

# CSE5DWD – Data Warehouse Concepts and Design

Assignment, Semester 1 2020

## Dimensional Modelling Business Cases

### Objectives

- To create a Data Warehouse conceptual design using Star Schema Modelling
- To document all steps during the design process

### Weight of the assessment

- 30% of total assessment.

### 1 The Instructions

**Due Date: 10.00am Friday, 22th May 2020**

***This is an INDIVIDUAL assignment. You ARE NOT permitted to work in a group when writing this assignment.***

#### **Submission Guidelines:**

- No hard-copy of the assignment is required. Only electronic copy of your assignment is to be submitted to CSE5DWD subject LMS web site.
- Please do not forget to attach a Statement of Authorship to your assignment, sign and date it.
- You can zip all files in one for the submission since the system does not accept multiple files.
- Assignment should be typed, not written/drawn by hand. Use the Word Drawing toolbar (or another tool of your choice) to draw figures in your assignment.

#### **Copying, Plagiarism:**

Plagiarism is the submission of somebody else's work in a manner that gives the impression that the work is your own. The Department of Computer Science and Information Technology treats plagiarism very seriously. When it is detected, penalties are strictly imposed. Penalties are normally a mark of **ZERO** for the assignment.

Students are referred to the Department of Computer Science and Information Technology's Handbook and policy documents with regard to plagiarism and assignment return, and also to the section on 'Academic Misconduct' in the unit outline distributed in the first lecture and available on LMS/WebCT.

### **No extensions will be given:**

Penalties are applied to late assignments (5% of total assignment mark given is deducted per day, accepted up to 5 days after the due date only). If there are circumstances that prevent the assignment being submitted on time, an application for special consideration may be made. See the departmental Student Handbook for details. Note that delays caused by computer downtime cannot be accepted as a valid reason for a late submission without penalty. Students must plan their work to allow for both scheduled and unscheduled downtime.

### **Marking Scheme**

The marking scheme is described in **section 3: Your task**.

### **Queries Regarding Assignment**

Any queries can be addressed to the Lecturer during the normal consultation time. You will have consultation helping time in the week before this assignment due date:

8 consultation hours in the week. The specific time will be announced on LMS.

## **2 Business description**

### **Heathcote Winery Group**

**Heathcote Winery Group** is offering Regional Wine Export and Tourism of vineyard. It aims to help transform the Australian grape and wine industry by showcasing the nation's wine tourism offering and driving demand for Australia's wine exports.

The management of **Heathcote Winery Group** for wine export department has a very big market with hundreds of worldwide business patterners such as northeast Asia (China, Japan, Korea etc), southeast Asian (Indonesia, Vietnam, Thailand, Singapore, Malaysia etc), Oceania (New Zealand, Fiji, Samoa, etc), Europe (Italy, France, UK, Germany, Poland, Greece etc), North American (United States, Mexico, Canada, Cuba, etc), Middle east (Iran, Israel, Turkey, Saudi Arabia, Egypt, Iraq, etc) and rest of world. In recent two years, Australian wine exports to China increase 18% in value and 12% to Canada in volume, but 5% declined to USA in value. The management wants to analyse the data every quarter to see

- the trends of the market such as the average value of the exports to each country/area (\*)
- the consumptions of wines in bottles/ cartons/volumes by each person in any country per quarter
- regular wine drinkers' consumption patterns, and
- the most popular types (sparkling, red, white) of wines among the different countries (\*)
- the sales between different countries and the largest volume sales of the country
- the holiday period in Easter month has increased the sales in Europe comparing with the same period in last year?

The winery and vineyard are managed by the local owners who supply wines to export department of **Heathcote Winery Group**. On another hand, the local winery generates wine sales from its cellar door, market stalls and sales to eateries in its local area. The winery owners want to expand the distribution of

their wines to eateries further afield. They currently have data files containing where each bottle of wine was sold, the quantity in which it was sold and the dollar sales amount. They hope that by looking at this data they will be able to tell which wines were most popular at particular market stalls and use this as a marketing tool to sell those wines to eateries in the country wide.

Each bottle of wine is described by the following information: the name of the winery/trade mark/business or brand name; name of wine, a grape variety name, i.e. 'Shiraz' or a generic wine style, i.e. 'Dry Red'; a Geographical Indication (zone and/or region and/or sub-region), i.e. 'Yarra Valley'; vintage (year of release), volume (in ml); alcohol content (as a percentage); the number of standard drinks the bottle is equivalent to; an allergens declaration (i.e. egg, milk, nuts etc.); the name and address of the responsible entity; and the country of origin.

The winery also wants to increase its bottle sales at the cellar door. To encourage people to visit the winery, management plans on holding special event days where they provide wine sales discount, light musical entertainment, complimentary cheese platters, etc. Managers want to be able to assess the success of these special event days so information regarding the event: event date; event day name; event day description; music type; food type (cheese platters, luncheon, etc); event cost; and an estimate of the number of people who attended is required. Managers want to analysis the event sales:

- What wines' vintage are most popular during the event?
- **Compare the total sales of the Shiraz wine style in last quarter of 2018 with the same period of quarter in 2019 (\*)**
- **What wines are not sold during the event discount? (\*)**
- Do the Yarra Valley wines sell better than Murray Darling wines?

Another department of Heathcote winery group is the tours management. Local tour management offers daily tours providing an entertaining, sociable and educational day tour. The winery entry fee is for a fixed fee of \$10 for each visitor. The visitors in a group of 10 people or more will get 10% discount. To provide the convince to tourist the winery has facilitated restaurants and cafe for tourist to order for foods and drinks. The tour management wants to keep the expenses of each visitor including the entry fee, dishes, wine and drinks purchased during the day tour in the local winery. Please notes that each visitor may order a number of dishes in the restaurant for a lunch. The number of dishes many not be predicted for each visitor.

Managers want to analyse visitors' habits in a day tour. They want to

- compare the percentage of grouped visitors with the individual visitors
- **find out the most popular group of dishes are ordered by the visitors (\*)**
- **find the most popular wine glasses ordered by the visitor (\*)**
- understand the winery tourist number between weekday or weekend.
- understand how many percentage more of tourists form holiday days against and non-holidays.
- understand which age group of tourists is most likely drink the Shiraz wine.
- **Has the holiday period in Christmas month increased the tourists number comparing with the same period in last year? (\*)**

You need to design the data warehouse schema to capture the business processes for the data analysis.

### 3 Your Task: (100 marks total)

Your task is to design a data warehouse for **Heathcote Winery Group** using **multidimensional Modelling**. Your design needs to encompass the following steps:

- (i) First construct a **Data Warehouse Bus Matrix** to identify the company's business processes and **any likely Data Marts**. **(10 marks)**
- (ii) Identify the grain of each **star/snowflake schema** to design the dimensional model for business processes you have identified, ensuring your dimensions are conformed, primary and foreign keys are clearly labelled, and that your attributes are named using verbose textual descriptions. You also need to provide **three rows of the fact table** your designed to explain the meaning of the rows. **(40 marks)**
- (iii) Create the following table with a row for each fact table in your design, indicating the granularity of each fact and a brief justification for choosing that granularity. **(12 marks)**.

Fact table name	Fact granularity	Fact table type	Brief justification

- (iv) Create the following table with a row for each dimension table in your design, giving a brief justification for **choosing that dimension, and indicating any attribute hierarchies that exist within the dimension**. **(12 marks)**.

Dimension table name	Brief justification	Attribute hierarchies

- (v) Create the following table with a row for each design feature you have used, such as handling of possible null foreign keys, and the inclusion of any fact-less fact tables, degenerate dimensions, role playing dimensions, junk dimensions, outriggers, mini-dimensions, bridge dimens or any other design techniques discussed in the lectures. Provide a brief description of each design feature used (how and where it is used – *not the theory behind the concept*), any possible problems & solutions with a justification for their use **(16 marks)**.

Design feature	Brief description	Brief justification

- (vi) Identify which fields from your facts/dimensions are required to answer 7 business questions listed as the Bold words and marked with (\*) above. **(10 marks)**

#### Lecturer's Notes:

- It is not enough to just draw the shapes of the facts and dimensions in the data warehouse/ data marts. Each fact/dimension should be completely defined, with keys and facts/attributes clearly described;
- All information from the given business problem should be recognisable in your final data warehouse;

- Your final data warehouse should be able to **answer all business questions listed not only marked (\*) bold text questions**. If one question cannot be answered using your proposed solution, it means the solution is not completely correct and you have to rethink it.