



Content-Oriented Application Architectures

Jean Barmash, Alfresco



IT Architect Regional Conference 2008



About Me / Alfresco

- Director, Technical Services
 - Blog – www.nywebguy.com
- Alfresco Software, www.alfresco.com
- Open Source Enterprise Content Management
 - Document Management
 - Web Content Management
 - Collaboration
 - Enterprise 2.0 / Social Computing Platform





Open Source = better ROI

- Lower Cost, no upfront licensing
- Lower Maintenance Costs
- Lower Training Costs
- Code Available
 - Easier to learn
 - Ability to fix bugs yourself



IT Architect Regional Conference 2008



Content Oriented Apps Examples

- Manage outsourcing process paperwork
- Email Archiving
- Today's Websites
 - Content Pages
 - Video
 - Images
 - Audio
- Records Management



IT Architect Regional Conference 2008



Requirements

- Store Large Files
- Granular Access Control
- User Management
- Transactions
- Query / Search
- Observation / Eventing
- Running Code
- Structured + Unstructured Data
- Versioning
- Exclusive Item Locking
- Content Lifecycle /

IT Architect Regional Conference 2008



What are some ways we store information?



IT Architect Regional Conference 2008



Systems for Storing Information

- Databases
- File Systems
- Directories
- CAS – Content Addressable Storage
- S3
- Google's BigTable



IT Architect Regional Conference 2008



Outline

- We Like FileSystems – store files
- We Love Databases – store data
- Content Repository – Files + Data
- Content-Oriented Applications
- Implications



IT Architect Regional Conference 2008



File Systems - Good

- Trees & Hierarchies
- Ability to Browse
- Storing Large Files
- Access Control



IT Architect Regional Conference 2008



File Systems - Bad

- No Query / Search
- No Versioning
- No Observation of Files
- No Exclusive Locking
- Limited metadata – no modeling ability
- No Transactions
- No Ability to Run Code



IT Architect Regional Conference 2008



Databases - Good

- Query / Search
- Observation / Eventing (Triggers)
- Transactions
- Ability to Model Data



IT Architect Regional Conference 2008



Databases - Bad

- Multi-Value Properties
- No Versioning
- BLOBs – for storing Large Files
- Trees & Hierarchies
- Access Control
- Only structured content
- Can't Browse



IT Architect Regional Conference 2008



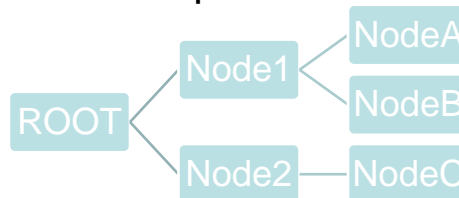
Content Repository Requirements

	File Systems	Databases
Hierarchies	Y	
Store Large Files	Y	
Granular Access Control	Y	
User Management		Y
Transactions		Y
Query / Search		Y / N
Observation / Eventing		Y
Running Code		Y
Data Modeling		Y
Store Structured Data		Y
Store Unstructured Data	Y	
Versioning		
Exclusive Item Locking	Y / N	
Content Lifecycle / Publishing		



Content Repository Abstraction

- One or more Workspaces



- Each workspace - Tree of Nodes
- Node has multiple Properties
 - Data is stored in properties



Content Modeling

- Content Types
 - Properties
 - Inheritance
- Mix-Ins / Aspects
- Associations
 - Child Associations
 - Peer Associations
- Data Integrity Checks

IT Architect Regional Conference 2008



Note on implementation

- Many ways to implement it
 - Typically involves combination DB and File System
 - Can be DB-only



IT Architect Regional Conference 2008



Review – Content Repository Features

- Store Large Files
- Granular Access Control
- User Management
- Transactions
- Query / Search
- Observation / Eventing
- Running Code
- Structured + Unstructured Data
- Versioning
- Exclusive Item Locking
- Content Lifecycle /



IT Architect Regional Conference 2008



Content Repository Deployment

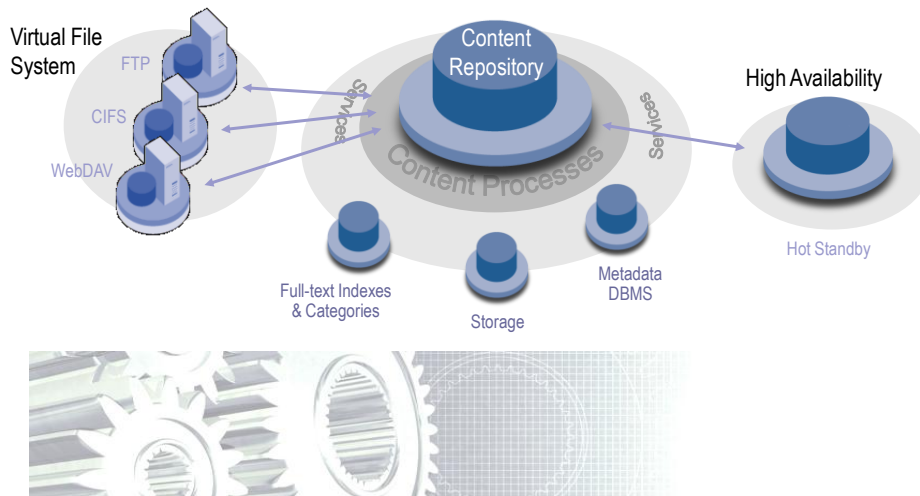
- Scalability
- High Availability
- User Directory Integration
- Backup / Recovery



IT Architect Regional Conference 2008



Content Repository



Standardization Efforts

- WebDAV – IETF Web-based Distributed Authoring and Versioning
- JSR-170 – Java Content Repository (JCR) API
- JSR-283 – Next generation of JCR
- iECM – AIIM Interoperable Enterprise Content Mgmt
- ATOM-PUB
- Other Standards Work
- No SQL-Equivalent for Content Repositories Yet



Content Repository is a great abstraction. But to enable content oriented applications, we want more.



IT Architect Regional Conference 2008



Additional Features

- Content Services
- Business Processes
- Collaboration Features
- Other



IT Architect Regional Conference 2008



Content Services

- Content Transformations, i.e. doc to pdf
- Plain Text Searching
- Multi-Format Publishing / Renditioning
- Metadata Extraction
- Library Services
 - Taxonomy
 - Classification
- Image Manipulation / Thumbnails

IT Architect Regional Conference 2008



Business Processes

- Task Management
- Workflow
- Scheduled Processes
- Auditing / Compliance
- Business Process
- Lifecycle management
- Membership Management

IT Architect Regional Conference 2008



Collaboration

- Pre-Built Functionality / Widgets
- Email Integration
- Social Networking
- Templated Spaces / Micro-Sites
- In-Context Editing



IT Architect Regional Conference 2008



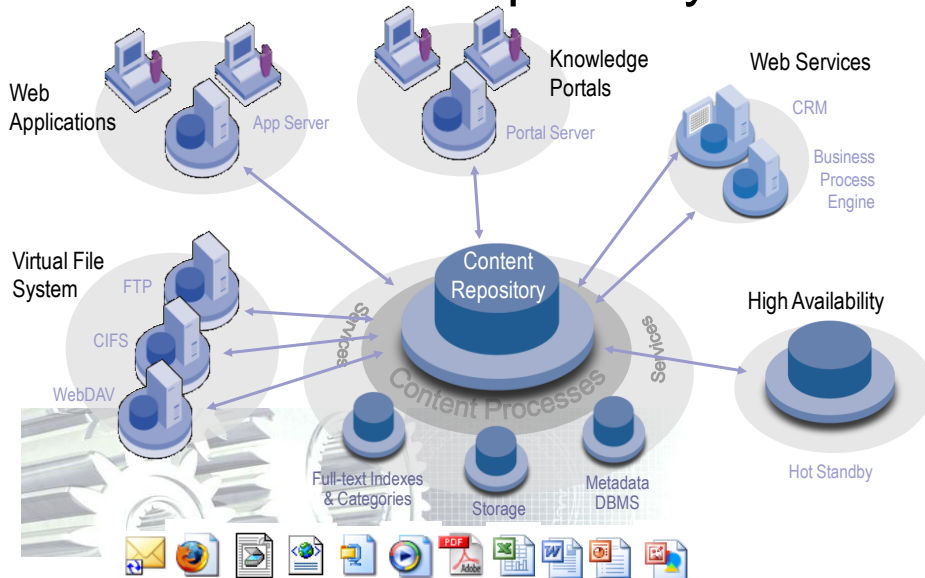
Other Features

- Multiple environments (dev, staging, prod)
- Specialized Versioning (i.e. Subversion-like)
- Scheduled Publishing / Expiration
- Extensibility and Integration
- Internationalization / Localization
- Digital Rights Management
- Ability to Expose Data Model to other applications

IT Architect Regional Conference 2008



Content Repository



Applications of Enterprise Content Mgmt

- Web and Portal Content Management
- Collaborative Development
- On Demand Publishing
- Compliance
- Records Management
- Document Management
- Digital Asset Management
- Source Control



Other Implications

- Great integration platform for content
 - Wiki, Blogs, Documents, Web Content
- A content repository should be client-runtime independent
 - Content as a Service in your SOA strategy
- Can expose Content Repository as Filesystem, FTP, Shared Drive

IT Architect Regional Conference 2008



Conclusion - 1

- Databases and Filesystems aren't well equipped to handle unstructured content
- Today's content explosion stretches the existing systems:
 - Content-oriented websites
 - User-generated Content
 - Scalability Requirements
 - Content Integration

IT Architect Regional Conference 2008



Conclusion - 2

- Though the content Repository offers a great abstraction, additional services are typically needed
- Using Content Repository as a backend allows encapsulation of business logic and processes, while exposing content services to other applications.



IT Architect Regional Conference 2008



Questions?

- Alfresco, www.alfresco.com
- Open Source Enterprise Content Management
 - Content Repository
 - Document Management
 - Web Content Management
 - Collaboration
 - Enterprise 2.0 / Social Computing Platform



IT Architect Regional Conference 2008



References

- <http://www.gadgetopia.com/post/5940?rl>
- <http://developers.sun.com/learning/javaoneonline/j1sessn.jsp?sessn=TS-4474&yr=2006&track=coreenterprise>
- <http://jcp.org/en/jsr/detail?id=170>
- <http://www.aiim.org/standards.asp?ID=29284>
- [http://en.wikipedia.org/wiki/Atom_\(standard\)](http://en.wikipedia.org/wiki/Atom_(standard))
- http://en.wikipedia.org/wiki/Content_management_system



IT Architect Regional Conference 2008



Scale Requirements for an Enterprise Platform

- **Scale in Information**
 - Complex search, structure & classification of information
- **Scale in Activity**
 - Complex information per activity with dynamic views with full object-level security
- **Scale in People**
 - Up to 100,000s of readers and writers of gigabytes and terabytes
- **Scale in Geography**
 - Sharing of information across continents in real-time

