

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

1) Name of three denormalization techniques

- Splitting Tables
- Pre-Joining
- Adding Redundant Columns

2) Ek doctor or patient data usko denormalized table me likhna tha

Denormalized Table: MedicalData

RecordID	DoctorName	DoctorSpecialty	PatientName	PatientAge	PatientCond
1	Dr. Smith	Cardiology	John Doe	45	Hypertensio
2	Dr. Johnson	Orthopedics	Jane Doe	30	Fractured Ar
3	Dr. Smith	Cardiology	Alice Brown	55	Arrhythmia

3) Ek tabe me additive and non-additive me difference krna tha

Additive facts are easy to work with Summing the fact value gives meaningful results

Additive facts: f Quantity sold f Total Rs. sales

Non-additive facts: f Averages (average sales price, unit price) f Percentages (% discount) f

Ratios (gross margin) f Count of distinct products sold

4) Ek schema wala tha

Schemas

Dimensions are the descriptive textual information and the source of interesting constraints on how we filter/report on the quantities (e.g., by geography, product, date, etc.). For the DM shown, we constrain on the clothing department via the Dept attribute in the Product table.

5) aik question transitive dependency use ha k ni es tra. Ka kch tha

A transitive dependency occurs when a non-key column that is a determinant of the primary key is the determinate of other columns.

6) Statements thi or un ko identify kr k lekhna tha en mein Sy Kon Se application kidr use Hoti.. statement given thi just btana ha ETL and ELT Kon sa Ha . Table given tha us mein facts or dimensions lekhna tha.

There are two fundamental approaches to data acquisition:

ETL: Extract, Transform, Load in which data transformation takes place on a separate transformation server.

ELT: Extract, Load, Transform in which data transformation takes place on the data warehouse server.

For any query feel free to ask princealvi9222@gmail.com

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

7) Enlist Two possible name of scalability problem in MOLA?

There are two possible, but limited solutions addressing the scalability problem i.e. Virtual cubes and partitioned cubes.

8) 3-OLAP ki btani thi jesay ROLAP, MOLAP

OLAP is Analytical Processing instead of Transaction Processing. It is also NOT a physical database design or implementation technique, but a framework.

OLAP Implementations

- 1. MOLAP:** OLAP implemented with a multi-dimensional data structure.
- 2. ROLAP:** OLAP implemented with a relational database.
- 3. HOLAP:** OLAP implemented as a hybrid of MOLAP and ROLAP.
- 4. DOLAP:** OLAP implemented for desktop decision support environments.

9) Basic task of transformation any three.

Data Transformation

Basic tasks *f* Selection *f* Splitting/Joining *f* Conversion *f* Summarization *f* Enrichment

10) Fact or dimension finout Karna tah.

OLAP: Facts & Dimensions

FACTS: Quantitative values (numbers) or “measures.” *f* e.g., units sold, sales \$, Co, Kg etc.

DIMENSIONS: Descriptive categories. e.g., time, geography, product etc. *f* DIM often organized in hierarchies representing levels of detail in the data (e.g., week, month, quarter, year, decade etc.).

11) Ak que m normalization ko denormalization m krna tha

For any query feel free to ask princealvi9222@gmail.com

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

Consider the following normalized tables:

1. Doctors Table:

DoctorID	DoctorName	Specialty
1	Dr. Smith	Cardiology
2	Dr. Johnson	Orthopedics

2. Patients Table:

PatientID	PatientName	Age
101	John Doe	45
102	Jane Doe	30
103	Alice Brown	55

3. Appointments Table:

AppointmentID	DoctorID	PatientID	AppointmentDate	Condition	Prescription
501	1	101	2023-01-15	Hypertension	Medication A
502	2	102	2023-02-02	Fractured Arm	Painkillers
503	1	103	2023-02-10	Arrhythmia	Medication B

Now, let's denormalize this data into a single table:

Denormalized Table: MedicalData

AppointmentID	DoctorName	Specialty	PatientName	Age	AppointmentDate	Condition	Prescription
501	Dr. Smith	Cardiology	John Doe	45	2023-01-15	Hypertension	Medication A
502	Dr. Johnson	Orthopedics	Jane Doe	30	2023-02-02	Fractured Arm	Painkillers
503	Dr. Smith	Cardiology	Alice Brown	55	2023-02-10	Arrhythmia	Medication B

12) De-normalization

Denormalization is a database optimization technique where you intentionally introduce redundancy into a table by merging or combining data from related tables.

13) Data warehouse

A Data Warehouse is not something shrink-wrapped i.e. you take a set of CDs and install into a box and soon you have a Data Warehouse up and running. A Data Warehouse evolves over time, you don't buy it. Basically it is about taking/collecting data from different heterogeneous sources.

14) Name the 3 molap cube operations

For any query feel free to ask princealvi9222@gmail.com

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

Cube Operations

f Rollup: summarize data *f* e.g., given sales data, summarize sales for last year by product category and region

f Drill down: get more details *f* e.g., given summarized sales as above, find breakup of sales by city within each region, or within Sindh

f Slice and dice: select and project *f* e.g.: Sales of soft-drinks in Karachi during last quarter

f Pivot: change the view of data

15) Statements de hoi thi un me select krna tha knsi second normal form ko represent krwa rhi or third normal form ko.

Normalization: 1NF

Update Anomalies

Normalization: 2NF

Every non-key column is fully dependent on the PK.

Normalization: 3NF

All columns must be dependent only on the primary key.

16) Table collapsing technique

Collapsing Tables.

- Two entities with a One-to-One relationship.
- Two entities with a Many-to-Many relationship

17) MOLAP cube 3 dimensions 3 marks

MOLAP Implementations

OLAP has historically been implemented using a multidimensional data structure or “cube”.

f Dimensions are key business factors for analysis:

f Geographies (city, district, division, province,...)

f Products (item, product category, product department,...)

f Dates (day, week, month, quarter, year,...)

18) ER 2 types. 3 marks

Conceptual ER Diagrams (High-Level ER Diagrams):

Physical ER Diagrams (Low-Level ER Diagrams or Database Schema Diagrams):

19) CDC 2 sources and 3 types. 5 marks

Two CDC sources

- Modern systems
- Legacy systems

Types

- Timestamps
- Partitioning
- Triggers

For any query feel free to ask princealvi9222@gmail.com

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

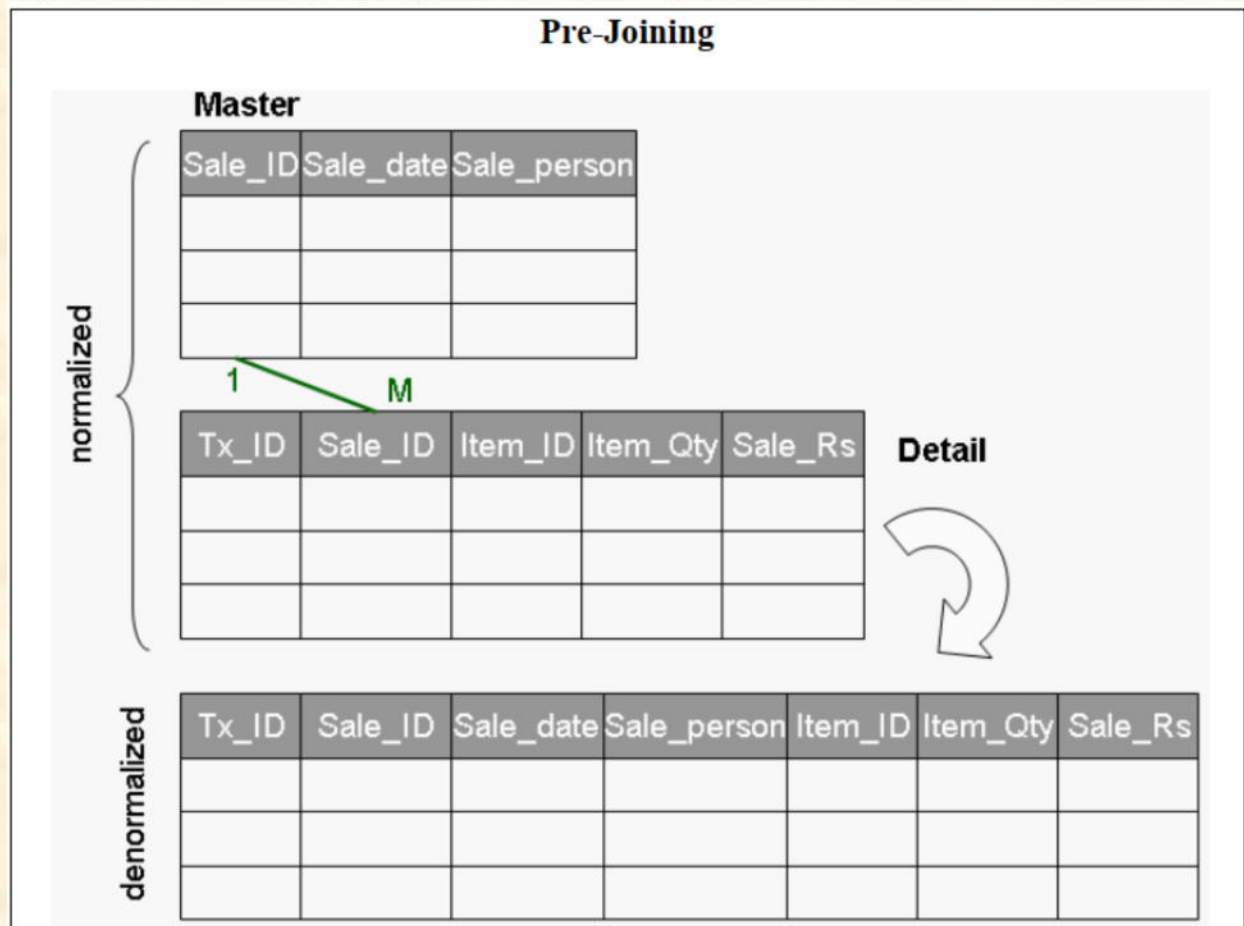
20) ER model Ki types batani thi do

1. Conceptual ER Model:
2. Physical ER Model:

21) MOLAP ko Kon C kin do type Mai David Hai aisa koi Tha question

MOLAP physically builds “cubes” for direct access - usually in the proprietary file format of a multi-dimensional database (MDD) or a user defined data structure. Therefore ANSI SQL is not supported.

22) Ek mormalized or prejoining ka table Tha long Mai



23) Capture picture Ki types likhni thi three

1. JPEG (Joint Photographic Experts Group):
2. PNG (Portable Network Graphics):
3. GIF (Graphics Interchange Format):

24) 3 tb to bytes

1 TB = 2^{40} or 10^{12} bytes = 50,000 trees chopped and converted into paper and printed

For any query feel free to ask princealvi9222@gmail.com

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

25) 5 pb to bytes conversion.

1 PB = 2^{50} or 10^{15} bytes = Academic research libraries across the U.S.

26) Find the 2nd or 3rd normal form of the table.

Consider a student database system to be developed for a multi-campus university, such that it specializes in one degree program at a campus i.e. BS, MS or PhD.

SID	Degree	Campus	Course	Marks	SID: Student ID
1	BS	Islamabad	CS-101	30	Degree: Registered as BS or MS student Campus: City where campus is located Course: Course taken Marks: Score out of max of 50
1	BS	Islamabad	CS-102	20	
1	BS	Islamabad	CS-103	40	
1	BS	Islamabad	CS-104	20	
1	BS	Islamabad	CS-105	10	
1	BS	Islamabad	CS-106	10	
2	MS	Lahore	CS-101	30	
2	MS	Lahore	CS-102	40	
3	MS	Lahore	CS-102	20	
4	BS	Islamabad	CS-102	20	
4	BS	Islamabad	CS-104	30	
4	BS	Islamabad	CS-105	40	

Figure-6.1: Part of example student database system

For any query feel free to ask princealvi9222@gmail.com

Database system in 1NF				
SID	Degree	Campus	Course	Marks
1	BS	Islamabad	CS-101	30
1	BS	Islamabad	CS-102	20
1	BS	Islamabad	CS-103	40
1	BS	Islamabad	CS-104	20
1	BS	Islamabad	CS-105	10
1	BS	Islamabad	CS-106	10
2	MS	Lahore	CS-101	30
2	MS	Lahore	CS-102	40
3	MS	Lahore	CS-102	20
4	BS	Islamabad	CS-102	20
4	BS	Islamabad	CS-104	30
4	BS	Islamabad	CS-105	40

Table FIRST

Figure-6.2: Part of example student database system in 1NF

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

REGISTRATION	SID is now a PK			PERFORMANCE	SID	Course	Marks
					1	CS-101	30
					1	CS-102	20
					1	CS-103	40
					1	CS-104	20
					1	CS-105	10
	SID	Degree	Campus		1	CS-106	10
	1	BS	Islamabad		2	CS-101	30
	2	MS	Lahore		2	CS-102	40
	3	MS	Lahore		3	CS-102	20
	4	BS	Islamabad		4	CS-102	20
	5	PhD	Peshawar		4	CS-104	30
					4	CS-105	40

PERFORMANCE in 2NF as (SID, Course) uniquely identify Marks

Figure-6.3: Part of example student database system in 2NF

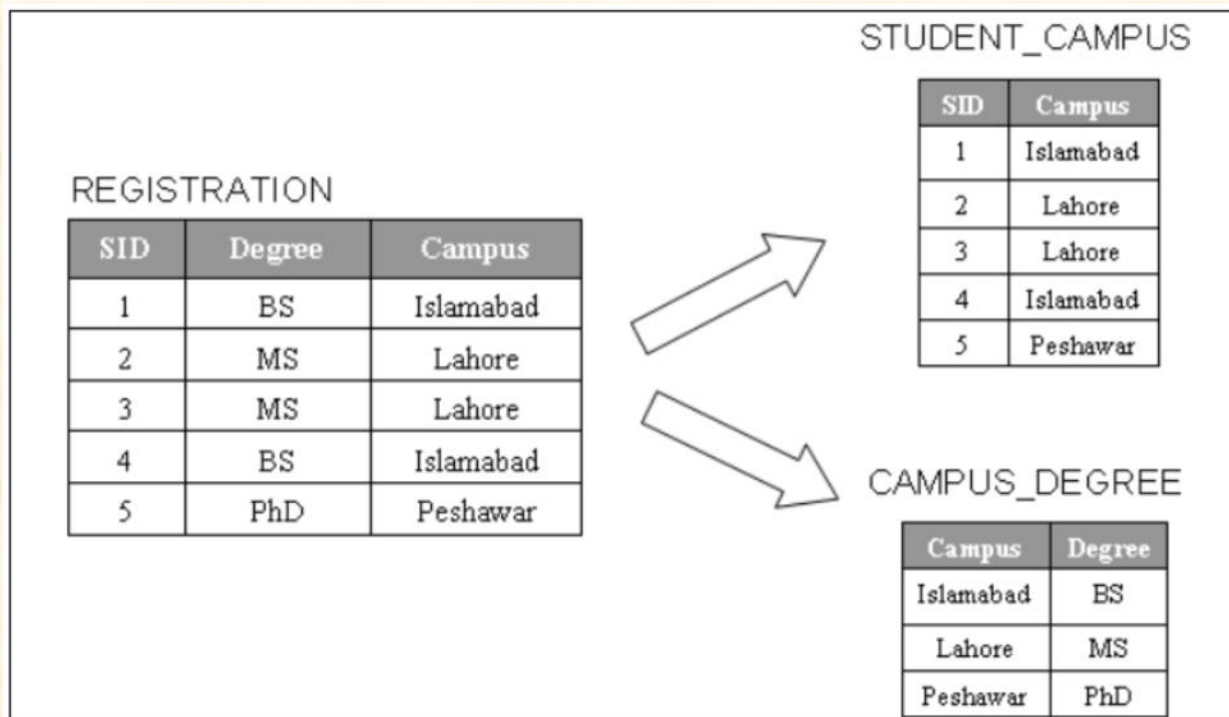


Figure-6.4: Part of example student database system in 3NF

27) SDLC isi trha kuch tha iski types ko arrange krna tha.
Classical SDLC

For any query feel free to ask princealvi9222@gmail.com

CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

f Requirements gathering *f* Analysis *f* Design *f* Programming *f* Testing *f* Integration *f* Implementation

DWH SDLC (CLDS)

f Implement warehouse *f* Integrate data *f* Test for biasness *f* Program w.r.t data *f* Design DSS system *f* Analyze results *f* Understand requirement

28) Jasy k btyea k offline logical extraction hai agr hai Tu disscuss krt

Offline Extraction

f Data NOT extracted directly from the source system, instead staged explicitly outside the original source system.

f Data is either already structured or was created by an extraction routine.

f Some of the prevalent structures are:

f Flat files

f Dump files

f Redo and archive logs

f Transportable table-spaces

29) MoLAP Jo hai wo mobile k liea use hota k ne

MOLAP is a technology that enables users to access and analyze data remotely using their mobile devices, providing a convenient and flexible way to work with OLAP data.

30) Feature OLTP OLAP

Level of data OLTP = Detailed OLAP = Aggregated

Amount of data retrieved per transaction OLTP = Small OLAP = Large

Views OLTP = Pre-defined OLAP = User-defined

Typical write operation OLTP = Update, insert, delete OLAP = Bulk insert, almost no deletion

“age” of data OLTP = Current (60-90 days) OLAP = Historical 5-10 years and also current

Tables OLTP = Flat tables OLAP = Multi-Dimensional tables

Number of users OLTP = Large OLAP = Low-Med

Data availability OLTP = High (24 hrs, 7 days) OLAP = Low-Med

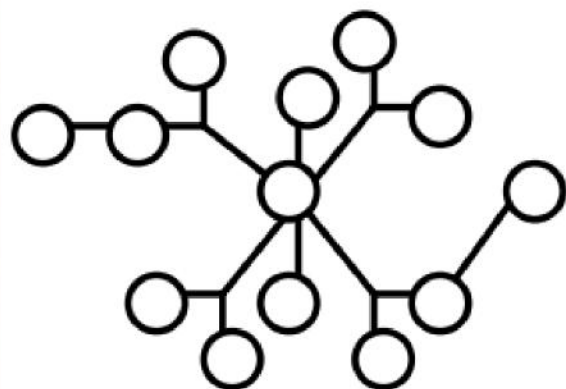
Database size OLTP = Med (GB- TB) OLAP = High (TB – PB)

Query Optimizing OLTP = Requires experience OLAP = Already “optimized”

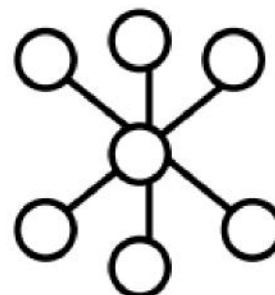
31) Diagram given thi btana tha k kn sa schema h

For any query feel free to ask princealvi9222@gmail.com

The two Schemas



Snow-flake



Star

32) btana tha or btana tha k frud detection kn sy scenario ma use hona credit risk kis ma use hona or yied management kis ma use hony yi

Fraud detection

- f By observing data usage patterns.
- f People have typical purchase patterns.
- f Deviation from patterns.
- f Certain cities notorious for fraud.
- f Certain items bought by stolen cards.
- f Similar behavior for stolen phone cards

Credit risk prediction

- f Who should get a loan?
- f Customer segregation i.e. stable vs. rolling.
- f Qualitative decision making NOT subjective.
- f Different interest rates for different customers
- f Do not subsidize bad customer on the basis of good.

Yield Management

- f Works for fixed inventory businesses.
- f The price of item suddenly goes to zero.
- f Item prices vary for varying customers.
- f Example: Air Lines, Hotels, Car rentals etc.
- f Price of (say) Air Ticket depends on:

33) Historical view of data processing page 12 k sab years 5 marks k table mn match krna

1) Master Files & Reports = 190 =

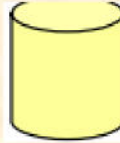


2) Lots of Master files! = 1965 =



CS614 MIDTERM CURRENT PAPER FALL 2023 SOLVED BY PRINCE ALVI

3) Direct Access Memory & DBMS = 1970 =



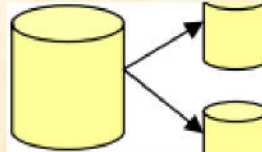
4) Online high performance transaction processing = 1975 =



5) PCs and 4GL Technology (MIS/DSS) = 1980 =



6) Extract programs, extract processing = 1985 =



7) The legacy system's web = 1990 =



34) Pre-Joining

- f* Typical of Market basket query
- f* Join ALWAYS required
- f* Tables could be millions of rows
- f* Squeeze Master into Detail
- f* Repetition of facts. How much?
- f* Detail 3-4 times of master

35) OLAP FASMI Test

Fast: Delivers information to the user at a fairly constant rate i.e. $O(1)$ time. Most queries answered in under 5 seconds.

Analysis: Performs basic numerical and statistical analysis of the data, pre-defined by an application developer or defined ad-hocly by the user.

Shared: Implements the security requirements necessary for sharing potentially confidential data across a large user population.

Multi-dimensional: The essential characteristic of OLAP.

Information: Accesses all the data and information necessary and relevant for the application, wherever it may reside and not limited by volume.

36) Typical Applications of DWH

- f* Fraud detection.
- f* Profitability analysis.
- f* Direct mail/database marketing.
- f* Credit risk prediction.
- f* Customer retention modeling.
- f* Yield management.
- f* Inventory management.

For any query feel free to ask princealvi9222@gmail.com