International Journal of Engineering Research and General Science Volume 3, Issue 2, March-April, 2015
ISSN 2091-2730

An Ensample for Proficient Insight of Data Visualization

Ambreen Khan[1], Apeksha Sakhare[2]

[1]Project Scholar, Department Of CSE, GHRCE- Nagpur,khanambreen11@gmail.com

[2]Assistant Professor, Department of CSE, GHRCE- Nagpur, apeksha7777@gmail.com

Abstract— Recent advancements in the technology have reconnoitered, thus leading to enormous generation and utilization of data. Data acts as a fundamental part for all the domains including the data warehouse domain. Data can be explored in various ways; one such way to explore and conceptualize the data is termed as "Data Visualization". This insight of data proffers an ease for the users to understand the scenario more precisely. But this visualization faces a lot of challenges as a correct visualization picture is of utmost importance. Along with visualizing data, the processing part should also be equally efficient. To deal with multidimensional data OLAP (Online Analytical Processing) appears to be a more feasible option but OLAP does not provide an efficient processing when pattern or sequential data is considered. Another issue that upraises over here is that there is no accurate processing of traditional and pattern data in the same OLAP. To deal with issues specified, a system is designed that deals with both traditional as well as sequential data and provides suitable visualization. The system incorporates a sequential OLAP and complexity design modeling technique to achieve the objective.

Keywords— Data warehouse, report, OLAP, datasets, multidimensional, Sequential, Visualization, complexity design modeling.

INTRODUCTION

As the technology is heading towards a wider array of knowledge, so there arises a need to improve the efficiency of the work performed. A data centric domain is always the substratum of all the technology. Data plays a pivotal role in the rapid emerging techsavvy globe and is used, provided and generated. In an effort to acquiesce decision makers the ease to analyze complex data for quick, efficient, interactive and meaningful exploration, a technology called Online Analytical Processing (OLAP) [3] has emerged. Due to its distinctive features OLAP has owned a remarkable demand in the business world as well as in the fields of academics and research, as it is competent enough to deal with huge datasets [1][5]. OLAP technology works on multi-dimensional databases or more pronouncedly referred to as MDB [2]. The use of these MDB's helps the firm to have a global view of the data as well assists the decision-makers or knowledge engineers to acquire an apprehensive perception of the scenario thus leading to collaborative access and surge in the enactment speed. The data in any industry is broadly classified into two categories, one is the transactional (operational) data and another is the historical data.

Data warehouse and the OLAP technology emphasizes on the historical data and the OLTP (Online Transaction Processing) deals with the transactional or operational part of the data [5].The data warehouse assimilates data from diverse sources storing its history so as to increase the proficiency of the decision support system [4]. Due to the voluminous data escalation rapid advancement of OLAP is required. Till now OLAP's performance and visualization have not been amalgamated in a way so that it can deal with the subject. The concern with the proficient visualization of MDB's generated by various domains including academics, scientific research, healthcare industries, stock market-commerce and statistics still subsists[2][7]. There is as such no measure taken to efficiently envision them. On the contrary there are certain commercially acknowledged tools for visualization, but the assimilation of OLAP with visualization goes missing. Another issue of concern is that the OLAP does not provide the most efficient solution for pattern or sequential datasets [6] such the RFID logs, biometric logs or server access logs. There is a need of emergence of a technique that deals with both traditional as well as sequential data and also visualizes the data effectively.

RELATED WORK

The domains that majorly deal with data have undergone remarkable shift since the era of its emergence. This is because of the constant updating in the technologies used. To deal with the rising need many researches have been commenced. Apart from researches many commercial tools have emerged. To enlist a few famous ones, such as Tableau [11] that is very much recognized globally, Spagobi [12], Actuate [13] etc. Majority of these commercial softwares work on the relational databases rather than on multidimensional database. The prime focus of these tools is only on the visualization part and the analysis and the processing part has a wide spectrum of improvement. Also there is a huge amount of literature that covers the functionality of OLAP in the data warehouse and various types of OLAP's proposed that may deal with varied customized situations. But none of the OLAP's focused on the issues of dealing with the massive datasets generated including the traditional and sequential ones. One such OLAP approach is the Visual OLAP [14] approach that outwits the drawback imposed by the conventional interfaces. The approach focuses on the exposure of the collaboration between the consummation-oriented techniques used foe BI (Business Intelligence) and the

accomplishments in the extents of Information Visualization. The authors in [15] [16] [17] and [18] also discussed regarding the various alternatives available that can provide an optimal solution. The issue of efficient visualization of both traditional and sequential data using a single system still exists.

SYSTEM MODEL

To deal with the problem of effective analysis and visualization an approach is designed. This system design or the model uses the combination of two influential techniques:

1. Sequential Analytical Processing [8][9]
2. Complexity Design Modeling [10]

The system model is described below in figure 1. The system is aimed towards providing an efficient visualization through an OLAP that is sequential in nature and manages to process the data that is sequential as well as the data that is non-sequential (traditional) in nature. The user here can be a manager or admin respective to the domain whose data is provided as the input to the system. The system is tested on various datasets such as academic dataset, toll-tax data, and biometric logs etc.

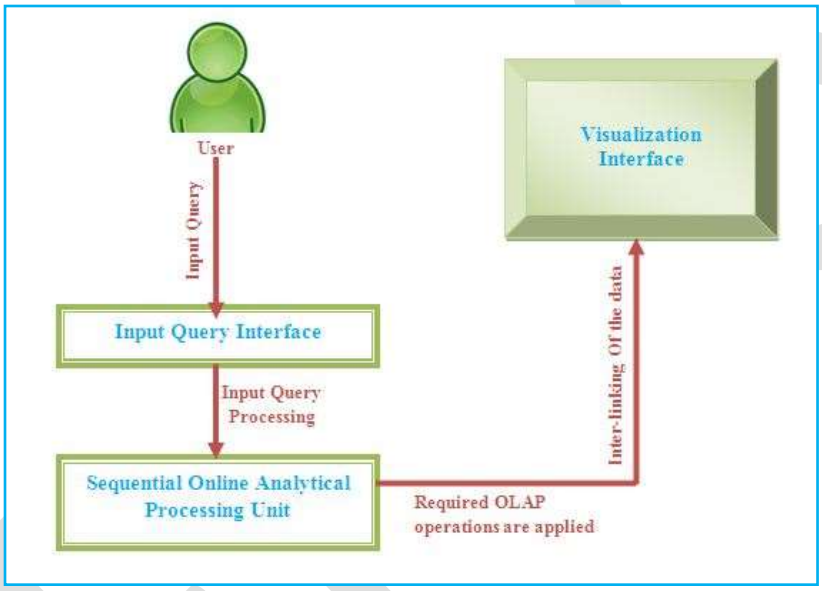

Figure 1: The workflow of the system model

As per the model the user interacts with the system by providing an input query to the system as shown in the figure 1**.** The input query is then processed using the sequential online analytical processing unit. This unit performs internal OLAP operations as per the query. After efficient query processing, the data is interlinked and then the final output is visualized in the visualization interface. The plan of work is shown in the below figure 2.

Plan of Execution

1. User enters the query
2. The query goes to the sequential online analytical processing unit
3. The OLAP units execution initiates:
   a. Processing of the query
   b. Applying suitable operation or sequence of suitable operations
4. The processed data is given in tabular format to the visualization unit
5. The necessary inter-linking of the data is performed
6. The final output is visualized

Figure 2: Plan of Execution

The Sequential Online Analytical processing unit is an addition to the OLAP so as to make it capable of effectively dealing with sequential as well as traditional data. More precisely Sequential OLAP [9] is a neoteric on-line analytical processing system created

for warehousing and analysing the sequential or pattern data. There are various conducive attributes in the sequential OLAP. The traditional OLAP systems groups the data in rows based on their attribute values but the sequential OLAP systems treats patterns as dimensions and it groups sequences based on the patterns they possess.The sequential OLAP consists of seven fundamental operations that are enlisted in below figure 3.

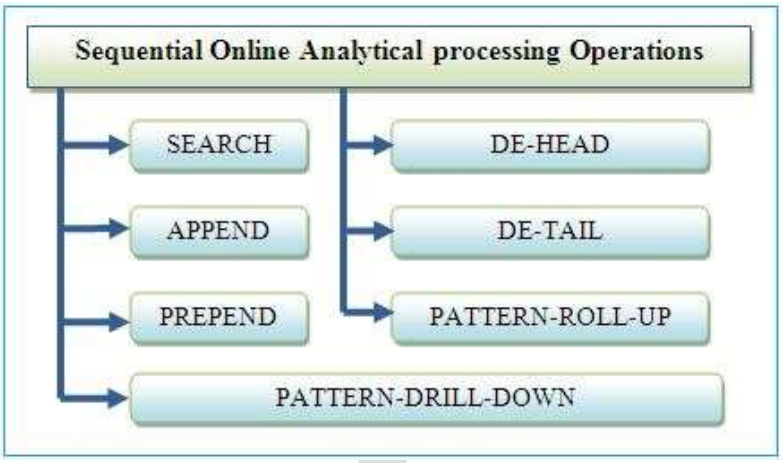

Figure 3: Sequential OLAP operations

The first operation i.e. SEARCH operation performs the basic task of finding or discovery of certain entry in the dataset and if a match exist it returns it as an output. For an instance when the academic dataset was considered that had the entries of students admitted in an institute in the year 2004 that contained details such as enrollment number, department, contact details, address and percentage of each semester etc. The snippet of the dataset is shown in below figure 4.

Admissio	Name	Gender	Departme	enroll no	roll no	contact nc	Address li	Address Li	ist Year	3rd Sem	4th Sem	5th Sem	6th Sem	7th Sem	8th Sem	Graduatio	Graduate	drop outs
2004	SNEHAL K	F	CSE	NU/A9/15	1	7.36E+09	Chhaoni	Nagpur	58	62	63.2	66	65	54	58	2008	Yes	No
2004	NISHIGAN	F	CSE	NU/A9/15	2	7.36E+09	Hingna	Nagpur	70	70	70	70	70	70	70	2008	Yes	No
2004	TANUSHRI	M	CSE	NU/A9/15	3	7.36E+09	Dharampe	Nagpur	70	70	70	70						Yes
2004	TEJAL NAI	M	CSE	NU/A9/15	4	7.36E+09	Gandhi Ch	Nagpur	58	62	63.2	66						Yes
2004	PRIYA VIN	F	CSE	NU/A9/16	5	7.36E+09	Itwara	Nagpur	70	70	70	70	70	70	70	2008	Yes	No
2004	AADITYA \	M	CSE	NU/A9/16	6	7.36E+09	Birdey	Nagpur	70	70	70	70	70	70	70	2008	Yes	No
2004	ABHISHEK	F	CSE	NU/A9/16	7	7.36E+09	Mahesh N	Nagpur	70	70	70							Yes
2004	AKSHAY V	F	CSE	NU/A9/16	8	7.37E+09	Jafar Naga	Nagpur	58	62	63.2	66	65	54	58	2008	Yes	No
2004	ASHOK RA	M	CSE	NU/A9/16	9	7.37E+09	kadbi Cho	Nagpur	70	70	70	70	70	70	70	2008	Yes	No
2004	BRAMHAN	M	CSE	NU/A9/16	10	7.37E+09	New Colo	Nagpur	58	62	63.2	66	65	54	58	2008	Yes	No
2004	PAWAN JA	M	CSE	NU/A9/16	11	7.37E+09	Ahbab Co	Nagpur	70	70	70	70	70	70	70	2008	Yes	No
2004	DEVAL SA	M	CSE	NU/A9/16	12	7.37E+09	Friends Cc	Nagpur	58	62	63.2	66	65	54	58	2008	Yes	No

Figure 4: View of academic dataset

The second operation is APPEND operation that performs the task of affixing the data to the end of the pattern or the template. The next operation i.e. PREPEND does the opposite of the second operation it prefixes at the beginning of pattern or template. DE-HEAD and DE-TAIL removes the first entry and the last at the template respectively. The last two operations works same as Roll-Up and Drill-Down in regular OLAP with an extension that they are adept to process the pattern data. The PATTERN-ROLL-UP operation rolls up the dimension to the upper hierarchy. On the contrary the PATTERN-DRILL-DOWN operation drills the data to the most detailed level in the hierarchy so that all the possible details can be understood. These operations in the processing unit are the most important measure because their efficient processing only guarantees that the output obtained after visualizing will be accurate.

EXPERIMENTAL RESULTS

The system is tested on academic dataset to explain the working of the model. The internal computation deals with the auto execution of the Sequential OLAP operations as in step 3a and 3b in the above fig 2. The query processing unit in the system model effectuates at the backend. Whenever a user inputs a query according to the requirement OLAP is executed at back end. To understand the functionality the internal computation is shown in figure 5. According to the demand an operation is performed or a combination of operations is executed.

Figure 5: Sequential OLAP operations processing

To test that whether the system works efficiently it has been tested for visualization as well. A test code is generated that works on the mappings of the technique described in [10]. The below figure 6 shows a static interface generated only for the academic dataset. The user first reads the dataset given as input to the system. The snippet of the dataset fetched is shown in the figure 7. As per the requirement the below steps are performed like visualize dropouts, visualize percentage, visualize locations, visualize graduates.

Figure 6: Visualization Test Code

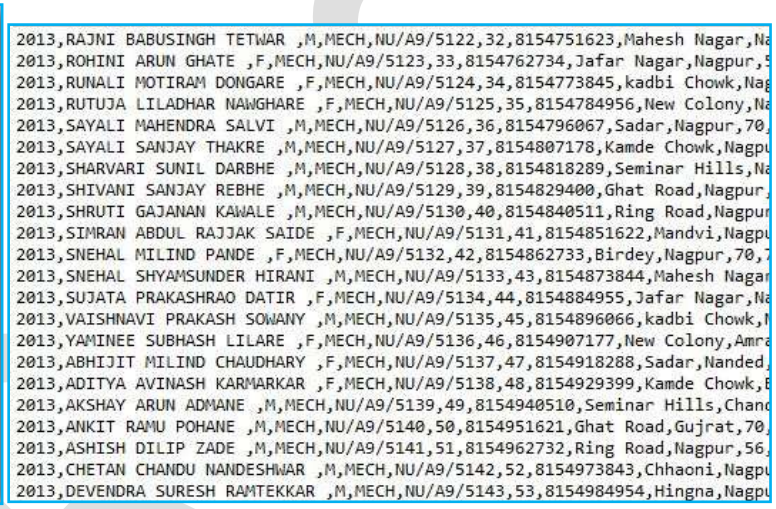

Figure 7 : Snippet of dataset fetched

If the user requires to visualize the dropouts the output generated is as shown in figure 7.

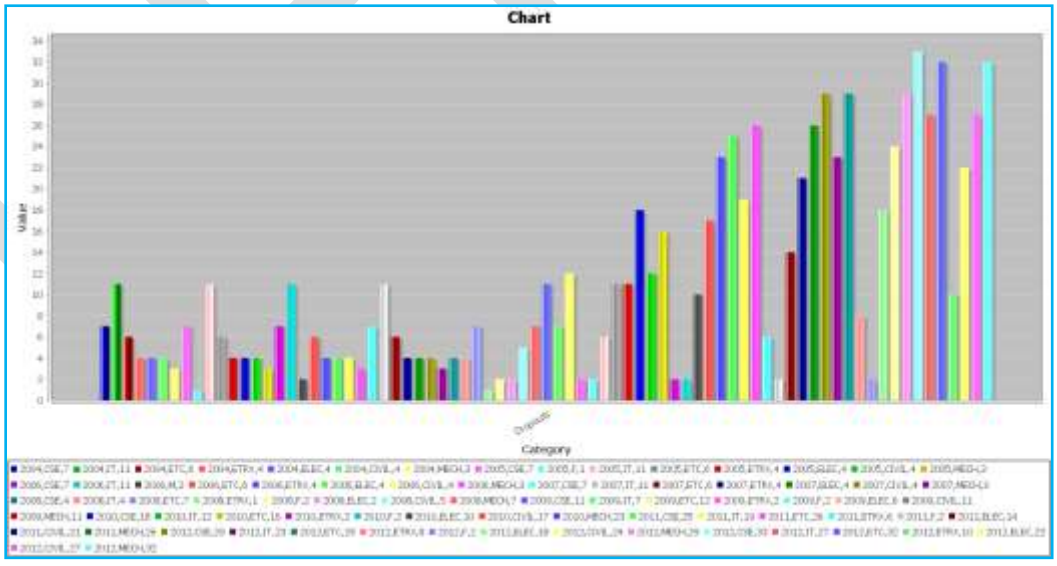

Figure 8: Output generated for dropouts visualization

The same can be generated for the rest of the options in figure 6 such as Visualize percentage, Visualize Locations and Visualize graduates. The working of the system involves the above described internal computations. The complete system will solve the problem of efficient visualization.

CONCLUSION AND FUTURE WORK

The system was created with a motivation so that the user can view the desired data in a single perception of an image. The system included two most important features including the sequential OLAP unit and the visualization done by using the complexity design approach. The work focused on the concept to accomplish the massive datasets so that it can match up to the pace of advancement and as well to present the data in a more human understandable form. The system's sequential OLAP was tested on various datasets such as academic dataset ranging from the year 2004 to 2014 including around 5500 records, biometric logs for around 50 employees for 14 months and toll tax data to check the performance of the system. After the successful accomplishment of the sequential online analytical processing, the system was tested to check for the visualization which also worked proficiently. The system can be used in academic institutes to manage their data and for the various visualizations generated for the documentation purpose. The complete visualization interface is yet to be achieved. Till now the system works for separate visualization, interlinking of data will be the next objective to be achieved.

REFERENCES:

[1] Muhammad Usman, Dr. Sohail Asghar Fong, S. "A Conceptual Model for Combining Enhanced OLAP and Data Mining Systems" NCM 2009 IEEE Fifth International Joint Conference on INC, IMS, and IDC , ISBN-number 978-0-7695-3769-6 , August 25th to August 27th , 2009.

[2] Ravat Franck, Teste Olivier, Tournier Ronan and Zurfluh Gilles "Algebraic and Graphic Languages for OLAP Manipulations" International Journal of Data Warehousing and Mining, Volume 4, issue 1, pg. no 17-46 , 2008.

[3] OLAP and OLAP server definitions , http://www.olapcouncil.org/research/glossaryly.htm

[4] Monica Caniupan, Alejandro Vaisman and Raul Arrendondo "Efficient repair of dimension hierarchies under inconsistent reclassification" Data & Knowledge Engineering, Volume 95, Pages 1-22, January 2015.

[5] Muhammad Usman, Dr. Sohail Asghar Fong, S "Data Mining and Automatic OLAP Schema Generation" ICDIM 2010, IEEE Fifth International Conference on Digital Information Management, July 5th to July 8th2010.

[6] Arian Bar and Lukasz Golab "Towards Benchmarking Stream Data Warehouses "In Proceedings of the fifteenth international workshop on Data warehousing and OLAP (DOLAP'12), ACM Conference on Information and Knowledge Management .

[7] Cuzzocrea, Alfredo, and Svetlana Mansmann "OLAP visualization: models, issues, and techniques." Encyclopedia of Data Warehousing and Mining, pg. no 1439-1446, IGI Publishing, 2009.

[8] Eric Lo, Ben Kao, Wai-Shing Ho, Sau Dan Lee, Chun Kit Chui, and David W. Cheung "OLAP on sequence data" In Proceedings of the 2008 ACM SIGMOD international conference on Management of data (SIGMOD '08). pg. no 649-660.

[9] Chui, Chun Kit, Ben Kao, Eric Lo, and David Cheung. "S-OLAP: an OLAP system for analyzing sequence data." In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, pg. no. 1131-1134. ACM, 2010.

[10] Risi, M.; Sessa, M.I.; Tucci, M.; Tortora, G. "CoDe Modeling of Graph Composition for Data Warehouse Report Visualization," IEEE Transactions on Knowledge and Data Engineering, vol.26, no.3, pg. no .563-576, March 2014.

[11] Tableau Software, www.tableausoftware.com

[12] Spagobi Software, www.spagobi.org

[13] Actuate Analytics Software, www.actuate.com

[14] S. Mansmann and M. Scholl "Visual OLAP: A new paradigm for exploring multidimensional aggregates" in Proceedings of IADIS International conference on Computer Graphics and Visualization, pg. no.59-66, 2008.

[15] Andreas S. Maniatis, Panos et al. "Advanced visualization for OLAP" In Proceedings of the 6th ACM international workshop on Data warehousing and OLAP (DOLAP '03), pg. no. 9-16 2003.

[16] M. Asif Naeem, Gillian Dobbie, Imran Sarwar Bajwa, and Gerald Weber "Resource optimization for processing of stream data in data warehouse environment" In Proceedings of the ACM International Conference on Advances in Computing, Communications and Informatics (ICACCI '12), pg. no. 62-68, 2012.

[17] Ladjel Bellatreche, Arnaud Giacometti et al. " A personalization framework for OLAP queries" In Proceedings of the 8th ACM international workshop on Data warehousing and OLAP (DOLAP '05pg. no. 9-18, 2005.

[18] Bhowmik, T. , Sarkar, A. and Debnath, N.C. " OLAP umbrella: Visualization model for multidimensional databases" IEEE/ACS International Conference on Computer Systems and Applications (AICCSA), pg. no. 1-8, May 16th to May 19th2010.