International Journal of Advanced Research in

Computer Science Engineering and Information Technology

Volume: 4 Issue: 3 18-Apr-2015,ISSN_NO: 2321-3337

Semantic updated Metadata Organization Paradigm in Next-Generation categorizer

K.Udhaya, M.Uvaneshwari

Student, Dept. of. Information Technology, Veltech University, India.

Assistant Professor, Dept. of. Information Technology, Veltech University, India.

ABSTRACT— *The namespace management is based on hierarchical directory trees. This tree-based namespace scheme is prone to severe performance bottlenecks and often fails to provide real-time response to complex data lookups. This paper proposes a semantic-aware namespace scheme, called sane, which provides dynamic and adaptive namespace management for ultra-large storage systems with billions of files. Associative access on the files is provided by an initial extension to existing tree structured file system protocols, and by the use of these protocols that are designed specifically for content based file system access. Access on the file details such as versions or any other concepts were interpreted as queries applied on our container engine, and thus provides flexible associative access to files. Indexing of key properties of file system objects and indexing/ caching on the file system is one of the fantastic features of our system. The automatic indexing of files and grouped based on relativity is called "semantic" because user programmable nature of the system uses information about the semantics of updated file system objects to extract the properties for indexing. The semantic correlations and file groups identified in sane can also be used to facilitate file perfecting and data de-duplication, among other system-level optimizations.*

Keywords—SANE,Semantic,encryption,encoding,metadata

1. INTRODUCTION

Fast and flexible metadata retrieving is a critical requirement in the next-generation data storage systems serving high-end computing. As the storage capacity is approaching Exabyte and the number of files stored is reaching billions, directory-tree based metadata management widely deployed in conventional file systems can no longer meet the requirements of scalability and functionality. Although existing distributed database systems can work well in some real-world data-intensive applications, they are inefficient in very large-scale file systems due to four main reasons. First, as the storage system is scaling up rapidly, a very large-scale file system, the main concern of this paper, generally consists of thousands of server nodes, contains trillions of files, and reaches Exabyte-data-volume (EB). Unfortunately, existing distributed databases fail to achieve efficient management of petabytes of data and thousands of concurrent requests. In real-world applications, cache-based structures have proven to be very useful in dealing with indexing among massive amounts of data. The main benefit of using semantic correlation is the ability to significantly narrow the search space and improve system performance. Here, we

propose a novel decentralized semantic ware metadata organization, called SmartStore to effectively exploit semantic correlation to enable efficient complex queries for users and to improve system performance in real-world applications. The new design is different from the conventional hierarchical architecture of file systems based on a directory-tree data structure in that it removes the latter's inherent performance bottleneck and thus can avoid its disadvantages in terms of file organization and query efficiency. Additionally and importantly, SmartStore is able to provide the existing services of conventional file systems while supporting new complex query services with high reliability and scalability. Our experimental results based on a SmartStore prototype implementation show that its complex query performance is more than 1,000 times higher and its space overhead is 20 times smaller than current database methods with a very small false probability.On the other hand, the semantic grouping can also improve system scalability and avoid access bottlenecks and single-point failures since it renders the metadata organization fully decentralized whereby most operations, such as insertion/deletion and queries, can be executed within a given group. In order to efficiently support complex queries with unpredictable attributes, we develop an automatic configuration technique to adaptively construct one or more semantic R-trees to improve query accuracy and efficiency. More Rtrees with each being associated with a different combination of multidimensional attributes provide much better query performance, but require more storage space. The automatic configuration technique thus must optimize the trade-off between storage space and query performance. Our basic idea is to configure one or more semantic R-trees to adaptively satisfy complex queries associated with an arbitrary subset of attributes.

2. OBJECTIVE:

The main aim of this project is storing the data in a custom repository in the file system. The data will be compressed and it will be stored in the container and retrieving the data faster the Indexing of data inside the container. Restring data access on the container and providing security and access rights on the documents in the container and data stored in a grouped manner based on the keywords and metadata information about the document.

3. SYSTEM ANALYSIS

3.1. Existing System

The relationship among the files often evolves; Rapport can fast identify their changes to update the namespace by exploiting the particular file semantic of provenance.The basic objective of the system is to create a Smart Store system with the basic CRUD options. In turn, the basic query operation of the users were tracked, top N Queries and Range of the top queries were retrieved. Semantic grouping of file data with a proper indexing is the core concept and the files were stored securely with a clear segregation of Storage Units and Index Units. Systems Petabytes or even Exabyte's. Our scheme logically creates metadata servers (MDS) into a multi-layered query hierarchy and exploits grouped filters Bloom to efficiently route metadata requests to desired MDSs existing solutions that build a separate meta-data database outside of the file system face consistency and management challenges at large-scales. To address these is-sues, we

developed Magellan, a new large-scale file system metadata architecture that enables the file system's metadata to be efficiently and directly searched.

Demerits Of Existing System:

Compression and encryption is not implemented in the existing approach

3.2. Proposed System

The scale of these systems gives rise to many problems: they will be developed and used by many stakeholders across multiple organizations, often with conflicting purposes and needs; Creating a customized container indicates the data will be compressed and it will be stored in the container. To enhance the process of retrieving the data faster the Indexing of data inside the container is achieved. Basic CRUD operations on the container cannot be logged in the existing system. This makes the admin to keep a track on the documents. Restoring data access on the container and providing security and access rights on the documents in the container is not available in the existing system. The first attempt at providing semantic namespace in large-scale file systems, to the best of our knowledge and this is different from existing namespace schemes. Indexing the container didn't specify the type of indexing which indicates it's a proposed one and providing effective search and indexing at the scale of billions of files is not a simple task and rely on relational databases, to provide search.

Merits of Proposed System:

Storage of data in Semantic Container is highly secured and very difficult to hack the data's in the semantic container

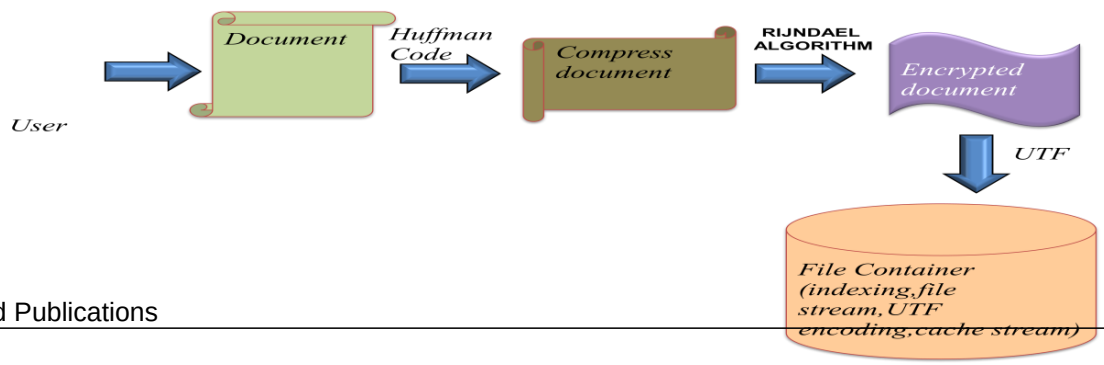
4. ALGORITHM AND DESIGN:

4.1. Rijndael Algorithm and Huffman Encoding

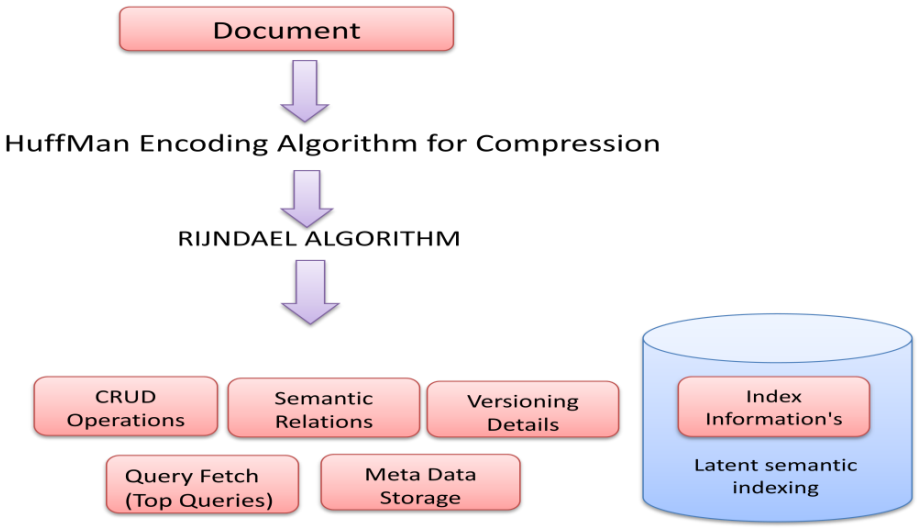
Rijndael algorithm are used for the encryption of the data for additional security purpose. The Rijndael algorithm is a new generation symmetric block cipher that supports key sizes of 128, 192 and 256 bits, with data handled in 128-bit blocks - however, in excess of AES design criteria, the block sizes can mirror those of the keys. Rijndael uses a variable number of rounds, depending on key/block sizes.Huffman encoding is used to compress the data for reduce the storage load.

4.2. Architectural and Functionality Diagram:

Functionality Diagram



Architectural Diagram



5. IMPLEMENTATION

5.1. Authentication Module

Authentication Module describes the Co-Ordination between the users and the system admin.The user is allowed to create his credentials to login into the system. An admin needs to approve the users created and after the login approval the users will be allowed to access the application.During new user creation, the options of Security questions and answers were added in the system.This module provides an interface to the permission system inside the Application.

5.2. Container Creator Module

This module enables the user to create a container in the system. A container format could mould/wrap any kind of data. Though there are some examples of such file formats like jpg, Doc. Most container formats are specialized for specific data requirements. A popular family of containers is found for use with different multimedia file formats. Since video and audio streams can be coded and decoded with our specified algorithms, a container format may be used to provide a single file format to the user.

5.3. Data Uploader Module

Users will have documents that contain confidential or sensitive data such that, the option of encrypt the data is one of an important factor in the field of network security/Data

security.To avoid max storage space, our container provides an additional option of compressing .

5.4. Container Visualize Module

This module provides an option for the user to view the documents stored in the semantic way. In turn, the documents will be stored in a grouped manner based on the keywords and metadata information about the document.The search starts by the user specifying keyword query which is then transformed into the corresponding semantic concept-based query.

5.5. Index Tune Module

To enhance the performance of the document storage, a clear index were placed on hold on the documents. The documents will be hierarchically stored and indexed based on the logically metadata information about the documents.

5.6. Operations Details view

This module provides a detailed view about the operations happening on the system.The terms offered by the autocompletion are concept labels from domain ontologies which had been used for the semantic annotation of the semantic documents from the repository.After the semantic search and the personalized ranking, the recommender shows the list of found document units to the user. For each of the retrieved document units the user can also see additional information that come from its annotation data (e.g., the list of annotation concepts, the number of reuses, the number and list of users, the list of documents in which it appears and the number of versions)

6. CONCLUSION

We presents a new paradigm for organizing file metadata for next-generation file systems, called SmartStore, by exploiting file semantic information to provide efficient and scalable complex queries while enhancing system scalability and functionality. The novelty of SmartStore lies in it matches actual data distribution and physical layout with their logical semantic correlation so that a complex query can be successfully served within one or a small number of storage units. Specifically, a semantic grouping method is proposed to effectively identify files that are correlated in their physical attributes or behavioral attributes. SmartStore can very efficiently support complex queries, which will likely become increasingly important in the next-generation file systems. Our prototype implementation proves that SmartStore is highly scalable, and can be deployed in a large-scale distributed storage system with a large number of storage units.

7. FUTURE ENHANCEMENT:

The future enhancement of this project is to indexing the data in the container and Version management of the data in the container. Easy access of data in the container using credentials check provides additional security for the container. Data are highly securable due to encryption and compression which can't be hacked so easily and will play a vital role in data security.

REFERENCES

[1] R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," J.Vis. Lang. Comput., vol. 20, no. 3, pp. 169–179, 2009.

[2] P. Johnson, B. Tan, and S. Schuckers, "Multimodal fusion vulnerability to non-zero effort (spoof) imposters," in IEEE Int'l Workshop on Inf. Forensics and Security, 2010, pp. 1–5.

[3] P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee, "Polymorphic blending attacks," in Proc. 15th Conf. on USENIX Security Symp. CA, USA: USENIX Association, 2006.

[4] G. L. Wittel and S. F. Wu, "On attacking statistical spam filters," in 1st Conf. on Email and Anti-Spam, CA, USA, 2004.

[5] D. Lowd and C. Meek, "Good word attacks on statistical spam filters," in 2nd Conf. on Email and Anti-Spam, CA, USA, 2005.

[6] C. Upendra ,G.Gopichand, " Survivability and Protection of Nodes and Links from Failures in WDM Mesh Networks",vol: 2-1 (Pages 51-59),), ISR Journal, Available:http://isrjournals.org/archives_abstract.php?id=28&t_n=ijarcseit&d_id=66&dm

[7]K Munivara Prasad,"Comparative Analysis of Performance of Ad-Hoc Wireless Routing Protocols Based On Topology Using Qualnet",vol: 1-1 (Pages 14-19),ISRJournal.Available:http://isrjournals.org/archives_abstract.php?id=28&t_n=ijarcseit&d_id=66&dm

[8] SANE: Semantic-Aware Namespace in Ultra-Large-Scale File Systems Yu Hua, Senior Member, IEEE, Hong Jiang, Senior Member, IEEE, Yifeng Zhu, Member, IEEE, Dan Feng, Member, IEEE, and Lei Xu