Tutorial – Week 10

Data Warehouse

# Objectives:

1. Understand the fundamentals of dimensional modelling – 20 mins
2. Design a dimensional model using Kimball’s four-step design process – 25 mins
3. Discuss the impact of grain on fact tables – 10 mins

# Key Concepts:

* + Data warehouse
  + Business events
  + Dimensions, dimension tables and hierarchies
  + Facts, fact tables and granularity
  + Dimensional modelling – the star schema

# Exercise:

## Designing a dimensional model

Wimmera Wines is a large company that takes deliveries of grapes from wine growers, produces and bottles wine, and sells those bottles to retailers and restaurants. They produce many different types of wine at a range of price points, from cheap cask wine to top-of-the-range vintage bottles.

Wimmera Wines’ day-to-day OLTP database uses the following ER model (see the Figure 1 on the next page).

The company is aiming to increase their product sales by 20% in comparison to the last 3 years. To help the business achieve their aim, you have been hired to design a data warehouse that can help business managers analyse data related to the sales theme.

该公司的目标是在过去3年的基础上增加20%的产品销售。为了帮助企业实现目标，公司聘请您设计一个数据仓库，帮助企业经理分析与销售主题相关的数据。

The company is keen to understand all the aspects of their business that contribute to strong sales. For example, two business measures that have been mentioned are “total number of units of each product sold” and “revenue generated by each employee per year”.

该公司渴望了解他们业务的所有方面，有助于强劲的销售。例如，前面提到的两个业务度量是“销售的每种产品的总单位数”和“每个员工每年产生的收入”。

Tasks.

* 1. As a class, brainstorm some more business measures that Wimmera Wines managers might need if they are to achieve their aim.
  2. Use Kimball’s four-step dimensional design process to design a dimensional model for

Wimmera Wines’ product sales subject area.

* + 1. Select and explain the business process.
    2. Declare the grain and justify your choice.
    3. Identify and explain the dimensions.
    4. Identify and explain the facts.



**Figure 1 - Wimmera Wines’ day-to-day OLTP database ER model**

## Fact tables in practice

Consider the following fact table:



Sale

Time key Geography key Product key Dollar sales Unit sales

Suppose the following sales data has been extracted from the business’s operational database:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SaleID** | **SaleDate** | **CustomerID** | **CustomerCity** | **ProductID** | **Price** | **Quantity** |
| 54 | 2003-12-13 14:13 | 788 | Melbourne | 9644 | $10.00 | 2 |
| 54 | 2003-12-13 14:13 | 788 | Melbourne | 8574 | $15.00 | 1 |
| 67 | 2003-12-13 15:05 | 903 | Melbourne | 9644 | $10.00 | 1 |
| 76 | 2003-12-13 17:26 | 322 | Sydney | 9644 | $5.00 | 4 |
| 77 | 2003-12-14 09:58 | 292 | Melbourne | 8229 | $15.00 | 2 |

* 1. Starting from this source data, how many rows will be inserted into the fact table if an hourly grain is selected?
  2. How many rows will be inserted into the fact table if a daily grain is selected?
  3. At which level of granularity can we answer questions about hourly sales? At which level of granularity can we answer questions about daily sales?