ISSN (ONLINE): 2349-7084
GLOBAL IMPACT FACTOR 0.238
ISRA JIF 0.351

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 5, MAY 2015, PP 319-321

A review on Data Compression Techniques in Cloud Computing

Supreet Kaur, Amanpreet Kaur

Abstract— Cloud Computing has become a crucial aspect in today's era of technology in the world and it has grown past all the boundaries. There is a need to connect resources and users without having physical connection. The high demand for data processing and leads to high computational requirement which is usually not available at the user's end. This has encouraged several companies to provide services over the cloud in the form of service, storage, platform etc. But along with its advantages cloud computing has brought with it several challenges like security, storage, scheduling etc. Storage in Cloud computing forms a very important part as the need of virtual space to store our large data has grown over these years. But the speed of uploading and downloading limits the processing time and there is a need to solve this issue of large data handling. This thesis aims at solving this problem using compression technique on multimedia data. A novel Genetic compression technique will be developed and applied on multimedia data and used in cloud computing for managing such large data. The implementation will be done in CloudSim toolkit and the results will be compared against the existing schemes.

Index Terms— Cloud Computing, Storage Space Bandwidth.

1 Introduction

CLOUD computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction .Cloud computing has emerged as a popular solution to provide cheap and easy access to externalized IT (Information Technology) resources. An increasing number of organizations (e.g., research centres, enterprises) benefit from Cloud computing to host their applications. Through virtualization, Cloud computing is able to address with the same physical infrastructure a large client base with different computational needs. In contrast to previous paradigms (Clusters and Grid computing), Cloud computing is not application-oriented but service oriented; it offers on demand virtualized resources as measurable and billable utilities [1].

2 Data Stroage Problem in Cloud Computing

Data storage is the one of the major challenge in cloud computing and this ensues two problems:
As the rate, scale and variety of data increases in complexity, the need for flexible applications that can crunch huge amounts of heterogeneous data (such as web pages, online transaction records, access logs, etc.) fast and cost-effective is of utmost importance [2].

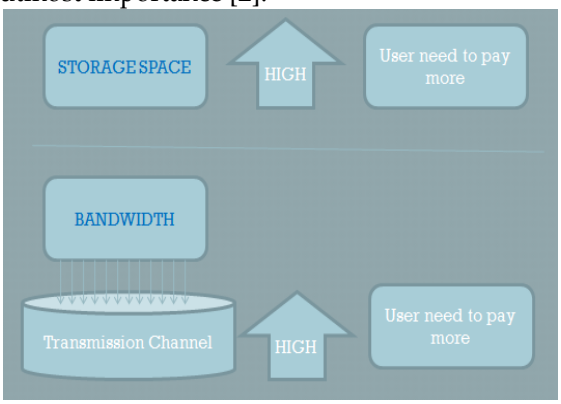

Fig. 1 Storage challenges

Since multimedia applications need to process huge amounts of data, a huge amount of storage space is required. Moreover, processing such huge amounts of data in a scalable fashion involves massively parallel data transfers among the participating nodes, which invariably leads to a high bandwidth utilization of the underlying networking infrastructure. In the context of cloud computing, storage space and bandwidth are resources the user has to pay for. It is therefore crucial to minimize storage space and bandwidth utilization for multimedia applications, as this directly translates into lower overall application deployment costs.

- ***Supreet Kaur*** *is pursuing M.Tech in Computer Science & Engineering in GIMET, Amritsar, Punjab, India.*
  E-mail: ksupreetk9@gmail.com
- ***Amanpreet Kaur*** *is working as Asst. Prof. in Department of Computer Science & Engineering in GIMET, Amritsar, Punjab, India.*
  E-mail: batthamanpreet@gmail.com@gmail.com

ISSN (ONLINE): 2349-7084

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING IN RESEARCH TRENDS
VOLUME 2, ISSUE 5, MAY 2015, PP 319-321

3 EFFICIENR STORAGE IN CLOUD COMPUTING THROUGH COMPRESSION TECHNIQUES

Several researchers have tried to solve this problem of storage through compression techniques. There are several compression techniques available in literature which can be applied to this problem to solve the storage issue.

Data compression squeezes data so it requires less disk space for storage and less bandwidth on a data transmission channel. Communications equipment like modems, bridges, and routers use compression schemes to improve throughput over standard phone lines or leased lines. Compression is also used to compress voice telephone calls transmitted over leased lines so that more calls can be placed on those lines. In addition, compression is essential for videoconferencing applications that run over data networks.

Most compression schemes take advantage of the fact that data contains a lot of repetition. For example, alphanumeric characters are normally represented by a 7-bit ASCII code, but a compression scheme can use a 3-bit code to represent the eight most common letters. Two important compression concepts are lossy and lossless compression.

Several researchers have tried to solve this problem of storage through compression techniques. There are several compression techniques available in literature which can be applied to this problem to solve the storage issue.

Data compression squeezes data so it requires less disk space for storage and less bandwidth on a data transmission channel. Communications equipment like modems, bridges, and routers use compression schemes to improve throughput over standard phone lines or leased lines. Compression is also used to compress voice telephone calls transmitted over leased lines so that more calls can be placed on those lines. In addition, compression is essential for videoconferencing applications that run over data networks.

Most compression schemes take advantage of the fact that data contains a lot of repetition. For example, alphanumeric characters are normally represented by a 7-bit ASCII code, but a compression scheme can use a 3-bit code to represent the eight most common letters. Two important compression concepts are lossy and lossless compression.

TABLE 1: COMPRESSION TECHNIQUE

No.	Compression Techniques	Description of techniques used for compression
1	Lossy compression	• With lossy compression, it is assumed that some loss of information is acceptable according to desire quality of any multimedia data.
2	Lossless compression	• With lossless compression, data is compressed without any loss of data. It assumes you want to get everything back that you put in.
3	Null Compression	• This technique replaces a series of blank spaces with a compression code, followed by a value that represents the number of spaces.
4	Run-length compression	• Expands on the null compression technique by compressing any series of four or more repeating characters. The characters are replaced with a compression code, one of the characters, and a value that represents the number of characters to repeat.
5	Keyword encoding	• Creates a table with values that represent common sets of characters. Frequently occurring words like *for* and *the* or character pairs like *sh* or *th* are represented with tokens used to store or transmit the characters.

4 HELPFUL HINTS

6	Adaptive Huffman coding and Lempel Ziv algorithms	• These compression techniques use a symbol dictionary to represent recurring patterns. The dictionary is dynamically updated during a compression as new patterns occur. • For ,data transmissions, the dictionary is passed to a receiving system so it knows how to decode the characters. • For file storage, the dictionary is stored with the compressed file.
7	DCT (discrete cosine transform)	• DCT is a common compression technique in which data is represented as a series of cosine waves.
8	Spatiotemporal Compression	• By exploring spatial correlation of data, this technique partition a data set into clusters so that, in one cluster all edges from the graph have similar time series of data. In each cluster, the workload can be shared by the inference based on time series similarity. • Based on it, a data driven scheduling will be developed to allocate the computation and storage on Cloud for better big data processing services[3]

5 CONCLUSION

In this paper we discuss the current problems and previous techniques used in cloud computing. The research has been taken place by long time but the problems are not solved completely. Due to the increase in the amount of data enormously, it is not acceptable for efficient storage in cloud computing. Hence, there is need to develop those technologies that will solve the storage problem which be solved by proposed methodology in future.

REFERENCES

[1] Srinivas, J., K. Venkata Subba Reddy, and A. MOIZ Qyser. "Cloud Computing Basics." International Journal of Advanced Research in Computer and Communication Engineering, 1 (5) (2012).

[2] Nicolae, Bogdan. "High throughput data-compression for cloud storage." *Data Management in Grid and Peer-to-Peer Systems*. Springer Berlin Heidelberg, 2010. 1-12.

[3] C. Yang et al., A spatiotemporal compression based approach for efficient big data processing on cloud, *J. Comput. System Sci.* (2014)