An Integrated Business Intelligence Framework for Healthcare Analytics

Prof. Dr Ayman Khedr, Prof. Dr Sherif Kholeif, Fifi Saad
Faculty of computers and information systems, Information systems dept., Helwan University, Cairo, Egypt

DOI: 10.23956/ijarcsse/SV7I5/0163

Abstract— Business Intelligence (BI) is a set of tools, applications, and techniques used to help organizations taking the right actions and decisions. Analytics that consolidate lab information with finance, pharmacy, radiology, and other offices might assume a critical part in the vital arranging necessary should move forward those caliber from claiming mind Also diminish Generally speaking expenses. Using BI for enhancing the analytics of healthcare sector will help improve the results obtained from such systems. This paper proposes an integrated BI framework for healthcare analytics. This framework composed of six tiers; data source, Extract Transform and Load (ETL), Data storage, analytics, optimization, and presentation tier. Moreover; the paper presents an approach for the analytics tier that will provide better results regarding data analytics.

Keywords— Business intelligence, healthcare, data analytics, framework, Electronic Patient Record (EPR)

I. INTRODUCTION

Business intelligence (BI) can be defined as an umbrella term that encompasses tools, architectures, databases, data warehouses (DW), performance management, methodologies, and so forth, all of which are integrated into a unified software suite [1]. The implementation of BI in health care industry has enabled data to be delivered beyond administrative offices and directly to clinical staffs who can make the most use of it. The use of BI in healthcare enables decision making process to become more effective where users can access any type of information with a fast and consistent response time. Healthcare enterprises use BI to build management dashboards that help in managing business processes and monitoring financial and clinical Key Performance Indicators (KPIs) [2]. Healthcare data can be categorized into four types: clinical data, patient behavior, Pharmaceutical, and insurance data. The clinical data which represents information about the health status of patients and the health care they receive over varying periods of time. Patient behavior data, collected through monitors and wearable devices to provide an accurate and detailed view. Pharmaceutical data contains the medications, doses, impacts, and clinical trial reports [3]. Insurance data includes total premiums, employer and employee share of premiums, cost-sharing arrangements, number of enrollees, and total health care expenditures among others. Based on the previous massive amount of data that should be saved in the databases; there is big need to use data analytics to help the healthcare sector take near optimum decisions.

This paper is organized as follows. Section 2 describes the data analytics and its types. Section 3 shows the proposed BI framework for healthcare analytics and is components. Besides it describes the structure of the analytics tier as a core for the proposed architecture. Finally; conclusion and future work are presented.

II. ANALYTICS TYPES

Data analytics is defined as a process that involves the use of statistical techniques, information system software, and operations research methodologies to explore, visualize, discover and communicate patterns or trends in data [4]. Generally, analytics convert data into useful information for better decisions accuracy. There are four types of analytics which are; Descriptive, Diagnostic, Predictive, and Prescriptive as shown in figure 1.
The goal of descriptive analytics is to discover patterns in data. Descriptive and predictive analytics together are often called “Knowledge Discovery in Data” (KDD), but literally that name is a better fit for descriptive analytics [5]. Predictive analytics is defined as a set of BI technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events [6]. Prescriptive Analytics goes beyond predicting the future outcomes by also suggesting actions to benefit from the predictions and showing the decision maker the implications of each decision option [6]. Diagnostic analytics is used to determine “why it happened” [7]. Inside this tier there will be a tools box containing the methods and types of data analytics. The big four data analytics considered to be clustering, classification, prediction and association rules.

III. PROPOSED BI FRAMEWORK FOR HEALTHCARE ANALYTICS

Kaiser Permanente has an international reputation for using innovative practices and technology, including embracing the concept of big data, to improve the quality and efficiency of patient care. He has shown how the use of consent-based integrated health information across all medical providers can improve efficiency in patient care. A 2009 study found that immediate access to complete; integrated, current patient information (including inpatient, outpatient, imaging and pharmacy data) was associated with a 26% reduction in patient visits and an eightfold increase in the number of consultations that could be conducted via telephone [8]. Throughout this introduction one can say that introducing an integrated BI framework for healthcare analytics will lead to better decisions in healthcare organizations which is the main objective of this paper. As figure 2 depicts, the proposed architecture is generally composed of six tiers which are; data source tier, Extract Transform and Load (ETL) tier, data storage tier, analytics tier, optimization tier, and presentation tier. The proposed framework is based on the basic architecture of BI framework.

Figure 2: Proposed BI framework for healthcare analytics

A. Data source tier

Numerous provision domains oblige the utilization about organized information and also unstructured Furthermore semi-structured information on aggravate successful and auspicious choice [9]. These data sources contrast starting with different requisition domains due to those structures of the medicinal records. The following sections clarify the different medical data sources and structures.

1) Electronic Patient Record (EPR)

There is approximately 500 petabytes of healthcare data in existence today and that number is expected to skyrocket to more than 25,000 petabytes within the next seven years [10]. An Electronic Patient Record (EPR) is an evolving concept defined as a systematic collection of electronic health information about individual patients or populations [11]. The proposed framework keeps in consideration the different structure of the EPR from one application to the others regardless the structure of each file which is the main function of the ETL component.

2) Clinic data source

The reason of adding such component ‘clinical data’ is to study and review new drugs by regulatory agencies as a market research which depends on a trust of clinical trials presented for sufficient integrity to ensure confidence in results and concluded by pharma companies. The Clinical Data Management (CDM) may be used here in the proposed model as
an application in the presentation tier (tier 6). CDM is involved in all aspects of processing the clinical data, working with a range of computer applications, database systems to support collection, cleaning and management of subject or trial data. CDM is the collection, integration, and validation of clinical trial data [12]. The challenge here is that the Clinical notes contain rich and diverse source of information. Such these Challenges are handling clinical notes, Ungrammatical, short phrases, Abbreviations, Misspellings, Semi-structured information and copy-paste from other structure source. These challenges could be handled in during the ETL tier functionality.

3) Pharmacy data source
Pharmaceutical care does not exist in isolation from other health care services [13]. It must be provided in collaboration with patients, physicians, nurses and other health care providers. Pharmacists are responsible directly to patients for the cost, quality and results of pharmaceutical care. So adding the pharmacy data source in a BI framework as proposed is a key factor to enhance the analytics produced from such systems. Moreover help in providing accurate decisions and effective healthcare research facts. Also there is an important reason to add such component in the proposed framework is what Saira et al., reported in [14] that in most of the developing countries are not fully executing their potential role. They are still struggling for the recognition of their role that can help improve the health care system. Besides; most of the medications provided to patients not recorded via an integrated system to support healthcare sector to estimate and evaluate the optimal treatment methodology.

4) Hospital data source
The data currently collected from hospitals is sufficient neither for commissioners to make properly informed decisions nor for clinicians and patients to understand the quality of care provided. Adding hospital data as a data source in the proposed BI framework confirmed by what has been documented by the National Health Service (NHS) England [15] that proposed directing the Health and Social Care Information Centre (HSCIC) to collect a far more complete data set from hospitals, beginning in April 2014. To have an accurate decision you have to have complete data sets and highly trusted analytics tool and algorithm. Using a data set from one hospital or any other health institute will give more accurate results in analytics; this is because the normalized form of data. On the other hand collecting data from many hospitals or health institutes may cause inconsistency of data which makes it difficult to be used as a data set for having analytics and decisions. So the need to integrate all types of data regarding healthcare sector will lead to having accurate and trusted decisions.

5) Patient Relationship Management (PRM)
PRM like Customer Relationship Management (CRM) is a vital component to have a complete technological solution about an organization. Through automated systems based on the concepts of loyalty and relationship-based marketing, consumers have grown to rely on helpful follow-up communications based on their preferences [16]. Similar concepts can be used to establish PRM systems in healthcare organizations. With such a system in place, clinicians can establish ongoing relationships with their patients, enabling them to increase focus on prevention, and giving them a new tool to help improve patient well-being and satisfaction. Analytics allows companies to monitor performance while resolving patient issues no matter which channels are used to report them [17]. Moreover; Mobile solutions provide real-time access to vital information, which eliminates information silos, streamlines processes, and increases staff and patient satisfaction. So during the proposed BI framework the healthcare sector can develop and maintain positive and profitable relationships with referring physicians by measuring and monitoring outcomes.

6) Insurance company data source
Healthcare insurance data represent a rich source of information and has the potential to contribute significantly in guiding business decision making [3]. A closure look at the insurance value chain suggests that BI can play a crucial role in almost every aspect of the chain. It can help identify the right customers for target marketing and analyze the reasons for customer attrition. It can help the insurer better manage its agents and sales force and improve the effectiveness of actuarial and underwriting functions. BI forms the most critical component of claims management, helping in fraud detection and claims estimation. On the asset management side, it can lower the insurers risk through sophisticated risk models developed using data mining tools. And most importantly, BI tools can help insurers provide crucial information to corporate clients, which can go a long way in cementing the insurers’ relations with the clients [18].

7) Legacy systems data source
Legacy data comes from virtually everywhere within the information system and support legacy systems. The many sources of legacy data include databases, often relational but hierarchical, network, object, XML, and object/relational databases as well. Legacy data is another term used for disparate data [19]. These systems are usually large and companies have invested so much money in implementing legacy systems in the past that despite some potentially problematic identified by IT professionals, many still want to keep them for several reasons. Legacy data and the problem regarding data disparity they bring to a data warehouse can be solved by the process of ETL. This is a mechanism of converting disparate data not just from legacy systems but all other disparate data sources as well before they are loaded into the data warehouse. Legacy data source component helps the BI healthcare architecture to save legacy data which is the most important factor in business success factors.

8) Laboratory data source
Combining lab data gives the most comprehensive impact to the pathologist or other laboratory leader which provides substantial improvements in patient care [20]. The laboratory is at the forefront of change. It interacts with almost all healthcare departments. The power to access, combine, and analyze large data sets can improve the ability to anticipate and treat illnesses. This data can be used to identify waste in the healthcare system and to hit the target of the
Triple Aim goals: to improve the health of the population, improve the patient experience, and lower the cost of healthcare across the board. The challenges for labs in Healthcare Information Systems (HIS) are; many lab systems still use local dictionaries to encode labs, diverse numeric scales on different labs which often need to map to normal, low or high ranges in order to be useful for analytics, missing data, and the order of a lab test can be predictive. The most important challenge here is inconsistency between different data structures from different lab management systems. But the solution is ETL effectiveness.

9) Radiology data source

The need to add such component in the framework is to have a complete scenario about all aspects of healthcare sector. The radiology software and hardware is very complicated regarding technology wise. Digital Imaging and Communications in Medicine (DICOM) protocol is a standard for handling, storing, printing, and transmitting information in medical imaging. It includes a file format definition and a network communications protocol. The traditional, film-based radiology system presents serious limitations for patient care [21]. These include forcing clinicians to make decisions based on information that is often less than optimal and making transfers of films and prior studies to other facilities more complicated than they need to be. So the need to add radiology data source in the proposed healthcare BI framework is big challenge.

10) Primary care data source

Primary care defined as care provided by general practice teams including health visitors, district nurses and mental health nurses, and excludes care given by dentists, pharmacists, opticians, and midwives. Primary care data means; all data recorded in a healthcare portal or any other medical website, mobile application, or sensors.

11) Human Resource (HR) data source

Individuals delivering care or working in Health Human Resources (HHR) are one of the greatest assets of our health care system. Evidence-based planning and management of HHR is a vital part of making sure Canadians can access health services when and where they need them. The Canadian Institute for Health Information (CIHI) Health Workforce Database includes information about the 29 groups of health care professionals listed in Table 1—information such as practice setting, regulatory environment, supply, and demographic, education and employment characteristics. This information is used by health care planners, decision-makers, policy-makers and researchers to support resource and use planning.

<table>
<thead>
<tr>
<th>NO.</th>
<th>Title</th>
<th>NO.</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Audiologists</td>
<td>2</td>
<td>Chiropractors</td>
</tr>
<tr>
<td>3</td>
<td>Dental assistants</td>
<td>4</td>
<td>Dental hygienists</td>
</tr>
<tr>
<td>5</td>
<td>Dentists</td>
<td>6</td>
<td>Dietitians</td>
</tr>
<tr>
<td>7</td>
<td>Speech–language</td>
<td>8</td>
<td>Genetic counsellors</td>
</tr>
<tr>
<td>9</td>
<td>Environmental public health Professionals</td>
<td>10</td>
<td>Health information management professionals</td>
</tr>
<tr>
<td>11</td>
<td>Medical laboratory technologists</td>
<td>12</td>
<td>Occupational therapists</td>
</tr>
<tr>
<td>13</td>
<td>Medical physicists</td>
<td>14</td>
<td>Licensed practical nurses</td>
</tr>
<tr>
<td>15</td>
<td>Opticians</td>
<td>16</td>
<td>Optometrists</td>
</tr>
<tr>
<td>17</td>
<td>Paramedics</td>
<td>18</td>
<td>Pharmacists</td>
</tr>
<tr>
<td>19</td>
<td>Pharmacy technicians</td>
<td>20</td>
<td>Medical radiation technologists</td>
</tr>
<tr>
<td>21</td>
<td>Registered nurses/nurse practitioners</td>
<td>22</td>
<td>Physician assistants</td>
</tr>
<tr>
<td>23</td>
<td>Physiotherapists</td>
<td>24</td>
<td>Psychologists</td>
</tr>
<tr>
<td>25</td>
<td>Registered psychiatric</td>
<td>26</td>
<td>Physicians</td>
</tr>
<tr>
<td>27</td>
<td>Social workers</td>
<td>28</td>
<td>Respiratory</td>
</tr>
<tr>
<td>29</td>
<td>Midwives</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Human resources in health sector reform also seek to improve the quality of services and patients' satisfaction. Health care quality is generally defined in two ways: technical quality and sociocultural quality. Technical quality refers to the impact that the health services available can have on the health conditions of a population. Sociocultural quality measures the degree of acceptability of services and the ability to satisfy patients' expectations [22].
12) Homecare data source

HomeCare Billing Solution, Carecenta, The Rosemark System, AdaCare [23], and much more software ranked and heavily used for homecare. Home care organizations include home health care agencies, home care aide organization, and hospices. Some of these organizations are Medicare certified, which allows providers to bill Medicare for reimbursement. Agencies that are not Medicare certified cannot be reimbursed through Medicare. So losing such data from such software and even healthcare portals and/or mobile applications may affect the results of healthcare analytics and by the way affect the decision and medications regarding large amount of similar cases.

B. Extract Transform and Load (ETL)

ETL is an essential and core modules of any BI solution. It is the process of getting data out from one source and loads it to data warehouse [24]. Data is extracted from the heterogeneous data source like figure 2 shows; healthcare data sources are many and highly complicated in structure. Extraction form different database platforms means that data need to be integrated into one and then resulted data should be captured. Data need to be integrated into one and then resulted data should be captured. The transformation process as shown in figure 3 is performed in the staging area it is main important step of the ETL where values are added by ETL Process. Data quality check is done for transformation process like, correct, unambiguous, consistent, complex. Cleaning data required the step which invoiced first anomaly detection.

![Figure 3: Data Transformation Process [25]](image)

ETL is the process used to extract data from the source system such as the Magnetic resonance imaging (MRI), Pharmacy, Lab, or Imaging, for loading into the data warehouse. Often, when this data is extracted, it needs to be converted, re-mapped, or transformed in some way. ETL applications are designed to do this efficiently. For example, an HL7 message might be received containing patient demographic information. The health system might prefer all states to be stored using the 2-letter code. The ETL would receive the HL7 message, transform Georgia to GA, and then load this revised data into the data warehouse. The ETL tools are the most important of all the capabilities because the healthcare data and their structures are extremely complicated. There are several standards and coding systems for each sub-domain: DICOM for radiology; HL7 for general health information exchange; SNOMED and ICD for coding diagnoses and procedures; LOINC for observations; RxNorm for pharmaceuticals. There are also areas for which there really are no standards whatsoever, such as: patient satisfaction, measuring accessibility to healthcare services, measuring the effects of improved outcomes, etc.

C. Data Storage Tier

Because of the various care practices, data types and definitions, and the perceived incompleteness of clinical information systems, the development of a clinical data warehouse is a challenge [25]. The storage tier is composed of two components; firstly; DW as a storage repository. The characteristics of any DW are: Subject-oriented, Integrated, Time-variant, and Non-volatile. Subject oriented means it concentrates on enterprise-specific concepts, such as customers, products, sales, and orders. An Integrated data means that the DW merges heterogeneous sources into one common schema. Time-variant means that data stored in the DW has time dimension to keep track of the changes or trends on the data. Non-volatile means that data stored in data warehouse are read-only and the users are not allowed modify or delete the stored data. Secondly; Data Mart which is a subset or an aggregation of the data stored to a primary data warehouse. It includes a set of information pieces relevant to a specific business area, corporate department, or category of users [26]. Here in the proposed framework; there are three types of data mart. Operational data mart; is a subset of the DW that contains data about the daily routines inside the healthcare organizations and it used mainly for analytical purposes. Medical claims data mart; it contains Claims data which consists of the billing codes that physicians, pharmacies, hospitals, and other health care providers submit to payers. Financial data mart; which contains data about expenses, medications cost, insurance companies and any other type payments.

D. Analytics Tier

The main goal of this tier is to provide some types of analytics regardless of the data types under investigation. The big four data analytics considered to be clustering, classification, prediction and association rules. As depicted in figure 1, analytics may be throughout ad hoc techniques or queries besides; Data Mining (DM) techniques (clustering,
classification…) could be implemented through algorithms or tools. The output of this tier should be optimized and enhanced to reach the optimum solution and results this leads to the next tier (optimization) of the proposed model.

The Analytics tier is considered to be a set of steps containing a toolbox for all types of analytics. As figure 4 shows; it starts from obtaining the data from any type of data sources and ending with the optimal solution agreed by the users (doctors, physicians, consultants …). There is a fact that can be extracted from figure 4 which is the existence of the optimization tier which is used in enhancing the results obtained from the analytics tier based on special metrics observed by the user.

The structure of this tier contains many algorithms and techniques regardless the type of analytics being used. As depicted in figure 2; query and reporting tools are very useful. There are many different types of reports including standard reports, ad hoc reports, budgeting and planning reports, and metadata reports. Both internal and external users can manage reports and other information easier and faster through BI portals. BI portal is a popular end user tool to deliver information. The analytics tier contains also OLAP functions which vary according to the stockholder’s requests. Roll-up or drill-up is a function which increases the level of aggregation by moving up to a higher level or reducing one or more dimensions from a given data cube. Drill-down is the opposite of roll-up. It decreases the level of aggregation. Slice and dice operation can be performed by selecting a specific value on a single dimension and performs a projection on a data cube by selecting a range of values on two or more dimensions.

**E. Optimization Tier**

Generally, Optimization process is any kind of process that systematically comes up with solutions that are better than the solution used before. Process optimization is the discipline of adjusting a process so as to optimize some specified set of parameters without violating some constraint. The main goal of this tier is to have the ability to have more modified results from the analytics tier. The analytics tier structure shown in figure 4 contains a block named optimization which contains a lot of techniques varies from mathematical programming, gradient method, and stochastic to distributed.

**F. Presentation Tier**

This tier consists of tools that display information in different formats to different users. These tools can be classified into web browser, web applications, mobile applications, dashboards, and alerts software. Data visualization tools such as dashboard and scorecards can be provided to managers and executives who need an overall view of their business performance. Dashboard is a useful tool that allows users to visualize data using charts, colored metrics or tables. Users can also view more detailed information about KPIs across their organizations. By doing so, managers can closely and more effectively monitor their business performance and progress toward defined goals. Web browsers are software that facilitates the way of surfing the internet. They can also be used to access information provided by web servers in private networks or files in file systems related to the Healthcare organization using the proposed BI architecture. Matt Gorbsky [27] reported that mobile applications are exploding. He has compiled a list of the best representative apps for health.
In developing countries: the current scenario in

Christopher, M. Starkey. 2014. Business Analytics Principles, for BI. But they actually concentrated on the

GraPhys: The huge amount of internet users especially mobile

Many trials have been achieved to reach an integrated framework for BI. But they actually concentrated on the

REFERENCES


SAP. 2014. Patient Relationship Management. Software solution


© 2017, IJARCSSE All Rights Reserved
ABOUT AUTHORS


Assoc. Prof. Dr. Sherif Kholeif, PhD. He was born in Cairo, Egypt. Head of the Department of Information Systems, faculty of computers and information, Helwan University, Cairo Egypt. Master in computer science and information systems, faculty of computers and information systems, information systems department Helwan University, Cairo, Egypt. PhD Degree in Information System and Computer Science, faculty of computers and information systems, information systems department Helwan University, Cairo, Egypt.

Fifi M. Hassan. She was born in Cairo, Egypt. BSc of computers and Information Systems, Helwan University, Egypt. 2000. Interested in business intelligence and big data analytics. She worked as a web developer in German chamber in Egypt. Besides, she worked as a web developer in Coldwell Banker. She is preparing her master degree named ”An Integrated Business Intelligence Framework for Enhancing Healthcare Analytics” in Helwan University, Cairo, Egypt.