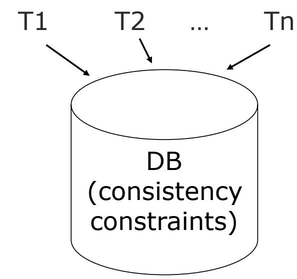
CSE 562
Database Systems

Concurrency Control

Some slides are based or modified from originals by
Database Systems: The Complete Book,
Pearson Prentice Hall 2nd Edition
©2008 Garcia-Molina, Ullman, and Widom

cse@buffalo

Chapter 18: Concurrency Control



Example:

T1: Read(A)
A ← A+100
Write(A)
Read(B)
B ← B+100
Write(B)

T2: Read(A)
A ← A×2
Write(A)
Read(B)
B ← B×2
Write(B)

Constraint: A=B

Schedule A

T1	T2	A	B
		25	25
Read(A); A ← A+100;			
Write(A);		125	
Read(B); B ← B+100;			
Write(B);			125
	Read(A);A ← A×2;		
	Write(A);	250	
	Read(B);B ← B×2;		
	Write(B);		250
		250	250

Schedule B

T1	T2	A	B
		25	25
	Read(A);A ← A×2;		
	Write(A);	50	
	Read(B);B ← B×2;		
	Write(B);		50
Read(A); A ← A+100			
Write(A);		150	
Read(B); B ← B+100;			
Write(B);			150
		150	150

Schedule C

T1	T2	A	B
		25	25
Read(A); A ← A+100			
Write(A);		125	
	Read(A);A ← A×2;		
	Write(A);	250	
Read(B); B ← B+100;			
Write(B);			125
	Read(B);B ← B×2;		
	Write(B);		250
		250	250

Schedule D

T1	T2	A	B
		25	25
Read(A); A ← A+100			
Write(A);		125	
	Read(A);A ← A×2;		
	Write(A);	250	
	Read(B);B ← B×2;		
	Write(B);		50
Read(B); B ← B+100;			
Write(B);			150
		250	150

Schedule E

Same as Schedule D but with new T2′

T1	T2′	A	B
		25	25
Read(A); A ← A+100			
Write(A);		125	
	Read(A);A ← A×1;		
	Write(A);	125	
	Read(B);B ← B×1;		
	Write(B);		25
Read(B); B ← B+100;			
Write(B);			125
		125	125

- Want schedules that are "good", regardless of
  - initial state and
  - transaction semantics
- Only look at order of read and writes

Example:

Sc=r1(A)w1(A)r2(A)w2(A)r1(B)w1(B)r2(B)w2(B)

Example:

Sc=r1(A)w1(A)r2(A)w2(A)r1(B)w1(B)r2(B)w2(B)

$$Sc' = \underbrace{r1(A)w1(A)\ r1(B)w1(B)}_{T1}\underbrace{r2(A)w2(A)r2(B)w2(B)}_{T2}$$

However, for Sd:

Sd=r1(A)w1(A)r2(A)w2(A) r2(B)w2(B)r1(B)w1(B)

- as a matter of fact,
  T2 must precede T1
  in any equivalent schedule,
  i.e., T2 → T1

- T2 → T1
- Also, T1 → T2

$$\begin{tikzcd} T1 \arrow[r, bend left] & T2 \arrow[l, bend left] \end{tikzcd}$$

⇨ Sd cannot be rearranged into a serial schedule

⇨ Sd is not "equivalent" to any serial schedule

⇨ Sd is "bad"

Returning to Sc

Sc=r1(A)w1(A)r2(A)w2(A)r1(B)w1(B)r2(B)w2(B)

T1 → T2   T1 → T2

☛ no cycles ⇒ Sc is "equivalent" to a serial schedule (in this case T1,T2)

Concepts

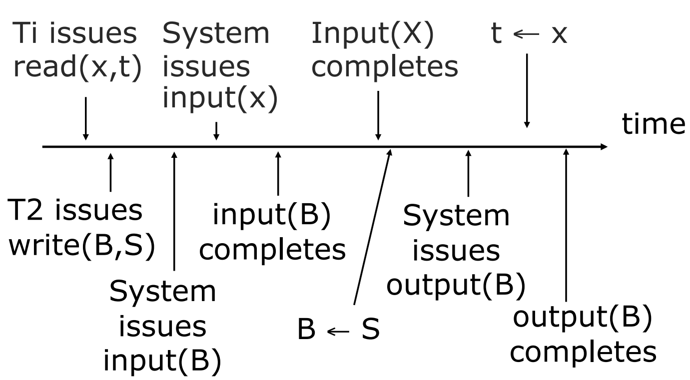
Transaction: sequence of ri(x), wi(x) actions

Conflicting actions: r1(A) / w2(A)   w2(A) / r1(A)   w1(A) / w2(A)

Schedule: represents chronological order in which actions are executed

Serial schedule: no interleaving of actions or transactions

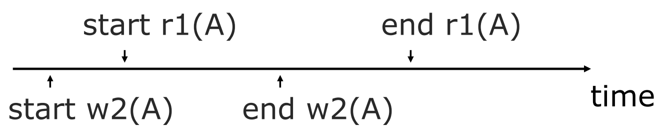
What About Concurrent Actions?



So net effect is either

- S=…r1(x)…w2(B)… or
- S=…w2(B)…r1(x)…

What about conflicting, concurrent actions on same object?



- Assume equivalent to either r1(A) w2(A)
  or w2(A) r1(A)
- ⇒ low level synchronization mechanism
- Assumption called "atomic actions"

Definition

S1, S2 are <u>conflict equivalent</u> schedules if S1 can be transformed into S2 by a series of swaps on non-conflicting actions.

Definition

A schedule is <u>conflict serializable</u> if it is conflict equivalent to some serial schedule.

Precedence Graph P(S) (S is schedule)

Nodes: transactions in S

Arcs: Ti → Tj whenever

- pi(A), qj(A) are actions in S
- pi(A) $<_S$ qj(A)
- at least one of pi, qj is a write

Exercise:

- What is P(S) for
  S = w3(A) w2(C) r1(A) w1(B) r1(C) w2(A) r4(A) w4(D)

- Is S serializable?

Another Exercise:

- What is P(S) for
  S = w1(A) r2(A) r3(A) w4(A) ?

Lemma

S1, S2 conflict equivalent $\Rightarrow$ P(S1)=P(S2)

<u>Proof:</u>

Assume P(S1) $\neq$ P(S2)

$\Rightarrow$ $\exists$ Ti: Ti $\rightarrow$ Tj in S1 and not in S2

$\Rightarrow$ S1 = …pi(A)… qj(A)…
S2 = …qj(A)…pi(A)…

$\left.\right\}$ pi, qj conflict

$\Rightarrow$ S1, S2 not conflict equivalent

Note: P(S1)=P(S2) $\not\Rightarrow$ S1, S2 conflict equivalent

<u>Counter example:</u>

S1=w1(A) r2(A) w2(B) r1(B)

S2=r2(A) w1(A) r1(B) w2(B)

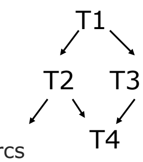
Theorem

P(S1) acyclic $\Longleftrightarrow$ S1 conflict serializable

($\Leftarrow$) Assume S1 is conflict serializable
$\Rightarrow$ $\exists$ Ss: Ss, S1 conflict equivalent
$\Rightarrow$ P(Ss) = P(S1)
$\Rightarrow$ P(S1) acyclic since P(Ss) is acyclic

Theorem

P(S1) acyclic $\Longleftrightarrow$ S1 conflict serializable



($\Rightarrow$) Assume P(S1) is acyclic
Transform S1 as follows:
(1) Take T1 to be transaction with no incident arcs
(2) Move all T1 actions to the front

S1 = …… qj(A) …… p1(A) …

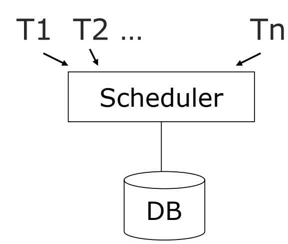
(3) we now have S1 = < T1 actions ><... rest ...>
(4) repeat above steps to serialize rest!

How to Enforce Serializable Schedules?

- Option 1: run system, recording P(S); at end of day, check for P(S) cycles and declare if execution was good

How to Enforce Serializable Schedules?

- Option 2: prevent P(S) cycles from occurring

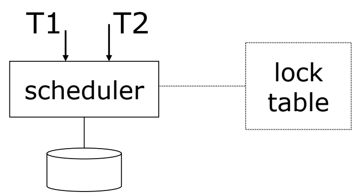


A Locking Protocol

Two new actions:

lock (exclusive): li (A)

unlock: ui (A)



Rule #1: Well-Formed Transactions

Ti: … li(A) … pi(A) … ui(A) …

Rule #2: Legal Scheduler

S = ……… li(A) …………… ui(A) ………

no lj(A)

Exercise:

- What schedules are legal?
  What transactions are well-formed?

  S1 = l1(A)l1(B)r1(A)w1(B)l2(B)u1(A)u1(B)
  r2(B)w2(B)u2(B)l3(B)r3(B)u3(B)

  S2 = l1(A)r1(A)w1(B)u1(A)u1(B)
  l2(B)r2(B)w2(B)l3(B)r3(B)u3(B)

  S3 = l1(A)r1(A)u1(A)l1(B)w1(B)u1(B)
  l2(B)r2(B)w2(B)u2(B)l3(B)r3(B)u3(B)

Schedule F

T1	T2
l1(A);Read(A)	
A←A+100;Write(A);u1(A)	
	l2(A);Read(A)
	A←Ax2;Write(A);u2(A)
	l2(B);Read(B)
	B←Bx2;Write(B);u2(B)
l1(B);Read(B)	
B←B+100;Write(B);u1(B)	

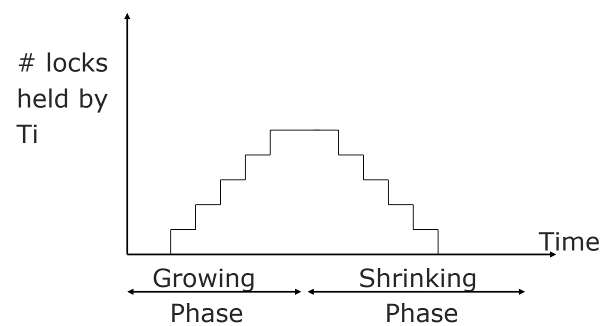
Schedule F

T1	T2	A	B
		25	25
l1(A);Read(A)			
A←A+100;Write(A);u1(A)		125	
	l2(A);Read(A)		
	A←Ax2;Write(A);u2(A)	250	
	l2(B);Read(B)		
	B←Bx2;Write(B);u2(B)		50
l1(B);Read(B)			
B←B+100;Write(B);u1(B)			150
		250	150

Rule #3: Two Phase Locking (2PL) for Transactions

Ti = …… li(A) ………… ui(A) ……

no unlocks ←| |→ no locks



Schedule G

T1	T2
l1(A);Read(A)	
A $\leftarrow$ A+100;Write(A)	
l1(B); u1(A)	
	l2(A);Read(A)
	A $\leftarrow$ Ax2;Write(A);l2(B) delayed

Schedule G

T1	T2
l1(A);Read(A)	
A $\leftarrow$ A+100;Write(A)	
l1(B); u1(A)	
	l2(A);Read(A)
	A $\leftarrow$ Ax2;Write(A);l2(B) delayed
Read(B);B $\leftarrow$ B+100	
Write(B); u1(B)	

Schedule G

T1	T2
l1(A);Read(A)	
A $\leftarrow$ A+100;Write(A)	
l1(B); u1(A)	
	l2(A);Read(A)
	A $\leftarrow$ Ax2;Write(A);l2(B) delayed
Read(B);B $\leftarrow$ B+100	
Write(B); u1(B)	
	l2(B); u2(A);Read(B)
	B $\leftarrow$ Bx2;Write(B);u2(B);

Schedule H (T2 reversed)

T1	T2
l1(A); Read(A)	l2(B);Read(B)
A $\leftarrow$ A+100;Write(A)	B $\leftarrow$ Bx2;Write(B)
l1(B) delayed	l2(A) delayed

- Assume deadlocked transactions are rolled back
  - They have no effect
  - They do not appear in schedule

E.g., Schedule H = $\underbrace{\qquad\qquad\qquad}_{\text{This space intentionally left blank!}}$

Next Step:

Show that rules #1,2,3 ⇒ conflict-serializable schedules

<u>Conflict rules for</u> li(A), ui(A):

- li(A), lj(A) conflict
- li(A), uj(A) conflict

Note: no conflict <ui(A), uj(A)>, <li(A), rj(A)>,...

<u>Theorem</u> Rules #1,2,3 ⇒ conflict serializable schedule
(2PL)

To help in proof:
<u>Definition</u> Shrink(Ti) = SH(Ti) = first unlock action of Ti

Lemma

Ti → Tj in S ⇒ SH(Ti) $<_S$ SH(Tj)

Proof of lemma:

Ti → Tj means that

S = … pi(A) … qj(A) …; p,q conflict

By rules 1,2:

S = … pi(A) … ui(A) … lj(A) ... qj(A) …

By rule 3: SH(Ti) SH(Tj)

So, SH(Ti) $<_S$ SH(Tj)

Theorem Rules #1,2,3 ⇒ conflict
(2PL) serializable
schedule

Proof:

(1) Assume P(S) has cycle

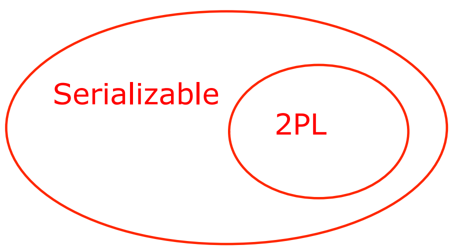
T1 → T2 →…. Tn → T1

(2) By lemma: SH(T1) < SH(T2) <…< SH(T1)

(3) Impossible, so P(S) acyclic

(4) ⇒ S is conflict serializable

2PL Subset of Serializable



S1: w1(x) w3(x) w2(y) w1(y)

- S1 cannot be achieved via 2PL: The lock by T1 for y must occur after w2 (y), so the unlock by T1 for x must occur after this point (and before w3(x)). Thus, w3(x) cannot occur under 2PL where shown in S1 because T1 holds the x lock at that point.
- However, S1 is serializable (equivalent to T2, T1, T3).

- Beyond this simple 2PL protocol, it is all a matter of improving performance and allowing more concurrency….
  - Shared locks
  - Multiple granularity
  - Inserts, deletes and phantoms
  - Other types of C.C. mechanisms

Shared Locks

So far:

S = ...l1(A) r1(A) u1(A) … l2(A) r2(A) u2(A) …

Do not conflict

Instead:

S=... ls1(A) r1(A) ls2(A) r2(A) …. us1(A) us2(A)

Lock actions

l-ti(A): lock A in t mode (t is S or X)

u-ti(A): unlock t mode (t is S or X)

Shorthand:

ui(A): unlock whatever modes
Ti has locked A

Rule #1: Well Formed Transactions

Ti =... l-S1(A) … r1(A) … u1 (A) …

Ti =... l-X1(A) … w1(A) … u1 (A) …

- What about transactions that read and write same object?

Option 1: Request exclusive lock

Ti = ...l-X1(A) … r1(A) ... w1(A) ... u(A) …

- What about transactions that read and write same object?

Option 2: Upgrade
(E.g., need to read, but don't know if will write…)

Ti=…l-S1(A)…r1(A)…l-X1(A)…w1(A)…u(A)…

Think of
- Get 2nd lock on A, or
- Drop S, get X lock

Rule #2: Legal Scheduler

S = … l-Si(A) … … ui(A) …

no l-Xj(A)

S = … l-Xi(A) … … ui(A) …

no l-Xj(A)
no l-Sj(A)

A Way To Summarize Rule #2

Compatibility Matrix

Comp

	S	X
S	true	false
X	false	false

Rule #3: 2PL Transactions

No change except for upgrades:

(I) If upgrade gets more locks

(e.g., S → {S, X}) then no change!

(II) If upgrade releases read (shared) lock (e.g., S → X)

- can be allowed in growing phase

Theorem Rules 1,2,3 for S/X locks ⇒ Conf.serializable schedules

Proof: similar to X locks case

Detail:

l-ti(A), l-rj(A) do not conflict if comp(t,r)

l-ti(A), u-rj(A) do not conflict if comp(t,r)

Lock Types Beyond S/X

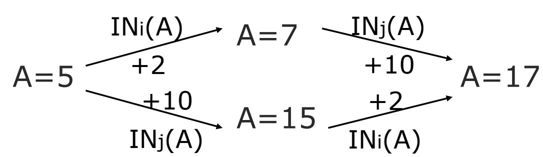
Examples:

(1) increment lock

(2) update lock

Example (1): Increment Lock

- Atomic increment action: INi(A)
  {Read(A); A ← A+k; Write(A)}
- INi(A), INj(A) do not conflict!



Comp

	S	X	I
S			
X			
I			

Comp

	S	X	I
S	T	F	F
X	F	F	F
I	F	F	T

Update Locks

A common deadlock problem with upgrades:

T1	T2
l-S1(A)	
	l-S2(A)
l-X1(A)	
	l-X2(A)

--- Deadlock ---

Solution

If Ti wants to read A and knows it may later want to write A, it requests update lock (not shared)

Comp

New request

Lock already held in	S	X	U
S			
X			
U			

Comp

New request

Lock already held in	S	X	U
S	T	F	T
X	F	F	F
U	TorF	F	F

-> symmetric table?

Note: object A may be locked in different modes at the same time...

S1=...l-S1(A)...l-S2(A)...l-U3(A)... { l-S4(A)...?
l-U4(A)...? }

- To grant a lock in mode t, mode t must be compatible with all currently held locks on object

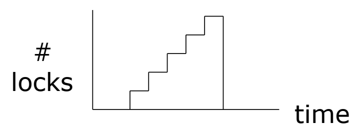
How Does Locking Work in Practice?

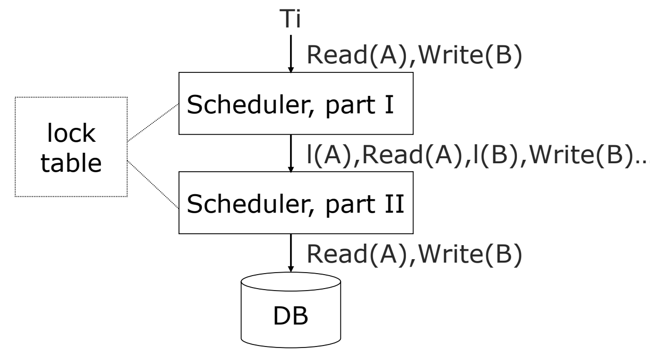
- Every system is different
  (E.g., may not even provide CONFLICT-SERIALIZABLE schedules)
- But here is one (simplified) way...

Sample Locking System

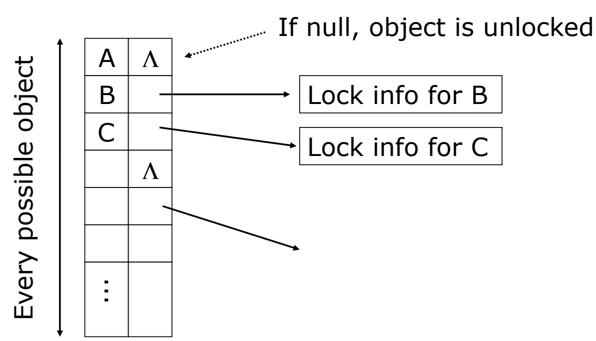
(1) Don't trust transactions to request/release locks

(2) Hold all locks until transaction commits

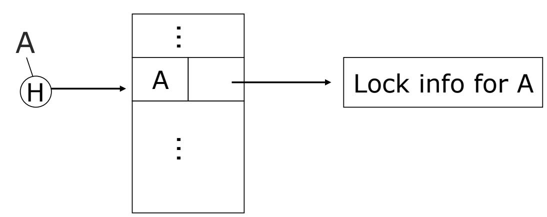




Lock Table: Conceptually

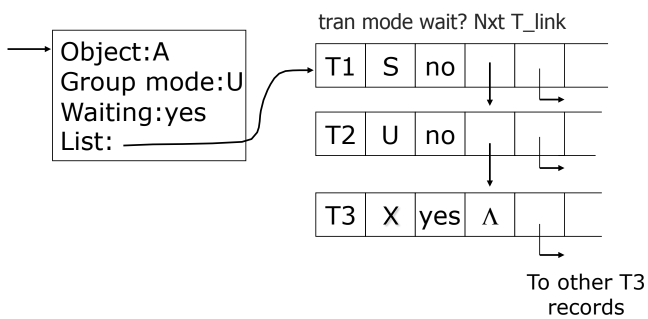


But Use Hash Table:

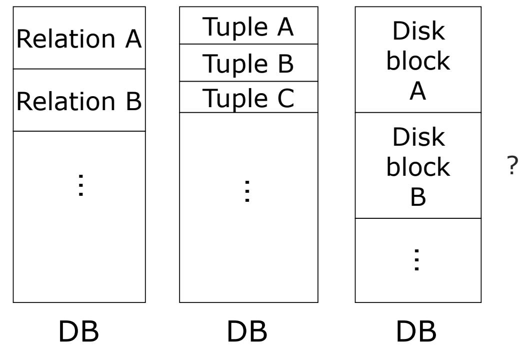


If object not found in hash table, it is unlocked

Lock Info for A: Example



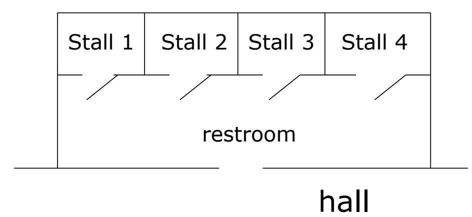
What Are The Objects We Lock?



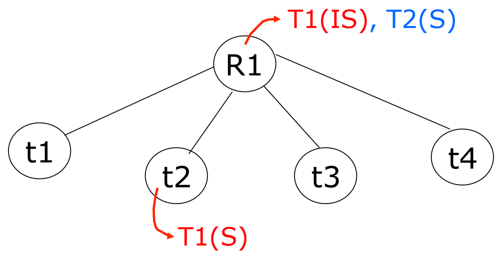
- Locking works in any case, but should we choose small or large objects?
- If we lock large objects (e.g., Relations)
  - Need few locks
  - Low concurrency
- If we lock small objects (e.g., tuples, fields)
  - Need more locks
  - More concurrency

We Can Have It Both Ways!!

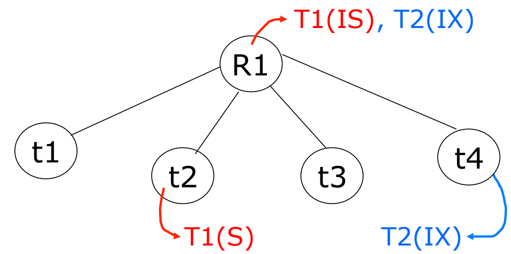
Ask any janitor to give you the solution...



Example



Example



Multiple Granularity

Comp

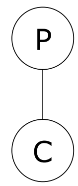
Holder \ Requestor	IS	IX	S	SIX	X
IS					
IX					
S					
SIX					
X					

Multiple Granularity

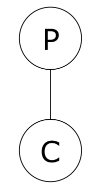
Comp

Holder \ Requestor	IS	IX	S	SIX	X
IS	T	T	T	T	F
IX	T	T	F	F	F
S	T	F	T	F	F
SIX	T	F	F	F	F
X	F	F	F	F	F

Parent locked in	Child can be locked in
IS	
IX	
S	
SIX	
X	



Parent locked in	Child can be locked in
IS	IS, S
IX	IS, S, IX, X, SIX
S	[S, IS] not necessary
SIX	X, IX, [SIX]
X	none

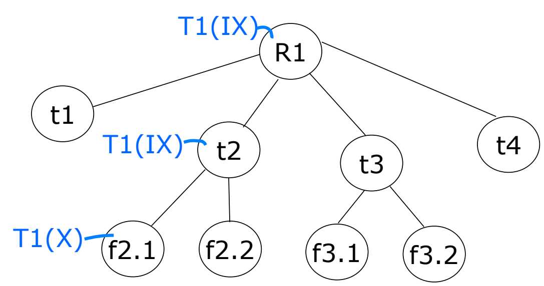


Rules

(1) Follow multiple granularity comp function
(2) Lock root of tree first, any mode
(3) Node Q can be locked by Ti in S or IS only if parent(Q) locked by Ti in IX or IS
(4) Node Q can be locked by Ti in X,SIX,IX only if parent(Q) locked by Ti in IX,SIX
(5) Ti is two-phase
(6) Ti can unlock node Q only if none of Q's children are locked by Ti

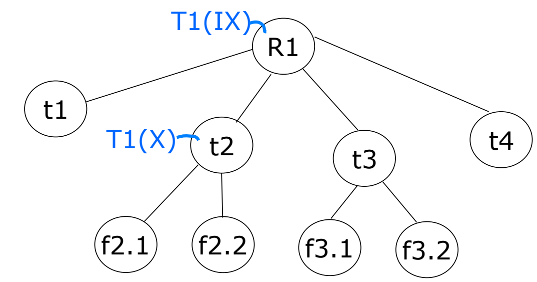
Exercise:

- Can T2 access object f2.2 in X mode? What locks will T2 get?



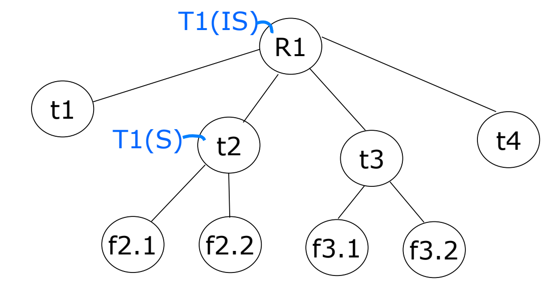
Exercise:

- Can T2 access object f2.2 in X mode? What locks will T2 get?



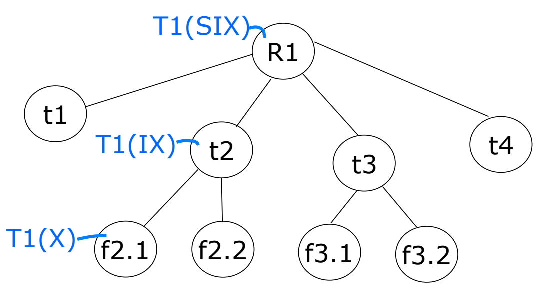
Exercise:

- Can T2 access object f3.1 in X mode? What locks will T2 get?



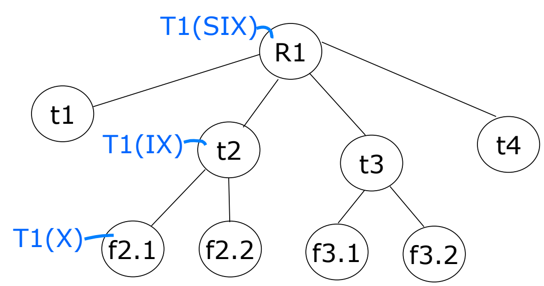
Exercise:

- Can T2 access object f2.2 in S mode? What locks will T2 get?

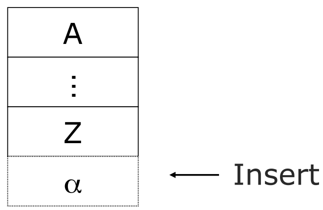


Exercise:

- Can T2 access object f2.2 in X mode? What locks will T2 get?



Insert + Delete Operations



Modifications To Locking Rules:

(1) Get exclusive lock on A before deleting A

(2) At insert A operation by Ti, Ti is given exclusive lock on A

Still have a problem: Phantoms

Example: relation R (E#,name,…)
constraint: E# is key
use tuple locking

R

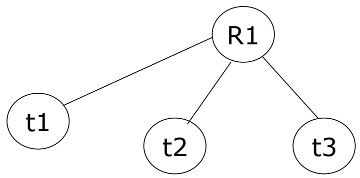
	E#	Name	…
o1	55	Smith	
o2	75	Jones	

T1: Insert <04,Kerry,…> into R
T2: Insert <04,Bush,…> into R

T1	**T2**
S1(o1)	S2(o1)
S1(o2)	S2(o2)
Check Constraint	Check Constraint
⋮	⋮
Insert o3[04,Kerry,..]	
	Insert o4[04,Bush,..]

Solution

- Use multiple granularity tree
- Before insert of node Q, lock parent(Q) in X mode

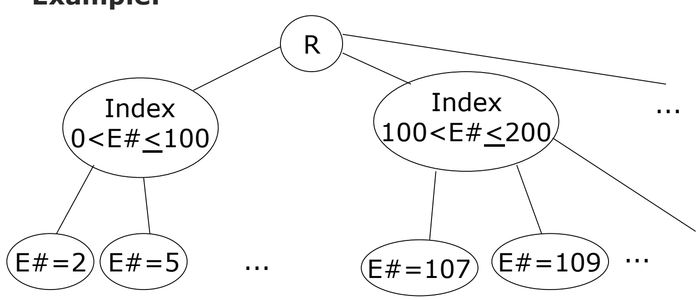


Back To Example

T1: Insert<04,Kerry>	T2: Insert<04,Bush>
T1	**T2**
X1(R)	
	X2(R) *delayed*
Check Constraint Insert<04,Kerry> U(R)	
	X2(R) Check Constraint Oops! e# = 04 already in R!

Instead of Using R, Can Use Index on R

Example:



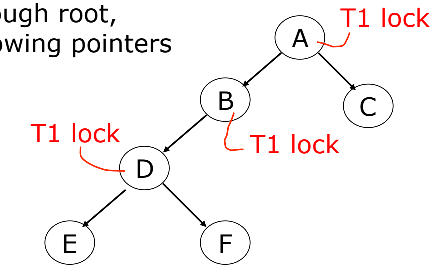
- This approach can be generalized to multiple indexes…

Next:

- Tree-based concurrency control
- Validation concurrency control

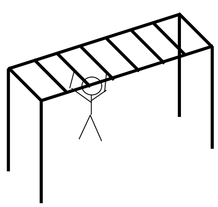
Example

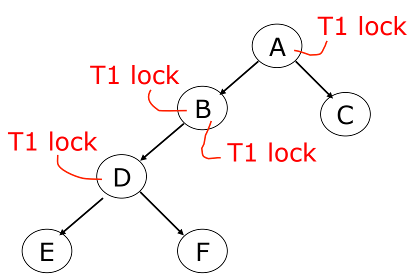
- all objects accessed through root, following pointers



☛ can we release A lock if we no longer need A??

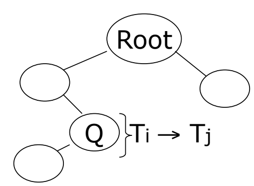
Idea: Traverse Like "Monkey Bars"





Why Does This Work?

- Assume all Ti start at root; exclusive lock
- Ti → Tj ⇒ Ti locks root before Tj



- Actually works if we don't always start at root

Rules: Tree Protocol (exclusive locks)

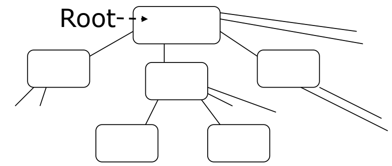
(1) First lock by Ti may be on any item

(2) After that, item Q can be locked by Ti only if parent(Q) locked by Ti

(3) Items may be unlocked at any time

(4) After Ti unlocks Q, it cannot relock Q

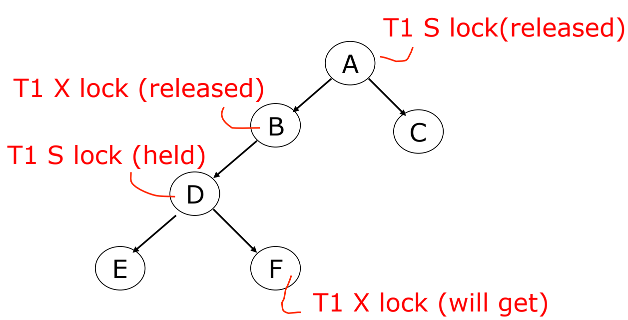
- Tree-like protocols are used typically for B-tree concurrency control



E.g., during insert, do not release parent lock, until you are certain child does not have to split

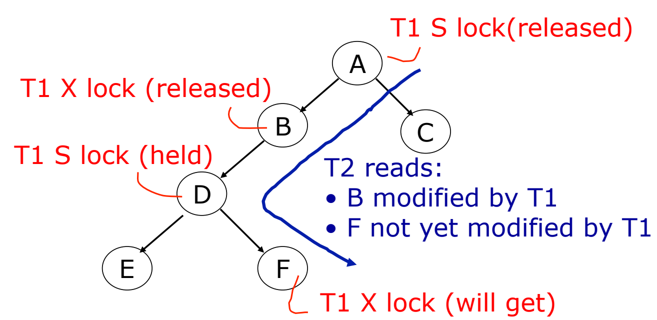
Tree Protocol with Shared Locks

- Rules for shared & exclusive locks?



Tree Protocol with Shared Locks

- Rules for shared & exclusive locks?



Tree Protocol with Shared Locks

- Need more restrictive protocol
- Will this work??
  - Once T_1 locks one object in X mode, all further locks down the tree must be in X mode

Validation

Transactions have 3 phases:

(1) Read
- all DB values read
- writes to temporary storage
- no locking

(2) Validate
- check if schedule so far is serializable

(3) Write
- if validate ok, write to DB

Key Idea

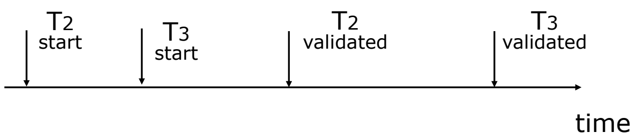
- Make validation atomic
- If T1, T2, T3,… is validation order, then resulting schedule will be conflict equivalent to Ss = T1 T2 T3…

To implement validation, system keeps two sets:

- FIN = transactions that have finished phase 3 (and are all done)
- VAL = transactions that have successfully finished phase 2 (validation)

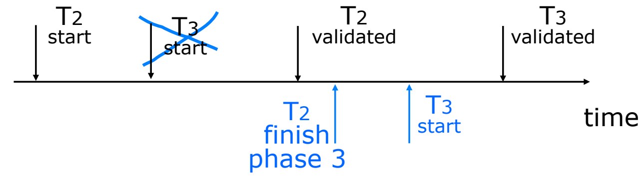
Example of What Validation Must Prevent:

RS(T2)={B} $\cap$ RS(T3)={A,B} $\neq \emptyset$
WS(T2)={B,D} WS(T3)={C}



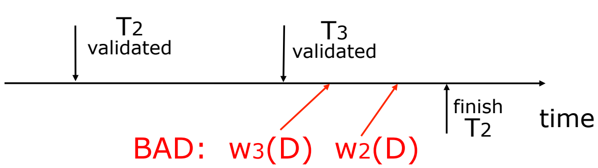
Example of What Validation Must ~~Prevent:~~ Allow

RS(T2)={B} $\cap$ RS(T3)={A,B} $\neq \emptyset$
WS(T2)={B,D} WS(T3)={C}



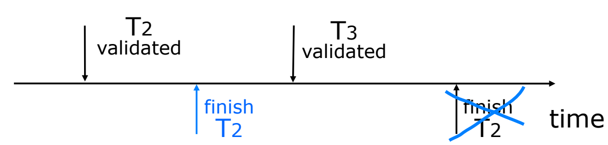
Another Thing Validation Must Prevent:

RS(T2)={A} RS(T3)={A,B}
WS(T2)={D,E} WS(T3)={C,D}



Another Thing Validation Must ~~Prevent:~~ Allow

RS(T2)={A} RS(T3)={A,B}
WS(T2)={D,E} WS(T3)={C,D}



Validation Rules For Tj:

(1) when Tj starts phase 1:
  ignore(Tj) ← FIN
(2) at Tj Validation:
  if check (Tj) then
   [VAL ← VAL U {Tj};
    do write phase;
    FIN ←FIN U {Tj}]

Check (Tj):

For Ti ∈ VAL – IGNORE (Tj) DO

IF [WS(Ti) ∩ RS(Tj) ≠ ∅ OR Ti ∉ FIN]

RETURN false;

RETURN true;

Is this check too restrictive ?

Improving Check(Tj)

For Ti ∈ VAL – IGNORE (Tj) DO

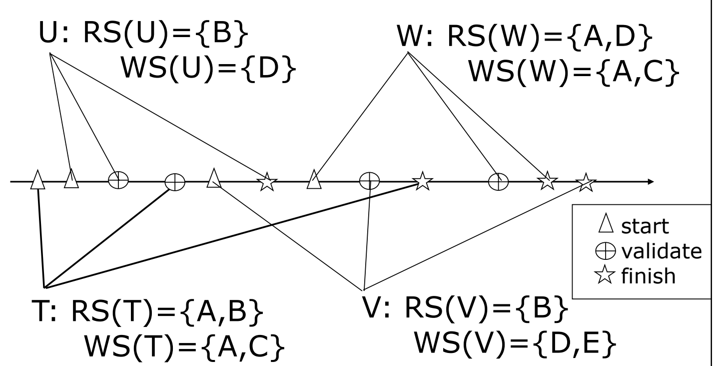
IF [WS(Ti) ∩ RS(Tj) ≠ ∅ OR

(Ti ∉ FIN AND WS(Ti) ∩ WS(Tj) ≠ ∅)]

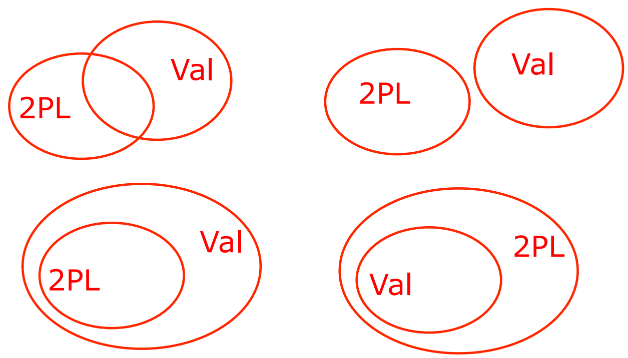
RETURN false;

RETURN true;

Exercise:



Is Validation = 2PL?



S2: w2(y) w1(x) w2(x)

- S2 can be achieved with 2PL:
  l2(y) w2(y) l1(x) w1(x) u1(x) l2(x) w2(x) u2(y) u2(x)
- S2 cannot be achieved by validation:
  The validation point of T2, val2 must occur before w2 (y) since transactions do not write to the database until after validation. Because of the conflict on x, val1 < val2, so we must have something like
  S2: val1 val2 w2(y) w1(x) w2(x)
  With the validation protocol, the writes of T2 should not start until T1 is all done with its writes, which is not the case.

Validation Subset of 2PL?

- Possible proof (Check!):
  - Let S be validation schedule
  - For each T in S insert lock/unlocks, get S':
    - At T start: request read locks for all of RS(T)
    - At T validation: request write locks for WS(T); release read locks for read-only objects
    - At T end: release all write locks
  - Clearly transactions well-formed and 2PL
  - Must show S' is legal (next page)

- Say S' not legal:
  S': ... l1(x) w2(x) r1(x) val1 u2(x) ...
  - At val1: T2 not in Ignore(T1); T2 in VAL
  - T1 does not validate: WS(T2) ∩ RS(T1) ≠ ∅
  - contradiction!
- Say S' not legal:
  S': ... val1 l1(x) w2(x) w1(x) u2(x) ...
  - Say T2 validates first (proof similar in other case)
  - At val1: T2 not in Ignore(T1); T2 in VAL
  - T1 does not validate:
    T2 ∉ FIN AND WS(T1) ∩ WS(T2) ≠ ∅)
  - contradiction!

Validation (also called optimistic concurrency control) is useful in some cases:

- Conflicts rare
- System resources plentiful
- Have real time constraints

Summary

Have studied concurrency control mechanisms used in practice

- 2PL
- Multiple granularity
- Tree (index) protocols
- Validation