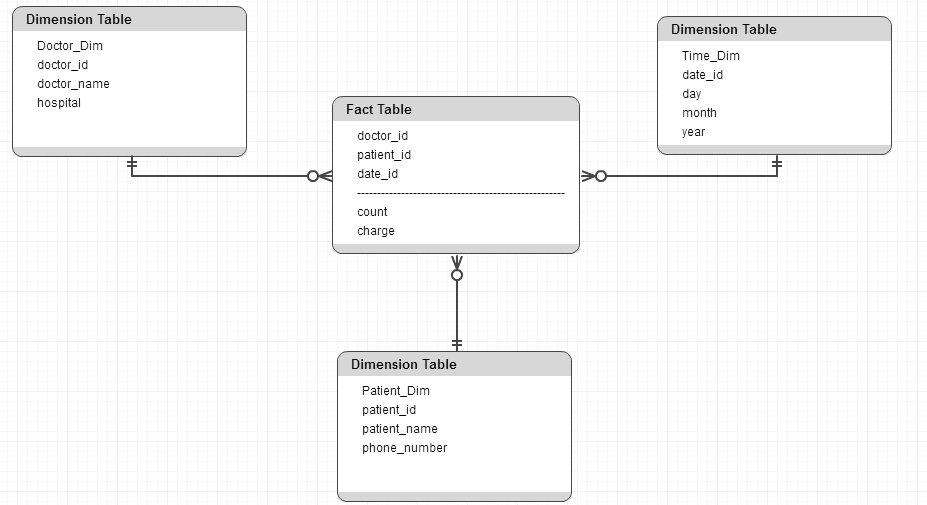
Tutorial 1 (WQD 7005, Data Mining)

Suppose that a data warehouse consists of the three dimensions *time*, *doctor*, and *patient*, and the two measures *count* and *charge*, where *charge* is the **fee** that a doctor charges a patient for a visit.

1. Draw a schema diagram for the above data warehouse.



1. Starting with the base cuboid[*day*, *doctor*, *patient*], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2018?

1. roll up from day to month to year

2. slice for year = “2018”

3. roll up on patient(all the patient available)

4. slice for all patient

5. get the list of total fee collected by each doctor in 2018

1. To obtain the same list, write an SQL query assuming the data is stored in a relational database with the schema (*day*, *month*, *year*, *doctor*, *hospital, patient*, *count*, *charge*).

Select doctor\_id, sum(charge) as total\_fee from fee\_table where the year = 2018 group by doctor\_id;