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 in giving 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	**Properties**	**Restrictions**
Location		State or Territory in Australia
Centre		3 Centres in each State/Territory
Tests		Only one test will occur per visit.
Diseases		Only one disease will be diagnosed per visit maximum.
Store	Interior Design Facility Type	3 restaurants in each country 3 different interior designs Facilities are 'dine in', 'drive through' and 'both'

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.

Warehouse Schema

A star schema was designed to make the data cube simpler, and the queries faster than a snowflake schema or fact constellation.

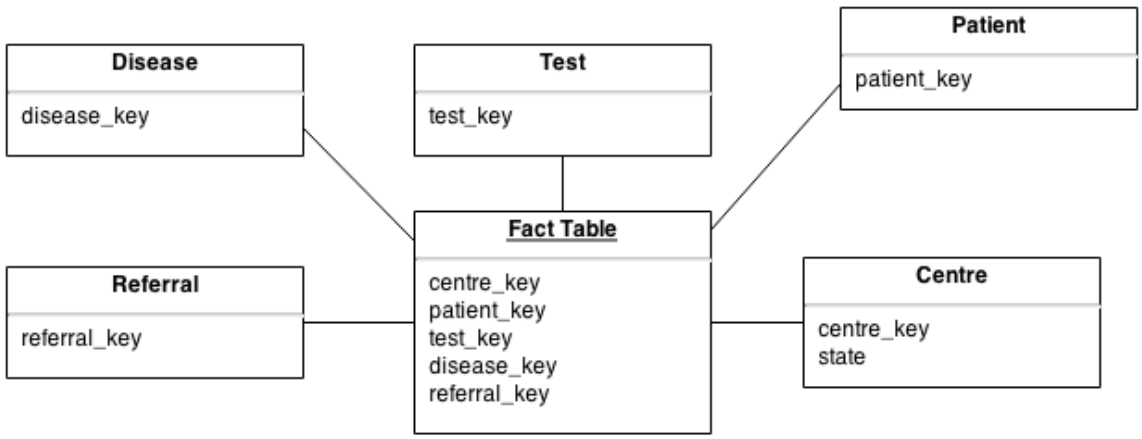


Figure 1: Fact Table in Snowflake Schema

Prototype Warehouse

Data Generation

Prototyape 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.

Features

Adds the ability to do...

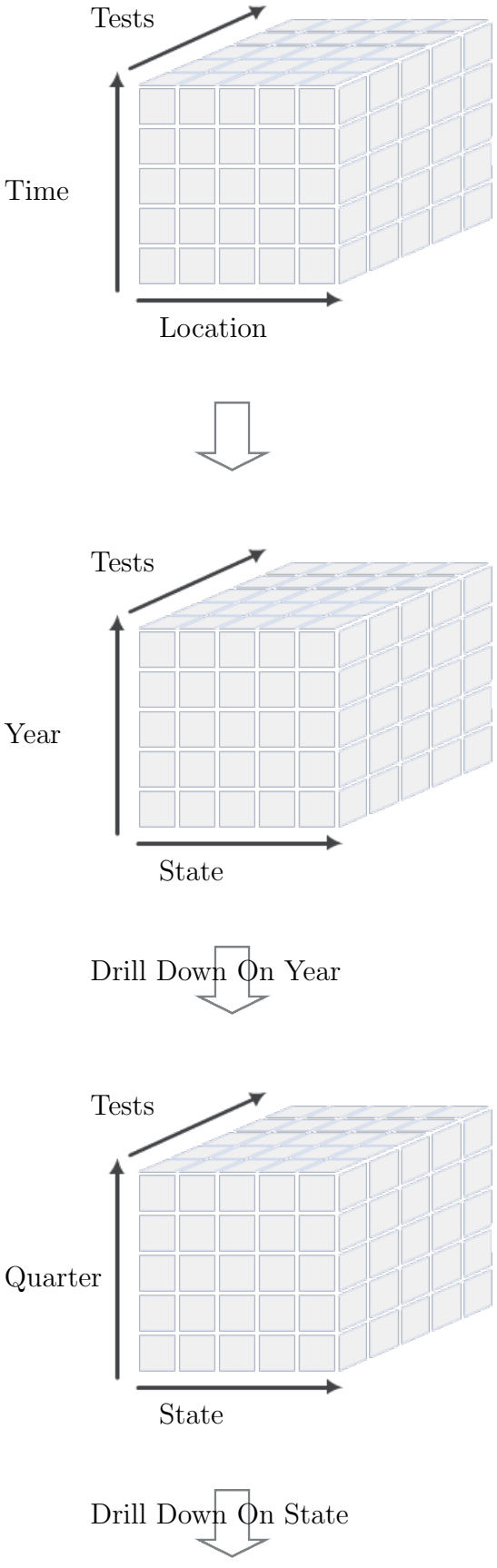
1. expenditure analysis
2. planning new infrastructure
3. detecting fraud
4. policy changes

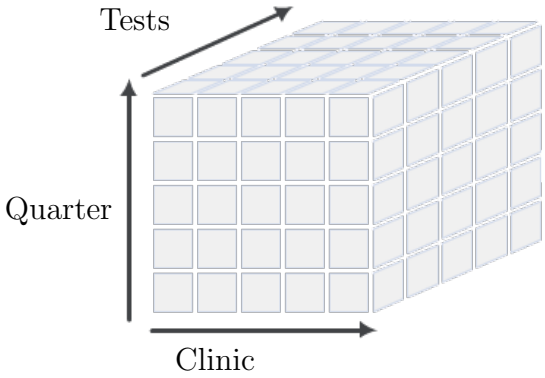
Expenditure Analysis

aa

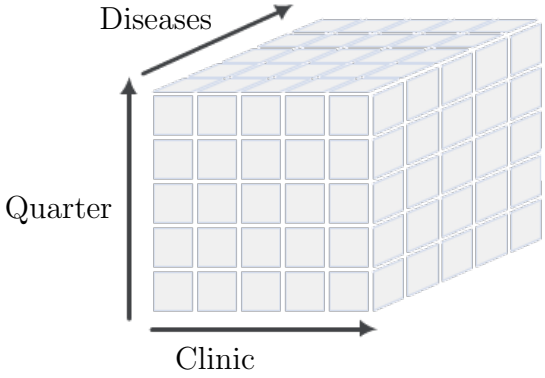
Data Cube

A visualization of the data cube is provided below:





Drill Down On Tests



Drill Down On Diseases