COMP4433 Data Mining and Data Warehousing

FAQ on Data Warehousing (Reference answers)

1. The MTR Corporation is starting a data warehouse project for storing and analyzing the data collected from the Octopus system. There are three dimension tables designed and their attributes are as follows:
   - Station: *Station_Code, Station Name, Region, Line, Station Manager, Number of Gates*
   - Period: *Period_Code, Year, Month, Quarter, Day of the Week, Day, Hour*
   - Class: *Class_Code, Class, Free Journey*

   The foreign keys of the fact table include *class key key* (linking to class dimension table), *period key key* (linking to period dimension table), *from_station key* (linking to station dimension table) and *to_station key* (linking to station dimension table). The facts being recorded by the fact table are *total dollars charged* and *total number of journeys*.

   a) Design a star schema for this specification.

   b) Record the following data in your design in part (a).
      - 3000 journeys of class code 1000 (Class=Adult, Free Journey=No) from station S5 (Causeway Bay Station) to station S9 (Central Station) for the period P4 (2002, May, Q2, Monday, 6, 9am) were found. The total dollars charged were 30K. The Causeway Bay Station is on the Hong Kong line and is located in region P. It has 25 gates and the station manager is Oscar Au. The Central Station is also on the Hong Kong line and is located in region Q. It has 35 gates and Peter Lau is the station manager.

   c) Draw a data cube (lattice of cuboids) according to the above specification.

   d) Starting from the base cuboid [From_Station_Code, To_Station_Code, Period_Code, Class_Code], what specific OLAP operations should be performed to list the total charge of all classes and free journey types of journeys from CWB station to Central station in 2018?