Assignment 02

ETL Processing with SQL Code

In this “hands on” activity you practice performing ETL processing using SQL code and the simple "Flush an Fill technique. *This activity will take you about 3 to 9 hours, so plan accordingly!*

The activity consists of the follow steps:

1. Watch some videos.
2. Read a chapter from a text book.
3. Study web resources.
4. Create a new data warehouse design.
5. Submit your work to the Canvas site.

# Watch some videos

Please watch some videos I created for a full ETL course. You will notice that there are a lot of other videos in this course, but you only need to watch two of them.

1. [Module01 02 ETL Overview (external site)](https://youtu.be/m7pK9sp1HNY)
2. [Module02 02 An Overview of ETL scripts (external site)](https://youtu.be/8fKVaGanK3k)

# Read a Chapter from a Textbook

To help you understand how to use SQL for ETL processing, I have included a chapter from a book I wrote. Please read though this chapter and apply what you learn to the next part of this assignment. (Note, that performing the chapter exercises are not required!)

• Chapter 6: ETL Processing with SQL (The PDF for this is in the folder)

# Study web resource

Since our goal is to perform ETL processing on a data warehouse, you will need to be familiar with common Business Intelligence, Data Warehousing, and ETL terminology and designs. I have included a couple of articles that cover to help with that, so please review them. (I have also included these in a PDF format for offline reading.)

1. [kimballgroup.com-Design Tip 81 Fact Table Surrogate Key (external site)](https://www.kimballgroup.com/2006/07/design-tip-81-fact-table-surrogate-key/)
2. [kimballgroup.com-Six Key Decisions for ETL Architectures (external site)](https://www.kimballgroup.com/2009/10/six-key-decisions-for-etl-architectures/)

# Review and Update a Spreadsheet

I have provided you a spreadsheet, called “Instructors BISolutionWorksheets.xlsx” that outlines the current data warehouse design. Please review the design and make special note of the Transformations column. I only recorded some of the ETL transformations! You need to figure out what other transformations are and update the “Instructors BISolutionWorksheets.xlsx” to note these.

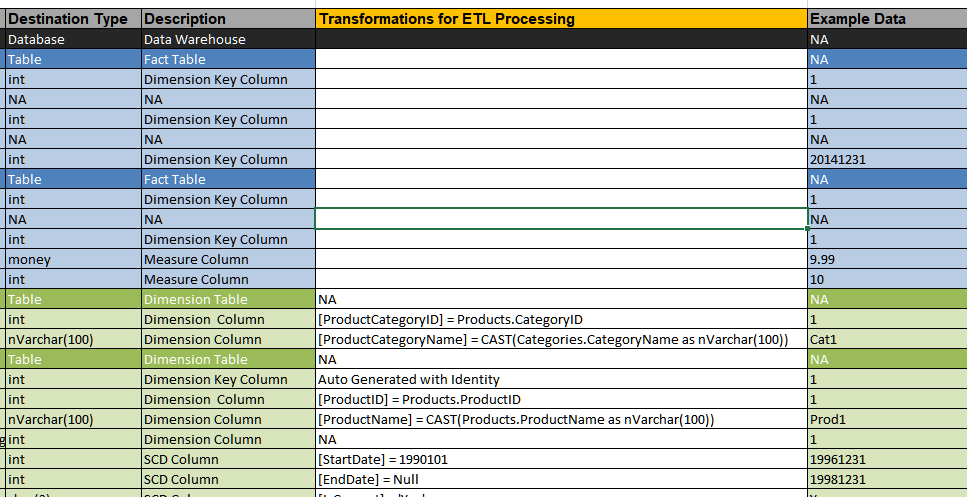


Figure 1: The Instructor's solution worksheet with the transformation column

# Create an ETL script

You need to create a SQL script that will move data from the source database to the destination database.

First, run the scripts ***01\_StarterScript Create the OLTP DB.sql*** and ***02\_SetupScript Create the DW DB.sql*** to the OLTP source and the OLAP destination databases. Then, create the ETL script using the starter file called “***03\_StarterETLScript.sql***.” Complete a SQL script to fill all the tables in the data warehouse database.

Please, **rename** the file **to** **<YourNameHere>ETLScript.sql**, add your own code ETL code to **flush and** **fill all tables** in the OLAP database, and make sure to **drop and recreate the foreign key constraints**.

**Important**:

* Place all your ETL code into views and stored procedures
* For you to get full points, your script must be able to run from beginning to end using the provided source and destination databases. So, test that your code runs as a complete script!

# Write a Paper

Create a Word document that outlines how your ETL Process works. This document should be written as a technical support document that new hire can read to understand what the process does, and technicians might use to troubleshoot the process. Include screen shots, please. The document should be about 2 to 4 pages (including any pictures) in length. Please use Microsoft Word or something compatible for your paper. I recommend that you start with the pictures, place them in your Word document, and then add a short paragraph to describe what you are seeing/doing.

**NOTES:**

1) Read this two page article on creating technical documentation before you start: <http://www.developer.com/tech/article.php/3848981/The-7-Rules-for-Writing-World-Class-Technical-Documentation.htm>

2) If you need help getting screenshots see this page ( <http://m.wikihow.com/Take-a-Screenshot-in-Microsoft-Windows>).

3) Make sure you format it like a college paper instead of the text message. Things like having your name, date, class, and citations are always expected of a college student!

**Note**: Not putting your name, course, and date at the top of the document you will cost your 25% of your grade.

# Submit your work to the Canvas site

After you have created your SQL ETL Processing Script file, updated the Excel Spreadsheet, and written the Word document, place them into one folder and ZIP that folder. Then upload the resulting ZIP file to the Canvas web site.

***You’re done!***