Faculty of Engineering & Information Technology

32113 ADVANCED DATABASE

Assignment - 2

SUBJECT COORDINATOR - DR. WEI LIU

SUBMITTED BY:

Vishnu Mohan Edala Dhanraj	13069909
Saminathan Raghavan	13075597
UjashKumar Leuva	12739975

Table of Contents

SOURCE - BACKGROUND ***3***

PROBLEM STATEMENT ***3***

AIM ***4***

INTRODUCTION ***4***

KEY DEFINITION ***5***

DATA MARTS ***5***

DATA WAREHOUSE - APPROACHES ***5***

INMON'S TOP-DOWN APPROACH ***5***

KIMBALL'S BOTTOM-UP APPROACH ***6***

DATA WAREHOUSE PLANNING ***6***

TECHNICAL REQUIREMENT ***7***

LITERATURE REVIEW ***8***

ARTICLE 1 ..... ***8***

ARTICLE 2 ..... ***8***

ERD DIAGRAM ***9***

STAR SCHEMA ***10***

ETL LAYER ***10***

DATA MART ***16***

QUERY ***20***

FUTURISTIC APPROACH OF CANCER AID - DATA WAREHOUSE ***21***

CONCLUSION ***22***

SOURCE - BACKGROUND

Cancer Aid, an application to empower cancer patient and caregiver developed by Australian oncologists. This application is configuration to help understanding battle malignancy by following their side effects, feelings and arrangements, and keeping them associated with family, companions and in addition worldwide network. The thought for an application started couple of years prior, now is enhancing the lives of in excess of 25,000 patients in 53 nations. Patient can log growth or treatment related side effects inside the application from anyplace, even while amid the chemotherapy.

The major problems are encounter during the data warehouse development for the Cancer Aid, as there is huge chunk of data of 25,000 patients from 53 countries from different stage of cancer i.e. chemotherapy, hormone therapy, targeted cancer drugs and so on. The data maintenance is major issues as patients have started using this application to greater extent that the users are increasing exponentially over the time. The organisation is unable to manage this enormous amount of data and so they require a decision support system to organise and manipulate the data as per the requirement and also endure the markets growth.

PROBLEM STATEMENT

Cancer Aid - An Application on Health Management System that has Users being Exponentially grown in the past two years would like to build a data warehouse to have an intensive monitoring unit in their Cancer treatment and rehabilitation centre. The CEO of that firm wants to have his investment worthy by analysing and predicting the past data history where the data mainly comprises of Patients who have already visited and patients who had been treated and had positive responses about the ambience during their stay inside the hospital. The CEO here wants to have a pattern of analysis and a general report based on the past history of data. But there are a few issues in getting things right.

Some of the problems listed are

- Integration of Multiple Sources of Data:

  Any Non- trivial repository of data will have information coming from variety of places like database, public records etc by 2019, three fourths of the data comes from multiple, external data sources hence the first thing to be checked against each other is to ensure their consistency and synchronising of data into a single centralised repository.

  To overcome this Metadata has to be created, as it gives all the necessary information of data source i.e. source table, target table, column description and incase of any discrepancies the data has to be converted to its prescribed or target format.

- Inconsistently Build OLAP reports due to Poor Data Quality:

  The main problem with manual process is inconsistent reporting. Manual built reports may be inconsistent either by faulty or missing information they contain or different standards of information they possess. This will raise issues in dependent functional teams.

  To subjugate this , the data has to be sent through the ETL layer number of times until the final data obtained is consistent and ready to use.

- Lack of Business alignment and Failure tolerance:

  Data warehouse design must be built in align with its business and its various technical departments. Top - down approach from top to lowest position possible can eliminate this problem and also help in providing information in hierarchical way to its users. In case of any Failure

happened the loss of entire data is catastrophic. Hence the technical teams should have a backup (azure snapshot) , or recovery with them, so that they can start from the failed step and complete them without any further issues .

To overcome alignment issues, various data warehouse approaches are suggested . for e.g. :Inmon's Top down approach and Kimball's Bottom Up approach.

AIM

To Build and Maintain a Data warehouse that helps the CEO and Management of CANCER AID to decide which area to be concentrated and development of further technical facilities , Medical treatment investments from the past history of data collected from the Application. These collective data can help in deriving conclusions of total revenue of the firm, doctors preferences and treatment, treatment procedures costs and statistics of showing cured patients from Past Patients History . These information aids the decision makers or the higher management to have a detailed report on the number of increase in cancer patients over the past years. If the Statistics shows that the number of cancer patients that has grown considerably over the past year , then the management will significantly work on investing the patient utilities , maintenance facilities and research units accordingly.

INTRODUCTION

Data Warehousing is the design and implementation of processes, tools, and facilities to manage and deliver complete, timely, accurate, and understandable information for decision making (Ballard et al. 1998). An information distribution centre is a subject-situated, incorporated, time-variation and non-unpredictable gathering of information in help of administration's basic leadership process. Data warehouse is a substantial store of information accumulated from an extensive variety of sources inside an organization and used to guide management decisions. The data warehouse is used as follows:

- Access of data from various heterogeneous data source
- Extract, Clean and filter the data into appropriate form of data
- Store the data into database used for query, reporting as well as data analysis

The modern web has rapidly become the viable platform for creating high quality and stunning applications for health to deliver better care and boost productivity while simultaneously reducing costs and waste due to the U.S. government's Affordable Care Act (ACA). Cancer Aid are celebrated universal portable application for tumor persistent which enables patient to get to customized, restoratively solid data about malignancy; track their indications, sentiments and in addition arrangement in a diary; designate loved ones who will bolster them; and access a news channel with data and strong counsel and network of growth tolerant and their parental figures.

The approach of this report is to discuss the case study about the data needs for an organisation names Cancer Aid and helps the company by providing various solutions for them.

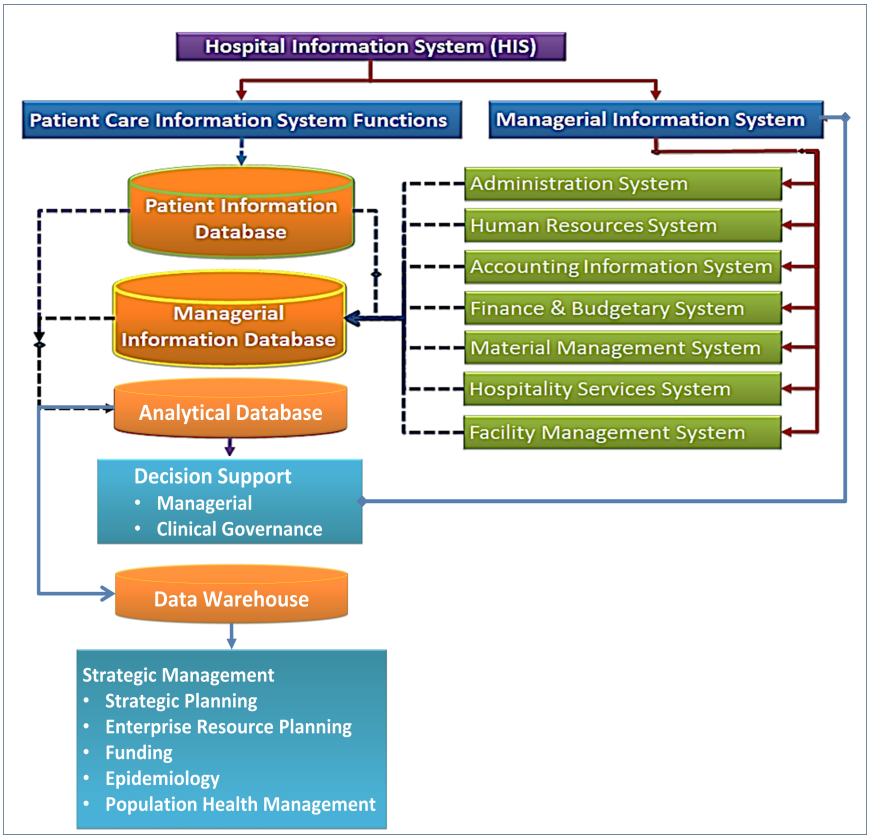


Figure 1. Data warehousing Architecture for CancerAid

KEY DEFINITION

DATA MARTS

A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales, Finance, or Marketing. Data marts are often built and controlled by a single department within an organization ('Data Mart Concepts', Oracle Business Intelligence Standard Edition one tutorial).

DATA WAREHOUSE - APPROACHES

When it comes to data warehouse approaches, there are two most widely discussed methods introduced by Bill Inmon and Ralph Kimball.

INMON'S TOP-DOWN APPROACH

This approach is defining the centralised repository for the entire enterprise application which is used to store the data at atomic/ lowest level. In this approach, dimensional data marts are developed only after the data warehouse.

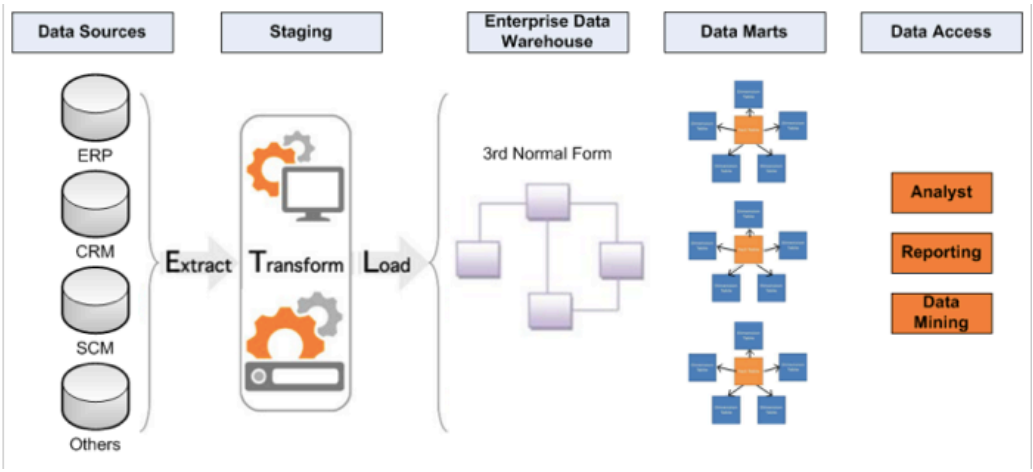


Figure 2: Inmon's Data Warehouse Architecture

KIMBALL'S BOTTOM-UP APPROACH

In Kimball's bottom-up approach, data marts are created before the data warehouse development. These providing a minimum view into the organisational data and, these combines in to the larger data warehouse. It defines the copy of transaction data specifically structured for query and analysis.

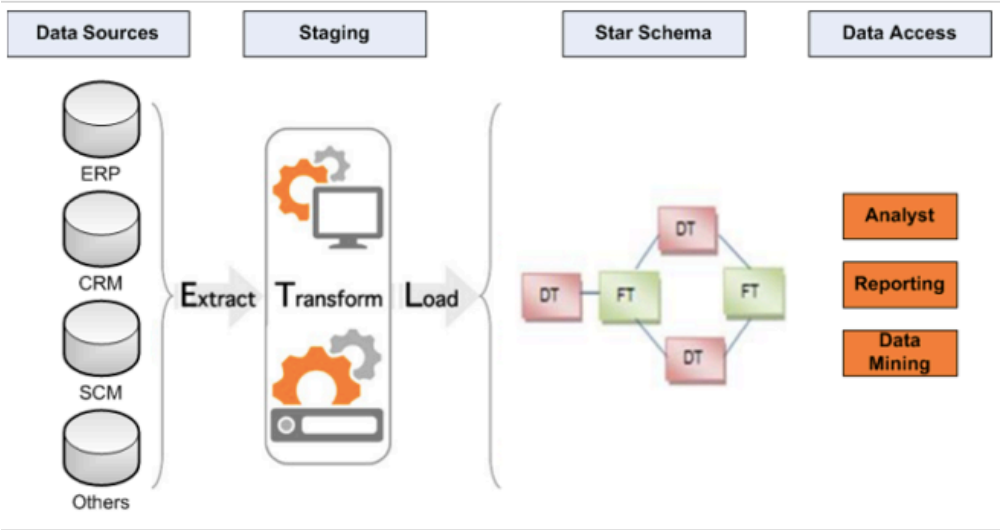


Figure 3: Kimball's Data Warehouse Architecture

DATA WAREHOUSE PLANNING

The Diagram illustrates the data warehouse system for Cancer Aid in figure 4.

Data Sources: Data will be available from various sources such as database (e.g. Oracle database), Flat files (e.g. CSV, Excel file), web services as well as other resources such as RSS feed.

Staging: All the accessible information is gone through the ETL (Extraction-Transformation-Load) layer. This layer filters the heterogeneous information and enhances the quality of information that enters the data warehouse. ETL is in charge for ensuring that the information entering in is exact, clean, and standardized.

EDW & Data Marts: The data is now load into enterprise Data Warehouse and can be drawing the data directly from operational, external or both sources. However, this data can be used by various department of organisation i.e. Finance, Maintenance, production and so forth.

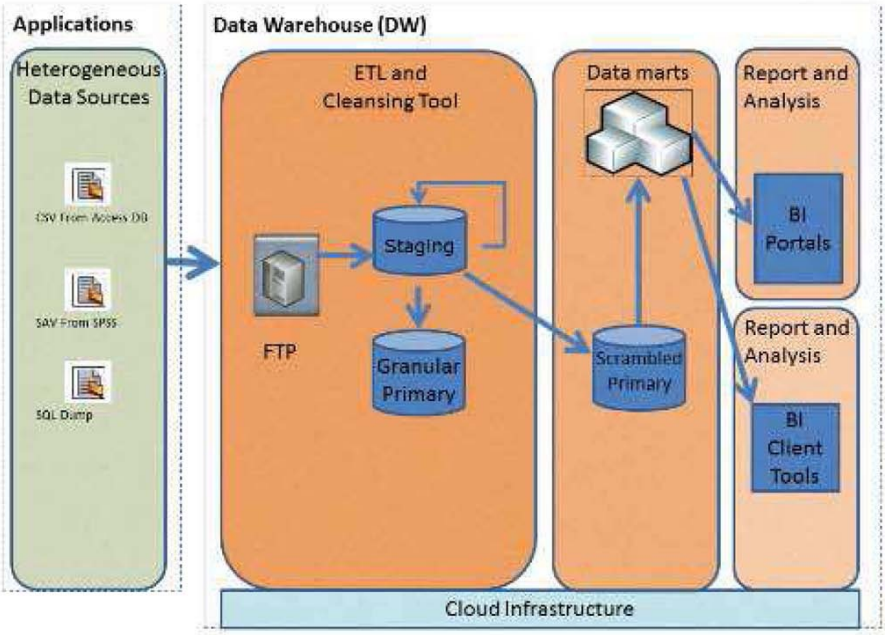


Fig 4: Architecture Diagram for CancerAid- Datawarehouse

Data Access: After loading the data into data warehouse, several other tools mostly used for data mining, data analysis as well as insight reporting.

TECHNICAL REQUIREMENT

It is mandatory to have certain requirements satisfied before building the data warehouse. Therefore better health management system for Cancer aid. Data warehouse should have the below enlisted requirements.

Operating System:

- The operating System to withstand all the other data entries should be around windows NT version - 4.0 and above
- Either Windows 9X should be supportable.

Database:

- My SQL with windows latest version along with SSIS (SQL Server Integration Service) packages for bulk uploading of data into the data warehouse.
- Visualisation technique tools to design and proper functioning of ETL packages.
- Design GUI, to have user friendly monitoring systems for better visualisation.

- Refreshment of data at regular intervals to have data up to date.

LITERATURE REVIEW

ARTICLE 1

Sahoo P., Mohapatra S., & Shih-lin W. 2016, 'Analyzing Healthcare Big Data with Prediction for Future Health Condition', *IEEE Access,* vol. 4, pp. 9786-9799

Healthcare Management is a significant area of interest for the future perspective. Advancements in this field is not only honoured but also is much needed. A lot of resources are already dedicated towards further research in this area. New technologies are being used as a catalyst for improving and upgrading Healthcare Management systems. Although there is no full proof technique or application available and accurate results are very difficult to guarantee. Ideas and improvements are being worked out perpetually towards this sole purpose. This paper likewise acknowledges the existing problems faced by real world in healthcare management systems and addresses some promising mechanism towards the improvement of this system. Now while design thinking about the entire idea of improving the healthcare management system there are a lot of distinct factors which need to be taken into consideration. All these factors are extremely sensitive and equally significant. In our case, the healthcare management system deals with several types of data chunks like clinical reports, notes that doctors develop, data from different body sensors constantly keeping a real time record of patient's body and other diagnosis. Data chunks are created in real-time perpetually, now it is very significant to manage these kinds of sensitive data with extreme care and integrity. These kinds of data are also time-sensitive as our real aim is to predict the problem as early as possible for its proper solution or treatment. Data here can take different forms like it could be structured or also unstructured, Big data is the best way to deal with these data chunks assuring its integrity. Big data can be clubbed with cloud services and other data analysis tools to get the best out of this data. This paper studies this particular idea of using big data with cloud services and data analysis to analyse the huge amounts of data created by healthcare management systems. It also suggests a specific data collection mechanism which makes the entire process efficient and then techniques like correlation analysis to make the data usable for analysis. It also describes in brief about a prediction model designed which can be actually used to predict the future health conditions based on the patient's current health status. Given the sensitivity of the area all of the proposed techniques and mechanism are firstly tested thoroughly and run in simulations with some real data in cloud environments making the best use of cloud service to fulfil our purpose. We can easily observe extreme improvements like this model gives about 98% accuracy in its predictions while making an efficient use of the Computer resources to give out quick, accurate and reliable results.

ARTICLE 2

Boon Keong S. 2013, 'An application of a Healthcare Data Warehouse System', *IEEE, vol. 1, no. 13, pp.269-273*

Data warehousing is a significant part of any organization or management system because it is responsible for making management decisions. It is responsible for collection of data from several distinct

sources and storing them in an organized or manageable way. Now Data warehousing also provides various features like data reports, data analysis and data management tool. The healthcare sector really develops a lot of data from distinguished sources which can be managed with the help of data warehousing. This paper focuses on various techniques to load data into the data warehouse in an effective way. These processes involve cleansing of widely collected data that being an essential task for accurate and qualitative data warehouse results. It also focuses on the process of data extraction and data transformation including description of various stages that the data is processed through in this entire process. This paper has given a data warehousing solution that can serve the purpose of generating statistical reports including ad-hoc reporting on a client-server platform and also on web-based platform. It does conclude with a very prospective and promising mechanism of using data warehousing for cleansing, extracting, transforming and managing huge amounts of data generated by healthcare systems.

ERD DIAGRAM

The below ERD diagram shows the entire relationships between different entities in an Hospital Management System i.e. Cancer Aid.

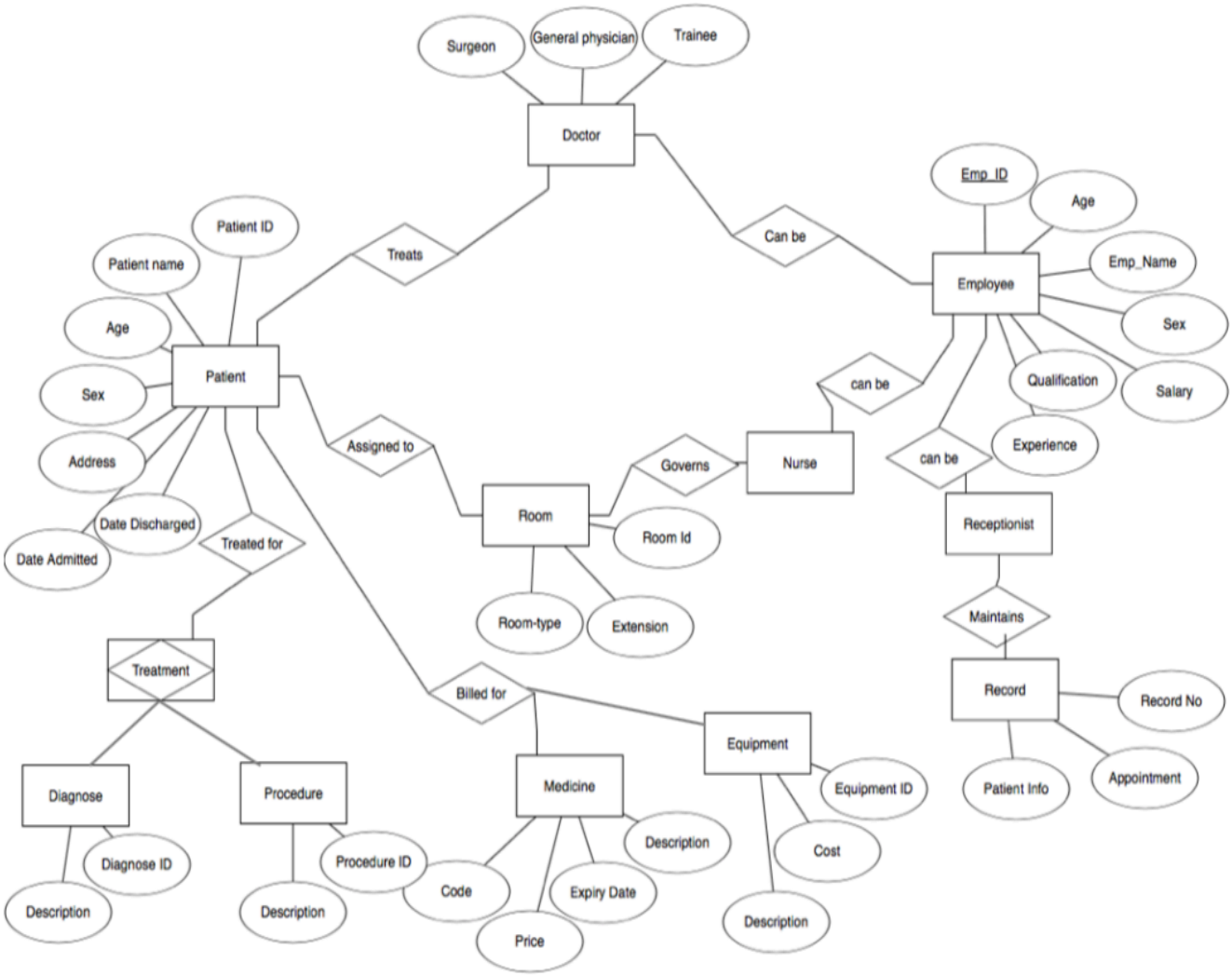


Figure 5 - ERD diagram for CancerAid - Management System.

STAR SCHEMA

The Star Schema for the Cancer Aid i.e. Patient Record Management System can be shows as below,

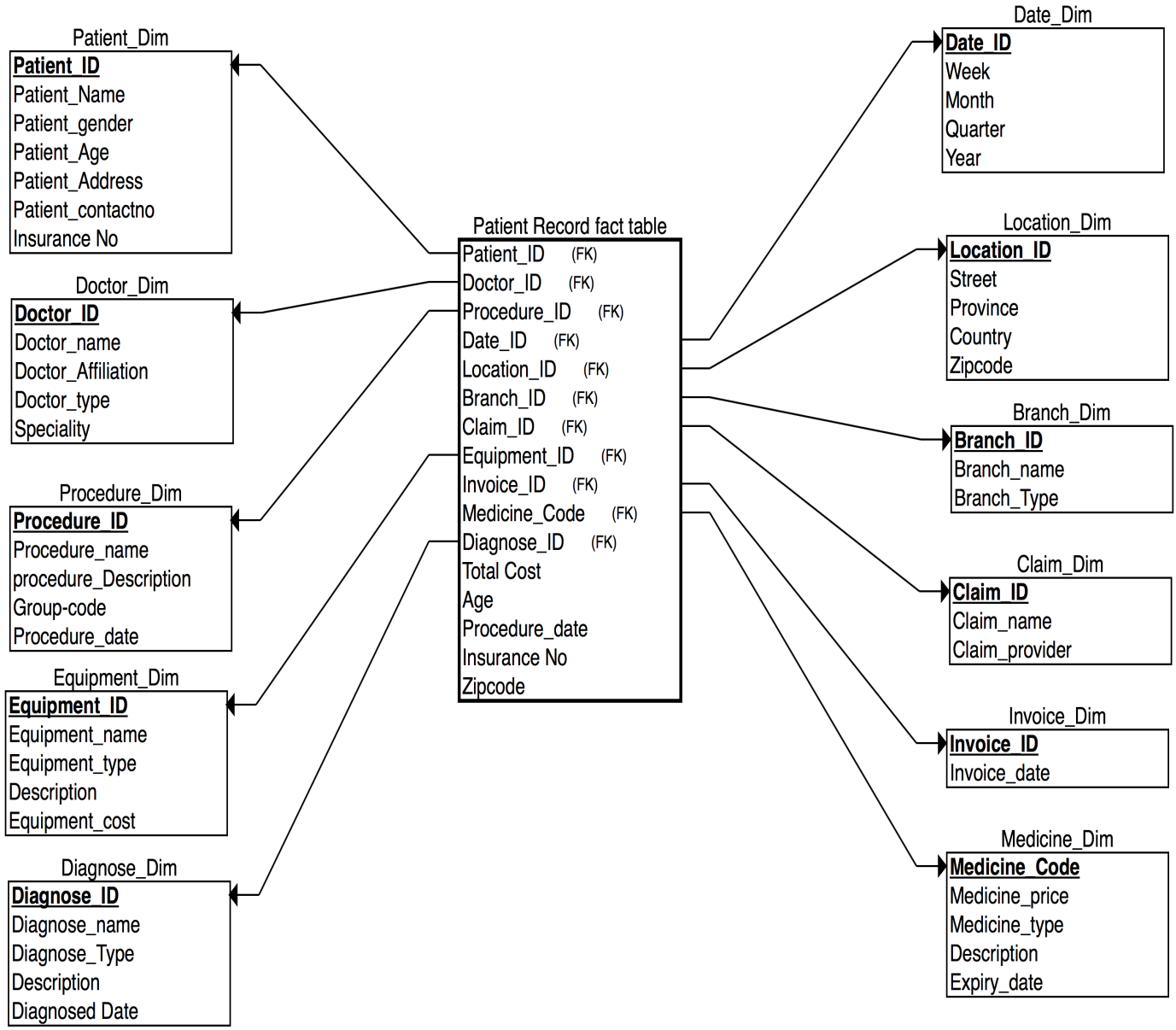


Figure 6 - Star Schema of Cancer Aid Health care management System.

ETL LAYER

ETL remains for Extraction, Transform and Load. Every one of the three-database works that are consolidated into one instrument to haul information out of one database and load it into another database. Information extraction is a procedure of perusing information from a database. The principle reason for information extraction is to recover the information from various datasets into one with least sources. While Transformation is a procedure of changing information from one table to another forbidden configuration. Information change contains information approval, information dismissal and in addition information mix. Stacking is the last procedure of ETL layer. Load work composes the subsequent information. i.e. the changed information to focused information dataset.

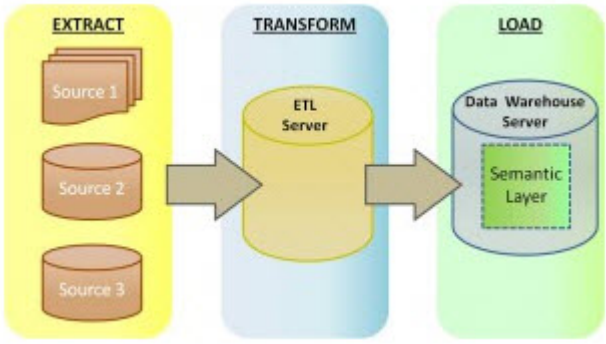


Fig 7: ETL Process

The following data gives a brief description of listed data sources,

DATA Source	Business Owner	IS owner	Platform	Location	Data Source Description
Cancer Aid	Operations Team	CEO	Corporate Server	Sydney	Data source for identifying the pattern of patients and their respective successful treatment.

The consolidated description of the source to target tables can be used to create and build the initial edited source and working table which is completely similar to that of one in the DataMart. These table contain the data in the source format as shown below.

Target Table	Target Column	Data Type	Length of Data	Column Description	Source System	Source Table	Data Source Field	Data Transform description
Patient	Patient	Integer	16	Unique ID				Surrogate Key
Patient	Patient Name	Text	20	Name of the patient	Patient Record	Patient	Name	
Patient	Patient Gender	Text	6	Gender of the patient	Patient Details	Patient	Gender	
Patient	Patient Address	Text	30	Address of the patient	Patient Details	Patient	Address	
Patient	Patient Contact no	Integer	10	Phone no of patient	Patient Details	Patient	Contact No	

Target Table	Target Column	Data Type	Length of Data	Column Description	Source System	Source Table	Data Source Field	Data Transform description
Patient	Insurance ID	Integer	8	Insurance ID of patient	Patient Details	Patient	Insurance ID	
Doctor	Doctor_ID	Integer	16	Unique ID	Doctor Details	Doctor	ID	Surrogate key
Doctor	Doctor_name	Text	20	Name of Doctor	Doctor Details	Doctor	Name	
Doctor	Affiliation	Text	20	Affiliation of doctor	Doctor Details	Doctor	Affiliation	
Doctor	Type	Text	10	Doctor type	Doctor Details	Doctor	Type	
Doctor	Speciality ID	Integer	10	Speciality _ID	Doctor Details	Doctor	Speciality Group	
Procedure	Procedure_ID	Integer	10	Procedure ID	Patient Details	Procedure	ID	Surrogate key
Procedure	Procedure_ Name	Text	20	Procedure name	Patient Details	Procedure	Name	
Procedure	Procedure_ Description	Text	20	Description of Procedure	Patient Details	Procedure	Description	
Procedure	Group Code	Integer	10	Procedure code	Patient Details	Procedure	Code	
Procedure	Date	Date	10	Date of procedure	Patient Details	Procedure	Date	
Equipment	Equipment_ ID	Integer	10	Unique ID	Patient Details	Equipment	ID	Surrogate Key
Equipment	Equipment_ Name	Text	20	Name of the equipment	Patient Details	Equipment	Name	
Equipment	Equipment_ type	Text	10	Type of the equipment	Patient Details	Equipment	Type	
Equipment	Description	Text	20	Description	Patient Details	Equipment	Description	
Equipment	Equipment_ Cost	Integer	6	Usage Cost	Patient Details	Equipment	Usage Cost	
Diagnose	Diagnose_ID	Integer	10	Unique ID	Patient Details	Diagnose	Unique Id for Diagnosis	Surrogate Key
Diagnose	Diagnose_Name	Text	10	Name of the Diagnosis	Patient Details	Diagnose	Name of	

							Diagnosis	
Target Table	**Target Column**	**Data Type**	**Length of Data**	**Column Description**	**Source System**	**Source Table**	**Data Source Field**	**Data Transform description**
Diagnose	Diagnose_Type	Text	10	Type of Diagnosis Method	Patient Details	Diagnose	Type of diagnosis	
Diagnose	Description	Text	20	Description of the diagnosis type	Patient Details	Diagnose	Description of the Diagnosis Method	
Diagnose	Diagnosed Date	Date	10	Diagnosis Date	Patient Details	Diagnose	Diagnosed Date	
Date	Date_ID	Date	10	Date Details	Patient Details	Date	Date_ID	Surrogate key
Date	Week	text	10	Week of admit details	Patient Details	Date	Week	
Date	Month	text	10	Month Details	Patient Details	Date	Month	
Date	Quarter	text	10	Quarter of the year details	Patient Details	Date	Quarter	
Date	Year	Integer	10	Year details	Patient Details	Date	Year	
Location	Location_Id	Integer	10	Unique ID of location	Patient Details	Location	Unique ID of location	Surrogate key
Location	Street	Text	10	Street Details	Patient Details	Location	street	
Location	Province	Text	10	Province Details	Patient Details	Location	province	
Location	Country	Text	10	Country details	Patient Details	Location	Country	
Location	Zip code	Integer	10	Zip code Details	Patient Details	Location	Zip code	
Branch	Branch_ID	Integer	10	Unique ID of the branch	Hospital Details	Branch	Branch_ID	Surrogate Key

Target Table	Target Column	Data Type	Length of Data	Column Description	Source System	Source Table	Data Source Field	Data Transform description
Branch	Branch_Name	Text	10	Name of the Branch	Hospital Details	Branch	Name	
Branch	Branch_Type	Text	10	Type of the branch	Hospital Details	Branch	Type	
Claim	Claim_ID	Integer	10	Unique Claim ID	Patient Details	Claim	Unique claim Id	Surrogate key
Claim	Claim_Name	Text	10	Claim provider Type	Patient Details	Claim	Claim Type	
Claim	Claim_Provider	Text	10	Claim Provider Name	Patient Details	Claim	Claim provider	
Invoice	Invoice_ID	Integer	10	Unique invoice ID	Management details	Invoice	Invoice_ID	Surrogate key
Invoice	Invoice_date	Text	10	Invoice Generated date	Management Details	Invoice	Invoice_date	
Medicine	Medicine Code	Integer	10	Medicine Unique Code	Patient Details	Medicine	Medicine_Code	Surrogate Key
Medicine	Medicine Price	Integer	10	Price of the medicine	Patient Details	Medicine	Medicine_Price	
Medicine	Medicine Type	Text	20	Type of the medicine	Patient Details	Medicine	Medicine_Type	
Medicine	Description	Text	20	Description of the medicine	Patient Details	Medicine	Medicine_description	
Medicine	Expiry Date	Date	10	Expiry date of the medicine	Patient Details	Medicine	Medicine_Expiry Date.	

Following table shows the entities in the Fact table as shown in the star schema ,

Derived Fact Name	Derived Fact Description	Type	Formula	Constraints	Transformation
Total Cost	The total bill that will be paid by the Patient	Integer	(Procedure Cost* no of times Visited)+Medicine Cost + Doctor fees + Admit charge* No. of days)	Out patients	
Patient_ID	Unique ID for patient	Integer			Surrogate key in Patient dimension table
Doctor_ID	Unique ID for the Doctor	Integer			Surrogate key in Doctor Table
Procedure_ID	Unique Id for the procedure	Integer			Surrogate key in Procedure Dimension table
Date_ID	Date details	Integer			Surrogate key in Date Dimension table
Location_ID	Unique Location Details	Integer			Surrogate key in Location Dimension table
Branch_ID	Unique Branch Details	Integer			Surrogate key in Branch Dimension table
Claim_ID	Unique Claim ID	Integer			Surrogate key in Claim Dimension table
Equipment_ID	Unique Equipment ID	Integer			Surrogate key in Equipment Dimension table
Invoice_ID	Invoice generated ID	Integer			Surrogate key in Invoice Dimension table
Medicine_code	Unique Code for medicine	Integer			Surrogate key in Medicine Dimension table
Diagnose_ID	Diagnose_ID	Integer			Surrogate key in Diagnose Dimension table

Age	Age of the patient	Integer			Patient Age from patient Fact table
Procedure _Date	Date of the procedure	Integer			Procedure date of patient from patient fact table.
Insurance_ ID	Unique Insurance ID	Integer			Surrogate key in Insurance Dimension table
Zip code	Zip code details	Integer			Zip code details from Patient fact table

DATA MART

Data mart is completely a subject oriented data repository that stores data and uses the extracted set of information to assist and support the needs of a particular business department. These exist within a single organisational data repository. The response time of the end users has been increased greatly to a different level , as the data marts provides all the necessary information needed particularly for the selective group of users.
It is basically a condensed and focussed view of data warehouse that displays the specifications needed for the business unit within an organisation. Thus different departments such as marketing , accounting , etc.. uses different data marts.

Here in our case study , Health care data mart is created to support the whole health management system which helps the firm to gather information of all the patients and predict patterns of their treatment and its measures.
The First and foremost reason of using data marts instead to Data warehouse is to make use of the information without any other resource consumption by the end users. It promotes the end user response time , as it primarily focuses on raw data that the department primarily look out for.

Data marts uses the similar kind of processes as that of a data warehouse. Both star schema and snowflake methodologies can be used to design a star schema. Both these tables has Fact tables and Dimensional tables linked together , which in turn establishes the relations between them and other related information. These association provokes which schema suits this perfectly.

Analysing the data structures and its comparative measures , Data marts and data warehousing falls under same categories which differs only by few measures. Data warehouse holds information of the entire firm, whereas data marts holds the collective information of that particular group or department in the firm.
The similarities and differences of data ware house and data marts are as follows,

- Data warehouse stores collective data to the system, whereas data mart stores specific kind of data to it.
- Data Mart usually integrates very few specific data into the system feed, whereas data warehouse usually integrates the entire data into it.

- It will be time consuming and need lot of work involved in building a data warehouse as the data and its link between respective dimensions are to be considered. But data mart follows simple process as it involves quite a few specifications as it focusses on small area.
- Maintaining Data warehouse is complex and the relationships between data stores is higher that the data marts scope.
- The names of each tables and its descriptions can be listed as below,

Table Name	Column Name	Data Type	Length	Null values	Column Description	PK	PK order	FK
Patient	Patient_ID	Integer	16	N		Y	1	
Patient	Patient_Name	Text	20		Name of the patient			
Patient	Patient Gender	Text	6		Gender of the patient			
Patient	Patient Address	Text	30		Contact Address of the Patient			
Patient	Patient Contact no	Integer	10		Phone No of the patient			
Patient	Insurance ID	Integer	8		Patient Insurance ID			
Doctor	Doctor_ID	Integer	16	N	Doctor ID	y	1	
Doctor	Doctor_name	Text	20		Name of the Doctor			
Doctor	Affiliation	Text	20		Affiliation of the Doctor			
Doctor	Type	Text	10		Type of Doctor			
Doctor	Speciality ID	Integer	10		Specialisation of doctor			
Procedure	Procedure_ID	Integer	10	N	Unique ID of procedure	y	1	
Procedure	Procedure_Name	Text	20		Procedure Name			
Procedure	Procedure_Description	Text	40		Description of procedure			
Procedure	Group Code	Integer	10		Group code for the procedure			
Procedure	Date	Date	10		Date of the procedure			
Equipment	Equipment_ID	Integer	10	N	Equipment-unique ID	y	1	

Equipment	Equipment_Name	Text	20		Name of the Equipment			
Equipment	Equipment_type	Text	10		Type of the Equipment			
Equipment	Description	Text	20		Equipment Description			
Equipment	Equipment_Cost	Integer	6		Equipment-Cost			
Diagnose	Diagnose_ID	Integer	10	N	Unique Diagnose ID	y	1	
Diagnose	Diagnose_Name	Text	10		Diagnose details			
Diagnose	Diagnose_Type	Text	10		Type of Diagnosis method			
Diagnose	Description	Text	20		Diagnosis Description			
Diagnose	Diagnosed Date	Date	10		Diagnosis Date			
Date	Date_ID	Date	10	N	Date ID	y	1	
Date	Week	text	10		Week details			
Date	Month	text	10		Month Details			
Date	Quarter	text	10		Quarter Details			
Date	Year	Integer	10		Year of the patient admittance			
Location	Location_ID	Integer	10	N	Unique Location ID	y	1	
Location	Street	Text	10		Street details			
Location	Province	Text	10		State Details			
Location	Country	Text	10		Country Details			
Location	Zip code	Integer	10		Zip code Info			
Branch	Branch_ID	Integer	10	N	Branch - unique ID	y	1	
Branch	Branch_Name	Text	10		Branch Name			
Branch	Branch_Type	Text	10		Type of Branch			
Claim	Claim_ID	Integer	10	N	Unique Claim ID	y	1	
Claim	Claim_Name	Text	10		Claim Name details			

Claim	Claim_Provider	Text	10		Claim Provider			
Invoice	Invoice_ID	Integer	10	N	Invoice Unique ID	Y	1	
Invoice	Invoice_date	Text	10		Invoice billed date			
Medicine	Medicine Code	Integer	10	N	Medicine Unique Code	y	1	
Medicine	Medicine Price	Integer	10		Medicine Cost			
Medicine	Medicine Type	Text	20		Type of medicine			
Medicine	Description	Text	20		Description			
Medicine	Expiry Date	Date	10		Expiry Date of medicine			

The Following table signifies the Fact table and its respective Data measures,

Table Name	Column Name	Data Type	Length	Null values	Column Description	PK	PK order	FK
Patient Record Fact Table	Total Cost	Integer	10		Total Cost Details of the Patient			
Patient Record Fact Table	Patient_ID	Integer	10	N	Unique Patient ID	Y	1	Y
Patient Record Fact Table	Doctor_ID	Integer	10	N	Unique Doctor ID	Y	2	Y
Patient Record Fact Table	Procedure_ID	Integer	10	N	Unique Procedure ID	Y	3	Y
Patient Record Fact Table	Date_ID	Integer	10	N	Unique Date_ID	Y	4	Y
Patient Record Fact Table	Location_ID	Integer	16	N	Unique Location ID	Y	5	Y
Patient_Record Fact Table	Branch_ID	Integer	16	N	Unique Branch ID	Y	6	Y
Patient_Record Fact Table	Claim_ID	Integer	10	N	Unique Claim ID	Y	7	Y
Patient_Record Fact Table	Equipment_ID	Integer	10	N	Unique Equipment ID	Y	8	Y
Patient_Record Fact Table	Invoice_ID	Integer	12	N	Unique Invoice ID	Y	9	Y

Patient_Record Fact Table	Medicine_code	Integer	16	N	Unique Medicine Code	Y	10	Y
Patient_Record Fact Table	Diagnose_ID	Integer	16	N	Unique Diagnose ID	Y	11	Y
Patient_Record Fact Table	Age	Integer	2	N	Age of the patient			
Patient_Record Fact Table	Procedure_Date	Integer	10	N	Date of the procedure			
Patient_Record Fact Table	Insurance_ID	Integer	10	N	Unique ID for insurance			
Patient_Record Fact Table	Zip code	Integer	10	N	Zip code details.			

QUERY

Some of the sample information required by the decision making personnel's can be derived from Queries relating fact and dimension tables as shown,

1. To get Patient information and their respective Treatment Details

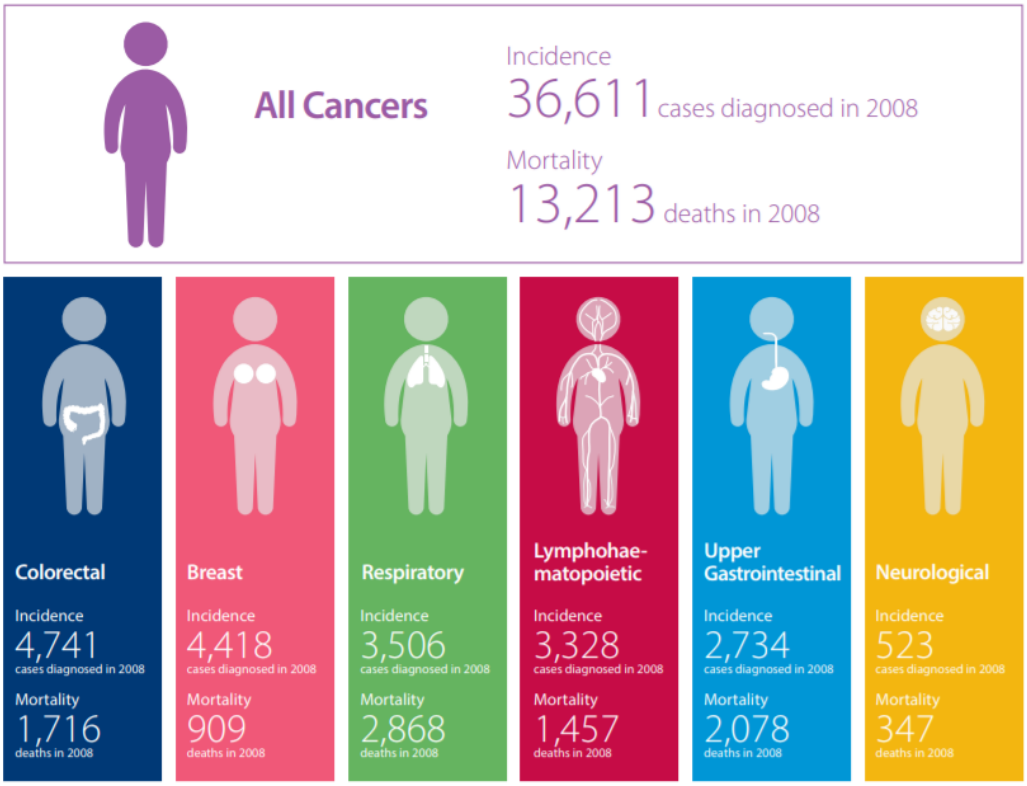
```
Select distinct pat_rec.Patient_ID, Patient_Name,Doctor_Name, pat_rec.Doctor_ID,Doctor_Type,
Patient_Age,
Speciality, pat_rec.Date_ID, Discharge_date, Admit_date, Insurance_no,pat_rec.Claim_ID,
Claim_provider, pat_rec.Invoice_ID, Invoice_Date
from Patient_Record as pat_rec
join Patient_Dim on pat_rec.Patient_ID = Patient_Dim.Patient_ID
join Doctor_Dim on pat_rec.Doctor_ID = Doctor_Dim.Doctor_ID
join Date_Dim on pat_rec.Date_ID = Date_Dim.Date_ID
join Claim_Dim on pat_rec.Claim_ID = Claim_Dim.Claim_ID
join Invoice_Dim on pat_rec.Invoice_ID = Invoice_Dim.Invoice_ID
Where
pat_rec.Patient_ID = &pat_id
order by
Admit_date desc
```

2.The following query helps in getting information of patients based on type of cancer diagnosed:

```
Select pat_rec.Diagnose_ID,Diagnose_type,pat_rec.Procedure_ID, Group_Code,
pat_rec.Doctor_ID,Doctor_Type,
Speciality, Procedure_Description, Diagnose_Description,pat_rec.Medicine_ID, Medicine_Type
```

from Patient_Record as pat_rec
join Doctor_Dim on pat_rec.Doctor_ID = Doctor_Dim.Doctor_ID
join Date_Dim on pat_rec.Date_ID = Date_Dim.Date_ID
join Procedure_Dim on pat_rec.Procedure_ID = Procedure_Dim.Date_ID
join Diagnose_Dim on pat_rec.Diagnose_ID = Diagnose_Dim.Claim_ID
join Medicine_Dim on pat_rec.Medicine_ID = Medicine_Dim.Medicine_ID
Where
Diagnose_type in (&diag_typ)

Thus, from the type of output obtained by the above query the decision makers or data analyst can compute visualisations as shown below,



FUTURISTIC APPROACH OF CANCER AID - DATA WAREHOUSE

Cancer have many sub- phases that can range from threat for life to fatal one. Of all most of them proved to be fatal. It would be great if someone identifies the cause that cause cancer and predict it in advance and help saving from them. The suggested Datawarehouse can help in predicting patterns and also can be used for complex data analysis that depicts the indications of cancer. Medical physicians can understand and analyse the patients case using various data analytics techniques.
The data warehouse possesses historical data of multiple patients which can help in efficient data mining techniques. The research team might take a look at these data and help in predicting the early cause of cancer and also help in preventing them.

There are several Health models that has been suggested by experts which will help in determining cancer cells growth using the data warehousing models. several modelling techniques like decision trees, naive bays, neural network clustering etc can help in providing right treatment of data at minimum cost and also improves the data visualisation and interpretation. The main advantage of using these techniques is that it helps in training nurses and other medical sophomores in determining and examining the patterns of cancer. Therefore, it acts as a great decision support tool in making decisions and second opinions.
One negative thing about these prediction systems is that it requires large amount of data to make the most appropriate decision. Also even by taking large amount of data to examine, it should surpass through several techniques to make a proper assumption and expected output.

CONCLUSION

The Proposed warehouse can be utilised by the CEO and chief doctors of cancer Aid for predicting the patterns of various fields like cancer treatment facilities and predominant branch locations, that can be generated with the help of right queries and parameters. These information can also help in figuring out the patient health state, as in which zone the patient is currently living in, does treatments helps in the betterment of the patient, the cost factors etc... from the previous data in the history. Owing to these predictions, several wrong choices can be avoided, as cost factor is one among them. This design helps in extracting the hidden information from the data warehouse which could be used to make strategic analytic decisions which promotes the value of the organisation. Moreover, with advancements in the data mining techniques like kNN algorithm, clustering etc. and the history of having large volumes of data the deadly disease can be predicted at the initial phase and patients can be made aware of those fatal factors.

REFERENCES

Sahoo P., Mohapatra S., & Shih-lin W. 2016, 'Analyzing Healthcare Big Data with Prediction for Future Health Condition', *IEEE Access,* vol. 4, pp. 9786-9799

Boon Keong S. 2013, 'An application of a Healthcare Data Warehouse System', *IEEE, vol. 1, no. 13, pp.269-273*

Rod Chester, 2017, *Apple highlights success of CancerAid, in its App Store,* Australia, viewed 19 sept 2018, < https://www.news.com.au/technology/gadgets/mobile-phones/apple-highlights-success-of-canceraid-in-its-app-store/news-story/028fa6f632af9395b49cd180cf34ec6e>