Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 847-854

http://www.ripublication.com

A Reassessment on Security Tactics of Data Warehouse and Comparison of Compression Algorithms

KM Divya and Jitendra Kurmi

Department of Computer Science Engineering, University Institute of Engineering & Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, India
(divyabharti1990@gmail.com, jitendrakurmi458@gmail.com)

Abstract

Objectives: Data is a major asset for any enterprise. Processed data not only helps in knowing the past, but also support today's business and predicts future trends. Since any system in this world works upon data and with the systems becoming complex, it became very difficult to manage this bulk data and hence this concept of warehouse needed to resolve but just storing the data doesn't help it. Data is anytime prone to security risks, hence it's very essential to secure this data and manage it well.

Methods/Statistical analysis: Various security techniques are designed to protect the data from various threats and this bulk data also needs a sort of compression in a way that the data quality is not compromised as well as the storage space is effectively managed. This paper introduces the data security metrices and the comparison of data compression techniques. Factors which are being considered are speed, applications, compression ratio, advantages and disadvantages.

Findings: Security attacks on data warehouse have been continuously increasing. Data security emphasizes on three main issues i.e Confidentiality, Integrity& Availability of data

To comply with the above three attributes, many techniques have been proposed. Proactive data security techniques protect the data in advance from security breaches or attacks. They used data masking, encryption, and data access policies. Reactive data security techniques effectively respond after the

security attack or security problem, .i.e. these techniques work after the security problem has occurred. These monitor and analyze the user actions and try whether the actions are harmful or not, thus protecting the data that bypass the. Preventive security techniques. These techniques use Intrusion detection system. Data compression has become the most crucial and essential part of data warehousing as it helps in saving the disk overhead and improves the query performance as well. This review paper conferencing different type of lossy and lossless compression tactics. Reassessment only depicts the broad idea of data compression. Comparison of various compression techniques are discussed in this paper.

Application/Improvements: Now many compression approaches are refined and some approaches are in progress. Some compression techniques are the best from the query processing point of view or the compression ratio.

Keywords: Data warehouse, security metrices, confidentiality, integrity, availability.

I. INTRODUCTION

Data warehouse is basically subject oriented, time variant, non volatile and integrated collection of data from various source system[8]. It is a relational database that is used for query and analysis purpose. It is the core component of business intelligence. It's composed of data from multiple heterogeneous that support adhoc queries, decision making and analytical reporting. This makes it a key business bonus for any enterprise, at the same time making it a target for attackers. Recently it has been published that the security attacks on data warehouse has been continuously increasing. Data security emphasizes on three main issues i.e Confidentiality, Integrity& Availability of data. Confidentiality emphasizes on protecting information from unauthorized users─Integrity emphasizes on protecting the data and no unauthorized party has altered the data. Availability ensures that the information should be available to all the authorized users whenever they want it[4,5,6] .

To comply with the above three attributes, many techniques have been proposed. These can be classified in two broad categories: preventive and reactive techniques[9].

II. SECURITY TECHNIQUES IN DATA WAREHOUSE

Proactive data security techniques- These techniques protect the data in advance from security breaches or attacks. They used data masking, encryption, and data access policies.

(A)Data masking- It is the procedure where the actual values are restored by bogus values or inexact abstract, so that our responsive and true facts unaccessible and hence not be conciliate by unlawful users.

(B)Data encryption- It is a method of interpreting abstract into another form so that only explicit users, who have the avenue to a unrevealed key (decryption key), can read it.It is one of the most accurate path to accomplish data security.

Reactive data security techniques-These techniques effectively respond after the security attack or security problem, .i.e. these techniques work after the security problem has occurred. These monitor and analyze the user actions and try whether the actions are harmful or not, thus protecting the data that bypass the preventive security techniques. These techniques use Intrusion detection system.

III DATA COMPRESSION

Data compression has become the most crucial and essential part of data warehousing as it helps in saving the disk overhead and improves the query performance as well. Objective of data compression is to contract the volume of data warehouse. Different compression techniques are used to deal with different types of attributes. Some compression techniques are the best from the query processing point of view or the compression ratio[7,10,11].

Two types of compression tactics are exist in data compression.

Lossless data compression-It grant all of its authentic data regained when the file is uncompressed again. It is a process in which after the reconstruction we get our authentic data [2,3].

Lossy data compression-A compression in which we did not get our authentic data after the reconstruction, because few bits are lost during the process, which means after the process the volume of the data is contracted conveniently but when we rebuild our data, few bits are lost, which may conciliate the quality of the data[2,3]..

Table No 1 Comparison between Lossless compression and lossy compression

FACTORS	DATA COMPRESSION	
	LOSSLESS COMPRESSION	**LOSSY COMPRESSION**
Definition	It grant all of its authentic data regained when the file is uncompressed again. It is a process in which after the reconstruction we get our authentic data[3].	A compression in which we did not get our exact data after the reconstruction. Data is contracted conveniently but when we rebuild our data, few bits are lost, which may conciliate the quality of the data[2].
Techniques	Run length encoding, arithmetic encoding, Huffman coding, Lempel Ziv.	Vector Quantization, DCT, DCW.
Uses	Text or programs	Images, audio and video.
Advantages	Compression is not efficient. But it maintain quality.	Compression is efficient. But it did not maintain quality. Suitable for compressing the multimedia file.
Compression Ratio	It does not generally acquire higher compression ratio.	Compression ratio is much better as compared to lossless.
Drawback	It doesn't reduce the file size as much as Lossy compression.	It degrade the quality of the data. We did not reconstruct our original data.

Run length encoding-It is seemingly the easiest way of compression. It is the appropriate approach for compressing the data, made of any combination of symbols[3]. Data consist string of identical unit. This technique is more feasible when only two symbols (0 and 1) used in its bit arrangement and one symbol is more repeated than the other. By substitution these repeated unit arrangement with the number of occurrences, a compelling reduction of data can be achieved. This process is known as Run length encoding.

Huffman coding- It is an effective technique of lossless data compression which induce no loss in information. This algorithm creates an array of frequencies of every character in the file. This array is there after used for determining an optimal way where each character is represented as a binary sequence[1]. In this the repetition of

every symbol in the sequence is helpful for creating a variable prefix code and then each symbol gets mapped to every binary sequence.

Adaptive Huffman coding-Normally we use huffman coding for compressing the data, the only drawback of Huffman coding is to send the probability table along with the compressed information because in decoding process, probability table play major role. For removing this drawback adaptive Huffman coding has been developed. The table needs insertion of '0', an additional byte to the output table, but as a outcome it doesn't make much asymmetry in the compression rate. In adaptive coding one pass encoding is used.

Lempel Ziv Welch- LZW conceal is a kind of dictionary based coding scheme. The aim is to construct a dictionary of strings and apply during the communication process. This code is not designed for any particular source but a large class of sources. Moreover, this algorithm befits for any fixed stationary and ergodic source in a way that it is designed for that very source, which means Lempel ziv performs very well for that source.

Table No 2 Comparison between Some of the Lossless compression algorithm

FACTORS	Lossless Compression Techniques		
	RLE	**LEMPEL ZIV**	**HUFFMAN CODING**
Advantages	Easy to implement.	It is simple and good compression. The Lempel ziv code is not designed for any particular source but a large class of sources.	It is easy to implement. It is ideal for compressing text or program files.
Speed	Rapid to execute.	Fast compression	Fast to execute.
Application	TIFF, PDF, BMP	TIFF, GIF, PDF	ZIP, ARJ, JPEG, MPEG
Drawback	This compression method cannot attain the high compression ratio as compared to another advanced method.	Difficulty while managing the string table. Dictionary is needed by everyone.	Decoding is hard to do due to dissimilar code lengths. Overhead due to Huffman tree.

Vector quantization-VQ is a strength way for compressing the data like audio, video and image. Because in this the vector representation over and over occupy only small portion of their vector spaces. Despite the fact that VQ serves more compression, up until now it is not universally implemented. This is due to two things. First one is that it takes time to constructing the codebook, and the second one is the searching time.

Discrete cosine transform- DCT works on the principle of partial separation of images into differing frequencies. The less important frequencies are discarded during the part of compression, called quantization. Hence few bits are loosed during the process, the term LOSSY is used. Thus only the most important part of frequencies of images is retrieved during decompression, resulting in distorted images.

Discrete wavelet transform-It is used for perform the wavelet transform with help of distinct set of the wavelet scales and subtitle that follow some principle. Preferably change completely and break up the signal into commonly orthogonal sets of wavelets. By this way it quite differs from the unfaltering wavelet transform. As once in a while it perform the discrete time, it is also named as the discrete time continuous wavelet transform (DT-CWT).DWT is used in the image blocks that are generated by the preprocessor. Two dimensions it often ends up in decomposing the estimation coefficient at level j in four components. The approximation is at level j=1 and the details are fed in the following orientations which are horizontal, vertical and diagonal.

Table No 3 Comparison between some of the Lossy compression algorithm

	Lossy Compression Techniques		
FACTORS	**Vector Quantization**	**DCT**	**DWT**
Advantages	This method is suitable for multi dimensional model. Flexibility in terms of design.	It gives a unique transformed value and so it would easy to convert back into its original value. Coefficients are nearly correlated.	Inherent multi-resolution nature, wavelet coding schemes. Used where scalability and tolerable degradation are important.
Computation	Technique is fast.	It is a fast computational approach.	It is computationally very fast.
Application	JPEG	JPEG, MPEG, MJPEG	JPEG, MPEG
Drawback	The designing of code book is a major problem in VQ.	The most important part of frequencies of images is retrieved throughout decompression, resulting in distorted images.	DWT is shift sensitive because input signal shifts generate unpredictable changes in DWT coefficients.

IV. CONCLUSION

Hence we can conclude how the addition of security tactics in the data warehouse affects the functioning of the data warehouse. This review paper conferencing different type of lossy and lossless compression tactics. Reassessment only depicts the broad idea of data compression. Comparison of various compression techniques are discussed in this paper. Now many compression approaches are refined and some approaches are in progress. This paper depicts comparison between lossless and lossy techniques. Comparison between some of the lossless algorithm such as Run length encoding, Huffman coding, Adaptive coding, Lempel Ziv coding as well as comparison of various lossy algorithms which are DCT,DWT and Vector quantization. This paper has been written to understand various compression techniques.

REFERENCES

Research Article

[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] Saumya Mishra, shraddha singh,"**A Survey Paper on Different Data Compression Techniques",** Indian journal of applied research, Volume: 6 | Issue: 5 | May 2016 | ISSN - 2249-555X

[3] Rajinder Kaur, Mrs. Monica Goyal, "**A Survey on the different text data compression techniques"**,International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 2, February 2013.

[4] Ricardo Jorge Santos; Jorge Bernardino; Marco Vieira," **A Survey on Data Security in Data Warehousing": Issues, Challenges and Opportunities** Conference Paper April 2011 DOI: 10.1109/EUROCON.2011.5929314 · Source: DBLP

[5] Anjana gosain; Amar arora,"**Security issues in data warehouse"**: A systematic review, International conference on intelligent computing, communication and convergence (ICCC-2015)

[6] Saiqa Aleem Luiz Fernando Capretz; Faheem Ahmed:**"**Security Issues in Data Warehouse**"** Recent Advances in Information Technology, 5th European Conference of Computer Science (ECCS'14), Geneva, Switzerland, pp. 15-20, December 2014.

Books

[7] Mark nelson, Sean –Loup Gailly, "The data compression book", second edition, "Publisher: IDG books Worldwide.

[8] H. Inmon, Building the Data Warehouse, 3rd Ed; John Wiley, USA, 2002.

[9] N. Yuhanna, You're Enterprise Database Security Strategy, Forrester Research, 2010.

[10] Introduction to Data Compression, Khalid Sayood, Ed Fox (Editor), March 2000.

[11] Blelloch, E., 2002. Introduction to Data Compression, Computer Science Department, Carnegie Mellon University