

CS4420/5420 Music DB Data Acquisition

In order to make the task of loading an interestingly large amount of chart data into the student databases for the SQL project, your tireless teacher has gone to great lengths to convert the Excel XLS data into an SQL table for you. In order for this to be as compatible as possible with various SQL implementations, the table is given to you as a 'do it yourself' kit, including all the parts to run on your particular SQL system.

The basic process is to create the table in SQL, then load the table from the provided comma-delimited file. The result is the flat file, an SQL table which exhibits very poor RDBMS characteristics. This will be the source table, from which you can build your nice, 3NF database, after which, this table should be discarded– it must not be part of your final db design, because it is REALLY ugly in spots.

Note that the SQL table has changed some of the original data fields, and omitted ones that are not needed for your implementation. The 'Featured' and 'Unfeatured' fields have been omitted entirely. The 'Written by' has been replaced with 8 'Written By1'..'Written By8' fields... a 'repeating group that should be eliminated in your implementation! The "Week1'..'Week76' is another repeating group to be eliminated.

There are three nice commands you can use with the SQL command line interface.

1) tee <filename>

This pipes all your activity to the named file as ascii text

2) notee

Turns #1 off.

3) source <filename>

Reads a plain text file and executes it as if you had typed it in.

Several files have been uploaded to the Canvas Files section. The file 'example.txt' from the website shows the exact command sequence for you to use to construct your raw data table. It is the result of using 'tee' to create the output of the session. The line 'source build.txt' reads the file called build.txt to create the table 'top100', ready to be filled in. The 'source load.txt' command then reads from the 'chart19641975.csv' file, which contains all the chart info for the years 1964 to 1975. You will need to make sure the path to the CSV file in load.txt is correct for your file placement on your machine.

So, within SQL, you will first create a database, then create and load the raw chart data, then use that data to build your tables.