

(1)

M	T	W	T	F	S	S
Page No.:	YOUVA					
Date:						

Name: Danyl Fernandes

Roll: 72

ID: 2020012004

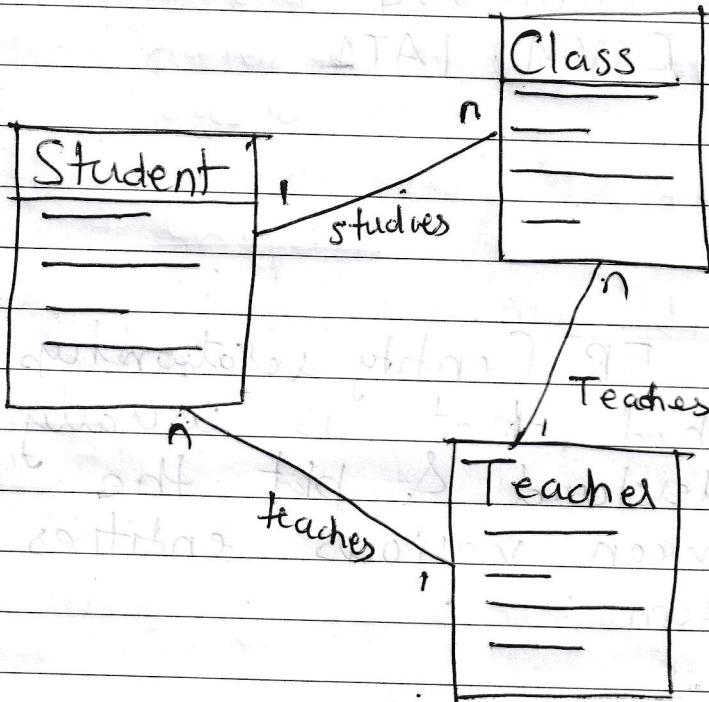
Sub: DWM IAT1

(Q2)

- 1) • The ER (entity relationship) model is a method that is usually used to understand & plt the flow & relations between various entities in a given system.
- Since, it is based on more recent data and regularly used for transactions, it is also called as OLTP (Online Transaction Processing) databases.
- However, in dimensional modelling the database / system must be optimised for read speeds and not writes as OLTP.
- Dimensional modelling is different because it is used as a data store, for historical data & for analysis. Hence it is also called OLAP (Online Analytical Processing) Databases.
- It has mainly fact tables & dimension tables & is optimised for analysis & faster reads.

Eg:

ER



Dimensional

Teacher
id
name
classes

Student-fact
study_hours
teacher_id
class_id
university_id

University
id
name

Class
id
name
taught_by

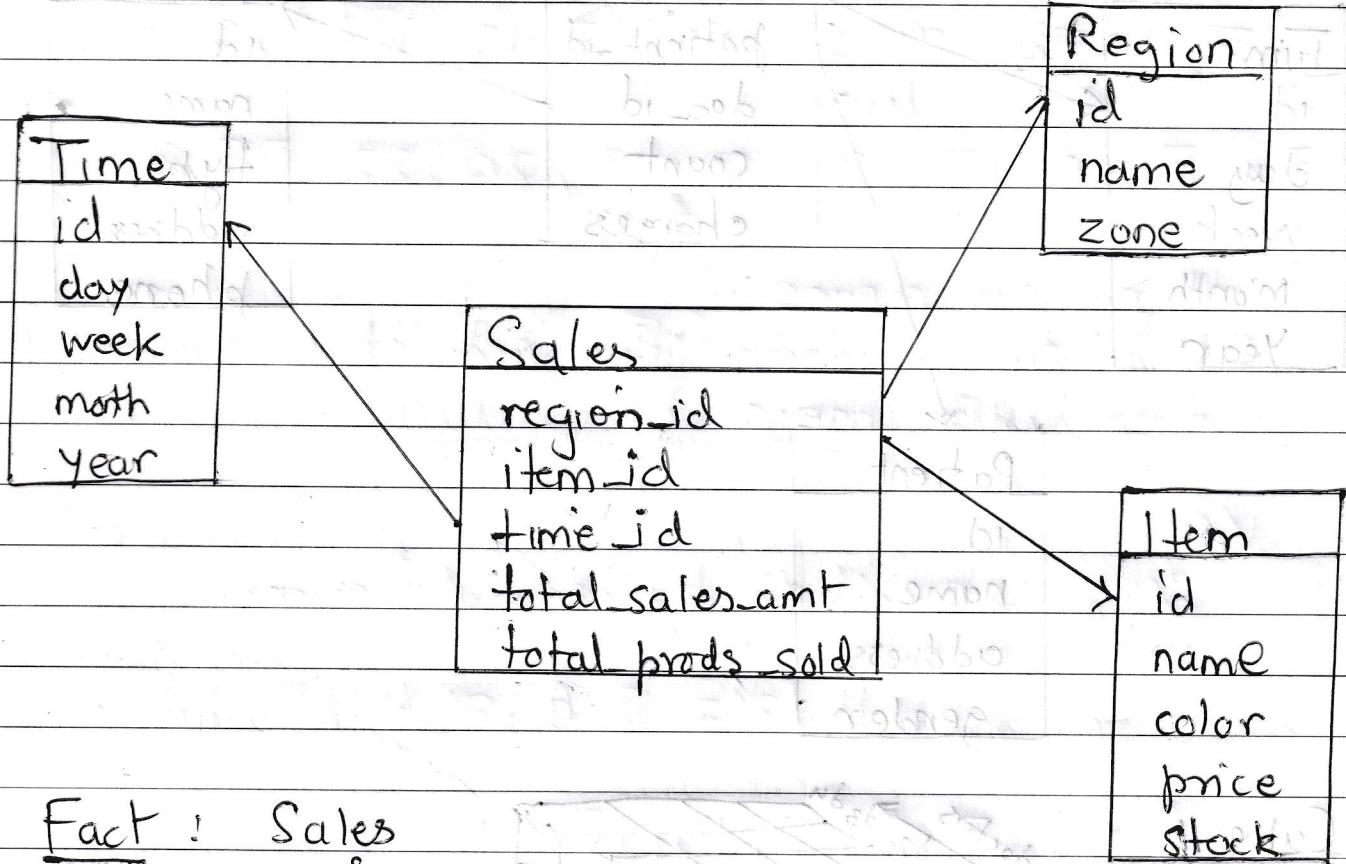
Dimension tables

(Q2)  
2)

3

M	T	W	T	F	S	S
Page No.:	YOUVA					
Date:						

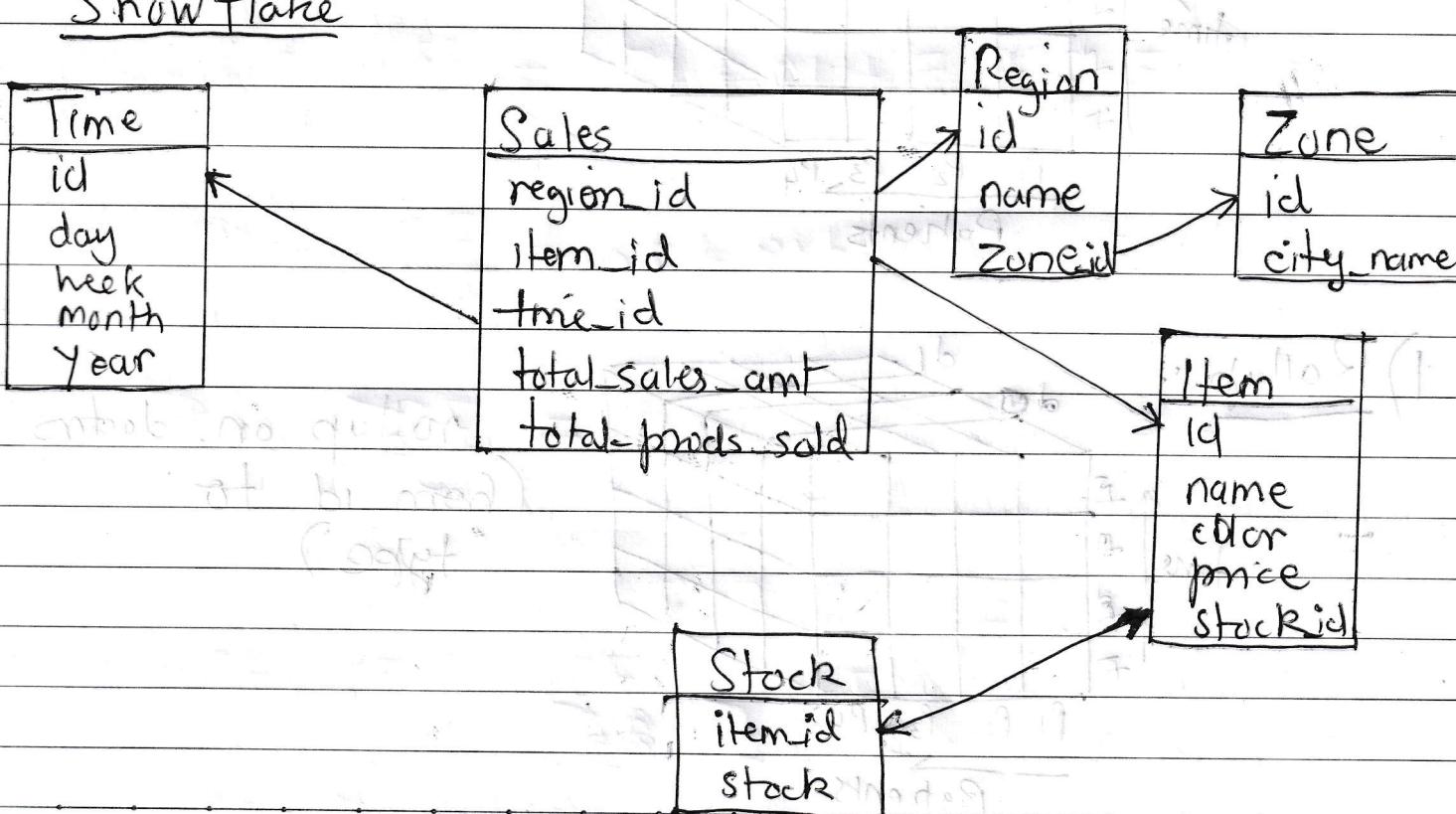
## Star Schema



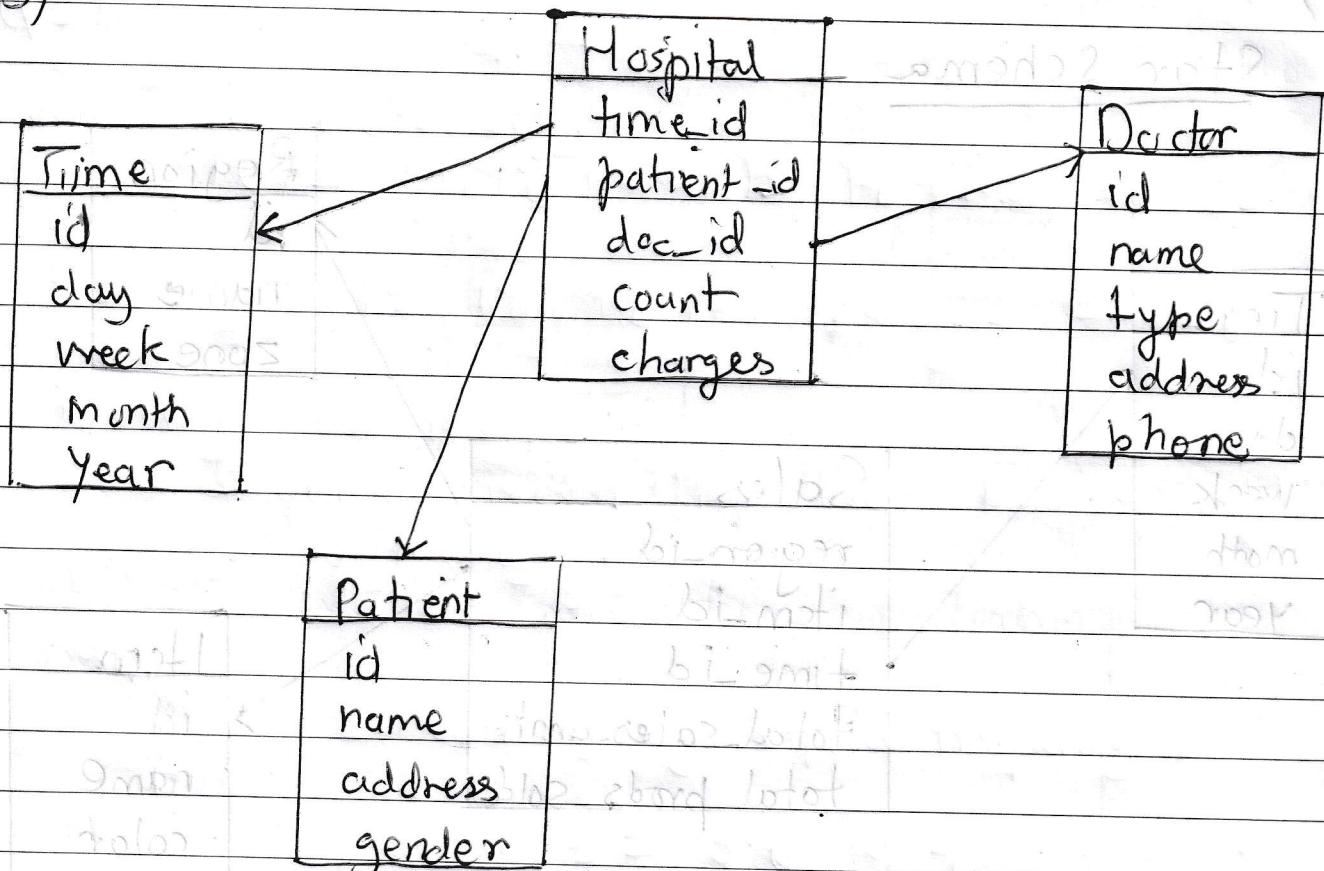
Fact : Sales

Dimensions : { Time , Item , Region }

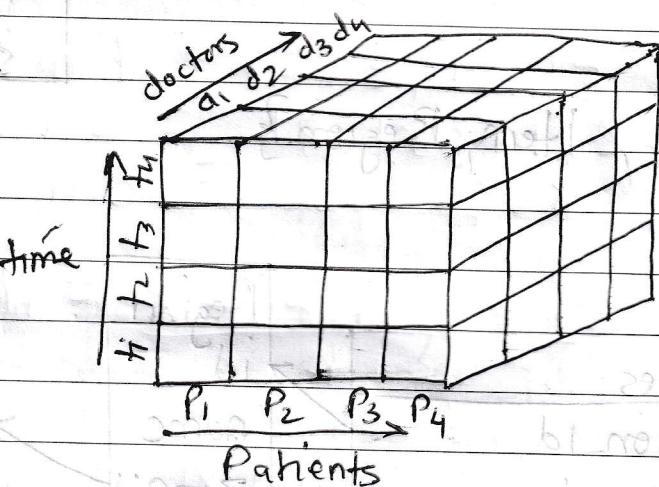
## Snowflake



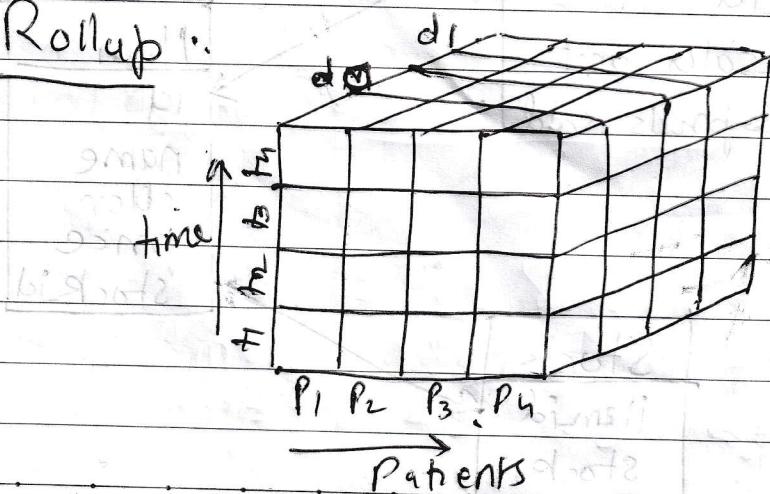
3) Q2)



Cubes:



1) Rollup:

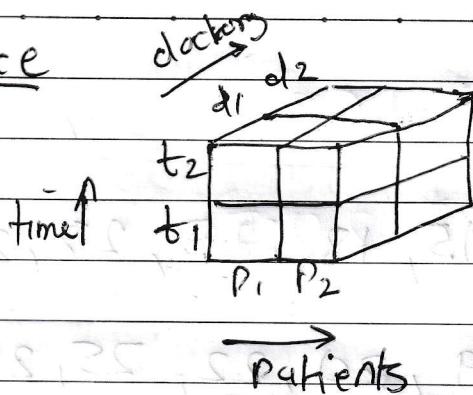


rollup on doctors  
(from id to type)

(5)

M	T	W	T	F	S	S
Page No.:	YOUVA					
Date:						

ii) Dice



dice for

doctors =  $d_1$  or  $d_2$ and time =  $t_1$  or  $t_2$ and Patients =  $P_1$  or  $P_2$ 

- Operations to find total charges of all doctors of Mumbai for year = 2020

- 1) Rollup on time from day to year
- 2) Rollup on doctors from id to address = Mumbai
- 3) Slice on time with year = ~~2020~~ 2020
- 4) Rollup on patient from individuals to all

Q3)

2) Given data set:

5, 10, 2, 19, 18, 20, 15, 18, 25, 28, 22

= 2, 5, 10, 15, 18, 18, 19, 20, 22, 25, 28

i) Using bin means:

- Calculating mean for each bin &amp; replacing values with mean

Ans

Bin / Smoothing by equal frequency

Bin = bins

No. of bins  $11/3 = 3.6 \approx 4$ 

$$\text{Bin 1} = 2, 5, 10, 15$$

$$\text{Bin 2} = 18, 18, 19, 20$$

$$\text{Bin 3} = 22, 25, 28$$

$$\bar{M}_1 = \frac{2+5+10+15}{4} = \frac{32}{4} = 8$$

$$\bar{M}_2 = \frac{18+18+19+20}{4} = \frac{75}{4} = 18.75$$

$$\bar{M}_3 = \frac{22+25+28}{3} = \frac{75}{3} = 25$$

$$\therefore \text{Bin 1} = 8, 8, 8, 8$$

$$\text{Bin 2} = 18.75, 18.75, 18.75, 18.75$$

$$\text{Bin 3} = 25, 25, 25$$

(7)

M	T	W	T	F	S	S
Page No.:	YOUVA					
Date:						

## ii) Using bin boundaries

- Calculate min & max box values & replace values with closest boundary

$$\text{Bin 1} := \min = 2$$

$$((\text{max} = 15) + 2) = 17$$

$$\text{Bin 2} (= \min = 18) + 2 = \\ \text{max} = 20$$

$$\text{Bin 3} := \min = 22 \\ \text{max} = 28$$

$$\therefore \text{Bin 1} = 22, 15, 15$$

$$\text{Bin 2} = 18, 18, 20, 20$$

$$\text{Bin 3} = 22, 28, 28$$

(Q3) 3)

25, 33, 15, 20, 32, 25, 30, 10, 35, 40, 46,  
52, 70

10, 15, 20, 25, 25, 30, 32, 33, 35, 40, 46, 52,

$$Q1 = 32$$

$$Q2 = \frac{20+25}{2} = 22.5$$

$$Q3 = \frac{40+46}{2} = 43$$

$$IQR = 43 - 22.5 \\ = 20.5$$

(Q3)3)

Data :-

10, 15, 20, 25, 25, 30, 32, 33, 35, 40, 46, 52, 70

$$\text{Median} = 32$$

$$Q_1 = 22 + 25 / 2 = \underline{\underline{22.5}}$$

$$Q_3 = 40 + 46 / 2 = \underline{\underline{43}}$$

$$IQR = 20.5$$

$$\begin{aligned} \text{Min} &= Q_1 - (1.5 * IQR) \\ &= 22.5 - 30.75 \\ &= -8.25 \\ &\approx 8.25 \end{aligned}$$

$$\begin{aligned} \text{Max} &= Q_3 + (1.5 * IQR) \\ &= 43 + 30.75 \\ &= 73.75 \end{aligned}$$

