



ENTERPRISE INFORMATION WORLD
Creating the Information-Centric Enterprise

Information Management Fundamentals and the Information Infrastructure

Presented by:

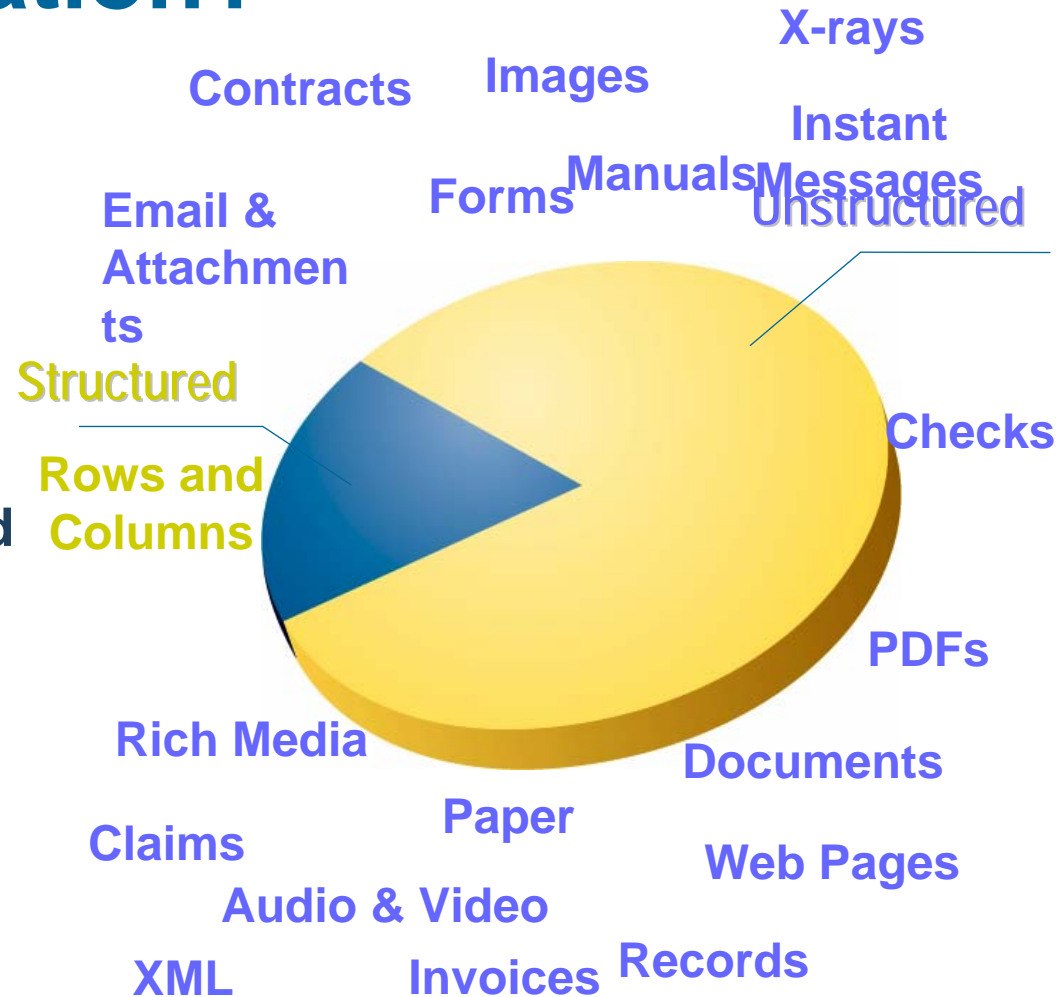
Cornelia Davis

Consulting Technologist, EMC CTO Office

EMC Corporation

What is Information?

- The amount of information is growing every year
- Includes structured, semi-structured, and unstructured data
- Over 80% of enterprise information is unstructured (*Fulcrum Research*)
- Content is mission-critical to many business processes



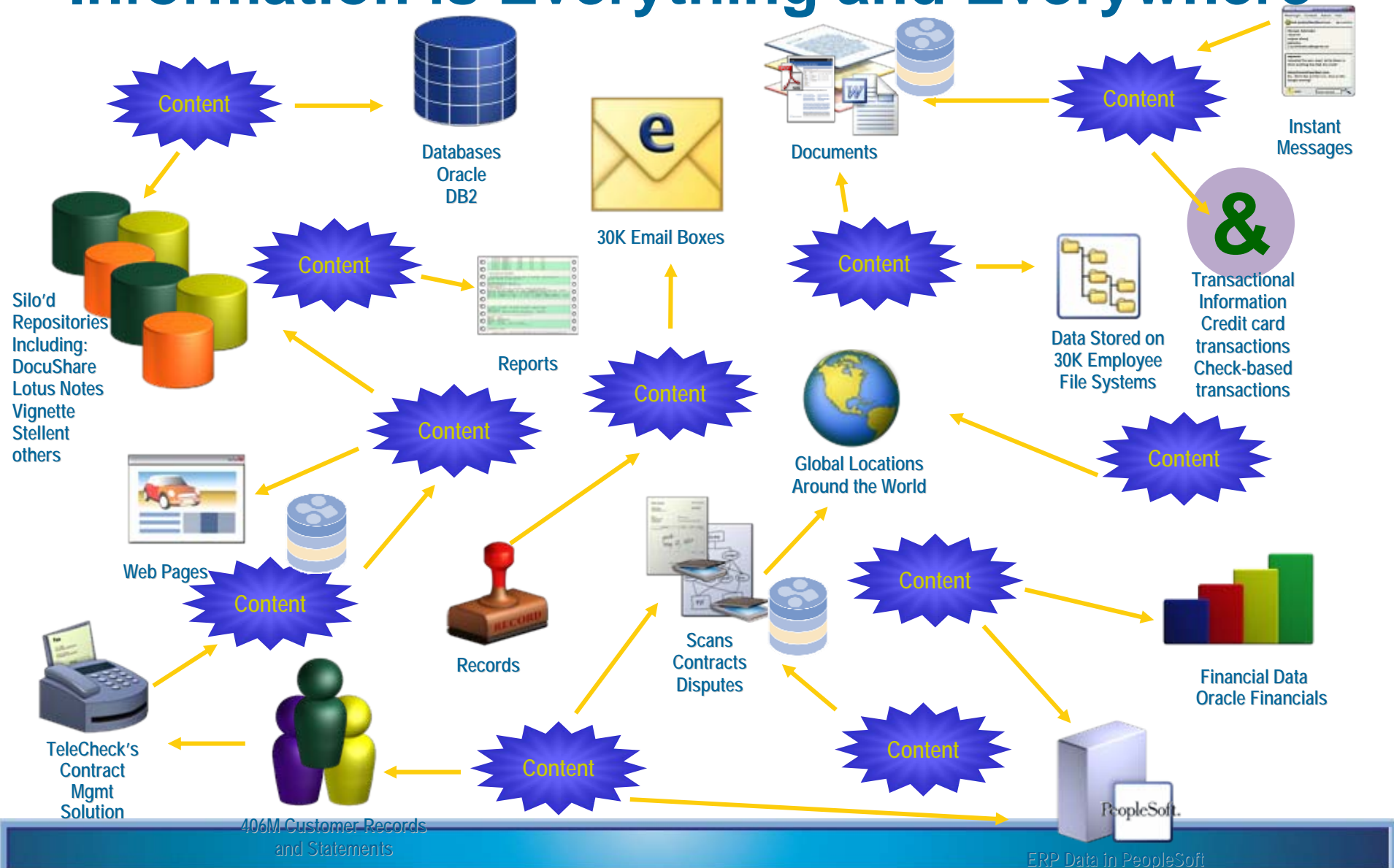
How Much Information is Out There?

*“Between 2006 and 2010, the information **added** annually to the digital universe will increase more than six fold from 161 exabytes to 988 exabytes.”**






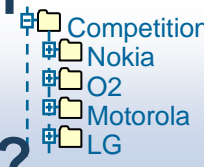


*“In 2007 the amount of information created will surpass, for the first time, the storage capacity available.”**

* According to “The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010” report issued by IDC

Information is Everything and Everywhere



How do we manage all of this?

- ✓ How do we collaboratively create it? 
- ✓ How do we track changes? 
- ✓ How do we secure it? 
- ✓ How do we facilitate approval?  And distribution? 
- ✓ How do we organize information so that we can later find it? 
- ✓ How do we keep track of when we no longer need it?
- ✓ How do we manage long-term archival and access? 
- ✓ And more... 

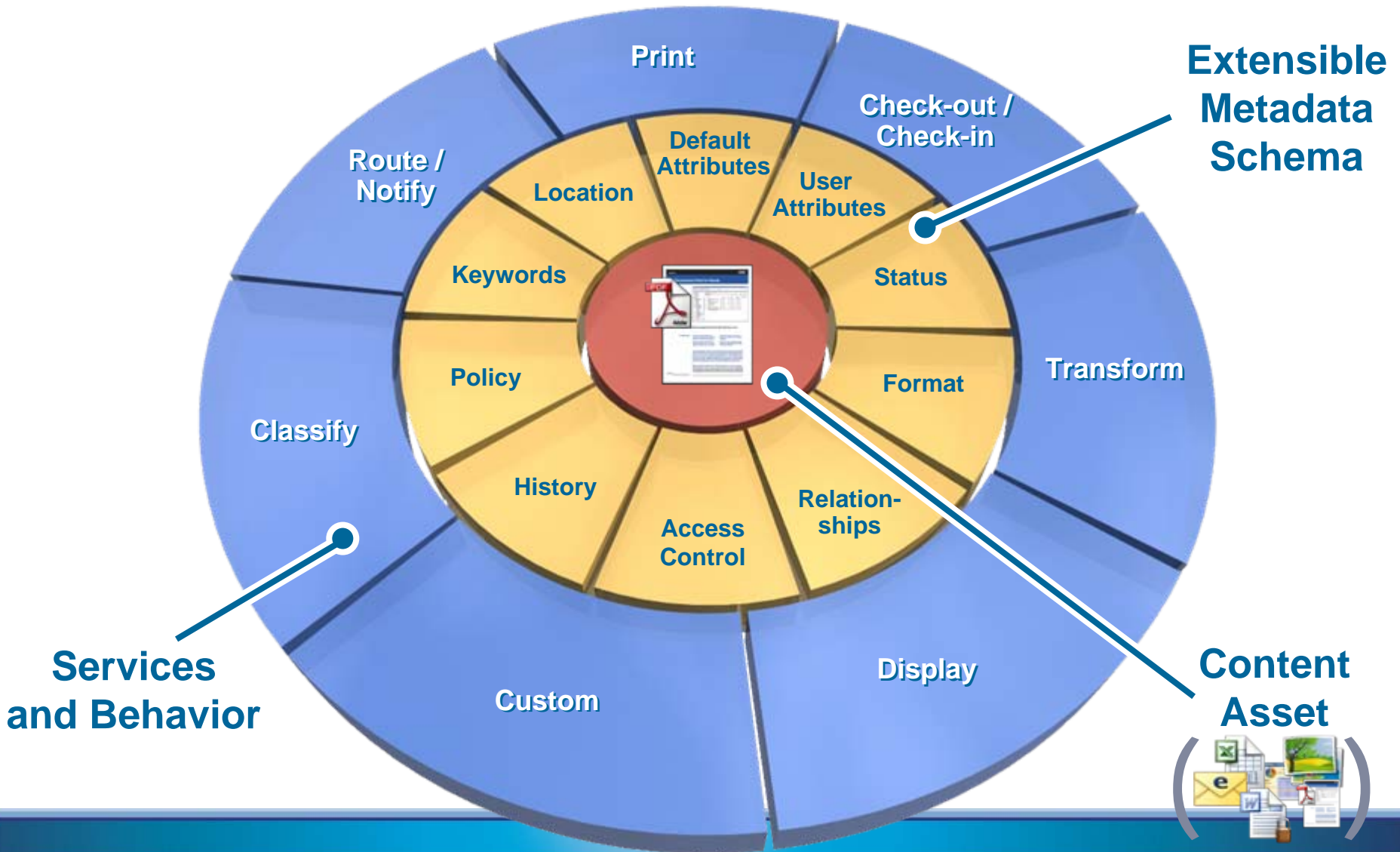
Agenda

- ✓ **Fundamentals of Managing Content**
- ✓ **Documentum Architectural Highlights**
- ✓ **ECM and the Information Infrastructure**
- ✓ **Q&A**

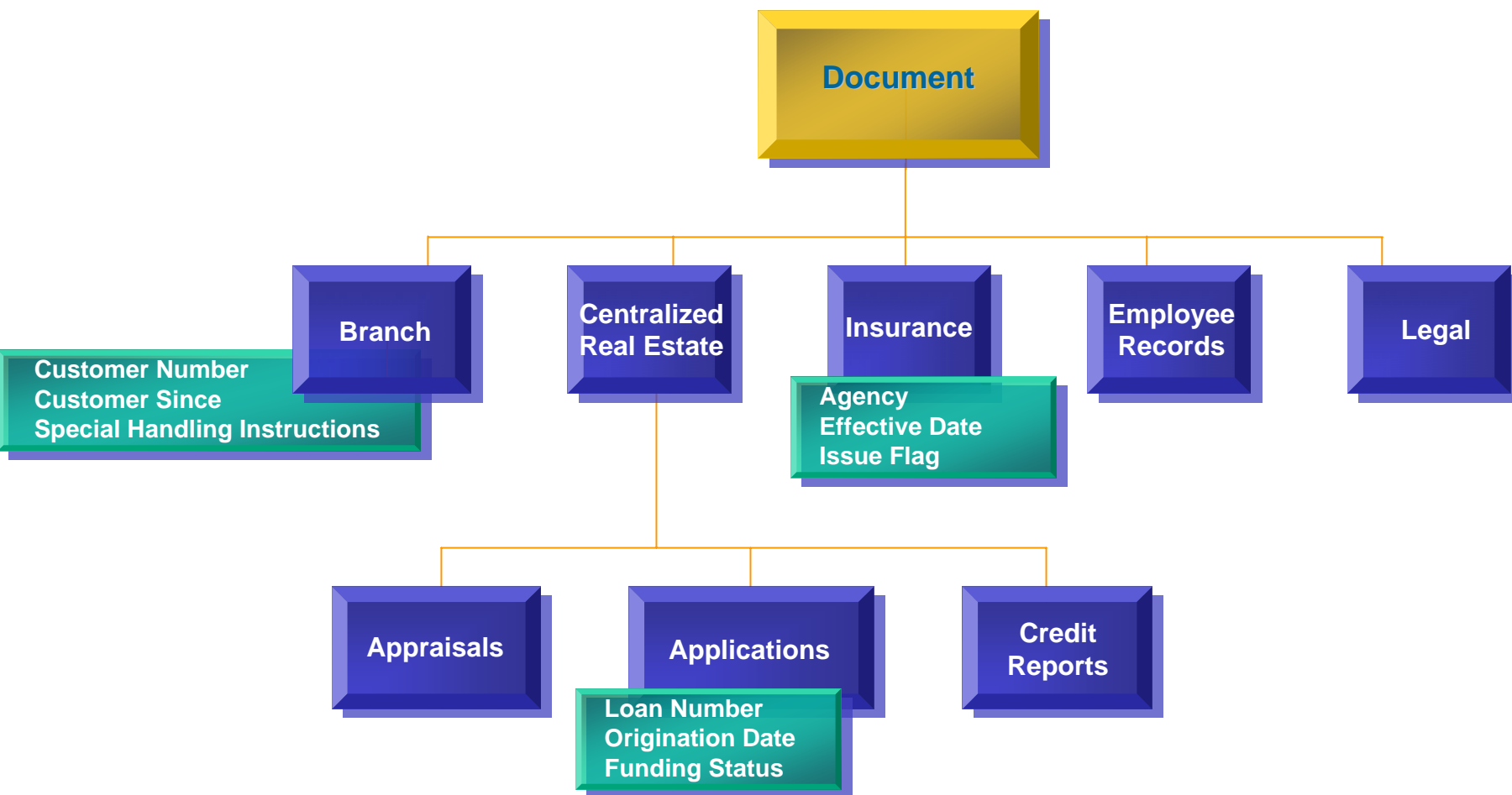
Unstructured Content



Information Object

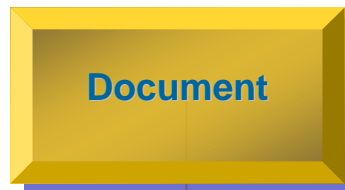


Information Objects are Strongly Typed

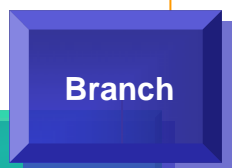


With Polymorphic Behavior

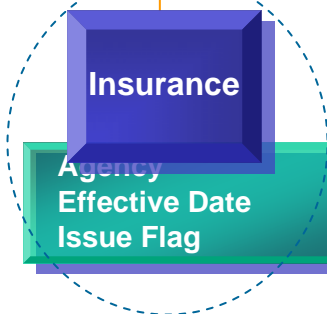
Behaviors are specialized to the content type



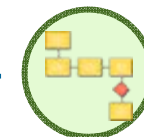
Checking in a document unlocks and versions content



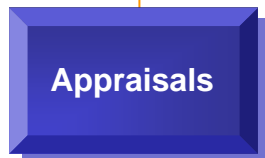
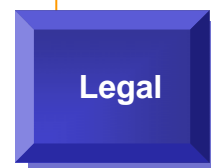
Customer Number
Customer Since
Special Handling Instructions



Agency
Effective Date
Issue Flag



...and kick off BP



Loan Number
Origination Date
Funding Status

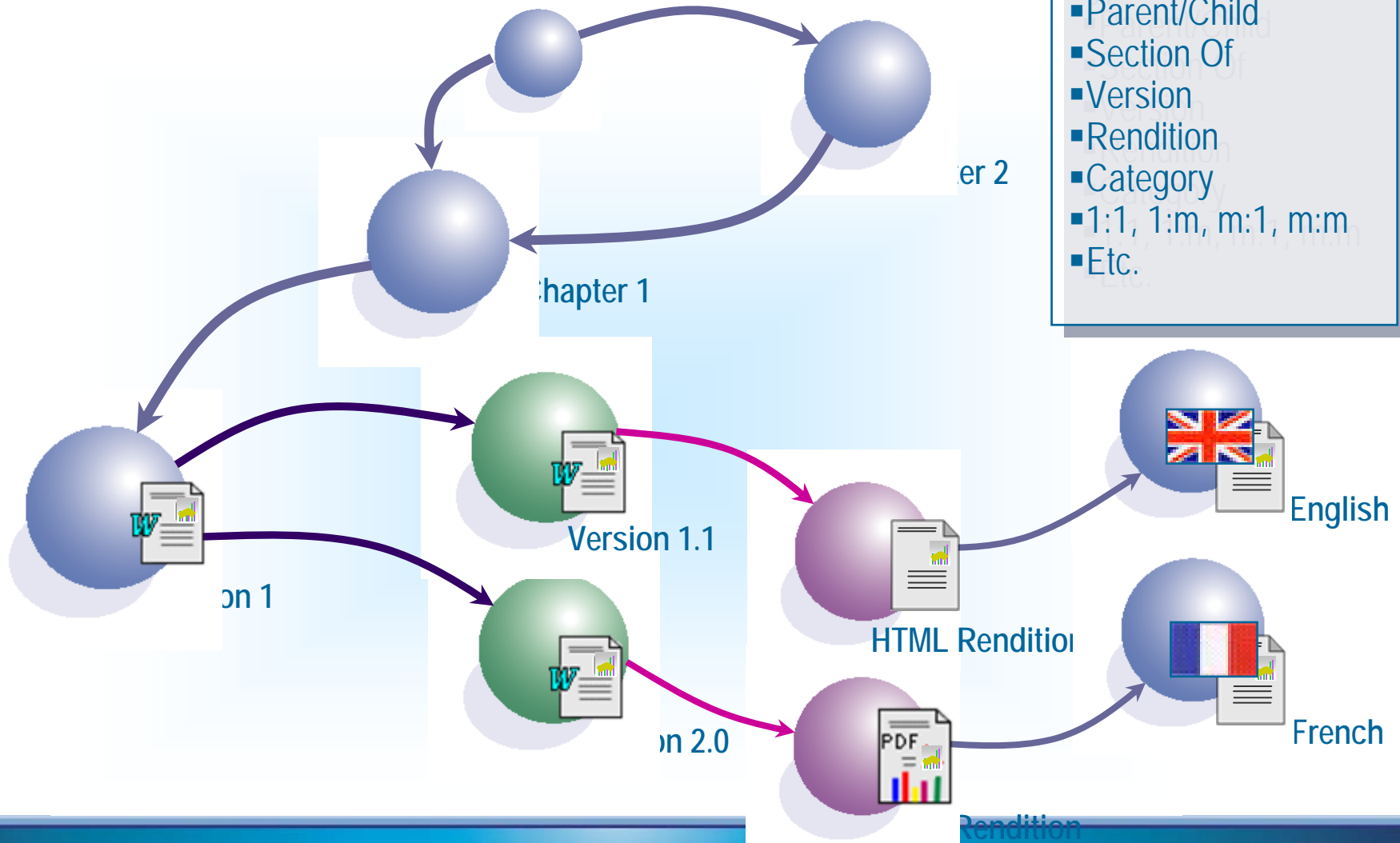


...and create PDF



...and route for approval

Objects Relate to One Another



Relationships are typed:

- Parent/Child
- Section Of
- Version
- Rendition
- Category
- 1:1, 1:m, m:1, m:m
- Etc.

EMC Content Management Platform Strategy

RICH SET OF CONTENT SERVICES

Basic
Content
Services

Business
Process
Management

Content
Capture

Information
Rights
Management

Email
Ingestion

Web
Publishing

Security &
Rights
Management

Enterprise
Reports
Archiving

Document
Image
Management

Automatic
Information
Classification

Compound
Document
Management

Lifecycle
Management

Unified
Desktop & Ent.
Search

Website
Management

Business
Process
Optimization

SAP Archiving

ALL CONTENT TYPES

Documents

Documents
Spreadsheets
Presentations
Notes



Web

HTML
SGML
WML
XML



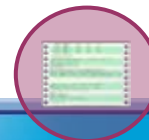
Rich Media

Images
Flash
Sound
Video



Fixed

Paper
Fax Images
Reports
Scanned Images



Messaging

Email Messages
IM Messages
Chats



Collaborative

Discussions
IM
Notes
Projects



MULTIPLE CLIENT ACCESS POINTS



RICH SET OF CONTENT SERVICES

Basic Content Services	Business Process Management	Content Capture	Information Rights Management	Email Ingestion	Web Publishing	Security & Rights Management	Enterprise Reports Archiving
Document Image Management	Automatic Information Classification	Compound Document Management	Lifecycle Management	Unified Desktop & Ent. Search	Website Management	Business Process Optimization	SAP Archiving

ALL CONTENT TYPES

Documents

Documents
Spreadsheets
Presentations
Notes



Web

HTML
SGML
WML
XML



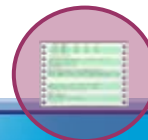
Rich Media

Images
Flash
Sound
Video



Fixed

Paper
Fax Images
Reports
Scanned Images



Messaging

Email Messages
IM Messages
Chats



Collaborative

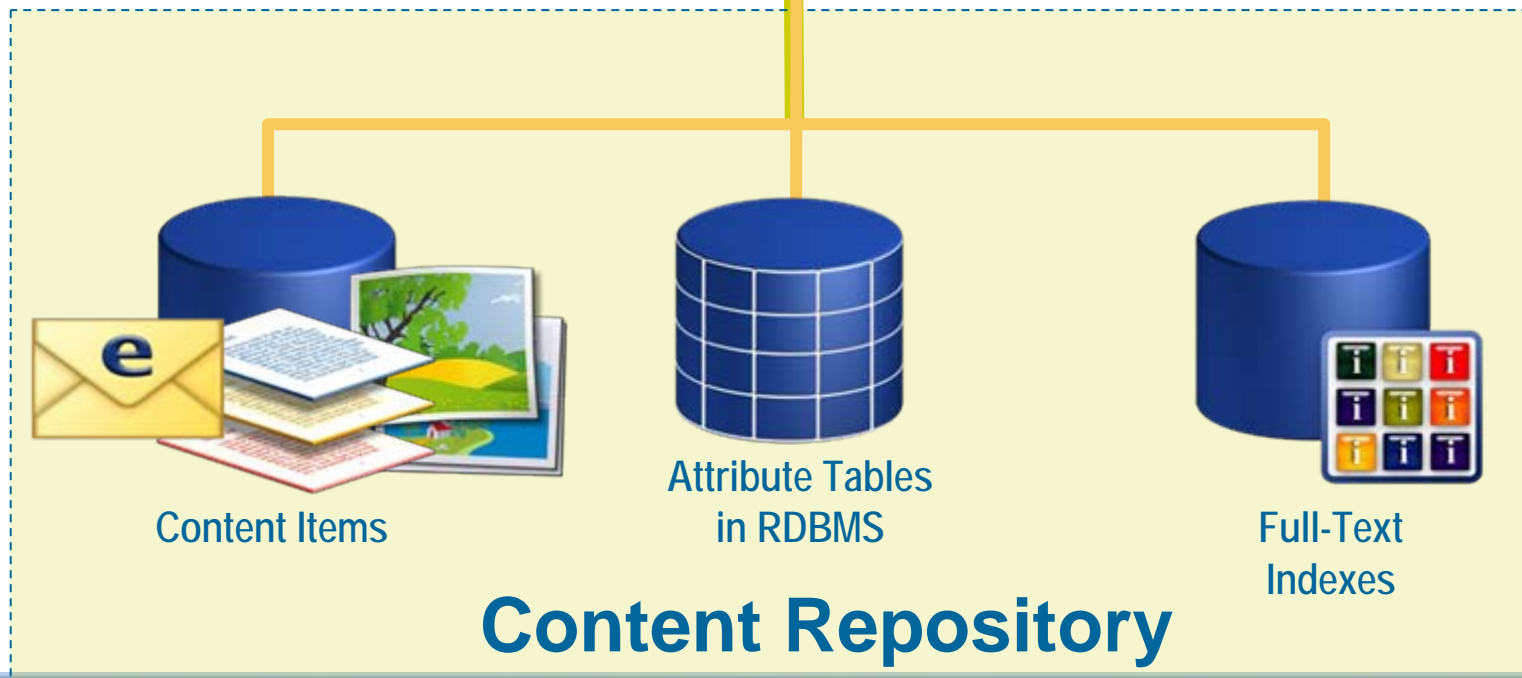
Discussions
IM
Notes
Projects



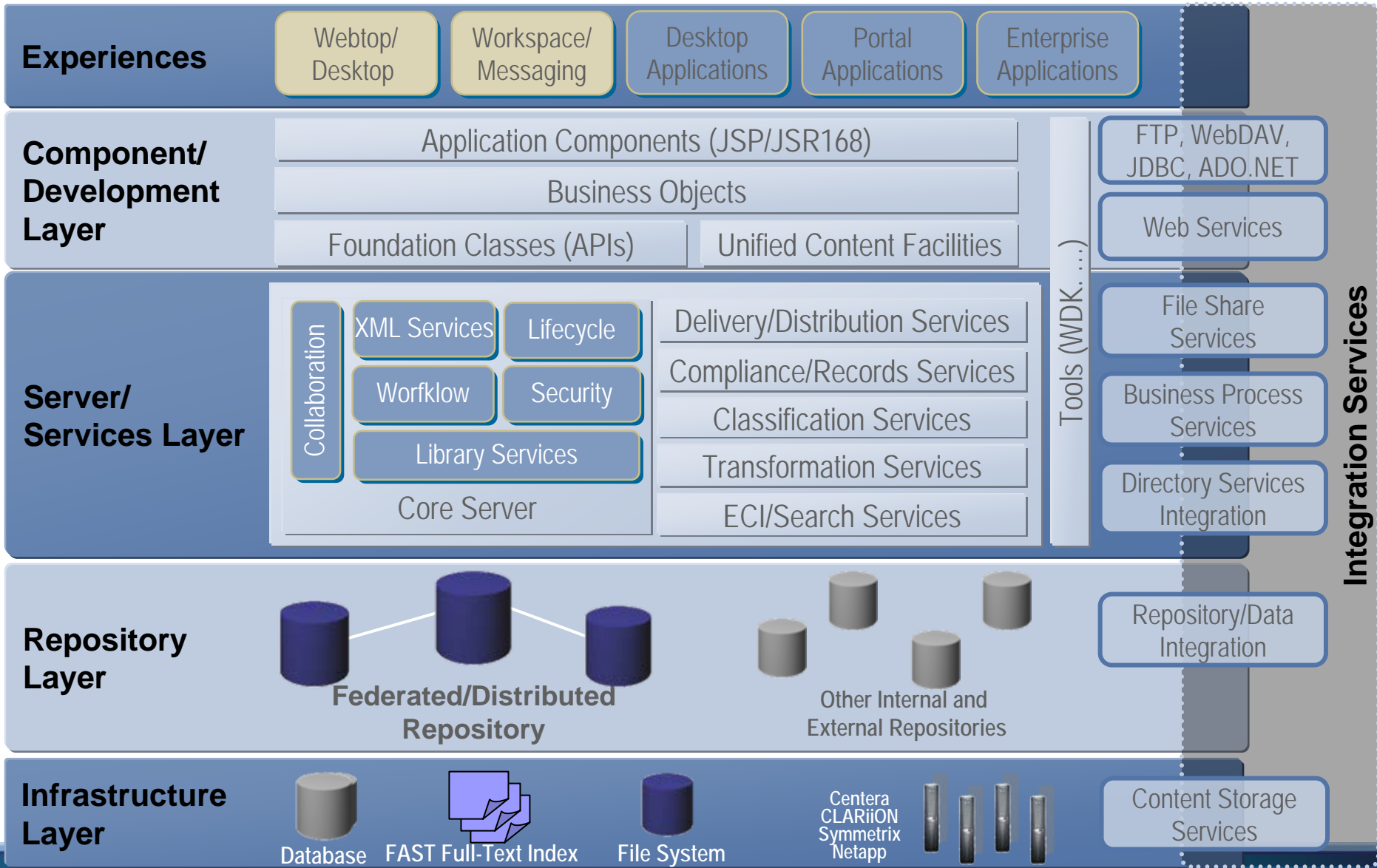
Agenda

- ✓ **Fundamentals of Managing Content**
- ✓ **Documentum Architectural Highlights**
- ✓ **ECM and the Information Infrastructure**
- ✓ **Q&A**

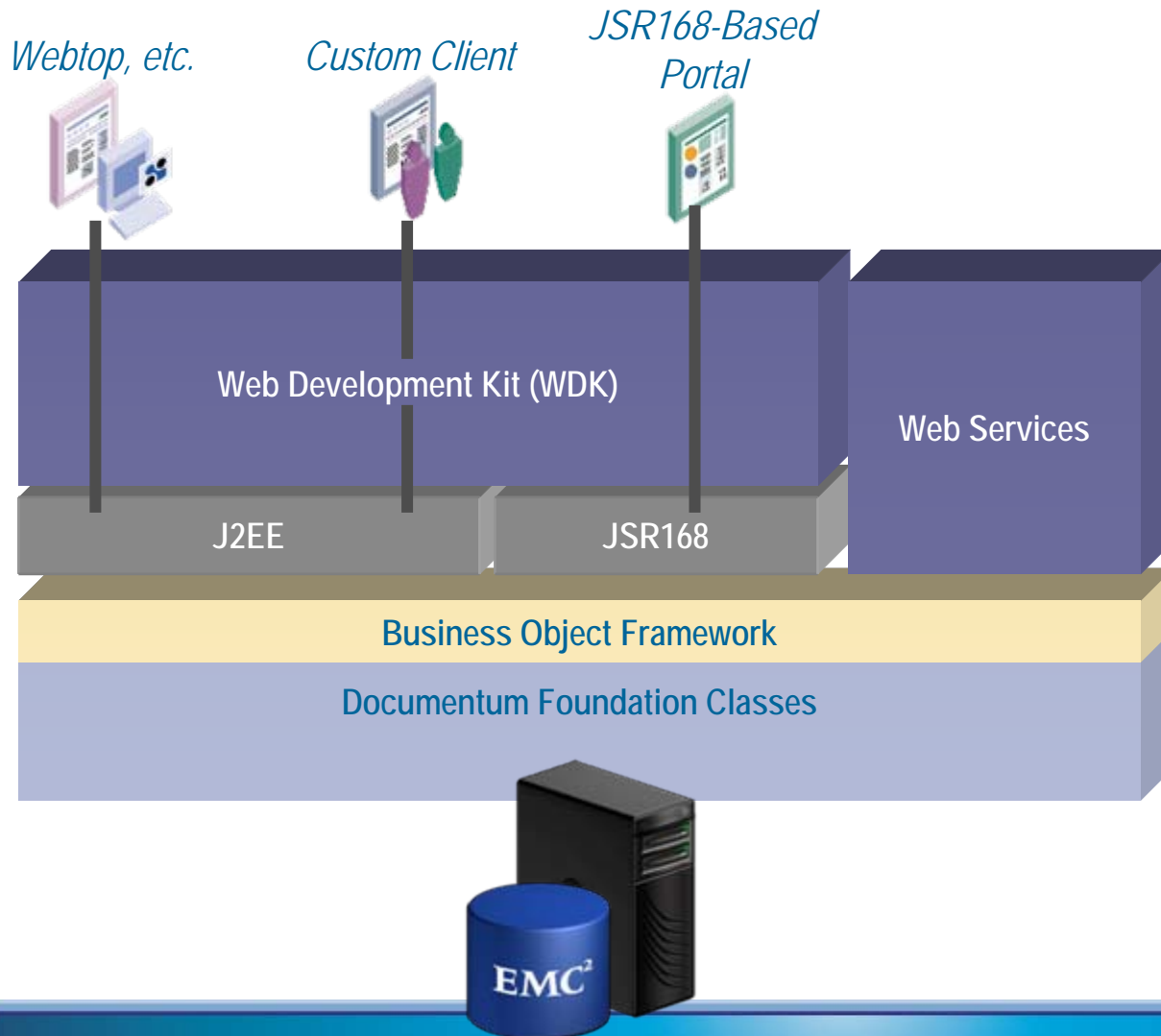
Content Server



Content Repository



Componentized Development Environment



Documentum Foundation Classes

- Lowest level API

Business Object Framework

- Mechanism for extensibility of core object behaviors and services

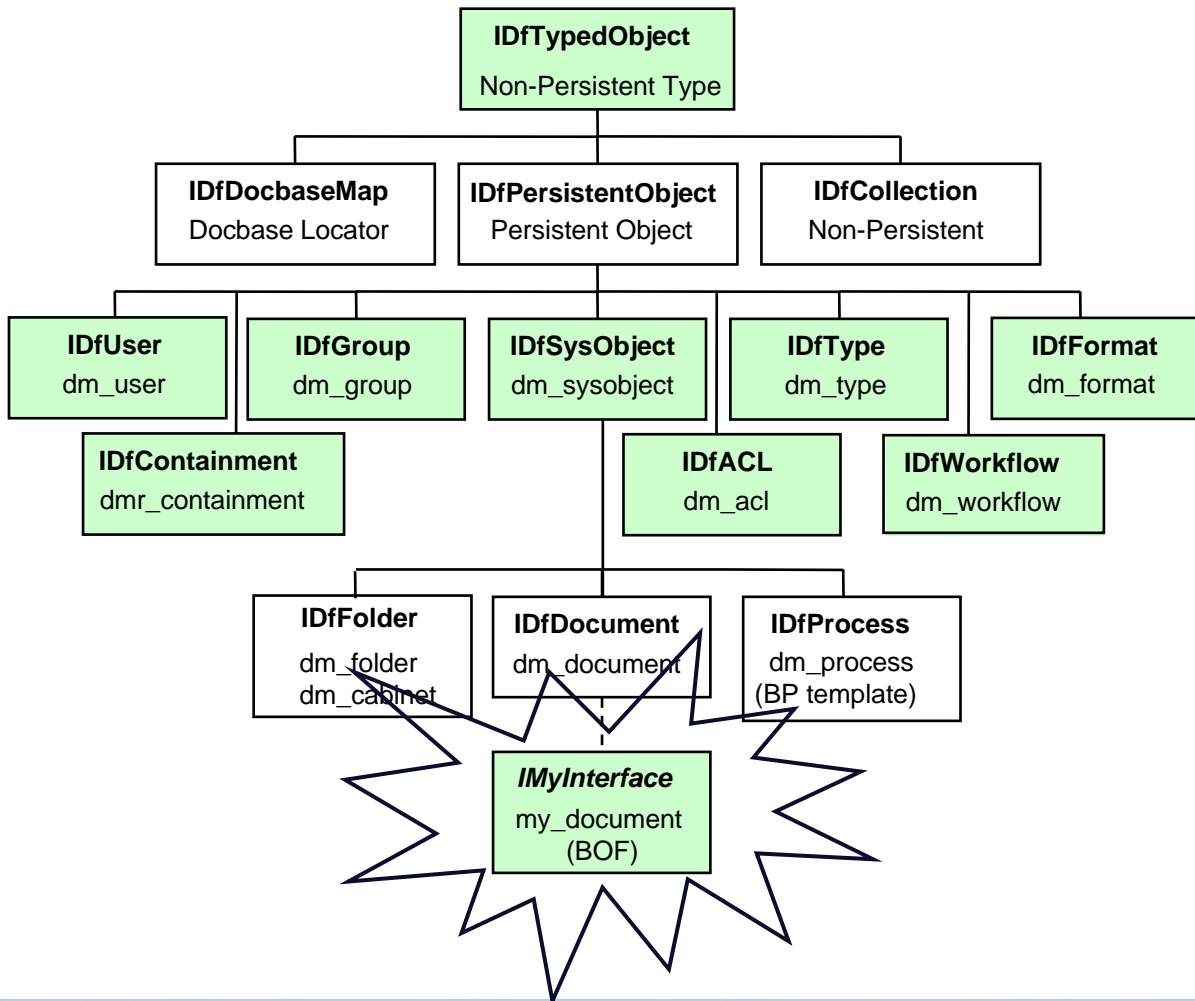
Web Services

- Allows services to be invoked from any platform

Presentation Customization

- Portals
- Web Development Kit
- Custom, AJAX, Mashup, etc.

Documentum Foundation Classes (DFC)



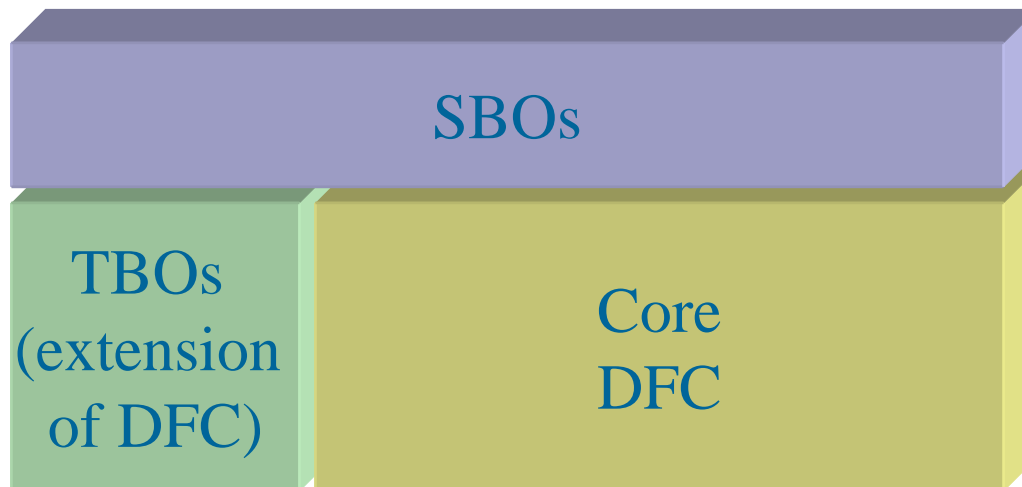
Object-oriented programming interface to the Documentum Content Server

Implemented as a set of Java classes and interfaces, along with a Java-COM bridge

- ◆ Implement object type specific behaviors
- ◆ Some basic operations
 - Client, session, persistent object, query collection
- ◆ Some provide high-level logic
 - Data validation, workflow management, version policy, Business Object Framework (BOF), XML functionality
- ◆ Extensibility through BOF

Business Object Framework

- ▼ **A Type-based Business Object (TBO) allows you to override DFC methods**
 - ◆ Enables DFC extension
 - ◆ A client-side 'trigger' to call your custom logic
- ▼ **A Service-based Business Object (SBO) provides functionality as a service to be used by applications**
 - ◆ A high-level layer on top of DFC, can use any methods
 - ◆ Could be wrapped as a Web Service



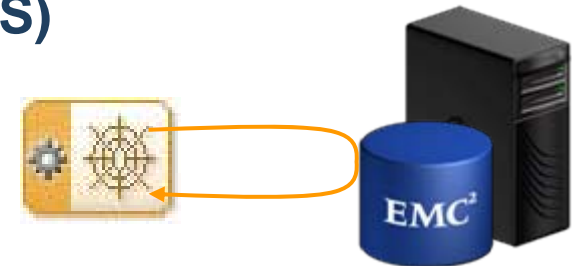
Service-Oriented Enterprise Content Management

✓ EMC Documentum Foundation Services (DFS)

- ◆ Part of forthcoming “D6” stack release this year

✓ DFS...

- ◆ Next generation of CMA SOA
- ◆ Consists of tools, services and runtime support
 - Deployable platform services OOTB
 - SDK for service and solution developers (internal and external)
- ◆ Facilitates via server-side and client-side libraries
- ◆ MTOM and UCF content transfer
- ◆ WS-* standards and relevant protocol support



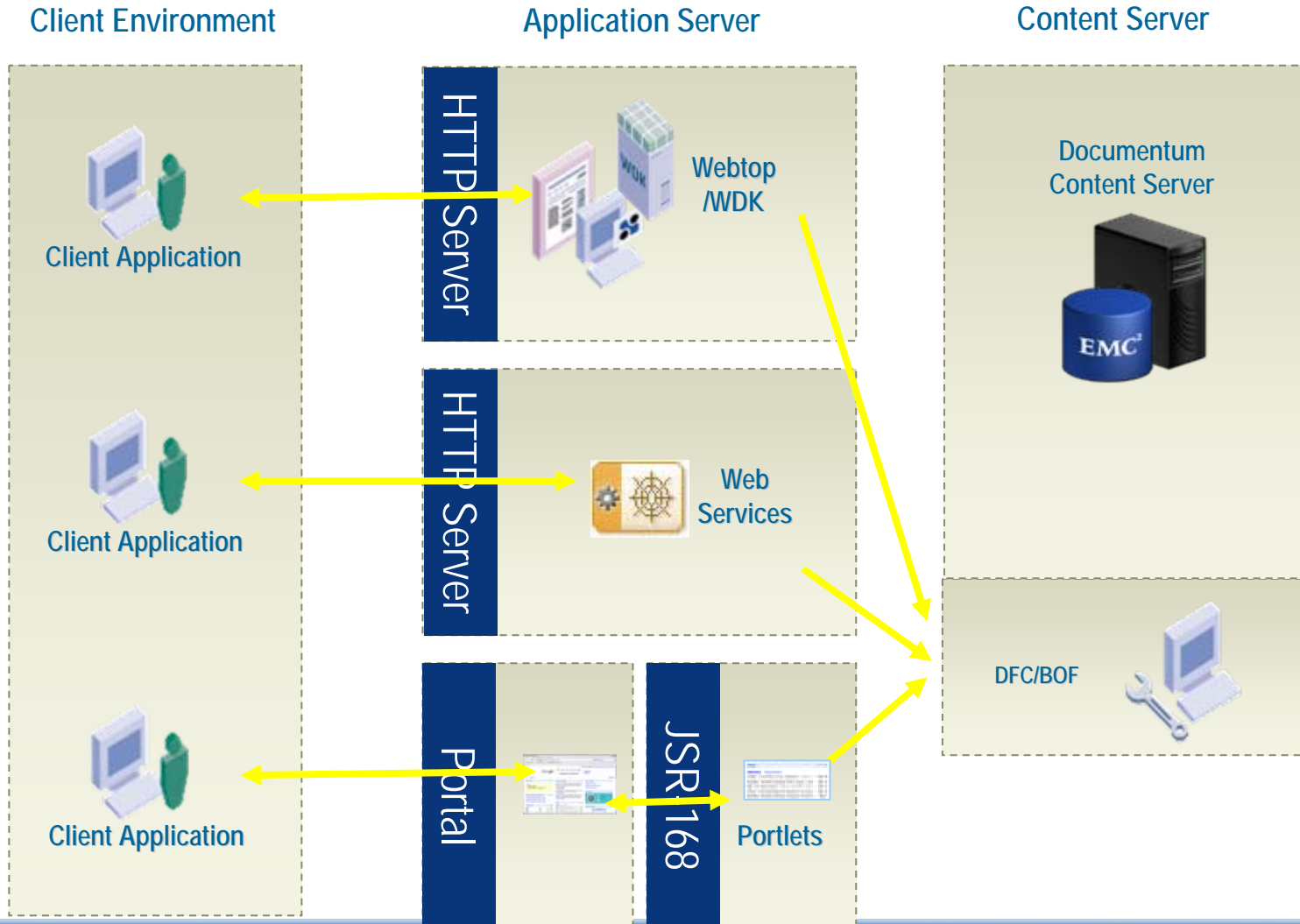
✓ Supports any client language and environment

- ◆ Java First! and WSDL First!
- ◆ Remote client process can be anything
- ◆ WSDL is the primary API

✓ Supports static and dynamic client invocation

- ◆ Static invocation uses generated client code
- ◆ Dynamic uses XML/DOM

Web Client Architecture

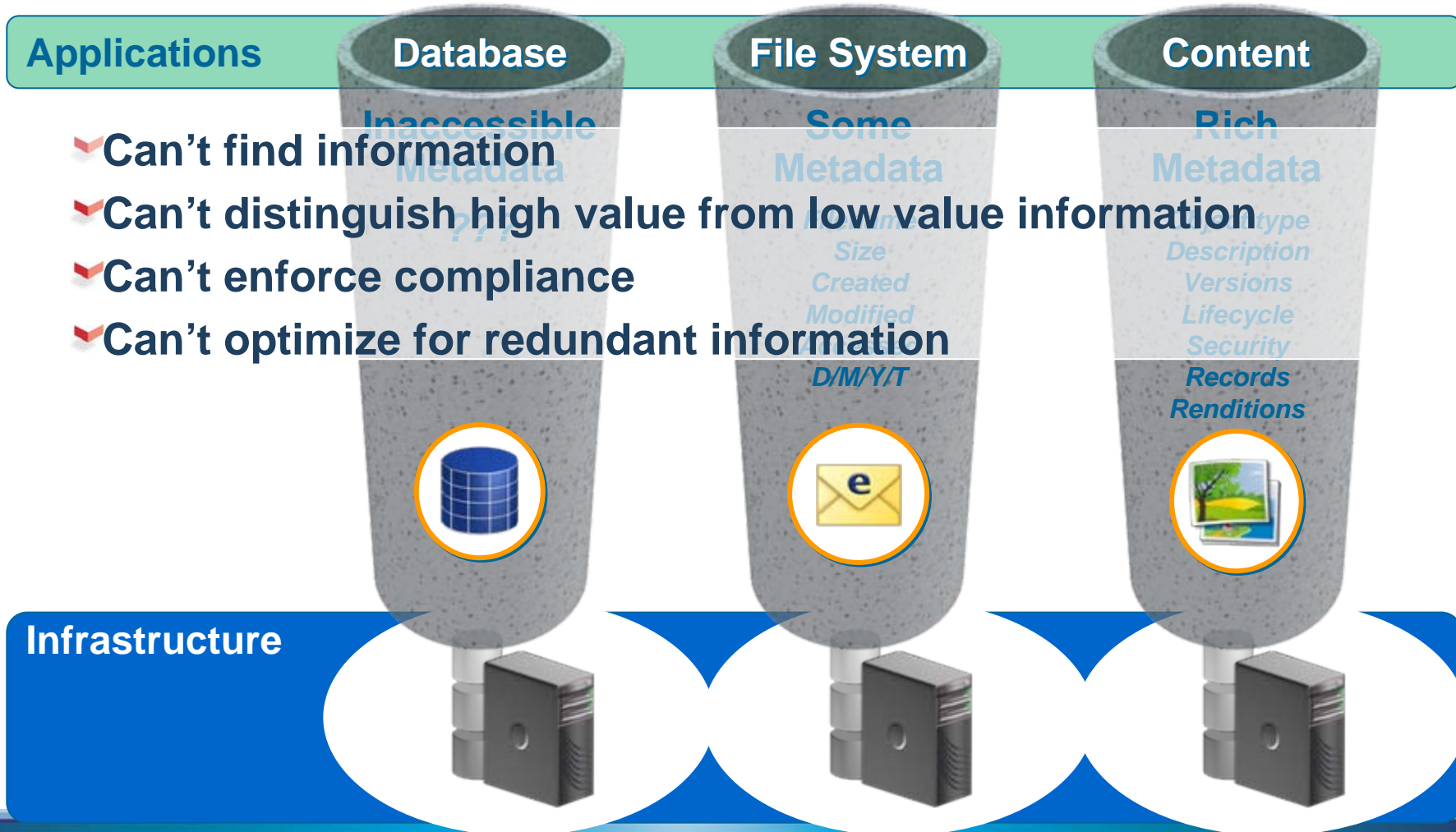


Agenda

- ✓ Fundamentals of Managing Content
- ✓ Documentum Architectural Highlights
- ✓ ECM and the Information Infrastructure
- ✓ Q&A

Information Lifecycle Management (ILM) Today:

Isolated, Incomplete, Purpose-built Metadata

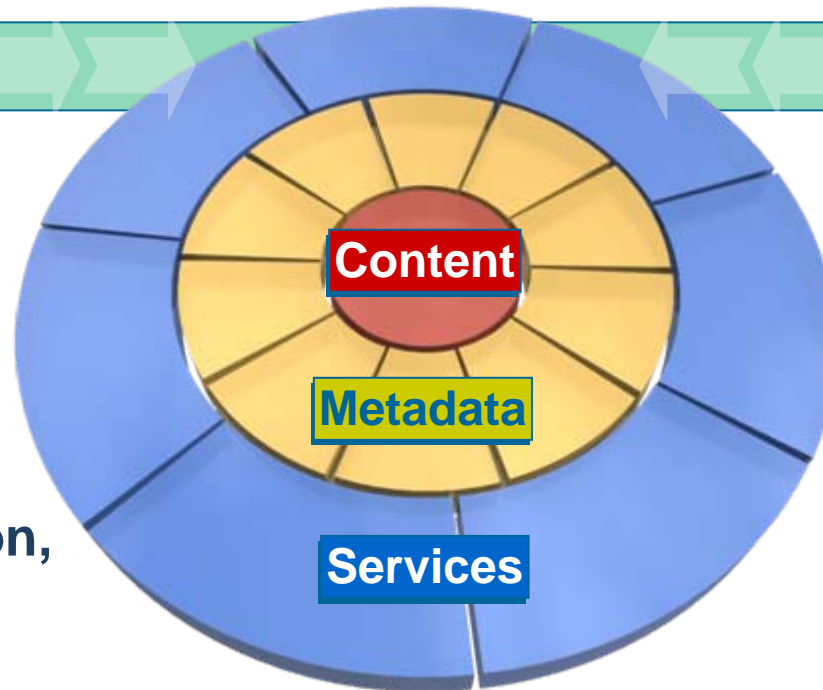


Intelligent Information Management

Unified information Metadata

Applications

- ✓ Repository independent
- ✓ Discover and classify information
- ✓ Guides intelligent placement protection, compliance and security



Tiered Infrastructure



EMC Infoscape Overview

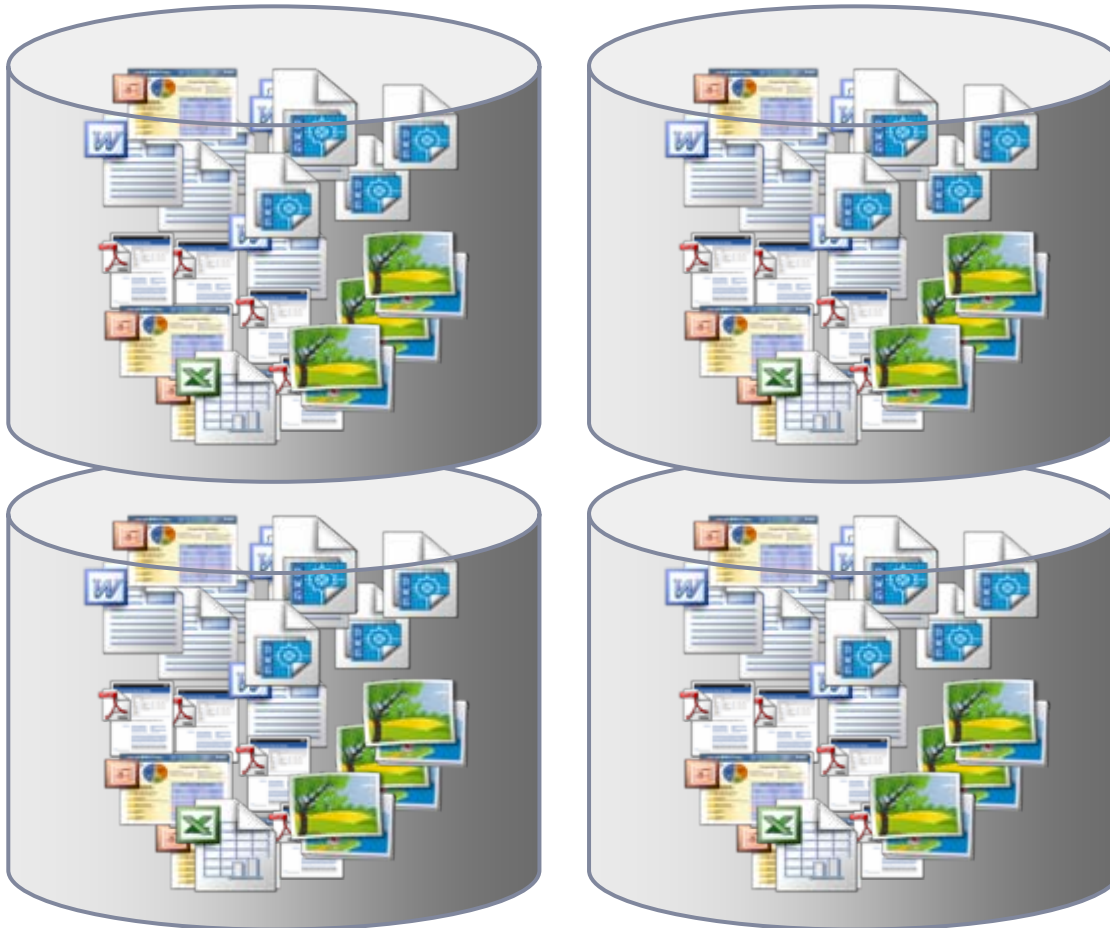
- ✓ **Enterprise information risk management solution**
 - ◆ Compliance, governance, security, eDiscovery
- ✓ **Better management of unstructured data**
 - ◆ Classify and take actions
- ✓ **Leveraging Best-of-Breed EMC Technologies**
 - ◆ Seamless integration with EMC solutions



Best Data Classification Solution



The Traditional Approach: *Before*



- ▼ No idea what's inside these files: Are they *legal* documents?
Customer information?
 Directions to the restaurant?
- ▼ Businesses at risk:
 Are files protected?
 Secured?
- ▼ All files are managed the same way due to a lack of knowledge needed to determine whether the content presents value or risk to the business

EMC Infoscapes: *After*

Files are Secured, Saved, and Stored Based on Business Value and IT Objectives



- 1 Unified EMC Infoscapes console
- 2 Search
- 3 Analysis and reports

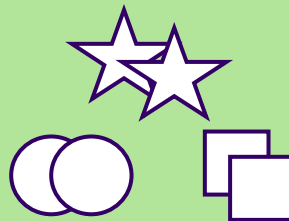
EMC Infoscapes Delivers Intelligent Information Management

Leverages User-defined Policies for Ongoing Information Management

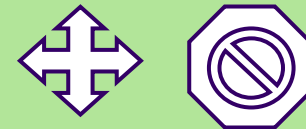
Information Management Policies



Discover



Classify



Act

Policy Analysis and Execution Platform

How It Works: Infoscape Discovery – Servers and File Shares

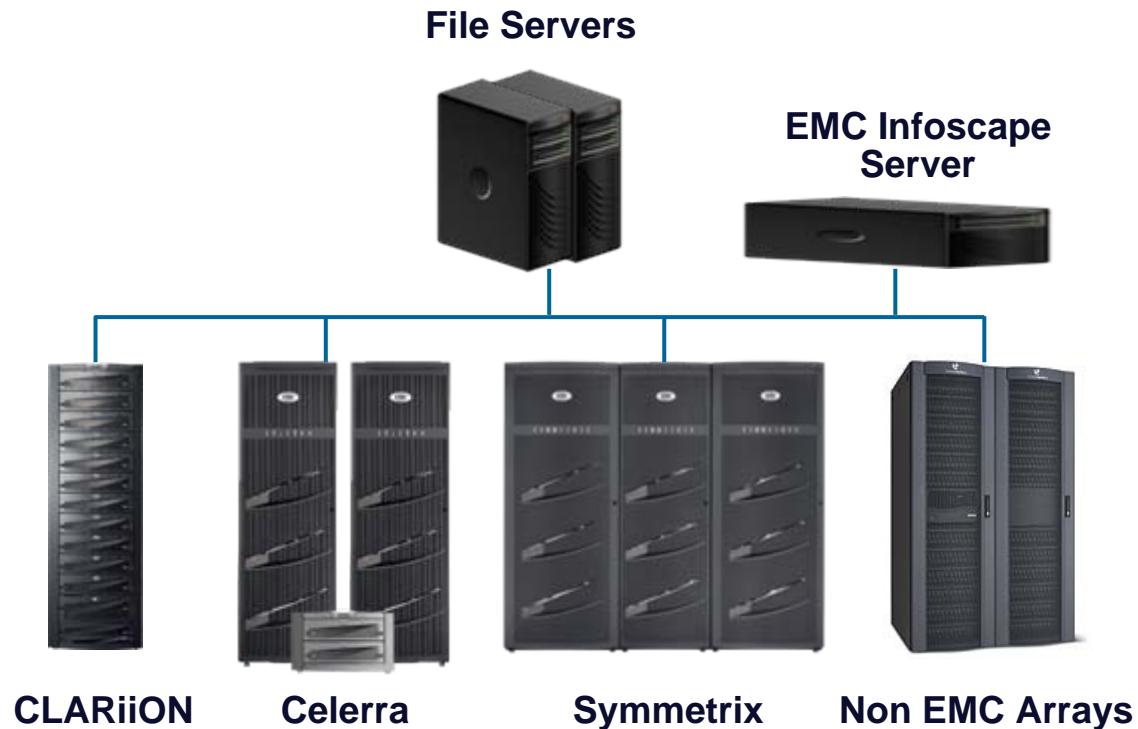
Infoscape discovers File Servers and File Shares

- ◆ CIFS-accessible
 - Support for Windows operating systems
- ◆ NFS support
- ◆ Back end storage is irrelevant

Infoscape “crawls” each file share, gathering file system attributes

Celerra “special case”

- ◆ Bulk load of file system metadata



How it works: Classification Engine

Classification is done by defining selection and exclusion rules per category

Super-fast metadata classification

- File-system-based attributes

Fast content-based classification

- Keywords and patterns

Deep content-based classification

- Content based with stemming, proximity, and weighting

File Attributes



Size:	555 KB (568,832 bytes)
Size on disk:	556 KB (569,344 bytes)
Created:	Friday, March 31, 2006, 12:46:22 PM
Modified:	Thursday, March 30, 2006, 12:48:56 PM
Accessed:	Today, April 06, 2006, 10:48:47 AM

File Content



Fourth-quarter continuing revenues fourth quarter 2004. Industrial sales in and the net effects of acquisitions. Fir about the same as last year, reflecting **Financial** services **revenue** growth ex SFAS 133 and the effects of disposition was 5%.

How it Works - Infoscapes Automated Data Orchestration Services

Policy-Driven, Automated Orchestration

- ◆ Policy-based actions bring classified data into compliance with stated service objectives and policies
- ◆ Broad range of orchestration actions supported

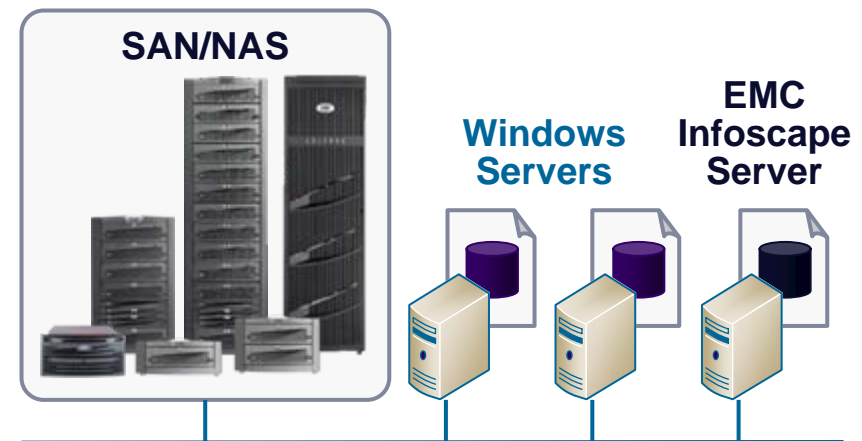
Move/Archive/Copy

- ◆ Automated migration from Celerra, and to Centera
- ◆ Copy to Documentum Repository
- ◆ Forensically Sound Copy

Apply Security

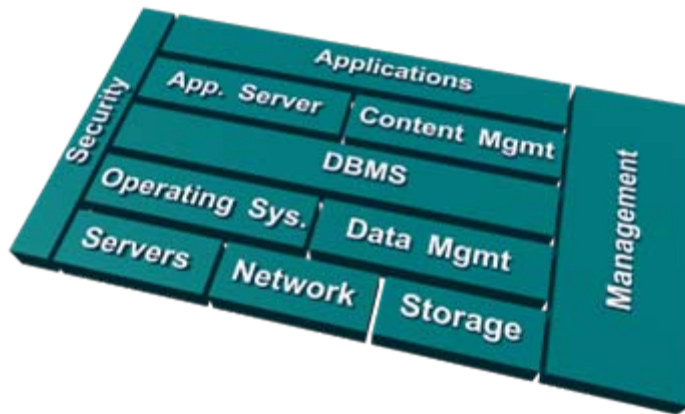
- ◆ Apply IRM/Authentica DRM

Targeted Index and Search



EMC Believes IT Strategy is Evolving

Yesterday



IT focus on the *technology*:
technology infrastructure

Tomorrow



IT focus on the *information*:
information infrastructure

Agenda

- ✓ **Fundamentals of Managing Content**
- ✓ **Documentum Architectural Highlights**
- ✓ **ECM and the Information Infrastructure**
- ✓ **Q&A**

EMC²[®]

where information lives[®]