ Integration implementation
- ▶ Hints and tips
- ▶ Summary

6.1 Introduction

Today's organizations face the challenge of archiving information for long-term retention. For many organizations, Content Manager OnDemand is already part of their overall enterprise content management for long-term retention of computer generated reports, customer documents (such as bank account statements, credit card statements, policy documents, and utility bills), and scanned images (such as customer correspondence and checks). Such organizations can now require a compliance storage solution for the ongoing management of their content collections. Content Manager OnDemand integrates with the Information Archive device to provide an end-to-end archiving infrastructure.

This chapter describes the integration between Content Manager OnDemand and the Information Archive device.

6.2 IBM Content Manager OnDemand overview

Content Manager OnDemand is a high performance, highly scalable system that allows for the archiving and retrieval of report data. Content Manager OnDemand is designed such that it can run on a variety of platforms including AIX, UNIX, Linux, z/Linux, z/OS, IBM i, and Windows systems. This platform diversity allows customers to implement a Content Manager OnDemand system that best suits their needs in terms of cost, scale, and technology.

A Content Manager OnDemand system is composed of a Library Server and one or more Object Servers. Figure 6-1 on page 105, provides an architectural view of Content Manager OnDemand, focusing on the storage management aspects of the system.

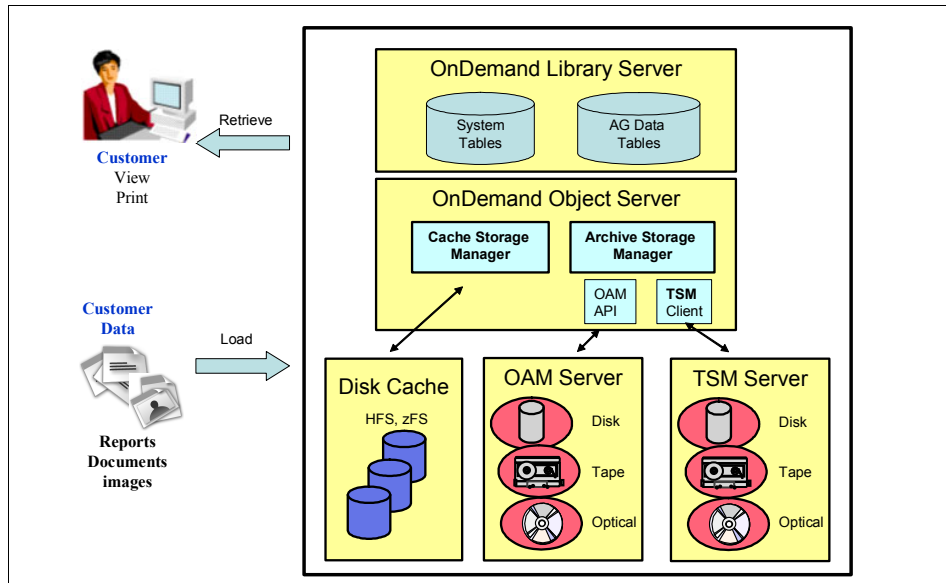


Figure 6-1 Content Manager OnDemand: Storage management architecture

Reports, documents, and images are loaded into the Content Manager OnDemand system archive using one of the Content Manager OnDemand *load* programs. The load process performs the segmenting, indexing, compressing, and storing of the report data. The report data is stored in an Object Server, and indexes to the report data are stored in the Library Server.

Authorized users can retrieve documents from a Content Manager OnDemand system using a Content Manager OnDemand client. To retrieve data from the Content Manager OnDemand system, users first submit a search for the document data that matches their criteria. This search is executed on the Library Server. The search results in a document hitlist from which the user selects the documents to view. The Content Manager OnDemand system then retrieves the document from the appropriate Object Server and displays (or prints) the document.

Within an Object Server, the document data can be stored and managed through a cache storage manager, an archive storage manager, or both.

6.2.1 Cache storage manager

The cache storage manager is used to maintain documents on disk storage. The cache storage manager uses a list of file systems to determine the devices that are available for storing and maintaining documents. Typically, each Content Manager OnDemand Object Server in the system has a defined set of (one or more) cache storage devices on which the report data can be maintained for a specified time period. These cache storage devices provide the fastest access times for system users.

6.2.2 Archive storage manager

The archive storage manager is used to store documents on archive media. The archive storage manager maintains one or more copies of documents and acts as the interface between the Content Manager OnDemand Object Server and archive storage media. Documents can be archived on a variety of media such as disk, optical, and tape. The archive storage devices must be configured and defined to the archive storage manager.

Content Manager OnDemand supports multiple archive storage managers:

- ▶ Tivoli Storage Manager is an archive storage manager provided with multi-platform versions of Content Manager OnDemand. Tivoli Storage Manager is also provided with the z/OS implementation of Content Manager OnDemand.
- ▶ Object access method (OAM) is an additional archive storage manager that is available only on the z/OS Content Manager OnDemand implementation. OAM uses the OSREQ macro interface and keeps online storage objects in OAM DB2 tables.

Both the Tivoli Storage Manager and OAM archive storage managers provide hierarchical data management, implying that without Content Manager OnDemand intervention, they are capable of migrating data from fast DASD to slow DASD to Optical to Tape. This migration is based on administrator specified data storage policy definitions.

This chapter focuses on the Tivoli Storage Manager implementation because it is common to all Content Manager OnDemand platforms.

For more information about IBM Content Manager OnDemand, refer to:

<http://www.ibm.com/software/data/ondemand/>

6.3 Integration architecture

The Information Archive device is the latest generation of information retention solutions. It is designed as a universal storage repository for all types of content (structured or unstructured) to help organizations of any size address complete information retention needs.

The Information Archive device is capable of storing data through multiple access methods. Content Manager OnDemand uses Tivoli Storage Manager as the access method to the Information Archive device. Content Manager OnDemand's use of the Tivoli Storage Manager APIs goes back to the DR450 and DR550 devices. Each Information Archive device is capable of storing 600 TB (raw capacity) of data. Content Manager OnDemand's horizontal scalability architecture allows for the integration of multiple Information Archive devices into a single Content Manager OnDemand implementation, thus providing for extremely large data archives.

Figure 6-2 illustrates the components that are involved in the Content Manager OnDemand and the Information Archive device integration (using Tivoli Storage Manager client API).

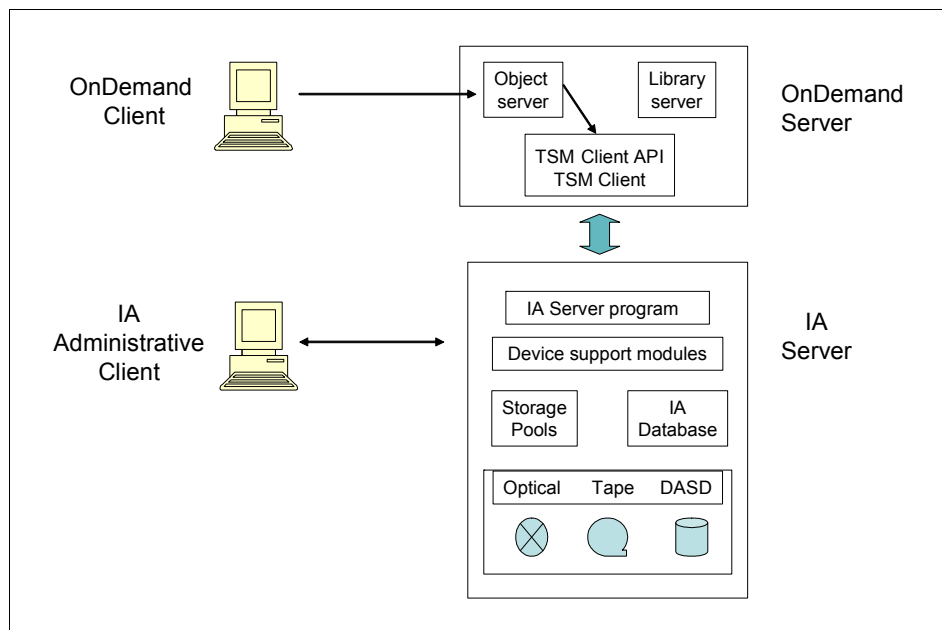


Figure 6-2 Content Manager OnDemand and the Information Archive device integration

Content Manager OnDemand includes the following components:

- ▶ The Content Manager OnDemand Library Server
Contains the Content Manager OnDemand system database tables and metadata database tables. The Library Server responds to user search requests for documents that meet specified search criteria. Embedded in the returned document hitlist is the location of the document.
- ▶ The Content Manager OnDemand Object Server
When a user selects a specific document to view (or print) off of the document hitlist, the Content Manager OnDemand client uses the embedded document location to direct the request to the appropriate Object Server that is attached to the Tivoli Storage Manager server that contains the document data. The document data is then retrieved from the Tivoli Storage Manager server and returned to the Content Manager OnDemand client.

Content Manager OnDemand and the Information Archive device enablement includes the following components:

- ▶ A Tivoli Storage Manager API that Content Manager OnDemand uses to communicate with the Information Archive device. The Tivoli Storage Manager API is required on the Library Server and all object servers that use Tivoli Storage Manager.
- ▶ A server program that maintains a database of information about the devices and data that it manages. The server program also controls the storage media and devices that you define to the Information Archive device.
- ▶ Device support modules, which provide support for storage devices and storage libraries.
- ▶ An administrative client program that you can use to control and monitor the following server program activities and define storage management policies:
 - Expiration processing, which is the process of deleting data that is eligible to be removed from the system.
 - Reclamation processing, which is the process of reclaiming the space taken by expired data. Storage volumes that have been reclaimed can be reused.
 - The storage management policies determine where data is stored and how long the Information Archive device maintains the data.

6.3.1 Supported configurations

A Content Manager OnDemand system is composed of a single Library Server and one or more Object Servers. When Tivoli Storage Manager is used as an archive manager logically attached to a Content Manager OnDemand object

servers, this configuration allows for the following basic Content Manager OnDemand system configurations to be supported:

- ▶ A single system Content Manager OnDemand implementation in which the Library Server, an Object Server, and the Tivoli Storage Manager client are installed.
- ▶ A multi-system Content Manager OnDemand implementation in which the Library Server is installed on one system and one or more other systems containing the Object Server and Tivoli Storage Manager clients are installed.

In both of these configurations, the Tivoli Storage Manager server can be local to the Object Server or can be on a remote system that is TCP connected to the Object Server.

For operating system specific information and details about the installation requirements, refer to the appropriate Content Manager OnDemand Installation and Configuration Guide.

6.3.2 Content Manager OnDemand data stored in the Information Archive device

The Information Archive device can be used to manage document storage for Content Manager OnDemand. Content Manager OnDemand accesses the Information Archive device through the Tivoli Storage Manager APIs. You can use Information Archive device to store the following information:

- ▶ Content Manager OnDemand report data (documents)
- ▶ Document indexes (metadata), which can be migrated from cache to the Information Archive device
- ▶ DB2 backup images that can be archived to the Information Archive device
- ▶ DB2 archived log files

6.3.3 Content Manager OnDemand data migration to the Information Archive device

Using the Content Manager OnDemand Administrative client, you can specify when document data is migrated to archive storage, which can occur in the following situations:

- ▶ The document is originally loaded into the system. The document can be stored to the Information Archive device (using the Tivoli Storage Manager APIs) and another copy of the document can be stored to cache storage or on both places.

- ▶ The next time that the migration maintenance process is run, the document or a copy of the document can be moved to the Information Archive device.
- ▶ If the document is originally saved to cache storage, it can be migrated to the Information Archive device after a defined number of days have passed.

6.3.4 Retention: Coordinating Tivoli Storage Manager and Content Manager OnDemand policies

The Content Manager OnDemand application group storage management settings determine how long report data and indexes are kept in cache storage before being expired. The following choices need to be made concerning how soon data is migrated to the archive storage after the report load is completed:

- ▶ Cache data: The cache data setting determines if the report data is stored in DASD cache and, if so, how long it is kept in cache before it is expired.
- ▶ Life of data and indexes: The life of data and indexes settings determine the length of time that report data, indexes and resources are maintained in the Content Manager OnDemand system before they are deleted from the application group.
- ▶ Expiration type: The following expiration types determine how report data, indexes, and resources are expired:
 - Load: If the expiration type is load, an input file at a time can be deleted from the application group. The latest date in the input data and the life of data and indexes determines when Content Manager OnDemand deletes the data from the Information Archive device. Data that has been stored in archive storage (the Information Archive device) is deleted by the storage manager based on the archive expiration date. Load is the recommended expiration type.
 - Segment: If the expiration type is segment, a segment of data at a time is deleted from the application group. The segment must be closed and the expiration date of every record in the segment must have been reached before the data is expired. If small amounts of data are loaded into the application group, and the maximum rows value is high, the segment can be open for a long period of time, and the data is not be expired for the period.
 - Document: If the expiration type is document, a document at a time is deleted from the application group. Storing with an expiration type of document causes the expiration process to search through every document in the segment to determine if the expiration date has been reached resulting in long processing times.

Using Load as expiration type: To ensure that data to be expired from the Content Manager OnDemand database and the Information Archive device are in synchronization, we recommend using Load as the expiration type.

6.3.5 Protecting data with the data retention protection protocol

To avoid the accidental erasure or overwriting of critical data, Content Manager OnDemand now supports the Tivoli Storage Manager Version 5.2.2 or later APIs related to data retention. Specifically:

- ▶ Data retention protection (DRP)

Prohibits the explicit deletion of documents until their specified retention criterion is met. Although documents can no longer be explicitly deleted, they can still expire.

Attention: DRP is permanent. After it is turned on, it cannot be turned off.

- ▶ Event-based retention policy

Retention based on an external event other than the storage of data. For Content Manager OnDemand, the retention event is the call to delete the data. A load, unload, application group delete, or expiration of data triggers the retention event.

Refer to your platform specific the Content Manager OnDemand Administration Guide for further details.

6.4 Integration implementation

This section explains how to integrate Content Manager OnDemand with the Information Archive device. The high-level steps are as follows:

1. Integrate Content Manager OnDemand with the Information Archive device.
2. Modify the configuration files.
3. Customize the setup.

6.4.1 Integrate Content Manager OnDemand with the Information Archive device

Content Manager OnDemand accesses the Information Archive device using the Tivoli Storage Manager APIs. This section describes the components of Tivoli Storage Manager that you need to configure to enable Content Manager OnDemand to store data in the Information Archive device. This configuration requires the following high-level steps:

1. Enable Tivoli Storage Manager support in Content Manager OnDemand.
2. Configure Tivoli Storage Manager to support Content Manager OnDemand.
3. Configure the Tivoli Storage Manager client node.

Enable Tivoli Storage Manager support in Content Manager OnDemand

To store Content Manager OnDemand archive data in Tivoli Storage Manager, you need to enable the Content Manager OnDemand Tivoli Storage Manager support. Usage of Tivoli Storage Manager allows Content Manager OnDemand to store data in the Information Archive device.

To enable Content Manager OnDemand to use Tivoli Storage Manager:

- ▶ A Tivoli Storage Manager server must be installed and running, and it must be configured to allow Content Manager OnDemand to store and retrieve objects. For information about installing and configuring a Tivoli Storage Manager server, refer to the Tivoli Storage Manager documentation. A starting point is the Tivoli Storage Manager Quick Start Guide for the Tivoli Storage Manager server you are using. For example, *IBM Tivoli Storage Manager for z/OS Quick Start*, GC32-0777.
- ▶ The Tivoli Storage Manager client API must be installed where the Content Manager OnDemand servers (either Library Server or Object Server) are installed. Details on configuring the API can be found in the *IBM Tivoli Storage Manager for UNIX Backup-Archive Clients Installation and User's Guide*, GC32-0789, and in the README.API delivered with the API.
- ▶ A Tivoli Storage Manager server must be configured and active that supports that level of the Tivoli Storage Manager API. The Tivoli Storage Manager server can reside on a different host or operating system than the Content Manager OnDemand server.

- ▶ Define the Content Manager OnDemand storage nodes that use Tivoli Storage Manager. The Content Manager OnDemand Administrative Client is used to define Content Manager OnDemand Storage nodes that use Tivoli Storage Manager. You can download the most recent Administrative client from:

<http://www.ibm.com/software/data/ondemand/mp/support.html>

We describe the actions that you need to complete to configure Content Manager OnDemand to use Tivoli Storage Manager in the following sections.

Configure Tivoli Storage Manager to support Content Manager OnDemand

The Content Manager OnDemand system uses the Tivoli Storage Manager API client to store data to and retrieve data from the Tivoli Storage Manager server. The Tivoli Storage Manager server is managed and administered independently of Content Manager OnDemand. Thus, the Tivoli Storage Manager administrator must ensure that the following conditions are met:

- ▶ All the normal requirements for Tivoli Storage Manager storage are monitored and managed accordingly.
- ▶ All required Tivoli Storage Manager policies, management classes, storage pools, and volumes are defined accordingly.
- ▶ All required Tivoli Storage Manager storage pools and volumes are online.
- ▶ All Tivoli Storage Manager storage pools and volumes have sufficient storage space to satisfy the needs of Content Manager OnDemand.
- ▶ The Tivoli Storage Manager server is active when Content Manager OnDemand needs to read from or write to its storage repository.

If your Tivoli Storage Manager configuration is not set up to support Content Manager OnDemand, then Content Manager OnDemand requests that require Tivoli Storage Manager services will fail. The Tivoli Storage Manager administrator should examine the system to ensure that it is set up to support the storage and retrieval of data by Content Manager OnDemand.

With Content Manager OnDemand, an application group definition includes a client node that is configured and registered in Tivoli Storage Manager. This node is the only link between Content Manager OnDemand and Tivoli Storage Manager. Within Tivoli Storage Manager:

- ▶ The client node is bound to one and only one Policy domain.
- ▶ Each policy domain has only one Active Policy set.

- ▶ A policy set can have many management classes, but only one of them is the default management class. The management class can contain a backup or archive copy group that will point to a storage pool.

Configure the Tivoli Storage Manager client node

Figure 6-3 is an example of a Tivoli Storage Manager client node called *ODNode*. It is assigned a policy domain called *STANDARD*.

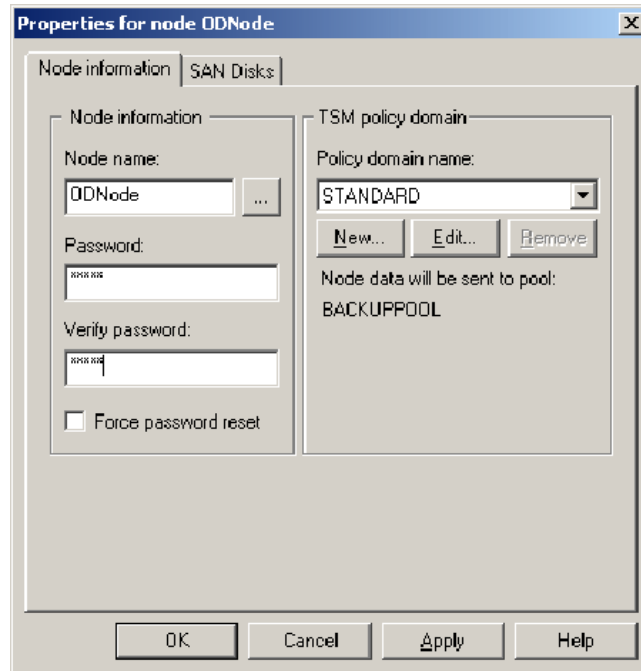


Figure 6-3 Tivoli Storage Manager client Node

The node name and password must match the Content Manager OnDemand storage sets that uses this Tivoli Storage Manager node. BACKUPPOOL is the default storage pool for this policy domain. Associate the client name with the storage pool that is set up to maintain the archive data on the desired device type for the period of time that is required. You can associate the new client node with a different storage pool by selecting **New** to create a new policy domain.

The Tivoli Storage Manager management policies define where to store the data, on which devices, and for how long.

Client nodes that you have registered can be configured to back up data to this Tivoli Storage Manager server instance. The backup data is managed according to the way you set up the client's associated storage pool hierarchy.

Note: Each client requires a client options file (`dsm.opt`) that contains options that identify the node, the server, and the communication method.

6.4.2 Modify the configuration files

There are two Tivoli Storage Manager client files and a Content Manager OnDemand file that need to be configured:

- ▶ Tivoli Storage Manager servers system options file (`dsm.sys`)
- ▶ Tivoli Storage Manager client options file (`dsm.opt`)
- ▶ Content Manager OnDemand configuration file (`ars.cfg`)

Modify the Tivoli Storage Manager servers system options file

Create the Tivoli Storage Manager servers file from the sample servers file, which is located in the `/usr/lpp/Tivoli/tsm/client/api/bin` directory:

```
cp dsm.sys.smp dsm.sys
```

Update the Tivoli Storage Manager servers file to identify the name of the server and the fully qualified TCP/IP host name on which the Tivoli Storage Manager server runs, enable TCP/IP as the communications protocol, set the TCP/IP port address of the server (for example, 1500), and turn off compression:

```
SERvername archive  
TCPServeraddress fully.qualified.tcpip.host.name  
COMMmethod TCPip  
TCPport 1500  
COMPRESSION OFF
```

Save any changes that you make to the `dsm.sys` file.

Note: The `dsm.sys` file can specify multiple servers for use by multiple users of the Tivoli Storage Manager client API.

Modify the Tivoli Storage Manager client options file

Create the client options file from the sample options file, which is located in the `/usr/lpp/Tivoli/tsm/client/api/bin` directory:

```
cp dsm.opt.smp dsm.opt
```

Update the client options file to identify the name of the server and set the processing mode:

```
SErvername archive  
QUIET
```

The name that you specify must be the same as the name of the server that you specified in the `dsm.sys` file.

As an alternative, you can copy the `dsm.opt` file to another directory (for example, `/etc/ars`) or even give it a name other than `dsm.opt` (`/ars/ars/archive.dsm.opt`). This method provides additional configuration flexibility. Each variant of the `dsm.opt` file can specify a different server in the `dsm.sys` file that is located in the `/usr/lpp/Tivoli/tsm/client/api/bin` directory.

Note: For further information about specifying the contents of the Tivoli Storage Manager API user options file (`dsm.opt`) and the Tivoli Storage Manager API server file (`dsm.sys`), see *IBM Tivoli Storage Manager for UNIX Backup-Archive Clients Installation and User's Guide*, GC32-0789.

Modify Content Manager OnDemand configuration file

This section describes the Tivoli Storage Manager and Content Manager OnDemand configuration file parameters that you need to modify to connect to the Information Archive device through Tivoli Storage Manager.

Library and Object Server parameters

You need to add the following parameters to the `ars.cfg` file for any Content Manager OnDemand library or Object Server that is to interface with Tivoli Storage Manager:

▶ **ARS_STORAGE_MANAGER**

This parameter determines whether the Content Manager OnDemand server program is linked to a cache-only storage manager or an archive storage manager. You must specify this parameter on both the Content Manager OnDemand library and object servers. You can specify one of the following values:

- `ARS_STORAGE_MANAGER=CACHE_ONLY...`

Causes the Content Manager OnDemand server program to use a cache-only storage manager. Note: This does not enable Tivoli Storage Manager.

- `ARS_STORAGE_MANAGER=TSM ...`

Enables the Content Manager OnDemand Tivoli Storage Manager support.

Note: Before Content Manager OnDemand can work with a Tivoli Storage Manager to maintain data, you must install and configure Tivoli Storage Manager.

▶ **ARS_DB2_TSM_CONFIG**

This parameter is required if you set the Content Manager OnDemand Library Server `ARS_DB_ENGINE` parameter to DB2 (the default) and if you plan to use Tivoli Storage Manager to maintain DB2 archived log files and backup image files. This parameter specifies the full path name of the Tivoli Storage Manager options file that identifies the Tivoli Storage Manager server that will maintain the DB2 files. For example:

```
/usr/tivoli/tsm/client/api/bin64/dsm.db2.opt
```

The `ARS_DB2_TSM_CONFIG` parameter is ignored on object servers.

Object Server parameters

You must define the following parameters in the `ars.cfg` copy for each Object Server that uses Tivoli Storage Manager to maintain Content Manager OnDemand data:

▶ **LIBPATH**

You must specify a LIBPATH containing the location of the `libApiDS.a` Tivoli Storage Manager client API DLL. For example:

```
LIBPATH=/usr/lpp/Tivoli/tsm/client/api/bin
```

▶ **DSMI_CONFIG**

Identifies the full path name of the Tivoli Storage Manager API options file. For example:

```
/usr/lpp/tivoli/tsm/client/api/bin64/dsm.opt
```

▶ **DSMI_DIR**

Identifies the directory that contains the Tivoli Storage Manager API files. These files are in the `dsmtca`, `dsm.sys`, and the `en_US` language directory. For example:

```
DSMI_DIR=/usr/lpp/Tivoli/tsm/client/api/bin
```

▶ **DSMI_LOG**

Identifies the directory in which Tivoli Storage Manager stores the Tivoli Storage Manager API error log (`dsmerror.log`). The default value is `/tmp`. You must set the `DSMI_LOG` parameter on each Object Server that uses Tivoli Storage Manager to maintain Content Manager OnDemand data. You must specify the directory in which Tivoli Storage Manager stores the Tivoli Storage Manager API error log. For example:

```
DSMI_LOG=/tmp
```

Note: If you are running with multiple Content Manager OnDemand servers, you might want to locate each server's log file in a different directory. If that is the case, you need to **mkdir** each directory and assign appropriate permissions to the directory to allow the Content Manager OnDemand server that is associated with that directory to create files in it.

▶ **DSM_CONFIG**

Identifies the full path name of the Tivoli Storage Manager client options file. For example:

```
/usr/tivoli/tsm/client/api/bin/dsm.opt
```

- ▶ **DSM_DIR**
Identifies the directory that contains the Tivoli Storage Manager client files.
For example:
`/usr/tivoli/tsm/client/api/bin`
- ▶ **DSM_LOG**
Identifies the directory in which Tivoli Storage Manager stores the space management client error log. The default value is `/tmp`.
- ▶ **DSMG_CONFIG**
Identifies the full path name of the Tivoli Storage Manager administrative client options file. For example:
`/usr/tivoli/tsm/client/api/bin/dsm.opt`
- ▶ **DSMG_DIR**
Identifies the directory that contains the Tivoli Storage Manager administrative client files. For example:
`/usr/tivoli/tsm/client/api/bin`
- ▶ **DSMG_LOG**
If you plan to use Tivoli Storage Manager, the directory in which Tivoli Storage Manager stores the administrative client error log. The default value is `/tmp`.
- ▶ **DSMSERV_CONFIG**
Identifies the full path name of the Tivoli Storage Manager server options file.
For example:
`/usr/tivoli/tsm/server/bin/dsmserv.opt`
- ▶ **DSMSERV_DIR**
Identifies the directory that contains the Tivoli Storage Manager server files.
For example:
`/usr/tivoli/tsm/server/bin`

6.4.3 Customize the setup

This section describes the steps that are required to customize the Content Manager OnDemand and the Information Archive device configuration. The following high-level steps are required:

1. Restart the Content Manager OnDemand server.
2. Register client nodes.
3. Create Content Manager OnDemand storage objects.

Restart the Content Manager OnDemand server

You must restart the Content Manager OnDemand server so that the configuration changes made in 6.4.2, “Modify the configuration files” on page 115 take effect.

Register client nodes

A client node links clients and their data with storage volumes and devices. Before Content Manager OnDemand can store data in Tivoli Storage Manager storage, you must register at least one client node. You must register at least one client node in each policy domain that will contain OnDemand data. You can use the example that follows as a guide when registering client nodes. The example presents the procedure with a minimum of customization. If you want to do more, refer to the Tivoli Storage Manager documentation. Enter the command at the Tivoli Storage Manager server command line interface.

For example, to register the client node PRI7YR and password and assign the client node to the OD7YRPD policy domain, enter:

```
register node PRI7YR password domain=OD7YRPD contact='your name'  
archdel=yes
```

An alternative to the command line is to use the Tivoli Storage Manager Administrative web interface as follows:

1. Expand Operation view.
2. Expand Work with client nodes.
3. Select Register a new node.
4. Enter the appropriate information.

Create Content Manager OnDemand storage objects

After updating Content Manager OnDemand to use Tivoli Storage Manager, you need to create the Content Manager OnDemand storage objects that will use archive storage. The storage management criteria that you specify on the Content Manager OnDemand Library Server determines where and when CMOC stores reports and how it maintains them:

- ▶ **APPLICATION:** An administrator creates an Content Manager OnDemand Application for each report that is to be stored on the system.
- ▶ **APPLICATION GROUP:** Applications with similar storage characteristics can be placed into a collection called an Application Group.

- ▶ **STORAGE SET:** Each application group identifies a storage set which is a logical container for one or more storage nodes.
- ▶ **STORAGE NODE:** A storage node identifies an Object Server on which data is stored. A storage Node can be defined as a cache storage node or as a Tivoli Storage Manager storage Node.

When you load a report into Content Manager OnDemand, you specify the application group into which it should be loaded. Content Manager OnDemand will segment, index and compress the report. The document indexes are stored into the AG data tables and the document data is stored in the active storage node for the specified storage group. If the storage node is a cache storage node then a copy of the report will be stored in cache storage on the Object Server. If the storage node is a client node in Tivoli Storage Manager, then Content Manager OnDemand stores a copy of the report in Tivoli Storage Manager.

One or more application groups can specify the same storage set. However, a storage set can write to only one (archive) storage node at a time. This means that all of the data that is written to a storage node will be maintained using the same policy, for example, the type of media, the devices, the length of time to maintain data on the system, and so forth.

Note: When using Tivoli Storage Manager to maintain reports, you should specify the same storage management criteria to Content Manager OnDemand and Tivoli Storage Manager. For example, the Life of Data and Indexes, which is used by Content Manager OnDemand, and the Retention Period, which is used by Tivoli Storage Manager, should be the same value.

For further information about the concepts regarding Applications, Application Groups, Storage Sets and storage nodes refer to the “Document Storage” section in the “Content Manager OnDemand Administration Guide” SC19-2940.

Add a new storage set

To add a storage set using the Administrative client:

1. Right-click **Storage Sets** in the left pane, and select **New Storage Set**.
2. Enter a name and description, and click **Add**.

3. Complete the information on the window, specifying the Tivoli Storage Manager logon and password that you specified to Tivoli Storage Manager when you registered the node to Tivoli Storage Manager. Make sure that **Load Data** is checked. See Figure 6-4.
4. Click **OK** on the Add a Primary Node pane.
5. Click **OK** on the Add a Storage Set pane.

You can now specify this new storage set when you create new application groups.



Figure 6-4 Add a storage set

Update an existing storage set

Alternatively, to update an existing storage set using the Administrative client:

1. Right-click the storage set you want to add a node to and select **Update**.
2. A Update Storage set panel similar to Figure 6-5 displays. You can add one or more Nodes to the storage set.



Figure 6-5 Update a Storage Set panel

Location of Content Manager OnDemand data

A storage set can contain one or more primary storage nodes. When Tivoli Storage Manager is used for archive storage, each storage node that is associated with Tivoli Storage Manager-managed storage must be registered as a client node in a Tivoli Storage Manager policy domain. The Tivoli Storage Manager policy domain properties determine the type of storage devices which are used to maintain the archived data and the length of time that the data is maintained.

The load type parameter determines where Content Manager OnDemand stores data. This parameter has the following possible values:

- **Fixed:** Content Manager OnDemand stores data in the primary storage node that has the load data field selected. When load type is set to fixed, you must select the load data check box for one primary storage node. Content Manager OnDemand loads data to only one primary storage node regardless of the number of primary nodes that are defined in the storage set.

- ▶ Local: Content Manager OnDemand stores data in a primary storage node on the server on which the data loading program executes. When load type is local, the load data check box must be selected for a primary storage node on each of the object servers which is identified in the storage set. A storage set can contain one or more primary storage nodes that reside on one or more object servers.

Note: Storage Manager Expiration is not supported for objects stored in Tivoli Storage Manager. Application Groups with an expiration type of Storage Manager should not be loaded into a Tivoli Storage Manager Storage Node. Because you cannot change the expiration type of existing Application Groups and you cannot change the Storage Node of existing Application Groups, exercise caution when adding Tivoli Storage Manager to existing Storage Sets.

Add a Tivoli Storage Manager storage node

One or more storage nodes are defined (as belonging to) in a storage set. Each storage node identifies a specific Tivoli Storage Manager node on a specific Content Manager OnDemand Object Server. The Tivoli Storage Manager Logon name must belong to a Policy Domain supporting the required Content Manager OnDemand storage management characteristics. The Content Manager OnDemand storage node name is an identifier within Content Manager OnDemand, it is not directly linked to the Tivoli Storage Manager client node. The storage node name can optionally be set the same as the associated client node name.

To add a Tivoli Storage Manager storage node to an existing Content Manager OnDemand Storage Set:

1. Click **Add** on the storage set panel.
2. Enter the appropriate information:
 - Log on information

The logon field must be a valid Tivoli Storage Manager client node name. This is the client node that has been defined on the Tivoli Storage Manager system through the wizard or command line. The password which follows the logon must be the same as the password created for the client node. Content Manager OnDemand uses a Tivoli Storage Manager API to connect and logon to the Tivoli Storage Manager server when data is being migrated to the Tivoli Storage Manager client node.

- Load data

This parameter determines the primary storage node into which Content Manager OnDemand loads data. When the load type is fixed, one primary

storage node must have load data selected. When load type is local, load data must be selected for one primary node for each Object Server that is associated with the storage set.

– Cache only

This parameter determines whether Content Manager OnDemand uses the archive manager for long term storage of data. After installing and configuring Tivoli Storage Manager, creating an Content Manager OnDemand storage set, and assigning it to a Tivoli Storage Manager client node, we are ready to consider how an application group uses the cache storage manager and the archive storage manager to store, maintain, and expire Content Manager OnDemand report data.

- If one of the Primary Storage Nodes has an asterisk (*) next to the name, the node is being used for loading. If you want to start loading to Tivoli Storage Manager, the existing node should not be loaded to.

To stop loading to the existing node:

- i. Select that node, and click **Update**.
- ii. Clear the **Load Data** field. See Figure 6-6.
- iii. Click **OK** on the Update a Primary Node panel.

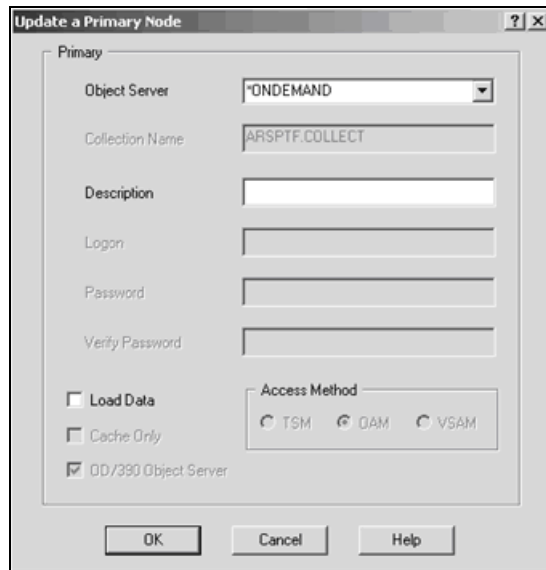


Figure 6-6 Updating a Storage Node

3. On the Update a Storage Set panel, click **Add**.

- Complete the information on the panel, specifying the Tivoli Storage Manager logon and password you specified to Tivoli Storage Manager when you registered the node to Tivoli Storage Manager. Make sure that **Load Data** is checked. See Figure 6-7.

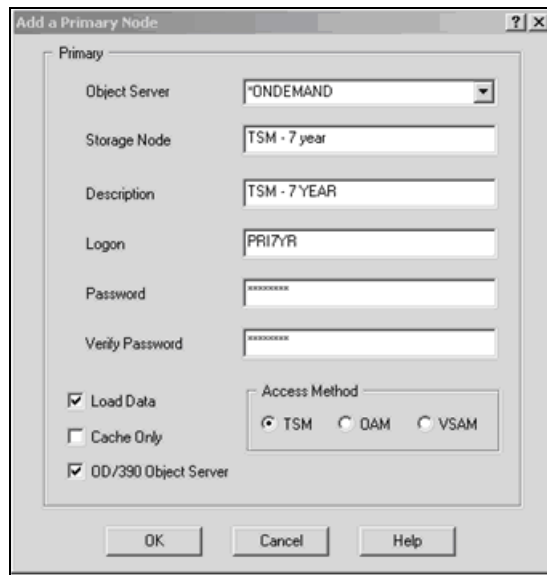


Figure 6-7 Adding a Primary Storage node

- Click **OK** on the Add a Primary Node panel.
- Click **OK** on the Update a Storage Set panel.

New loads to Application Groups using that Storage Set will now load to Tivoli Storage Manager.

6.4.4 Test the integration

Now that you have defined a storage set and a storage node as belonging to a specified Content Manager OnDemand Application group, you can run a test to verify that the settings are correct. To test the integration:

- Load data into the Information Archive device.
- Retrieve the loaded data from the Information Archive device.

During both steps, data is passed from Content Manager OnDemand to the Information Archive device through the Tivoli Storage Manager API, according to the Storage Set and Storage Node definitions that you created for the Application Group.

Load data into the Information Archive device

To load the data:

1. Select a report that is suitable for loading into the Application Group.
2. Load the report using the `arsload` command, for example:

```
arsload -I CMODinstance -nvf -u userID -p passWord -c indexDirf -g  
applicationGroupName loadFileName
```

You can find a more detailed explanation of the `arsload` command in the Content Manager OnDemand Administration guide for your specific platform.

The `arsload` command produces an output file that includes text similar that shown in Example 6-1. The last four lines in this example indicate that the report data was successfully loaded into Content Manager OnDemand.

Example 6-1 Sample output file for the `arsload` command

```
arsload: 08/31/10 11:31:17 -- Indexing started, 2895748561 bytes to process  
.....  
0425-440 ACIF AT PM06753 HAS COMPLETED NORMALLY WITH RETURN CODE 0.  
arsload: 08/31/10 11:38:04 Indexing completed  
arsload: 08/31/10 11:38:04 -- Loading started, 3110905584 bytes to process  
.....  
OnDemand Load Id = >5072-2-0-1FAA-11118-11118<  
Loaded 240075 rows into the database  
Document compression type used - 0D77. Bytes Stored = >3519822< Rows = >2400<  
arsload: 08/31/10 11:40:46 Loading completed  
arsload: Processing successful for file >/afpfc/perf-afpfc<
```

Retrieve the load data from the Information Archive device

To retrieve the load data from the Information Archive device:

1. Log in with the Content Manager OnDemand client (either the GUI client or the web client).
2. From the folder list, select the folder that contains the application group that you just loaded.
3. Open the folder, and perform a search, using search criteria that matches the report that you just loaded.
4. A list of hits displays. Make sure that the total number matches the number of rows loaded into the database (as shown in Example 6-1).
5. Highlight several rows from the list, and click **View all**.

For each highlighted row, the document data will be retrieved from the Information Archive device and displayed.

This concludes the testing scenario. We have successfully tested the integration between IBM Content Manager OnDemand and the Information Archive device by loading a report containing thousands of documents into the Content Manager OnDemand system. These documents were stored to the Information Archive device as specified by the storage set and storage node defined to the Application Group. The document data was then retrieved from the Information Archive device and displayed on the Client screen.

6.5 Hints and tips

This section provides hints and tips for your integration planning and implementation.

6.5.1 Coordinating Tivoli Storage Manager and OnDemand policies

You want to make sure your retention policies are coordinated between what is set in Tivoli Storage Manager and what is set in Content Manager OnDemand. Refer to 6.3.4, “Retention: Coordinating Tivoli Storage Manager and Content Manager OnDemand policies” on page 110 for details.

6.5.2 Ensuring that the Information Archive device is properly installed

You need to make sure that the Information Archive device is operational and is properly installed prior to starting the Content Manager OnDemand and Tivoli Storage Manager configuration process. If the Information Archive device is not properly installed, then during the Content Manager OnDemand integration process, whenever you receive an error message from Content Manager OnDemand, you will not be able to tell if this is an error in the installation process or an error due to the Information Archive device being nonoperational.

6.6 Summary

This chapter provided an overview of the integration of Content Manager OnDemand with the Information Archive device as an archive storage system. We explained how to use the Tivoli Storage Manager APIs as a communication method between Content Manager OnDemand and the Information Archive device.

We also covered the steps that are required to set up Content Manager OnDemand and the Tivoli Storage Manager APIs to use the Information Archive device as a storage archive.

The final section provided some basic hints and tips to help with the installation and configuration process.



IBM Information Archive for Email, Files, and eDiscovery

This chapter describes a specific IBM Smart Archive strategy solution—IBM Information Archive for Email, Files, and eDiscovery. It provides an overview of the solution, the software and hardware in the bundle, its capabilities, configuration, and two use cases. The intention of the chapter is to introduce you the special preinstalled, preconfigured, easily-to-be implemented solution and demonstrates how IBM Enterprise Content Management products and offerings combined with the IBM Information Archive device can effectively and efficiently meet your business requirements.

The chapter includes the following topics:

- ▶ Introduction
- ▶ The solution bundle
- ▶ Solution configuration
- ▶ Use cases
- ▶ Summary

7.1 Introduction

As part of the IBM Smart Archive strategy, Information Archive for Email, Files, and eDiscovery is one specific solution offering that helps to simplify and accelerate the implementation of an end-to-end archiving and eDiscovery solution from weeks to days. This solution combines the Information Archive device and the IBM Enterprise Content Management products, and it helps to ensure the success of information archiving projects.

Figure 7-1 shows how the Smart Archive strategy combines software, hardware, and services to allow organizations to intelligently and systematically identify what information needs to be collected, classified, and archived and to do it in a unified way. There are many products, offerings, and services available that materialize the Smart Archive strategy to deliverable solutions. The items in the figure that are highlighted with a square show the elements that this particular solution uses to help organizations with email and file system archiving and eDiscovery challenges.

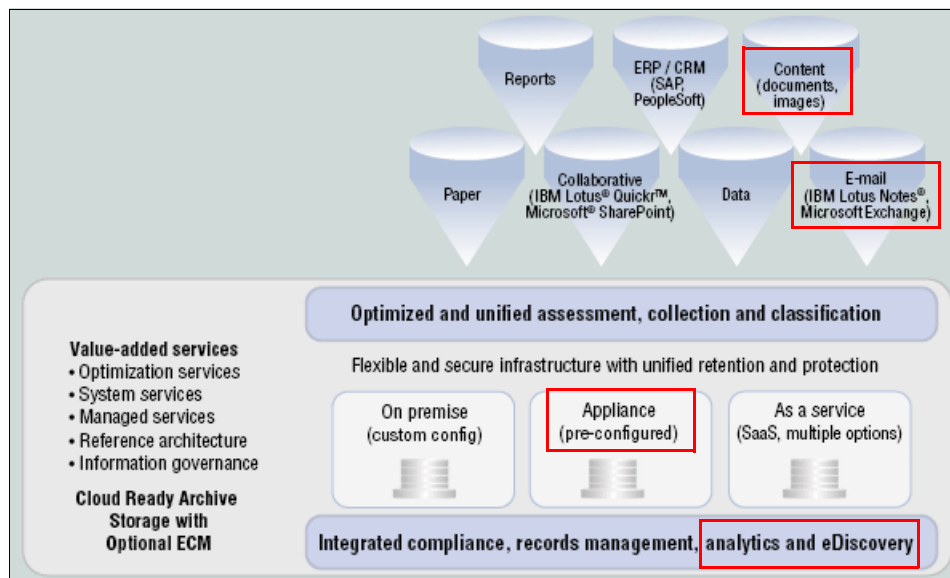


Figure 7-1 IBM Smart Archive strategy

This solution is unique because it bundles everything an organization needs to implement the solution in-house from beginning to end from a single vendor. This solution provides an end-to-end archiving and eDiscovery solution that delivers the following features and functions:

- ▶ IBM software and hardware preintegrated, preconfigured, and preinstalled. Organizations can simplify implementation and get the system up and going in a short time period. It is optimized to support up to 1,400 users.
- ▶ Built-in archiving and eDiscovery functions, based on IBM archiving and eDiscovery products. It offers comprehensive content collection, archiving, and eDiscovery capabilities for email and file systems.
- ▶ Flexible, high performance archiving infrastructure that is scalable, secure, with multiple-protection levels. It can be used as the base for this offering and can also be used for other types of archiving.

The archiving and eDiscovery built-in functions include mailbox management and journaling, file systems archiving, litigation, audit and investigation readiness, and advanced content analytics for early case assessment.

7.2 The solution bundle

This solution is a bundle that includes preinstalled, preconfigured, and pretested IBM software, hardware, and implementation services all from IBM. Specifically, the IBM Information Archive for Email, Files, and eDiscovery bundle includes the following package and products:

- ▶ IBM Content Collector Discovery Analytics Starter Pack (CCDA), which includes the following products:
 - IBM Content Collector for Email
 - IBM Content Collector for File Systems
 - IBM eDiscovery Manager
 - IBM eDiscovery Analyzer
- ▶ IBM Content Manager Enterprise Edition
- ▶ IBM Information Archive device
- ▶ IBM System x® servers (x2), already sized and configured to run the software. The CCDA software is preconfigured onto the System x servers for quicker implementation at the customer site.

Figure 7-2 shows the solution architecture.

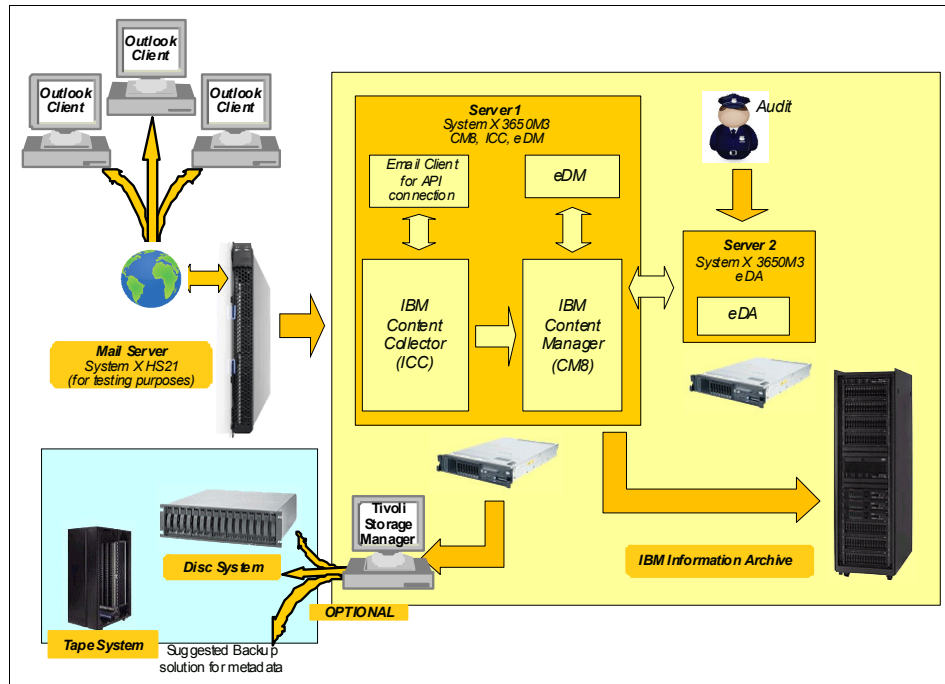


Figure 7-2 IBM Information Archive for Email, Files, and eDiscovery

The total solution configuration uses existing experience in archiving deployments, simplifying implementation and providing quick time to value. In addition, proven services packages are available to manage the ongoing administration and maintenance. This solution is ideal for organizations, with up to 1,400 employees, that have:

- ▶ A requirement to capture and retain email and file system content
- ▶ A requirement to demonstrate compliance with retention mandates
- ▶ Experienced soaring storage costs and need to bring those costs under control
- ▶ Requirements to quickly and efficiently respond to eDiscovery requests

The solution is expandable to meet changing business requirements. It includes comprehensive yet straightforward archiving and eDiscovery capabilities for emails and file system content.

The solution bundle's capabilities include:

- ▶ Mailbox management and journaling
- ▶ File system archiving
- ▶ Litigation, audit, and investigation readiness and response
- ▶ Advanced content analytics for early case assessment

The solution can be expanded to include archiving and eDiscovery of additional information types, such as Microsoft SharePoint, as well as additional capabilities, such as records management and advanced classification, without costly or risky custom integrations.

Information Archive for Email, Files, and eDiscovery uses an infrastructure that supports multiple types of information across industry-standard interfaces from multiple sources.

The same infrastructure can integrate with other Smart Archive solutions, including other on-premise solutions, hosted cloud solutions, or hybrid configurations.

7.3 Solution configuration

The Information Archive for Email, Files, and eDiscovery solution delivers a highly scalable, secure, high performance infrastructure (the IBM Information Archive device) with the flexibility to meet multiple requirements in a single solution. This is a purpose built, universal, scalable, and secure storage repository for all content (structured and unstructured) with up to 608 TB of disk capacity.

System X server 1 uses the following configuration:

- ▶ IBM software
 - IBM Content Collector for Email
 - IBM Content Collector for File Systems
 - IBM eDiscovery Manager
 - IBM Content Manager
- ▶ Base system configuration
 - Operating system: Windows Server 2008 R2 Enterprise Edition 64 bit fully updated
 - Static IP address assigned (preferably)

- Hard disk is partitioned with:
 - C for OS and software stack
 - D for disk image and temp files
 - E for DB2 databases
- ▶ Hardware
 - System X 3650 M3
 - 48 GB RAM
 - 3.4 TB HDD

System X server 2 uses the following configuration:

- ▶ IBM software: IBM eDiscovery Analyzer
- ▶ Base system configuration
 - Operating system: Windows Server 2008 R2 Enterprise Edition 64 bit fully updated
 - Static IP address assigned (preferred)
 - Hard disk is only partitioned with C: drive for operating system and software stack
- ▶ Hardware
 - System X 3650 M3
 - 16 GB RAM
 - 135 GB HDD

Inside the IBM Information Archive device, Figure 7-3 shows the hardware, software, and security set up.

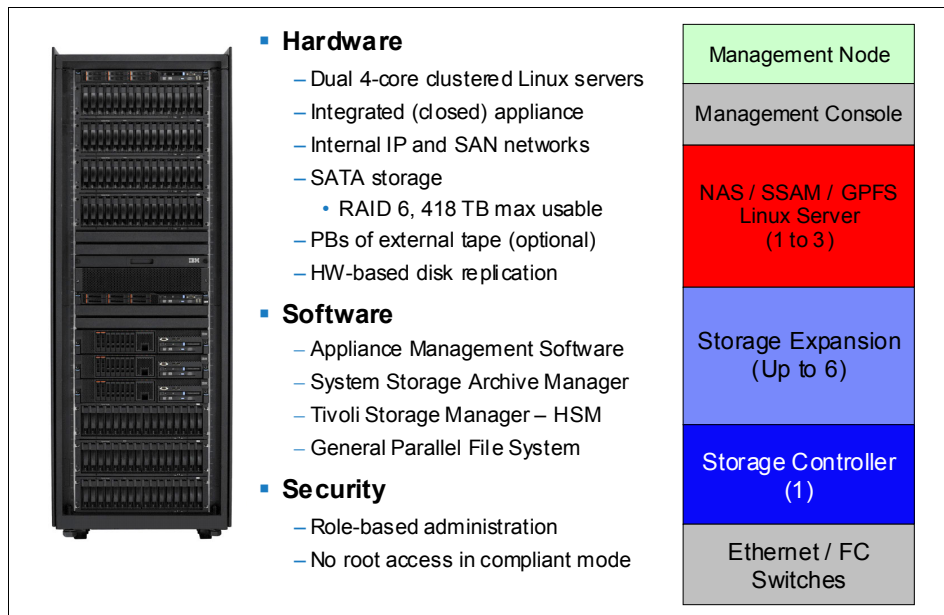


Figure 7-3 Inside the IBM Information Archive device

7.4 Use cases

This section describe two ideal scenarios for using Information Archive for Email, Files, and eDiscovery solutions.

7.4.1 Use case 1: Mailbox archiving, file archiving, and eDiscovery for compliance

A company has a requirement to capture and store all incoming and outgoing e-mail messages and to gain control over file system content for compliance purposes. The company experiences surmounting storage requirements and shortened response deadlines for information inquiries, especially in regulated and heavily litigated industries.

Authorized users have the requirement to perform both simple and sophisticated searches. Comprehensive archiving capabilities (while keeping the very same user experience) and fundamental eDiscovery capabilities, such as litigation hold, early case assessment, culling, and analytics.

If the company meets the following criteria, the company can use the preinstalled, preconfigured, flexible, secure archiving and eDiscovery offering for a practical quick start and be assured that it has a flexible strategic growth path:

- ▶ A short archiving and eDiscovery project implementation time window
- ▶ Not enough expertise and experienced personnel in-house
- ▶ Not enough time to deal with complexities of pulling together the required hardware, software and services for the solution

7.4.2 Use case 2: Archiving for storage and business optimization

A company has identified soaring storage costs and the associated management issues (that is, time and cost of backups), which come with increasing volumes. The company has a requirement to offload seldom accessed but important information, typically older information, from the production e-mail servers and file systems to an online archive repository. The company requirement is that content meeting a particular profile is automatically captured, deduplicated and archived. Users must have the capability to search for and retrieve their own content; authorized users have the ability to search for and retrieve content across mailboxes.

If the company meets the following criteria, the company can use the preinstalled, preconfigured, flexible, secure archiving and eDiscovery offering for a practical quick start and be assured that it has a flexible strategic growth path:

- ▶ A short archiving and eDiscovery project implementation time window
- ▶ Not enough expertise and experienced personnel in-house
- ▶ Not enough time to deal with complexities of pulling together the required hardware, software and services for the solution

7.5 Summary

This chapter provided an overview of IBM Information Archive for Email, Files, and eDiscovery, the preinstalled and preconfigured solution that provides an end-to-end archive and eDiscovery solution from a single vendor that can be deployed in days. The solution offers flexibility for companies to choose any of the delivery methods—on-premise, preconfigured, or cloud—to match their

investment and time to market, based on a unified software code base. It supports the IBM Smart Archive strategy and creates a growth path to manage the life span of structured and unstructured information from creation to disposition. In addition, Information Archive for Email, Files, and eDiscovery offers a long-term solution and the reassurance of vendor viability for companies that are facing archiving and eDiscovery challenges,

Related publications

We consider the publications that we list in this section particularly suitable for a more detailed discussion of the topics that we cover in this paper.

IBM Redbooks publications

For information about ordering these publications, see “How to get Redbooks publications” on page 142. Note that some of the documents referenced here might be available in softcopy only.

- ▶ *IBM Information Archive Architecture and Deployment*, SG24-7843
- ▶ *Content Manager Implementation and Migration Cookbook*, SG24-7051

Online resources

These websites are also relevant as further information sources:

- ▶ The Information Archive device home page:
<http://www.ibm.com/systems/storage/disk/archive>
- ▶ The Information Archive device Information Center:
<http://publib.boulder.ibm.com/infocenter/tivihelp/v37r1/index.jsp>
- ▶ IBM Content Manager home page:
<http://www.ibm.com/software/data/cm/cmgr/mp/edition-enterprise.html>
- ▶ IBM Content Manager product documentation:
<http://www.ibm.com/support/docview.wss?rs=86&uid=swg27015910>
- ▶ IBM Content Manager Information Center:
<http://publib.boulder.ibm.com/infocenter/cmgmt/v8r4m0/index.jsp>
- ▶ IBM Content Manager OnDemand home page:
<http://www.ibm.com/software/data/ondemand>

How to get Redbooks publications

You can search for, view, or download Redbooks, Redpapers, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks publications, at this website:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services



IBM Enterprise Content Management and IBM Information Archive Providing the Complete Solution



Redpaper™

Discusses the concept behind the IBM Smarter Archive strategy

Provides an overview of Information Archive and ECM offerings

Details the benefits and technical aspects of the product integration

The need to archive information is on the rise, driven by content and data growth, regulatory compliance, legal discovery, and data protection requirements. The IBM Smart Archive strategy is a comprehensive, unified, and integrated archive strategy that combines IBM software, systems, and service capabilities that are designed to help organizations extract value and to gain new intelligence from information by collecting, organizing, analyzing, and using that information.

IBM Enterprise Content Management (ECM) products and offerings combined with the IBM Information Archive device provides the type of end-to-end Smart Archive solution that is a critical component of the IBM Smart Archive strategy.

This IBM Redpaper publication focuses on the benefit and technical details of the integration of ECM products and offering with the Information Archive device. The paper introduces technical sales people and IT specialists to the IBM Smart Archive strategy and the integration of Information Archive and ECM products and offerings. At the same time, it provides IT specialists specific guidance about performing the integrations.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks