INFO90002 Tutorial – Week 11

(Tutorial: Data warehousing)

Objectives:

This tutorial will cover:

I. Understand the fundamentals of dimensional modelling – 20 mins
II. Design a dimensional model using Kimball's four-step design process – 25 mins
III. 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

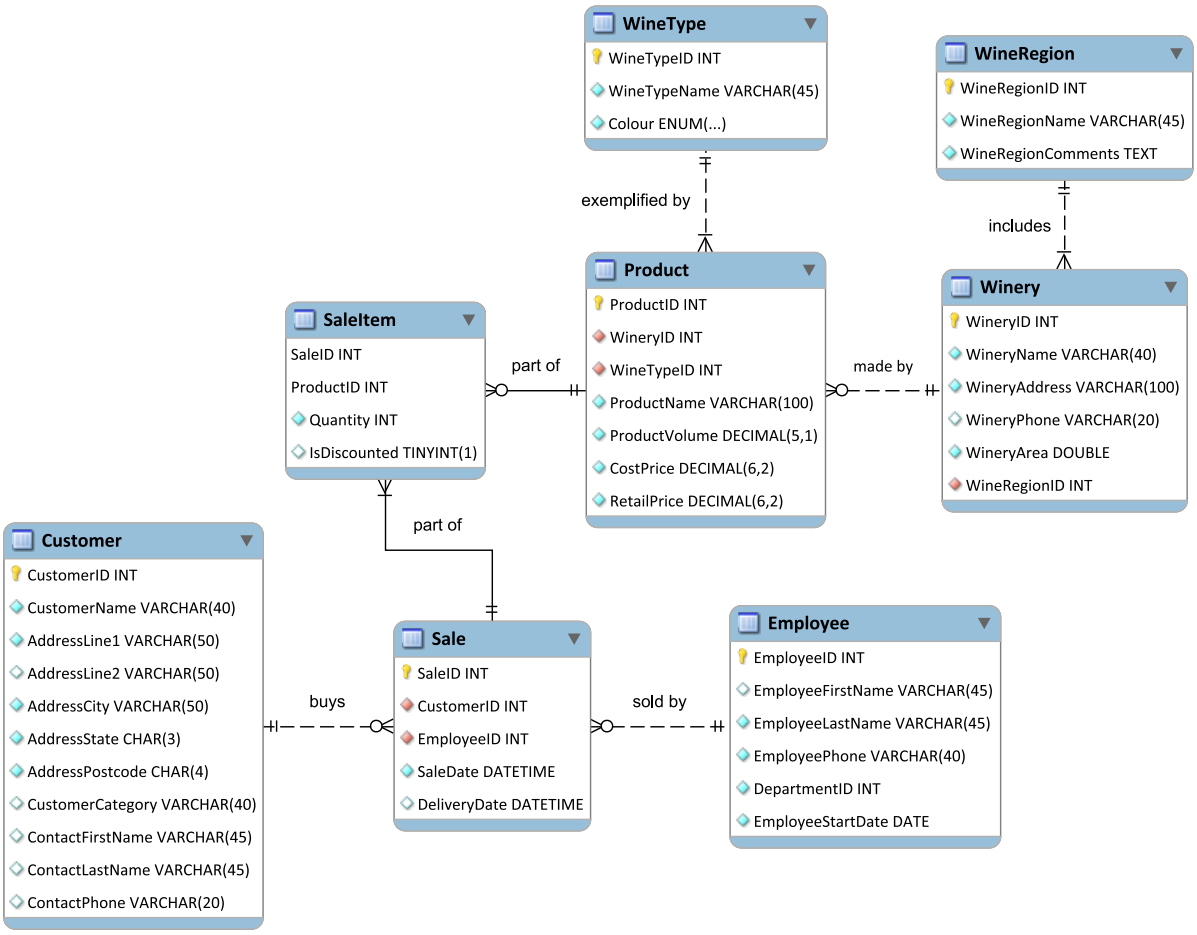
Exercise:

1. 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:

Continued over 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.

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".

a. As a class, brainstorm some more business measures that Wimmera Wines managers might need if they are to achieve their aim.
b. Use Kimball's four-step dimensional design process to design a dimensional model for Wimmera Wines' product sales subject area.

   i. Select and explain the business process.
   ii. Declare the grain and justify your choice.
   iii. Identify and explain the dimensions.
   iv. Identify and explain the facts.

Continued over page

2. Fact tables in practice

Consider the following fact table:

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

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