UNIT-I

→ A database management system (DBMS) is a collection of interrelated data and a set of programs to access those data.

→ The collection of data usually refered as a database contains information relavent to an enterprises

→ The primary goal of a database is to provide a way to store and retrive database information i.e both convenient and efficient.

database system applications :-

→ databases are widely used. Here are some of the applications

i) enterprise information
   - Sales
   - accounting
   - human resources
   - manufacturing.

ii) Banking and finance
   - Deleting accounting
   - credit and transfer the amount
   - loans
   - opening an account

iii) universities/ institutions
   - course information systems
   - student information system
   - ~~stock~~ information system
   - human resource information system
   - staff information system.

Scanned by CamScanner

iv) airlines -

- ticket booking / cancelling
- reservation

v) telecommunication

Purpose of the database systems :- (database system Vs file processing system)

The purpose of the database systems is computerised one commercial office. In earlier 1960's all the data are maintained manually.

→ In this manual process their may be a chance of corruption, problems to maintain (or) store the data like huge manpower involvement multiple copies are maintained manually.

→ That leads to delay of the process. to overcome above disadvantages we should follow database systems in our organisations.

There are 2 methods to computerize one office

i) file processing system.

ii) data base system.

Before database system file processing systems are used to computerize one office

→ In file processing systems there are so many disadvantages exists

i) data redundency and inconsistency

↓ duplication of file

ii) difficulty in accessing data

Scanned by CamScanner

iii) data isolation.

iv) integrity problems

→ The data values stored in the database must satisfy certain types of consistency constraints.

constrain - is a rule which is not violated.

v) Automicity Problems :-

Atomicity maintaing data any where else.

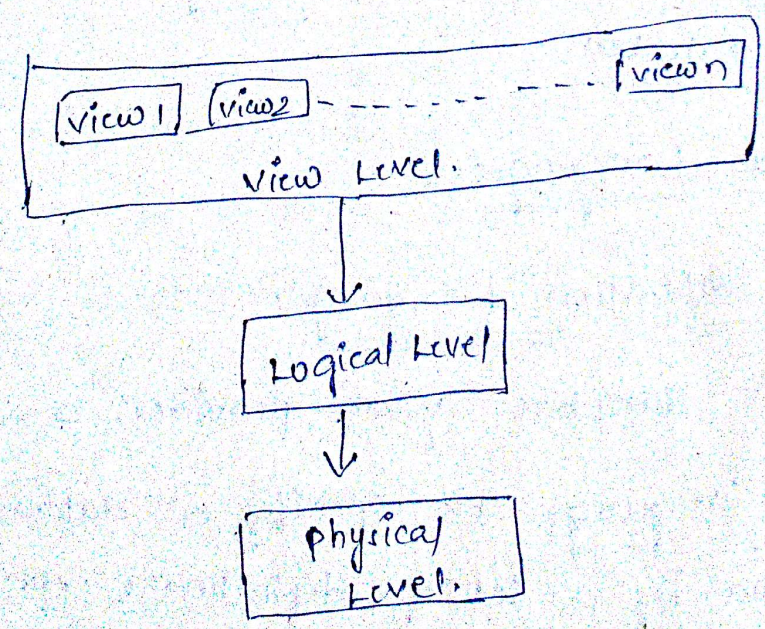
To rectify the problems dbms is used in atomicity

vi) concurrent access anomyles :-

anomoly is a problem in a database while insertion, deletion and updation of the data.

vii) security problems :- Hiding personal information.

views of data:



→ physical level describes the data design, type at physical appearence of the data. that means it discuss about the storage and how to store the

Scanned by CamScanner

data in a hard disk.

→ Logical level the next higher level of extraction describes what data are stored in the database and what relationship exists among those data.

→ The logical level describes the entire database in terms of a small no. of relatively simple structures.

→ View level

The highest level of abstraction describes only a part of the entire database

→ Many users of the database system does not need all the information inserted into the database,

→ they needed only a part of the database. the system will provide many views for the same database.

database Languages:

data definition Language:

A database system provides a data definition language. We specify a database schema by a set of definitions expressed by a special language called "ddl" (data definition language).

Scanned by CamScanner

The "create" command is used to specify different datatypes and size of each field and constraints for the data entry.

datamanipulation language:

data manipulation is

i) the retrival of information stored in the database.

ii) the insertion of new information into the database

iii) the deletion of information from the database.

iv) the modification of information stored in the database.

Note

The combination of "DML" and "DDL" called structured Query Language ('SQL'). (oracle mysql)

Eg:

→ A DML is a language that enables users to access OR manipulate data as organized by the appropriate data models.

→ there are basically 2 types.

1. procedural DML's.
2. declarative DML's.

Scanned by CamScanner

1. Procedural DML's:

It requires a User to specify what data are needed and how to get those data.

2. Declarative DML's:

also called as non procedural DMLs. require a User to specify. what data are needed without specifying how to get those data.

Query :-

A Query is a statement that requesting the retrival of information.

Database access from application programs:

Application programs are programs that are used to interact with the database

application programs are written in a high level languages such as COBOL, C, C++, Java, PHP.

→ The open database connectivity (ODBC) standards defined by the microsoft for use with the C language.

→ The JAVA database connectivity (JDBC) standards provided by the Sun microsoft to connect with "JAVA Language".

Scanned by CamScanner

Database Users and administrators:

People who work with database can be categorized as database Users and administrators

Database Users & User interface:

There are 4 different types of database system Users differentiated by the way they expect to interact with the systems.

1. Naive Users. (or) unsophisticated users & who interact with the systems by invoking one of the application problems that have been written previously.

Eg: A bank User who needs to Transfer Rs.600/- from an account A to account B. involves a program called Transfer.

→ The typical user interface for navie users is a forms interface, where the user can fill in appropriate fields of the forms.

Data storage and Query:

The functional components of a database system can be broadly divided into two categories

1. storage manager and 2. the Query processor components.

Scanned by CamScanner

→ The storage manager is important because database requires a large amount of storage space.

→ It is implemented that to define data structure to store the data and speed access of the data.

→ and it specifies how to move the data between disc and main memory.

i) Storage management :

→ A storage manager is a programme module that provide the interface between the low level data stored in the database and the application programs.

→ storage manager is responsible for interaction between file manager.

→ the storage manager transfers various DML statements into low level file system commands

→ the storage manager is responsible for storing, retrieving, updating data in the database.

ii) storage manager components :

1. authentication and integrity manager :

which tests the satisfaction of integrity constraints and checks authorization of users to access data

Scanned by CamScanner

Transaction manager :-

Which ensures that the data base remains in a consistent state, system failure and the component transaction, executes without any conflicting.

File manager :-

Which manages the allocation of space on disc storage and data structures used to represent information stored on disc.

Buffer manager :

Which is responsible for fetching data from disc storage into main memory and deciding what data is to catch in main memory.

→ The storage manager implements several data structures as part of the physical system implementation

i) data files. :-

Which stores the database itself.

ii) data dictionary :

Which gives the information about the data type and length of the data item.

iii) indices :-

Which provide fast access of data items.

Scanned by CamScanner

Query processor :-

The Query processor components includes

i) DDL Interpreter which represents DDL statements and records in the data dictionary.

ii) DML compiler which translates DML statements in a Query language into an evolution plan.

iii) it also perform Query optimization- i.e it picks the low cost evolution plan from the alternatives.

iv) Query Evolution engine :-

Which executes low level instructions generated by DML compiler.

Transaction management :- (unit-V)

Database Users and Administrators :-

People who work with a database can be categorized as : DB Users and DB administrators.

ii) Application (Users) programmers :- are computer Professionals who write application programs.

Application programmers can choose from many tools to develop user interface.

iii) Rapid Application development (RAD) :- tools are those that enable an application programmer to construct forms and reports without writing a Program.

Scanned by CamScanner

iv) Sophisticated Users:

interact with the system without writing programs, instead, they form their request in DB query language. they submit such query to a query processor whose function is to breakdown DML statements into instructions that the storage Manager Understands.

v) Specialized Users:

Specialized Users who write specialised database applications, that do not fit into the transactional data processing team work.

Database administrator:

One of the main reiponse for using database is to have control of both the data and programs that access those data. A person who has such control over the system is called a Database administrator (DBA). The functions of DBA include Schema definition.

=

DBA creates Original DB schema by executing a set of data definition statement in the DDL. Storage structure and access method definition.

Schema and physical organization Modifications:

The DBA carries out changes to the schema and physical organization to reflect the change needs of the organization or alter the physical

Scanned by CamScanner

organisation to improve performance.

Granting of authorization for data Access:

By generating different types of authorizations for the DB. The DB administrator can regulate which parts of the DB can be accessed by various users.

Routine Maintenance:

The routine maintainance activities are

i) periodically backup the DB.

ii) upgrading the space as required.

iii) Maintaining jobs running on the DB and enhancing the performance.

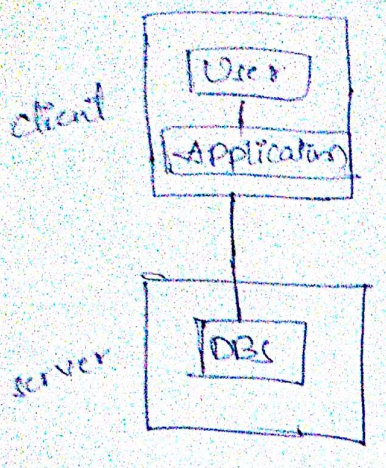
If it is not required task (or) authenticated task submitted by user then it is resubmitted to the same user.

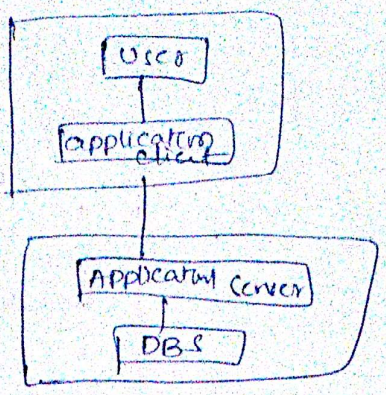
DB Architecture:

We are now in a fascination to provide a single picture of the various components of the DB systems and the connections among them. There are 2 types of DB architectures.

i) 2 tier architecture ii) 3-tier architecture.

Scanned by CamScanner



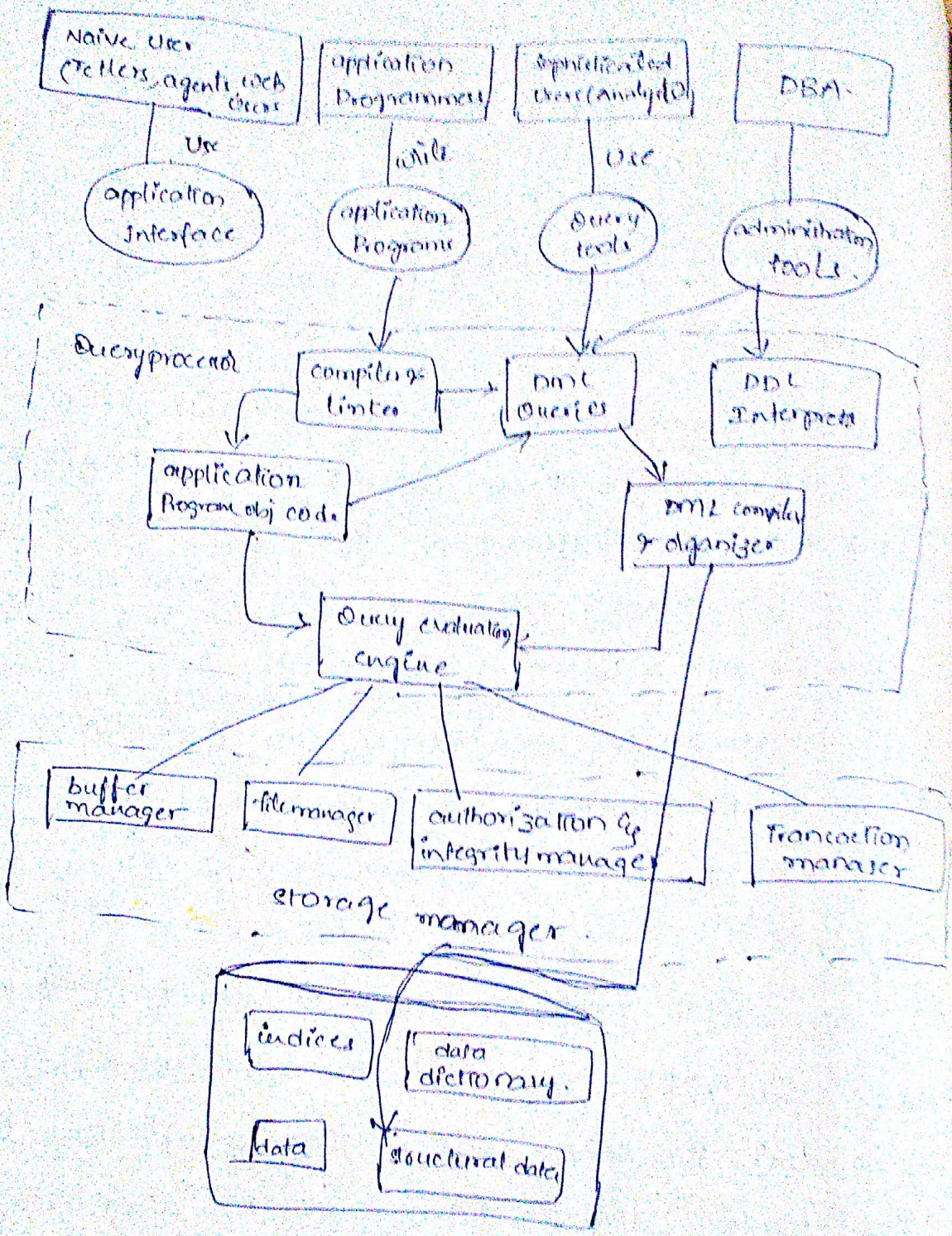


Most users of DBs today are not present at the sight of the DBs but connect to it through a network. We can differentiate blw client machines on which remote DB users work and server machines on which the DBs runs.

In a 2-tier architecture the application resides at the client machine where it invokes DBs functions at the server machine through Query lang. statements.

Application program interface standards like ODBC and JDBC are used for interaction blw client and the server. In a 3-tier architecture the client machine acts as front end and does not contain any direct DB calls. Instead the client end communicates with an application server usually through a form-interface.

Scanned by CamScanner



Speciality databases:

There are so many evolutions are done in application programe writing style. 90% of the data base users are webUsers. In Olden days data base and application programs are resides in the same system. so that in those days relational database model is sufficient to develop a

Scanned by CamScanner

database system in an organization.

There are 2 types of special databases are introduced in now-a-days.

1. object oriented database model (object based data model).

2. semi structured data model.

Object based data model:

Object oriented has become the dominate software development methodology. So that in E-R model we use the notations of encapsulation, methods, inheritance and object identity.

In E-R model (entity relationship model) when we use above object oriented features then it is termed as Enhanced E-R model (EER model).

Major database vendors presently support the object relational data model.

The object relational database model contains object oriented database model features and relational database model features.

Semi structured data model:-

Semistructured data models permit the specification of data where individual dataitems of the same type may have different sets of attributes.

The XML (Extensible mark up language) was initially designed as a way of adding markup information to text document.

Scanned by CamScanner

different ways of executing Queries on data stored in XML and transforming XML data from one form to another.

Introduction to relational model:

The relational model is today the primary data model for commercial data processing applications. In this chapter we study functions of the relational model that helps in the design of relational database schemas.

structure of relational databases

A relational database consists of a collection of tables each which is assigned a unique name for eg: staff table in a college database. the table look like.

ID	Designation	name	Dept name	sal.

→ In general arrow in a table represent a relationship among a set of values.

→ Since a table is a collection of such relationships
→ table and the mathematical concept of relation forms which the relational data model takes

Scanned by CamScanner

its name.

→ In the relational model the terms relation is used to refer to a table, tupple is used to refer to a row., attribute is used to refer a column. and relation instance is used to refer to a specific instance of a relation that contains specific set of rows.

Domain:

For each attribute of relation there is a set of permitted values called the Domain.

Atomic domain:

A domain is atomic if elements of the domain are considered to be indivisible units.

The NULL:

The NULL value is a special value that signifi-cant that the value is unknown (or) does not exists.

NOTE:

NULL in database is not equal to zero (0).
zero is itself a value.

Database schema:-

database schema is a logical design of the data base.

→ the database instance which is a snapshot of data in the database at a given instant in time.

Scanned by CamScanner

→ Design the college information System table (atl. for our PVPSIT.

Keys:

A key is nothing but specialised column in a tab. There are several types of keys exists in a table based on their requirements.

1. Super key :-

A super key is a set of one (or) more attributes. that taken collectively, ~~identify~~ allowed to identify uniquely a tupple in a relation.

courseId	coursename	Branchname	Intake	year of comm.
B105	Btech	IT	120	1998
B106	Btech	CSE	120	1998
B107	Btech	EEE	60	1999
B108	Btech	ECE	190	1998
B109	Btech	ME	60	2000
m105	mtech	IT	30	2000
m106	mtech	CSE	18	2004
m107	mtech	ME	18	2005
P101	MBA	MBA	60	2008

$SP_1 \longrightarrow$ CID

$SP_2 \longrightarrow$ course name & Branch name

$SP_3 \longrightarrow$ Branch name & year of commence

$SP_4 \longrightarrow$ Course ID, Course name

&

Scanned by CamScanner

SP5 → CID & Bname.

SP6 → CID & Intake.

SP7 → CID & Year of commence

SP8 → Bname, intake, coursename, year of commence.

Candidate key:-

The minimal subset of a super keys is called a candidate key.

Primary key:

Among all candidate keys database deigner can choose one candidate key as primary key

for eg:- In above table there are so many no. of super keys are present if their is a courseId column. if their is no courseid column then

SP_1 – coursename & branchname.

SP_2 – B_name & year of commence

SP_3 – B-name & intake.

SP4 – ~~Corse Id~~ ~~coursename~~, Branchname, intake, year of commence

SP5 – B-name & year of commence, coursename.

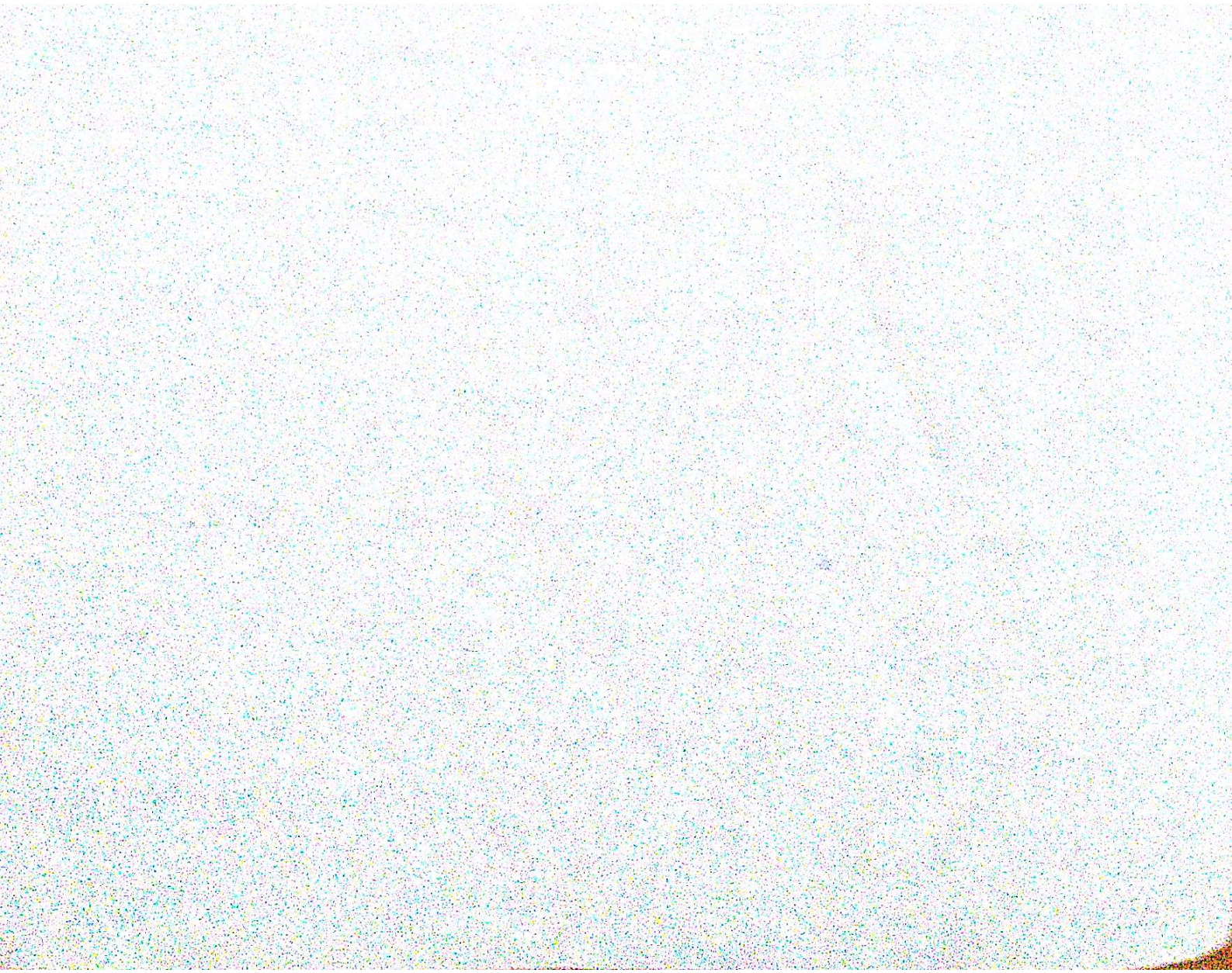
SP6 – course name & Branch name, Intake

SP7 – Branch name, Intake, year of commence.

SP8 – Coursename, Branchname, Intake, year of commence

In above Super keys SP1, SP2, SP3 contains minimal no. of attributes i.e 2. so that SP1, SP_2 and SP_3 are the candidate keys of the given table (if their is no courseId in the table).

Scanned by CamScanner



Scanned by CamScanner

Based on the situations demand in the database. ~~designer~~ the database designer choose one of the candidates key as primary key.

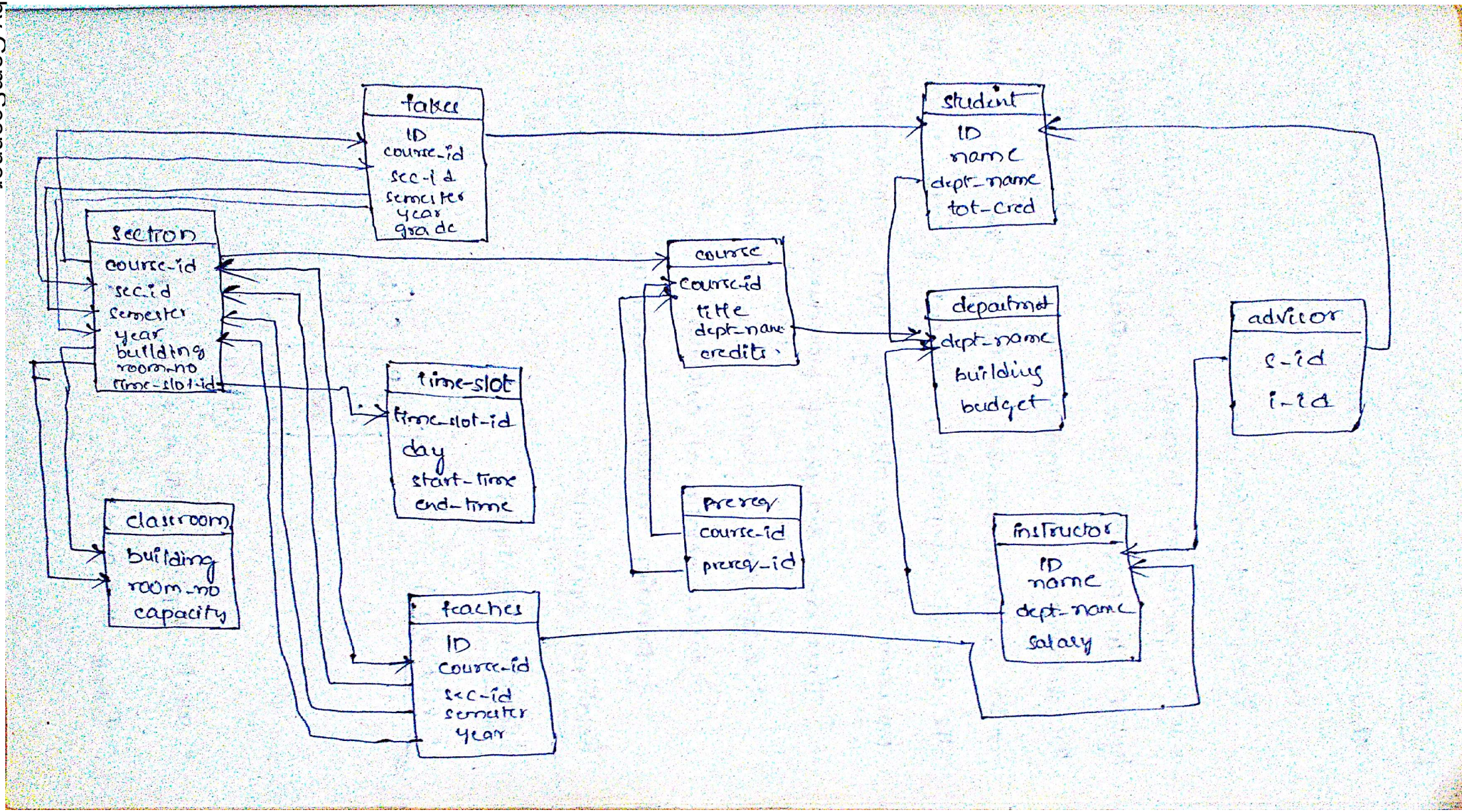
→ In above 3 candidate keys we are choosing SPI as a primary key.

Schema diagram:

→ In schema diagram each relation appears as a box with attribute listed inside and the ~~under~~ relation name above it. if there are primary key attribute a horizontal line is drawn under the primary key foreign key dependences appear as arrows from the foreign key attributes of the referencing relation to the primary key of the referenced relation.

Scanned by CamScanner

Scanned by CamScanner



Relational Query languages:

There are 2 types of Relational Query language

1. procedural
2. nonprocedural (or) declarative.

→ In a procedural language the user instructs the system to perform a sequential of operations on the database to compute the desired result.

→ In a nonprocedural language the user describes the desired information without giving a specific procedure for obtaining that information.

There are 3 types of procedural language

1. Relational algebra.
2. tuple relational calculus.
3. Domain relational calculus.

1st formal Procedural language and remaining 2 are non Procedural languages.

Relational operations:-

Selection of tuples:-

- relation(r.)

A	B	C	D
α	α	1	7
β	β	5	7
B	B	12	3
B	B	23	10

- select tuples with A = B and D > 5
- $\sigma_{A=B \text{ and } D>5}(r)$

Scanned by CamScanner

A	B	C	D
α	α	1	7
β	β	23	10

Selection of columns (attributes):-

- relation r.

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

- select A and C
  - projection
  - $\Pi_{A,C}(r)$

A	C
α	1
α	1
β	1
β	2

=

A	C
α	1
β	1
β	2

Joining two relations – cartesian product:

- relation r, s:

r

A	B
α	1
β	2

s

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

r x s

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

A = {1, 2, 3}
B = {1, 2, 3, 4}
Ans-

Scanned by CamScanner

Union of two relations:-

- Relations r, s

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

r ∪ s

A	B
α	1
α	2
β	1
β	3

q = {a, b, c}
z = {b, c, d}
q ∪ z = {a, b, c, d}

Set difference of two relations:

Relations r, s:

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

→ r - s:

A	B
α	1
β	1

s - r

A	B
β	3

A = {1, 2, 3}
B = {2, 3}
A - B = {1}

Set intersection of two relations:-

Relation r, s:

Scanned by CamScanner

A	B
α	1
α	2
β	1

A	B
α	2
β	3

r ⋈ s

A	B
α	2

Joining two relations - Natural join :

→ Let r and s be relations on schemas R and S respectively. Then, the "natural join" of relations R and S is a relation on schema R ∪ S obtained as follows:

- Consider each pair of tuples t_r from r and t_s from s.
- If t_r and t_s have the same value on each of the attributes in R ∩ S, add a tuple t to the result, where :
  - → t has the same value as t_r on r.
  - → t has the same value as t_s on s.

Natural join Example :

Relations r, s :

A	B	C	D
α	1	α	a
α	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

r

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	ε

s

Scanned by CamScanner

Natural join

r ⋈ s

A	B	C	D	E
α	1	α	a	α
α	1	α	a	γ
α	1	γ	a	α
α	1	γ	a	β
β	2	β	b	β

Synopsis:

Symbol (Name)	Example of use
σ (selection)	$\sigma_{salary >= 85000}(instructor)$ return rows of the input relation that satisfy the predicate.
Π (projection)	$\Pi_{ID, salary}(instructor)$ output specified attributes from all rows of the input relation remove duplicate tuples from the output.
⋈ (natural join)	instructor ⋈ department output pairs of rows from the two input relations that have the same value on all attributes that have the same name.
× (cartesian product)	instructor × department output of all pairs of rows from the two input relations (regardless of whether or not they have the same value on common attribute

Scanned by CamScanner

∪ (union)

- Π_{name} (instructor) ∪ Π_{name} (student)
- output the union of tuples from the two input relations.

Scanned by CamScanner