International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Rapid Image Search with Tags Using Semantic-Aware Namespace

V. Archana[1], M. Tamil Thendral[2]

[1]ME-CSE, Kingston Engineering College, Vellore, India

[2]Assistant Professor –CSE, Kingston Engineering College, Vellore, India

Abstract: ***Thousands of images are uploaded in social networking sites daily. Technically there is no proper way for searching the files like images in the ever growing social sites. For example a user might have uploaded many images in his profile and when a friend requests a particular image from the uploaded list. it becomes a problematic one, since this requires us to search the complete profile and time to search for particular image is high. An access control mechanism is used in Semantic aware namespace (SANE). Using this mechanism the user can set privileges to each friend as what he/she views. A new compression scheme is used in SANE. Which is used to compress the uploaded images and store as a single file in a ultra file system. The images are stored in a hierarchical structure along with list of tags. The image search can be done with the help of tags. The file can be shared to any user.***

Keywords: Compression method, Tags, SANE- Semantic-aware namespace, Rapid Search, Hierarchical File.

1. Introduction

Storage systems are facing great challenges in handling the files from many data intensive application such as business transactions, scientific computing, and social network webs, mobile applications, information visualization, and cloud computing. Approximately 800 Exabyte of data were created in 2009 alone. According to a recent survey, 1,780 data center in 26 countries. The hierarchical directory tree based metadata management scheme used in almost all file systems today. The most important functions of namespace management are file identification and lookup File system namespace as an information-organizing infrastructure is fundamental to system's quality of service such as performance, scalability, and ease of use.

2. Problem Domain

Network Security relies on layer of protection and consists of multiple components including network monitor and security software addition to hardware and appliance. This all increase security of computer network. Network security covers variety of computer network like both public and private.

3. Problem Definition

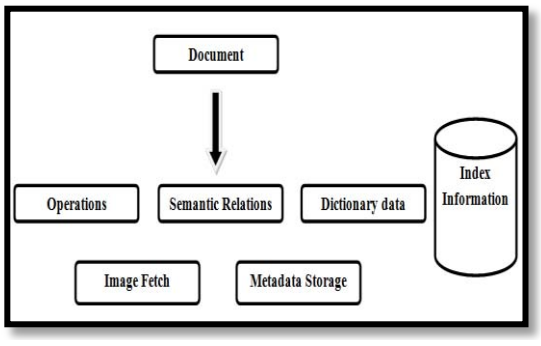
Petabyte or Exabyte-scale data sets and Gigabit data streams are the frontiers of today's file systems. Storage systems are facing great challenges in handling the deluge of data stemming from many data-intensive applications such as business transactions, scientific computing, social network webs, mobile applications, information visualization, and cloud computing. Approximately 800 Exabyte of data were created in 2009 alone. This reflects a reality in which we are generating and storing much more data than ever and this trend continues at an accelerated pace. This data volume explosion has imposed great challenges to storage systems, particularly to the metadata management of file systems. For example, many systems are required to perform hundreds of thousands of metadata operations per second and the performance is severely restricted by the hierarchical directory-tree based metadata management scheme used in almost all file systems today.

4. Objective and Scope of the Paper

The most important functions of namespace management are file identification and lookup. File system namespace as an information-organizing infrastructure is fundamental to system's quality of service such as performance, scalability, and ease of use. Almost all current file systems, unfortunately, are based on hierarchical directory trees. This namespace design has not been changed since it was invented more than 40 years ago. As the data volume and complexity keep increasing rapidly, conventional namespace schemes based on hierarchical directory trees have exposed the many weaknesses.

5. Problem Characterization

Explosive growth in volume and complexity of image data makes it difficult to manage and find images. Ultra large-scale file systems rely on hierarchically structured namespace that leads to severe performance bottlenecks and renders it impossible to support real-time queries on multi-dimensional attributes. SANE supports building an hierarchical structure with the help of tags. The search is made upon tags.



6. Features of Sane

The Semantic-aware namespace has a rich set of features. It includes;

1) Semantic aware namespace paves way for efficient search, automatic organization and pre-fetching.
2) Ability of reduplication and pre-fetching makes the file system a more efficient one.
3) Possibility of automatic organization of files using locality sensitivity hashing (LSH).
4) Provide a namespace that is flat, small and easily manageable.
5) Efficient lookup is archived with the help of multi dimension attributes.

7. Literature Survey

A body of literature has been conducted by several authors and a list of them is given below;

1. Semantic-Aware Metadata Organization Paradigm in Next-Generation File Systems

Data storage systems based on the hierarchical directory-tree organization do not meet the scalability and functionality requirements for exponentially growing data sets and increasingly complex metadata queries in large-scale, Exabyte-level file systems with billions of files. The decentralized semantic-aware metadata organization, called SmartStore, which exploits semantics of files' metadata to judiciously aggregate correlated files into semantic-aware groups by using information retrieval tools. The key idea of SmartStore is to limit the search scope of a complex metadata query to a single or a minimal number of semantically correlated groups and brute-force search in the entire system. The decentralized design of SmartStore can improve system scalability and reduce query latency for complex queries. The main disadvantage is it will take more time to search the entire design.

2. Security Aware Partitioning for efficient file system search

Index partitioning techniques where indexes are broken into multiple distinct sub-indexes a proven way to improve metadata search speeds and scalability for large file systems. A partitioned metadata index can rule out irrelevant files and quickly focus on files that are more likely to match the search criteria .To meet the goals, a new partitioning algorithm, Security Aware Partitioning, that integrates security with the partitioning method to enable efficient and secure file system search. Our results show that Security Aware Partitioning can provide excellent search performance at a low computational cost to build indexes. Based on metrics such as information gain. Also, in a large file system that contains many users. User's search should not include confidential files the user doesn't have permission to view.

3. Just-In-Time Analytics on Large File Systems

As file systems reach the petabytes scale, users and administrators are increasingly interested in acquiring high level analytical information for file management and analysis. Two particularly important tasks are the processing of aggregate and top-k queries which, unfortunately, cannot be quickly answered by hierarchical file systems. Existing pre-processing based solutions, e.g., file system crawling and index building, consume a significant amount of time and space (for generating and maintaining the indexes) which in many cases cannot be justified by the infrequent usage of such solutions. The user interests can often be sufficiently satisfied by approximate - i.e., statistically accurate - answers. We develop Glance, a just-in-time sampling-based system which, after consuming a small number of disk accesses, is capable of producing extremely accurate answers for a broad class of aggregate and top-k queries over a file system without the requirement of any prior knowledge. We use a number of real-world file systems to demonstrate the efficiency, accuracy and scalability of Glance.

4. A New algorithm for Data Compression Optimization

In many free servers, users are permitted to store data in the servers. For example, in Google, each user is given 15GB of free space. When the limit is reached, the user should either delete files or use some compression methods to save disk space. This paper provides a method for compression using bits. Each bit inside a file is taken and the duplicate bits are removed from it. This saves space .The advantage of this method is the compression is loss less compression. There is no data loss while decompression. The disadvantage is the Bit coding is split the zero bits and non-zero bits. This algorithm stores the bits into 2 different files. If one file is deleted, data can't be regained.

5. Real-time Semantic Search using Approximate Methodology for Large-scale Storage System

The challenges of handling the explosive growth in data volume and complexity cause the increasing needs for semantic queries. The semantic queries can be interpreted as the correlation-aware retrieval, while containing approximate results. The advantage is the real time property of FAST enables rapid identification of correlated files. Fast have some disadvantage improved by using semantic aware namespace.

6. Distributed File System of Expandable Metadata Service Derived from HDFS

To store and manage data efficiently is the critical issue which modern information infrastructures confront with. To accommodate the massive scale of data in the Internet environment, most common solutions utilize distributed file systems. However there still exist disadvantages preventing these systems from delivering satisfying performance. In this paper, we present a Name Node cluster file system based on HDFS, which is named Clover. This file system exploits two critical features: an improved 2PC protocol which ensures consistent metadata update on multiple metadata servers and a shared storage pool which provides robust persistent metadata storage and supports the operation of distributed transactions. Clover is compared with HDFS and its key virtues are shown. Further experimental results show our system can achieve better metadata expandability ranging from 10% to 90% by quantized metrics when each extra server is added, while preserving similar I/O performance.

7. A New Algorithm for Data Compression Optimization

In many free servers, users are permitted to store data in the servers. For example, in Google, each user is given 15GB of free space. When the limit is reached, the user should either delete files or use some compression methods to save disk space. This paper provides a method for compression using bits. Each bit inside a file is taken and the duplicate bits are removed from it. This saves space. The advantage of this method is the compression is loss less compression. There is no data loss while decompression. The disadvantage is the JBit coding is split the zero bits and non-zero bits. This algorithm stores the bits into 2 different files. If one file is deleted, data can't be regained.

8. A New Lossless Image Compression Technique Based on Bose, Chandhuri and Hocquengham (BCH) Codes

A bit level compression scheme called BCH. This scheme divides the data into chunks of 7 bits. Then 3 parity bits are added. At the decoder side whenever a chunk is inputted, it removes the parity and checks the correctness of data. This can also recover data error during 1 bit error. The advantage of this paper is data error detection and correction. The disadvantage is redundancy causes extra space management.

9. Avoiding the Disk Bottleneck in the Data Domain Deduplication File System

Data deduplication is an emerging method to store data. Many users will use only a portion of data frequently. This creates a locality of storage. The same data will be used again and again in many areas. Each and every area, the same data is duplicated. These duplicate storage quickly fills up the memory. This paper identifies the duplication areas of a data. Only once data is stored and all the duplicate copy is removed and a reference is made. This advantage is duplicate data is removed. The disadvantage is reference may create cycle during deletion of data.

10. Semantic-sensitive Namespace Management in Large-scale File Systems

Many file systems rely on hierarchical data model where the datas are stored in the tree like format. This paper focus on semantic relationship to build the tree. Other file system uses the date of creation or the file name to build the tree. If two files are in same cluster, then the files are placed in a subtree. Each subtree contains only similar files.

8. Conclusion and Future Enhancement

This paper has focused on describing the basis of SANE-Semantic Aware Namespace. Also a broad list of literature survey has been discussed. We propose to extend this paper by implementing a new namespace management system for exploiting semantic association among the images to create a flat, small and accurate semantic aware namespace for each file. SANE is a precious tool for both system developers and users.

9. Acknowledgement

I would like to take this opportunity to express my profound gratitude and deep regard to my guide, Assistant Professor M.Tamil Thendral CSE, Kingston Engineering College, for his exemplary guidance, valuable feedback and constant encouragement in completing this paper. His valuable suggestions were of immense help in getting this work done. Working under him, was an extremely knowledgeable experience. Also, I would like to extend my sincere gratitude to my parents for their constant support and encouragement in completing this paper.

Reference

[1] A.W. Leung, M. Shao, T. Bisson, S. Pasupathy, and E.L. Miller, "Spyglass: Fast, Scalable Metadata Search for Large-Scale Storage Systems," Proc. Seventh USENIX Conf. File and Storage Technologies (FAST), 2009.

[2] H. Huang, N. Zhang, W. Wang, G. Das, and A. Szalay, "Just-In- Time Analytics on Large File Systems," Proc. Ninth USENIX Conf. File and Storage Technologies (FAST), 2011.

[3] K. Veeraraghavan, J. Flinn, E.B. Nightingale, and B. Noble, "quFiles: The Right File at the Right Time," Proc. USENIX Conf. File and Storage Technologies (FAST), 2010.

[4] Y. Hua, H. Jiang, Y. Zhu, D. Feng, and L. Tian, "SmartStore: A New Metadata Organization Paradigm with Semantic-Awareness for Next-Generation File Systems," Proc. ACM/IEEE Supercomputing Conf. (SC), 2009.

[5] D. Beaver, S. Kumar, H. Li, J. Sobel, and P. Vajgel, "Finding a Needle in Haystack: Facebooks Photo Storage," Proc. Ninth USENIX Conf. Operating Systems Design and Implementation (OSDI), 2010.

[6] PVFS2. Parallel Virtual File System, Version 2, http:// www.pvfs2.org, 2013.

[7] Hadoop Project, http://hadoop.apache.org, 2013.

[8] C. Maltzahn, E. Molina Estolano, A. Khurana, A. J. Nelson, S. A. Brandt, and S. Weil, "Ceph as a Scalable Alternative to the Hadoop Distributed File System," login: The USENIX Magazine, vol. 35, pp. 38-49, Aug. 2010.

[9] "Symantec. 2010 State of the Data Center Global Data.," http:// www.symantec.com/content/en/us/about/media/pdfs/ Symantec_DataCenter10_Report _Global.pdf, Jan. 2010., 2013.

[10] I. Gorton, P. Greenfield, A. Szalay, and R. Williams, "Data- Intensive Computing in the 21st Century," Computer, vol. 41, no. 4, pp. 30-32, 2008.

Author Profile

V.Archana received the B.Tech (IT) degree in 2014 from Priyadarshini Engineering College, India. She is a post graduate student in the Computer Science Department, Kingston Engineering College, India. Her research interests are Network Security.

M.Tamil Thendral, Assistant professor, Department of computer science, Kingston Engineering College. He received his B.Tech (IT) degree in 2005 from SKP Engineering College, India and He then complete his M.E degree in 2011 from Madha Engineering college, India, respectively. His area of interest is Network Security. He has published two papers and has attended 2 conferences.