ISYS90086 Data Warehousing – Summer Semester 2019

Assignment 2 – Data Warehouse Load Assignment (25 marks)

This assignment uses the same case study that you used in assignment one. We have provided you with the following extra details (from the LMS):

- Production System
    - The Product and ProductOrder Tables
- Sales System
    - The Customer Table (pre-processed)
    - The ProductPriceList Table
    - The Store Table
    - The Sale Table
    - The SaleItem Table
- A date dimension table that has been set up based on the data for the assignment

In groups of 2, you are required to carry out the ETL process to get the data from these source tables into the data warehouse that you designed. You will design 1 transformation per fact or dimension table in the data warehouse. You are then asked to write a report about the process (see requirements below).

Assessment

This assignment is worth 25% of your assessment for the subject. Historically, the average mark for this assignment has been around 17/25 marks.

Case Study – Report Requirements

You will write a 2000 word report outlining the main issues and concerns you came across in performing the data load. A suggested outline for the sections of the report is (Please note that this structure is a guide only and can be changed if you wish, word limits are suggestions):

1. **Executive summary:** An executive summary is a short summary of the key information in the document and is written for a business person. It should contain clear details in summary form of the issues you came across in designing the ETL process. (200 words)
2. **Design of the ETL Process:** You should discuss and justify the design of your ETL processes. Include diagrams of the design of your ETL transformations (for each fact and dimension table in your data warehouse) to use as evidence for the discussion. (1200 words)
3. **Design of the Data Warehouse**: If you had to do some redesign work on your data warehouse, detail what you changed and give reasons for the changes. (300 words)
4. **Data Dictionary**: Describe the **types of additional** meta-data that would be required in the Data Dictionary that you developed in Assignment 1, based on the ETL process. You should list and describe the **kinds of metadata** you need to store in a data dictionary in addition to what was required for assignment 1 – do not simply reproduce the assignment 1 data dictionary (300 words). NOTE: we are not asking for a complete data dictionary here – we are just asking for the TYPES OF ADDITIONAL Meta Data required (see the meta data lecture).
5. **Appendix 1 – Work Breakdown**: Detail the breakdown of work of the team members for this assignment. This should be a detailed account of what each team member accomplished as part of the assignment.

Assessment Criteria

Aspect of review	Poor	Good	Very Good	Excellent
Executive Summary 5%				
Design of the ETL Process 60%				
Dimensional Model Changes 15%				
Data Dictionary Additions 20%				
Work Breakdown (Hurdle)				

Submission

This assignment is due at **9:00am on Wednesday 6th Feb**

You will submit only an electronic version. We will endeavour to get your assignment marked within a 2 week period. You will receive feedback on your assignment via the marking scheme. The marker will also give written feedback in the form of comments on the marking sheet.

You are required to submit an electronic version of your work through the LMS (See the Assessment Link on the LMS and look for the link Assignment 2 Submission. You will be submitting this version of your work through TurnItIn (this is a plagiarism checking platform) and will be able to submit more than once. Also, by submitting the Assignment via the LMS you are agreeing that the Assignment is your own work (**thus you do not need an assignment cover sheet**). Please make sure the **Names** of each team member appear on the **Front Page** of the assignment.

You will also submit a single zip file (See Assignment 2 ZIP Submission link in the LMS). This file will contain all of the Pentaho transformations completed to load the data warehouse. When marking some or all of these transformations may be executed!

Academic Honesty

Academic misconduct occurs when students portray someone else's work as their own. There are many ways in which Academic misconduct can occur. Some of these are:

- **Sham Paraphrasing**: Material copied verbatim from text and source acknowledged in-line but represented as paraphrased.
- **Illicit Paraphrasing:** Material paraphrased from text without in-line acknowledgement of source.
- **Other Plagiarism**: Material copied from another student's assignment with the knowledge of the other student.
- **Verbatim Copying:** Material copied verbatim from text without in-line acknowledgement of the source.
- **Recycling:** Same assignment submitted more than once for different subjects.
- **Ghost Writing:** Assignment written by third party and represented as own work.
- **Purloining:** Assignment copied from another student's assignment or other person's papers without that persons knowledge.

The university is committed to graduating students with "a profound respect for truth, and for the ethics of scholarship... we want our graduates to be capable of independent thought, to be able to do their own work, and to know how to acknowledge the work of others" (Prof. Peter McPhee). As such, the university takes a dim view of students to are not able to correctly acknowledge the work of others, or who try to pass this work off as their own. All students should be aware of the following web site: http://academichonesty.unimelb.edu.au/advice.html which provides practical advice to students about how not to be involved in academic misconduct. In particular look at the Quick Checklist section (paraphrased below)

Quick checklist

(Reproduced with permission of Dr Stephen Morgan, Faculty of Economics and Commerce)

To be certain to acknowledge sources fairly and avoid plagiarising, review this checklist before beginning to write your essay and again after you have completed your first draft.

1. What type of source are you using: your own independent material, common knowledge, or someone else's independent material?
2. If you are quoting someone else's material, is the quotation exact? Have you used quotation marks for quotations run into the text? Have you set off block quotes with an extra space before and after the quote, single spacing within the quote, and left indenting of all lines of the block quote? Are omissions shown with ellipses and additions with square brackets?
3. If you are paraphrasing someone else's material, have you rewritten it in you own words and sentence structures? Does your paraphrase employ quotation marks when you resort to the author's exact language? Have you represented the author's meaning without distortion?
4. Have you acknowledged each use of someone else's material?
5. Do all references contain complete and accurate information on the sources you have cited