

Design and Data Compression Techniques to Reduced Time in Data Warehouse with Tested Algorithms

Nirmal Sharma¹, S. K. Gupta²

¹Research Scholar, Computer Science, Teerthanker Mahaveer University Moradabad

²Computer Science. & Engineering. Department of BIET Jhansi

¹nirmal1709@rediffmail.com

²guptask_biet@rediffmail.com

Abstract— Present scenario of the industry is very fast functioning but if we are using this concept so takes the technique data compression. Data compression concept is decreasing the unique facts in your data warehouse. In this research paper design two techniques one is size reduction of data warehouse and second is reduced time consuming algorithm. We are merged the number of practices in this paper. This supports to decrease the feasting of costly assets for analysis. Lossless data compression technique is measured text and program and lossy data compression technique is measured multimedia text. Above these techniques show a bit decrease process used text information for data compression. Data compression and data decompression practice is simple which is permitted as of time complication

Keywords— Algorithm Design, Data compression, Decompression, Lossy and Lossless

I. INTRODUCTION

The main purpose of data compression is to decrease the large volume of facts to be communicated. In this scheme is contemporary and evaluate a data compression practice. Data compression is very useful concept for data warehouse [1]. Numerous data warehousing requests need for huge amount of techniques such as increasing the growth of application. At the identical period, the statement systems are having in enormous assignment of data through statement stations.

Data compression has two conceptual techniques-

1. *Lossy data compression*- Lossy data compression is compared by lossless data compression. It is suitable for loss of statistics. These things is reliant on the request, element can be released from the statistics to except loading space.

Lossless data compression- It is probable reduced utmost practical facts in statistical data warehouse. There are so many techniques to decrease dimension in data warehouse by eliminating redundancy methods as Zero Run length Coding [2]

Lossy data compression technique is moreover to advance of lossless data compression. Lossless data compression has used text and program and decreases bits formation in numerical termination. If we are decrease the text and program so that do not lost information. Lossy data compression has used video and audio multimedia information reduces bits and pixel point by detecting marginally information and removing it. There is also not lost effective bits information [3].

In this algorithm, we are taken comparison of text column in data warehouse and same matching text column is used for decompression data in proposed algorithm. We can attain the special outcomes on well define system in data warehouse which type of data compression algorithm is used to particular optimal. For instance, Teradata 13.10 can compress two byte character i.e. Unicode to design an algorithm when two times more data and text can be compress. Numbers of texts compress important characteristic are:

- If we are need to little multimedia text and program cost is decreased for free storage in data warehouse.

- This is enhanced structure I/O routine which is underestimating statistics transferred in memory and warehouse.
- Enhanced data will be increase the speed for storing in data warehouse by the data compression technique.
- Multi-value compression is used Teradata for last few years in data warehouse.
- Multi-value compression is used to data compression and repetitive principles.
- In other hand main assistance of refining routine unit your data will not save in your disk. Practically CPU sequences cannot be needed for compression or decompression. It is used to bit pattern that switches values in data warehouse [4].

II. MOTIVING DATA COMPRESSION PRACTICES

a) *Zero Run Length Coding*: The Zero Run-Length Coding proposed the algorithm to compress the data by arranging the most and least attainable data. The entire data is divided into two parts in such a way that their numbers of step are used. It is coded sequences of zeros to shorter strings. Now we explain the technique of the Zero Run-Length Coding step by step 233300000/9 (output of the example from the Move-To-Front Transform) as example input.

233300000/9

REPLACING 2 BY 3 AND 3 BY 4

OUTPUT = 3444

Process Steps:

Step 1- Increase all characters of the input which are greater as 0 by 1

Input: 233300000 = 344400000

Step 2- Coding the sequences of zeros with string combinations of 0 and 1

We use at this point an example coding table.

Number of zeros coding string

1	0	
2	1	
3	0	0
4	0	1
5	1	0
6	1	1

344400000 = Output: 344410

It reduced the size of data warehouse character is 9 to 6 bits per string [5].

b.) *Huffman Encoding Algorithm*: In this techniques test the string character reduced encoded the size of data warehouse. The main advantage of occurring character in memory and practices the little bit loading and reducing cost also. It gives an instance of a variable-length encrypting—certain characters can first need 2 or 3 bits and additional characters can need 7, 10, or 12 bits. String instance is “MISSISSIPPE STATE”.

II. PROBLEM STATEMENT

All The compression data achieve the time needs for reduction of text and program rate. The essential complication of scientific for temporary data radically used for each limited application both lossless and lossy data compression practices The main challenge is generated design the new feature developed for in this field. In this method are not used real number values. But we can use uncertainty and distinct bit patterns. Therefore, we are using two data compression techniques but we analysis by lossless data compression and reduced the size of database and time also [6].

III. COMPRESSION AND DECOMPRESSION METHODOLOGY

Proposed Algorithm of data compression for Time Performance Tool:

Data compression algorithm is compressed the text by bit decoding and encoding practise steps are following:

Step1- Take illogical string for example “MISSISSIPPE STATE”. By the standard ASCII encoding, this 17 character string involves $13 \times 8 = 104$ bits total. The table below displays the relevant subset of the standard ASCII table.

Character	ASCII Number	Binary Number
M	77	01001101
I	73	01001001
S	83	01010011
P	80	01010000
T	84	01010100

A	65	01000001
E	69	01000101
Space	32	00100000

The string "**MISSISSIPPI STATE**" would be encoded in ASCII as **77 73 83 83 73 83 83 73 80 80 73 32 83 84 65 84 69**.

Step 2- These 13 characters are written as the binary code as under box.

0100 1101	0100 1001	0101 0011	0101 0011	0100 1001	0101 0011	0101 0011	0100 1001
0101 0000	0101 0000	0100 1001	0010 0000	0101 0011	0101 0100	0100 0001	0101 0100
0100 0101							

Step 3- The given string we have only seven distinct characters and we will set up a special coding table using three bit of each characters.

Character	Number	Bit Pattern
M	0	000
I	1	001
S	2	010
P	3	011
T	4	100
A	5	101
E	6	110
Space	7	111

Step 4 using the three bits table the string "**MISSISSIPPI STATE**" encoded as **0 1 2 2 1 2 2 1 3 3 1 7 2 4 5 4 6**

Represented as binary numbers:

000 001 010 010 001 010 010 001 011 011 001
111 010 100 101 100 110

Using three bits per character, the encoded string requires 51 bits instead of the original 136 bits, compressing to 62.5% to its original size [7].

Just in time, already presented quantity of profitable data analytical tools it produced cost of a lot for manufacturing time, it is run by two main mechanisms for testing as –Test fact and figures and size of data text and information request. This proposed tool and algorithm

reduced the text and data over the 10 pixel and bits information time [8].

IV. CONCLUSION

This algorithm stretches improved compression ratio when inserted between Zero Run length coding (ZRLC), Huffman Encoding Code (HEC). These algorithm compress 62.5% data has been compressed. Using Huffman coding, we can translate the communication into a string of bits and send it to you. However, you cannot decompress the communication. We observed at the structure of database and the numerous phases of compression and the decompression, in which the phases are track in opposite direction compared to the compression. Designed TPA tool, and reduced the size of database contents 62.5% for data warehousing the using algorithm compress technique and bit reduction 9 to 6 bits. Although exploiting storage space operation and enhancing presentation of facts time to time grow in data warehouse.

REFERENCES

- [1] Govind Prasad Arya, Prince Kashyap, Nilika Kumari, Mitali Hembrom,"CAPSULE A Programming Language Code Compression Technique", International Journal of Computer Science and Information Technologies, Vol. 4 (6), 2013, 883-885.
- [2] Ajit Singh and Yogita Bhatnagar, "Enhancement of Data Compression Using Incremental Encoding", International Journal of Scientific & Engineering Research, Volume 3, Issue 5, May-2012 1 ISSN 2229-5518.
- [3] Rupinder Singh Brar Bikramjeet singh,"A Survey on Different Compression Techniques and Bit Reduction Algorithm for Compression of Text/Lossless" Data International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 3, March 2013 ISSN: 2277 128X.
- [4] <http://www.teradatamagazine.com/New-Options-for-Compression> "New Options for Compression reducing data volume puts the squeeze on system cost".
- [5] Daniel Schiller, "The Burrows Wheel Algorithm", International Journal, August 5, 2012.
- [6] Sriram Lakshminarasimhan, Neil Shah1, at., el,"Compressing the Incompressible with ISABELA: In-situ Reduction of Spatio-Temporal Data", USA International Journal, 2011. www.cs.cmu.edu/~neilshah/research/papers/ISABELA.EuroPar.2011.pdf
- [7] Julie Zelenski, Keith Schwarz, "Huffman Encoding and data compression", handouts, Spring, May 23, 2012.
- [8] Anshuman Chandra, Krishnendu Chakrabarty, "Analysis of Test Application Time for Test Data Compression Methods Based on Compression Codes" JOURNAL OF ELECTRONIC TESTING: Theory and Applications 20, 199–212, 2004 c 2004 Kluwer Academic Publishers. Manufactured in The United States.