

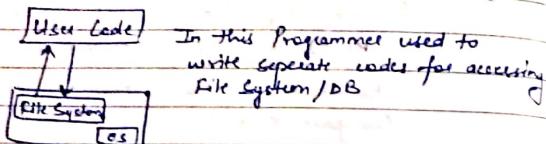
DBMS

Data Base Management Systems

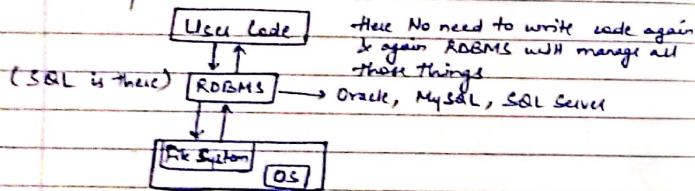
DBMS - Software System that provides you a quick way to access & modify the data

3 Types of Evolution of DBMS :-

1) File Based - Both File use write the file



2) Relational DBMS - Acts as an Interface b/w user & DB. It is a DBMS where data is stored in the form of Tables & there are multiple Table with Schema.



This method is used a lot & everywhere. But it has some problems that they require a proper structure to store the data & we need that much space & there is Scalability issues there, as we need to insert & update in the same DB for the Query so if we need to copy that DB than the write "open" becomes very costly

3) NoSQL (Non Relational DB) - Those DB which do not have proper structure in them & they could be stored like Key-value Pair

or like Objects or like Documents but they are not Replacements for SQL ones because

* It has been seen that there are consistency issues in them (Consistency in Transaction) (Data that we Read should be latest one) But these don't guarantee that. They are Horizontally Scalable.

e.g- Mongo DB, Dynamo DB (By Amazon), Cassandra etc

→ Entity Relationship Model (ER Model)

It is a way to design the Database. It is divided into 3 Parts :-

- 1) Entity Set : Student, Teacher, Course
- 2) Relationship Set : Teaches, Gives, Joins
- 3) Attributes : Upar ke dono chiso ki description like Name, Roll no. etc in case of Student or Joining date/ Duration etc in Teacher & so on

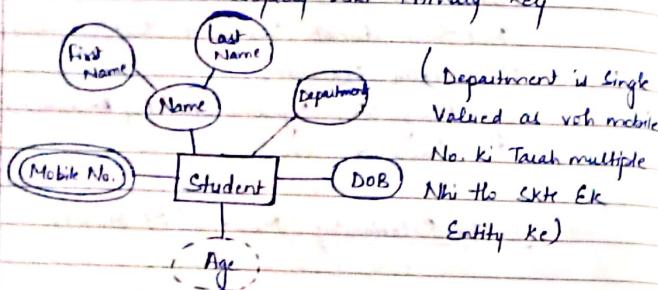
Representations

- Entity Set → weak Entity Set
- Relationship Set → weak Relationship Set
- Attributes → (oval/ellipse)
 - Composite (ek hi attribute me multiple chise)
 - Name (First name, last name)
 - Address (S.No., City, Society, State)
 - Multivalued (ek hi ke multiple thi tho like)
 - Mobile No.

Page No. _____
Date _____

→ Derived (Mtb Jiske aage de chalte ki zarurat Nhu Raha)
→ Age
→ Duration (Begin & End date)

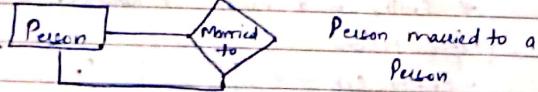
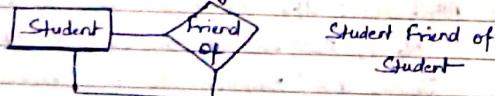
→ Key → which defines all the entities uniquely like Primary key



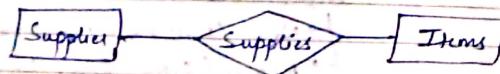
→ Degree of Relationship Sets

3 Types

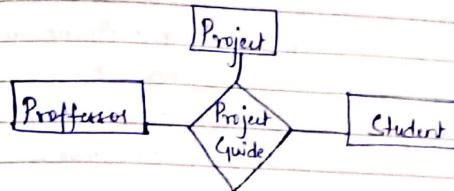
1) Unary - Ek hi Type ki Entity pe kaam tho



2) Binary (Most Common)

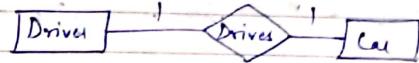


3) n-ary (Not very common) (More than 2)

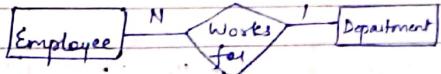
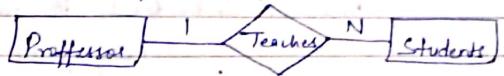


→ Cardinality (Ek Entity Kitni Baar use tho Rhi)
3 Types

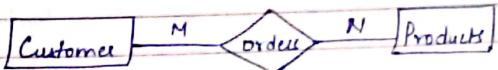
1) One to One (dono one time use tho)



2) One to Many (Many to one Rhi isilne ata)



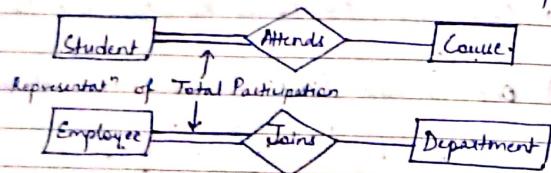
3) Many to Many



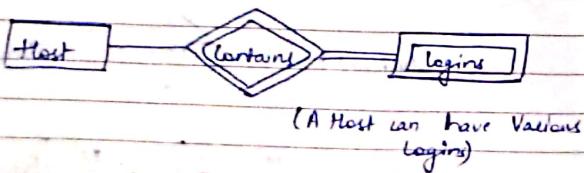
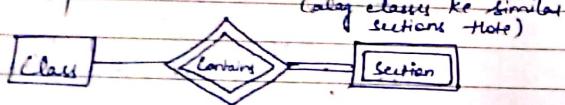
→ Participation & Weak Entity Set

1) Total Participation → Every entity of one side participates in the Relationship

e.g., These things which are compulsory to do like Every student must be involved in some course or Every employee must Join a department



2) Weak Entity Sets → Don't have their own key. Always Total Participation & to become a key weak Entity set has to take help of other Entity present. Individually it could not become a key.



→ Keys - An Attribute or set of Attributes that uniquely Identifies set of Entities

Candidate Key ; Primary Key ; Super Key ; Alternate Key

→ Primary

Enroll No.	Name	Adhar No.	Addr	PAN No.	Ph.No
.

Now, here in this database

(Enroll No. , Adhar No. & PAN No.) are

one or more ←
anyone
• Candidate Keys as they are the minimal set of attribute that derives all other attribute (Candidate Key must be minimal, Kaise kisi attribute ko to Identify)

• Primary Key Kai Ek Chose tho Jaati hai
From the options of Candidate keys & it is most convinient & Baaki Tables me Bhi use tho Rha Hota hai
• ~~Primary~~ Alternate Key - Candidate ki Bhi Kuch Options AftaH Choosing primary.

• Super Key - (Greater than or equal to, Candidate Keys) Any set of attribute that can uniquely identify the Table contents.
If no. is minimal like 1 than Candidate key & from them we shortlist primary key.

* In worst case, then saare attribute ko ikahe krde hai as vo to har Row ke liye unique hi honge. These will represent all the keys in that case.

`CREATE TABLE Order (`

;

`Cust_ID REFERENCES Customers (Cust_ID)`

ON DELETE CASCADE

`); for foreign` ↓
`Key` (For Referential Integrity)

* Foreign Key same table ke primary key ko bhi point kr skta hai. Referential Integrity means agar ek Jagah se entry delete kro to Baaki Sab Jagah pr Bhi pta lg Jaaye ki abh iska kaam Nhi Koi

∴ CASCADE line is solution for that, more solutions could be Restrict/NULL.

→ Database Normalization

Motive is to decrease Data Redundancy & Increase Data Integrity

- Redundancy is that same data is occurring at multiple places specially string ones

- Integrity means less erroneous data i.e Kam se Kam error vala

(Related data is Together)

We divide data into multiple Tables so the Redundancy is minimised.

- * Objectives of Good DB Design

- a) No updation, insertion & deletion anomalies
- b) Easily Extensible
- c) Good Performance for all Query Sets
- d) More Informative

Page No.	
Date	

Page No.	
Date	

(This is Bad DB as there is Redundancy)

→ Anomalies

1) Updation anomaly

Enroll No.	Student-ID	Student-Name	Subject-ID	Subject-name
-	- 1001	- ABC	- CS101	- Database
-	- 1002	- XYZ	- CS101	- Database
1001	ABC	ABC	CS102	OS
1003	BED	BED	CS102	OS
5 th → 1004	PQR	PQR	?	?

if we update database to DBMS than due to Redundancy Khi Khi toh Jayega Khi Khi Nhi Hoga, crash ki vjah se ya Bad Query ki vjat se

2) Insertion anomaly → as upar 5th entry insert kri toh as Bad DB so ho skta hai Bas Kuchh entries hi toh add karne ko & iski vjah se possibly "Koi Restrict" toh kii adhi adhvi entries enter Nhi kr skte toh Problem hoga;

3) Deletion Anomaly - As Bad DB toh let's suppose humne 1st & 2nd Entries delete krdi (---)(is line se) toh Database subject ka toh saara data thi ud Jayega i.e. subject hi Nhi Raha.

∴ to Reduce Redundancy we Break data into multiple Tables But Bohot Jyada, Bhi Nhi Hori Chahiye Ki Performance issues aa Jaye so there are Rules for normalization

→ Functional Dependency

Enroll No.	Name	Add ^x	Ph. No.
"			
"			
"			

Enroll No → Name

" → Add^x

" → Ph. No.

Student-ID	Subject-ID	Marks	Grade
Subj-ID	Student-ID	Marks	Grade

A → B ⇒ B is Functionally dependent on A

A	B
x	1
y	1
z	2

(A → B ✓) Here A is unique values
(B → A X) & could be used to uniquely Identify all the B values

Enroll-No.	Name	Add ^x	Enroll-No. → Name, Add ^x
101	Raj	ABC	
102	Ravi	ABC	Name → Add ^x X
103	Raj	XYZ	Add ^x → Name X

* Why do we study Functional dependencies

Stu-ID	Name	Dept-ID	Dept-Name	Dept.
101	ABC	10	CS	
102	BED	11	ECE	
103	ABC	10	CS	
104	XYZ	11	ECE	
105	CDE	10	CS	

Dept-ID → Dept-Name
(so due to this we could divide it into multiple Tables)

Dept-ID	Dept-Name	Dept
10	CS	
11	ECE	

Divided to Separated the data into other Table.

Functional Dependency

Trivial	Non-Trivial
$AB \rightarrow A$	$A \rightarrow B$
$A \rightarrow A$	$AB \rightarrow C$
$ABC \rightarrow AC$	$BC \rightarrow DEA$

Database Normalization

Data ko much more efficient way me store krne ka Tarika do that ke ek chiz efficient way me store ho & operations bhi like it is said to divide data into multiple tables. (1NF)

1) First Normal Form → If Every attribute contains only single value (Atomic)

Cust-ID	Name	Mob. No.	
101	ABC	886..., 9960...	Not in 1NF
102	BCD	8527...	

a) \rightarrow \downarrow solut'

"	"	Mob. No. 1	Mob. No. 2
		886...	9960...
		-	99...

(solut' but not always possible as if you are reqd. to fill then the skills than kitne attribute brangle)

b) \rightarrow solut'

Cust-ID	Name	Mob. No.	
101	ABC	Mob1.	(This is better solution but here Cust-ID
102	BCD	99...	would not remain unique)
:	:	:	

Table A

(Not in 3NF)

Table A me ek attribute aur jisko

ID Customer-ID

Mob. No.

This is not only in
First Normal Form

Page No.	
Date	

Second Normal Form (2NF)

Must be in 1NF & No Partial Dependency

No Non-Prime Attribute Should depend upon partial Candidate Key

User-ID	Course-ID	Course-Fee
1	CS101	5000
2	CS102	2000
1	CS102	2000
3	CS101	5000

(User-ID, Course-ID) \rightarrow Total Prime Key

But as course-ID \rightarrow course-fee

\therefore Partial Candidate key par ek Non Prime Key Depend Kr Rhi Hai Toh Not in 2NF

Solut' is

User-ID	Course-ID	Course-Fee
1	CS101	5000
2	CS102	2000

* * * a table is *

Third Normal Form (3NF)

Must be in 2NF & Non-Prime \rightarrow Non Prime

(a) (b) Not Allowed

No Non Prime Should depend on Non Prime

Stud-No.	Stud-Name	Stud-State	Stud-Country
101	Ram	Haryana	India
102	Ramneek	Punjab	"
103	Sureesh		"

Stud-No. \rightarrow { Every Attribute }

(Stud-State \rightarrow Stud-Country). Rule No (b) Violated

Page No.	
Date	

Stud-ID	103	IT - 10	Subject	Redundancy is gone
1001	110	103	DBMS	
1001	111	110	OS	
1002	111	111	DBMS	
1003	104	104	DBMS	

* If a table is in 2NF & one non Prime is defining other Non Prime than we take these non prime & make a New Table But Jaise Chiz define Kr Rhi Hati. Taisi use then original Table pe Bhi RKHte Hai

T₁: Stud-No, Stud-Name, Stud-State
T₂: Stud-State, Stud-Country

Now these 2 Tables are in 3NF

eg- Exam Name | Exam Year | Topper Name | Topper DOB

(Exam Name, Exam Year) → Topper name, Topper DOB
Topper DOB → Topper DOB

Now To fix it to be in 3NF we put First 3 Attr. in one Table & last 2 Attr. in Other Table to divide

Definition- d ⇒ (1) In 2NF

(1) Non Prime attr. not Transitively dependent on prime attributes

4) BCNF → More strict than previous 3

2NF ⇒ P → NP X 3NF ⇒ NP → NP X

BCNF ⇒ P/NP → P X

1, 2 & 3 Normal form me to ho hi But yeh upar vali cond' Bhi Satisfy ho

Only Superkeys on the left hand side ($x \rightarrow \text{Superkey}$)

eg- Stud-ID | Subject | Prof-ID (Stud-ID, Subject) → Prof-ID

Prof-ID → Subject

This is not BCNF → Partial Cand. Key → NP

Stud-ID Subject Prof-ID

1001	DBMS	103
1001	OS	110
1002	DBMS	111
1003	DBMS	103



Sometimes it creates problems as above. Dependency is gone but some Tables are not correct according to DB so some decompositions might not be ~~acceptable~~ acceptable in BCNF me lya paana guli Tarah se kaafi mushkil Hai

→ Indexing in Databases

i) Clustered Indexing - It is to put Indexing on Hard disk Block for data of Database & could be on Basis of any attribute but there is one based on primary key.

eg- Disk Blocks

A-Block	→	101 ...	102 ...
		;	;

Clustered Index on Order-ID

Order-ID	Order-date	Cost	Cust-ID
101	15-06-18	2000	102
102	15-06-19	5000	103

Clustered Indexing
Prime Index → Primary Key

(Now if we don't do Indexing than it would act like a heap ~~ki~~ ki Bai eye Jai Rha Hai But for performance issue Hoga as Searching Time in HDD is more.)
Clustered Index problem is that it is on Basis of one key only for mae. comes under cost of iron clustered

2 types → Space & density

search key | Pointer

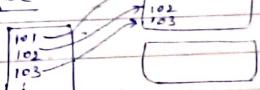
Stores Reference to disk block containing given key

Dense doesn't mean ki

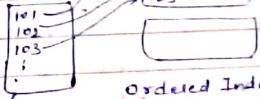
All the Rows must be present but all unique rows Bohat hi densely Present also index file me yeh entry ek Hoga Hui ek Entry ka → hai key ke liye Bohat ek Index file Faster But Update & Delete ke Time

Bohat dhyaan Rakhna Pata hai to maintain

i) Dense



Ordered Index file



Hashing File origin

Ek Ek Index

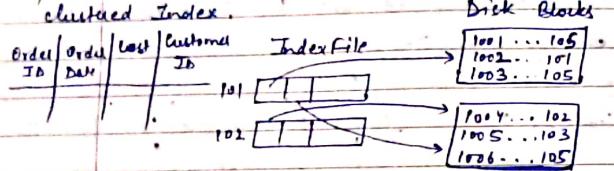
Bohat hi densely Present

also index file me yeh entry ek Hoga Hui ek Entry ka → hai key ke liye Bohat ek Index file Faster But Update & Delete ke Time

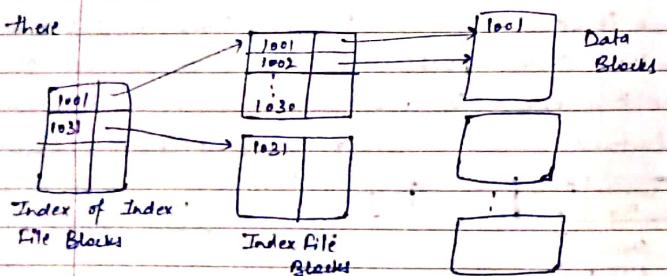
Bohat dhyaan Rakhna Pata hai to maintain

(ii) Space → as indexing is sequential so each block (disk) ke base address ki ~~key~~ ko store karta hai as basi sab derive ho Jayegi (Slow but update & delete very convenient).

2) Non Clustered Indexing (only one) → Clustered Index in DB are by default made on primary keys, also we should not make many secondary indexes as then it would create a problem of performance & overhead/ updation so only that entry which is used very oftenly should have secondary Indexes. If we want to index the DB according to another attribute to increase performance than we need secondary ~~index~~ ^{address} of non clustered Index.



3) Multilevel Indexing → Main purpose is to reduce disk block access time. But eventually comes a case where size goes a lot then we need B+ Tree. Due to multilevel less access to blocks are there.

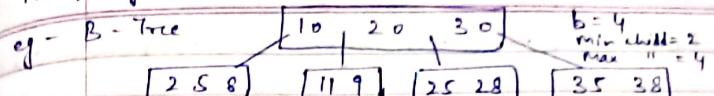
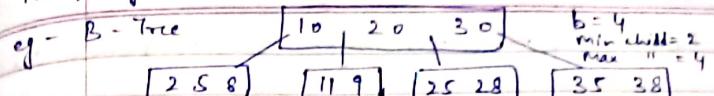


4) B-Trees → n-way search trees

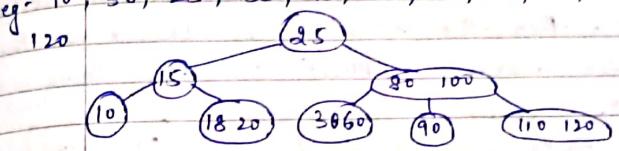
↳ If Branching Factor is b, then every node has $\lceil \frac{b}{2} \rceil$ to b children

↳ In B+ Tree leaf node contains all the data present in internal nodes so that sequential access to data is possible.

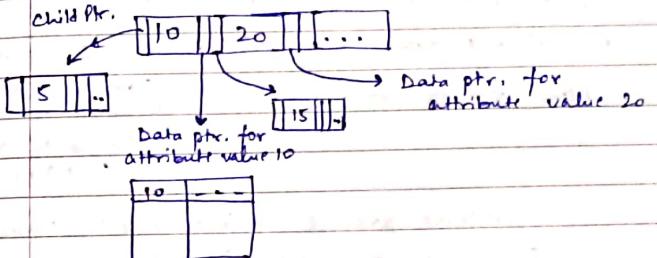
These are self Balancing Trees so height don't exceed (logn). Also No. of Keys in node = no. of Children - 1



Structure ek dum BST Jaisa Root ke children left me Bda Right me. (Also yeh Restructure karta hai & iski height upwards Bati hai). Insert " 7 sees the size & length whereas vice versa for delete" eg: 10, 30, 25, 80, 90, 100, 15, 18, 20, 60, 110, 120

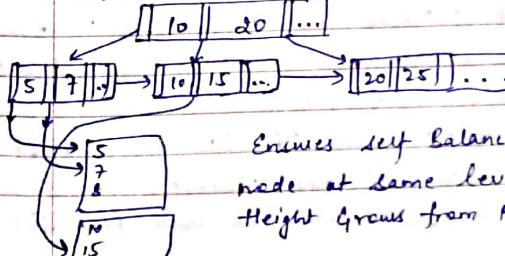


B-Tree Node Structure



* B+ Trees → Internal nodes have pointers to index blocks only. Leaf nodes store pointers to actual data blocks. Only the last ptr ~~last~~ in the leaf nodes point to next Index Block (leaf node me sequential access & mostly companies use kri-hai)

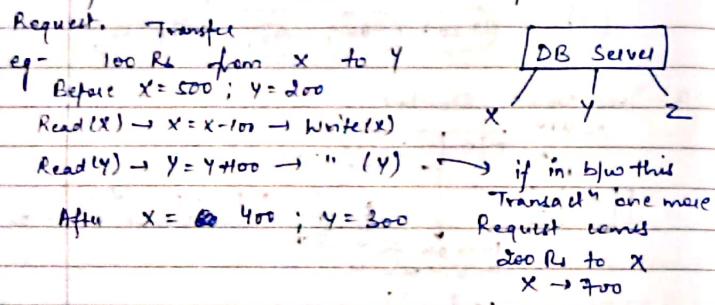
eg - 2 level B+ Trees



Ensures self balancing & all leaf node at same level & here Height Grows from Root Node.

Advantage → All the nodes has only one type of pointer unlike B-Trees & has sequential access.
Disadv. → Key value pair might be more than 1 so duplicate key value pair might occur in parent & children

→ Transaction & Concurrency Control (ACID Properties)
 "Transact" is a set of process instructions that should happen all together at same time or nothing should happen. Concurrency is just like in OS preemptive nature like if some process working on I/O then the other process should use the CPU for maximum utilization same for server request.



i) Interleave process might lead to inconsistency
 so might not get expected result

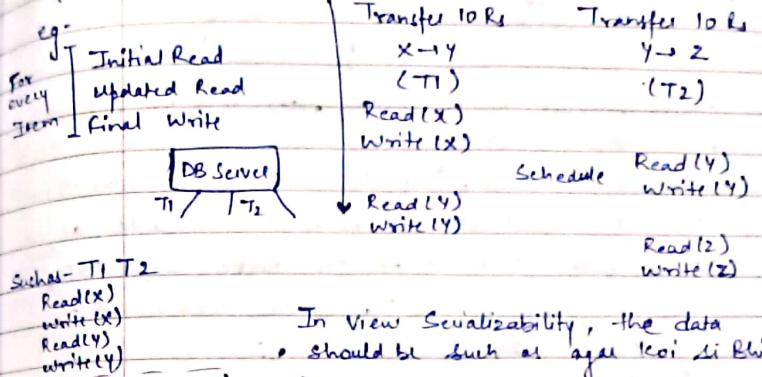
* ACID Properties

- i) Atomicity → Either whole thing executes or Nothing executes
- ii) Consistency → means ki in total amount should remain constant before & after Transaction
 eg- $500, 200 = 400, 300$ (Sum)
- iii) Isolated → Means esa lgna chahiye ki pura process akeli me chahiye Rhi Hai without anyone's Interference
- iv) Durability → means ki DB me latest updated values Rhi Chahiye.

eg- Bank Transaction me yeh sab Properties

→ Serializability → It is used to validate interleaved schedule that whether it is valid or not

① View Serializability → T1 Ke Baad T2 Ho ya ulta effect same hi ana chahiye. i.e. different Permutation's of T (Transact) ke liye same effect ana chahiye



In View Serializability, the data should be such as agar koi si Rhi "Permutation" me agar process tho such write(Z) as T1 T2 or ulta or if b Transaction tho b! ways tho dekhne se esa hi lgna chahiye ki processing same manner me ho Rhi Hai. But To check it is very expensive operat' very time consuming

② Conflict Serializability → In this we only talk about ~~conflicting serializability~~ Conflicting process for serializability
Conflicting Operat's → 2 Operat's from 2 different Transact's but on same data items & one of them must be "write"

Given Schedule		Serial Schedule (T1 T2)	
T1	T2	T1	T2
Read A Write A	Read A Write A	Read A Read B Write A Write B	As usual Niche Ko ke Rhi different data items so conflict serializa- bility is fine
Read B Write B	Read B Write B	Read A Write A Read B Write B	

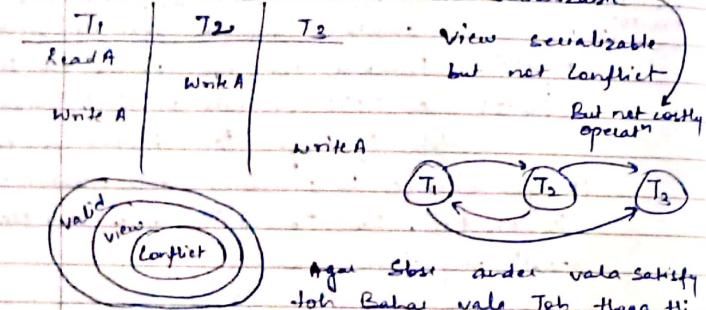
Given Schedule	
T1	T2
Read A	
	Read A Write B
Write A Read B Write B	

This is neither conflict serializable with T1T2 nor T2T1

* Conflict Serializability could be checked using Precedence Graph

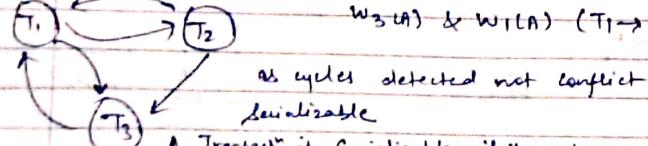


- * This concept is to check after that DB is consistent even after interleaving.
- * If we draw Graph & there is a cycle b/w them then it is not at all conflict serializable as it khega phle sare T1 pe kro durr khega phle T2. This is more strict than view serializable.



Precedence Graph thru kewal conflicting processes ka bante hai & cycle detect karte hai

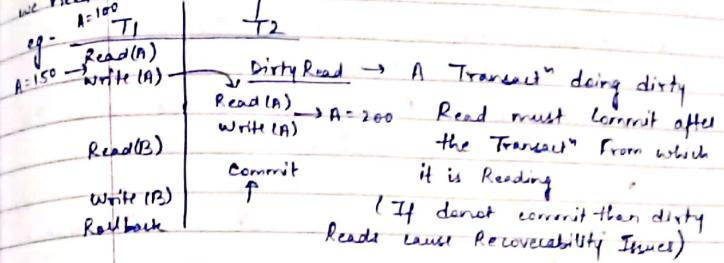
T1	T2	T3	R1(A) & W2(A) [T1 → T2]
R(A)			R1(A) & W2(A) [T1 → T3]
	W(A)		R1(A) & W3(A) [T1 → T3]
W(A)			W2(A) & R3(A) [T2 → T3]
		W(A)	W2(A) & W3(A) [T2 → T3]
			R3(A) & W1(A) [T3 → T1]
			W3(A) & W1(A) [T1 → T3]



A Transaction is Serializable if its outcome is equal to outcome of its Transaction executed serially i.e sequential without overlapping in time

* Recoverable, Cascadecess & Strict

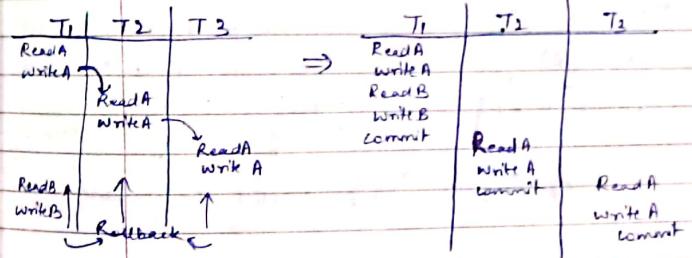
- (1) As we know in DB, Transaction may not execute completely either due to hardware failure, system crash, software issue in that case we have to Roll Back failed Transaction, but if some other Transaction also have used same values so we have to Roll Back those Transaction as well. So we need Recoverability.



i. For Recoverable schedule transaction must be committed in order.

Dirty Read → when a Transaction Reads a data from uncommitted write in another Transaction

(ii) Cascadecess Schedule → Dirty Reads are not allowed i.e reading data written by an uncommitted Transaction not allowed.



(iii) Strict → This takes into consideration the Blind writes that is the writes without reading so if Rollback happens then value would not be updated properly. Therefore, Neither Dirty Reads nor lost update problem allowed. Rule is that Blind writes are only allowed after Transaction is committed i.e. Read, write from uncommitted to Transaction not allowed

→ 2 Phase Locking Protocol → These are the protocol that has the objective that we want to ensure consistency, interleaved Transaction must happen & to ensure that this is conflict Equivalent to a Serial Stamp.

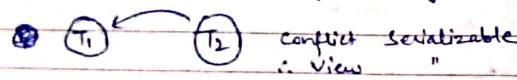
* Lock → it is to lock something in DB & if something is locked then no other Transaction can take place. lock is same as the one in OS

2 Types → Shared → as we need to apply concurrency to Reading Re Time pe share krt sat leg hence shared lock.

, Exclusive → Whereas during writing or modify we need exclusive so that no interference is there.

	T1	T2	
S - Shared			
X - Exclusive			
Lock - S(A)	T1		
Read(A)		T2	
Lock S(A)			T1
Read(A)			
Lock - S(B)			
Read - B)			
Lock - X(A)			
Wait for T2 to Release			
Shared lock on A			
Write(A)			
unlock - S(B)			
unlock - X(A)			

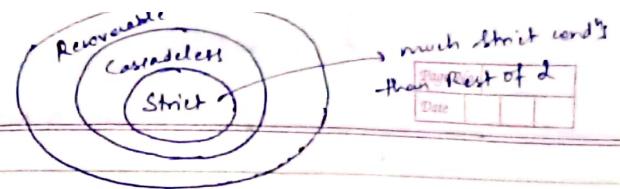
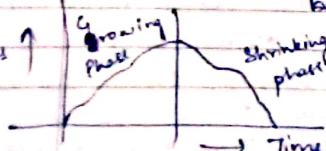
lock point is when all the acquiring of the locks have been done & all locking Nahi hogi only releasing



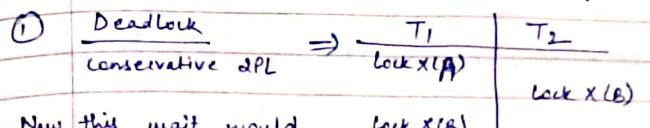
2 Phases in Every Transaction

→ Growing phase → New locks on data items may be acquired but none can be released

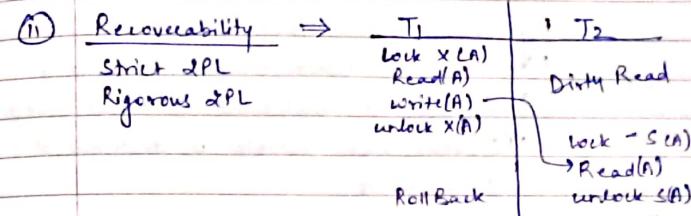
→ Shrinking phase → Existing locks may be released but no new locks can be acquired



* Problems with 2PL (2 Phase Locking)



New this wait would be oo hence case of a deadlock



Rollback

lock - S(A)

→ Read(A)

unlock S(A)

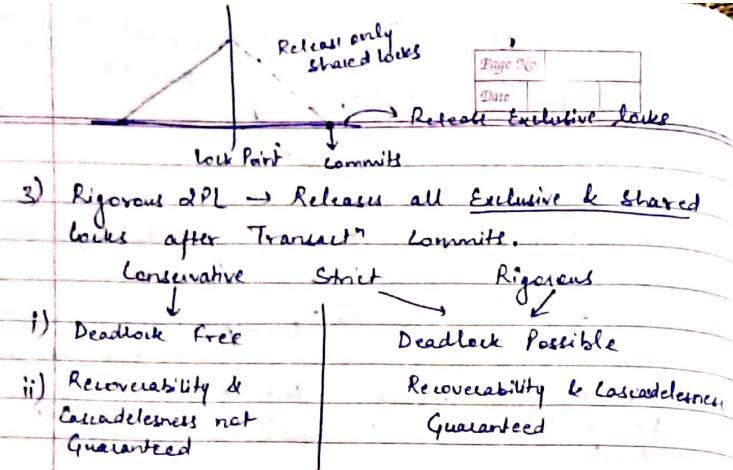
commit

* Types of 2PL

1) Conservative 2PL → Handles deadlock → Remove hold & wait, this protocol Requires Transaction to lock all the items it access before the Transaction begins executing. If any of the predeclared items needed can't be locked, the Transaction does not lock any of the items, instead, it waits until all the items are available for locking.

Note → deadlock is not that big issue as it happens rarely, also this doesn't take care of recoverability which is more important.

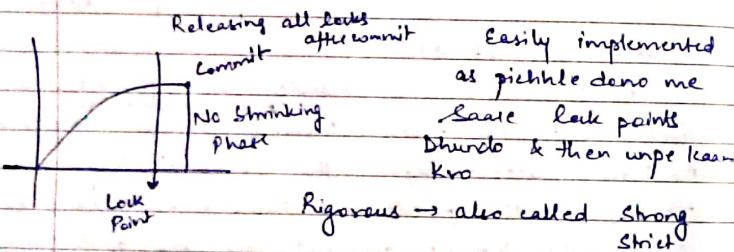
2) Strict 2PL → Releases all Exclusive locks only after Transaction commits. Ensures that schedule is Recoverable, Cascaded.



Rigorous v/s Strict

Impl as most used protocol

- \rightarrow Strict & allow more concurrency
- \rightarrow Simple to Implement & most used
- \rightarrow Rigorous also suitable in distributed environments



\rightarrow Time Stamp Based Protocol \rightarrow To protocol has the same objective i.e., Timestamp is a unique identifier created by DBMS. Main idea is to order the Transaction's based on these Timestamps. This makes it serializable & equivalent to Serial Schedule.

$TS(T_i)$ \rightarrow Time stamp of Transaction T_i

$RTS(x)$ \rightarrow Time stamp of latest Transaction that performed Read on object (x) \therefore max timestamp

$WTS(x)$ \rightarrow Timestamp of the latest Transaction that performed write on object (x) \therefore max timestamp

$T_1(10) \mid T_2(20) \mid T_3(30) \mid T_1 T_2 T_3$

eg-	$T_1(10)$	$T_2(20)$
$TS(T_1) = 10$	Read(A)	
$TS(T_2) = 20$	Write(A)	
$RTS(A) = 10$		
$WTS(A) = 10$		
$RTS(A) = 20$		
$WTS(A) = 20$		
$RTS(B) = 20$		

yha Problem Ayi so new system will Roll Back & instruct krega ki let them come in a new Time stamp which is conflict free

Now, there is no possibility of deadlock But Recoverability could be an issue as there occurs Dirty Read (as in above example)

* Simple Rules

i) Rule for Read(x) in T_i

if $TS(T_i) < WTS(x)$ then abort T_i
else Allow Read(x) \rightarrow current one
 $RTS(x) = \max(RTS(x), TS(T_i))$

ii) Rule for write(x) in T_i

if $TS(T_i) < RTS(x)$ OR $TS(T_i) < WTS(x)$
then abort
else Allow Write(x)
 $WTS(x) = TS(T_i)$

NOTE \rightarrow Extra work / Book Keeping could be done in this Protocol to ensure Recoverability & Cascadeness by applying Restriction to all dirty Read operat's. This work is done by Scheduler whose job is to Schedule Transaction in an interleaved manner & then protocol checks for these Rules.