Data Analysis in Trade

R. Bhuvaneswari[1*], M.Malathi[1], D. Ravi[2]

[1] Associate Professor, Department of CSE,Kathir College of Engineering,Coimbatore,TN, India
[2]UG Scholar, Department of CSE,Kathir College of Engineering,Coimbatore,TN, India

*Corresponding author E-Mail ID: bhuvaneswari030@gmail.com Mobile: +917373156689

ABSTRACT

Data Warehousing is arisen in order to overcome the issues faced by the excel file. When several database is maintained by the excel file a conflict will arise such that which source file belongs to which excel sheet. In order to overcome the problem a data warehousing is used and a new technique used in data warehousing is soft concatenation mapping which depicts a relationship between column during transformation and concatenation .At that time many redundancy and loss of data may occur in order to overcome the issue an efficient method is used that is used to derive an approximate algorithm. This approximation is used both in real time and synthetic data sets.

Keywords-Database management, data profiling, data integration, data analysis, Concatenation

1. INTRODUCTION

A Data Warehouse is a form of central repository which stores the data that are collected from various internal and external source file .Data Warehousing that emphasizes and capture the data for processing in an legitimate manner. Typically a Data Warehousing is a form of relational database which is mounted on an enterprise main framework. Dataset from various online transaction process and other online system they are selectively extracted in order for efficient analysis on Business activities ,for better decision support and also enhance the growth of easy analysis task. The process of collecting all the required data about the small or large scale organisation and storing in the central repository and by the use of data analysis task using the certain tools an enhanced analysis process can be taken out from the output of the data analysis an better decision can be taken out or that make an intimate for users as kind of awareness before proceeding the process.

1.1 EXISTING SYSTEM

The Data are stored in the form of excel where the searching process and their performance is very low. Since the data is stored in the form Excel sheets and Spreadsheet the data is information that is stored in any spreadsheet program such as Excel or Google Sheets or any other format. The Data stored in cells in a worksheet can be used in calculations, displayed in graphs, or sorted and filtered to find specific information. The maintenance of the Excel sheet for collection of storage of all amount data in an large database is that the performance is very low and issue will arise such that which data belongs to which data source .similarly, the maintenance of the collection of huge data in the Excel sheet will result in an many redundancy and loss of data.
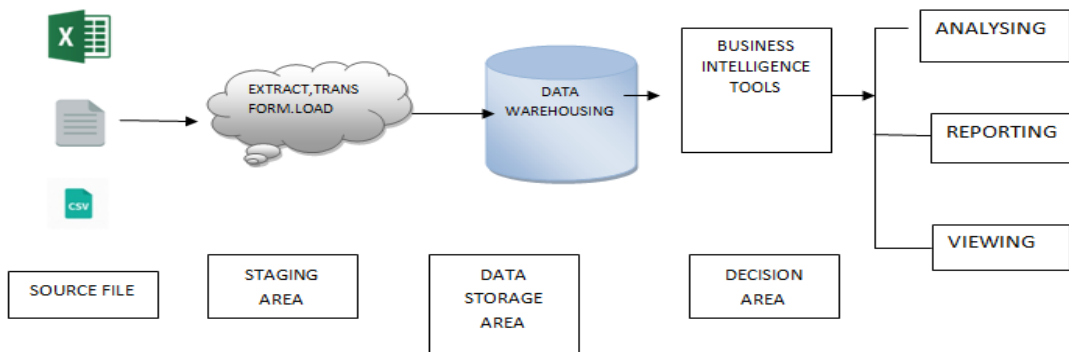
2. ARCHITECTURE OF PROPOSED SYSTEM

Fig 1. Architecture of proposed system

2.1 Working of the Proposed Model

Step 1: initialization process.

Step 2: The source file can be of any format such that excel,word,csv etc.., which can be given as an input to the staging area .

Step 3: In the staging area the database is gone under some operations where the required data that are needed by the users are the extracted .The extracted data will be the same as before.

Step 4: After the extraction process the data is then transformed by using a certain conditions and then converted into the output .It is made to connect the source and destination data.

Step 5: The data is cleaned such that by reducing the redundancy and loss of data. The process of cleaning the data and integrating the data all done under the same platform .So therefore, the data cleaning and data integration takes place.

Step 6: After the cleaning and the integration process is done the entire original, cleaned data is stored under the central repository which is the data warehousing.

Step 7: The data in the warehousing is gone under some analyzing process so which gives the user better and clear view so that would support the user to have a clear understanding and help them to take better decision and that would predict the future scenario.

Step 8: Finally the data is then mined and new information would be derived and an efficient reporting can be made.
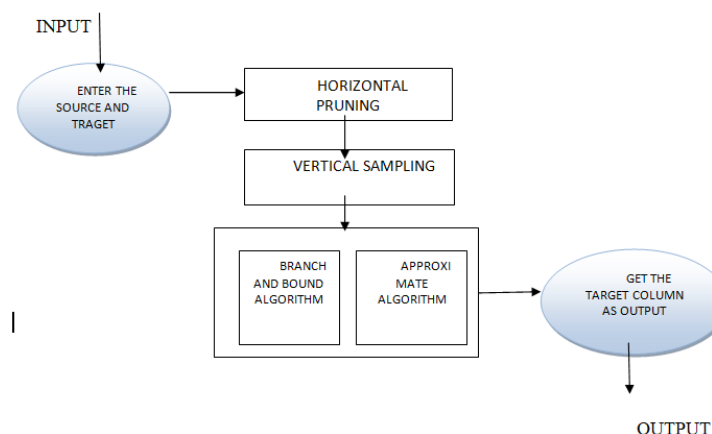
2.2 Proposed Algorithm

Fig 2. Algorithm working module

2.3 Process of Proposed Algorithm

Step 1: The user during the analysis process should give the required column as input from the database.

Step 2: The source column is taken as input and is made to split it up as an horizontal manner which is done according to the convince of the user requirement.

Step 3: The required output produced from the horizontal pruning is then again went on some under process where the rows are the made to sample in a vertical order.

Step 4: The required column is then again going under some process where the column is made to search the perfect data without any redundancy and loss .This kind of operation is done by using the branch and bound algorithm.

Step 5: After an efficient search is made the data are gone under the approximate algorithm which is used to reduce the redundancy of the columns efficiently.

Step 6: desired output for the required columns will be obtained.

3. PERFORMANCE AND EVALUATION

Performance of the Data Warehouse is a highly comprehensive and complex data storage, information management solution, hosted or on other criteria's, that integrates with the enhancing analytical and reporting tools of Business Analytics and Performance Analytics from very a powerful business intelligence tool which could be then viewed with the help of an dashboard. It provides for the organization with a "effective decision making and reporting " – a collection of sharing metrics and dimensions, facts etc,.. that can be used throughout our organization for an efficient handling process – using leading-edge method. This kind of maintenance on huge collection of data make an efficient and high performance on fast retrieval and reporting needs.
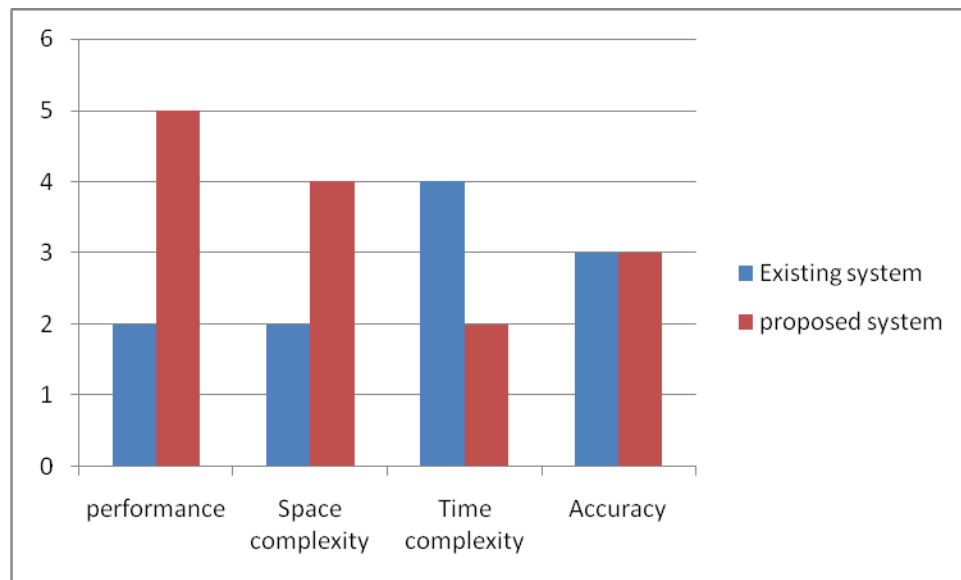

Fig 3. Performance factors

4. ADVANTAGE

- Performance is enhanced
- Effective utilization
- Provide the generic method
- Readable by the users
- Proper reporting
- Runtime is low

5. CONCLUSION

Therefore, we propose the concept of mapping of soft concatenation, a relationship between database columns would be observed. The two mapping detection problems had been observed . We then proposed an Branch and bound, approximate algorithm. Both algorithms can scale to hundreds of columns. In particular, the approximate algorithm targets at detecting the optimal SCM only, which makes it orders of magnitudes faster than Branch and bound. Our experimental evaluation demonstrates that our algorithms are both efficient and scalable and it would be useful for the future use.

REFERENCE

[1] https://www.oracle.com/datawarehousing.

[2] W. A. Giovinazzo, Object-Oriented Data Warehouse Design: Building a Star Schema, Englewood Cliffs, NJ, USA:Prentice Hall, 2000.

[3] P. Bohannon, W. Fan, F. Geerts, X. Jia, A. Kementsietsidis, Conditional functional dependencies for data cleaning, *IEEE 23rd Int. Conf. Data Eng.*, , 2007.pp. 746-755.

[4] P. G. Brown and P. J. Hass, Bhunt: Automatic discovery of fuzzy algebraic constraints in relational data, in VLDB, 2003, pp. 668–679

[5] H. Liu, J. Xiao, X. Guo, H. Tan, Q. Luo, and L. M. Ni, Cuttle: Enabling cross-column compression in distributed column stores, in APWeb-WAIM 2017, 2017, pp. 219–226

[6] J. Liu, J. Li, C. Liu, and Y. Chen, Discover dependencies from data a review, IEEE Trans. on Knowl. and Data Eng., 2012, vol. 24, no. 2, pp. 251–264.

Conflict of Interest

None of the authors have any conflicts of interest to declare.

About the License