RESEARCH ARTICLE OPEN ACCESS

A Novel Way of Deduplication Approach for Cloud Backup Services Using Block Index Caching Technique

G.Gayathri[1], A.Soundarrajan[2]

[1]Assistant professor Department of Computer Science Ponnaiyah Ramajayam Institute of Science & Technology (PRIST) Vallam, Thanjavur.

[2] Master of computer application Department of Computer Science Ponnaiyah Ramajayam Institute of Science & Technology (PRIST) Vallam, Thanjavur.

-------------------------------------*************************-------------------------------

Abstract:

Data Deduplication describes approach that reduces the storage capacity needed to store data or the data has to be transfer on the network. Cloud storage has received increasing attention from industry as it offers infinite storage resources that are available on demand. Source Deduplication is useful in cloud backup that saves network bandwidth and reduces network space Deduplication is the process by breaking up an incoming stream into relatively large segments and deduplicating each segment against only a few of the most similar previous segments. To identify similar segments use block index technique The problem is that these schemes traditionally require a full chunk index, which indexes every chunk, in order to determine which chunks have already been stored unfortunately, it is impractical to keep such an index in RAM and a disk based index with one seek per incoming chunk is far too slow. It describes application based deduplication approach and indexing scheme contains block that preserved caching which maintains the locality of the fingerprint of duplicate content to achieve high hit ratio and to overcome the lookup performance and reduced cost for cloud backup services and increase dedulpication efficiency.

Keywords — **Data mining, Duplication, Novel.**

-------------------------------------*************************-------------------------------

INTRODUCTION

The explosive growth of the digital data, data deduplication has gained increasing attention for its storage efficiency in backup storage systems. Today, in the context of user data sharing platforms the challenges for large scale, highly redundant internet data storage is high. Due to this redundancy storage cost is reduces. Storage for this increasingly centralized Web data can be getting by its de duplication. Data deduplication describes a class of approaches that reduce the storage capacity needed to store data or the amount of data that has to be transferred over a network. These approaches detect coarse-grained redundancies within a data set, e.g. a file system; Data deduplication not only reduces the storage space requirements by eliminating redundant data but also minimizes the network transmis duplicate data in the network systems. It splits files into multiple that are each uniquely identified by a hash signature called a fingerprint. It removes duplicate chunks by checking their fingerprints, which avoids byte by byte comparisons. Mainly data deduplication focused on different terms like throughput, advance chunking schemes, other type of storage capacity and clustering method and system workload. As data passes through a cache on its way to or from the storage/processing/networking device, some of the data is selectively stored in the cache. When an application or process later accesses data stored in the cache that request can be served faster from the cache than from the slower device. The more requests that can be served from cache, the faster is the overall system performance. There is a trade-off in cache cost and performance.

This project contain following modules

ADMIN LOGIN

To make the system more secure, the admin has an unique id and password. If an unauthorized person tries to login, the system will omit them by giving an message as ”Login Failed”.

VIEW UPLOAD FILES

` In this module admin can view uploaded files and make a decision to accept a file or not. Only accepted files can be uploaded in the server. Files doesn’t accept by admin cannot be uploaded in server. These files can be processed further by users.

VIEW USER DETAILS

Admin can view the all user details using this module. This module contains information about user like their name, mailid contact number, etc... This information are stored in admin database and used further by admin.

REGISTRATION

In this module the user enters their details including the id, name, Address, DOB, contact number and other details. User must register the registration form. Otherwise the user cannot login into the system.

3.1 EXISTING SYSTEM

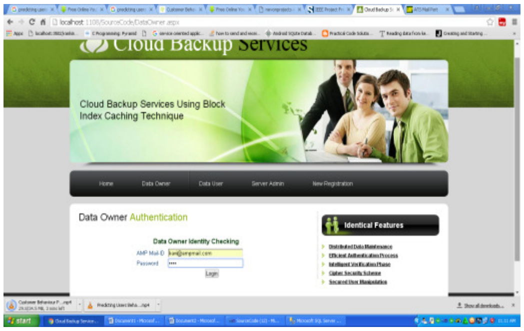
The increasing popularity of the cloud backup services has a great attention to the industry. cloud backup services has become a cost effective choice for data security of personal cloud environment and also for improving deduplication efficiency Existing method that are introduces for deduplication technology for backup service only focus on removing redundant data from transmission during backup operation to reduce backup time and there is no attention in restore time this paper introduces CAB Architecture that captures the casual relationship among dataset used in backup and restore operation. It is integrated into existing backup system.

3.2 PROPOSED SYSTEM

Data Deduplication has emerged as an attractive lossless compression technology that has been employed in various network efficient and storage optimization systems so that it proposed a new approach for Application based Deduplication for cloud backup services using Block Locality Caching contain backup data as shown in figure 1.From backup files as input files having redundant or copied data files that want to deduplicate and for improve storage efficiency this system uses different chunking method base on file type. Files are filtered because of containing tiny files having less than 10 KB Size. So that after making group of files in MB Get filter and then different chunking strategy is used in this system. Chunk with file type are then deduplicate by calculating hash value name as fingerprint using different hash algorithm this fingerprint is then stored in container of cloud having new entries. Fingerprint which it stored for finding duplicate copies are get indexing by using block locality method index entries are name by their block number and chunk id. All this information is stored in block and blocks are stores in cache. If we search fingerprint in block and a match is found, the block for the file containing that chunk of fingerprint is updated and point to the location of the existing chunk of fingerprint. If there is no match, then new fingerprint is stored based on the container management in the cloud, the metadata for the associated file is updated to point to it and a new entry is added into the application aware index to index the new chunk of fingerprint. Due to this performance of system increases and system overload is reduced.

This Architecture remove the redundant data from transmission not only backup operation but also restore operation and improve the backup and restore performance and also reduce both the reduction ratio.

If there is no match, then new fingerprint is stored based on the container management in the cloud, the metadata for the associated file is updated to point to it and a new entry is added into the application aware index to index the new chunk of fingerprint.


localhost:1108/SourceCode/DataOwner.aspx
Cloud Backup Services
Cloud Backup Services Using Block Index Caching Technique
Home
Data Owner
Data User
Server Admin
New Registration
Identical Features
Distributed Data Maintenance
Efficient Authentication Process
Intelligent Verification Phase
Cipher Security Scheme
Secured User Manipulation
Login
start

Data Owner Authentication

Data Owner Identity Checking

AMP Mail-ID: kani@ampmail.com

Password: ****

CONCLUSION

For cloud storage, using deduplication techniques and their performance and suggests a variation in the index of block level deduplication and improving backup performance and Reduce the system overhead, improve the data transfer efficiency on cloud is essential so that, it presented approach on application based deduplication and indexing scheme that preserved caching which maintains the locality of the fingerprint of duplicate content to achieve high hit ratio with the help of the hashing algorithm and improve the cloud backup performance. This project proposed a novel variation in the deduplication technique and showed that this achieves better performance. Currently, optimized cloud storage has been tested only for text files and pdf files .In future, it can be further extended to use files of other type i.e. video and audio files.

Reference:

1.Steven Holzner, "BLACK BOOK ASP. NET" Published By DreaTech (New Edition)

2.Shaum Tec Media, "TEACH YOURSELF ASP.NET IN 21 DAYS" Published By G.C.JAIN (2000)

3.Patrick Dalton, " MICROSOFT ADO.NET " Published By (2007) Morgan Skinn

4. Shaum Series, "TEACH YOURSELF SQL-SERVER IN 21 DAYS"Published By G.C.JAIN (2005)

5. C# 2012 Programming, Covers .Net 4.5, Black Book Paperback – 7 Nov 2013 by Kogent Learning Solutions Inc. (Author)

6. Professional C# 2005 Paperback – 17 Nov 2005 by Christian Nagel (Author), Bill Evejen (Author), Jay Glynn (Author),