CM/ECM Portfolio Review, Architecture, & Directions

Jim Reimer
Distinguished Engineer
Chief Architect, Content Management

IBM Software Group
Content Management basics

- Streamline business processes and communications
- Capture physical media and electronic documents
- Capture computer output instead of printing
- Store, manage, and deliver information digitally rather than physically
Business re-engineering with CM

Deployment Phases

- Business processes built around physical information handling
- Same business processes, with information handled electronically
- Simplified business processes, exploiting electronic information handling

Business impact:
- Increased competitiveness from refined business processes
- Cost savings in reduced physical handling
Content Management usage by media type

- Bilevel documents
- Grey/color documents
- Computer output
- Audio
- Video
- Film

Year:
- 85
- 90
- 95
- 00
- 05
e-business solutions leverage all content

Relevant, comprehensive information

Structured Data

Business Data

Business Content

Unstructured Data

Structured Data

Unstructured Data
e-business content

- Operational Content
  - Scanned images
  - Facsimiles
  - Computer generated output

- Workgroup Documents
  - Word processors
  - Spreadsheets
  - Presentation graphics

- Web Content
  - HTML
  - Dynamic content

- Media Assets
  - audio
  - video
e-business content usage

Industry Solutions

- Banking and Finance
- Insurance
- Retail
- Government
- Telecom

SCM
- Purchase Orders
- Invoices
- Proof of Delivery Forms

ERP
- Employee Information
- Invoices
- Job Applications

E-COMMERCE
- Product catalogs
- Transaction records
- Pictures, music, video, books...

CRM
- Self-service information
- Bills, statements
- e-mails, letters, fax
- Loan applications
- Phone conversations

Business Integration (inter- and intra-enterprise)
Elements of e-business Content Management

**Digitized content**

- Facsimiles
- PC documents
- Incoming paper documents
- Web content
- Audio & Video
- Computer generated output

**Applications**

- Customer Service
- e-commerce
- Supply Chain
- ERP

**Application Development**

- APIs
- Components
- Workflow

**Workflow**

- Capture
- Index
- Search / Access
- Create
- Rights Mgmt.
- Workflow

**Storage**

- Library server
- Object server
- Device support
- Mgmt. Utilities

**DB2 Data Management Software**
IBM Content Manager structure

or why don't we just do this with a database?

data modelling & doc mgmt
large number of entries
all indices on disk for search
complex data models
metadata and full text
fine grained access controls
CPU intensive

access control

referential integrity

integrated hsm & caching

large total data volume
large size per object
still and streamed objects
only recently used on disk
transparent migration
object level hsm
I/O intensive

Library Server

Object Server

Object Server

Object Server
IBM Content Manager highlights

**e-business application enablement**
- Thin client or workstation configurations
- Popular visual builders / development tool support
- Java Beans, ActiveX Components, Java, C++ APIs
- Folder management

**Advanced Search**
- Parametric, full-text
- Boolean
- Fuzzy, phonetic
- Relevance ranking
- Image Content
- SBCS, DBCS

**Rights Mgmt.**
- Access control
- Watermarking
- Encryption

**Mgmt. Utilities**
- Administrative GUI
- Catalog config. & maint.
- Hierarchical Storage Management
- Tivoli Storage Manager
- Tivoli Ready

**Industry standards**
- XML, HTML
- TIFF, JPEG, MPEG...
- Device driver support
- Popular tool interfaces
- Scanning support

**Scale**
- Cross platform
- Multiple, distributed object servers
- Hard drives to storage servers
- Computer generated output support
- Auxiliary storage to tape and optical

**DB2** Data Management Software

**Library Server**

**Object Server**

**Object Server**
WCM Author time and Run time

Authoring and Collaboration Application

Run time Websites

Webmaster Publish Controls

Author time Repository

Website File System

EIP

CM and other sources
- rich media
- documents

- page content
- JSPs
- servlets
- authortime controls

- page content
- JSPs
- servlets
- rich media
- documents
Content Manager VideoCharger Version 8

Real Time Encoding with Multicast/store

Video on Demand

Multicast

VideoCharger

Stream A

Stream B

Stream C

Play/Rew/FF/Stop

Multicast

Data Management Software
Enterprise Information Portal Version 8

Production Clients

Thin Clients

Java, C++, Beans, JSP

Federated Operations
- Create
- Retrieve/View
- Update
- Delete

Workflow
- Parallel Routing
- Ad Hoc Routing

Information Mining
- Multi-Lingual
- Categorization
- Summarization
- Clustering
- Language ID

DB2 Data Management Software
Content Management Portfolio Highlights

- **Convergence**
  - OnDemand family converged with OD V7 (GA 9/01)
  - ODWEK being converged with EIP
  - Content Management server family converged with CM V8
  - Family direction to converge OnDemand and Content Manager servers

- **Solutions**
  - Enhanced browser, windows, and ODMA clients with CM V8
    - extended function
    - reworked usability and asthetics; improved consistency
    - improved performance
  - Family direction to converge OnDemand and Content Manager clients
  - Enhanced enabling components (Java/Beans and C++/ActiveX)
  - Greatly increased business partner/solution activity
    - Siebel
    - Green Pastures
    - SAP
    - Exchange
    - Lotus/Domino
    - Lotus/Notes
Content Manager V8 Architectural Overview

**Windows Client**
- **CM API**
- **DB2 CAE**

**Library Server**
- **SQL** (Direct call if local)
- **DB2**
  - Stored Procedures
  - DB2 Extender
  - Security, RI

**Application Server**
- **CM API**
- **DB2 CAE**

**Thin Client**
- **HTTP, FTP, Files**
- **Listener**
- **Security, RI**
- **Filters**
- **DB2, HSM, Obj’s**

**Resource Manager**

---

**DB2** Data Management Software
V8 Key Objectives

- Enhanced browser, windows, and ODMA client function
  - expose all the function of the servers to the client
- Application development toolkit for popular OO languages
  - C++, Java, Beans
- Improved performance and scalability
  - push function down the system stack
  - improved use a parallelism
- Improved architecture for application topology flexibility
  - 2 or 3 tier models
- Improved object delivery options
  - HTTP, FTP; file system
- New function for additional applications
  - data modelling extensions, enhanced versioning, advanced workflow
- Improved exploitation of companion middleware
  - Websphere, MQSWF, Tivoli Storage Manager, OAM
- Expanded platform coverage
  - NT, AIX, MVS, Sun
New Features and Functions

- Improved data modelling services (XML Repository)
  - Hierarchical documents (objects)

- Improved Flexibility in Item relationships - Linking, References
  - Representation of Folders, Compound Documents, Shared Parts/Objects, Maps, Graphs

- Open coupling to Resource Managers
  - Indexing of and access to data owned by other applications
  - VideoCharger

- Improved Retention Management
  - Each ItemType can have a retention period
  - Expiration date calculated when Item is created
    - Number of days, months, years
  - Can be updated by applications
  - Removal performed by custom application or client
  - Expect to provide automatic removal in the future

- Improved Event log / Audit trail
  - Who did what and when
  - Recorded in 2 tables (one for Items, one for Definitions)
  - Level controlled by SysAdmin
New Features and Functions

- **Versioning** - new versioning services for metadata
- **Access Control**
  - Enabled at the Item and Part level
- **Simplification**
  - Less than 50 OO Classes, including SysAdmin
  - Enabled by exploiting data model changes
- **API Performance**
  - Multiple Items per call -- Create, Delete, Checkout/In, etc.
  - Validation at the API where possible
  - Single Library Server round trip per call
- **Tracing and Error Logging**
  - Entry/Exit, detailed trace, data sent/received
  - All errors (including SQL message) logged
  - CM errors will have at least "name" logged
- **Improved Transaction Management**
- **Administrative Domains**
New Features and Functions

- **Workflow**
  - Serial Work Baskets
  - Advanced document routing via MQSeries Workflow
  - Simplified graphical builder

- **Auto-Foldering**
  - When Item is indexed, automatically add it to a folder
    - i.e., Claims added to Customer folder
  - Implemented on the server, not the client

- **ItemID - based on timestamp so continuously increasing**
  - Reduces fragmentation/reorg of ITEMS table

- **Segmentation**
  - At some size (i.e., 50 million records), automatically create a new set of tables
  - ItemID will contain Segment number
  - Architected, but not implemented in first release

- **VideoCharger**
  - MPEG 4 support
  - Filter architecture
Content Manager Version 8.1

- Platform coverage CM and EIP with V8.1
  - NT, AIX, Sun with first release
  - Operating with DB2 and Websphere
  - Coming release to include
    - Oracle support for NT, AIX, Sun
    - Support for DB2 V8.1
    - MVS support for Library Server and Object Server
  - MVS support for EIP planned for release in 2003

- Lan Cache and Replication
  - Will be released with first V8.1 fixpack, before y/e 2002
CM Access Control Model

- Efficient run-time checking
- Support of multiple protection models

**Access Control List**
- Users
- Privileges
- (Conditions)

**User General Privileges**

**Special Privileges**
- Administrator
- Owner
CM V8 Windows Client
CM V8 Browser Client
ODMA Client

- Support for additional server function
  - Versioning (1.1.x)
  - Child components (single level)
  - Multi-value attributes

- Autofoldering (server function)

- Full text search and index capabilities
VideoCharger - New Features

MPEG 4 Server Support
- Transparent "Click and Play" client - no installation required
- Standards-based support - audio, video, image, rich media
- Provides support for highly interactive environment

Server Filter Architecture
- IBM Filter technology 'future proofs' server architecture
- Pluggable support for new codecs, new features without having to replace server or impact existing installation
- Allows customized support for enhancements like encryption, live capture, watermarking, proprietary codecs and formats

Basic Server Enhancements
- Automatic calibration
- Warm boot support
- Enhanced server to server transfers
- Improved performance, scaling, and reliability
System Admin Preview
CM and EIP - One Admin Client
CM V8 Sys Admin: Text-search Attributes
### Item Type Properties - Doc

<table>
<thead>
<tr>
<th>Auto-linking Definition</th>
<th>Foreign Keys Access Control</th>
<th>Locking Attributes</th>
<th>Document Management Default Storage</th>
</tr>
</thead>
</table>

**Name:** Doc

**Display name:** Document (for Document data model)

**Language:** US_ENGLISH

**New version policy:**
- Always create
- Prompt to create
- Never create

**Maximum total versions:**
- Unlimited
- Limited to 1

**Item type classification:** Document

**Media Object (XDO) Class:** DBImageICM

**Text Search:**
- Text Searchable

**Item retention period:**
- Forever
- For [ ] Years

[Image of CM V8 Sys Admin: Text-search Documents]

**DB2 Data Management Software**

[IBM Logo]
Advanced Workflow

- Begin
- Work Basket
- Collection Point
- Decision Point
- Sub Process
- Join
- End
EIP Workflow Builder
Imbedded into CM Administrator
Document Routing Definitions imbeded into CM Administration
Defining a Process

### Process Properties - P1

<table>
<thead>
<tr>
<th>From node</th>
<th>Action</th>
<th>To node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>Continue</td>
<td>WN1</td>
</tr>
<tr>
<td>WN1</td>
<td>Continue</td>
<td>CP1</td>
</tr>
<tr>
<td>CP1</td>
<td>Continue</td>
<td>End</td>
</tr>
</tbody>
</table>

Name: * P1
Description: 

Add Row | Remove  | OK      | Cancel | Apply | Help
Defining a Node

![WorkNode Properties - CP1]

- **Definition**
  - **Work Node**: Name: *CP1*
  - **Collection Point**
  - **Description**:
  - **Access control list**: *
Step 2 of 6: Test communication and verify migration authorization

This wizard must connect to your earlier Content Manager Library server and to DB2 Universal Database with super user access privileges before migration can proceed. Enter the library server name and your DB2 user name and password. Click the **Verify** button. Once the status fields indicate both a successful connection and authorized access, click the **Next** button to continue.

- **Content Manager Version 6.1 or 7.1**
  - **Library server name**: ABCSRVRN

- **DB2 user name**: migrate
- **DB2 password**: ********

**Communication status**: Communication successful

**Authorization status**: Authorized
## EIP Text Mining Languages supported

<table>
<thead>
<tr>
<th>Language</th>
<th>Language Identification</th>
<th>Summarization</th>
<th>Categorization</th>
<th>Information Extraction</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Japanese</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>French</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simp. Chinese</td>
<td>x *)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trad. Chinese</td>
<td>x *)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norwegian - Bokmal</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norwegian - Nynorsk</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finnish</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*) language identification does not distinguish simplified and traditional Chinese
EIP Extended Searching

- Lotus Databases/Products
- Document Mgmt Systems
- Microsoft Indexes
- Mail Systems
- Directories
- Relational Databases
- File System Content
- Commercial Content
- World Wide Web

* An information access toolkit is provided that enables you to develop your own custom ES link to a data source.
What Can Be Searched (Cont)

- Notes / Domino databases (.nsf vers. 4.5 and higher).
- Domino.Doc
- Domino R5 Domain Index
- File System content not in compressed format
- IBM EIP connected data sources
- IBM Secureway
- IBM DB2
- Oracle
- Sybase
- MS SQL Server
- MS Exchange
- MS Index Server
- MS Site Server

- LDAP directories
- Intranet Web Sites using a third party web access toolkit
- ES Link Toolkit for additional sources
CM z-series Family Highlights

- Ongoing support for ImagePlus/390 and IAFC
  - IWPM support extended to XP

- OnDemand family converged 9/01
  - OD V7.1 and ODWEK on zOS
  - OD V2 index and object data supported without migration, via compatibility feature
  - Intent to provide tighter integration of OD V2 legacy data with future OD release
    - migration utility for index data
    - access support for current OD V2 object data without data migration

- CM V8 on zOS in forthcoming release
  - Library and object server/resource manager on zOS
  - Existing CM index data migrated to CM V8 library server
  - Existing CM object data assessed without data migration
  - Supported through CM V8 browser and Windows clients
  - Intent to provide EIP on zOS
    - focus on providing support for access to CM family on zOS
    - this release to permit two tiered zOS to browser topology
CM/zOS V8 Specifics

- Common code base with other platforms
- Unique exploitation of zOS companion products - OAM
- Conforms with platform
  - Install via SMP/E
  - Monitor via SMF, RMF, SDSF, or 3rd party products
  - Manage via WLM
  - Security via SAF (RACF)
- Text Search functions not available until DB2 V8
- LDAP exit sample not provided on 390
- LUM not supported on 390
- Resource Manager
  - 390 Web Server (vs WAS)
  - No object migration from 390 to other resource manager
  - 50 meg maximum object size vs >2 gig (current OAM restriction)

- Prereqs
  - OS/390 V2R8 or higher, or zOS V1.1 or higher
    - Including USS Tools, OAM, IBM HTTP Server
    - SecureWay Communication Server - Security Level 1
  - DB2/390 V7.1 or later
    - Exploiting features such as UNICODE, temp tables, UNION, UNION ALL