

CA270 Coursework

(due for completion in week 11)

Design a data warehouse and construct the data marts to solve 2 specific problems.

Approach:

Begin by picking a domain area (education, healthcare, sport, gaming, movies etc.).

What type of problem are you trying to solve?

- I want to understand how many prescriptions are written by county, by town, by pharmacist, by patient in a 6 month period.
- Imagine a similar breakdown for DCU
- Movies: How much money do we earn by genre, studio, actor for movies made in California, for movies made in 2018. Can we do an analysis based on IMDB rankings?

What to do:

Step 1. Create a Star Schema (constellation) of your choice (40 marks)

Step 2. Spec 2 OLAP queries (20 marks). These must be well motivated (from above).

OLAP creates the datasets by a series of Joins. Identify the attributes you need from across the dimensions.

Step 3. Code the 3 queries (20 marks) to provide the queries and datasets.

Deliverables:

Report (6-8 pages) [An extra 20 marks]

Warehouse Schema

Data Mart queries (description and SQL)

Final dataset (post OLAP query)

Suggested Timetable:

(Week 6) Use the break week to locate your dataset, and come up with some ideas for analysis

(Week 7) After Warehousing lecture: de-construct your dataset into dimensions and a fact table.

(Week 7/8) Email me with your idea for feedback and go-ahead.

(Week 11) Complete the project.