

IBM Content Collection & Archiving

IBM Enterprise Content Management
Information Lifecycle Governance

eDiscovery
Management



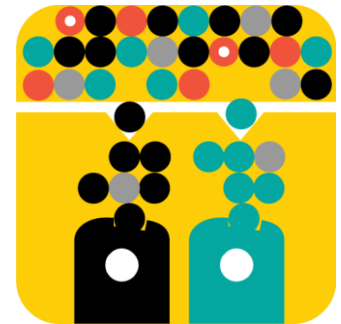
Records and Retention
Management



Archive and
Storage



Defensible
Disposition



Information proliferates...

Volume

12 terabytes
of Tweets create daily

Velocity

5 million
trade events per second

Variety

4 terabytes/
average surveillance video
site/day

15 petabytes
of new information daily

500 million
call detail records per day

80% info
growth
is unstructured content ...

How do you keep up?

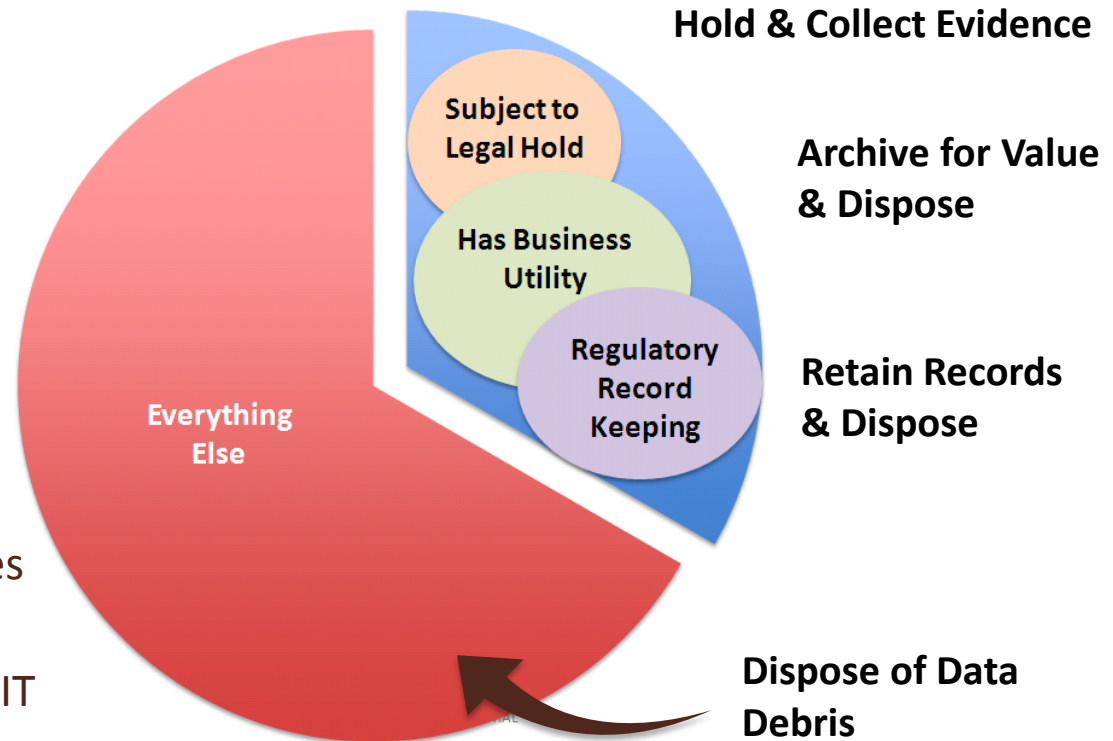
Value-based Archiving with Smart Archive for IT

Value-based Archiving puts the brakes on unchecked growth by providing instrumentation that helps:

- Eliminate duplicate data
- Reduce over retention
- Lower storage costs
- Reduce risk

Smart Archive for IT helps solve common operational & tactical issues

- Ever growing storage infrastructure – ever shrinking IT budget
- Backup used as Archive
- Overwhelmed eDiscovery processing



Across Structured & Unstructured Data

IBM Content Collector

A family of Content Collection and Archiving offerings based on a modular, extensible architecture

A critical part of implementing a value based archiving program to meet multiple requirements across multiple content sources

IBM Content Collector for Email

IBM Content Collector for File Systems

IBM Content Collector for Connections

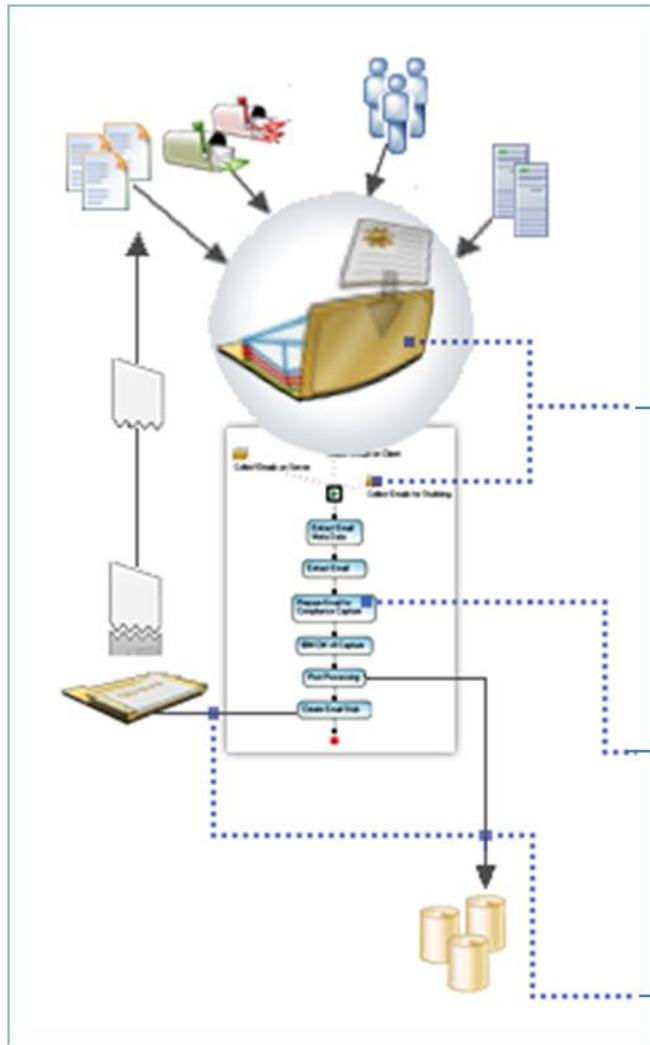
IBM Content Collector for SharePoint

IBM Content Collector for SAP



- 12+ years in market
- 2011 Gartner & Forrester “Leaders”
- IDC 2011 Market Share Leader
- 1000’s of customers archiving millions of documents

IBM Content Collector Overview



Configuration Manager is used to administer ICC.

Task Routes are the principal configuration elements and can be thought of as a workflow. The administrator creates Task Routes that implement the required archiving logic by connecting Collectors and Tasks.

Collectors contain definitions of data sources and surface the options available for that source. Multiple connectors can be installed and configured through one interface

Tasks define the processing steps that can be applied while content passes through the archiving process.

Post-Processing tasks define the processing that occurs after content has been added to the repository.

Task Routing Engine: Much More Than “Straight through” Archiving

Content Collector Task Routing Architecture enables flexible routing and configuration for processing of internal and external instructions (or tasks)

Configuration Manager:

- Drag & Drop Designer
- Define source/target connector configuration & archiving logic
- Pre-configured task route templates

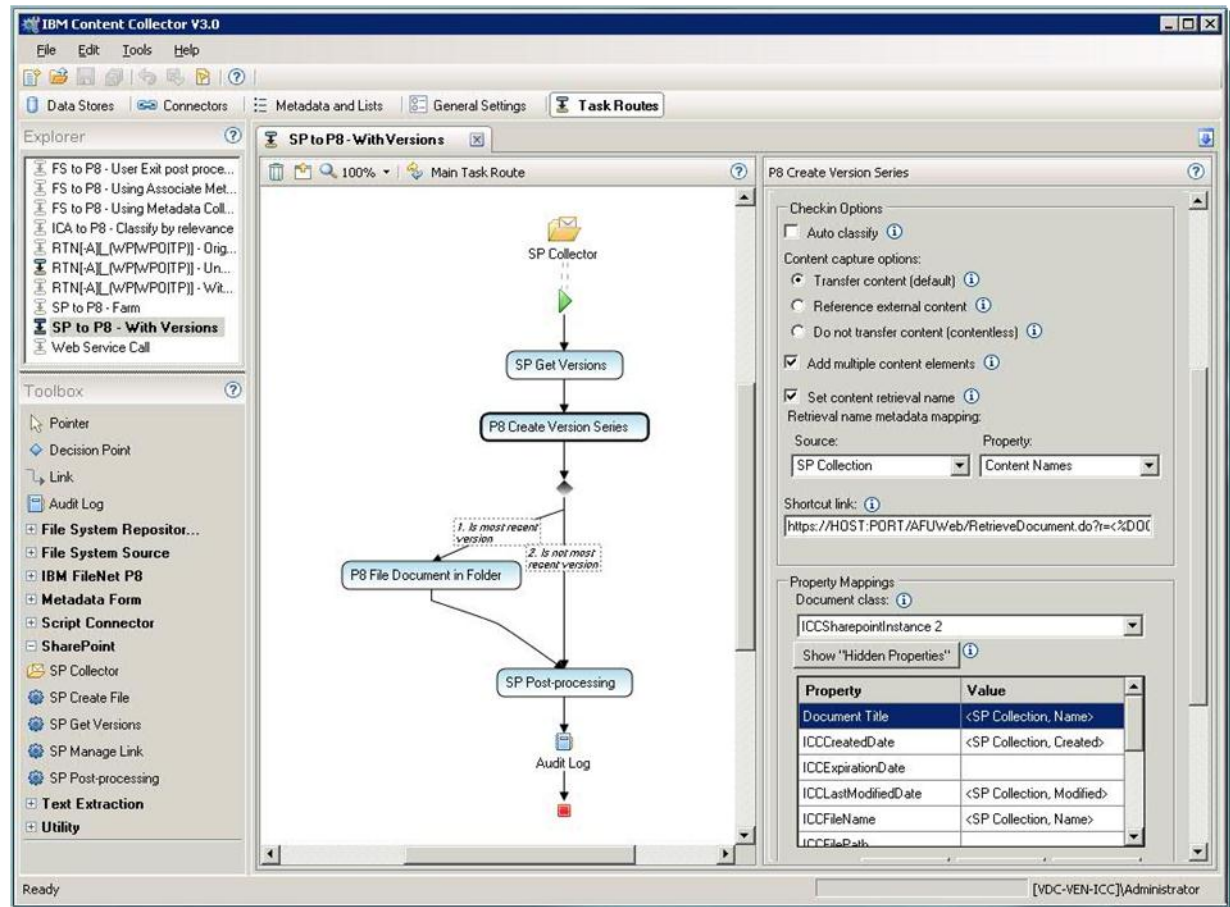
Implement decision points based on meta data

- Selection criteria
- True/False
- Greater than, Less than etc

Error Handling

Provides Administrator visibility to:

- Status Dashboards
- “Problem” documents



Collect, Process and Archive

Across multiple sources and targets

Connectors

- **MESSAGING / E-MAIL**
 - Microsoft Exchange
 - Lotus Domino
 - SMTP Mail Systems
 - PST/NSF content
 - Instant Messaging (3rd party)
- **FILE SYSTEMS**
 - NTFS/DFS File System
 - Novell File Systems
 - Desktops (3rd party)
- **COLLABORATION**
 - IBM Connections
 - Microsoft SharePoint
 - Social Media (3rd party)
- **SAP**
 - Data Archiving
 - Incoming Documents
 - Outgoing Documents

Collect

Policy & Task Engine

- **INTERNAL TASKS**
 - Copy / Move / Delete / Stub
 - De-Duplication
 - Rules-Based Classification
 - Text Extraction
- **EXTERNAL TASKS**
 - Call external applications for:
 - Records Declaration
 - Advanced Classification
 - Rendering (3rd party)
- **CUSTOM TASKS**
 - API Support / Partner Apps
 - Encryption
 - System Lookups
 - 3rd Party Rules

Process

Target Connectors

- **IBM NATIVE**
 - FileNet Content Manager
 - IBM Content Manager v8.4
 - FileNet Image Services (ICC for Files only)
- **OTHER**
 - IBM SmartCloud Archive*
 - Microsoft File System

Archive

Value-Based Archiving with IBM Content

Collector

Multiple options to help meet archiving requirements

- Identify specific content within a source system to process
 - Archive Memos, but not Calendars
 - Wikis but not Blogs...
- Define filters to further refine what gets Archived
 - Dates, Users, file size or properties...
- Apply Advanced Post-processing options native to ICC
 - Delete original; Leave a stub
 - De-duplicate to reduce volume
 - Apply Version control
- Options for expiring content from the repository
 - Expiration Manager can delete content from the repository based on schedules
- Integration with other IBM applications for even more capabilities

The screenshot displays the 'Filter' tab of the 'File System Collector' application. It includes several sections for configuring search criteria:

- Libraries and Lists:** Options to 'Select libraries and lists' and a list of library types such as 'Announcements', 'Calendar', 'Links', 'Shared Documents', 'Tasks', and 'Team Discussion'.
- Filter file extensions:** A dropdown menu set to 'Exclude files' with a text input field containing '.mov;.jpg'.
- Ignore documents:** Checkboxes for 'Already processed', 'Already captured', and 'Where access is denied'.
- File attribute filter:** A table with columns 'Attribute' and 'Action'.

Attribute	Action
Read only	Process
Hidden	Do not process
System	Do not process
Archive	Not used
Temporary	Do not process
Sparse file	Not used
- Date filters:** A table with columns 'Property', 'Type', and 'Date'.

Property	Type	Date
Latest create d...	Not used	Not used
Earliest modified...	Not used	Not used
Latest modified ...	Not used	Not used
Earliest access ...	Not used	Not used
Latest access d...	Relative	90 days ago
- Minimum file size in bytes:** A checkbox that is checked, with a value of '5,000,000'.

Expiration Manager

Expiration Manager enables deletion of content from the repository

Expiration dates set when content is archived

– Stored as metadata & configured at the Task Route level by:

- Specified time period provided or calculated
- User/ LDAP Group Membership

- Retention Periods are day forward and apply to new content but not for already archived content
- Expiration Manager runs against the repository and checks for documents that are older than their retention date.
 - Has options to list, count, or delete documents

Calculate Expiration Date

General

Name: Calculate Expiration Date

Description: This task sets a retention (expiry) date on content based on input metadata, which can include system metadata, user

Configuration

Date set by metadata matching

Metadata and date for evaluation:

Evaluation metadata: <Conditional Expression>

Source: File Property: Created Date

Retention periods:

Match Term	Retention Period (days)
Executive	720
Transitory	180

ICC for File Systems, eMail and SharePoint all support stub removal from the source when content is deleted from the repository

Integration with IBM eDiscovery Manager

eDiscovery processing against the data collected from multiple sources in place, without further collection

Legal can create cases, apply holds or export case data across all archived data directly - without requiring IT involvement

Reduce process steps and avoid duplication of data.

The screenshot displays the IBM eDiscovery Manager interface. At the top, the user is logged in as 'administrator' and the current case is '--Credit CARD Act Investigation'. The interface includes a navigation pane on the left with options like 'Case Information', 'Search', 'Folders', 'Case Details', 'Task Status', 'Case Searches', 'Saved Searches', and 'Scheduled Searches'. The main area shows a list of 128 items with columns for 'Type', 'Native view', 'Sender', and 'Subject'. An 'Export Content' dialog box is open in the foreground, allowing users to configure export settings. The dialog includes fields for 'Export format' (set to 'EDRM XML'), 'Export file name prefix' (set to 'SomeCompany'), 'Notification email addresses' (set to 'carl@somecompany.com'), 'Export task name' (set to 'Export_Task_20110623T17010047'), 'Export directory' (set to 'Export_Task_20110623T17010047'), and 'Export task description' (set to 'Items needing Secondary Legal Review for case'). There is also a 'Schedule Task' checkbox and a 'Start time' section with 'Date' (6/23/2011) and 'Time' (20:00) fields.

Export data in multiple formats, including native format or in the industry standard EDRM XML

Adding IBM Content Classification

Language processing technology that can analyze digital content and apply a predefined taxonomy to discover possible categories and relevancy scores.

Can be called from an ICC task route to:

- **Assign a folder path for the archived e-mails or documents on file system**

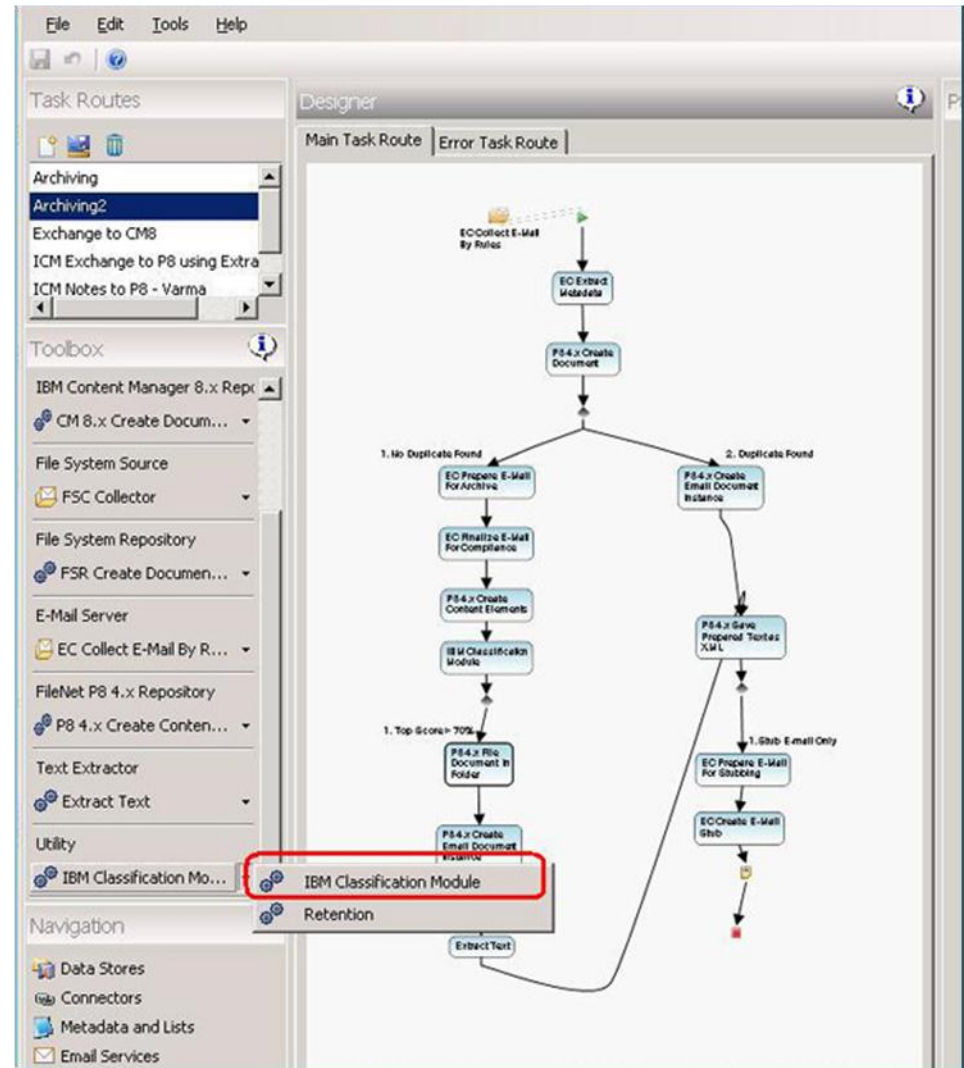
Archiving e-mails or documents on file systems into known folders

- **Identify e-mails that should be declared as records**

To ensure enterprise compliance with record retention and disposition policies, automatically analyze e-mails for business-sensitive data and declare such an email as a record - with no user input.

- **Add metadata properties**

The metadata of the archived e-mails can be populated with the top category names or top relevancy scores.



IBM Content Collector The Connectors

IBM Content Collector for email: Microsoft Exchange Archiving

IBM Content Collector collects, processes and archives content from Exchange servers.

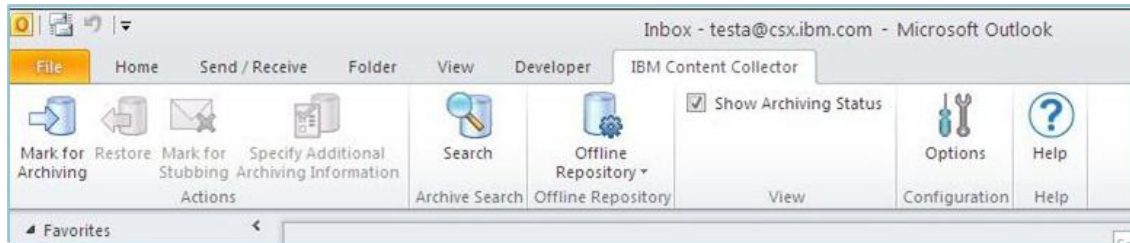
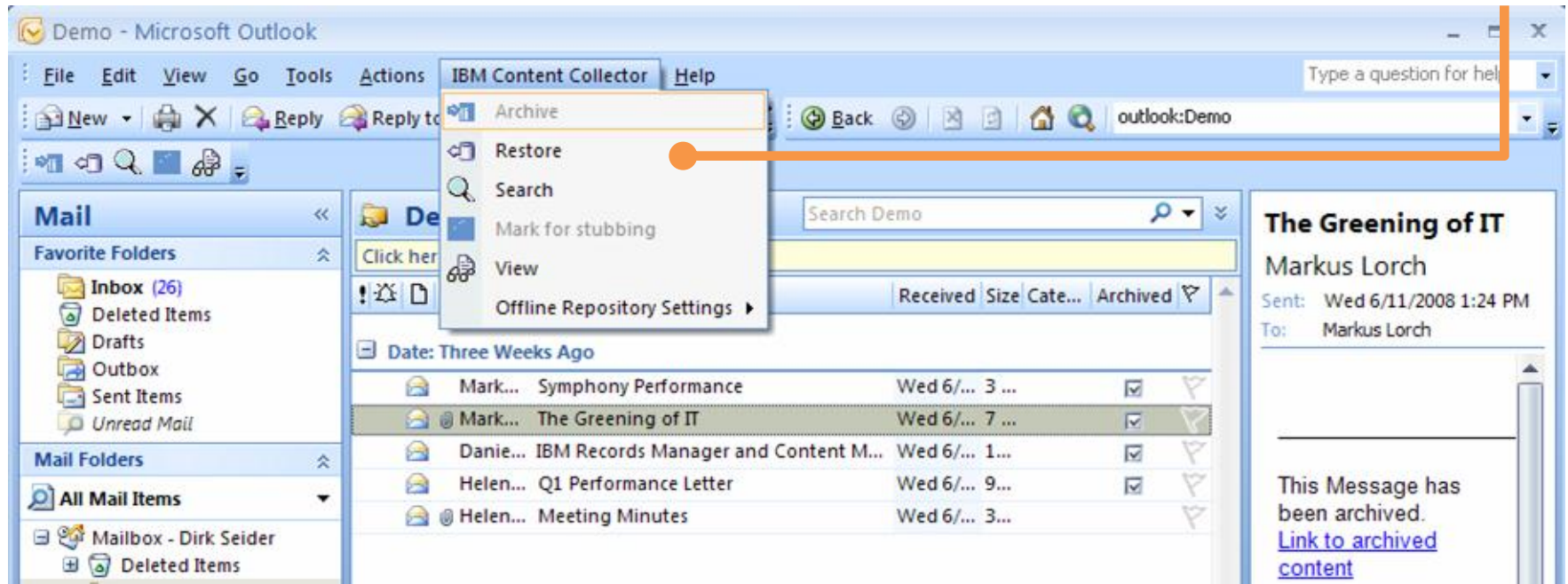
- Supports both email AND public folders
- Supports multiple archiving paradigms
 - Journaling for compliance
 - Mailbox management for storage savings
 - User Driven mailbox archiving for business value
 - Advanced PST processing capabilities to help eliminate personal archives
- Supports Offline access
 - Creates a local instance of archived data that can be used while disconnected.
 - Synchronized with official archive per schedule
- Supports “maintenance” of archived content
 - When content is retrieved, it will be re-stubbed (not re-archived) in the next archive run
 - Full clean up of mail client when expired from repository
- Respects “private” tag in delegation scenarios
 - Data marked “private” by owner not returned in search by delegate

Outlook Integration

End-user retrieval of stubbed emails is transparent

Reply / forward / view emails stored in the archive with no additional steps

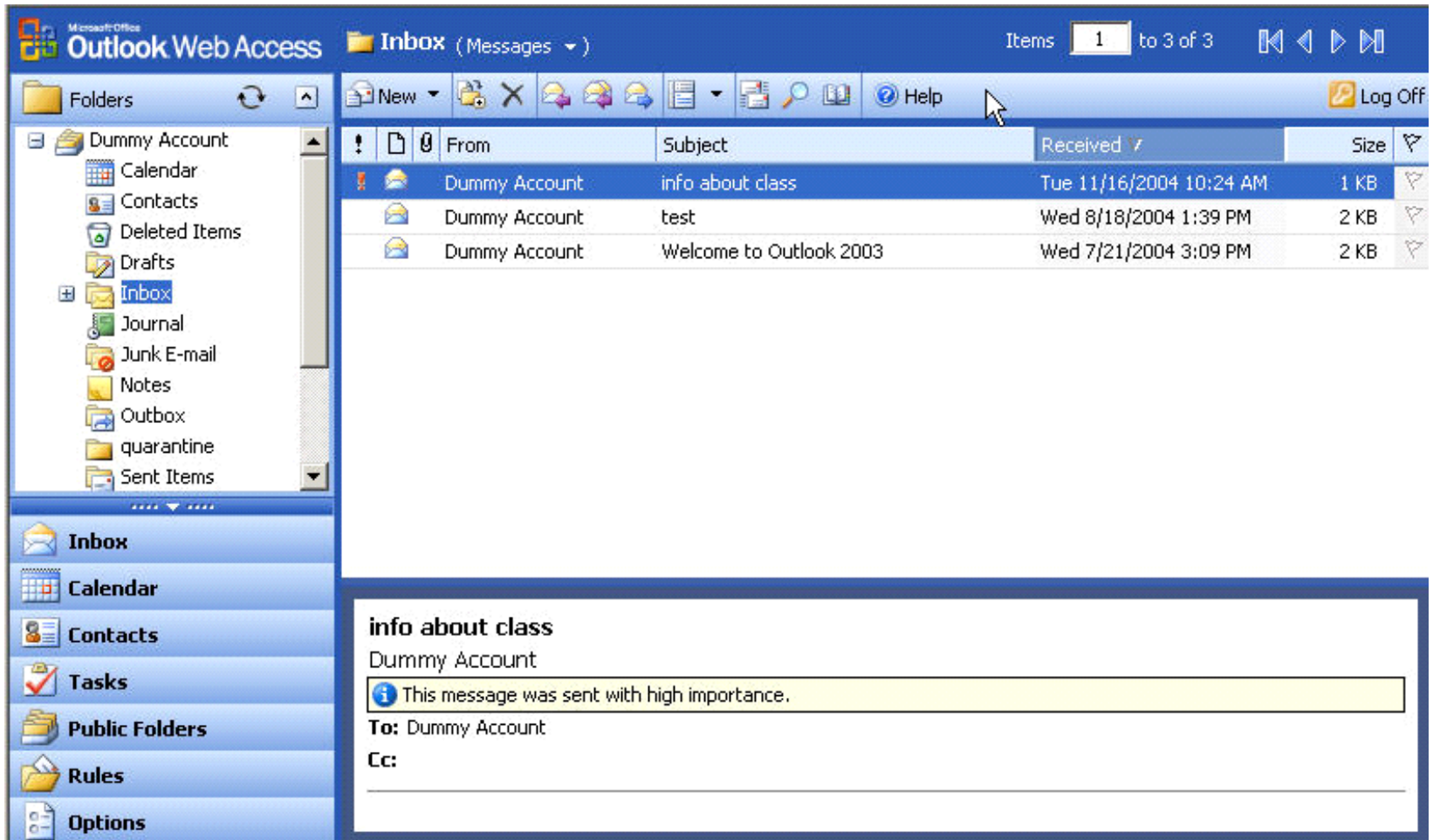
ICC
extensions
deployed to
Outlook



Outlook 2010
look & feel

Microsoft Outlook Web Access

End-users can select documents to be collected, restore documents from the repository, and search collected email



The screenshot displays the Microsoft Outlook Web Access interface. The top navigation bar shows "Outlook Web Access" and "Inbox (Messages)". The left sidebar contains a "Folders" list with "Dummy Account" selected, and a secondary list with "Inbox", "Calendar", "Contacts", "Tasks", "Public Folders", "Rules", and "Options". The main area shows a table of messages:

From	Subject	Received	Size
Dummy Account	info about class	Tue 11/16/2004 10:24 AM	1 KB
Dummy Account	test	Wed 8/18/2004 1:39 PM	2 KB
Dummy Account	Welcome to Outlook 2003	Wed 7/21/2004 3:09 PM	2 KB

The selected message "info about class" is displayed in the bottom pane. It includes a warning icon and the text: "This message was sent with high importance." The sender is listed as "To: Dummy Account" and "Cc:" is empty.

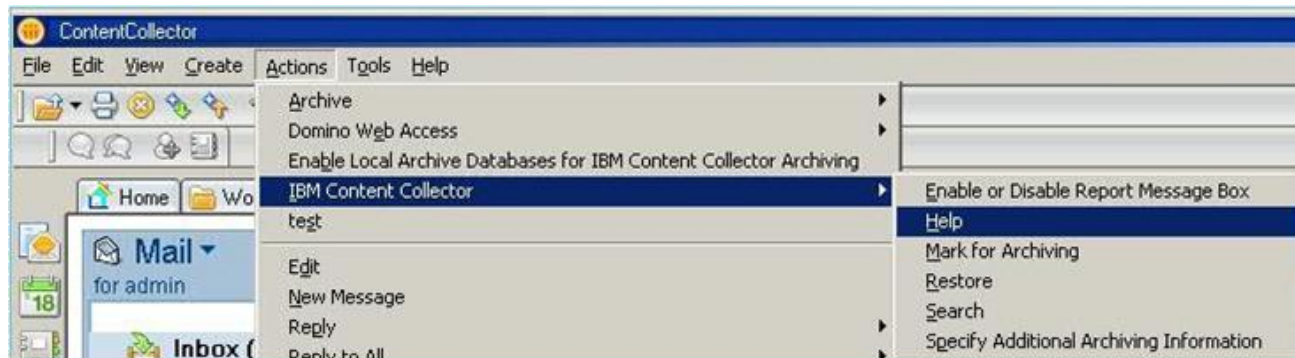
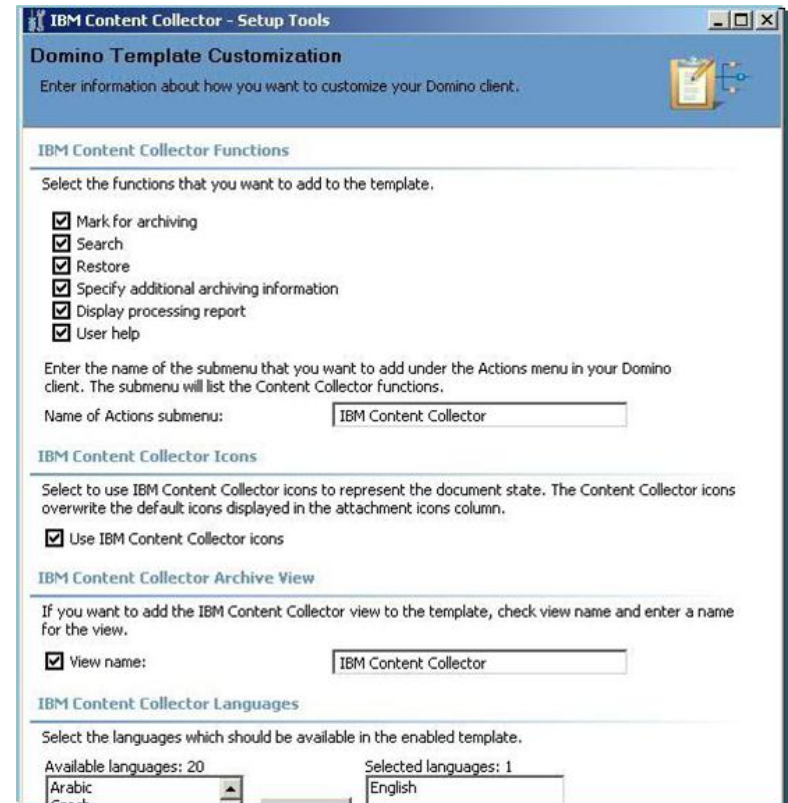
IBM Content Collector for email: Lotus Notes Archiving

IBM Content Collector collects, processes and archives content from Domino servers.

- Supports both email AND applications
- Supports multiple archiving paradigms
 - Journaling for compliance
 - Mailbox management for storage savings
 - User driven mailbox archiving for business value
 - Advanced NSF processing capabilities to help eliminate personal archives
- Offline access
- “Round tripping” of content
 - Archived content; retrieved; re-stubbed (not re-archived)
 - Clean up in mail client when expired from repository

Integration with Lotus Notes

- Client integrations with Notes and iNotes
- Transparent retrieval of content from archive “on demand”
- ICC Setup Wizard applies required elements to meet the archiving use case
 - If enabled, provides user with archive and search functions directly from the email client
 - Search, Restore, Folder integration, Add Metadata...
 - Appearance
 - Languages



Content Collector Search Results

Location of Message

Single search across Notes mail file and archived messages.

The screenshot shows the IBM Content Collector Search View interface. The search term 'Smarter' has been entered, resulting in 37 hits. The results are displayed in a table with columns for From, To, Subject, Date, and Location. An orange arrow points from the 'Location' column header to the 'Mail Database' entry in the first row.

From	To	Subject	Date	Location
Smarter Commerce	WinWire-Group2	WinWire: Saint-Gobain	03/12/2012 05:56:45	Mail Database
Smarter Commerce	WinWire-Group2	WinWire: Saint-Gobain	03/12/2012 04:30:29	Mail Database
Smarter Commerce	WinWire-Group2	WinWire: Saint-Gobain	03/09/2012 10:04:09	Mail Database
Steve Cowley - Vice President Industry and Solution Sales	WinWire-Group2	WinWire: Saint-Gobain	11/04/2011 03:59:32	Repository
Zhu Hao	SC26	GCG Smarter City mailcast	03/12/2012 04:30:29	Mail Database
Zhu Hao	SC26	GCG Smarter City mailcast	03/08/2012 06:25:00	Mail Database
Shuang S Zhang	R&D Group D	B2B&Smarter Commerce NingBo Recruiting Promotion	03/12/2012 04:30:29	Mail Database
CDL HR	R&D Group D	B2B&Smarter Commerce NingBo Recruiting Promotion	07/15/2011 02:19:07	Repository
IBM Newswire	Industry Solutions Market...	CMO August Newsletter	03/12/2012 04:30:29	Mail Database
IBM Newswire	Industry Solutions Market...	CMO August Newsletter	03/08/2012 06:57:00	Mail Database
VP Industry Solutions Marketing	Industry Solutions Market...	CMO August Newsletter	08/12/2011 03:53:07	Repository
IBM Newswire	IBM Newswire	Industry Products and Frameworks Newsletter - Technical Solution Design Assets	03/12/2012 04:30:29	Mail Database
IBM Newswire	IBM Newswire	Industry Products and Frameworks Newsletter - Technical Solution Design Assets	03/08/2012 03:55:17	Database/Repository
GMU CTO	GMU1 - 111, GMU1 - 112, G...	GMU TECHNICAL COMMUNITY - Hero profiles, new DE video, blogs & more	03/12/2012 04:30:29	Mail Database
GMU CTO	GMU1 - 111, GMU1 - 112, G...	GMU TECHNICAL COMMUNITY - Hero profiles, new DE video, blogs & more	08/06/2011 11:45:00	Repository
Aurelie Pairon	European Conferences	IBM Exceptional Web Experience Conference 2011 Session Directory available!	03/12/2012 04:30:29	Mail Database

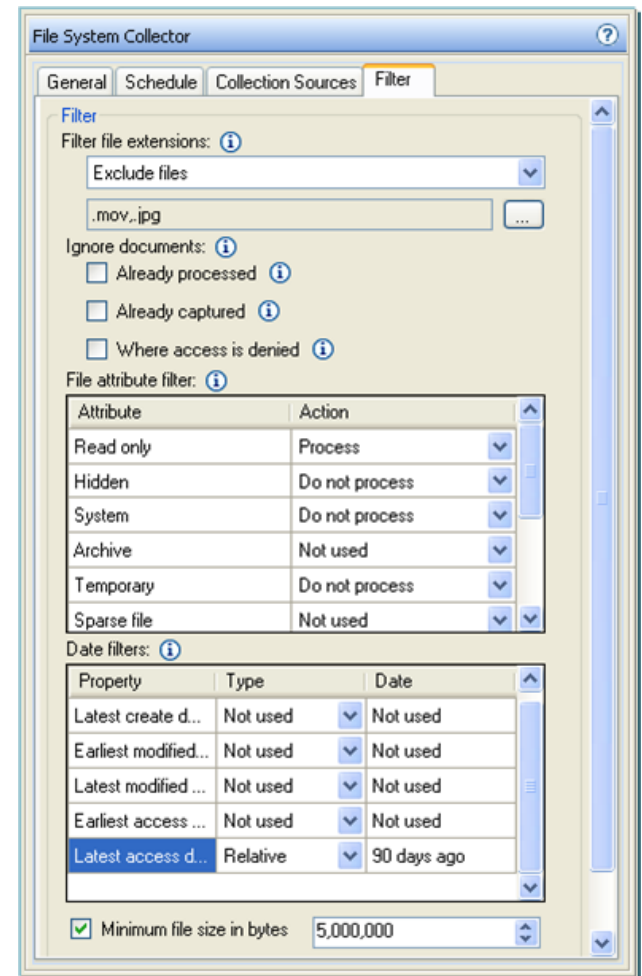
IBM Content Collector for File Systems

ICC for File Systems provides solutions for several use cases:

- Application integration tasks
 - Output from an application generates multiple files & associated control (metadata files) collected, processed and stored in ECM repository
 - BPM...
- File server consolidation projects
 - Harvest, de-duplicate & store
- Supports Microsoft NTFS and other file systems that can emulate NTFS
 - Can carry over NTFS Access control to the archived content
- Include/Exclude files based on extensions, access control, file attributes, dates, file sizes (Max/Min)
- De-duplicates across multiple sources
- Archive repositories: FileNet Content Manager, IBM Content Manager, FileNet Image Services, File System

IBM Content Collector for File Systems

- Configurable post-processing behavior
 - Leave original
 - Replace file with link to archived content
 - Leaves icon in place
 - Delete original entirely
- Supports re-collection, stub cleanup upon deletion from Archive
- Supports versioning of archived documents
 - Updates determined by a combination of modified date, file size and uniqueness hash.
- Supports mimicking file folder structure in the repository
- Supports maintaining file system Access Control settings in the repository



IBM Content Collector for IBM Connections

New in IBM Content Collector v3.0

- Collect from files, blogs, wikis, activities, forums, bookmarks, profiles
- Supports collecting from one or more IBM Connections deployments
- Focused on eDiscovery use-cases
 - IBM provides a single eDiscovery search capability across Connections and any other data types supported by ICC



Configuring ICC for IBM Connections

The screenshot displays the IBM Content Collector V3.0 interface. The main window shows a task route diagram with the following steps: CX Collector, CX Pre-processing, P8 Create Version Series, and P8 File Document in Folder. An 'Edit Collection Source' dialog box is open, showing the 'Collection Sources' tab. The 'Site URL' is set to 'http://vdc-ven-cx301.ytgdev.com'. The 'Site Connection' is set to 'CX Connection'. The 'Applications' section includes checkboxes for Files, Wikis, Blogs, Activities, Profiles, Forums, and Bookmarks, all of which are checked. The 'Users' section includes an 'Add User' field and a 'Selected Users' list.

Original Content in IBM Connections

How to install Lotus Connection - Mozilla Firefox

File Edit View History Bookmarks Tools Help

How to install Lotus Connection

vdcqa.com https://vdc-vqa-cx301a.vdcqa.com/forums/html/topic?id=0667945d-281f-4196-bf34-ce149bdf7461

Most Visited IBM eDiscovery Mana... IBM Web Interface fo... Log In to IBM Connec... Team80 - Home

IBM Connections Home Profiles Communities Apps Moderation seedlist Settings Help Log Out

My Forums Public Forums This Forum Search

LotusConnect Project Forum Follow

Topic Tags Find a Tag howto lotusconnection qa Cloud List

LotusConnect Project Forum > How to install Lotus Connection

2 replies Last Post - Oct 28 2011 by kerry xue1

kerry xue1 2 Posts **How to install Lotus Connection** Flag as Inappropriate

Oct 28 2011 | Tags: howto, lotusconnection, qa Add or Remove Tags

This question has not been answered yet.

Please referce the doc attached

Attachments

DOC.doc 1009 KB

Add Your Answer Reply Edit Delete Move Lock Topic Unpin this Topic

Kerry Xue 1 Posts **Re: How to install Lotus Connection** Flag as Inappropriate

Oct 28 2011 in response to kerry xue1

I have some issue when install the LC, a, b, c,

Accept this Answer Reply Edit Delete

kerry xue1 2 Posts **Re: How to install Lotus Connection** Flag as Inappropriate

Oct 28 2011 in response to Kerry Xue

Please do A. B.

Accept this Answer Reply Edit Delete

Reply to this Topic Start a Topic Feed for this topic

Content retrieved with eDiscovery Manager

The screenshot displays the IBM eDiscovery Manager web interface within a Windows Internet Explorer browser. The browser's address bar shows the URL `http://vdc-vqa-s11:9080/EDMClient/Main.do`. The page title is "IBM eDiscovery Manager".

At the top right, the user is logged in as "administrator". Below this, the current case is identified as "test-cx" with a link to "Change case". The case phase is "Active". A task status bar indicates: "Error(0) | Inactive(0) | Processing(0) | Complete(0)".

The main interface is divided into several sections:

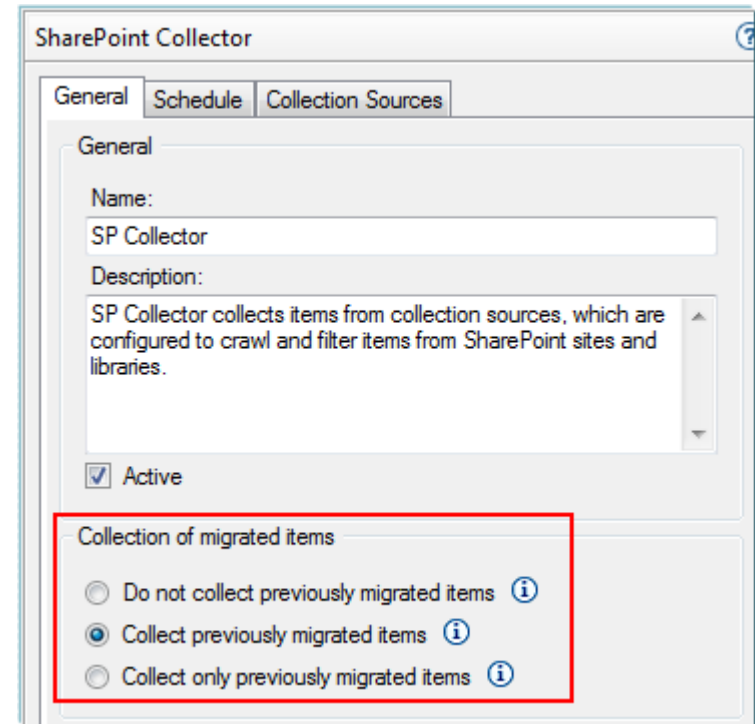
- Search Panel (Left):** Contains search filters and options. The search template is "Search-CX". There are radio buttons for "Archive", "Case", and "Folders", with "Case" selected. A checkbox "Show only results not already in the case" is present. Input fields for "Author", "Contributor", "Doc title" (containing "How to install Lotus Connection"), and "Title" are visible. A "Content or attachment" field is also present. Buttons for "Search", "Clear", and "Schedule Search" are at the bottom.
- Results List (Top Center):** Shows "1 Total" results. A table lists the search results with columns for "Type", "Native view Doc title", and "External id". One result is shown: "How to install Lotus Connection" with external ID "{DE8CB674-479C-4EB4-A37A-C8CE...}".
- Document Preview (Center):** Displays the content of the selected document. The title is "LotusConnect Project Forum > How to install Lotus Connection". It shows a forum post by "kerry xue1" dated "28/10/2011 16:54 UTC" with the title "How to install Lotus Connection". The post includes a question: "This question has not been answered yet." and an attachment "DOC.doc 1009 KB". Below this, two replies are shown: one by "Kerry Xue" and another by "kerry xue1".

The status bar at the bottom left shows the time "1/9/12 10:23 AM" and a message: "[INFO] eDiscovery Manager successfully connected to the content server." The bottom right of the browser window shows "Trusted sites | Protected Mode: Off" and "100%" zoom.

IBM Content Collector for Microsoft SharePoint

Identify and collect content from Microsoft SharePoint for compliance, storage management and long term retention of valuable content

- Supports all list and library types
 - Enterprise wiki page library, Asset Library, Record Library...
 - Announcement, Contacts, Project Tasks, Discussion Boards, Surveys....
- Define collection at multiple levels & determine depth
 - Farm, Web Application, Site Collection, Sub-site
- Configuration options available for archiving and re-collection of restored items
- Cleanup links in SharePoint when content deleted from the repository



“Look & Feel” Maintained on Archived Content

- Non-binary content stored as HTML renderings
- Displayed in “high fidelity”

Original “Discussion” in SharePoint

As Archived by ICC; Rendered in browser

Started: 2012-04-19T00:07Z by System Account (SHAREPOINT\system)

Post photos of your systems/rooms here! (Part 9)

Share photos of your systems/rooms in the discussion thread.

Posted: 2012-04-19T00:09Z by System Account (SHAREPOINT\system)

Some serious rigs going on in this thread. I'd thought I'd share my work in progress.



From: System Account
Posted: 4/18/2012 12:07 PM
Subject: Post photos of your systems/rooms here! (Part 9)

Share photos of your systems/rooms in the discussion thread.

Posted: 2012-04-19T00:10Z by System Account (SHAREPOINT\system)

Very nice. Clean. Refined.

Threading

Started: 4/18/2012 12:07 PM by System Account

Post photos of your systems/rooms here! (Part 9)

Share photos of your systems/rooms in the discussion thread.

Posted: 4/18/2012 12:09 PM by System Account

Some serious rigs going on in this thread. I'd thought I'd share my work in progress.



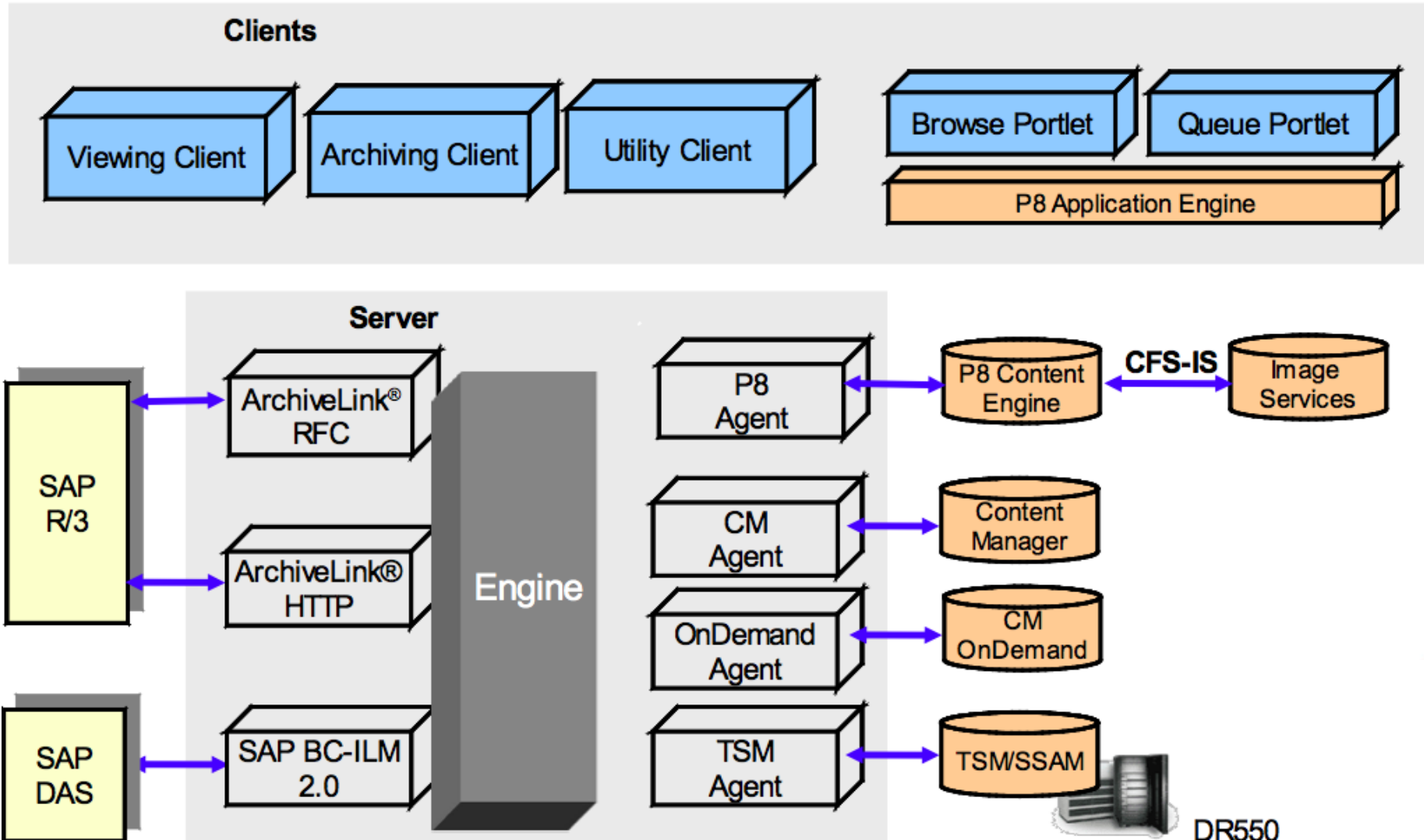
⌵ Show Quoted Messages

Posted: 4/18/2012 12:10 PM by System Account

Very nice. Clean. Refined.

⌵ Show Quoted Messages

IBM Content Collector for SAP



IBM Content Collector for SAP

The screenshot displays a SAP HR system interface. The main window shows the profile of Dr. Karin Höllerer with the following details:

- Personnel No.: 888888
- Name: Dr. Karin Höllerer
- Employee Group: 1 (Active)
- Employee Subgroup: DT (Non-PS staff)

A document viewer is open, displaying a document titled 'PERSONALINFORMATIONEN' with the following content:

xft
PERSONALINFORMATIONEN

Datum: 11.08.2011
Mitarbeiter: Karin Höllerer
Personalnummer: 888888

Für die korrekte Gehaltsberechnung bitten wir Sie um Übermittlung bzw. Mitteilung folgender Daten:

- Anschlag und Tarifplan
- Aktuelle Lohnsteuerkarte (falls noch nicht vorhanden, bitte Angabe von Steuerklasse, Anzahl Kinder, Freibetrag)
- Bankverbindung
- Kirchenzugehörigkeit und Kirchenversteherungs-Nr. des Verzeichnisses

The interface also shows a navigation pane on the left with a tree structure including 'Outlook Connector', 'Hiring Process', 'Personal Documents', and 'Personal information'. The status bar at the bottom indicates the current path is '/XFT/EPA_FIND' on 'server056'.

Thank You!

eDiscovery
Management



Records and Retention
Management



Archive and
Storage



Defensible
Disposition

