Clinical Data Warehousing
A Business Analytics approach for managing health data

Lekha Narra, Tony Sahama and Peta Stapleton
Science and Engineering Faculty, Queensland University of Technology,
Brisbane, Australia

Introduction

- Changing lifestyle trends are contributing to an increase in chronic diseases as well as healthcare costs
- Effective prevention and management of chronic diseases is only possible through quality information
- Disparate health data silos and mounting worldwide health data pose a big challenge
- BI & A approach enables analysis of enormous datasets to gain invaluable insights which can improve the understanding of various health issues

Research Scenario

- BI & A approach to study the effect of lifestyle on obesity
- The overall cost of obesity to Australian society and government was estimated as $58.2 billion in 2008 by Australian Bureau of Statistics (ABS)
- To enable effective and efficient analysis of obesity in the Australian context with a focus on levels of physical activity, nutrition, sex and age
- Questions of interest:
  1. What is the relationship of weight to age and sex?
  2. What is the relationship of weight to fruit and vegetable intake?
  3. What is the relationship of weight to type of milk consumed?
  4. What is the relationship of weight to amount of physical activity for fitness, recreation or sport?
  5. What is the relationship of weight to number of days exercised for fitness, recreation or sport?

Approach

- Challenges with conventional tools and techniques with respect to capacity and performance
- BI & A relies on a data warehousing approach where extraction, transformation and loading (ETL) is utilised for conversion and integration of enterprise-wide data
- Why data warehouse?
  - ➢ Enables effective integration of disparate health data as well as efficient multi-dimensional analysis
- What is a data warehouse?
  - ➢ A data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data that supports managerial decision making (Inmon 2005)

Data Sources

- Published Australian Bureau of Statistics (ABS) data in 2011 based on 2007-08 National Health Survey (NHS) and the 1995 NHS
- The source data is in summary form presented in Excel spread sheets. Access to raw data is restricted.
- Hence data aggregated from the source files is used as raw data
- Table 1 presents the summary of source data in terms of the number of levels for each factor as well as the nature of data for each level

Factors	Levels of the Factors & Attributes			
	No. of levels	Sub-Levels	Nature of options	
			Check compliance with standards	Selections among pre-defined criteria
Nutrition	4	1	✓	
		2	✓	
		3	✓	
		4		✓
Physical Activity	3	1	✓	
		2		✓
		3		✓

Table 1: Summary of Data Sources

Tools and Techniques

- Centralised data warehouse architecture is implemented which provides consistent, integrated and flexible source of data (Moody and Kortink 2000)
- Star schema as shown in Figure 1 is utilised comprising of a fact table and multiple dimensions with no direct connection between the dimensions. This offer faster query execution times due to less number of joins.
- Extract, Transform and Load methodology is used to load data into data warehouse
- A single cube was built using all of the dimensions in the data warehouse. This is to allow for the analysis of weight factor with respect to each category of nutrition, physical activity, age and sex as well as location, time and individual

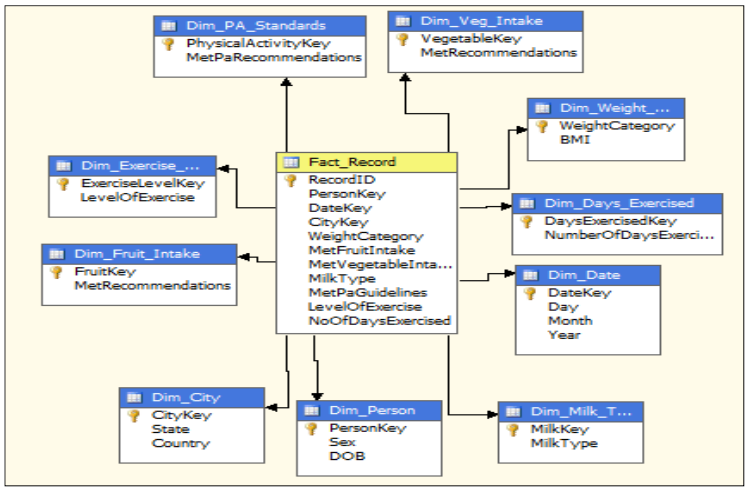

Figure 1: Schema Diagram of Obesity data warehouse

Discussion and Conclusion

- Ability to perform dimensional analysis of person's weight with respect to various lifestyle factors
- High capacity and performance which is crucial with increasing volume of health care data
- This approach enables analysis of large amounts of historical data there by allowing identification of relationships between various aspects, for example the relationship between lifestyle with various types of cancers.

Queensland University of Technology
Brisbane Australia