Page No.
Date

Steps for the design & construction of Data warehouse

- A Business Analysis Framework ⇒
What can business analysis gain from having a data warehouse.
  - Win over competitors.
  - enhance business productivity.
  - Data warehouse facilitates customer relationship management.
  - cost reduction by tracking trends, patterns, & exceptions over long periods in a consistent and reliable manner.

- Four different views regarding the design of a data warehouse must be considered:
  - top-down view
  - data source view
  - data warehouse view
  - business query view.

* Top-down view ⇒ allows the selection of the relevant information necessary for the data warehouse. This information matches the current & future business needs.

* Data-Source View ⇒ exposes information being captured, stored & managed by operational systems.

* The data warehouse view: includes fact tables & dimension tables. It provide historical context.

* Business Query view ⇒ is the perspective of data in the data warehouse from the viewpoint of the end user.

- In general, Warehouse design process consists of following steps.

1. Choose a business process model:
   - for example, order, invoices, shipments, inventary, sales etc.
   - if the process is departmental, a data mart model should be choosen.

2. Choose the grain of the business process
   - for example, ~~order~~, individual transaction, individualy daily snapshot (represent in fact table)

3. Choose the dimension that will apply to each fact table record.
   - typical dimensions are time, item, customer, supplier, transition type & status.

4. Choose the measures that will populate each fact table record
   - typical measures are numeric additive quantities like dollars_sold & units_sold.

Page No.
Date / /

⇒ Metadata Repository ⇒

- Metadata are data about data
- Metadata are data that defines warehouse objects.
- Metadata repository contains
  - A description of the structure of the data warehouse
  - operational metadata
  - Algorithms used for summarization
  - mapping from the operational environment to the data warehouse.
  - ~~Data~~

Shrikrupa

* The
tables
list

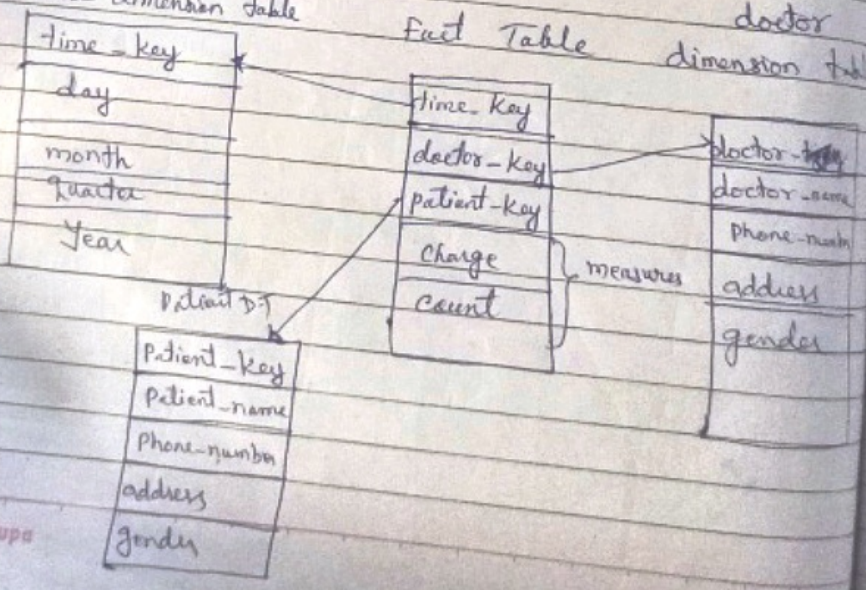
* ... warehouse consists of three ... patient & two measures ... charge is the fee that ... a patient for a visit.
... Star schema diagram for above data ...

b) Starting with base cuboid [day, doctor, patient] ... specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2010?

c) Starting with base cuboid [day, doctor, patient] what specific OLAP operations should be performed in order to list the total fee collected by doctors D1, D2 in 2010, 2011 by patient P1, P2?

Sol": Star Schema

(a) Time dimension table



Page No.
Date

Exercise:

① Suppose that data warehouse consists of three dimensions time, doctor, patient & two measures count and charge, what charge is the fee that a doctor charges a patient for a visit.

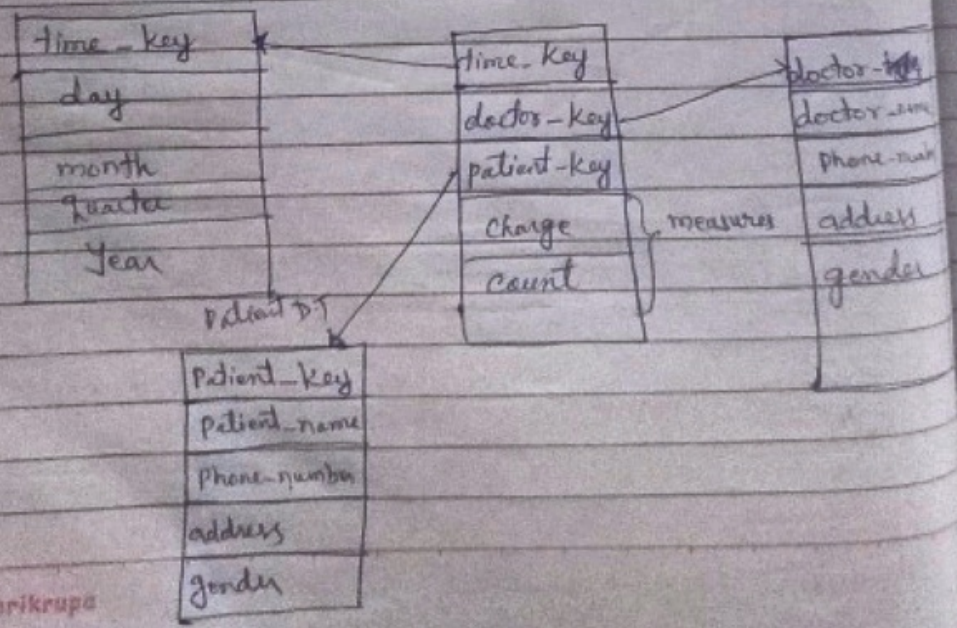
a) Draw a star schema diagram for above data warehouse.

b) Starting with base cubaid [day, doctor, patient] what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2010?

c) Starting with base cubaid [day, doctor, patient] what specific OLAP operations should be performed in order to list the total fee collected by doctors D1, D2 in 2010, 2011 by patient P_1, P_2?

Ans: Star Schema

(a)



b) The operations
Roll-on
Slice
Roll-on
patient

c) The operations
Roll up
Dice on
(doctor
"P1"

b) The operations to be performed are:
Roll-on time from day to year
Slice for time = "2010"
Roll-on patient from individual patient to all

c) The operations performed are.
Roll up on time from day to year
Dice on the
(doctor = "D1" or "D2") and (patient = "P1" or "P2") and (year = "2010" or "2011")

Exercise 2:

Suppose that a data warehouse consists of the four dimensions (date, spectator, location and game) and two measures, count and charge where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults or seniors with each spectators having its own charge rate.

a) Draw star schema

b) Starting with base cuboid [date, spectator, location, game] what specific OLAP operation should be performed in order to list the total charge paid by student spectators at Mumbai place in 2010?

c) Starting with base cuboid [date, spectator, location, game] what specific OLAP operations should be performed in order to list the total charge paid by student, adult spectators at Mumbai, Pune place in 2013, 2014?

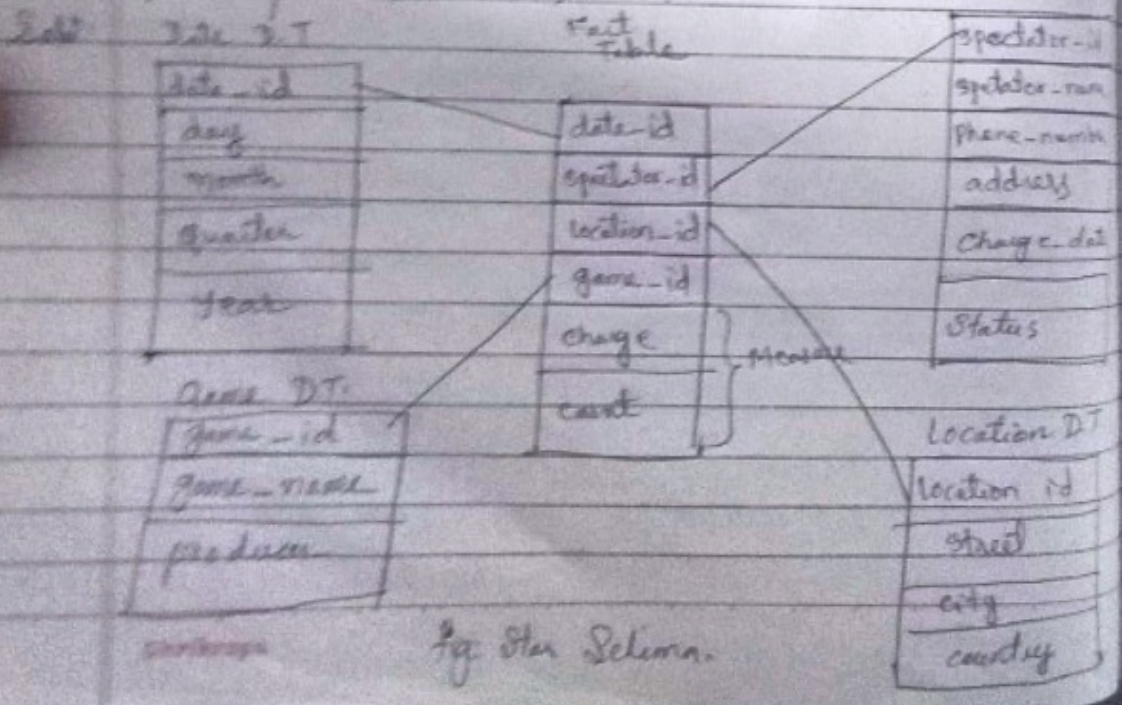


fig: Star Schema.

b) OLAP operations are
Roll-up on date from date to year
Set location as cities
Dice with status = "Students" and location = "Mumbai" and year = "2010"

c) OLAP operations are
- Roll up date to year
- Rollup on location from location_id to cities
- Rollup on spectator from spectator_id to status.
- Dice with
(status = "student" or "adults") and (location = "Mumbai" or "pune") and (year = "2013" or "2014")