CITS3401 Data Exploration and Mining Project 1

Medicare Australia Data Warehouse

Mitchell Pomery
21130887

April 28, 2014

Abstract

This document outlines the design choicies for a data cude created to assist Medicare, an Australian Healthcare Supporter, in providing the best service possible. The original requirements were incomplete and so assumptions were made where necessary.

Introduction

Medicare Australia wishes to use data from it's previous years to assist in making decisions to improve their services, analyse expenditure and detect individuals who are abusing their system. Each centre stores information about visits in an Online Transaction Prorocessing (OLTP) database, these are then collated at a state and country wide level. The patient, doctor, treatment and prescriptions for each visit are stored. This document outlines the data cube designed facilitate in the decision making processes of Medicare.

Requirements

The authors' interpretation of the requirements are listed below.

Object	Restrictions
Location	State or Territory in Australia
Centre	3 Centres in each State/Territory
Patient	
Tests	Only one test will occur per visit.
Diseases	Only one disease will be diagnosed per visit maximum.
Referrals	Occur when a disease has been diagnosed.
Date	2006 to 2011, broken down into quarters.

Table 1: Requirements

Assumptions

Assumptions were made where the requirements were incomplete or insufficient, to simplify the schema and keep it managable, and to make the scenario as realistic as possible.

1. Only a small number of patients, diseases, physicians, hospitals, specialists and pathology clinics exist.

2. Doctors are irellevant, only the name of the clinic matters.

3. Patients will always visit a General Physician before seeing a specialist.

4. The cost of treatment, as well as the person or company who pays for the treatment is irrelevant.

5. People only visit medical centres in their own state.

6. All data is complete and easily available in the desired format.

Warehouse Schema

A star schema was designed to make the data cube simpler, and the queries faster than a snowflake schema or fact constellation. The date of the visit was broken into two dimensions, Year and Quarter, to allow for comparison between different years. This allows us to see which diseases reoccur at high rates each year, and when they occur.

The requirements also state that Medicare is interested in using this data for analyzing several areas of their bussiness. They would like to be able to determine which patients frequently return for diagnosis and perscriptions and which doctors frequently refer patients to the same specialists. The discovery of trends in diseases and referrals, and the discovery of outliers in patients, medical centres and treatment times would be benificial for Medicare.
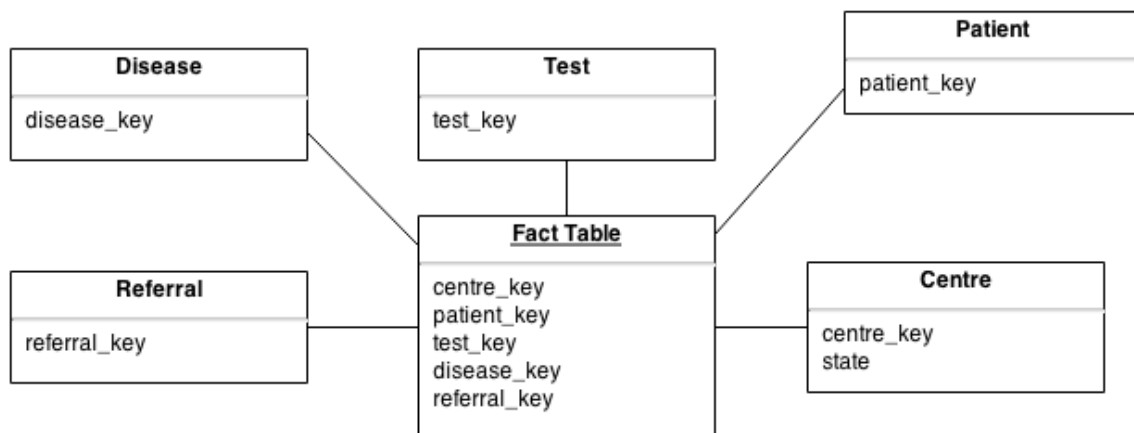

Figure 1: Fact Table in Star Schema

Prototype Warehouse

Data Generation

Prototype data was generated using the Python script `gendata.py`. It takes sets of words for diseases, clinics, names, medical tests and states, creates random people and outputs information about their visits from 2006 to 2011. An attempt has been made to make it reflect reality, by restricting the average persons visits to a couple of times a year, charging based on the location visited and making a small percentage of users abuse the system each year.

Data Analysis

`Palo.xlsx` was created to import the data into the OLAP database for analysis, however due to issues with PALO, it was soon discarded. `search.py` was written to do data analysis on the data generated by `gendata.py`, and the output was manually entered into `Python.xlsx` .

`search.py` enumerates all the possible elements for each dimension into several arrays, then counts the number of transactions that satisfy certain conditions, such as state, and places it into a table. Medicare Australia is then able to view the output data and determine what action, if any, needs to occur. By looking at the visits per patient per year, Medicare will be able to determine if any individuals are using their services abnormally, they will then be able to investigate if the visits are legitimate.
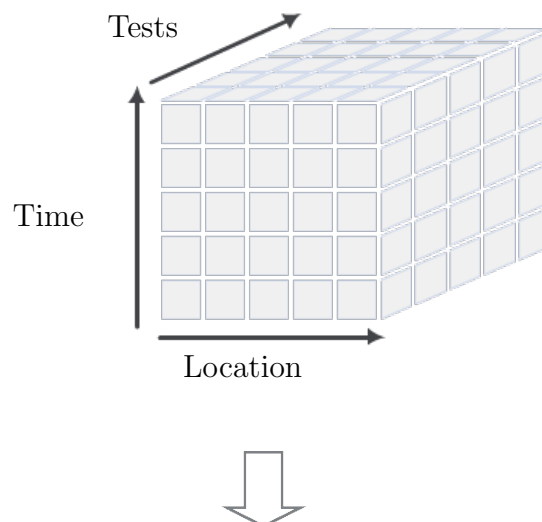
Features

Adds the ability to do...

1. expenditure analysis
2. planning new infrastructure
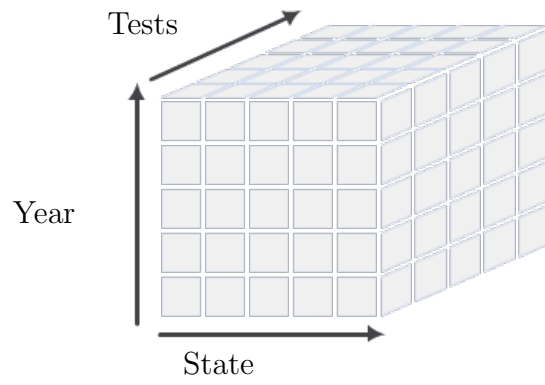3. detecting fraud
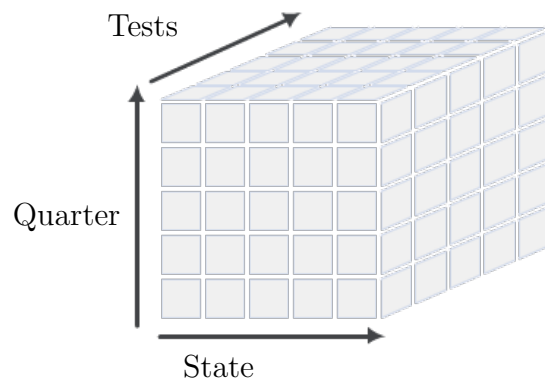4. policy changes

Expenditure Analysis

aa

Data Cube
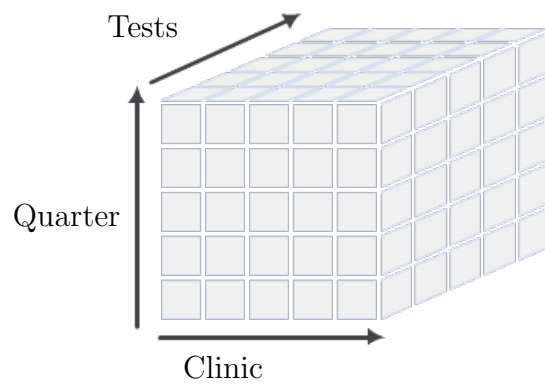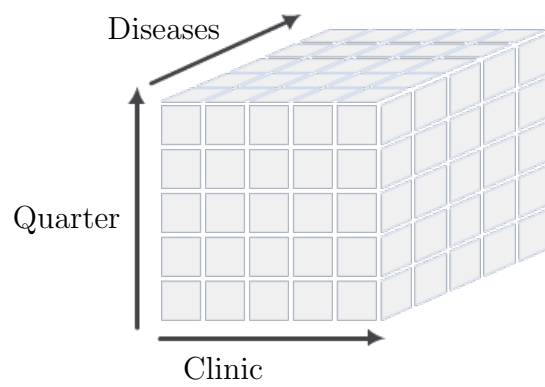
A visualization of the data cube is provided below:

Tests

Year

State

Drill Down On Year

Tests

Quarter

State

Drill Down On State

Tests

Quarter

Clinic

Drill Down On Tests

Diseases

Quarter

Clinic

Drill Down On Diseases