Development of National Health Data Warehouse for Data Mining

Shahidul Islam Khan, Abu Sayed Md. Latiful Hoque

Dept. of Computer Science and Engineering (CSE),
Bangladesh University of Engineering & Technology (BUET), Dhaka, Bangladesh
nayeemkh@gmail.com , asmlatifulhoque@cse.buet.ac.bd

Health informatics is currently one of the top focuses of computer science researchers. Availability of timely and accurate data is essential for medical decision making. Health care organizations face a common problem with the large amount of data they have in numerous systems. Researchers, health care providers and patients will not be able to utilize the knowledge stored in different repositories unless amalgamate the information from disparate sources is done. This problem can be solved by Data warehousing. Data warehousing techniques share a common set of tasks, include requirements analysis, data design, architectural design, implementation and deployment. Developing health data warehouse is complex and time consuming but is also essential to deliver quality health services. This paper depicts prospects and complexities of health data warehousing and mining and illustrate a data-warehousing model suitable for integrating data from different health care sources to discover effective knowledge.

Keywords: *Data Mining, Data Warehouse, Health Informatics, Clinical Database, Data Preprocessing*

1 Introduction

Health informatics or healthcare informatics is an intersection of computer science and health care services. It deals with resources and methods needed to optimize the acquisition, storage, retrieval and use of information in medical research and applied to the areas of health care management, diagnosis, clinical care, pharmacy, nursing and public health [1, 2]. Knowledge discovery from data (KDD) is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. Data mining, a major part in KDD, consists of applying data analysis and learning algorithms to produce potential interesting patterns over the data [3, 4, 5].

Health data refers to any information that is contained in a patient's medical record. This information may be acquired from notes derived from a hospital admission or a doctor's visit or diagnostic report. This data comes in various forms such as text or numbers (patient identification, demographics, history, laboratory data, etc), analog or digital signals (ECG, EEG, EMG, ENG etc), images (histological, radiological, ultrasound, etc), and videos. Further complicating the storage of this data is the fact that because patient identification information cannot be publicly used. Such identifiers must be removed from other clinical parameters. Difficulty in storing this type of data is that each disease and species can only be effectively described using greatly different vocabularies and data elements. [2, 6, 7, 8].

One of the major Information Technology challenge in medical practice is how to integrate several disparate, isolated information repositories into a single logical repository to create consistent information for all users. A massive amount of health records, related documents and medical images created by clinical diagnostic equipment are generated daily. These valuable data are stored in various medical information systems such as HIS (Hospital Information System), PACS (Picture

Archiving and Communications System), RIS (Radiology Information System) in various hospitals, departments and diagnostic laboratories. Data required to make informed medical decisions are trapped within fragmented, disparate, and heterogeneous clinical and administrative systems that are not properly integrated. As a result health care suffer because medical practitioners and health care providers are unable to access this information to perform activities such as diagnostics, and treatment optimization to improve patient care [1, 6, 7].

Successful healthcare data management is an important factor in developing support systems for the clinical decision-making process. Traditional operational database system does not satisfy the requirements for critical data analysis tasks of the clinical decision-making users. It contains detailed data but do not include important historical data, and since it is highly normalized, it performs poorly for complex queries that need to join many relational tables or to aggregate large volumes of data in order to generate various clinical reports. A health data warehouse is a data store that is different from the hospital's operational databases. It can be used for the analysis of consolidated historical data [7, 8].

According to Inmon [9] A data warehouse (DW) is a subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management's decisions.

Subject-oriented: as the warehouse is organized around the major subjects of the enterprise (such as customers, products, and sales).

Integrated: as DW is constructed by integrating multiple heterogeneous sources usually, such as relational databases, flat files etc.

Time-variant: as data in the warehouse is only accurate and valid at some point in time or over some time interval.

Non-volatile: as the data is not updated in real time but is refreshed from on a regular basis from different data sources.

The advantages and disadvantages of DW are given below [9, 10, 11]:

Advantages of DW:

1. Standardize data across the organization
2. Improve turnaround time for analysis and reporting
3. Easy Sharing of data
4. Remove informational processing load from operational database
5. Enhance Data Quality and Consistency
6. Provide historical intelligence and reduce cost to access historical data
7. Integrate data from multiple sources into a single repository
8. Improve data quality by providing fixing noisy data
9. Restructure the data so that it delivers excellent query performance
10. Make decision–support queries easier to write.

Disadvantage of DW:

1. Long initial development time and associated high cost
2. Data owners lose control over data, raising ownership and privacy issues

Implementing a Health DW is a complex task containing two major phases. Firstly, in the configuration phase, a conceptual view of the warehouse is specified according to user requirements (DW design). Secondly, the related data sources and the Extraction-Transform-Load (ETL) process (data acquisition) are determined. After the initial load during the operation phase, warehouse data must be regularly refreshed that is, modifications of operational data since the last DW refreshment must be propagated into the warehouse such that data stored in the data warehouse reflect the state of the underlying operational systems [5, 8, 12].

The main aim of this research is to identify the obstacles for healthcare data integration and to propose a data-warehousing model suitable for integrating fragmented data in respect to Bangladesh as well as anywhere else. The result will contribute to the advancement of knowledge in the field of medical informatics. In this paper "Health", "Clinical" "Pathological" and "Medical" these terms are used for similar meaning. The rest of this paper is organized as follows. In Section 2 we have presented selected literature reviews on DW, Health DW and KDD techniques. Section 3 describes briefly some design issues of National Health DW. In Section 4 we have shown the calculation of approximate size of our DW. Some preprocessing techniques that we have used are illustrated in Section 5. Section 6 gives readers ideas about how our DW will be used for knowledge discovery and mining. Finally Section 7 concludes the paper.

2. Literature Review

DW unifies the data scattered throughout an organization into a single centralized data structure. It is a repository of integrated information available for querying and analysis. DW may be considered a *proactive* approach to information integration, as compared to the more traditional *query driven* approaches where processing and integration starts when a query arrives [5, 6]. A health data warehouse is a repository where healthcare providers can gain access to medical data gathered in the patient care process. Extracting medical domain information to a data warehouse can facilitate efficient storage, enhances timely analysis and increases the quality of real time decision making processes. Today's healthcare organizations require not only the quality and effectiveness of their treatment, but also reduction of waste and unnecessary costs. In order to construct an operational and effective DW it is essential to combine process work, domain expertise and high quality database design [7, 8].

Electronic Health Record (EHR) describes the diseases and treatments of patients, are normally stored in hospitals or clinics, where they are created. Patients may be treated in different hospitals, clinics and, therefore, there is a need for integrating health records from different hospitals to enable any hospital to obtain a total overview of a patient's health history. Different heterogeneity problems have to be solved in order to integrate EHR systems from different hospitals and health service providers in a consistent way. The first problem is that different hospitals normally do not use a same DBMS and therefore, the traditional ACID properties of databases are missing across the different hospital locations. This may cause performance, autonomy, and consistency problems. Another heterogeneity problem is that there are several incompatible standards for EHR entries [12].

The trend of adopting data warehouses for health systems in presented in [13], where the design experience in the University of Virginia Health System is reported. Here the data warehouse is used to provide clinicians and researchers with direct, rapid access to desired patients' data. In addition they use DW also for educational and research aims, as it serves to face informatics issues−such as data capture−and to perform exploratory analyses of healthcare problems.

Medical domain has certain unique data requirements such as high volumes of unstructured data and data confidentiality. There are huge constraints and issues that limit the way the data mining is performed for medical datasets. Some of these issues are the way the data is collected; accuracy of the data, ethical, privacy and social issues that comes with patient's records [2]. Research is also done to find out impact of missing values and explore the impact of noise and how this can influence the output. Zhu et al. classified noises into class noise and attributes noise. Attribute noise include incorrect attribute values, missing or don't know attribute values and incomplete attributes or don't care values [14].

Several researches have focused on the techniques that have built in mechanism to handle noise and missing values and which are more appropriate to use for medical applications. Few techniques that have been applied and are more suited to medical data sets are studied in [15, 16]. For example decision tree, logic programs, K-nearest neighbour, and Bayesian classifiers. Lee et al recommended that Bayesian networks and decision trees are the primary techniques applied in medical information systems [17]. Obenshain claimed that that neural networks performed better then logistic regression, but the decision tree did better in identify active compounds most likely to have biological activity [18]. Wang and Wang discussed that most process models do not focus in gaining new knowledge. Medical data mining applications should follow a five stage data mining development cycle: planning tasks, developing data mining hypotheses, preparing data, selecting data mining tools, and evaluating data mining results [19].

Handling Missing Data in Pathology Databases using Multiple Imputation technique is discussed in [20]. Optimizing public health data collection for KDD using feature selection is studied in [21]. Cubillas et. al. proposed a model for improvement in appointment scheduling in health care centers [22]. Hoque et. al. discussed present structure of pathological data, requirements to formulate efficient models and the necessity to reform the present structure for predicative data mining in [23]. Kumari and Singh used Neural Network for the diagnosis of diabetes [24]. Yilmaz et. al. proposed a modified K-means Algorithm based data preparation method for diagnosis of heart and diabetes diseases [25]. Herland et. al. present recent research using Big Data tools and approaches for the analysis of Health Informatics [26].

3. Design Issues of National Health DW

The architecture of national health DW model is illustrated in Fig. 1. Health data from different govt. and private sources such as hospitals, clinics, diagnostic centers, research centers will be collected. Using ETL process data will be integrated into a temporary data repository [27].

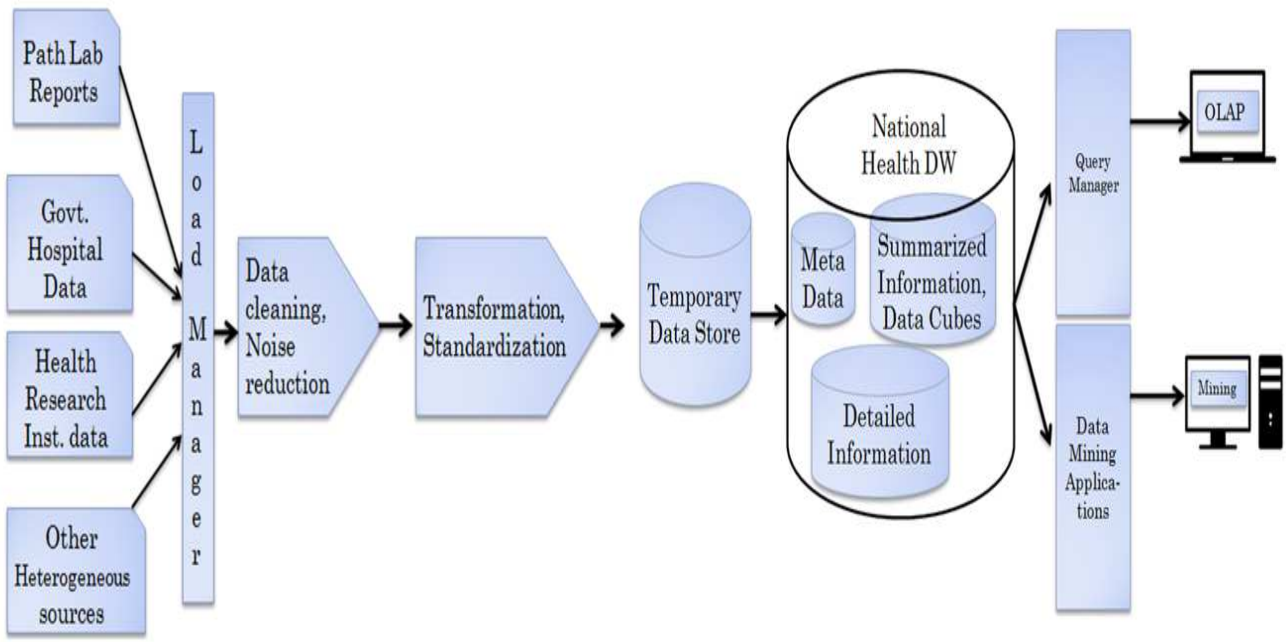

Fig. 1 Brief Architecture of Health Data Warehouse

Cleaning, noise reduction, normalization techniques will be applied next. After that data will be loaded into DW. Online Analytical Processing (OLAP) queries and mining operations can be easily performed over the pathological DW.

4D Health data cube used for national health DW development is shown in Fig. 2. Here 0-D apex cube will provide highest level of summarization of national health data. Partial materialization is used rather than full materialization of cuboids to reduce huge space requirements [9, 10].

Logical design of DW involves the definition of structures that enable an efficient access to information. There are many logical models like Flat schema, Star schema, Star Cluster schema, Snowflake schema, Fact Constellation schema etc. Among them, star schema, snowflake schema and fact constellation schema are mostly used commercially. Efficiency is the most important factor in DW modeling because many queries access large amounts of data that may involve multiple join operations. Most suitable Logical Data Warehousing Model is the Star Schema [9, 12, 13]. We have used Star Schema in our design, illustrated in Fig. 3.

Using the building blocks of the fact table and the various dimension tables, one has thousands of ways to aggregate the data. For clinical analysis purposes, frequently needed aggregated datasets should be created in advance for the users. Having data readily and easily available is a major tenet of data warehousing. For our DW, some aggregated datasets could be:

- Patient count by Diagnosis, Gender, Age, Date
- Count of Procedures by Provider and Date
- Billing and discount information.
- Count of retesting

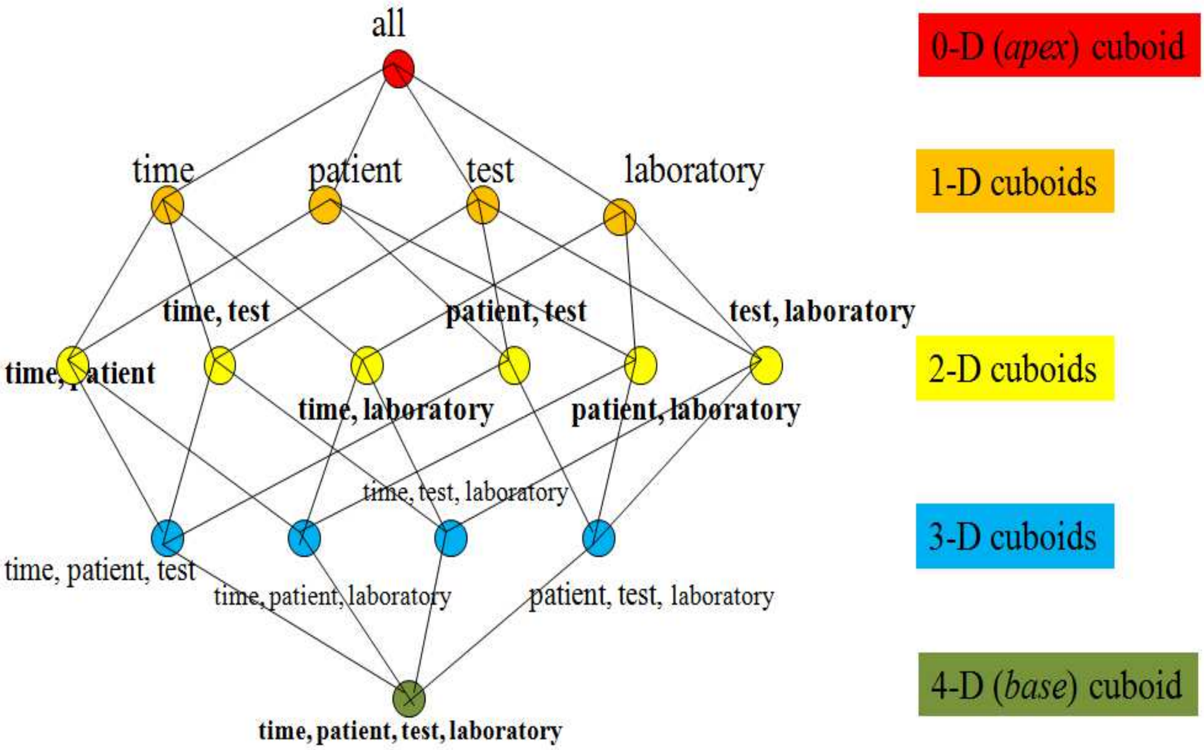

Fig. 2 Health Data Cube

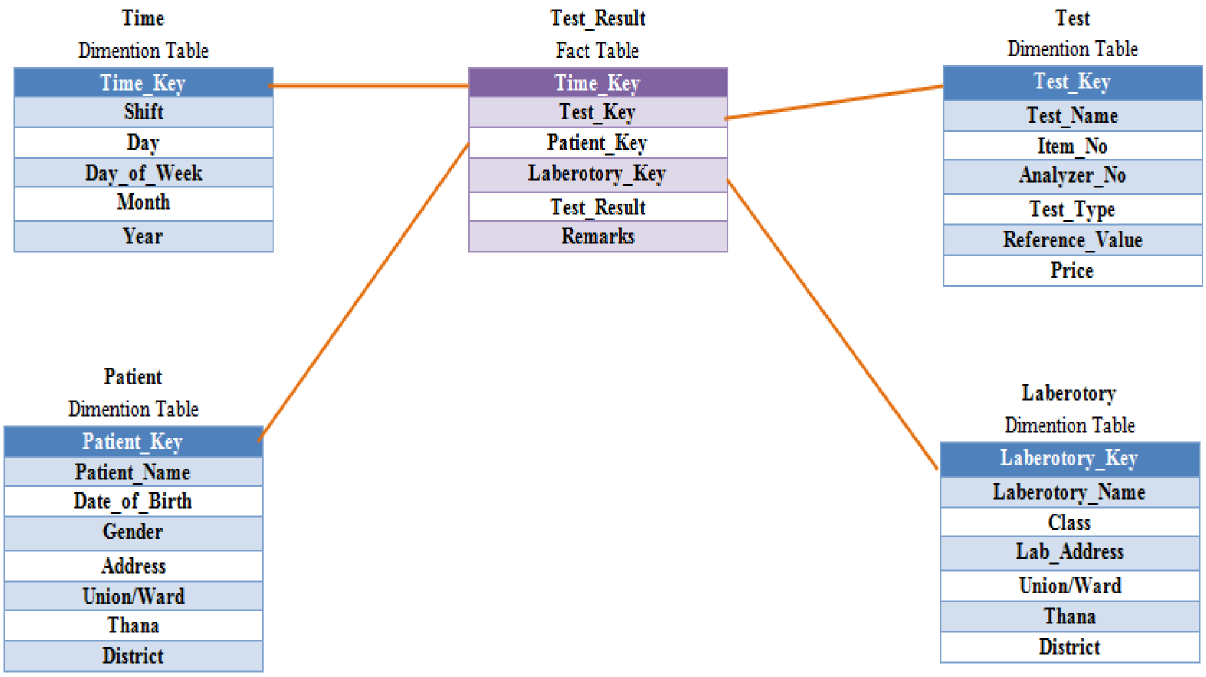

Fig. 3 Fact Table and Dimension Tables of National Health DW

4. DW Size Analysis

Let, test for a single patient = t_p ; total patients in class i = p_i

So total test for p_i patients = $\sum_{P=1}^{P=P_i} t_p$

total reports in class i = r_i ;

number of test in r_i = $\sum_{j=1}^{j=r_i} t_{rj}$

where t_{rj} is number of test in report j

Total number of tests T is given by

n=number of classes in laboratories

l_i = number of laboratories in class i

r_i =number of test reports in class i

t = number of test per report

We have derived the following equation to count the total number of test records (tuples in fact table) are generated by the different healthcare organizations such as hospitals and diagnostic centers.

$$T = \sum_{i=1}^{i=n} l_i \times r_i \times \left(\sum_{j=1}^{j=r_i} t_{rj} / r_i \right)$$

According to Directorate General of Health Services (DGHS) under the Ministry of Health and Family Welfare (MoHFW): Total number of government hospitals under DGHS is 592 and Government hospitals of secondary and tertiary levels under DGHS is 125[28], [29]. According to the Bangladesh Private Clinic and Diagnostic Owners Association (BPCDOA), 8,000 diagnostic centers of the country have DGHS approvals till December 2014 [30], [31]. So minimum number of places where pathological tests are performed is 8717. If for simplicity of calculation we consider average 500 patients reports are produced every day, each report consists 15 test attributes then putting the values in above equation we get:

T= 65377500 records(tuples /test attributes) per day.

So more than 65 million records will be added in the fact tables of National Health DW. Considering 1 record takes 0.2KB memory space than the DW will consume 12.50 GB memory/day. It is certainly falls under big data category and Bangladesh Government should go for Cloud storage and services for this [26, 32]. For fragmentation of big database there are several techniques such as CRUD matrix based fragmentation proposed by the authors of this paper [33, 34].

5. Data Preprocessing

Data preprocessing is one of the major task for developing a DW from heterogeneous sources. It includes data cleaning, missing values imputation, normalization, transformation etc. As for National Health DW, data are coming from different public and private hospitals, diagnostic centers and other sources, different preprocessing steps has been performed on data. Followings are some data preprocessing that we have performed. Table 1 and 2 present few of PCV Hct Red Cell Indices test data. Here reference values for female are 36-46 and for male are 40-50. The full dataset for this particular test contains 13296 records where the minimum and maximum results are 0.1 and 64 respectively. The results and the reference values are normalized using the Min-Max and Z-Score normalization techniques. Missing data are replaced by *class mean* method. Table 3 shows partial metadata for the same test dataset.

Table 1 Attribute subset selection and normalization of numeric data

TESTRESULT_ID	RESULT	Z Score normalized result	Min Max Normalized result
114080000000002	39.9	0.58	0.6228
114080000000204	39	0.41	0.6088
114080000000283	0.1	-6.91	0
114080000000609	0.2	-6.89	0.0016
114080000000755	0.1	-6.91	0
114080000000834	28.3	-1.6	0.4413
114080000000913	43	1.16	0.6714
114080000001138	29.7	-1.34	0.4632
114080000001279	37.8	0.18	0.59
114080000001436	37	0.03	0.5775
114080000001650	39	0.41	0.6088
114080000002071	39	0.41	0.6088
114080000002248	35	-0.34	0.5462
114080000003618	42	0.97	0.6557
114080000003766	41	0.79	0.6401
114080000003900	46	1.73	0.7183

Table 2 Reference values and their normalization

Reference values	Min-Max Normalization	Z Score Normalization
Female Lower: 36	0.561815336	-0.154685736
Female Upper:46	0.718309859	1.727135868
Male Lower: 40	0.624413146	0.598042906
Male Upper:50	0.780907668	2.479864509

Table 3 Metadata for the test result

Type	Value
Average	36.812
Maximum	64

Minimum	0.1
Standard Dev.	5.3070

Table 4 presents normalization technique of nominal data where metadata of Table 5 are used to replace result data for Urine colour diagnosis.

Table 4 Preprocessing of nominal data

TESTRESULT_ID	Result Type_ID	Min-Max Normalization
114080000002446	0	0.0000
114080000013324	1	0.1111
114080000098717	4	0.4444
114080000487792	6	0.6667
114080000743386	2	0.2222
114090000792554	3	0.3333
114090000822074	3	0.3333
114090000902763	7	0.7778
314080000143669	8	0.8889
314080000652184	5	0.5556
414080000046596	9	1.0000

Table 5 Metadata for Type_ID generation from Urine colour

Colour	Sample_Count	Type_ID
Straw	9439	0
Yellow	58	1
Reddish	32	2
L. Yellow	29	3
Milkly white	2	4
D. Yellow	2	5
Yellowish white	1	6
Reddish black	1	7
L.Reddish	1	8
Hazy	1	9

6 Data Mining from National Health DW

National Health DW can be used in many ways to improve national health standard, to provide better and prompt services to the patients and to facilitate health related research among the doctors, clinical researchers etc. In this section we are describing the use of National Health DW with two examples.

Example 1: National Reference level threshold finding

Table 6 WHO's Hemoglobin thresholds to define anemia

Age/Gender group	Hb threshold(g/dl)
Children(0.5-5yrs)	11.0
Children(5-12yrs)	11.5
Teens(12-15yrs)	12.0
Women, non-pregnant(>15yrs)	12.0
Women, pregnant	11.0
Men(>15yrs)	13.0

In Bangladesh, rule of thumbs is for *Woman > 15 years, non pregnant,* Hb> 11 Good; Hb >=10.5 ok and if Hb < 10 (g/dl), medication for the patient is needed. This is slightly different from WHO threshold [35]. Using National Health DW National Reference level for different clinical values can be found by data mining.

Example 2: Fraud Testing Awareness
If For a Costly Test T1:
Age(X,<30) ^ (Gender='M') => Negative (X, T1)
[Support =85%, Confidence=98%]
From confidence value it can be clearly identified that this test T1 has almost no impact of disease diagnosis. National awareness can be developed not to perform the test at initial level for Young Males.
In this way many other interesting patterns can be derived from National Health DW by using various data mining algorithms like association or clustering.

7. Conclusions

This paper presented the developmental stages of National Health DW platform for the management, processing and analysis of large-scale Health data modeled for e-health system. In this paper, widely accepted conceptual and logical design approaches in DW design are discussed. Considering the quality factors and the information requirements, star schema was chosen as the most suitable logical model for the purpose. Establishing a data warehouse gathering huge data from existing Health databases should give easier and better access to interesting data for researchers, health service providers and govt. authorities. In order to get maximum benefit from the model presented in this research, the conditions mentioned below should be satisfied.

1. There should be a document which clearly defines the structure of the data tables currently used by the concerned Pathological centers.
2. Stakeholders should clearly know what the data retrieval operations are going to be executed using the data warehouse.
3. There should be strong cooperative mind among different health service providers to help the governmental bodies for successful implementation.

Acknowledgment
This research is funded by ICT Division, Ministry of Posts, Telecommunications and Information Technology, Government of the People's Republic of Bangladesh.
Authors are grateful to Mr. Mamunur Rashid and Muhammed Dastagir Husain of Dept. of CSE, BUET for their valuable discussions.

References

[1] Roddick JF, Fule P, Graco WJ (2003) Exploratory medical knowledge discovery: experiences and issues. SIGKDD Explor. Newsletter, 5(1): 94-99
[2] Cios K (2002) Uniqueness of medical data mining. Artificial intelligence in medicine. 26:1-24
[3] Fayyad UM, Shapiro GP, Smyth P (1996) From Data Mining to Knowledge Discovery: An Overview. Advances in Knowledge Discovery and Data Mining 1–36
[4] Khosla R, Dillon T (1997) Knowledge Discovery, Data Mining and Hybrid Systems. Engineering Intelligent Hybrid Multi-Agent Systems, Kluwer Academic Publishers 143–177
[5] Inmon WH (1992) EIS and the data warehouse: a simple approach to building an effective foundation for EIS. Database Programming and Design, 5(11): 70-73
[6] Stolba N, Banek M, and Tjoa AM (2006) The Security Issue of Federated Data Warehouses in the Area of Evidence-Based Medicine. First International Conference on Availability, Reliability and Security (ARES'06, IEEE)

[7] Sahama TR, Croll PR (2007) A Data Warehouse Architecture for Clinical Data Warehousing. Australasian Workshop on Health Knowledge Management and Discovery (HKMD 2007)

[8] Lyman JA, Scully K, Harrison JH (2008) The development of health care data warehouses to support data mining. Clin Lab Med. 28(1):55-71

[9] Inmon, W (2005): Building the Data Warehouse, 4th edition, Wiley-New York.

[10] Jiawei H, Micheline K, Jian P (2012) Data Mining Concepts and Techniques 3rd Edition, Elsevier

[11] Kimball R, Ross M (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling 3rd Edition, Wiley

[12] Nugawela S (2013) Data Warehousing Model For Integrating Fragmented Electronic Health Records From Disparate And Heterogeneous Clinical Data Stores, M.Sc. Thesis, Queensland University of Technology

[13] Mullins M, Siadaty MS, Lyman J et al (2006) Data mining and clinical data repositories: Insights from a 667,000 patient data set. Comput. Biol. Med. 36: 1351–1377

[14] Zhu X, Khoshgoftaar T, Davidson I, Zhang S (2007) Special issue on mining low-quality data, Knowledge and Information Systems, 11:131-136

[15] Brown ML, Kros JF (2003) Data mining and the impact of missing data. Industrial Management & Data Systems 103: 611-621

[16] Lavrač N (1999) Selected techniques for data mining in medicine. Artificial intelligence in medicine 16(1): 3-23

[17] Lee IN, Liao SC, Embrechts M (2000) Data mining techniques applied to medical information. Medical Informatics & the Internet in Medicine 25(2): 81-102

[18] Obenshain MK, Application of Data Mining Techniques to Healthcare Data, Infection Control and Hospital Epidemiology, vol.25, no 8, pp. 690-695, 2004

[19] Wang, H, Wang S (2008) Medical knowledge acquisition through data mining,. IEEE International Symposium ITME.

[20] S. FU (2011) Missing Data in Pathology Databases. MSc Thesis, Australian National University.

[21] Partington SN, Papakroni V, Menzies T (2014) Optimizing data collection for public health decisions: a data mining approach. BMC Public Health 14: 593-598

[22] Cubillas JJ, Ramos MI, Feito FR, Ureña T (2014) An improvement in the appointment scheduling in primary health care centers using data mining. J. Med. Syst., Springer 38: 89

[23] Hoque ASML, Galib S, Tasnim M (2013) Mining Pathological Data to Support Medical Diagnostics. Workshop on Advances on Data Management: Applications and Algorithms, Department of Computer Science and Engineering, BUET, Dhaka, 71-74

[24] Kumari S, Singh A (2013) A data mining approach for the diagnosis of diabetes mellitus. IEEE 7th International Conference on Intelligent Systems and Control

[25] Yilmaz N, Inan O, Uzer MS (2014) A New Data Preparation Method Based on Clustering Algorithms for Diagnosis Systems of Heart and Diabetes Diseases," J. Med. Syst., Springer, 38

[26] Herland M, Khoshgoftaar TM, Wald R (2014) A review of data mining using big data in health informatics. J. Big Data, Springer 1: 2

[27] Khan SI, Hoque ASML (2015) Towards Development of Health Data Warehouse: Bangladesh Perspective. Accepted in 2nd International Conference on Electrical Engineering

and Information & Communication Technology (iCEEiCT 2015).
[28] HEALTH BULLETIN 2014, 2nd Edition, DGHS, Ministry of Health and Family Welfare, Government of the People's Republic of Bangladesh
[29] http://www.dghs.gov.bd/index.php/en/health-program-progress/hpnsdp-2011-16/84-english-root/ehealth-eservice/497-hpnsdp-2011-16-brief. Accessed 20 Feb 2015
[30] http://www.bpcdoa.com/clinics_and_diagnostics.html. Accessed 22 Feb 2015
[31] http://www.thefinancialexpress-bd.com/2014/12/15/71077/print Accessed 22 Feb 2015
[32] Liang Z, Sherif S, Anna L, Athman B (2014) Cloud Data Management-Springer Switzerland
[33] Khan SI, Hoque ASML (2010) A New Technique for Database Fragmentation in Distributed Systems, International Journal of Computer Applications 5 (9), 20-24.
[34] Khan SI, Hoque ASML (2012) Scalability and Performance Analysis of CRUD Matrix Based Fragmentation Technique for Distributed Database, in Proceedings of 15th International Conference on Computer and Information Technology (ICCIT), 562-567.
[35] World Health Organization (2008). Worldwide prevalence of anaemia 1993–2005. Geneva: World Health Organization. ISBN 978-92-4-159665-7

Shahidul Islam Khan obtained his B.Sc. and M.Sc. Engineering Degree in Computer Science and Engineering (CSE) from Ahsanullah University of Science & Technology (AUST) and Bangladesh University of Engineering & Technology (BUET), Dhaka, Bangladesh in 2003 and 2011 respectively. He is now a Doctoral Student in the Department of CSE, BUET, which is the highest ranked technical university of Bangladesh. His current field of research is data mining and health informatics. He has more than 10 published papers in international conferences and journal. He is also an Assistant Professor (Currently in study leave) in the Dept. of CSE, International Islamic University Chittagong (IIUC), Bangladesh.

Abu Sayed Md. Latiful Hoque graduated from the Dept. of Electrical and Electronic Engineering (EEE), Bangladesh University of Engineering & Technology (BUET), Dhaka, Bangladesh in 1986. He obtained Post Graduate Diploma in 1992 from Asian Institute of Technology (AIT), Thailand and Ph.D. in CS from University of Strathclyde, Glasgow, UK in 2003. He is a professor of the Dept. of CSE, BUET and a prominent international researcher in the field of Database, Data Mining, E-Learning. He has near about 50 published papers in reputed international journals and conferences. He is also author of a book on Database Systems which is taught in Universities.