FileNet Integrated Document Management Database Usage and Issues

Daniel S. Whelan, Ph.D.
FileNet Corporation
3565 Harbor Blvd.
Costa Mesa, CA 92626
(01).714.966.3509
dwhelan@filenet.com

ABSTRACT

The FileNet Integrated Document Management (IDM) products consists of a family of client applications and Imaging and Electronic Document Management (EDM) services. These services provide robust facilities for document creation, update, and deletion along with document search capabilities. Document properties are stored in an underlying relational database (RDBMS); document content is stored in files or in a specialized optical disk hierarchical storage manager. FileNet Corporation's Visual WorkFlo® and Ensemble® workflow products can be utilized in conjunction with FileNet's IDM technologies to automate production and ad hoc business processes respectively.

This talk will discuss how Integrated Document Management requirements affect an IDM system's usage of a RDBMS. Some of the areas to be discussed include:

1. Schema and Schema Evolution

FileNet's IDM products provide a single inheritance class model (see AIIM's Document Management Alliance 1.0 Specification [1]) object oriented API. IDM objects, such as documents, have a class and an associated set of properties. It is necessary to provide for dynamic extension of both the properties in a class and the set of classes. This schema evolution must behave reasonably in terms of behavior and performance under conditions where there may be large numbers (tens of millions or more) of extant persisted objects. IDM schema and schema evolution requirements impact RDBMS physical table layout considerations.

2. Security

The FileNet IDM services provide fine-grained security on IDM objects (documents, folders, etc.). The user views independently securable IDM objects has having possibly distinct access control lists (ACLs). This results in a need to implement authorization at the object (row) level that cannot be implemented utilizing traditional RDBMS table or view based security models. FileNet's IDM services products store ACLs in the RDBMS and relate those ACLs to secured objects. ACL space management and the efficiency of authorization checks are important characteristics of the IDM architecture.

3. Events and Workflow

IDM systems provide event frameworks, such that an action on an IDM object results in a server- or client-side event notification. Event notifications must often be durable, reliable, and transactionally scoped. Workflow mechanisms are provided to manage document life cycles and automate business processes. Workflow utilizes the event framework in that a workflow may be triggered by an event and may generate other events. The workflow software typically relies on a persistable queuing infrastructure, often provided by the RDBMS and itself places unique demands on the database system.

4. Distributed Access

IDM customers typically deploy departmental solutions, e.g. an IDM system for the Legal Department and a separate IDM system for the Engineering Department. It is commonplace for an IDM user to want to locate documents (a) without specific knowledge of which IDM system a document is in or (b) across all of an organization's IDM systems. IDM systems must provide feature rich mechanisms to search across multiple IDM systems.

Documents may also be cataloged in multiple databases, e.g. a RDBMS and one or more Full Text databases. IDM systems must provide capabilities to query across RDBMS and Full Text databases, ideally allowing arbitrary construction of WHERE clauses involving traditional and content operators.

5. Manageability

IDM systems impose manageability requirements on the underlying RDBMSes. In addition to traditional space management issues, IDM systems may be characterized by more frequent updates of table and index definitions. IDM systems also tend to have large user communities (many thousands of potential users) whom tend to generate ad hoc queries. It is useful to be able to meter or restrict the resource consumption of ad hoc queries to deliver an optimal response time distribution across the user community.

6. References

[1] AIIM International's Document Management Alliance 1.0 Specification. http://www.aiim.org/dma