Project Portfolio

CSC-457 / CSC-557

Points: 100

Due: Monday, May 4th (before 10:00 PM)

1. The aim of the “project portfolio” is to create a document that 1) discusses the main stages in the design and implementation of a data warehouse and 2) describes the data warehouse you developed as part of this course.
2. See attached template for the content that should be included in your project portfolio.
3. Grading criteria
   * Completeness and organization of document
   * Quality of writing
   * Document formatting

Project Portfolio

Template

Note: Your portfolio document should include the following section headings. You are welcome to add sub-section headings.

# Data Warehouse Requirements

*In this section discuss the general process of gathering the user information requirements for a data warehouse. In addition, make sure to discuss 1) the target application domain of your data warehouse, 2) potential users of your data warehouse, and the information requirements for your data warehouse.*

# Data Warehouse Logical Design

*In this section discuss 1) key decisions that need to be made in designing a data warehouse (e.g. selecting the “business” process, declaring the granularity, identifying measures, and identifying dimensions), 2) types of measures (i.e. additive, semi-additive, and non-additive), 3) importance of using surrogate keys, and 4) process of creating a logical star scheme. In addition, make sure to incorporate your project design work and artifacts into the discussion.*

# Extraction Transformation and Load (ETL)

*In this section discuss 1) why the ETL process is challenging, 2) transformations that may need to be performed on extracted data, 3) general process of populating dimension tables, and 4) general process of populating a fact table. In addition, make sure to incorporate the Talend jobs created in Part-3 of your project (e.g. screen shots of your jobs along with descriptions).*

# Physical Data Warehouse Design

*In this section discuss 1) motivation for creating aggregate fact tables / generalized cuboids and 2) process for deciding what aggregate fact tables / generalized cuboids to create. In addition, describe the generalized cuboids you would create for your data warehouse and the reason why. (I recommend you look at your information requirements).*

# OLAP and Intelligence Applications

*In this section discuss the types of user applications that might be developed in the context of a data warehouse. In addition, make sure to include the Tableau “visual analytic” reports that you created in Part-4 of the project (e.g. screen shots of your reports along with descriptions).*

# 