

FOUNDATION STEPS FOR A WELL-DEFINED CLINICAL BUSINESS INTELLIGENCE SYSTEM - AN ANALYSIS OF DATA WAREHOUSE APPROACHES WITH RESPECT TO HEALTHCARE DATA

Joseph George *

Department of Computer Science and Engineering,
Noorul Islam Centre for Higher Education Kumaracoil, Tamilnadu, India.
jg.joseph@hotmail.com
<http://www.niuniv.com/>

Dr. M.K Jeyakumar

Professor, Department of Computer Applications,
Noorul Islam Centre for Higher Education Kumaracoil, Tamilnadu, India
<http://www.niuniv.com/>

Abstract - Turing data into information and further to knowledge and wisdom is essential for any organization to perform better in today's era. Healthcare industry is not an exception, but it demands more of data transformation than any other industry. "Analysis Paralysis" is the term, which is often used to relate or define the data chaos situation. Clinical Decision Support and Clinical Business Intelligence are the approaches which capitalize the wealth amount of the data collected. Consolidation of fragmented data from different sources to have a holistic view and analysis is the fundamental idea behind Data Warehouse. This intention behind this research is to compare and analyze the suitability of different Data Warehouse architectures for healthcare domain focusing on the data utilization output, such as healthcare Business Intelligence and Clinical Decision Support.

Keywords: Data Warehouse, Healthcare Business Intelligence, Decision Support Systems

1. Introduction

Data growth is in a multi exponential path now a days. As per the statistics compiled by Dell EMC, healthcare organizations are witnessing near about 878% of data growth rate in the last few years[Dell Technologies, (2019)]. Digital transformation and technology enabled care delivery are among the top priority list of all modern healthcare organizations, to improve process efficiency and to reduce operational cost, resulting in improved and quality care delivery.

Healthcare uses a handful of systems for operational activities starting from HIS (Hospital Information Systems) for clinical activities to ERP (Enterprise Resource Planning) for back office operations. Apart from these, the point of care devices and associated data also play a vital role in the care delivery cycle.

One of the biggest challenges in healthcare is the integration of all the collected data to infer intelligence for efficient care delivery. Here comes the role of Clinical Business Intelligence (CBI), which consolidates and analyze the various operational and clinical data captured at various touch points of care delivery process to produce informed decisions. Business intelligence utilizes the historical data and can provide predictive views to improve the operational efficiency of the organization. Similar way Clinical Decision Support (CDS) systems heavily depends on the refined historical data to provide support for clinicians to take clinical decisions based on evidences.

The core of CBI and CDS is a well architected Data Warehouse (DW), which transforms fragmented operational data from different applications to a well formulated repository. Data warehouse provides integrated and consistent information across various subjects, often referred to as "subject oriented". It supports the conventional reporting and ad-hoc reporting as well [Inmon et al., (2008)].

Top down approach by Bill Inmon and Bottom up approach by Kimball are the prominent methodologies exist in DW spectrum. Data Vault is the third approach by Linsterdt but it is based on Inmon's approach. Thousands of articles exist which compare and point out the differences between these doctrines.

Many Healthcare DW implementations got either failed or closed down, and if you analyze the reasons behind it, it reveals few key elements such as missing of a solid business imperative and the planning and selection of the implementation methodology [Barlow, (2014)]. There exists various DW development approaches and it is mandatory to choose the right approach based on the data volume and business requirements to have a successful DW project. The selection of the modelling approach must consider various parameters like project management, return of investment, deployment methods etc. [Yessad et al., (2016)].

HIMSS Analytics' EMRAM maturity model acknowledges the importance of DW as one of the most important requirements for the highest maturity level (Level 7) achievement. The guideline says "Data warehousing is being used to analyze patterns of clinical data to improve quality of care, patient safety, and care delivery efficiency." [HIMSS, (2019)]

2. Methods

In this section, first we will analyze how healthcare data is unique and distinct. Secondly, we will define the DW concept, followed by an analysis of the widely used DW frameworks with respect to healthcare.

2.1. Why healthcare data is unique?

Healthcare industry is unique compared to any other industry in many ways and thus the data and processes associated with it too [Lesueur, (2017)].

Health information for decision making is generated from the combination of the atomic and fragmented data, spread across the pathway of care delivery process. For better decisions, we need better information. In a hospital environment, we can broadly classify the data generated into 3 categories – Clinical Data – Administrative Data– Machine Generated Data (a subset of clinical Data). Clinical data can be categorized again into the below categories[Washington, (2018)].

- (1) Electronic health records
- (2) Process data
- (3) Claims data
- (4) Patient / Disease registries
- (5) Health surveys
- (6) Clinical trials data

2.1.1. Spread, variety and complexity:

Single comprehensive healthcare record is the vision of almost all healthcare regulatory bodies and organization. But still the achievement of the target is far away for many organizations. Consider a patient encounter on a specific day. The data associated with that specific visit lies in different fragmented systems (often), in various format like videos, waves, DICOM, structured, unstructured etc.[Lesueur, (2017)]. The no of variables and its format involved in a care delivery process are enormous.

2.1.2. Changing Data definitions and regulations

Healthcare is the industry bound to most no of regulatory guidelines and rules, which changes in a very fast pace. Based on the changes in regulatory clauses and thus the data definitions, the corresponding data produced also getting changed.

2.1.3. Context dependency.

In healthcare data definition, context awareness comes before the value of the parameter. [Corbin, (2014)]. Data without context definition is not at all usable here. The best example to back this statement is the measure of systolic and diastolic blood pressure (BP) values. The tolerance value or the accepted range varies based on the medical condition of the patient. The accepted BP range of a patient who is on Metformin could be in a different level than the one who is on non-medication. A value of 140/90 has different meanings for a patient who is already on a blood pressure lowering medication and for another patient who is not on any medication. Data without proper context is invalid here.

2.2. The Data Warehouse

The concept of DW was introduced by Bill Inmon, known as the Father of DW. He introduced the term 'CIF- Corporate Information Factory'. The term was first introduced in his seminal work "Building the Data Warehouse"[Inmon., (2002)]. The dimensional model was proposed by Ralph Kimball through his book "The Data Warehouse Toolkit" [R, Kimball;R, (2002)]. Inmon and Kimball are known to be the authentic words of DW.

The objective of any DW is to facilitate informed decision making, irrespective of the framework it is built on. The six types of data systems mentioned above are coming under the transaction system category, which is meant for day to day operations. In other words, it is designed to record, not analyze [Ponniah, (2001)]. These systems are made in such a way that, it is response intensive and optimized for real-time transactions. For example EMR (Electronic Medical Record), CPOE (Computerized physician order entry), LIS (Laboratory Information System) etc. are designed in such a way that, it is providing quick patient information to the caregiver, specific to a particular patient in no time.

A definition of DW from Inmon is “Data Warehouse is a subject oriented, integrated, nonvolatile and time variant collection of data in support of management’s decisions.” [Inmon., (2002)]. In a healthcare perspective, “subject” means the specific business processes or care activities performed with respect to specific care delivery areas. To make it simple, ED (Emergency Department) activities can be taken as a subject. Another example could be long term treatment like Diabetes treatment and related outcomes. (We will consider these are Data Marts, which are the miniature of DW, in later sections). The word “integrated” means, the consolidation of data from different fragmented systems to get the single version of truth.

3. A comparison of DW models for Healthcare

3.1. Inmon’s top down approach

Inmon Approach also known as Corporate Information Factory(CIF) is based on a holistic data approach of the entire organization[Breslin, (2004)]. A holistic “data driven” approach is essential here. Inmon propose a 4-level data environment. They are namely “Operational Data”, “Atomic DW”, “Data Marts” and “Visualization”. The individual Data Marts are fed from the atomic DW.

The atomic DW is in 3rd normal form in a relational data base, which contains the entire data of the organization, fed from different operational systems integrated through an ETL (Extraction Transformation Loading) and a staging area. Inmon makes it as a data driven approach than a user or process driven. User involvement will come only when the data marts are generated from the DW. This model is based on the belief that, the data should be stored in relational model till the moment it reaches the authorized user[Breslin, (2004)]. Let us see how the 4 layers of CIF is mapped to a healthcare environment (Fig. 1)

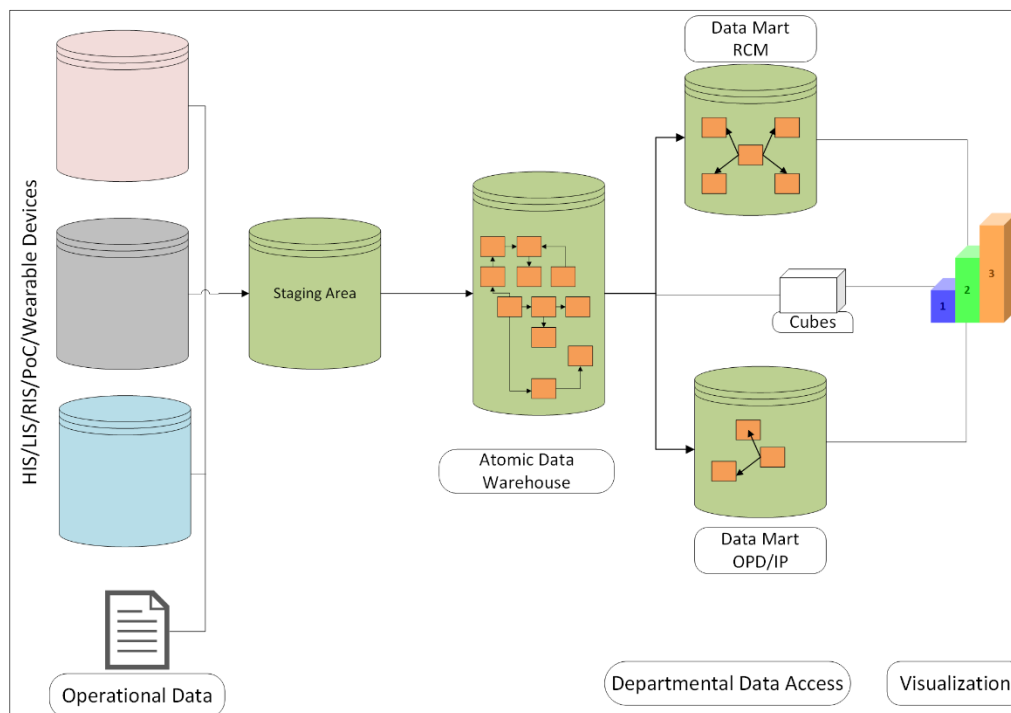


Fig. 1. Healthcare Data Warehouse with respect to CIF

Operational Data:

HIS, LIS, RIS, PACS, ERP, Mobile Application: all these front facing systems comes in the operational data layer of healthcare. Most of the cases, these systems could be either independent or integrated. Since CIF is proposing a data intensive model, during the ETL process, data from all these systems need to be scrubbed and moved to the staging system and then to DW[Hamoud, (2018)].

Atomic DW:

This is the consolidation layer which holds the transformed data from the operational systems. (Example, your HIS could be Cerner/Epic/Intersystem etc. and the ERP could be SAP/Infor/Oracle etc. Data fetched from all these systems resides in this DW with a common data dictionary).

Departmental Data Marts:

An analysis of claims and revenue could be one of the primary interests of the RCM and finance team. To perform business intelligence activities on those “subjects”, the team would be needing a subset of the DW (data mart) which will be created from the Atomic DW. End user will never have access to the Atomic DW. They will be interacting with data only through the generated data marts. That’s why CIF is known as top down approach. Another examples of departments data marts could be Lab Data Mart, OPD, IP Data Mart etc.

Individual Data:

This represents the reports generated from the respective data marts. For example, an analysis of LOS, OPD waiting time etc.

3.2 Kimball’s bottom up approach

We discussed in previous section that, top down approach believes that data should remain in 3NF until it reaches the end user. But in Kimball’s approach, he overwrites all these concepts by introducing the dimensional modelling. In simple words, data is made redundant in many instances to provide increase in performance and to hide the data complexity from end users.

Dimensional model contains 2 key elements, facts and dimensions. Fact is the event, or the measurable element and dimensions are the slicing and dicing angles. Facts are connected with Dimensions and forms the corresponding Data Mart. Star schema and Snowflake schema are the most popular among dimensional modelling. Fig.2 represents Kimball model made on a healthcare environment.

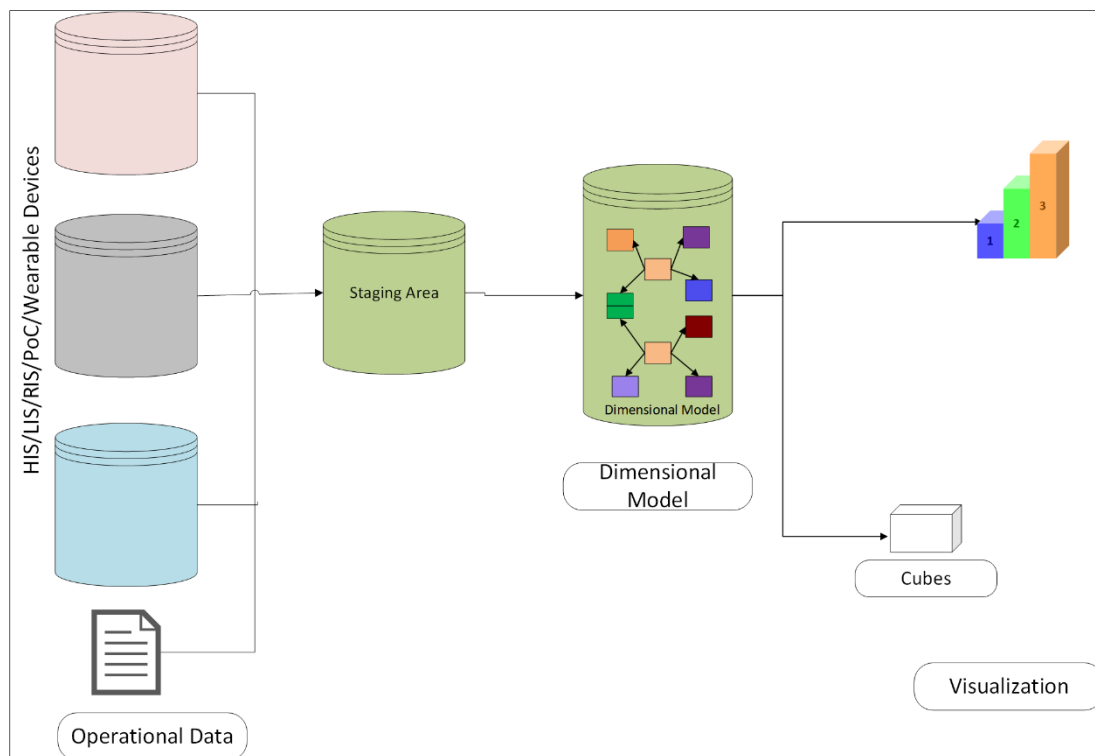


Fig. 2. Healthcare Data Warehouse with respect to bottom up approach

Kimball model is deeply rooted in the involvement of business users from the very beginning. It is highly business process oriented.

Kimball's bottom up approach is essentially an agile development method. Data marts are created and eventually all the data marts together formulate the enterprise DW. Let's say, we start with the OPD analysis Dimensional model, followed by an RCM model there after an IP analysis and eventually once completed, the enterprise organization wide DW is created.

Bottom up modelling consists of 4 steps for the designing of the dimensional model[R, Kimball;R, (2013)]. They are

- Selection of the Process (Business Process or the subject of interest)
- Grain selection (Decision on to what extend data to be granularity saved)
- Decide the facts (the measurable elements or the events, or even the fact less facts)
- Confirm the dimensions

Let us equate this for the creation of an RCM Data Mart

- Selected business process is Revenues cycle management. We would like to analyze the rejection factors and to analyze the revenue.
- To analyze the data, selected grain shall be individual care item and bills.
- Fact is the encounter
- Dimension shall be based on the workflow. For example, Provider/Payer/Clinician/Patient etc.

4. Results

4.1. Compare the models

Inmon and Kimball, both agree on the fundamental need for a consolidated DW to facilitate a holistic data view, although the frameworks are different. Both frameworks acknowledge the importance of ETL, which will accumulate data from various operational systems and the utilization of timestamps. Mary Breslin categorize the difference between Inmon and Kimball into 3 categories based on[Breslin, (2004)]:

- The methodologies and Architecture
- The way how data is modelled
- The philosophical approach on the implementation of the DW

Let's us compare the models from a healthcare perspective now.

4.1.1. Process/Data Orientation

Healthcare industry is a very much process-oriented environment. Every provider tunes the processes in order to make the output the best. The same output could be produced in another organization through another processes. For example, have witnessed different organizations tuning the OPD appointment process in different ways to reach the same goal – “reduced waiting time.”

CIF model is not favoring process-oriented approach, it is a data driven methodology. Bottom up model is in line with this and it is completely a business process-oriented approach. Business users are involved from the very beginning for the business intelligence solution. But the issue here is the enterprise data reporting. Bottom up model is oriented towards processes than the enterprise as a whole, CIF model the other way.

4.1.2 Enterprise/Agile/Data redundancy

The CIF is a very complex model which involves the design and maintenance of the huge centralized DW. In this model, it is necessary that we decide on all the data elements in advance, which we might use once the DW is implemented. Healthcare industry is very dynamic and drawing the DW data elements beforehand is something like a mission impossible. The reason is the specific nature of the data and the changing regulatory requirements around the healthcare industry.

Kimball model is very much modular. It gives the freedom to start very small scale and build from it. Like, start with an OPD mart then extend to IP then center of excellences like rehabilitation, diabetes etc.

Inmon's enterprise view ensure that we have the single version of truth always and there is no data redundancy. In top down approach, the departmental marts are fed from the atomic EDW. That means, the Data Mart's created say OPD and IP, there is no data duplication and no chance of data integrity issue.

In Kimball's approach, the flavor of the 'single source of truth' could be lost as data is not fully integrated before it arrives the end user for analysis. Kimball model is facilitated with redundant data which can cause data update issues over time [Rangarajan, (2016)].

4.1.3 Accommodation of changing requirements.

One of the healthcare specific DW model called “Late Binding Model” by health catalyst is stressing on the need of late binding of data in the DW. Both top down and bottom up model are mapping data into a predefined model in the DW. This leads to an inability to make changes in the data model. Industries where fixed Facts and dimensions exists, this is not an issue. In healthcare business, which undergoes constant changes, early binding results in performance and capability issues.

5. Discussions and Conclusion:

We discussed the pros and cons of various DW approached with respect to healthcare. It is clear that the traditional DW and Business Intelligence approaches are not well suited for healthcare industry to adopt as it is.

If we consider industries like manufacturing, aviation, banking etc., the data analytics requirements are well defined, and top down approach or bottom up approach could be deployed as it is in their contexts. In healthcare, we lack that luxury and the industry is very dynamic and volatile.

Healthcare has two main directions, clinical is one and the other is administrative and commercial. Both these two directions to be catered equally, with respect to data analytics needs while adopting the framework.

This is leading to the requirement of a healthcare specific Data Warehouse architecture, which can cater all aspects of the healthcare industry. A combination of the existing frameworks concentrating more on the healthcare industry's specific demands is the solution. For example, the well-defined features such as RCM and financial aspects could be explored with top down model and the agile business process with a bottom up approach.

5.1. Future work of new style

Based on the conclusions made, A layered data warehouse approach with high level of integrity and holistic data view is something which healthcare demands for its long run. Authors are working on that directions to arrive into a healthcare specific DW model.

Acknowledgements

We would like to thank Mr. Ramachandran and Mr. Gopakumar for their support and motivation throughout the research work.

Conflict of Interest:

The authors confirm that there are no known conflicts of interest associated with this paper.

References

- [1] Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis*, 16(1), 3–9. doi: 10.5840/du2005155/629
- [2] Barlow, B. S. (2014). 6 Reasons Why Healthcare Data Warehouses Fail. Retrieved from <https://www.healthcatalyst.com/why-healthcare-data-warehouses-fail/>
- [3] Breslin, M. (2004). Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models. *Business Intelligence Journal*, 1(1), 6–20. Retrieved from [http://olap.it/Articoli/Battle of the giants - comparing Kimball and Inmon.pdf](http://olap.it/Articoli/Battle%20of%20the%20giants%20-%20comparing%20Kimball%20and%20Inmon.pdf)
- [4] Corbin, K. (2014). What Does All That Healthcare Data Really Mean? | CIO. *CIO.Com*. Retrieved from <https://www.cio.com/article/2854599/what-does-all-that-healthcare-data-really-mean.html>
- [5] Dell Technologies. (2019). Solutions for Healthcare.
- [6] Hamoud, A. A. K. (2018). CLINICAL DATA WAREHOUSE: A REVIEW. *Iraqi Journal for Computers and Informatics*, 44(2), 1–11.
- [7] HIMSS. (2019). Electronic Medical Record Adoption Model | HIMSS Analytics - Middle East. HIMSS. Retrieved from <https://www.himssanalytics.org/middle-east/electronic-medical-record-adoption-model>
- [8] Inmon, W. H. (2002). Building the data warehouse. In Wiley (3rd ed., Vol. 41, Issue 9). 605 Third Avenue, New York: Wiley Computer Publishing. Retrieved from <http://portal.acm.org/citation.cfm?doid=285070.285080>
- [9] Inmon, W. H., & Neushloss, D. S. and G. (2008). DW 2.0. In Morgan Kaufmann (Vol. 43, Issue 6). Burlington, Massachusetts: Morgan Kaufmann. doi: 10.1016/B978-0-12-374319-0.X0001-2
- [10] Lesueur, D. (2017). 5 Reasons Healthcare Data Is Unique and Difficult to Measure. *Health Catalyst*, 1(1), 1–6. Retrieved from <https://www.healthcatalyst.com/wp-content/uploads/2014/08/5-Reasons-Healthcare-Data-Is-Unique-and-Difficult-to-Measure.pdf>
- [11] Ponniah, P. (2001). Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals (Vol. 6). 605 Third Avenue, New York: John Wiley & Sons.
- [12] Pusic, M. (2005). Clinical decision support systems | *British Columbia Medical Journal*. *BC Medical Journal*, 46(5), 236–239. Retrieved from <https://www.bcmj.org/articles/clinical-decision-support-systems>
- [13] R, Kimball; R, M. (2002). The Data Warehouse Toolkit - Second Edition (2nd ed.). 605 Third Avenue, New York: Wiley Computer Publishing. Retrieved from [http://www.dsc.ufcg.edu.br/~sampaio/Livros/alph Kimball. The Data Warehouse Toolkit.. The Complete Guide to Dimensional Modelling \(Wiley,2002\)\(ISBN 0471200247\)\(449s\).pdf](http://www.dsc.ufcg.edu.br/~sampaio/Livros/alph%20Kimball.%20The%20Data%20Warehouse%20Toolkit..%20The%20Complete%20Guide%20to%20Dimensional%20Modelling%20(Wiley,2002)(ISBN%200471200247)(449s).pdf)
- [14] R, Kimball; R, M. (2013). The Data Warehouse Tool Kit - Dimensional Modelling (3rd ed.).
- [15] Rangarajan, S. (2016). Data Warehouse Design – Inmon versus Kimball. [Http://Tdan.Com](http://Tdan.Com). Retrieved from <http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>
- [16] Warner, & Diana. (2013). IG 101: Developing Clinical Business Intelligence. *Journal of AHIMA*, 1(1), 1–10.
- [17] Washington, U. of. (2018). Clinical Data. Data Resources in the Health Sciences. Retrieved from <https://guides.lib.uw.edu/hsl/data/findclin>
- [18] Wasylewicz, A. T. M., & Scheepers-Hoeks, A. M. J. W. (2019). Clinical Decision Support Systems. In *Fundamentals of Clinical Data Science* (pp. 153–169). Cham: Springer International Publishing. doi: 10.1007/978-3-319-99713-1_11
- [19] Yessad, L., & Labiod, A. (2016). Comparative study of data warehouses modeling approaches: Inmon, Kimball and Data Vault. 2016 International Conference on System Reliability and Science (ICSRS), 95–99. doi: 10.1109/ICSRS.2016.7815845