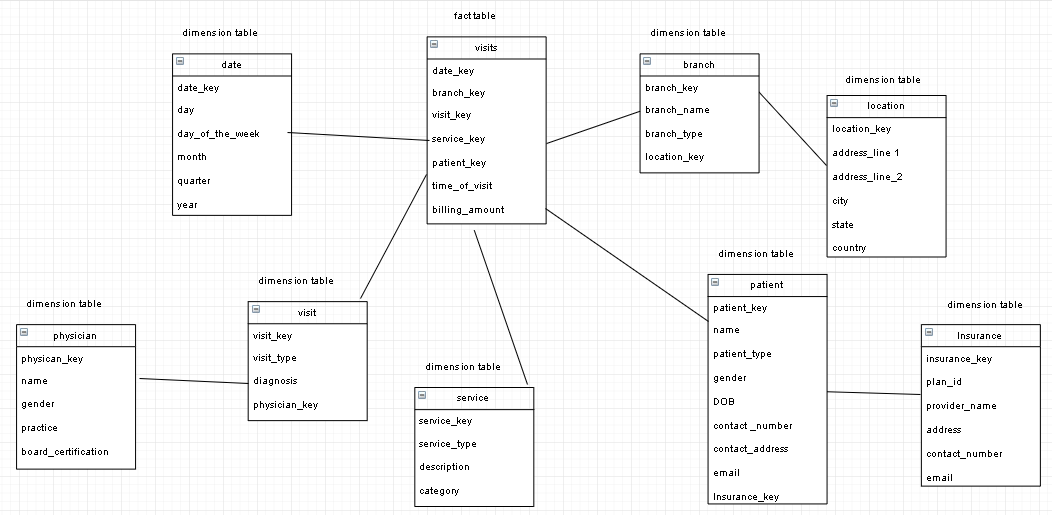
1. Suppose that a data warehouse for a Hospital chain consists of the five dimensions’ date, Hospital branch, visit type, services, patient, and the two measures time of visit and billing amount, where billing amount is the amount that a patient pays for a particular visit starting on a given admission date. Patients may be children, adults or elderly. Hospitals are in US and the visit type include Emergency, Consulting or Admission

a) Draw a snowflake schema diagram for the data warehouse.

b) Starting with the base cuboid, what specific OLAP operations should one perform in order to list the total billing amount of children patients in the Charlotte branch in 2015 for all Emergency type of visits.

c) If each dimension has four levels (excluding all), how many cuboids will this contain (including the base and apex cuboids)?

**Response:**

**a) **

**b)** Base cuboid consists of (date, branch, patient, visit, service). Perform the following OLAP operations:

1. Slice for service on service\_type = All (Since nothing is mentioned about services, assuming that all types of services are to be included)
2. Roll-up on date from date\_key to year
3. Roll-up on branch from branch\_key to location\_key to city
4. Roll-up on patient from patient\_key to patient\_type
5. Roll-up on visit from visit\_key to visit\_type
6. Dice with (year = 2015 and city = Charlotte and visit\_type = Emergency and patient\_type = Children) to get the total billing amount

**c)** If each dimension has four levels (excluding all), the number of cuboids (including the base and apex cuboids) are:



n = 5, Li = 4(excluding all, for each dimension)

T **=** (4+1) x (4+1) x (4+1) x (4+1) x (4+1) = 55 = 3125 cuboids.

1. Suppose that a data warehouse contains 20 dimensions, each with about five levels of granularity.

a) Users are mainly interested in four particular dimensions, each having three frequently accessed levels for rolling up and drilling down. How would you design a data cube structure to efficiently support this preference?

b) At times, a user may want to drill through the cube, down to the raw data for one or two particular dimensions. How would you support this feature?

**Response:**

**a)**  Since the users are interested only in 4 dimensions out of 20, we can use Partial Materialization to selectively compute a proper subset (in this case 44 ) of the whole set of possible cuboids to create a shell cube. By pre-computing the cuboids for only a small number of dimensions, we can minimize the storage space and increase the response time.

**b)** To support this user requirement, any additional combinations of the dimensions can be computed on-the-fly. Since the user may only need this feature occasionally, the extra time required to compute one or two particular dimensional aggregates on-the-fly should be okay.