

HEALTH CARE DATA WAREHOUSE SYSTEM ARCHITECTURE FOR INFLUENZA (FLU) DISEASES

Rajib Dutta

Department of Computer Science & Engineering, Global Institute of
Management & Technology, Krishnanagar, Nadia, West Bengal, India.
rajibdutta2007@gmail.com

ABSTRACT

Data Warehouse is the most reliable technology used by the company for planning, forecasting and management. Critical business management data was contained in several unrelated and disconnected databases, both internally managed and from external sources. Client was unable to view the data from an integrated viewpoint. The data warehousing is one of the best technique to integrate data. This paper presents the Influenza (Flu) diseases specific data warehouse architecture for health care. This could be used by the database administrator or executive manager, doctors, nurses, other staff members of the health care. Health care data warehouse is mostly important to integrate different data format from different data source. All information about patient including their medical test reports are store in the database, the executive manager needs to access those data and make a report. By seeing the report, the doctor takes action.

KEYWORDS

Influenza or Flu, Health care data warehouse, Extract Transform and Load (ETL), and Decision Support System (DSS).

1. INTRODUCTION

There are relatively few institutions that have developed clinical data warehouses, containing patient data from the point of care. 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 [1]. The world largest, fast growing and most information's are available in the health care industry. In health care industry data may be recorded as doctor's name, patient's name, patient's record, individual patient pathology report, physician order entry, doctor's decision support system, medicine. Most of the health care centers are still stand along, they are not communicating with other health care center, and they don't share their documents with others. But now in my design health care data warehouse, doctors can also shared patient record to others, they can take decision from others. Most of the health care center design their patient record in individual group but in my design group are created based on disease i.e. same disease parent are in the same group.

In this paper based on the influenza (flu) disease, the cost of treatment, treatment using drug and vaccine, risk factor, death rate.

Influenza, commonly known as the 'flu', is an infectious disease of birds and mammals caused by RNA viruses of the family Orthomyxoviridae, the influenza viruses [9]. The most common symptoms are chills, fever, sore throat, muscle pains, headache (often severe), coughing, weakness/fatigue Irritated, watering eyes, Reddened eyes, skin (especially face), mouth, throat and nose, Petechial Rash and general discomfort. Influenza may produce nausea and vomiting, particularly in children. Typically, influenza is transmitted through the air by coughs or sneezes, creating aerosols containing the virus. Influenza can also be transmitted by direct contact with bird droppings or nasal secretions, or through contact with contaminated surfaces. Influenza spreads around the world in seasonal epidemics, resulting in about three to five million yearly cases of severe illness and about 250,000 to 500,000 yearly deaths [9].

People who suspected influenza are investigated with medical tests. These commonly include Blood test (white blood cell differential), Chest x-ray, Auscultation (to detect abnormal breath sounds), Nasopharyngeal culture.

In general, treatment with the antiviral medication like oseltamivir (Tamiflu) or zanamivir (Relenza) or amantadine or rimantadine may make disease less severe if the patient starts taking the medicine within 48 hours after the symptoms start. Several vaccines may be use according to the doctor's recommendation [10].

2. DATA WAREHOUSE ARCHITECTURE

Data warehousing is the only viable solution for providing strategic information, it is a simple concept for information delivery. Data warehouse provides an integrated and total view of the enterprise to make the enterprise's current and historical information easily available for decision making. It also presents a flexible and interactive source of strategic information. It is an ideal environment for data analysis and decision support. It is cent percent user-driven, fluid, flexible, interactive, subject oriented, integrated, and nonvolatile and time variant collection of data in support of management's decisions.

Architecture is the proper arrangement of the all software and hardware components. As per requirements of organization arrange those building blocks in a certain way that gives maximum benefit.

Just like any software development project, data warehouse also follow a set of steps to ensure that a working system is delivered on time and to ensure that all user requirement has been fully captures by the design. Architecture of a data warehouse system builds with Source Data components in the left side. The Data Staging component present at the next building block. Those two blocks is under Data Acquisition Area. In the middle Data Storage component that manages the data warehouse data. These components also with Metadata, that also keep track of the data and also with Data Mart. Last component of this architecture is Information Delivery component that shows all the different ways of making the information from the data warehouse available to the user.

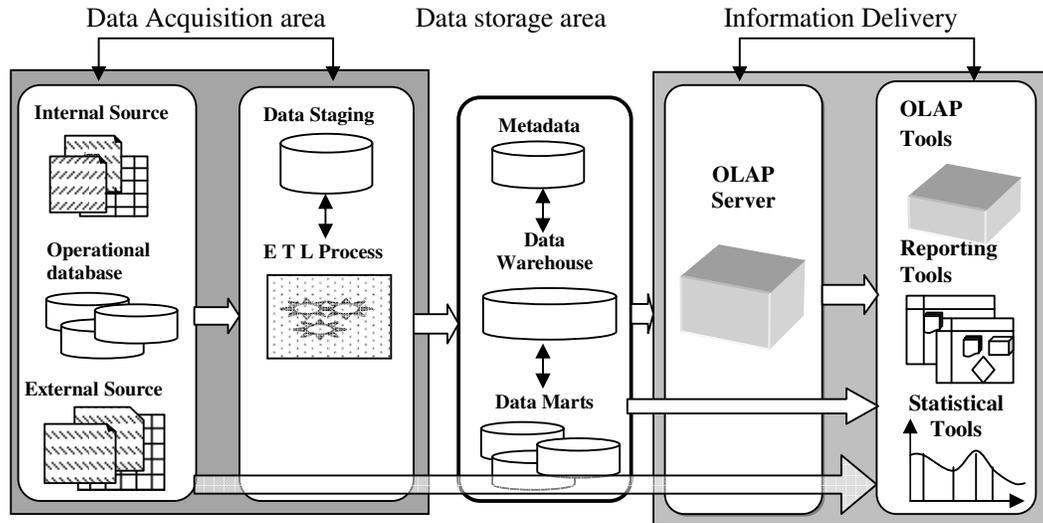


Figure 1: Data Warehouse Architecture

So, there are several major components in data warehouse architecture as follows:

- 2.1. Data Acquisition
- 2.2. Data Storage
- 2.3. Information Delivery

Now those areas are explaining in details as following:

2.1. Data Acquisition

This portion data are coming from multiple sources into the data warehouse, moving all the extracted data to the staging area and preparing the data for loading to the data warehouse. In this portion two major architectural components are source data and data staging. Source data are grouped into four categories – Production Data, Internal Data, Archived Data and External Data. Data are comes from those four source data categories in different formats, so its need to convert those data into data warehouse suitable formats. This is done by the Data Staging portion.

The function and services for this area are:

- 2.1.1. Data Extraction
- 2.1.2. Data Transformation
- 2.1.3. Data Loading

2.1.1. Data Extraction

In this portion, source data may be from different source machines in different data format. Sometimes it may be part of relational database system. Some time it may be in hierarchical data model or just flat files or from spreadsheets. So extract data from source data into a separate physical (group of flat files, or a data-staging relational database, or a combination of both) environment from which moving the data into data warehouse.

2.1.2. Data Transformation

Data conversion is an important function in every system development. In data warehouse data are comes in different formats. So perform a number of individual tasks as of part of data transformation. First clean the data extracted from each source. Cleaning may be correction of misspelling may be conflict between state code and zip code, may be providing default values for missing data elements or elements duplication. Combining pieces of data from different data source is a part of data transformation. When data transformation function ends then the collection of integrated data that is cleaned, standardized and summarized.

2.1.3. Data Loading

Data loading function is done by two distinct groups of task. First time, when complete the design and construction the data warehouse, initially loaded the large amount of data in the data warehouse. And after the data warehouse starts functioning, continue to extract the changes the data source.

2.2. Data Storage

In this stage, data are loaded in to the data warehouse in the day-to-day basic. Data repositories contain the data structure in highly normalize form for fast and efficient processing. Large amount of historical data are needed in data warehouse for analysis. The data storage in data warehouse is kept separate for quick retrieval of individual pieces of information. Data warehouse are read-only data repositories.

The function and services for this area are:

2.2.1. Load data for full refreshes of data warehouse tables.

2.2.2. Optimize the loading process.

2.3 Information Delivery

In this stage the user collect information from data warehouse. To collect the information from data warehouse, information delivery components is use to make it easy to access. Different levels of user are able to collect information from data warehouse. There are different information delivery methods for different user. Ad hoc reports are predefined reports primarily meant for novice and casual user. Provision for complex queries, multidimensional analysis and statistical analysis cater to the needs of the business analysts and power users. Information fed into Executive Information Systems is ment for senior executive and high level managers.

The function and services for this area are:

2.3.1. Allow user to collect information from data warehouse.

2.3.2. Aggregate table for faster execution of results and analysis by using queries enable.

2.3.3. Using Online Analytical Processing (OLAP), it is easy to perform complex analysis.

3. BUILDING HEALTH CARE DATA WAREHOUSE FOR INFLUENZA

For any health care organizations both effective health care and to financial survival is the most important things. Data about the accuracy of diagnoses, effectiveness of treatments, efficient doctors, and proper cost is the crucial things about a health care center. The health care industry is unique in that it needs to bring together efforts to improve the quality of individuals' health with

the effort to cut costs to employers and governments. Recent in India there are several types of influenza viruses spreads in seasonal epidemics. It is quite different from other industry to build the health care center data warehouse for diagnoses influenza. But just like software development project, data warehouse maintain the stage by stage procedure.

Two basic stages for building a health care data warehouse for diagnoses influenza is describe in this paper as following.

- 3.1. Business Analysis.
- 3.2. System Architecture Design.
- 3.3. Data Architecture Design.

3.1. Business Analysis

A data warehouse is an information delivery system for business intelligence. It is solving users' problems and providing strategic information to the user. In the phase of defining requirements, need to concentrate on what information the users need. With the purely top-down approach, the data warehouse will be developed based on the third normal form relational data model. This relational database will form the data warehouse.

The business analysis stages consist of business requirement analysis and business process analysis.

3.1.1. Business Process Analysis

There are four actors in the following case diagram. In that process four actors are Patient, Doctors, Pathologist, and executive manager. Here all patients are coming to the health care center for treatment. According to the symptoms of the patient, the doctors send them to the pathology for some medical test. As per their medical test report, doctors decide the type of flu and then started the treatment. The influenza data warehouse use case diagram is shown in the figure-2 below.

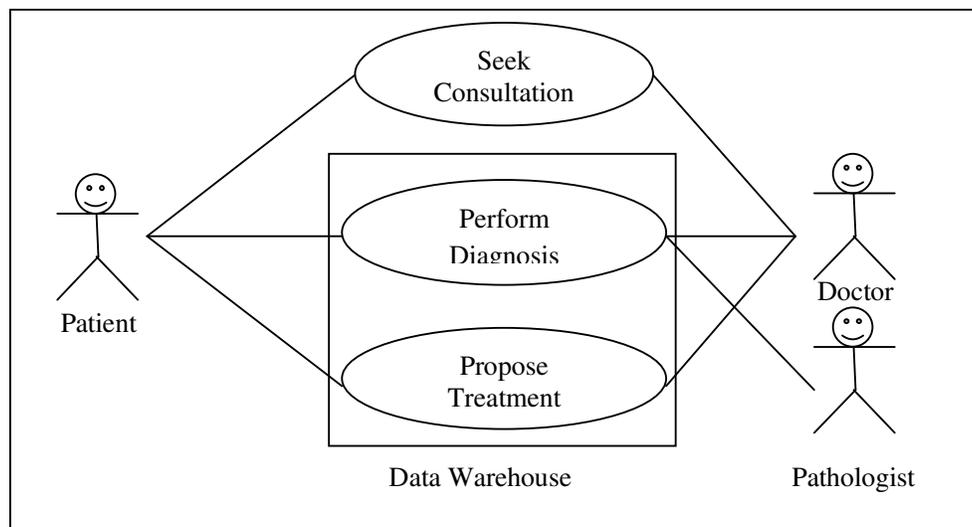


Figure 2: Influenza data warehouse use case diagram.

The interaction between the actors and cases are described in below.

Case 1: Seek Consultation

Patients go to doctor when certain symptoms are noticed by the patient.

Case 2: Perform Diagnosis

The doctors and the pathologist will together perform a series of test like Blood test (white blood cell differential), Chest x-ray, Auscultation (to detect abnormal breath sounds), Nasopharyngeal culture to determine the type of influenza.

Case 2.1: Determine the type of Influenza

These viruses are only distantly related to the human parainfluenza viruses, which are RNA viruses belonging to the paramyxo virus family that are a common cause of respiratory infections in children such as croup but can also cause a disease similar to influenza in adults. There are three types of flu viruses: A, B, and C.

Influenza virus A

Wild aquatic birds are the natural hosts for a large variety of influenza A. Occasionally; viruses are transmitted to other species and may then cause devastating outbreaks in domestic poultry or give rise to human influenza pandemics. The type A viruses are the most virulent human pathogens among the three influenza types and cause the most severe disease. The serotypes that have been confirmed in humans, ordered by the number of known human pandemic deaths, are:

H1N1, which caused Spanish Flu in 1918, and Swine Flu in 2009

H2N2, which caused Asian Flu in 1957

H3N2, which caused Hong Kong Flu in 1968

H5N1, which caused Bird Flu in 2004

H7N7, which has unusual zoonotic potential

H1N2, endemic in humans, pigs and birds [9]

Influenza virus B

Unlike type A flu viruses, type B flu is found only in humans. Type B flu may cause a less severe reaction than type A flu virus, but occasionally, type B flu can still be extremely harmful. Influenza type B viruses are not classified by subtype and do not cause pandemics.

Influenza virus C

This genus has one species, influenza C virus, which infects humans, dogs and pigs, sometimes causing both severe illness and local epidemics. However, influenza C is less common than the other types and usually only cause's mild disease in children

Case 3: Propose Treatment

After determine the types of flu, the doctor starts treatment. People with the flu are advised to get plenty of rest, drink plenty of liquids, avoid using alcohol and tobacco and, if necessary, take medications such as acetaminophen (paracetamol) to relieve the fever and muscle aches associated with the flu. Also gives some antiviral medication to care from flu. The two classes of antiviral drugs used against influenza are neuraminidase inhibitors (oseltamivir and zanamivir) and M2 protein inhibitors (adamantane derivatives). For Influenza prevention uses some vaccine. Prevention for high-risk groups, such as children, the elderly, health care workers, and people who have chronic illnesses such as asthma, diabetes, heart disease, or are immuno-compromised among others.

3.1.2. Business Requirement Analysis

Here some important requirements for health care data warehouse to support Influenza diagnosis and treatment recommended by the doctors. The propose requirements are:

Minimum level of dimensional nature of business data about the patient required where patient details is stored. In the record included portion are: Full Name, Date of Birth, Gender, Age, Marital Status, Address, Contact Number, Occupation, Disease Details, Treatment under which Doctors and etc.

All patient record details must be recognized by unique it may be using ID number which may be declaring as a primary key. Thus may prevent data duplication, and easy to search.

The medical diagnosing function requires to updates patient medical history report, symptoms, drug interaction before and after the patient treatments.

The system must be able to display at both summary and details levels, by which the use (doctors) may get specific idea about the disease and analyze the result.

3.2. System Architecture Design for Influenza Disease

Figure 3 shows that total proposed architecture for the health care data warehouse specific to Influenza disease. Architecture of Influenza specific health care data warehouse system builds with Source Data components in the left side where multiple data are comes from different data source and transform into the Data Staging area before integrated. The Data Staging component present at the next building block. Those two blocks is under Data Acquisition Area. In the middle Data Storage component that manages the data warehouse data. This component also with Metadata, that also keep track of the data and also with Data Marts. Last component of this architecture is Information Delivery component that shows all the different ways of making the information from the data warehouse available to the user for further analysis.

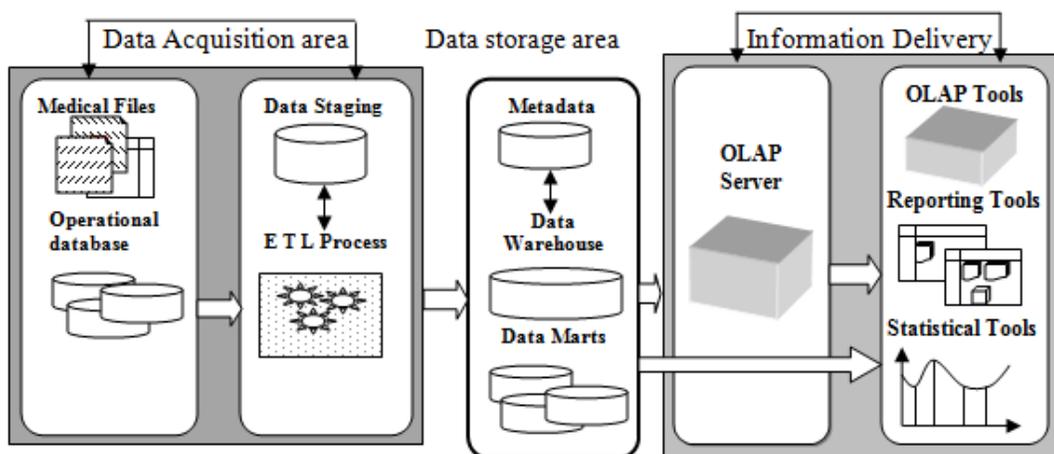


Figure 3: Data Warehouse Architecture for Influenza Disease.

3.2.1. Data Acquisition

This portion data are mainly medical files which are store in Microsoft Access database. Medical files such as patient medical reports, blood tests result, x-ray results, auscultation test results, Nasopharyngeal culture reports, etc. Those data are coming from multiple sources. In data extraction, select data from those source data and moving all the extracted data to the staging area. Now in the Data Transformation portion, map extracted data for the data of the data warehouse. Combining pieces of data from different data source is a part of data transformation. When data transformation function ends then the collection of integrated data that is cleaned, standardized and summarized.

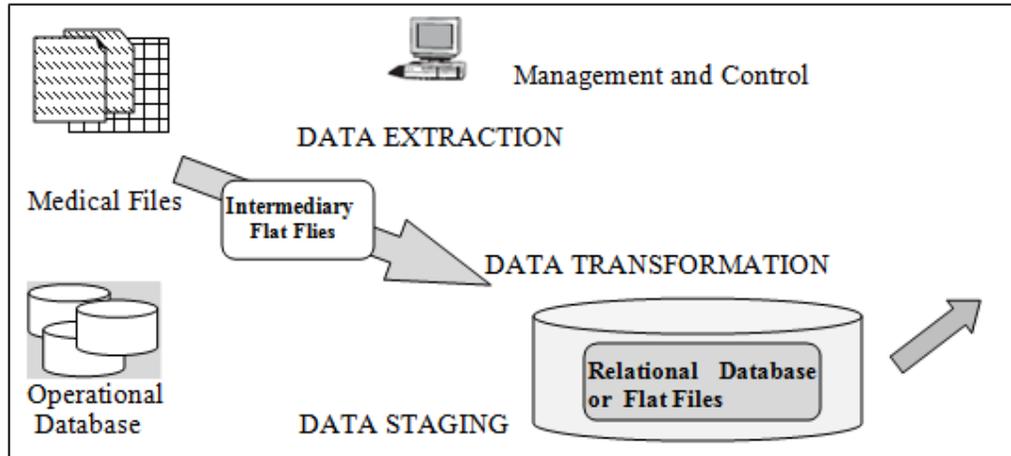


Figure 4: Data acquisition: Influenza Architecture.

In this stage there are a set of functions and services such as:

Data Extraction

Select data from Medical files and determine the types of filters to be applied to individual sources. Generate automatic extract files from operational systems using replication and other techniques. Create intermediary files to store selected data to be merged later. Transport extracted files from multiple platforms. Provide automated job control services for creating extract files. Generate common application codes for data extraction. Resolve inconsistencies for common data elements from multiple sources.

Data Transformation

Map input data to data for data warehouse repository. Clean data, deduplicate, and merge/purge. Denormalize extracted data structures as required by the dimensional model of the Influenza disease data warehouse. Convert data types, Calculate and derive attribute values, Check for referential integrity, Aggregate data as needed. Resolve missing values, Consolidate and integrate data.

3.2.2. Data Storage

This portion, the data from the staging area load into the data warehouse repository. Medical Files data and Microsoft Access Data are loaded in to the data warehouse in the day-to-day basic. Data repositories contain the data structure in highly normalize form for fast and efficient processing. Large amount of historical data of the influenza patients are needed in data warehouse for analysis. The data storage in data warehouse is kept separate for quick retrieval of individual pieces of information. Data warehouse are read-only data repositories.

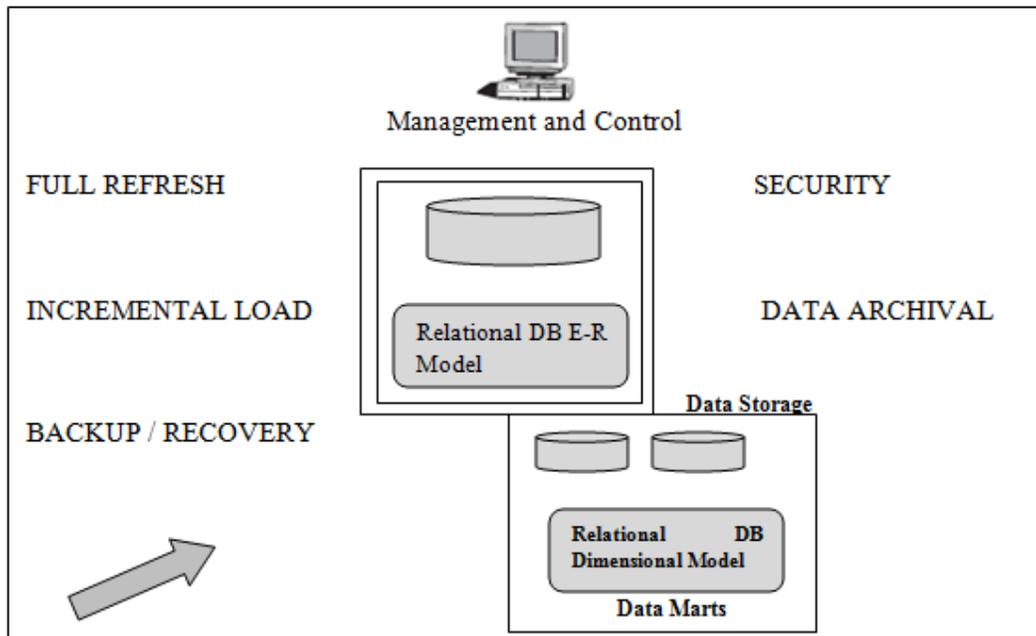


Figure 5: Data Storage: Influenza Architecture

In this there are a set of function and services such as:

Load health care data for full refreshes into data warehouse tables. Perform incremental loads at regular prescribed intervals. Loading details and summarized levels of patient's data into multiple tables. Optimize the data loading process.

3.2.3. Information Delivery

In this stage the Doctors collect information from data warehouse. To collect the information from data warehouse, information delivery components is use to make it easy to access and decision making to access the information directly from the health care data warehouse. There are different information delivery methods for different user. Ad hoc reports are predefined reports primarily meant for novice and casual user i.e. staffs of the health care. Provision for complex queries, multidimensional analysis and statistical analysis cater to the needs of the business analysts and power users i.e. Doctors are this type of user. Information fed into Executive Information Systems is ment for senior executive and high level managers. The primary data warehouse feeds data to proprietary multidimensional databases (MDDBs) where summarized data is kept as multidimensional cubes of information. In influenza data warehouse, when source data is located that time the influenza data warehouse is design. Based on fact table and multiple dimensions table, the star model is adopted. Then from dimension table create hierarchy which is useful to create reports.

These stages there are a set of functions and services such as:

It allows the doctors and other decision maker to brows Influenza data warehouse content. It also provides security for unknown user to access data. It totally hides the complexities of data storage from the user and allows them simply access the data. Reformat the queries automatically for optimal execution. Provide self-service report generation for users, consisting of a variety of

flexible options to create, schedule, and run reports. Queries and reports result set are store for future use. Provide event triggers to monitor data loading and multiple levels of data granularity.

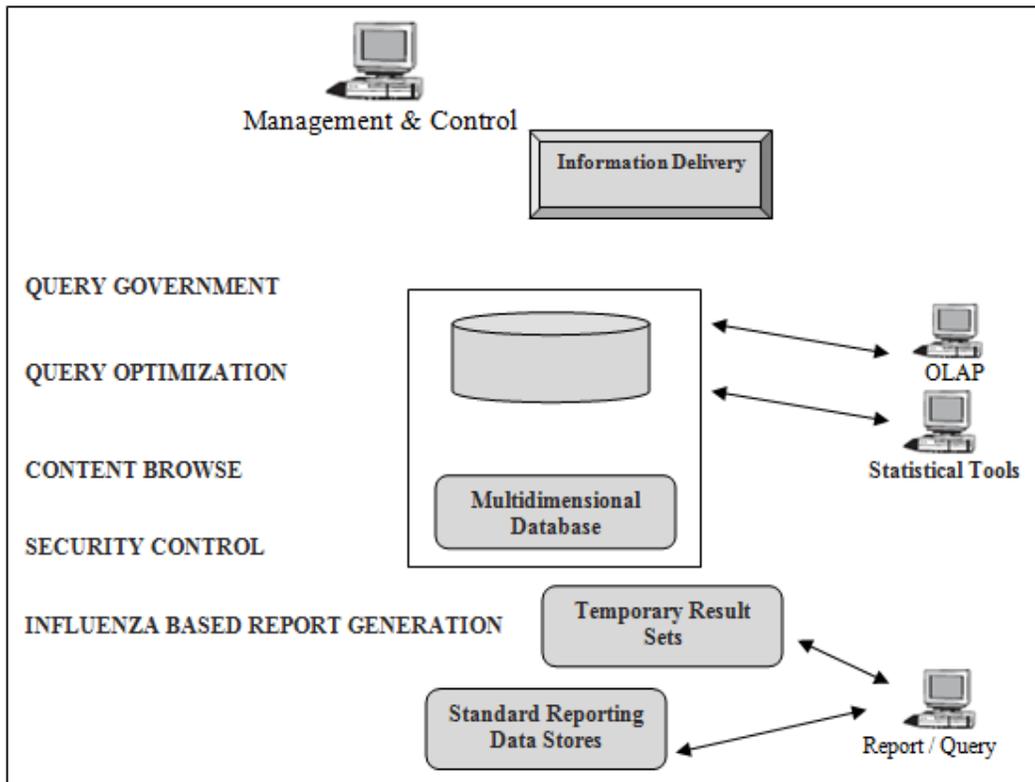


Figure 6: Information Delivery: Influenza Architecture

3.3. Data Architecture Design

The star schema demonstrates the data layer architecture of the health care data warehouse which is shown in Figure 7. The design star schema of health care data warehouse uses a de-normalize schema which contain de-normalize or redundant data. Business Intelligence techniques and data mining may uses facilitate of de-normalized data.

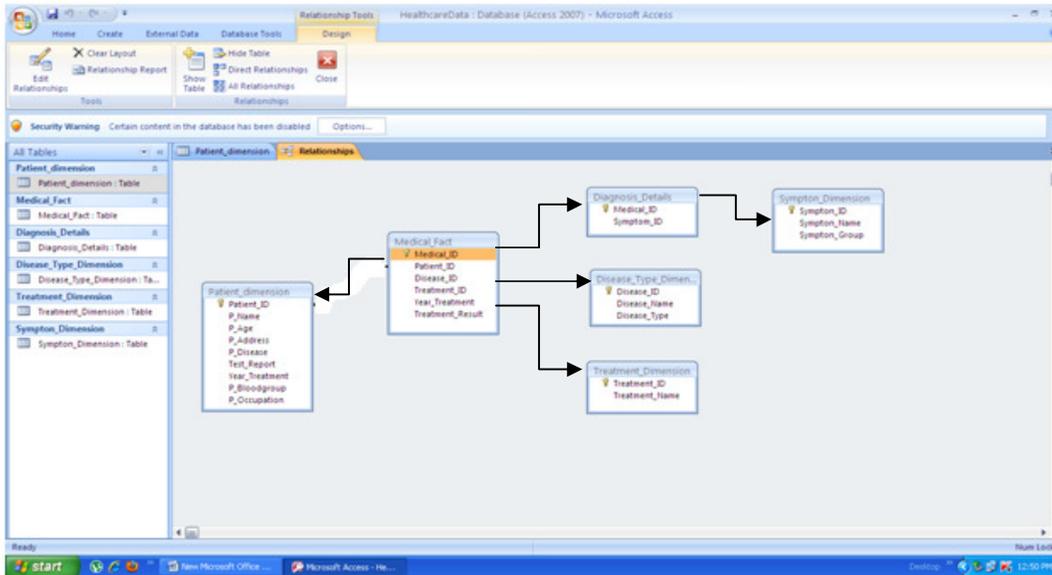


Figure 7: Health Care Data Warehouse Star Diagram

The fact table that describe all about medical report is named Medical Fact. The table consists of Medical_ID, Patient_ID, Disease_ID, Treatment_ID, Year_Treatment, Treatment_Result etc. The symptoms of a patient diagnosed condition and result are stored in Diagnosis_Details table.

Field Name	Data Type	Description
Medical_ID	AutoNumber	
Patient_ID	Number	
Disease_ID	Text	
Treatment_ID	Text	
Year_Treatment	Text	
Treatment_Result	Text	

Figure 8: Medical_Fact Table.

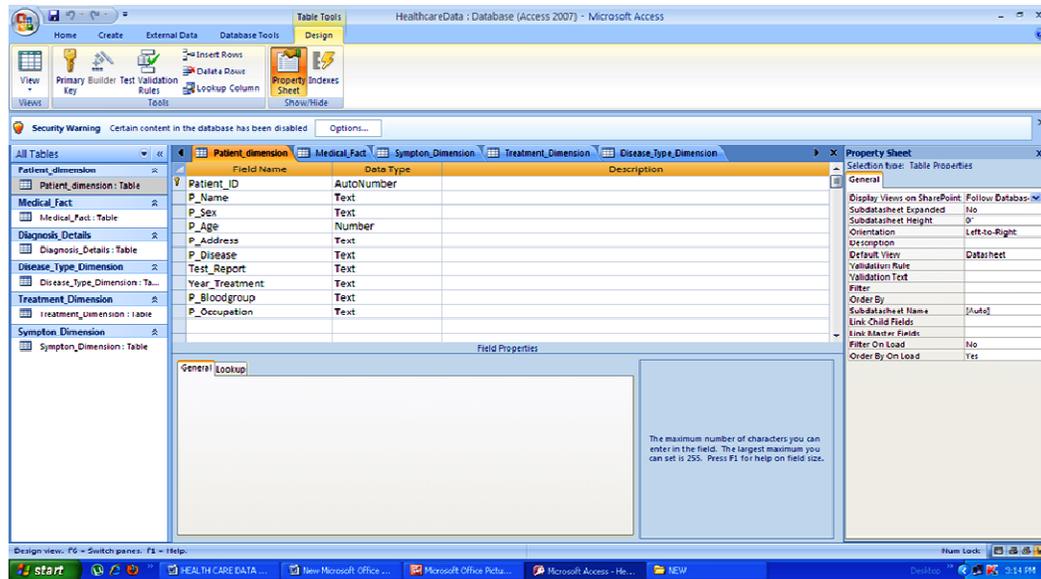


Figure 9: Patient_Dimension Table.

The dimension table that store all details of each entity of every table. In Medical_Fact table the entity are Patient_Dimension, Treatment_Dimension, Symptom_Dimension, Disease_Type_Dimension.

Patient Dimension:

A table that stores patient information, such as patient name, sex, age, address, disease, test report, blood group etc. This data are use for find out which type of influenza (flu) disease.

Treatment Dimension:

A table that is used to stores all possible treatment option.

Symptom Dimension:

A table that is used to stores the entire symptom, the normal condition values and abnormal condition values.

Disease Type Dimension:

A table that is used to stores all the diseases and the types of the diseases. In this case only all types of influenza (flu) related data are stores.

Patient_ID	P_Name	P_Sex	P_Age	P_Address	P_Disease	Test_Report	Year_Treat	P_Bloodgr	P_Occupation
1	Ananda Sarkar	M	28	Dumdum	H5N1	P	2009	A+	Salesman
2	Mainak Ghosh	M	42	Kolaghat	H1N2	P	2009	B+	Businessman
3	Mitali Barman	F	33	Ranaghat	H1N1	P	2010	A+	House wife
4	Somnath Dutta	M	63	Bardhaman	H1N1	P	2010	O+	Businessman
5	Barnali Sen	F	22	Shyamnagar	H5N1	N	2010	AB+	Student
6	Khakon Sarkar	M	58	Kolkata	H5N1	P	2010	A-	Businessman
7	Pradip Nath	M	16	Kolkata	H1N2	P	2010	B+	Student
8	Nabin Ghosh	M	38	Barbhaman	H1N1	P	2010	O+	IT
9	Rita Halder	F	42	Katwa	H1N1	P	2010	AB+	House wife
10	Mousumi Pal	F	27	Berhampur	H5N1	P	2010	A+	Teacher

Figure 10: Data in Patient_Dimension table.

4. CONCLUSION

Developing health care data warehouses, places data quality high on the agenda. Health care related data warehouses is challenging because definitions for individual items must be clear and unambiguous throughout the organization while in practice shared data elements have alternative definitions, owing to a range of different users with a variety of different information needs. This health care industry is the fast developing, most data sufficient industry. For taking that advantage, build an Influenza disease oriented data warehouse. In influenza related data warehouse, integrate between medical files and operational data. Then analysis on patient's medical report data make easy by using OLAP cubes. By using those multilevel viewing data, anyone can analysis the influenza diseases, cost of the treatment, date rate of specific type of influenza and impact of particular drug.

The proposed classifier has been implemented using JAVA (JDK 1.6_16), Microsoft Sql Server 2008, Microsoft Office Access 2007, Microsoft Sql Server Integration Service 2008, Microsoft Sql Server Reporting Service 2008, Microsoft Sql Server Analysis Service 2008.

REFERENCES

- [1] Marleen de Mul , Peter Alons , Peter van der Velde , Ilse Konings , Jan Bakker , Jan Hazelzet (2010) "Development of a clinical data warehouse from an intensive care clinical information system" Computer Methods and Programs in Biomedicine
- [2] Kimball, R. And M. Ross. (2002) "The Data Warehouse Toolkits: The Complete Guid to Dimensional Modelling. 2 nd ed". New York: Wiley.
- [3] Paulraj Ponniah, (2010) "Data Warehousing Fundamentals for IT Professional", John Wiely & Sons.
- [4] Scheese R. Data warehousing as a healthcare business solution. Healthcare Financial Management. 1998.
- [5] Muranaga F, Kumamoto I, Uto Y. Development of hospital data warehouse for cost analysis of DPC based on medical costs. Methods of Information in Medicine. 2007.
- [6] Teh Ying Wah, Ong Suan Sim (2009) "Development of a Data Warehouse for Lymphoma Cancer Diagnosis and Treatment Decision Support".
- [7] Golfarelli, Matteo; Rizzi Stefano (2009). "Data Warehouse Design: Modern Principles and Methodologies.
- [8] Mohamed Salah and Amine Farhat (2010) "Building A Data Warehouse National Social Security Fund Of The Republic Of Tunisia " IJDMS, Vol2, No2.
- [9] Detailed Guide: all about Influenza disease related data are collect from Web site, <http://en.wikipedia.org/wiki/Flu>
- [10] Detailed Guide: all about Influenza disease and types of Influenza related data are collect from Web site, <http://www.webmd.com/cold-and-flu/flu-guide/advanced-reading-types-of-flu-viruses>

Author

Rajib Dutta received M.Tech. in Computer Science and Engineering under West Bengal University of Technology, India in 2011. He has more than 3 years of teaching and industry experience and presently working as an Assistant Professor at Global Institute of Management and Technology, Krishnanagar, Nadia, West Bengal, India. His fields of interest are Database Management System, Data Mining, Data Warehouse and Distributed Database.

