



Data Compression Techniques in Cloud Computing

Monika Soni¹, Dr Neeraj Shukla²

Gyan Ganga Collage of Technology in Jabalpur^{1,2}

Abstract: Cloud Computing has become an important aspect in today's world as technology has grown past all the boundaries and 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. 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.

Keywords: Compression algorithms, increasing effective data density, data storage space, resource usage or transmission capacity

INTRODUCTION

Data compression is one of the enabling technologies for multimedia applications. It would not be practical to put images, audio and video on websites if do not use data compression algorithms. Mobile phones would not be able to provide communication clearly without data compression. With data compression techniques, we can reduce the consumption of resources, such as hard disk space or transmission bandwidth.

Data Compression is the process of encoding data so that it takes less storage space or less transmission time. Compression is possible because most of the real world data is very redundant. In this survey, first we introduce the concept of lossy and lossless data compression techniques.

Classification of compression methods-

We have two types of compression methods:

Lossless compression: - It is used to reduce the amount of source information to be transmitted in such a way that when compressed information is decompressed, there is not any loss of information.

Lossy compression: - The aim of lossy compression is normally not to reproduce a exact copy of the information after decompression. In this case some information is lost after decompression

Lossless Compression Methods:-

Run Length Encoding: -

The first step in this technique is read file then it scans the file and find the repeating string of characters [6].when repeating characters found it will store those characters with the help of escape character followed by that character and count the binary number of items it is repeated. This method is useful for image having solid black pixels. This algorithm is also effective for repeating of characters. But it is not effective if data file has less repeating of characters. We can compress the run-length symbols using Huffman coding, arithmetic coding, or dictionary based methods.

Huffman Coding:- The Huffman coding algorithm is named after its inventor, David Huffman, who developed the method as a student in a class on information theory at MIT in 1950[1]. Huffman Coding Algorithm— it is a bottom-up approach 1. Initialization: Put the old nodes in a list sorted according to their frequency counts. 2. Repeat the following steps until the sorted list has only one node left: (1) From the list pick two nodes with the lowest frequency counts.

Form a Huffman sub tree that has these two nodes as child nodes and create a parent node. (2) Assign the sum of the children's frequency to the parent node and insert it into the list such that the order is maintained. (3) Delete the children from the sorted list. 3. Assign a 0 and 1 codeword to the two branches of the tree on the path from the root.

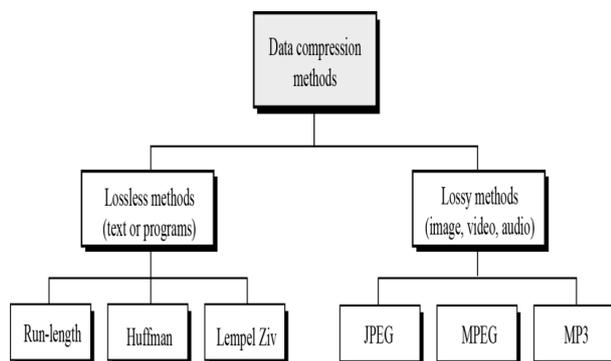


FIG: - TREE REPRESENTATION OF COMPRESSION METHODS



After the Huffman tree, the method creates a prefix code for each node from the alphabet by traversing the tree from the root to the node. It creates 0 for left node and 1 for a right node.

LZW (Lempel-Ziv Welch) compression method:- LZW is the most popular method. This technique has been applied for data compression. The main steps for this technique are given below:-

Firstly it will read the file and given a code to each character.

If the same characters are found in a file then it will not assign the new code and then use the existing code from a dictionary.

The process is continuous until the characters in a file are null.

RELATED WORK

Shanmugasundaram and R. Lourdasamy, "A Comparative Study of Text Compression Algorithms" There are lot of data compression algorithms which are available to compress files of different formats. This paper provides a survey of different basic lossless data compression algorithms. Experimental results and comparisons of the lossless compression algorithms using Statistical compression techniques and Dictionary based compression techniques were performed on text data [1].Md. RubaiyatHasan, "Data Compression using Huffman based LZW Encoding Technique" A method and system for transmitting a digital image (i.e., an array of pixels) from a digital data source to a digital data receiver. More the size of the data be smaller, it provides better transmission speed and saves time. In this communication we always want to transmit data efficiently and noise free [2]. R.S. Brar and B.Singh, "A survey on different compression techniques and bit reduction algorithm for compression of text data" This paper provides a survey of different basic lossless and lossy data compression techniques. On the basis of these techniques a bit reduction algorithm for compression of text data has been proposed by the authors based on number theory system and file differential technique which is a simple compression and decompression technique free from time complexity S.Porwal, Y.Chaudhary, J.Joshi, M. Jain, "Data Compression Methodologies for Lossless Data and Comparison between Algorithms"

This research paper provides lossless data compression methodologies and compares their performance. Huffman and arithmetic coding are compared according to their performances. In this paper the author has found that arithmetic encoding methodology is powerful as compared to Huffman encoding methodology. By comparing the two techniques the author has concluded that the compression ratio of arithmetic encoding is better and furthermore arithmetic encoding reduces channel bandwidth and transmission time also.

U. Khurana and A.Koul, "Text Compression And Superfast Searching" A new compression technique that uses referencing through two-byte numbers (indices) for the purpose of encoding has been presented. The technique is efficient in providing high compression ratios and faster search through the text. It leaves a good scope for further research for actually incorporating phase 3 of the given algorithm. The same should need extensive study of general sentence formats and scope for maximum compression. Another area of research would be to modify the compression scheme so that searching is even faster. Incorporating indexing so as to achieve the same is yet another challenge.

CONCLUSION

In this paper, author presents the review on various lossless text data compression techniques. Authors represent literature survey of various latest papers analysis. Authors conclude that text data can be compressed more than that of by the existing methods and better compression ratio can be achieved in the new proposed method.

REFERENCES

- [1] S. Shanmugasundaram and R. Lourdasamy, "A Comparative Study of Text Compression Algorithms" International Journal of Wisdom Based Computing, Vol. 1 (3), December 2011
- [2] V.K. Govindan and B.S. Shajeemohan,"IDBE - An Intelligent Dictionary Based Encoding Algorithm for Text Data Compression for High Speed Data Transmission Over Internet "
- [3] R.S. Brar and B.Singh, "A survey on different compression techniques and bit reduction algorithm for compression of text data" International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE) Volume 3, Issue 3, March 2013
- [4] S. Porwal, Y. Chaudhary, J. Joshi, M. Jain, "Data Compression Methodologies for Lossless Data and Comparison between Algorithms" International Journal of Engineering Science and Innovative Technology.
- [5] U .Khurana and A. Koul, "Text Compression And Superfast Searching", Thapar Institute Of Engineering and Technology, Patiala, Punjab,
- [6] D.E. Asuquo, E.O. Nwachukwu and N.J. Ekong, "Application of Run-length Encoding Technique in Data Compression