"database"

* database definition :- a collection of data (tables) that stores organized information. most of database contain ~~contain~~ of multiple tables. each of which include several different Fields.

* data base classification.
DB classified into different types according to organized approach.
"organizational approach".

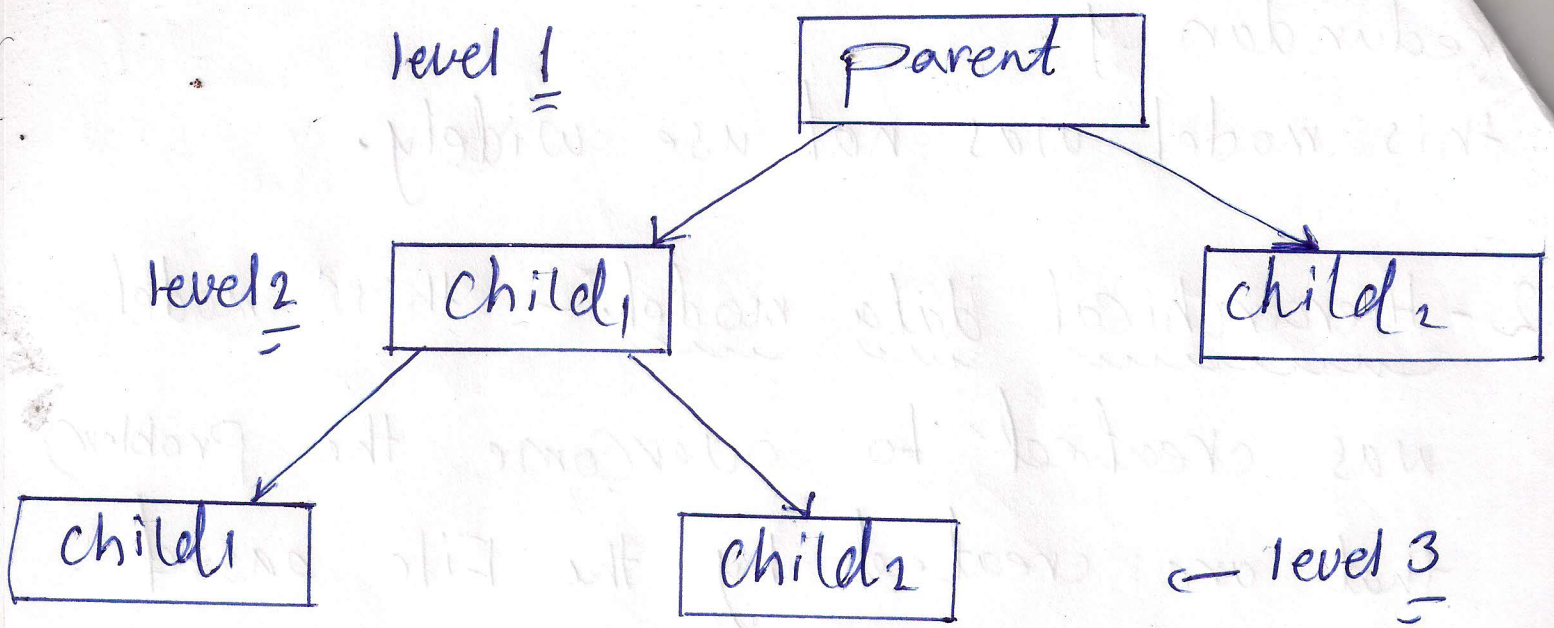
1- Flat File based data model.
2- Hierarchical data model.
3- Network data model.

1- Flat File based data model :- in this DB model data are stored in numerous Files. and Files are not linked So, data are repeated in more than one file.

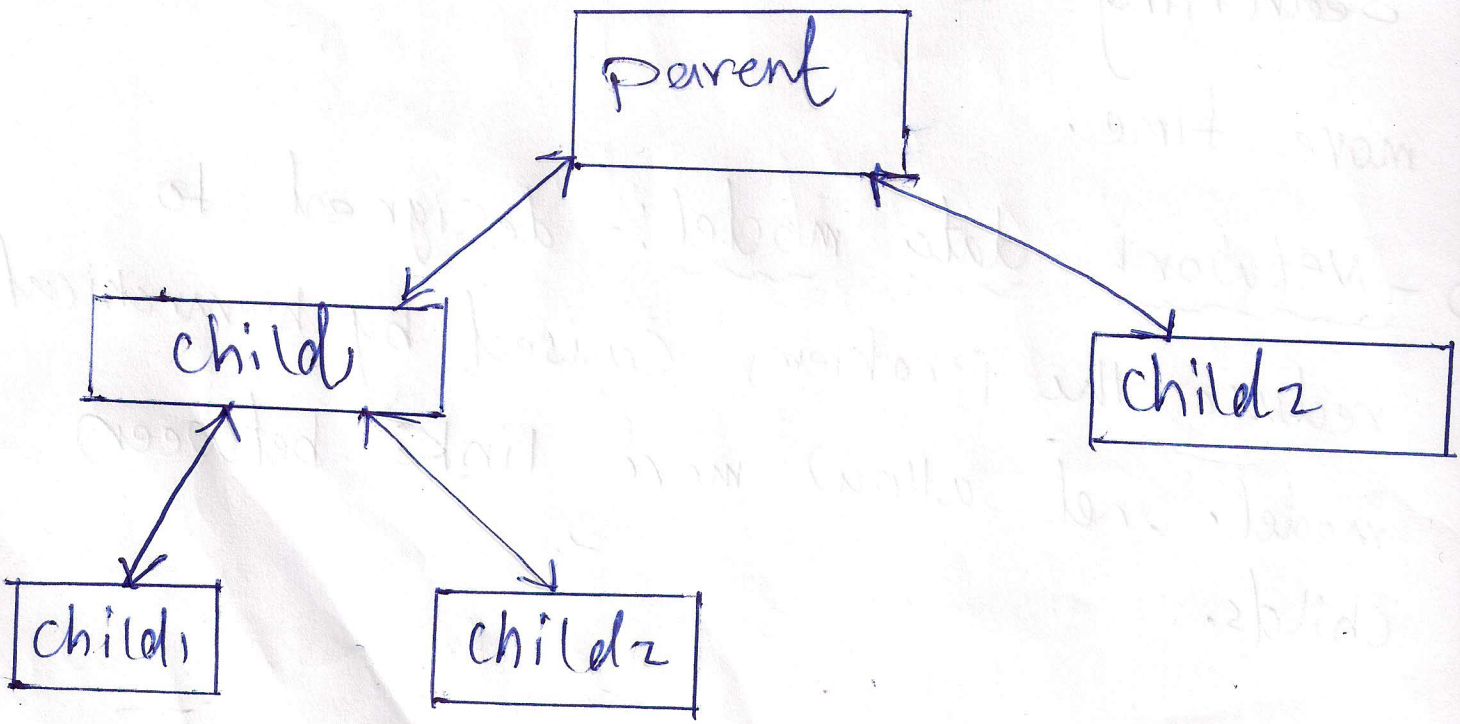
ore problem in this model is redundancy.
this model was not use widely.

2- Hierarchical data model:- this model was created to overcome the problem that are created by the File based data model.
the main problem in this model is searching data is extremely difficult. Searching data in this model need more time.

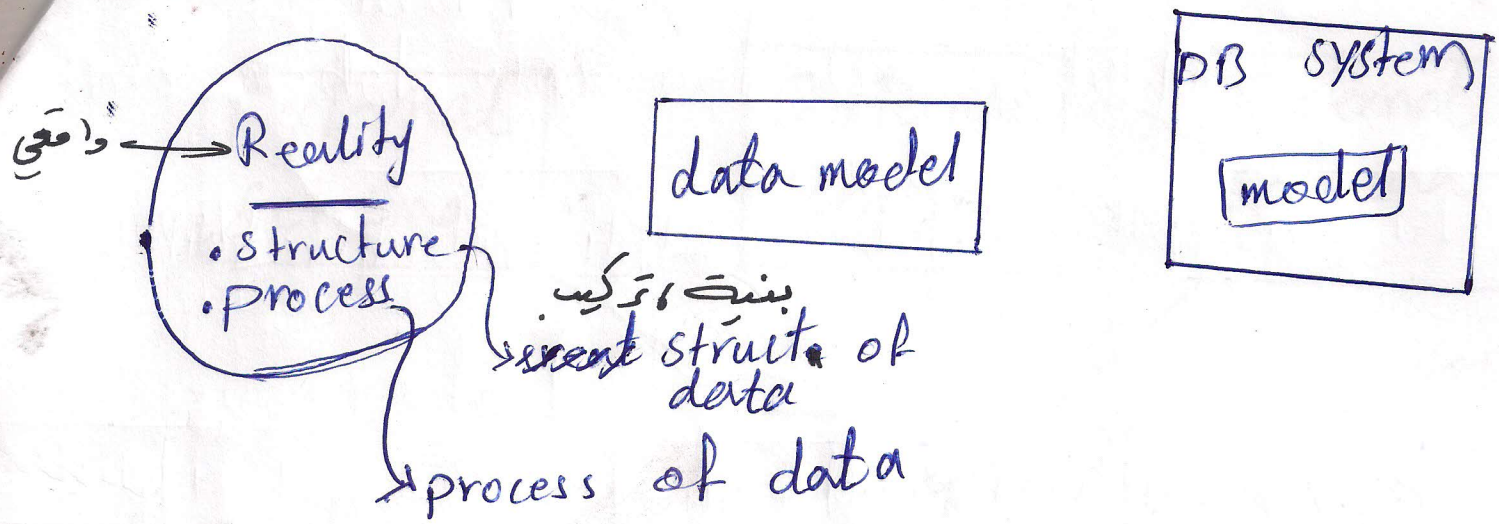
3- Network data model:- designed to reduce the problem caused by hierarchical model. and allow more links between childs.



" hierarchical "



"DB modeling"



---

* why use model?

model is a meaning of communication

user of DB must have a certain amount

of knowledge.

db design.

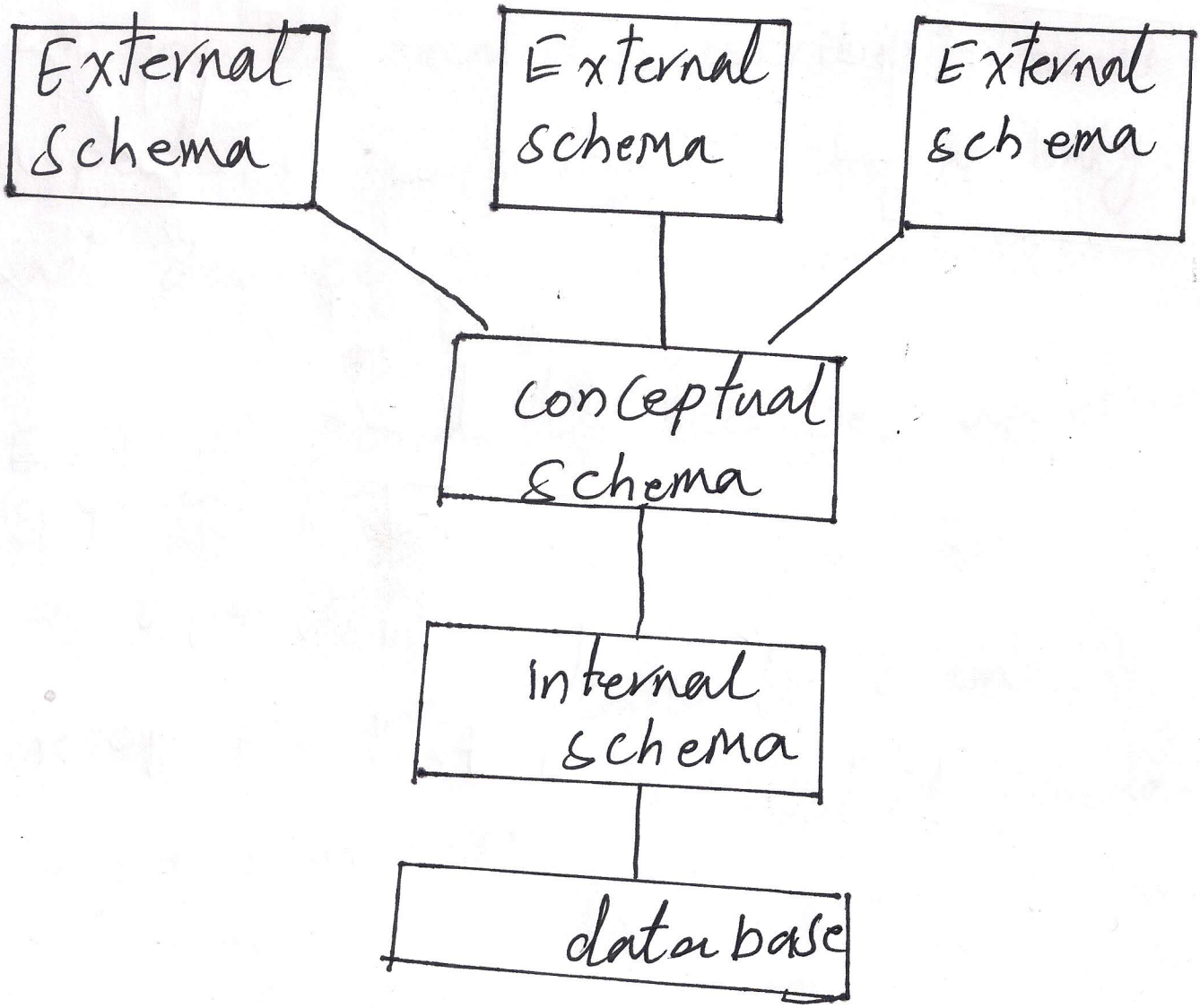
- the purpose of DB design is to

create data base which is:

→ is a model of structure of reality.

→ run efficiently.

DB Architecture"



خيالي، وهمي، مفاهيمي

→ Conceptual schema: is ameaning of data. (مشروع، مخطط)

→ external schema: is a use of data.

→ internal schema: Storage of data.

- Conceptual Schema:

an object oriented describe Process aspects in Conceptual Schema only. anel describe the data structures.

2- External Schema:- describes aparticular part of information that are described in Conceptual schema. it's derived from Conceptual Schema.

3- internal Schema:-

describe how the information described in conceptual schema, physically represented to provide best performance.

→ File system :-

the file system is typically described as various files and numbers of different application programs are written to extract records from and add records to the appropriate files.

→ File system disadvantages :-

1- data redundancy and inconsistency existed.

2- Difficulty in accessing data.

3- data Isolation.

4- Security problems existed.

5- Integrity problems existed.

data base advantages!-

1- Reduction in data redundancy.

2- the ability to operate on different data structure.

3- Independent of data From the program.

4- minimum Cost.

5- In Consistent Can be avoided.

6- integrity Can be maintained.

7- Security restriction Can be applied.

ational model:-

is Important model which is represent data relationships among data by a collection of tables each of them has number of columns and rows with unique table names.

- Columns represent the fields or the attributes in the table.
- rows represent the records or entity in the tables.

Relational DB

المحاضرة الثانية

"data stored in a table that are associated by the shared attributes.

→ any data element (entity) can be found in the relational data base through the name of tables, attributes name, and the value of Primary key.

→ Entity :- object, Subject, event.

→ attribute :- characteristics of an entity.

→ Row or Record :- Specific characteristics of an entity. (collection of fields)

→ Tables :- a Collection of records.

→ D.B :- a Collection of Tables.

"DB table key"

Keys:- are very important part in the relational data base, they are used to establish, and identify relation between tables.

there are two types of keys.

1- Super key

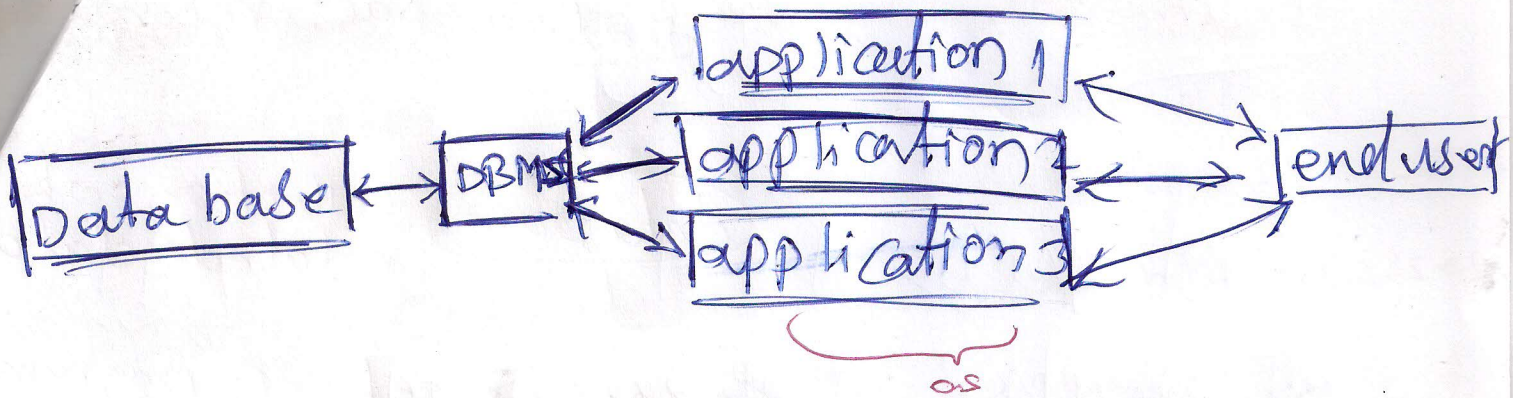
is defined as aset of attributes with in the table that uniquely identifies each record with in atable super key is asuperset of Candidate key.

2- Candidate key:- Candidate keys are defined as aset of fields from which primary key can be selected.

it is an attribute or set of attributes that can act as a primary key for atable to uniquely identify each record in that table.

- Primary key :- is a candidate key that is most appropriate to become main key of the table, it is the key that uniquely identify each record in a table. relation DB must have one and only one of Primary key.

→ Foreign key :- a columns or a set of columns within the table that are required to match those primary key of second table.



DBMS

* advantages of DBMS :-

1- DB development :- it allows organizations to place control of the DB development in the hand of DB administrators (DBAs) and other specialists.

2- data independance :- application programs should be as independent as possible from details of data representation and storage. the DBMS can provide an abstract view of the data.

3- Data integrity and security.

4- Data administration and concurrent access. →

when Several users share the data (more than one user access the database at the same time).
DBMS Schedules (جدول) Concurrent allesses to the data. in such manner that users Can thinks of the data as being accessed by only one user at atime.

* RDBMS :- Relational database mangement system, in which different tables are related to each other by common fields, so that information From several table Can be (وصل) combined.

* what are queries?
once a data base has been created and data has been entered into it, we want to be able to get this information back (طلع) out again.
query is the question [Comand] that we send to the database in order to Select and view

DBMS

DBMS:- is a set of computer programs that controlls the -

1- Creation of the data base
2- the storing and organization of the data in the databases.
3- maintenance the database.
4- searching and data retrieval of a database.

the DBMS accepts requests for data From an application program and instructs the operating system to transfer the appropriate data.

as shown in the Following Figure.

~~figure~~.

* E/R diagram.

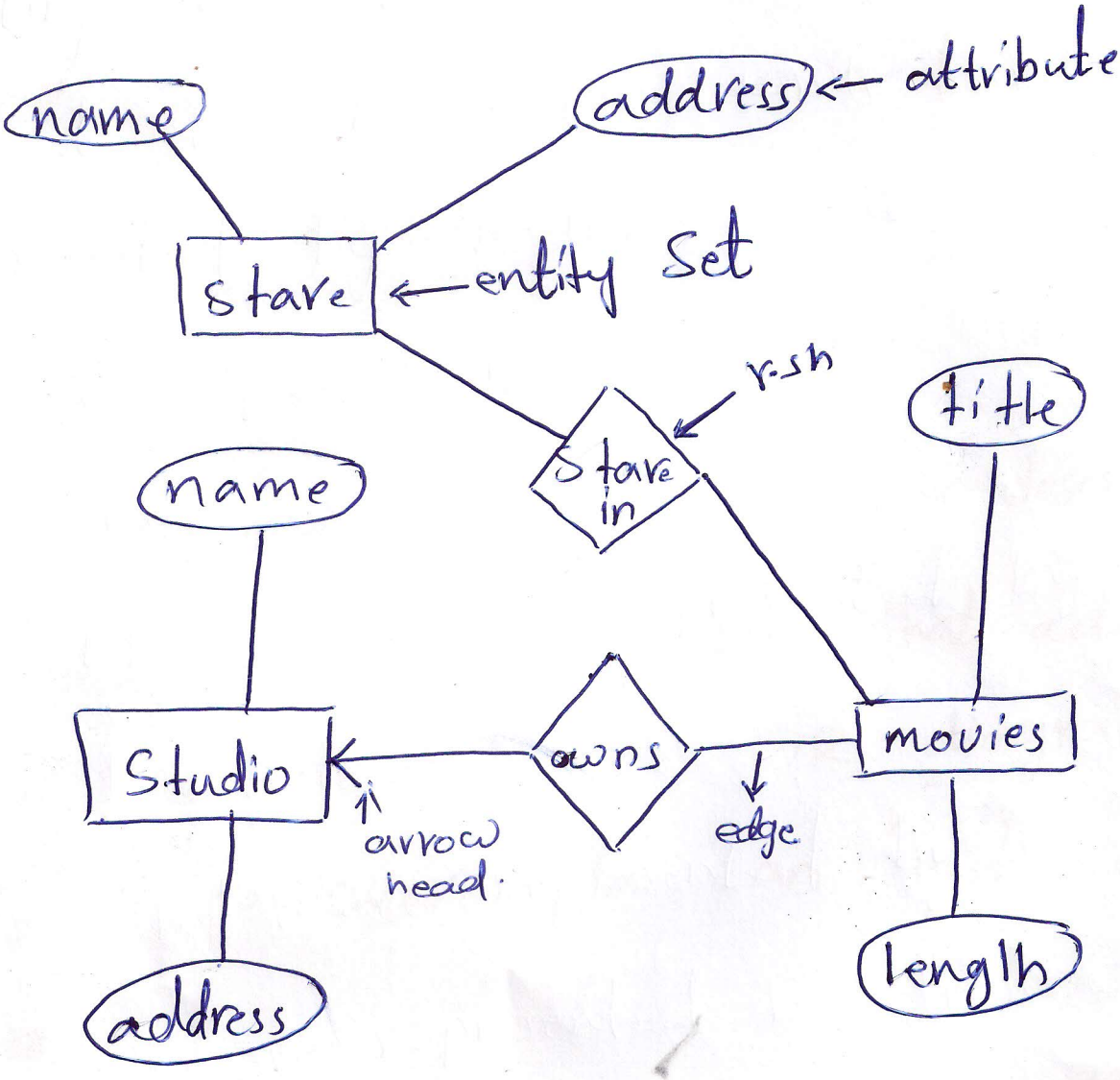
E/R diagram is a graph representing entity sets, attributes, relation ships. element of each of these kinds are represented by node of the graph. we used special shapes to indicate the kinds as follows.

1- entity sets are represented by rectangle.

2- attributes are represented by vals.

3- r.sh. are represented by diamond.

E/R diagram representing simple data base about movies, the entity set are movies stars and studio.



Stare in is represented as r.sh.

Connect each movie to the stare of that movie.

, Connect each movie to the Studio

* the arrow head pointing to entity Set Studio.

indicate each movie may owned by more than one Studio.

---

* instance For E/R diagram.

E/R diagram are notation for describing schema in db we imagine that a data base described by an E/R diagram Contain a particular data, an "instance" of db.

→ For each entity set the instance may Contain a Finite set of entities.

ultiplicity of Binary r.sh-

in general a binary r.sh can connect any member of one of its entity set to any member of another entity set.

→ Suppose R is relationship connecting entity set E and F then.

→ if each Member of E can connected by R to at most one member of F then we say that R is a many-one r.sh. From E to F.

i.e " each entity set in F can be connected to many member of E ".

instead each member of F can be connected by R to at most one member of E then we say that R is many-one from F- to - E.

→ if R both many-one from E to F and many to one from F to E then we say that R is one-one.

(" an entity at either or entity set can be connected to at most one of the other."

→ if R is neither many-one from E to F nor from F to E then we say that R is many--many.

many to many.

* videos — customers

* Students — subjects

* one to one.

husband — wife

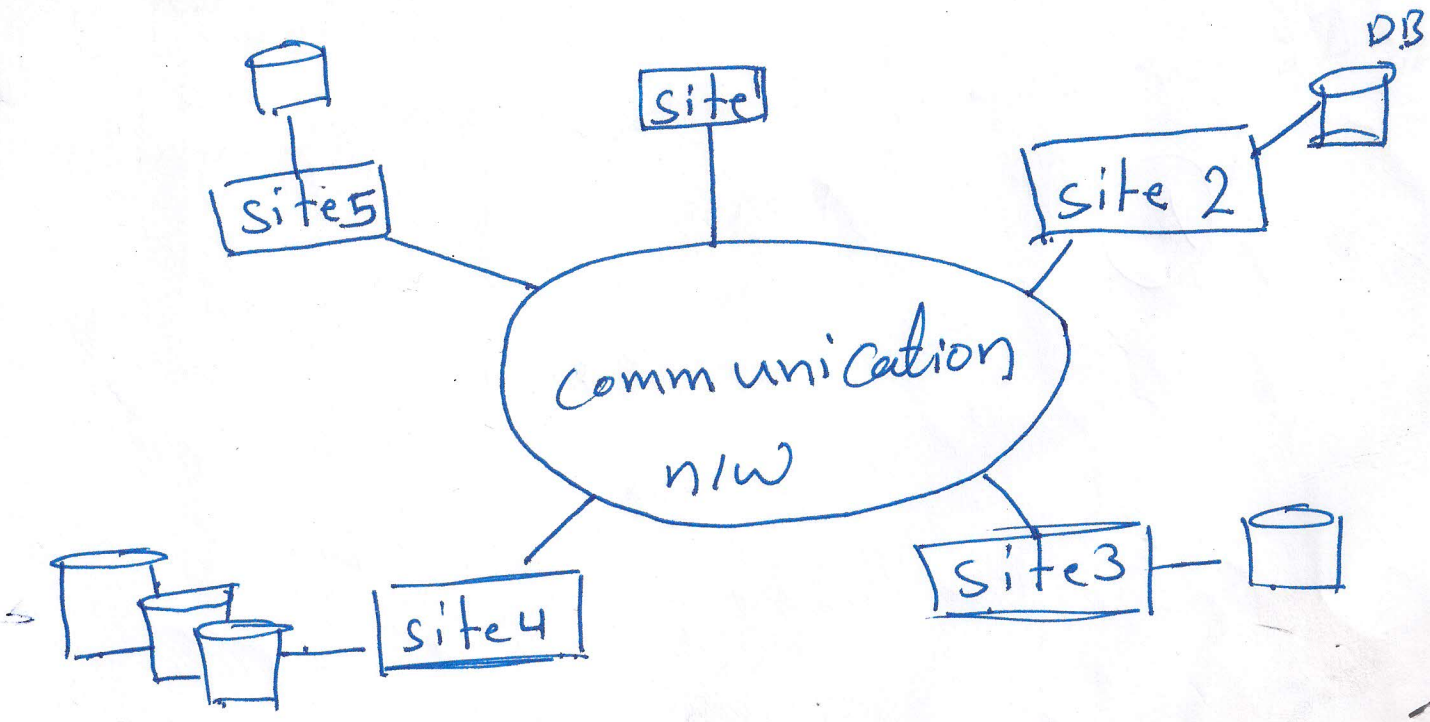
* many-one.

mother — children

"distributed data base" - que 1,3,1

*DDB:- is a collection of multiple, logically interrelated data bases distributed throug Computer n/w.

*DDBMS:- is the s/w that manages the DDB and provides an access mechanism that makes this distribution transparent to the users.



n concept of DDB

→ data stored at the several location.

→ each manages by DBMS can run autonomously.

*distributed system:- a collection of computers that appears to it's users as a single coherent system.

*distributed database types:-

1- homogeneous :- every site run at the same DBMS.

2- hetrogeneous :- different sites run at the different DBMS.

the Following conditions must be satisfy in homogeneous system.

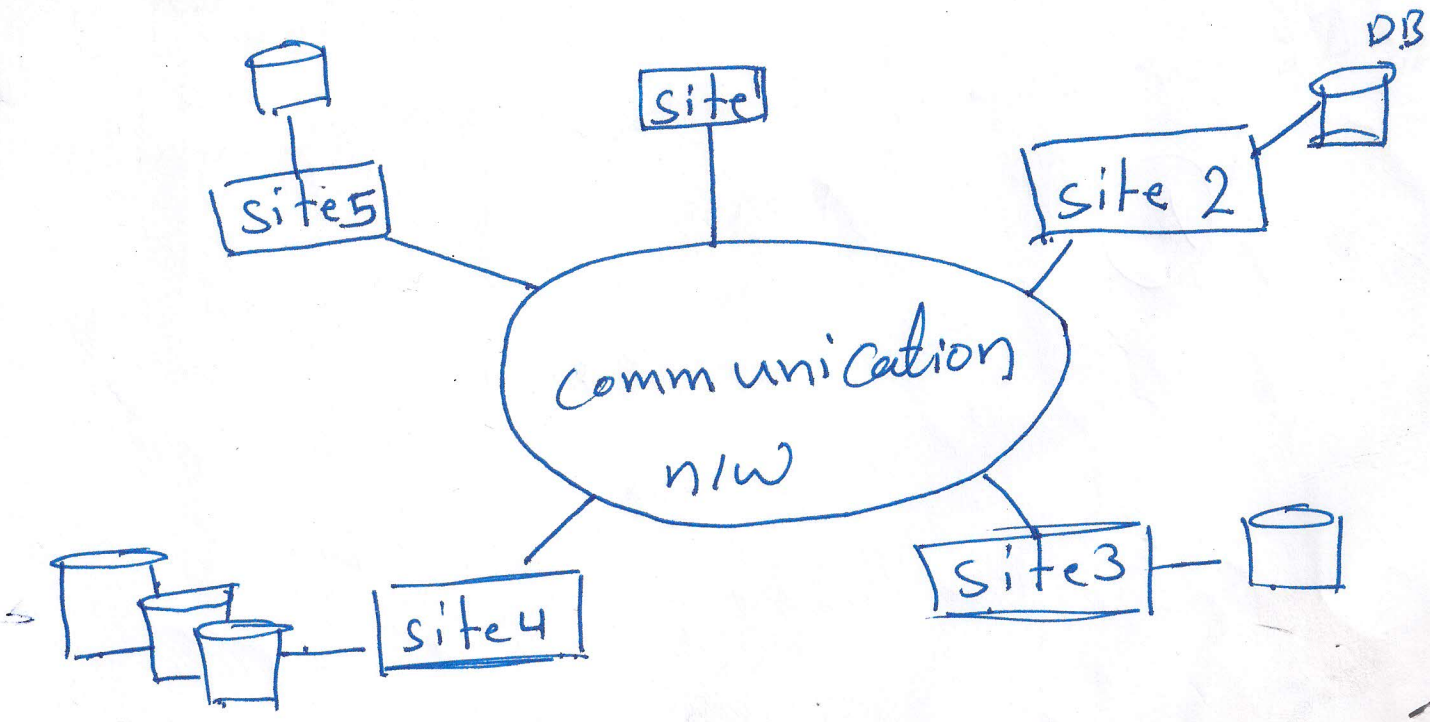
1- operating system useed in every location must be same.

2- data Structure and DB application must be same

"distributed data base" - que 1,3,1

*DDB:- is a collection of multiple, logically interrelated data bases distributed throug Computer n/w.

*DDBMS:- is the s/w that manages the DDB and provides an access mechanism that makes this distribution transparent to the users.



n concept of DDB

→ data stored at the several location.

→ each manages by DBMS can run autonomously.

*distributed system:- a collection of computers that appears to it's users as a single coherent system.

*distributed database types:-

1- homogeneous :- every site run at the same DBMS.

2- hetrogeneous :- different sites run at the different DBMS.

the Following conditions must be satisfy in homogeneous system.

1- operating system useed in every location must be same.

2- data Structure and DB application must be same

"distributed DB Issues"

1- data base design.

2- Concurrency control. (in access of data).

3- Reliability.

4- Query Processing.

1- db design: how to distribute the data over n/w. (replicate, non replicate)

2- Isolate of transaction Effect.

3- reliability:- how to make system more (flexible) to the Fail Failures.

4- Query Processing:- Convert the user transaction to manipulating instruction.

dvantages of DDB.

1- management of DDB with the different level of transparancy, like n/w transparancy & fragmentation transparancy and replication transparancy.

2- increase reliability and availability.

3- easier expansion.

4- improved performance.

* disadvantages.

1- complexity.

2- Security. (remote sites must be secure).

3- difficult to maintain integrity.

4- analysis.

distributed database design:-

1- data Fragmentation:- how to Fragment data, (partition data) into Fragments.

2- data replication: which Fragment to replicate.

3- data allocation:- where to locate this Fragment.

(1) data Fragmentation:-

- break the object into two or more Fragment.
- each Fragment can be Store in any site over n/w.
- the object might be a user's database table, or system data base.

data Fragmentation strategies:-

There are three types of fragmentation strat:-

1- horizontal 2- vertical 3- mixed.

1- horizontal Fragmentation:- refer to the division of a relation into subset (Fragments) of tuples (rows). each Fragment is stored at the different node.

2- vertical fragmentation:- refer to the division of arelation into attributes (column). each subset (fragment) is stored at ~~the~~ different node.

3- mixed Frag:- refer to combination of horizontal and vertical strategies. in other word, atable may be division into several horizontal (rows, each one having asubset of the attributes.

* distributed db design :-

(2) Data Replication :- data copies stored at multiple sites served by computer n/w.
(Fully, No Replicating, partial).

(3) Data allocation :- deciding where to locate data. (is a process of assigning a particular Fragment in a particular site. in d.s.).

there are 3 types of data allocation.

1- Centeralized data allocation.
entire data base stored in one site only.

2- Partition data allocation :-
data base divided into several Fragments and stored in any site.

3- replicate data allocation.
copies of one or more DB are stored in any site over computer network.

note: data allocation also called data distribution.

ata Replication types.

1- Fully replicated db.
→ replicating the whole db at every site.

2- NO-Replicated db.
→ each Fragment is stored exactly at one site.
→ All Fragments must be disjoint except primary key.

3- partial Replication.
→ Some Fragment may be replicated while others may not.

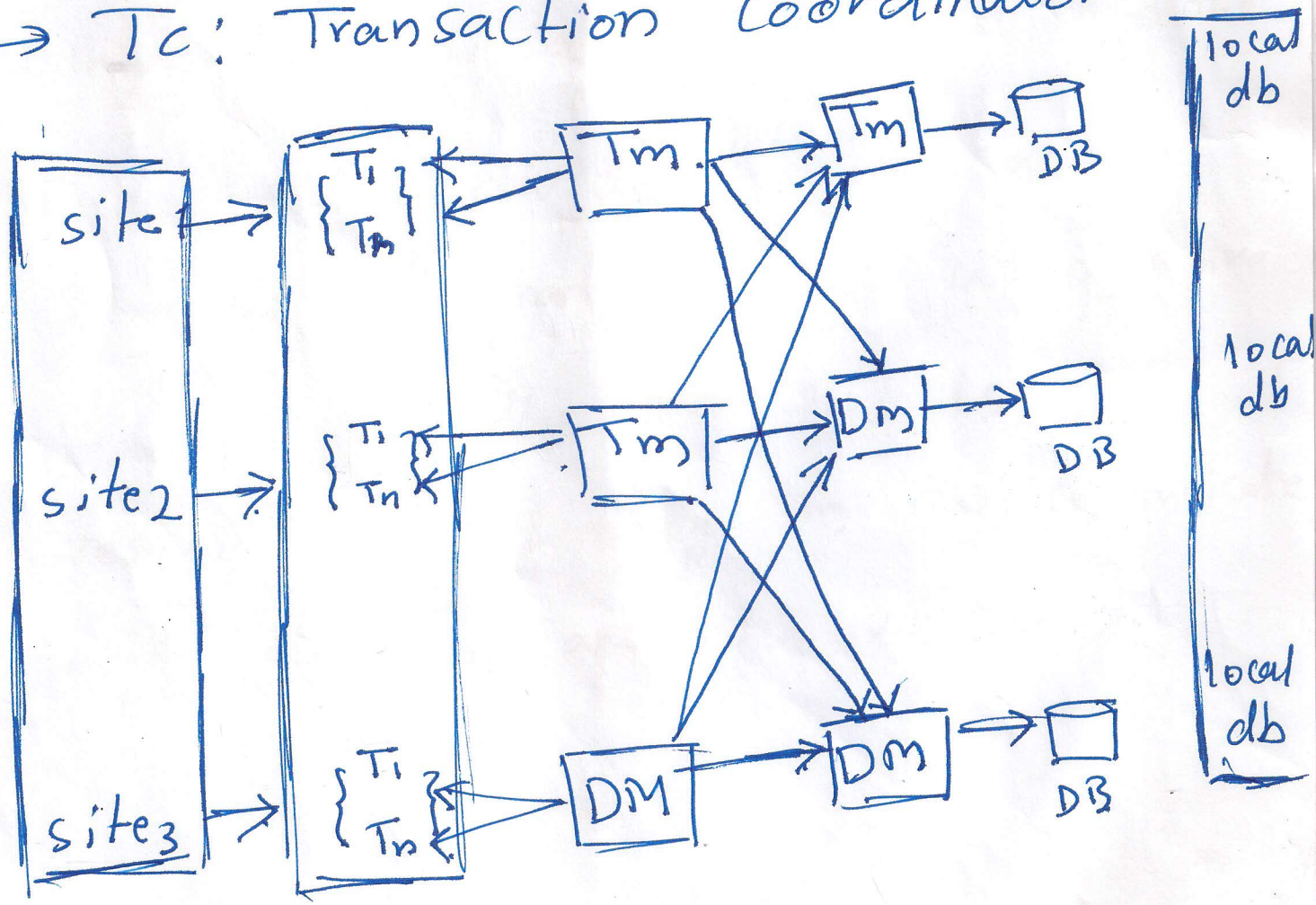
'architecture of DDBMS'

→ transaction'.- is a sequence of actions (read, write) or programs must be executed completely or leave the db in consistent state. (not executed).

→ TM: transaction manager.

→ DM: data manager.

→ Tc: Transaction coordinator.



{ DDBMS architecture. }

each site (computer) may contain (Tms) transaction manager and data manager (DM) and transaction coordinator (TC). the transaction manager responsible for transaction that are received by computer, the DM is manages the data base access on the local computers. when the transaction arrives into the TM the Tm divides the transaction into the sub transactions which are transmite to the DM. this is TC responsibility. any Tm can communicate with all DMs and vice versa.

distributed schema

→ can be classified into :-

1- distributed external schema.
2- distributed conceptual schema
3- distributed description.
4- local external schema
5- local conceptual schema.
6- local internal schema.

DBMS Failure & Commit

a data base management system is susceptible to anumber of failures. in d db system, failure can be classified into soft, Hard, and network Failures.

1- Soft Failure.

is the type of Failures that causes the loss in volatile memory of the computer and not in the persistent memory. the various types of soft Failures are as:

1- operating system Failures.
2- main memory failure.
3- transaction Failure.
4- power Failure.

- Hard Failure:
is the type of Failure that causes loss of data in the non-volatile memory.

3- Network Failure.
are prevalent (common, general) in distributed data base. these Comprises of the errors induced in the data base system due to the distributed the data over n/w.

types of its.
1- Communication link Failure.
2- n/w congestion.
3- information Corruption during transfer or losting
4- Site Failure.
5- n/w partitioning.

Commit protocols:

any data base should guarantee that the desirable properties of a transaction are maintained even after Failures. if Failure occurs during the execution of the transaction, it may happen that all the changes brought about by the transaction are not committed. this make the db inconsistent.

commit protocols prevent this by using either transaction undo (roll back) or transaction redo (roll Forward).

* transaction undo:

the process of undoing all the changes made to a data base by a transaction. this is mostly applied in case of soft Failure.

Transaction Redo!.. إعادة تطبيق

the process of reapplying all the changes made to a transaction. this is mostly apply on hard Failure.

Commit protocols:

the different distributed commit protocols are:

1- one phase commit.
2- two phase commit.
3- three phase commit.

---

* distributed one phase commit:

is the simplest commit protocol. let us consider that there is controlling site and a number of slave sites. when the transaction is being executed. the steps in distributed commit are.

1- after each slave has locally completed its transaction, it sends a "DONE" message to the controlling site.

the slaves wait For "commit" or "Abort" message From the controlling site. ~~this waiting time is called~~

3- when the controlling site receive the "DONE" message From each slave it makes adecision to commit or abort. then, it sends this msg to all the slaves.

4- on receiving this msg, a slave either commit or abort and then sends an acknowledgement msg to the controlling site.

distributed two-phase commit:
this is used for reduce the waiting time for sending msg between the controlling site and slaves in one-phase commit protocols.

the steps performed in this protocol as:

1- prepare phase.

- after each slave has locally completed its transaction, it sends a "DONE" message to the controlling site.
- when the controlling site has received "DONE" msg from all slaves, it sends a "prepare" msg to the slaves.
- the slave votes on whether they still want to commit or not.
- if a slave want to commit, it sends "Ready" msg.
- A slave that does not want to commit sends a "Not Ready" msg.

2- Commit/abort phase:

* after the controlling site has received "Ready" msg From all the slaves:

1- the Controlling site sends a "Global Commit" msg to the slaves.

2- the slaves apply the transaction and send a "Commit ACK" msg to the Controlling site.

3- when the Controlling site receives "Commit ACK" msg From all the Slave, it considers the transaction as committed.

*after the Controlling site has received the First "Not Ready" msg From any slave:

1- the Controlling site sends a "Global abort Commit" msg to the slaves.

e slaves abort the transaction and send "Abort Ack" msg to the controlling site.

3 - when the controlling site receives "Abort Ack" msg from all the slaves, it consider the transaction as aborted.

* distributed three-phase commit:-

the steps are.

1- prepare phase

→ same as in 2 phase.

2- prepare to commit phase.

3- Commit / Abort phase.

Same in 2 phase.

except the "commit Ack" / "Abort Ack" message is not required.