

# Data Warehouse Design Considerations for a Healthcare Business Intelligence System

Joseph George, *Member, IAENG*, B. Vijay Kumar, V. Santhosh Kumar

**Abstract—** This paper deals with the importance of Business Intelligence (BI) in healthcare and its vital perspectives. It is an effort to identify the difference between traditional BI approaches and the one needed in healthcare. It also talks about the unique nature of clinical data. When it comes to Business Intelligence, Data Warehouse (DW) can't be omitted. This paper is an effort to identify the answers of, why traditional BI concepts are not suitable for healthcare industry. It also gives specific directions towards the design of an optimal DW structure for healthcare domain.

**Index Terms—** Clinical Business Intelligence, Healthcare Information System Maturity Level, Data Warehousing in Healthcare, Decision Support in Healthcare, Business Intelligence in Healthcare

## I. INTRODUCTION

The rate at which data getting accumulated in healthcare domain is much more than any other industry. The sources of data are of wide variety and still the question remains opened, whether we are using the data in a proper way or not. Data warehouse is the answer for data handling, as everyone presume. But the applications of traditional DW methods in healthcare domain requires considerable attention, due to the unique business nature of the industry.

The business nature of healthcare industry makes it stand out from other industries. The data sources of healthcare industry varies from health records, pathology equipment to point of care devices and national registers. Most of the time, the healthcare institutions are unable to make use of the pile of data, they have accumulated. Clinical Business Intelligence is a term, which refers specially to the BI methods applied in healthcare industry.

Healthcare Information and Management Systems Society (HIMSS) is a global, cause-based, not-for-profit organization, focused on better health through information technology. HIMSS leads efforts to optimize health engagements and care outcomes using information technology. HIMSS defines maturity levels for healthcare information systems' usage and capability. The top maturity level demands the need of data warehouse and business

intelligence applications in the unit. As per the data from HIMSS Analytics Database 2014, even in the United States, only 3.1% organizations got eligible for the top maturity model. This indicates that, still the healthcare industry is lacking the facilities of data warehousing and business intelligence. The most prominent reason is the unique and complex nature of the industry. The application of traditional data warehousing methods, which we use in other industries, do not fit for the business intelligence needs of the healthcare domain. There are many Hospital Information Systems (HIS) available in the market and they all offer one or other mode of Manage Information System (MIS) functionality. Needless to say, most of them are still using the traditional pre-defined reporting models, based on the operational database.

## II. HEALTHCARE BI REQUIREMENTS

Strategic Information is essential to any industry. It is derived from a combination of historical and current data. While operational systems are essential for the day to day operations of the organization, data warehouse and the derived business intelligence provide the input for the strategic decision making. Time is always a dimension in this case.

Let us take an example of a manufacturing company. The information, which is of top management's interest, is well defined. The production, sales, purchase, debt etc. If we take another example from a hospitality industry, the facts remains almost same, with minor alterations. Now let us have a look into a healthcare firm. The administration requirements are almost same like any other industry like income, insurance claims etc. But the second aspect, which is of clinical and research, varies a lot. No well-defined matrices exist and the potential data sources are many. In such a scenario, traditional data manipulation techniques will satisfy the administration aspects. The clinical data, which is the key for research and treatment, needs a special approach.

The motivating factor for the proposed research work is, the unique nature of the healthcare data and its analysis requirements. Healthcare industry is now realizing the need of a proper BI infrastructure and moving towards it. There are BI tools available in health industry. But they also couldn't retain the unique nature of healthcare industry. If we look at the cost of these tools, they are not less than a few hundred thousand USDs, which makes it unattainable by many organizations, as the return of investment is too less. Figure 1 shows a traditional data warehouse steps. The dotted box are the areas, where the healthcare industry

Manuscript received March 15, 2015; revised April 08, 2015.

Joseph George is with Data Warehousing and Clinical Business Intelligence, Birla Institute of Technology and Science (BITS). (email: JG.JOSEPH@HOTMAIL.COM)

Dr. B. Vijayakumar is Head of Computer Science Department at BITS Pilani, Dubai Campus. (email: VIJAY@DUBAI.BITS-PILANI.AC.IN)

Dr. V. Santhosh Kumar is Assistant Professor of the department of Computer Science at BITS Pilani, Dubai Campus. (e-mail: SANTHOSHKUMAR@DUBAI.BITS-PILANI.AC.IN)

demands a unique solution. A good decision support system can not only improve the quality of care but also reduce the cost of improved care.

### III. DATA WAREHOUSE FOR HEALTHCARE DOMAIN

HIMSS, in its resource library by Paul Henchey, clearly defines Clinical business intelligence as the use of data analysis to improve care delivery using a wide category of information technologies, that gather data during the various

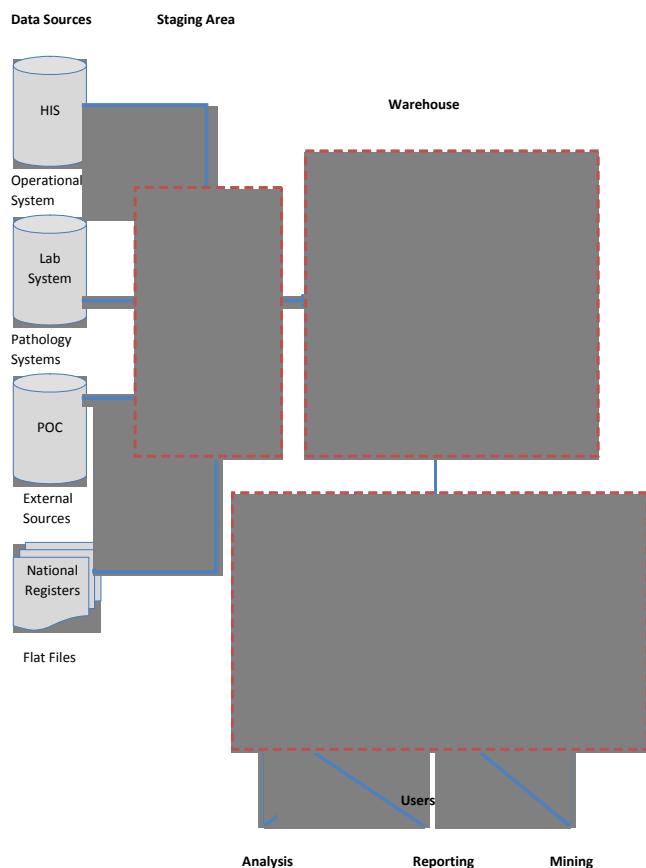


Fig. 1. Data Warehousing Architectural Diagram for Healthcare Application.

steps of delivery of care and analyze that clinical data, to make clinical and strategic decisions[1]. According to Tobias and Vivian, the margins achieved by healthcare organizations, who make use of integrated technologies is higher than that compared to its counterparts [2]. Still the question remains on the Return On Investment (ROI). Many organizations might not be capable of affording huge investments.

It is interesting to note the main characteristics of healthcare industry, pertaining to Business Intelligence, which make it stand out from other industries. Processes and Actors makes huge impact on this distinguished nature. Management's need, for most of the sectors, are same. They are have well defined goals. But in healthcare, there are mainly two dimensions: namely, clinical and administrative. Most industrial DW and BI has well defined hard matrices, but in healthcare patients feelings itself is a matrix [2].

When we analyze the actors /users of any BI system, there are a clear group of customers to the application. But in healthcare the actors are multiple and the needs also vary a lot[2]. According to Paulraj Ponniah, the frequency of the usage in DW system is less compared to an operational

system[4]. But when we need to use a healthcare BI and DW, to make clinical decisions, for example in the case of value based healthcare, the frequency of usage is not as less, compared other DW applications, and thus we need to assure that the healthcare DW/BI response is not affected by the design.

BI is always a challenging topic and exercise in any industry. But when it comes to healthcare the complexity increases as there are added layers of issues such as privacy issue. Healthcare industry is always dealing with sensitive data which is governed by privacy rules [3]. As Hanson stated, the healthcare industry is going through ever increasing pressure, which demands more with less[5]. This scenario requires an essential use of the existing information, which is spread in both clinical and administrative directions.

Jonathan Palmer, senior director for clinical warehousing and analytics at Oracle, describes clinical data warehouse as a mission critical hub. From being a data gathering and analytics tool, Clinical Business Intelligence is moving to a new era, to become a business critical platform[6]. When we discuss about a framework for DW and BI for healthcare organization, the first consideration is the comments from Scott Wanless and Thomas Ludwig. They are pointing out five challenges which need to be addressed by healthcare data analytics[7], namely:

- Demand
- Resource Shortage
- Compliance requirements
- Financial pressure
- Integration issue

Star schema, or in another words the dimensional modeling, is the solution we might need to implement in healthcare domain as well. Dimensional model of any business process has two distinct aspects, measurement and its context. Or in more data warehousing terms, facts and dimensions[8]. But the level of aggregation and the level of flaking needs to be decided.

Business Intelligence can never be treated isolated from Data Warehouse. When we apply BI in healthcare industry, several factors to be considered. These same factors we might be using in other industries as well. According to Pungrat Jinpon, there are six major considerations[9].

- The matrices we are choosing must meet the desired goal of the project
- The selection of data sources. As we mentioned earlier, there are hundreds of potential data sources available for healthcare DW. The method deployed to clean the data also place a vital role.
- The architecture which we are choosing, one of my favorite direction, must be suitable for the domain.
- The data load frequency and the method of update activity.
- Spot out the users and actors
- Security and permissions.

Mostly the healthcare / biological DW is based on a star schema. Star and Snow flakes are widely used in DW

applications. But as mentioned above, the unique nature of healthcare industry deserve something better. Torben Back Pedersen proposed a multi dimensional schema, an extended form of star, for healthcare DW [15]. The clinical data space is huge and among them, obviously, Patient is the central point of interest. There are situations, where might analyze the situation in a different way, say provider centric. For example, to concentrate on the productivity angle of an organization, the resources(who provides care) will become the center of attraction.[14]

Carol Muirhead and Brian Gaffney in their research paper on how and where to apply BI, provides a reference model for a clinical Data warehouse. A generalized structure, which can be used as starting guide for defining a clinical DW[24]. This reference model consists of 6 components namely:

- 1) Data Source
- 2) Data Intake
- 3) Database
- 4) Security
- 5) End user analytical tools
- 6) Analytical use cases.

#### IV. PROPOSED SYSTEM

It is clear from the analysis of current scenario that , even though we are talking about Clinical Business Intelligence in Data Warehouse context, there is a considerable difference between traditional Business intelligence applications and Clinical Business Intelligence. There is no well defined structure and framework for Clinical Business Intelligence yet. Still we are using the same techniques and structure, used in other industries, for healthcare as well. We need innovative methodologies, which can satisfy the increasing demand of BI in healthcare industry which is catering all the different dimensions and actors[12].

The proposed research work identified the following gaps, related to the Data Warehousing in Clinical Business Intelligence context.

- 1) In healthcare analytics, there are no well defined matrices. This makes the DW design tedious and ambiguous. Measuring facts are getting regenerated and dynamically arrives. Simple dimensional modeling gets a hit back in this scenario.
- 2) In healthcare data, two major aspects or classes of analysis are clinical and administration. These two aspects acts in a parallel line. This has a direct impact on the underlying architecture.
- 3) Traditional hospitals and specialized centers differ a lot in measuring matrices or facts.
- 4) An additional layer of complexity in terms of data confidentiality exists in healthcare data.

The above gaps will be addressed in the proposed research work, as mentioned below.

- 1) This will be handled by a modified star schema of dimensional modeling.
- 2) Administration and clinical directions, need to be

treated separately. Administration requires Star based scheme. Clinical track will be handled by the modified schema as mentioned in point 1. A hybrid approach need to be employed for the interoperation.

- 3) An SOP will be defined for standardizing the entire BI process.
- 4) A modified ETL procedure will be employed. Staging area will be treated with utmost confidentiality aspect.

#### V. FUTURE WORK

Researcher is trying to elaborate and do individual studies on all the mentioned points in the paper. Different approaches might be needed in different scenarios. But if we can come up with a standardized approach, this will help the healthcare industry a lot. Lack of standard operational procedure for a well define BI, this is the aim of the project.

#### VI. CONCLUSION

Value based healthcare(VBH),e-health, m-health, these are the current and future directions of healthcare industry. This will increase the data volume to the next level. The information from the collected data, with utmost accuracy and the best possible response time, this is what everyone needs. The proposed research work will significantly contribute to these directions by its objectives.

#### ACKNOWLEDGMENT

I would like to thank Mr.Bengt Ternstrom, CEO-Rashid Centre for Diabetes and Research (RCDR) for his support and encouragement. This project would not have been possible without the inputs and feedback from Dr.Salah Abusnana, CMO and Mr.Benil Thambi, Systems Admin – RCDR.

#### REFERENCES

- [1] Paul Henchey. "HIMSS Clinical & Business Intelligence Primer" 2013 April [cited 2014 September 10]. Available from: <http://www.himss.org/library/clinical-business-intelligence/clinical-business-intelligence-primer/what-is-clinical-business-intelligence>.
- [2] Tobias Mettler, Vivian Vimarlund. "Understanding Business Intelligence in the Context of Health Care" Proceedings of the 13th International Symposium for Health Information Management Research (ISHIMR 2008), At Auckland, New Zealand,2008,Oct 20-22,61-69.
- [3] Noushin Ashrafi, Lori Kelleher, Jean-Pierre Kuilboer. "The Impact of Business Intelligence on Healthcare Delivery in the USA" Interdisciplinary Journal of Information, Knowledge, and Management. 2014,9:117-127.
- [4] Paulraj Ponniah. "Data warehousing Fundamentals for IT Professionals",2nd ed.Wiley,Hoboken, New Jersey,USA.2010,10-15.
- [5] Hanson. R. "Good health information – An asset not a burden!" AustralianHealth Review .2011, 35(1): 9–13.
- [6] 6. Jonathan Palmer. "The Clinical Data Warehouse – a New Mission-Critical Hub" Oracle Health Sciences Journal. 2013,1:5-12.
- [7] Scott Wanless , Thomas Ludwig. "Business Intelligence and Analytics for Healthcare Organizations",1st ed. Ark Conference Group,London,UK.2011,5-9.
- [8] Christopher Adamson."Star Schema – The Complete Reference",1st ed. The McGraw-Hill Companies,New York,USA.2010,250-260.
- [9] Jinpon P., Jaroensutansinee M., Jaroensutansinee K. "Business intelligence and its applications in the public healthcare system" Walailak Journal of Science and Technology.2011, 8(2):97-110.

- [10] Koh H. C., Tan G. "Data mining applications in healthcare" *Journal of Healthcare Information Management*.2011, 19(2): 64-72.
- [11] Kimball ,Ross . "The Data Warehouse Toolkit - The Complete Guide to Dimensional Modeling",2nd ed.Wiley, ,Hoboken, New Jersey,USA.2002,255-275.
- [12] Alfredo Cuzzocrea,Ladjel Bellatreche,Il-Yeol Song. "Data warehousing and OLAP over big data: current challenges and future research directions" *Proceedings of 22nd ACM International Conference on Information and Knowledge Management (CIKM '13)*, San Francisco, CA, USA,2013, October 27 - November 01,67-70.
- [13] Tony R., Sahama,Peter R. Croll. "A Data Warehouse Architecture for Clinical Data Warehousing", *Proceedings of First Australasian Workshop on Health Knowledge Management and Discovery (HKMD 2007)*, Ballarat, Australia,2007, January 10 -13,227-232.
- [14] Nan Du, Suxin Guo, Supriya D. Mahajan, Stanley A. Schwartz, Bindukumar B. Nair, Chiu Bin Hsiao, Aidong Zhang. "BioStar+: a data warehouse schema for integrating clinical and genomic data from HIV patients" *ACM SIGBioinformatics*.2012,2(3):6-16.
- [15] Torben Bach Pedersena,Christian S. Jensen, Curtis E. Dyresonb. "A foundation for capturing and querying complex multidimensional data" *Pergamon Information Systems*.2001,26:383-423.
- [16] Inmon W. "Building the Data Warehouse", 3rd ed.,Wiley,Hoboken, New Jersey,USA.2002,25-30.
- [17] Kimbal.R., Reeves. L., Ros. M, Thronthwaite. W. "The Data Warehouse Lifecycle Toolkit", 1st ed.,Wiley,Hoboken, New Jersey,USA.1998,52-61.
- [18] Steven A. Demurjian, Michael Blechner, Rishi Kanth Saripalle. "A proposed star schema and extraction process to enhance the collection of contextual & semantic information for clinical research data warehouses",*Proceedings of the 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*,EEE Computer Society Washington, DC, USA 2012,October 7-10,798-805.
- [19] Riccardo Lora, Alberto Sabaini, Carlo Combi, Ugo Moretti."Designing the reconciled schema for a pharmacovigilance data warehouse through a temporally-enhanced ER model",*Proceedings of the 2012 international workshop on Smart health and wellbeing (SHB '12)*,Maui, Hawaii, USA.2012,October 29,17-24.
- [20] James P. McGlothlin, Latifur Khan."Managing evolving code sets and integration of multiple data sources in health care analytics",*Proceedings of the 2013 international workshop on Data management & analytics for healthcare(DARE '13)*,San Francisco, CA, USA.2013,November 1-3,9-14.
- [21] Abubakar Ado, Ahmed Aliyu , Saifullahi Aminu Bello , Abdulra'uf Garba Sharifai , Abubakar Sulaiman Gezawa."Building a Diabetes Data Warehouse to Support Decision making in healthcare industry" *IOSR Journal of Computer Engineering (IOSR-JCE)*.2014,16(2):138-143.
- [22] P. DI BITONTO, F. DI TRIA, T. ROSELLI, V. ROSSANO, F. TANGORRA."A Data Warehouse in an E-Health System",*Proceedings of the 3rd WSEAS International Conference on Biomedicine and Health Engineering (BIHE '14)*, Tenerife, Spain, 2014, January 10-14,87-92.
- [23] David Stodder."This article explores important BI and analytics technology trends and how organizations are capitalizing on them" *The data warehouse institute Journal(TDWI Research)*.2013,3:2-9.
- [24] Carol Muirhead,Brian Gaffney. "Clinical & Business Intelligence: An Analytics Executive Review.Industry Capabilities – Available Tools, How & Where Applied" 2013 February[cited 2014 September 5] Available from:  
[http://www.himss.org/files/HIMSSorg/Content/files/HIMSS\\_CBI\\_Analytics\\_Exec\\_Review\\_Building\\_the\\_Business\\_Case\\_module\\_2013-02-19\\_FINAL.pdf](http://www.himss.org/files/HIMSSorg/Content/files/HIMSS_CBI_Analytics_Exec_Review_Building_the_Business_Case_module_2013-02-19_FINAL.pdf)
- [25] Wided Oueslati,Jalel Akaichi. "A survey on data warehouse evolution",*International Journal of Database Management Systems ( IJDMs )*.2010,2(4):11-24.

## Bibliography

Joseph George is currently pursuing his Doctor of Philosophy (Ph.D) in Data Warehousing and Clinical Business Intelligence from Birla Institute of Technology and Science (BITS). His research interest includes Clinical Analytics and Cloud Technologies. He is a trained quality and project management professional and holds various technology certifications. He is the Head of Information Technology at Rashid Centre for Diabetes and Research (RCDR), Ministry of Health UAE. He was appointed to his present position in November 2009 and was one of the core team members during RCDR's inception. He started his career in Information Technology fifteen years ago, and has held lead and senior managerial positions in Software Development, IT Architecture and Hospital Information Systems. Starting his career as a Software Developer and Database Engineer, he then moved into Business Intelligence. He received his Master of Engineering degree from BITS and Master of Science in Information Technology from Manipal University. email: JG.JOSEPH@HOTMAIL.COM

Dr. B. Vijayakumar is Head of Computer Science Department at BITS Pilani, Dubai Campus. His research interests include Multimedia Systems, Web Data Mining, Component Based Software Development and Distributed Database Systems .He has 30+ publications in referred journals and conferences. He has served as technical program chair for the International Conference on Cloud Computing Technologies, Applications and Management-ICCCTAM-2012 held at BITS Pilani, Dubai Campus in December 2012 and reviewer for the conference SKM-2014 at BITS Pilani, Dubai Campus held in Dec 2014. email: VIJAY@DUBAI.BITS-PILANI.AC.IN

Dr. V. Santhosh Kumar is currently working as Assistant Professor in the department of Computer Science at BITS Pilani, Dubai Campus. His research interests include data warehousing, and performance evaluation of computer systems. e-mail: SANTHOSHKUMAR@DUBAI.BITS-PILANI.AC.IN