_____

IDART DATA MART

By

ZUKILE RORO

A report submitted in partial fulfillment of the
requirements for the degree of BSc Honours
(Computer Science)

University of the Western Cape

2010

_____

University of the Western Cape
Department of Computer Science
Supervisor: Dr William D. Tucker

ABSTRACT

IDART DATA MART

by Zukile Roro

Supervisor: Dr William D. Tucker
                Department of Computer Science

Any online transaction processing (OLTP) data contains information that can help in making informed decisions about businesses. For example, you can calculate your net profits for last quarter and compare them with the same quarter of the previous year. The process of analyzing your data for that type of information, and the data that results, are collectively called *business intelligence*. Because most operational databases are designed to store your data, not to help you analyze it, it's expensive and time consuming to extract business intelligence information from your database. The solution: an online analytical processing (OLAP) database, a specialized database designed to help you extract business intelligence information from your data. The purpose of this project is to combine data from multiple instances of iDart into a single data mart that can be used by Cell-Life for analysis and reporting. In my data mart design I use the star schema. The advantage of using this schema is that it reduces the number of tables in the database.

ACKNOWLEDGMENTS

First and foremost I would like to thank my family for their support, without them I wouldn't be where I am today. Then I wish to thank my supervisor Dr William D. Tucker for his kind supervision, advices and support.

Table of Contents

LIST OF FIGURES

LIST OF ACRONYMES

ARV–antiretroviral

iDart – Intelligent Dispensing of ART

IDE - integrated development environment is a software application that provides comprehensive facilities to computer programmers for software development.

Data mart - It is a simple form of a data warehouse that is focused on a single functional area.

ETL - Extract, transform, and load is a process in database usage.

GUI - Graphical User Interface

HIV – Human immunodeficiency virus

RA – Requirement Analysis

Star schema - is the simplest style of data warehouse schema.

Talend - is an open source data integration software vendor which produces several enterprise software products, including Talend Open Studio.
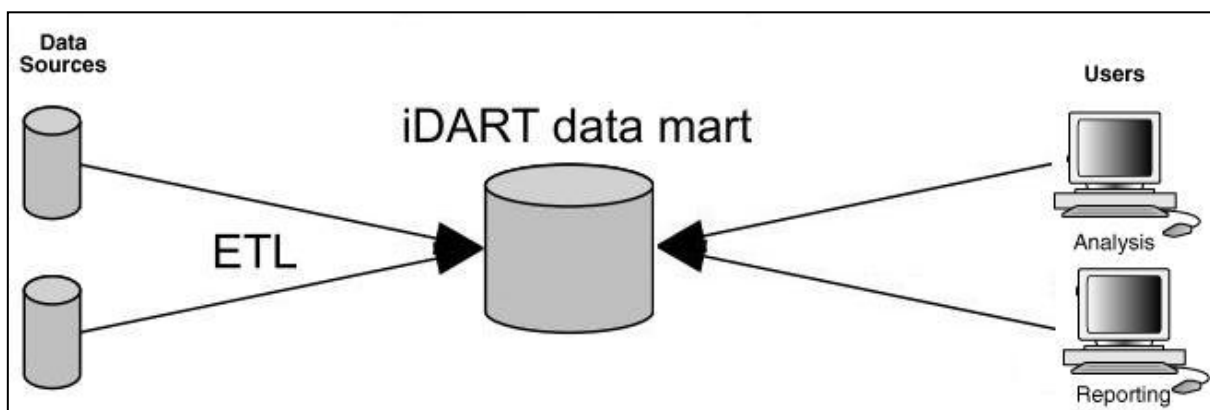
UR – User Requirements

INTRODUCTION

Concerning the data mart design, two commonly used schemas are the star and snowflake schema. In star schema the fact is denormalised, all dimension tables are normalise and will be primary foreignkey relationship between fact and dimension tables. For better performance we use star schema when compare to snow flake schema where fact table and dimension tables are normalised every dimension table there will be a look table meaning that we have to dig from top to bottom in the snowflake schema. The main advantages in star schema: 1) it supports drilling and drill down options, 2) fewer tables, and 3) less database.

In response to a request from the Desmond Tutu HIV Foundation to assist the management of ARV dispensing, the Intelligent Dispensing of ARVs (iDART) system was developed by Cell-life which in 2009 is in over 20 clinics dispensing drugs to more than 45,000 patients. This system is used by pharmacists to manage the supply of ARV stocks, print reports and manage collection of drugs by patients.

One of many iDart sites is the ARV pharmacy at the Tsepong Wellness Centre which became the third Elton Aids Fundation sponsored health care facility to receive the iDart system. The Tsepong Wellness Centre is currently servicing over 6000 HIV+ patients.

What is a data mart?: It is a simple form of a data warehouse that is focused on a single functional area. Data marts represent the retail level of the data warehouse, where data is accessed directly by end users.

The goal of this project is to combine data from multiple instances of iDart into a single data mart that can be used for reporting and analysis by Cell-life.

Figure 1: iDart data mart concept.

USER REQUIREMENTS (UR)

Overview

This chapter contains the user requirements of iDart data mart. These requirements have been derived from Cell-life's project specification. This chapter is intended to guide development of iDart data mart. This also will give overview of the project, including why it was conceived, what it will do when complete, and the types of people we expect will use it.

User's view of the problem

The time and expense involved in retrieving answers from databases means that a lot of business intelligence information often goes unused. Some organizations use a dozen different software packages to produce simple reports. Also, if the report doesn't have the proper information, its creators have to start over. The iDart data mart will help minimize cost of extracting business intelligence information from iDart instances around the country.

What is expected from a software solution?

Cell-Life expects a software solution that can be used for analysis and reporting purposes.

Cell-life would like to be able to generate the following statistics on a monthly/annual basis.

- Number of patients treated(based on packages created)
- Number of patients enroll on treatment
- Number of patients terminating treatment(including reason for termination)

by date, site, gender and age groups.

What is not expected from a software solution?

The software solution is not expected to be deployed to all the Cell-Life branches and it is not expected to be able to function in times of power failure unless a backup power supply is in place.

Also the software solution is not expected to be used by multiple business units except what it's designed for.

General Constraints

We will work under a few number of constraints such as development environment which in thi case has to be the integrated development environment (IDE). Also the database we'll have to use is Postgresql, to make sure that our product (iDart data mart) is compatible with existing database which is currently in use.

REQUIREMENT ANALYSIS (RA)

Overview

Requirements analysis is critical to the success of a development project. [2] Requirements must be documented, actionable, measurable, testable, related to identified business needs, and defined to a level of detail sufficient for system design. Requirements can be functional and non-functional.

Designer's interpretation of the user's requirements

Cell-Life has clearly expressed the requirements for the iDart data mart in the previous chapter (Chapter 1). Now we will focus on the busuness and technical requirements needed to implement the given user requirements. Existing solutions will also be considered.

A basic desktop computer running Windows/Linux will work prefect and a PostgreSQL Database Management System with Java. For data integration, ETL (Extract, Transform, Load) tool (Talend) will be used. Pentaho server and Pentaho Dashboard Designer will form part of the system.

The basic building block I'll use in data mart design is the star schema. A star schema consist of one large central table called *fact table,* and a number of smaller tables called *dimension tables* which radiate out from the *fact table*.

After classifying data from the requirements in Chapter 1, I have the following:

- Date, location/site and patient are dimensions
- Number of patient treated, enrolled for treatment, terminating treatment are facts.

Suggested solution

The suggested solution will make use of a desktop personal computer (PC) running Windows/Linux and can be broken down into various parts. The first stage uses ETL (Extract, Transform, Load) tool Talend Open Studio to retrieve data from stand alone iDart databases to the iDart data mart. Second stage is accessing data in the data mart, analyzing it, creating reports, graphs, and charts using a Pentaho dashboard.
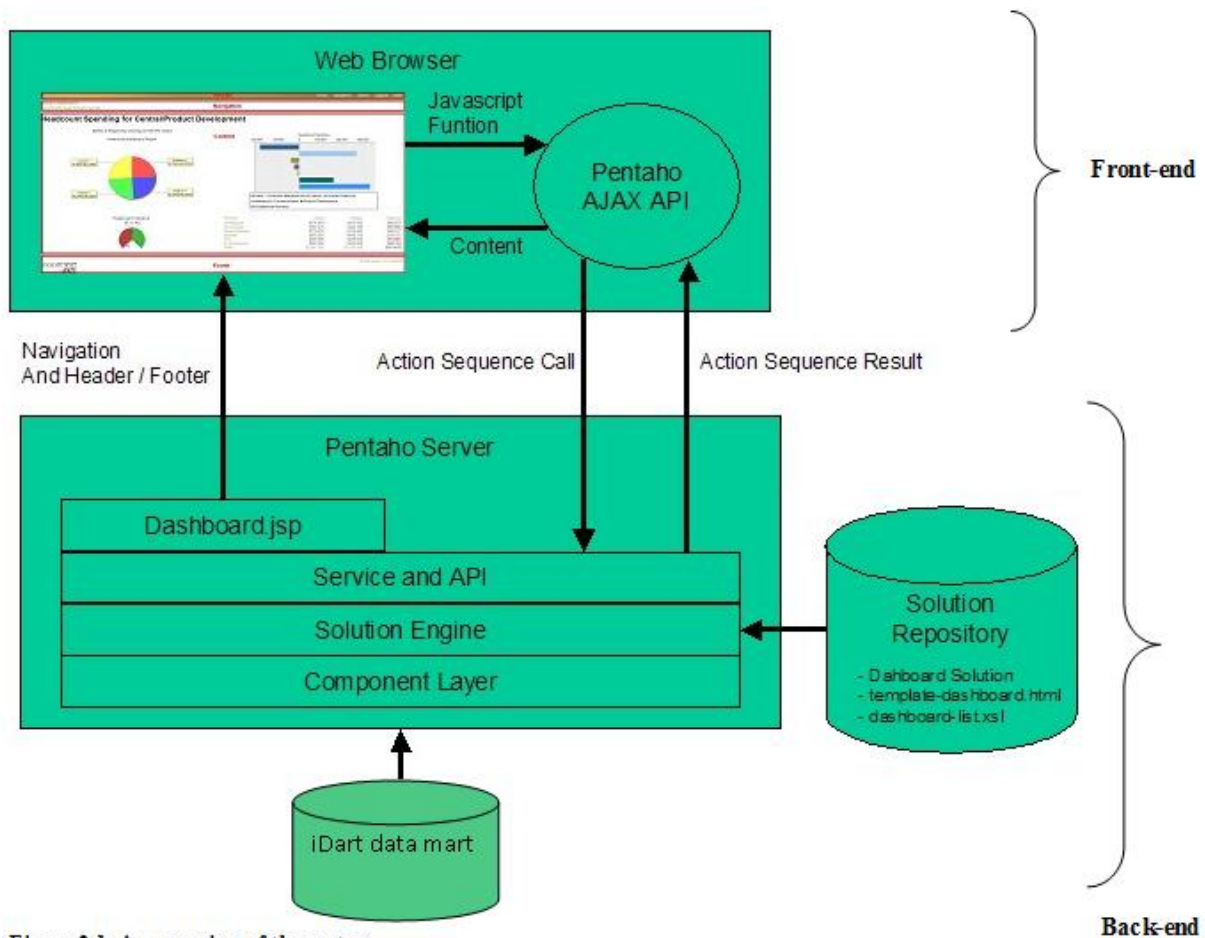
Figure 2.1: An overview of the system.

Testing the suggested solution

There are many different approaches to test software. For this project, functional and usability testing will be performed.

1. Functional Testing:

This is a new system and critical, so I must ensure its functional quality. All the features will be tested to ensure all functions provide the expected output.

2. Usability Testing:

Usability testing of this system will evaluate the potential for errors and difficulties involved in using the system for Cell-Life related activities.

Conclusion

This document is intended to guide development of iDart data mart. It also will give overview of the project, including why it was conceived, what it will do when complete. Screenshots showing how the final product will look like and behave will be provided in our next chapters for the second term.

User requirements were gathered and analyzed, then documented. From this documentation will follow the user interface specification (UIS), high level design (HLD) and low level design (LLD).

Bibliography

[1] BATIN, C., SERI, S., AND NAVATE, S.B, (1994) *Conceptual Database Design: An Entity Relational Approach*, Redwood City, California

[2] Executive editors: Alain Abran, James W. Moore; editors Pierre Bourque, Robert Dupuis, ed (March 2005).

[3] InfoManagement Direct, November 1999. *Data Mart Does Not Equal Data Warehouse* [online]. Available http://www.solent.ac.uk/Library/informe/law/referencing/solution2how24.stm [accessed 7 March 2010]

[4] KIMBALL, R.,(1996): *The Data Warehouse Toolkit*, New York: J. Wiley & Sons.
[5] KIMBALL, R.,(1997): *DBMS Online*,A Dimensional Manifesto August, 1997.