Cloud Storage Optimization using Compression Technique

G. Uma[1*], L. Jayasimman[2]

[1*]Srimad Andavan Arts and Science College (Autonomous), Trichy.
[2] Srimad Andavan Arts and Science College (Autonomous), Trichy

Available online at: www.ijcseonline.org

Abstract— Data Storage in the new era of storing data anywhere and everywhere has paved way to find use and access the available data in an efficient manner hence it is no longer an warehousing issue. Moreover for various cloud applications invariably providing sufficient space for the available data in an user friendly manner is a very important task of Cloud Storage. However studies are on, on the issue of Data Storage as to make it more competitive and user friendly nowadays.

*Keywords—*Compression, Decompression, Cloud, Storage, Techniques

Introduction

A collective endeavour in optimal utilization of various process, frameworks , technologies   pave way for the efficient storage of infrastructure and resources. In a wider perspective,   it is a concept   which works with all technological platforms and management layers of storage management that ensures the cost effectiveness of the entire system. As the storage data size increases, it should be noted that the storage is optimized so that the link doesn't become weak in the process.

Cloud Storage Types

Consumer centered cloud storage services focus on various services such as back up files online, enable sync files across multiple devices,   enable sharing of files, photos, music … across various devices thro cloud environment.

Yet another type of cloud storage services are those focused towards the needs of businesses, various **Public cloud** storage services shared and are sold to various multiple customers in the open market[2].

a. **Private cloud** storage service is developed and maintained for the exclusive use of a sole single organization. Obviously organizations   develop a private cloud due to high concern for data security in cloud computing.

b. **Hybrid cloud** storage refers to the use of a combination of private cloud storage and public cloud storage. An organization on any business may prefer to store their highly sensitive data in a private cloud on the other   hand processing their less sensitive data into low-cost public cloud storage.
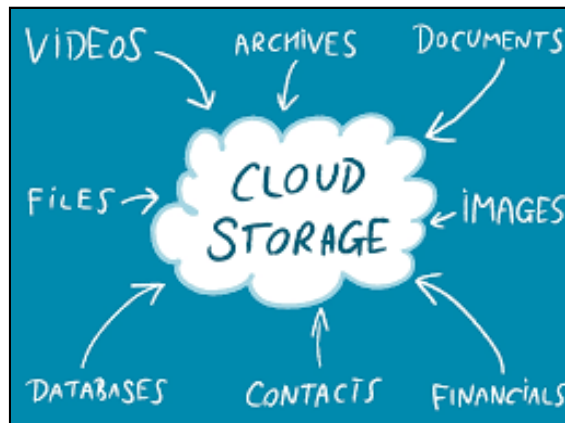
Features of cloud-Optimized storage
Optimal utilization of cloud storage be broadly classified into three methods viz Optimizing data, Optimizing database, Implementing software-defined storage.

Data optimization
Recently   Deduplication,   Compression   and thin provisioning are utilized in data optimization techniques. It can be referred to be the very significant development in the handling of space management process. It helps in storage of a voluminous data in in a comparatively smaller space.

Deduplication
With deduplication (also known as single-instance storage), you save space by eliminating duplicate copies of repeating data[5]. Here we are continuously processing the data for uniqueness in the pattern with reference to the byte pattern. Such unique patterns are identified for replacement with a reference   in the new version. Maximum replacement of

the repeated patterns and its replacement helps in reduction of used space.

Compression

Data compression (also known as intelligent compression) is a familiar term that simply means encoding information in fewer bits than the original[6].Normally we follow two methods of compression techniques viz lossy and lossless. In lossy compression unnecessary information is identified and removed wherein lossless refers to the statistical redundancy, here the bits transferring data and not the data as such are removed.

Cloud computing types are service deployment models which let you choose the level of control over your information and types of services you need to provide. There are three main types of cloud computing services, sometimes called the cloud computing stack because they build on top of one another.

Types of cloud computing Services

The first cloud computing type is infrastructure-as-a-service (IaaS), which is used for Internet-based access to storage and computing power. The most basic category of cloud computing types, IaaS lets you rent IT infrastructure - servers and virtual machines, storage, networks and operating systems - from a cloud provider on a pay-as-you-go basis.

The second cloud computing type is platform-as-a-service (PaaS) which gives developers the tools to build and host web applications. PaaS is designed to give users access to the components they require to quickly develop and operate web or mobile applications over the Internet, without worrying about setting up or managing the underlying infrastructure of servers, storage, networks and databases.

The third cloud computing type is software-as-a-service (SaaS) which is used for web-based applications. SaaS is a method for delivering software applications over the Internet where cloud providers host and manage the software applications making it easier to have the same application on all of your devices at once by accessing it in the cloud.

Literature Review

1 Huffman Coding

Huffman coding refers to a lossless data compression algorithm. Length codes allotted to characters we input and the same is determined by the frequency in which the characters are utilized. Naturally in the process of reduction of space, the smallest code is assigned to the character that is most frequently used and vice versa [7]. The process is completed by a specific method in choosing the representation for every symbol, as such it results in a prefix-free code which expresses most common characters

using shorter strings of bits than that are used for less common source symbols.

2. LZ77

LZ77 algorithms target achieving compression of data by a methd of replacing most repeated occurrences of data with references to a single copy of data existing already in the uncompressed data stream[8]. Thus a match is encoded with a pair of numbers referred as a *length-distance pair*, which is equivalent to the statement "each of the next *length* characters is equal to the characters exactly *distance* characters behind it in the uncompressed stream".

3.LZ78

LZ78 algorithms achieve data compression by replacing repeated occurrences of data with references to a dictionary that is built based on the input data stream. Each dictionary entry is of the form `dictionary[...] = {index, character}`, where *index* is the index to a previous dictionary entry, and character is appended to the string represented by `dictionary[index]`

4.LZW

LZW compression is the compression of a file into a smaller file using a table-based lookup algorithm invented by Abraham Lempel, Jacob Ziv, and Terry Welch. Two commonly-used file formats in which LZV compression is used are the GIF image format served from Web sites and the TIFF image format. LZW compression is also suitable for compressing text files.

A particular LZW compression algorithm takes each input sequence of bits of a given length (for example, 12 bits) and creates an entry in a table (sometimes called a "dictionary" or "codebook") for that particular bit pattern, consisting of the pattern itself and a shorter code[9]. As input is read, any pattern that has been read before results in the substitution of the shorter code, effectively compressing the total amount of input to something smaller.

Algorithm - LZW_COMPRESS

1. STRING = get input character

2. WHILE there are still input characters DO

3. CHAR = get input character

4. IF STRING+CHAR is in the string table then

5. STRING = STRING + character

6. ELSE

7. output the code for STRING

8. add STRING+CHAR to the string table

9. STRING = CHAR

10.END of IF

11.END of WHILE

12.Output the code for STRING.

On applying this algorithm on a text file, the size of the file had reduced to 1.6kb from 2.30kb. The application of the algorithm is proposed to be executed on other types of files and to be checked for reduction in size of the resultant file.

We can also use this same algorithm to compress pdf files xls files and video files also.

Similarly we can also decompress the data using decompress algorithm. We can get the actual data after decompression.

Conclusion

We can successfully compress data by implementing the LZW method of compression so as to reduce space in the cloud. It can also be ensured that the transfer time for the data and the storage space for the data could be considerably reduced in the process.This procedure can also be implemented to various files like pdf, xls,doc and etc.

References

[1] A.P. Mohod, A. Pankar, A Mohite, A. Dondalkar, A. Kadhe, Manish Naik, "*Digital Health Record and Prescription*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.1, pp.60-62, 2018.

[2] Anitha H M, P. Jayarekha , "*Security Challenges of Virtualization in Cloud Environment*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.1, pp.37-43, 2018.

[3] Sakshi kathuria, "*A Survey on Security Provided by Multi-Clouds in Cloud Computing*", International Journal of Scientific Research in Network Security and Communication, Vol.6, Issue.1, pp.23-27, 2018.

[4] Amazon Inc. Amazon Web Services EC2 site[URL].

[5] Naresh Kumar, Ishu Devi, "*Various Chunking and Deduplication Techniques in Big Data*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.129-131, 2017

[6] Z. Li, X. Zhang, and Q. He, "*Analysis of the key technology on cloud storag*e," in International Conference on Future Information Technology and Management Engineering, 2010, pp. 427-428.

[7] B. Sowmya, K. Madhavi, "*Secured Cloud Storage via Attribute-based Encryption*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.7, pp.96-100, 2017.

[8] L. B. Costa and M. Ripeanu, "*Towards automating the configuration of a distributed storage system*," in Proc. 11th IEEE/ACM Int. Conf. Grid Comput., Oct. 2010, pp. 201–208.

[9] J. Dinerstein, S. Dinerstein, P. K. Egbert, and S. W. Clyde, "*Learning based fusion for data deduplication*," in Proc. 7th Int. Conf. Mach. Learning Appl., Dec. 2008, pp. 66–71.