International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 23 Issue 8 –NOVEMBER 2016.

A REVIEW FOR VARIOUS DATA COMPRESSION ALGORITHNS AND METHODS IN CLOUD STORAGE

V.Sherine Vargheese[#1] and Dr.D.Maruthanayagam[*2]

[#] *Research Scholar, Periyar University, Salem, Tamilnadu.India.*

[*] *Head cum Assistant Professor, PG and Research Department of Computer Science, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri,Tamilnadu,India.*

***Abstract*— Cloud computing is a type of internet based computing, where different resources such as servers, data storage, and applications are delivered to the organizations through the internet as pay per use system. Cloud Computing reduces the cost. It is a location independent technology which leads access the services from anywhere anytime. The high demand for data processing and leads to high computational requirement which is usually not available at the user's end Compression algorithms reduce the redundancy in data representation thus increasing effective data density. Data compression is a very useful technique that helps in reducing the size of text data and storing the same amount of data in relatively fewer bits resulting in reducing the data storage space, resource usage or transmission capacity. In this paper, we discussed about some existing compression techniques for cloud storage problems employed in cloud computing.**

***Index Terms*— Cloud Storage, Compression algorithm, Bandwidth, Image Compression, Wavelet and Decompression.**

I. INTRODUCTION

Cloud storage is a new concept come into being simultaneously with cloud computing, which generally contains two meanings: It's the storage part of the cloud computing, virtualized and high scalable storage resource pool. Cloud users access to cloud computing services based on the cloud storage resources pool, but not all storage part can be separated in cloud computing. Cloud storage means that storage can be provided as a service over the network to the user. User can use storage pass through a number of ways, and pay by the use of time, space or a combination of both. Obviously, such statement is not tightly defined this new concept of cloud storage. In addition, the relationship between the concepts of Cloud Storage, Storage Cloud [1][2], Storage as a Service, Cloud-Based Storage should be cleared.

Cloud storage is divided into public cloud storage, private cloud storage and hybrid cloud storage. Public cloud storage is designed specifically for large-scale, multi-user cloud storage. All components are built on a shared infrastructure, and public storage devices were logical partitioned through virtualization technology, data access, data management technology, according users need. Also known as internal cloud storage, private cloud storage is designed for a specific user. Unlike the public cloud storage, private cloud storage running on a dedicated storage devices in the data center so as to meet safety and performance requirements. However, it's obvious disadvantage is the relatively poor scalability. The hybrid cloud storage is the cloud storage to integrate public cloud storage and private cloud storage. Generally, hybrid cloud storage was case-based in private cloud storage, supplemented by public cloud storage.

Compression techniques are most important concept in cloud storage environments. This technique is used to reduce the size of data by removing repetitive bits of data or duplicates of data. This process helps save bandwidth and space during data transmission and storage. While the technique does not resolve bandwidth bottleneck issues, it certainly helps in the medium term to conserve bandwidth for enhanced network performance. Compression can help individual organizations handle their bandwidth requirements by compressing data and regulating flows of information. It is extremely useful in communication as it improves performance.

There are several kinds of compression techniques in use. However, only a few compression techniques are standardized, and popularly used to compress data transmitted over networks. The CCITT has defined a standard for data compression and for data communications that use modems. Compression is also widely used in backup and recovery and most Cloud based storage applications. Some compression technologies used by Cloud vendors are discussed below:

File Compression: As the name suggests, this compression reduces the size of the file. The approach is pretty straightforward. The file is compressed as it flows into the storage disk. However, file system compression technologies are still evolving and existing systems create performance issues when the files are written once and accessed occasionally. It is recommended that this system should be deployed selectively on files that are large and infrequently used.

Storage Array Compression: This type of compression is

very difficult to implement at block level, and is used to compress file systems in a storage array. The compression can be inline or post processed. Inline compression means data is compressed at source. Post processed means data is compressed after it reaches the destination. Bandwidth saving happens only with inline compression.

NAS Storage Compression: This is an inline compression system that uses compression appliances to compress files. The device will support both CIFS and NFS, and will sit between application servers and NAS arrays, performing the role of a front-end for the NAS.

Data De-duplication: This is considered to be the "holy grail" of compression. Duplicate bits of data are identified at the byte or file level and eliminated from the backup to reduce on bandwidth and storage. De-duplication can be performed at source or destination. Source de-duplication saves bandwidth.

II. LITERATURE REVIEW

M. Vasavi, et al. [3] in this research paper presents architecture with objective to access easily the large files in the cloud using light-weight metadata. The size of metadata for every file is the same. It has been found that proposed system requires less uploading and downloading time as compared to BFC, Drop box and normal technique. Also size of metadata is less as compared to other strategies. System also detects duplication of files using SHA value. Overall performance is improved using proposed system. In this research tested this system for different document and image files like doc, txt, pdf, jpeg, audio, video files like mp4 etc. Store the big file in chunks format and detect the duplicate files as well as increase the storage space of cloud with file compression.

Bikramjeet Singh Bumrah, et al. [4] in this paper reduces the storage on the cloud loss less compression has been implemented that reduces the file size. The file size has been get reduced and stored on the cloud. After this process when ever user want to access that data then the de-compression have to be done to decrypt the data. After decryption the logical and arithmetical formulas have to be implemented on the user side. These formulas have been implemented on user side. After this process the user can get the plain text and use this for different operations. The proposed work provides better storage management of the data on the cloud environment. It reduces the storage capacity of the data and can be easily stored on data.

Amanjot Kaur Sandhu, et al. [5] proposed Genetically Modified Compression Algorithm showed the better results by keep both the PSNR and compression size in balance. The execution time of the algorithm is also manageable to run it over the cloud to compress images and hence saves the storage space. The quality of image is preserved after compression .The present research work is done on different type of image data so in future work can be extended to audio and video data.

Gabriel Alatorre, et al. [6] in this paper investigated the feasibility of deploying compression techniques in large scale data centers. This research showed that poorly selecting data volumes to compress could result in significantly increasing response time latency while yielding marginal capacity savings. Moreover, limited computation resources allow only the compression of a limited number of volumes so selecting the right volumes to compress is of great importance from a data center administrator's standpoint. In addition, the transformation of volumes was a process that needed to be executed carefully to not lose data, waste resources, or disrupt actively running applications. To address these challenges, the researcher developed RTCA which is a tool for deploying compression within a data center that used a host heuristic based compression prediction algorithm to compress a near optimal set of volumes.

Ravichandran et al. [7] argued that healthcare researchers are moving towards their efforts to the cloud platform in order to process, store, exchange and use a large amount of medical image data which are generated and acquired through various advance medical modalities. One of the challenges that arise in hospitals and medical organizations is the difficulty of transmitting such a large volume of medical images with relatively limited bandwidth. Image compression techniques have been incorporated in order to reduce the bandwidth requirement and economically transfer of medical images for primary diagnosis.

Muneshwara M.S, et al. [8] this paper work studies the problem of ensuring the integrity of data storage in Cloud Computing. The challenges associated with the retrieval of data from cloud in an appropriate manner. As the data gets compressed, it leads to a more optimized way of retrieving data from cloud. The use of compression in cloud computing leads to effective use of storage disks and bandwidth. This work enables the user to fine-tune the trade-off between storage costs, computation time and bandwidth costs. Different computations of characters can be represented by fewer numbers of bits in compression, which is an efficient way of retrieving data in the cloud environment. The technique of compression, AES, RSA and KP-ABE algorithm are used to monitor the tasks simultaneously. Unlike most prior works, the new scheme further supports secure and efficient dynamic operations on data stored in the cloud, including: data compression, data encryption, data update, delete and append. Unique access structure enhances security. Implementation simplifies the development efforts and provides direct support for security where data can be managed effectively in cloud. Extensive security and performance analysis shows that the proposed scheme is highly efficient and provably secure.

Chi Yang et al [9] discussed the problem of managing large size data on cloud such as big data, multimedia data etc. that raise hard challenges like memory bottlenecks and storage space. To overcome these issues large size data is being compressed by using spatiotemporal features on cloud and gives result in terms of data size and fidelity loss.

S. Girija Lakshmi1, et al. [10] in this research Haar Wavelet Transform is used for compressing the image in smaller size and the compression ratio is as small as the other techniques. This compression is deployed in cloud environment as software where it can be used and shared among the multiple users. Image compression is performed using haar wavelet transform which reduces the

computational cost and efficiently performs the compression of the image. Memory space is reduced since the compression ratio is greatly reduced by proposed technique.

Sri kanth.S, et al. [11] in this paper using different wavelet families and after that compares the PSNRs and bit rates of these families. These algorithms were tested on different images, and it is seen that the results obtained by these algorithms have good quality and it provides high compression ratio as compared to the previous exist lossless image compression technique. In which we use the different embedded Wavelet based image coding with Huffman-encoder for further compression. In this paper they implemented the SPIHT and EZW algorithms with Huffman encoding.

Mukherjee , Tilak et al [12] proposed better lossless compression approach using wavelets in which both approximation and detailed contents of the image are decomposed and yields better compression parameters. Using transform methods the wavelet coefficients are rounded off to integer values to finite precision and lifting scheme is used obtaining high PSNR and compression ratio for real time image compression systems where speed is a deciding factor.

Sherif E. Hussein, et al. [13] in this research, a hybrid compression technique was presented that selectively compresses significant regions lossless compression using Huffman compression. The remaining of the image was compressed using lossy JPEG at a compression level of 50. The significant regions were manually selected with a radiologist. The compression results showed compression ratios in the range from 4 to 10 which are higher than those reported in literature. The compression algorithm was implemented in a web application front end to compress/decompress the medical images before to be sent to and after to be received from the back end. The back end utilized Scalia which is a system that continuously optimizes the placement of data stored at multiple cloud providers, based on their access statistics. Scalia mediates data placement across multiple public cloud providers and private cloud resources. It helps the data owners to avoid vendor lock in and satisfy certain availability and durability constraints in a cost-effective way.

Kagadis et al. [14] has observed that cloud computing has been introduced only recently but is already one of the major topics of discussion in research and clinical settings. They conclude that healthcare researchers are moving their efforts to the cloud, because they need adequate resources to process, store, exchange, and use large quantities of medical data. Among the potential driving forces for the increased use of cloud computing in medical imaging are raw data management and image processing and sharing demands, all of which require high-capacity data storage and computing.

Suhasini Kalki1, et al. [15] The sharing of medical information resources is a key factor playing a fundamental role towards the effective and rapid medical diagnosis, especially when doctors are faced with new clinical cases or whether they should take into account diseases that are not properly part of their domain of competence or their own specialization. In this work we presented a Virtual infrastructure-less Cloud solution for secure management of 3D medical images, which operates in an almost completely transparent manner, regardless of computational and networking capabilities which users can avail in any given moment.

K. Govinda et al [16] discussed the problem with Simple storage service (S3) in july 2008 suffers from great loss because it is failed to route user's request to appropriate physical storage and proposed lossless LZW compression for optimizing the cloud storage that achieves 50 percent of compression.

Wen Sun, et al. [17] proposed a value of interest VOI-based compression scheme for medical images with extended bit depth, which considered the user side display during the compression process. Given the recommended VOI mapping parameters in the display, the original image was decomposed into a VOI layer and an enhancement layer. The VOI layer image only contained information of the VOI pixels to reduce the transmission delay of these pixels. The enhancement layer image was encoded using Gray-Golomb codes to support VOI scalability. They applied pixel domain bit plane coding for the both layers, which enabled pixel level distortion estimation for arbitrary truncated bit-stream. Compared to JPEG2000, there scheme could significantly reduce the bandwidth cost and transmission delay, with very slight overall bit-rate increase.

Yao-tien Chen, et al. [18] Proposed a wavelet-based compression scheme with an adaptive prediction for medical images. They proposed as well the adaptive prediction approach to overcome the multicollinearity problem and employ three prediction equations for different wavelet sub bands to achieve a more accurate prediction. The proposed WCAP method and four famous lossless compression methods: SPIHT, JPEG2000, CALIC, and SSM, were experimentally compared. Their experimental results showed that the proposed WCAP method almost achieved the highest compression rates for CT, MRI, and ultrasound images; moreover, WCAP had individually improved 20.7, 12.6, 8.02, and 3.47% compression rates with respect to SPIHT, JPEG2000, CALIC, and SSM methods in average.

Zukoski, et al. [19] proposed an expansion to the residual approach is still experimental, this paper has shown that the overall approach of clinically relevant regions has clearly demonstrated advantages over both traditional lossless compression and simple lossy compression, and that the residual approach holds some promise. A larger scale study in conjunction with Penn State University’s Hershey Medical Center is ongoing. This study will include a qualitative component to examine the effects of the hybrid compression methodology in a clinical setting.

Armen Dzhagaryan, et al. [20] Based on this research result analysis this author was developed the following guidelines for transferring mHealth data between personal devices and the cloud. When uploading the mHealth data, the low-complexity utilities such as *lzop* should be used for data from inertial sensors, accelerometers, as well as log files. Certain types of mHealth data such as time intervals between successive breaths and R-peaks should be uploaded uncompressed. Downloading compressed files with decompression almost always helps reduce latency and

increase energy efficiency. The compression utilities such as *gzip* and *xz* outperform other utilities as they combine good compression ratios with fast decompression. The researcher also demonstrated that the type of monitored activity impacts the effectiveness of the compression utilities. These findings may guide mHealth application developers in developing frameworks for optimizing data transfers between the personal devices and the cloud with potential to improve user experience, reduce energy requirement, and reduce required storage.

Niles B, et al. [21] this research presents various techniques of image compression. In which we comparing the performance of compression technique is difficult unless identical data sets and performance measures are used. It is found that the lossless image compression technique is most effective over the lossy compression technique. In which we analyze the different type of existing compression methods. In present time some other techniques are added with basic method. In some area the neural network genetic algorithm are used for image compression.

III. CONCLUSION

The use of compression techniques are in the cloud computing leads to effective use of storage disks and bandwidth. This work enables the user to fine-tune the trade-off between storage costs, computation time and bandwidth costs. Different computations of characters can be represented by fewer numbers of bits in compression, which is an efficient way of retrieving data in the cloud environment. In this paper, presents the review on various data compression algorithms as well as clear analysis for literature survey of various latest papers belongs to cloud storage compression techniques.

REFERENCES

[1] Robert L.Grossman, Yunhong Gu, Michael Sabala,Wanzhi Zhang. Compute and storage clouds using wide area high performance networks. Future Generation Computer Systems, 2009,25(2):179-183.
[2] Brandon Rich, Douglas Thain. DataLab: Transactional data-parallel computing on an active storage cloud. In: Proc. of the 17th International Symposium on High Performance Distributed Computing,2008, 233-234.
[3] M. VASAVI, J. NARASIMHA RAO, IMPROVE THE PERFORMANCE OF BIG FILE CLOUD STORAGE WITH FILE COMPRESSION International journal of research, Volume 5, Issue 4 SEP 2016
[4] Bikramjeet Singh Bumrah, Gurbind Kaur, Security and Storage Management of Data on Cloud using Fully Homomorphic Encryption and Adaptive Compression Approach International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Volume 5 Issue 6, June 2016.
[5] Amanjot Kaur Sandhu, Er. Anupama Kaur, Genetically Modified Compression Approach for Multimedia Data on cloud storage, IJCSC, Vol 6 • Number 2 April - Sep 2015 pp. 258-263
[6] Gabriel Alatorre, Nagapramod Mandagere, Yang Song and Heiko Ludwig, On Selective Compression of Primary Data IBM Almaden Research Center, 978-3-901882-77-7 c 2015 IFIP
[7] Ravichandran D, Ashwin Dhivakar M R and Dakha V, " Medical Image Compression based on Discrete Wavelet Transform (DWT) " in the International Conference on Technical Advancements in Computer Science and Engineering (ICTACSE- 2015), Hyderabad, pp. 1-7, 2015
[8] Muneshwara M.S , Swetha M.S , Dr. Anil G.N, A Smarter Way of Securing and Managing Data for Cloud Storage Applications Using High Throughput Compression in the Cloud Environment , International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 9, September 2014.
[9] Yang, Chi, et al. "A spatiotemporal compression based approach for efficient big data processing on Cloud." Journal of Computer and System Sciences 80.8 (2014): 1563-1583.
[10] S. Girija Lakshmi1, N.R. Gayathiri, A FRAMEWORK OR HOSTING IMAGE COMPRESSION IN CLOUD IJCSMC, Vol. 3, Issue. 3, March 2014, pg.845 – 848.020217192
[11] S.Srikanth, Sukadev Meher,"Compression Efficiency for Combining Different Embedded Image Compression Techniques with Huffman Encoding" IEEE, 2013.
[12] Mukherjee, Tilak, and M. Koteswara Rao. "Efficient Performance of Lifting Scheme Along With Integer Wavelet Transform In Image Compression."International Journal of Engineering Research and Applications (IJERA) 3.4 (2013): 1950-1953.
[13] Sherif E. Hussein, Sherif M. Badr, Healthcare Cloud Integration using Distributed Cloud Storage and Hybrid Image Compression, International Journal of Computer Applications (0975 – 8887) ,Volume 80 – No3, October 2013
[14] Kagadis, G C, Kloukinas C, Moore K, and Hendee, W R, "Cloud Computing in Medical Imaging," American Association of Physicists in Medicine, vol. 40, no.7, pp.1-11, July 2013
[15] Suhasini Kalki1, Jayashree Agarkhed2 , Provisioning of Compression and Secure Management Services for Healthcare Data in Cloud Computing International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14.
[16] Govinda, K., and Yuvaraj Kumar. "Storage Optimization in Cloud Environment using Compression Algorithm."
[17] Wen Sun , Yan Lu, Feng Wu, Shipeng Li "Level embedded medical image compression based on value of interest,"Proceedings of the 16th IEEE international conference on Image processing Pages 1749-1752, NJ, USA, 2009.
[18] Yao-Tien Chen, Din-Chang Tseng "Wavelet-based medical image compression with adaptive prediction," Computerized Medical Imaging and Graphics, Pages 1–8, Volume 31, Issue 1, January 2007.
[19] Matthew J Zukoski; Terrance Boult; Tunç Iyriboz "A novel approach to medical image compression," International journal of bioinformatics research and applications 2(1):89-103, 2006.
[20] Armen Dzhagaryan, Aleksandar Milenkovic, On Effectiveness of Lossless Compression in transferring mHealth Data Files, U.S. National Science Foundation under grant CNS-1205439 and CNS-1217470.
[21] Bhonde, Nilesh, Shinde sachin, Nagmode Pradip and D.B.Rane "Image compression using discrete wavelet transform" in A national level conference held at Prava engineering college, Maharashtra

V.Sherine Vargheese received her B.Sc degree from Avs College of Arts and Science affiliated to Periyar University in 2002 and her Master's degree from Bishop Heber College affiliated to Bharathidasan University, Tiruchirapalli in 2004. She received her M.Phil degree from Periyar University, Salem. She is pursuing her Ph.D degree in Periyar University, Salem. She is working as HOD Cum Assistant Professor in the PG Department of Computer Science at Kailash Women's College, Nangavalli, Salem. Her current research interests include Cloud Computing and Networks.

Dr.D.Maruthanayagam received his Ph.D Degree from Manonmanium Sundaranar University, Tirunelveli in the year 2014. He has received his M.Phil, Degree from Bharathidasan University, Trichy in the year 2005. He has received his M.C.A Degree from Madras University, Chennai in the year 2000. He is working as Head Cum Assistant Professor, PG and Research Department of Computer Science, Sri Vijay Vidyalaya College of Arts & Science, Dharmapuri, Tamilnadu, India. He has 15 years of experience in academic field. He has published 1 book, 20 International Journal papers and 23 papers in National and International Conferences. His areas of interest include Grid Computing, Cloud Computing and Mobile Computing.