**Edmonds Community College**

**Community College Survey of Student Engagement (CCSSE)**

**Winter 2011.**

ETL transform documentation

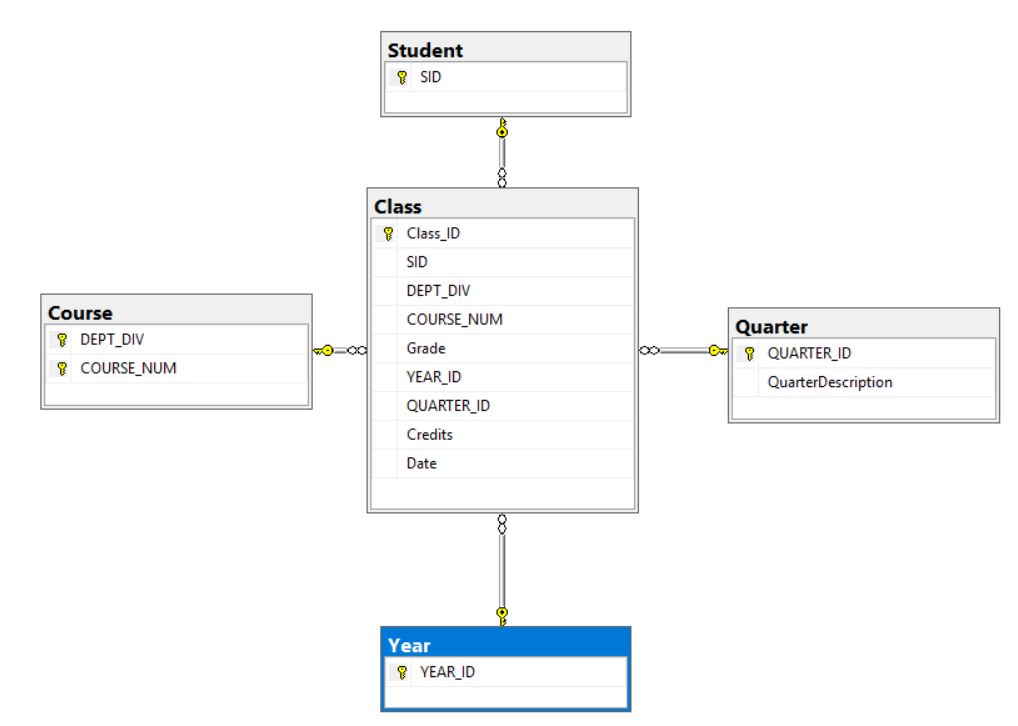
This is the documentation for the ETL transformation for the data set for the Winter2011CCSSE data. The requirements was to import the text file Winter2011CCSSE.txt, transform it and load it into Sql Server.

I gave some thought to what the data would look like once it was in SQL Server. The Data Dictionary gave me one part of the methods. I had to look over the state of the data to handle the rest. For the Department and Course number, the Data Dictionary suggested that there should be a separate table for Department and Course, but the data was in such a state that it could not have been useful until further input.

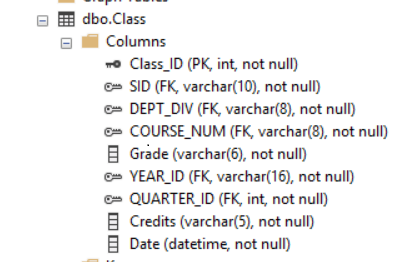
In general there were several hundred cases of bad or corrupt data. Lacking further input, I decided to create a working transformation and handle the .1 percent of bad data once further guidance was given.

I must apologize. I was limited to the capabilities of a home computer. With better tools, I would have used the Microsoft Sql Server Integration Services. However, the data provided was in such a state that exceptions would have been the rule. I feel that writing this as a CSharp program will give us the gist of what was needed.

The database design ended up like so:



The table design on the main table had the following definition.



I do not list the other tables. The columns have the same definition and use Foreign Key Constraints to maintain consistency.

The system is composed of two parts.

C# Console program to create a database with required tables and import the data a file specified on the command line. If the file is not specified, it uses Winter2011CCSSE.txt.

A SQL script that creates the database and table definitions. This script is run whenever the CSharp program is run.

The Data Dictionary was very specific about some details. Information would be needed for further work. The following elements were discovered while I was working on this.

* non-standard Departments
* course numbers with letters
* Year ID is supposed to be a three letter code. The list of codes were not supplied
* Grade included “missing data”
* Credits contained values not in the Data Dictionary
* Dates were inconsistent with the school calendar year.

**select \* from EdmondsInterview.dbo.Course**

**select \* from EdmondsInterview.dbo.Student**

**select \* from EdmondsInterview.dbo.Year**

**select \* from EdmondsInterview.dbo.Class**

**select \* from EdmondsInterview.dbo.Quarter**

**select Count(\*) as Class from EdmondsInterview.dbo.Class**

**select Count(\*) as Course from EdmondsInterview.dbo.Course**

**select Count(\*) as Quarter from EdmondsInterview.dbo.Quarter**

**select Count(\*) as Student from EdmondsInterview.dbo.Student**

**select Count(\*) as Year from EdmondsInterview.dbo.Year**

Thank you for the chance to do this project.