

Enterprise Web based File Management - A System Architecture and Solution

Dharmendra Choukse
Institute of Engg. & Science
IPS Academy
Indore, India

Umesh Kumar Singh
Institute of Comp. Science
Vikram University
Ujjain, India

ABSTRACT

In the expansion of web Application system architectures, software programmers, product test engineers, and clients who may be distributed over the world need brilliant communication and collaboration. They require easy access to necessary data and sharing of relevant documents of the project which are distributed in a network of resources and users. Hence, web-based applications, decentralized repositories, and databases are needed to store and manage project and process development information. However relational databases work poorly in performance, scalability, and reliability while processing large amounts of big data spread out across many servers. This paper works for managing Web Based File Management project documents in an Internet-based collaborative environment using NoSQL database technology. The research work focuses Web Based File Management using Apache Couch DB which is an open source document-oriented database with REST web services interface.

Keywords

Servers, Web services, Distributed databases, Relational databases, Computer architecture, Collaboration

1. INTRODUCTION

Every large organization or enterprise deals with millions of files or documents on regular activities of both technical and non-technical. The teams or groups involved in activities always transact their communications through standards and protocols. And these protocols require documents. Hence the files are transacted across hundreds of teams and thousands of employees through storage media or e-Mails. But all these documents go un-available once if they served their purpose. Subsequently, dependency on source arises when the documents are in need. Also, the access to documents in need may not be highly available when the following:

1. The author of documents are busy,
2. No proper data structure or classifications exists
3. Largetime consumption in searching and recovering the documents and
4. Reliability is less for hard disks and media.
5. Such scenario dampens the spirit of team collaboration and keeps team members waiting indefinitely for access to right documents at the right time. Thus an exclusive architecture with solutions to provide access to right users' with right documents at the right time for team collaboration and sharing documents is essential to lead the way for knowledge management.

2. ARCHITECTURE

The system architecture is designed in IPS ACADEMY as described here. It facilitates document management such that it enables users to find the right documents at theright instant of time from right sources. This architecture contains three layers and all perform its well-defined functions to facilitate storing, retrieving, Sharing, online editing and collaborating documents for teams.

Layer 1: Application Server with Database

Layer2: Organization/Product Structure (FTP Server)

Layer 3: Physical Storage Layer

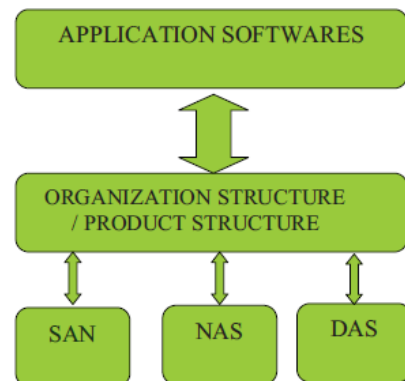


Fig 1: Architecture

The concept behind the architecture is to bring an object-oriented approach in practical scenario of our Organization and library. This concept provides the basis for containing all data and its associated documents at Storage Area Network (SAN) in accordance with information transaction schema of an organization.

2.1 Model

The functional requirements of our organization are transformed into business logic and deployed at the Application layer as workflows.

2.2 Architecture & Solution

The whole architecture is envisaged at our premises with application server transacting over a network through a unified file system at FTP-server. This unified file system is exhaustive and amendable for Present scenario and scalable for future needs of our organization.

3. APPLICATION SERVERS

Application servers are servers that contain the business logic or workflows being implemented in web servers. It functions as a controller listening requests of users and executes the work flows Required for it. It routes requests to the 'intermediate file system server' for access to the corresponding Documents when the access list permits users from database.

In VSSC, there are multiple-application servers – each meant for its own set of requirements with a local Database. Some of the applications are as follows:

1. Documents Management
2. Design Collaboration
3. Library Server

3.1 Documents Management

Technical documents as uploaded for publications. These documents go to the respective experts who suggest their corrections. The corrections or suggestions are feed back to the authors. Authors update the document with required corrections and suggestions. The then updated documents are processed for final approval. The approved documents get stored and become accessible to respective teams.

3.2 Design Collaboration

The design documents as posted are reviewed online by domain experts. The corrections and suggestions would be updated online. It facilitates design collaboration among designer with its associated teams.

3.3 Library Server

The articles of various domains and systems can be grouped as per product tree structure. The server routes users' requests to the respective domains of product structure. The tree structure brings all technical documents specific to the product so that insightful knowledge of a given product could be used for its advancements in other missions.



Fig 2: 1/2 layer description of architecture

Local Database

Personnel access rights information is maintained at local database of the respective application server. User's requests are checked with Access Control List (ACL) defined at local database. Based on the Authorization rights, it is routed to 'File System Server' for access. The local database can also be used to keep a log of all users' transactions at application servers.

4. GENERIC FILE SYSTEM

The generic file system can be categorized into two parts.

- a. Organization/Functional Structure
- b. Product Structure

Functional Meta-Data Structure

The generalization observed at organizations applications are organization structure or a product structure - the basic structure that essentially defines the communications and transactions between entities so as to realize the sustained growth in goals and objectives.

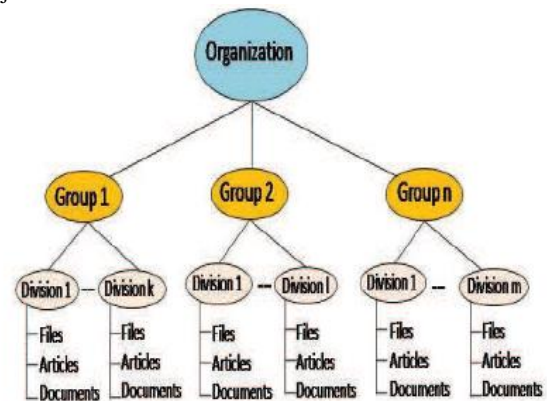


Fig 3: Organization Structure

Since the structure is pervasive in the whole organization, the storage and access to the documents in accordance with the enterprise structure - facilitates users to make transactions and communications in the wake of missions and goals of the organization.

Such a partial or full implementation of enterprise structure – as a file system has brought out a system of storing and retrieving documents at a single point of access for the whole organization.

Product Life Cycle Meta-Data Structure

The product life cycle information of any system necessarily advocates systems Engineering approach where each system contains a set of sub-systems. The product is made by a set of defined processes right from requirements collection to till it is realized.

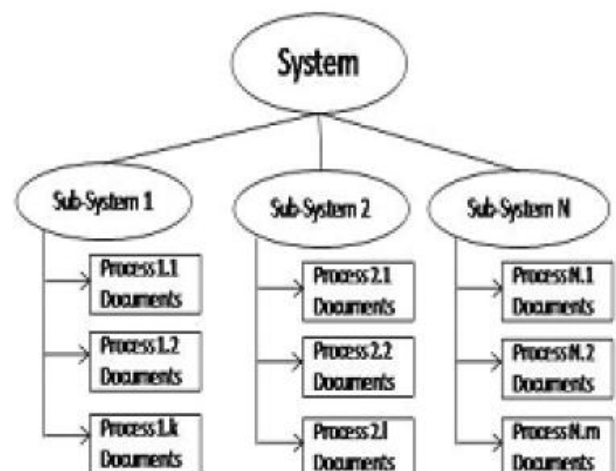


Fig 4: Product Structure

The successful outcome of Organization files management is promoting Product Lifecycle structure to get it implemented as a file system. This would make users store, retrieve and view documents as per the processes requirements from a single point.

Further, it enables us to find status quo of a product and comprehensive execution of processes of an Overall system.

5. ORGANIZATION/ PRODUCT STRUCTURE AT SECURE MULTI-PROTOCOL SERVER

The organization structure or a product structure is a vital graph that determines to route personnel requests to its corresponding folders and documents. The whole structure is contained in a file system. The file system enlarges whenever the organization expands goals of its vision. And the huge file system would be serving enormous users for sharing documents and collaborating users online.

If such a file system is shared among various disciplines or application servers in full or in partial, then it should have an efficient server supporting multiple secure protocols such as (*HTTP, https, ftp, ftps, telnet, sftp, ssh2 secure shell*) for sharing and collaborating files among huge users. Nowadays, secure multi-protocol servers are available at free cost.

6. USERS OF SYSTEM

Users become the object of the system since the data/documents and files are encapsulated in the form of Product / Organization structure under the purview of ACL – Access control list. The whole structure acts as a skeleton or backbone for the system. The workflows of applications act as methods / procedures executed by the system.

7. CASE STUDY

A study is carried out to find out the feasibility of stated architecture. A *server* was configured on Linux platform and made to contain the unified file system for studying architecture.

7.1 Apache Tomcat Container

HTTP Web server is contained in all business logic and workflows. These workflows are required for operationalization of enterprise management. Business logic includes modules such as authentication, authorization, data upload, data download, viewing the organization structure, Viewing the files, copying the files, moving the files and etc.

7.2 MySQL

The access control list for reading and writing the documents are maintained at '*MySQL*' database. Also, separate tables are maintained to keep logging the details of users operations on enterprise files. These logs are useful for audit trails.

7.3 FTP software

FTP is widely used for transferring files between computers over a TCP/IP network. FTP runs on the upper layers of the OSI model and uses the Transport Control Protocol ((TCP)) to transport the transferred data.

The server was configured with FTP for the Merits of

- A. Three-way handshake synchronization
- B. Error-free transmission
- C. Flow--control mechanism

The organization structure or PLC Management Structure (unified file system) is maintained at the intermediate server. The unified file system is ensured to be common to all application servers. The FTP server is mounted on physical storage.

7.4 Physical Storage

Physical storage is comprised of NAS or SAN or hybrid of SAN and NAS.

NAS - Network Attached Storage is file-level computer data storage connected to a computer network providing data access to heterogeneous clients. NAS uses file-based protocols such as NFS on UNIX systems, SMB on MS-Windows systems.

SAN - Storage area network contrasted to NAS – it only provides block-based storage & leaves file

System concerns on the client side. It is an architecture to attach remote computer storage devices (such as disk arrays, tape libraries, and optical jukeboxes) to servers in such a way that the devices appear as locally attached to the operating system.

Hybrid SAN-NAS is a system that contains both block-level protocol and file level protocol through

NFS.

7.5 Mounting

A huge volume of space was made out of hybrid SAN-NAS configuration. Also, it is made accessible

to the application servers of enterprise through mounting the unified file system as defined earlier. This has resulted in storing and retrieving enterprise data on a hybrid System for high availability.

8. INFORMATION LIFE CYCLE

The user's access to documents and their acknowledgment by downloading the documents, Confirms that documents are served at end-users. Moreover, a log is created and maintained in the database to know the application of documents as a function of time, which elucidates the usage of documents from its inception till to-date.

9. OBSERVATIONS

A group of users is transacting their documents through this architecture. It is observed that bulk requests of storing, viewing, retrieving and collaboration requests (on documents) are served among the group online.

The overall performance of intermediate *FTP* server is yielding satisfactory results. The performance of Intermediate FTP Server is tabulated as follows: End users found the system as effective in getting prompt responses to their online requests for shared documents. The throughput is good even for large files.

10. CONCLUSION

Enterprise files management for transacting files, documents among the teams' for sharing, co-operation and collaborations is implemented in IPS ACADEMY as per the stated organization' structure through ftp solution.

It is found that transactional documents of personnel were aligned towards organization requirements and processes. This architecture and ftp solution is viable and feasible to an organization, especially where large quantities of resourceful informational documents are transacted over an intranet.

The architecture is implemented through software called 'Campus Active' operationalized at IPS Academy. The software is witnessed at center-level, various committees, and review boards. The users found the architecture effective to share their documents for collaborative works.

11. REFERENCES

- [1] Jui-Hung Kao, Chien-Yeh Hsu, Yu-Ping Sung, Wei-Pan Liao, DICOM-Based Multi-Center Electronic Medical Records Management System, *International Journal of Bio-Science and Bio-Technology*, Vol. 2, No. 2, June, 2010.
- [2] Loc Helouet and Albert Benveniste, Document based modeling of Web services choreographies using ActiveXML, 2010 IEEE International Conference on Web Services.
- [3] Shaolong Zhang, Application of Document-Oriented NoSQL Database Technology in Web-based Software Project Documents Management System, Third International Conference on Information Science and Technology, March 23-25, 2013; Yangzhou, Jiangsu, China.