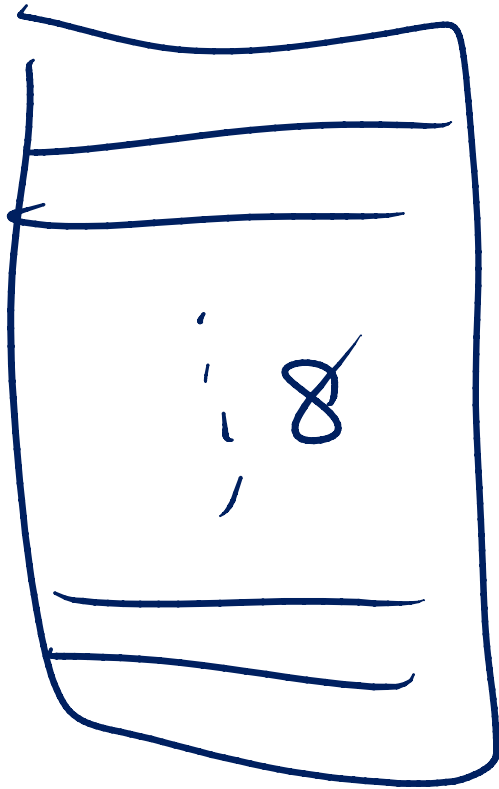
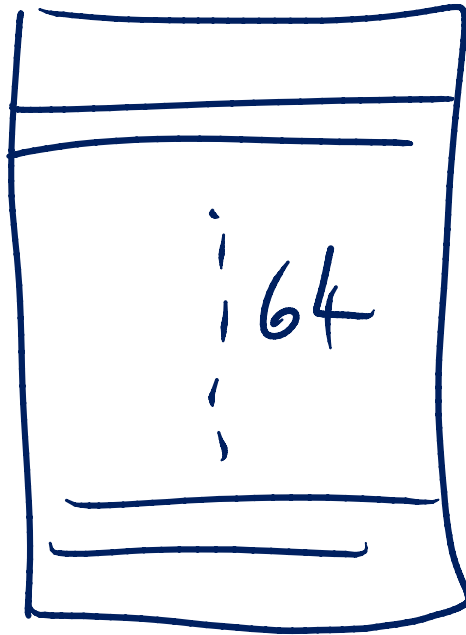
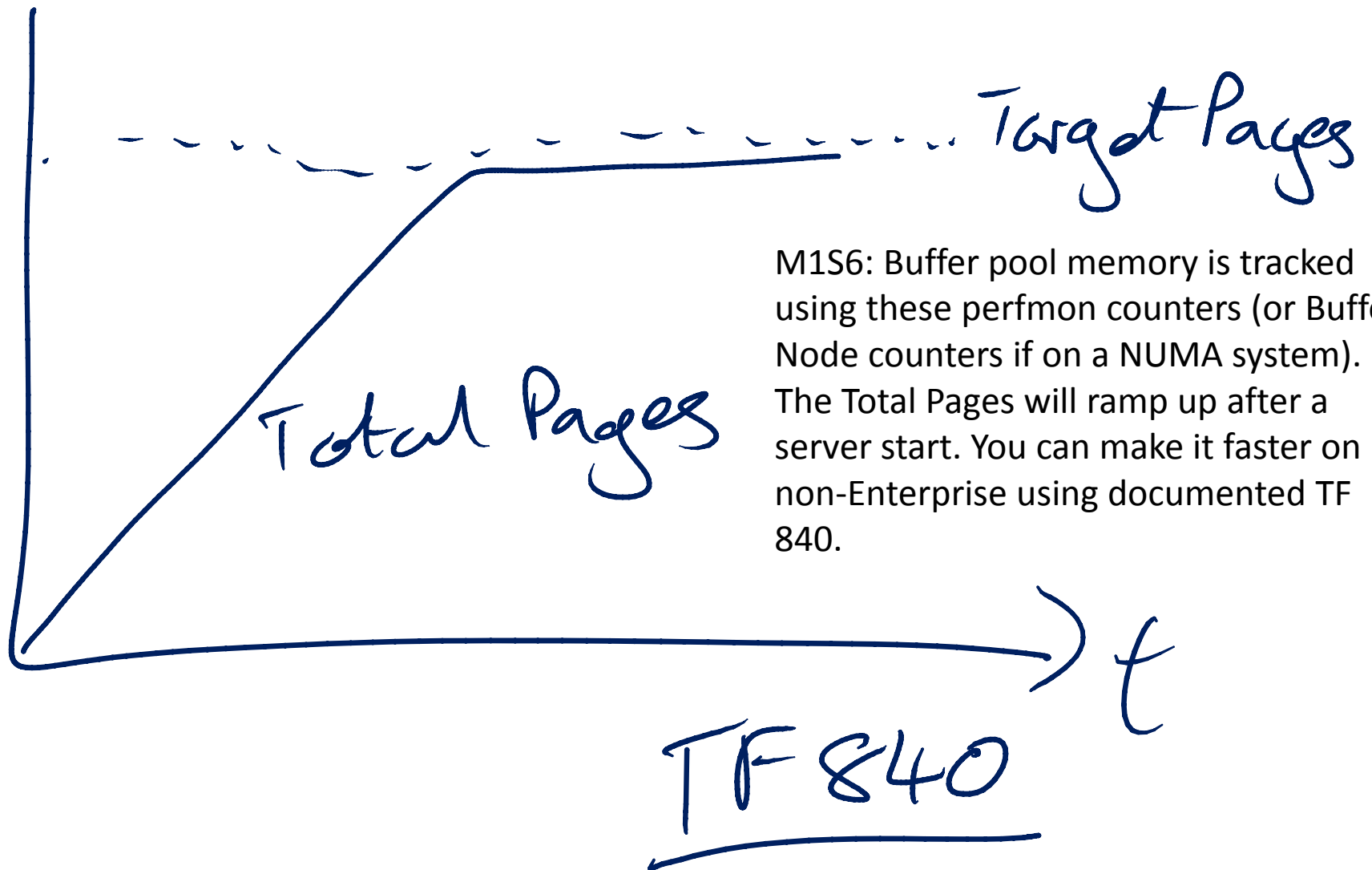




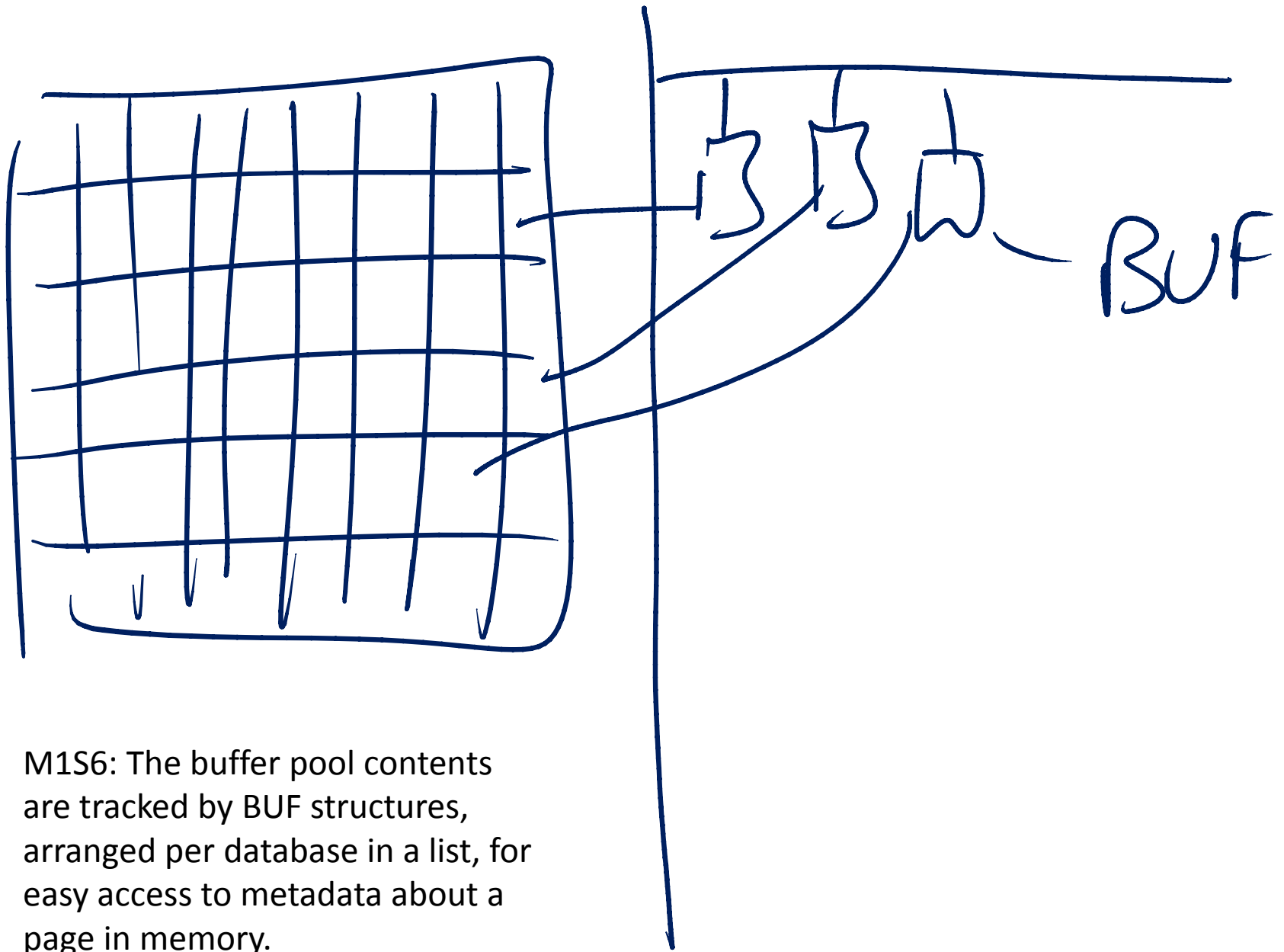
100 24 → OFF Row ^{0.5%}



M1S5 Having a large LOB value in row that's not used much makes for large rows and low data density. Choose how to store LOB data wisely.



M1S6: Buffer pool memory is tracked using these perfmon counters (or Buffer Node counters if on a NUMA system). The Total Pages will ramp up after a server start. You can make it faster on non-Enterprise using documented TF 840.



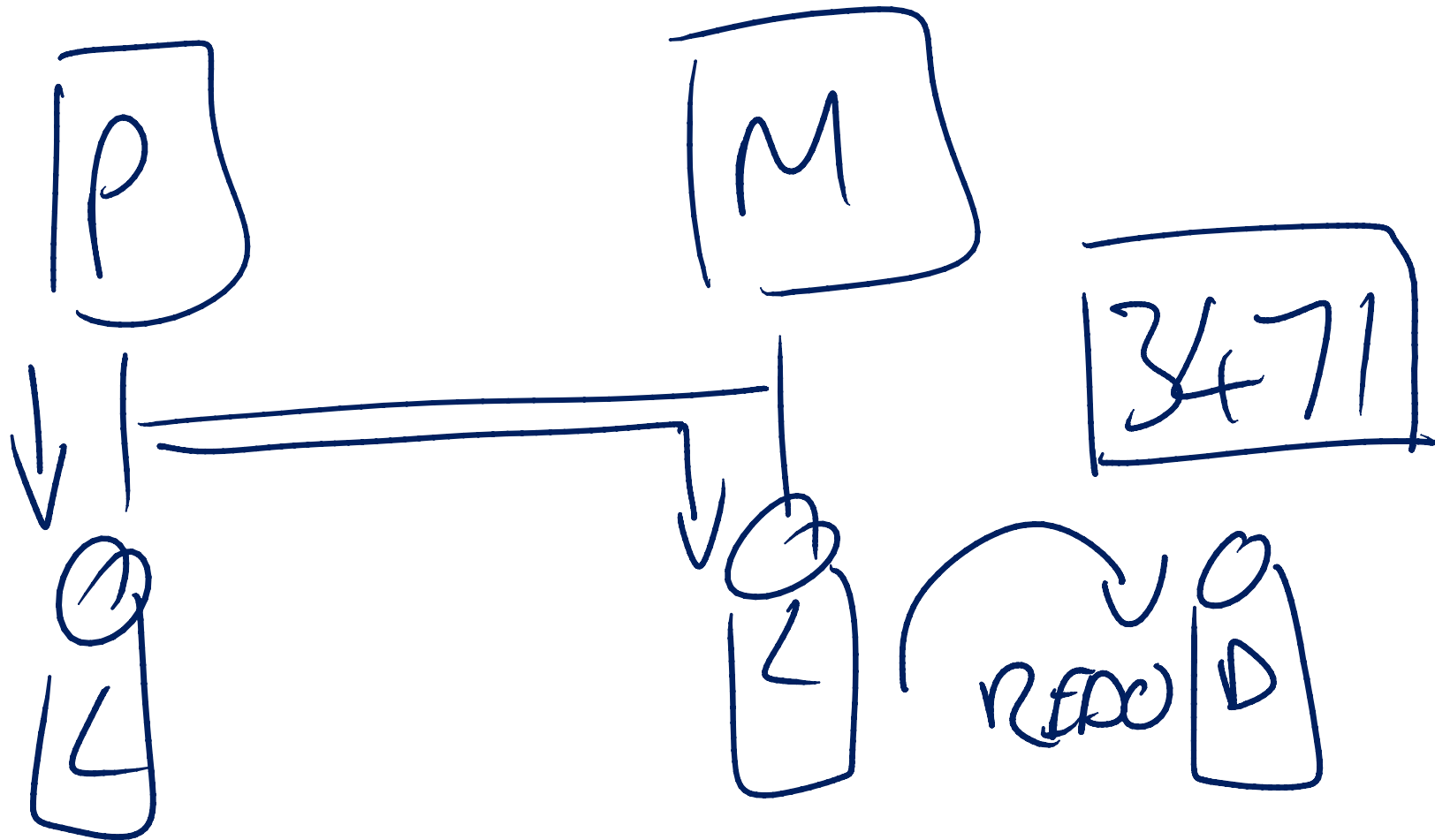
M1S6: The buffer pool contents are tracked by BUF structures, arranged per database in a list, for easy access to metadata about a page in memory.

BUFFER
DISFAVORING

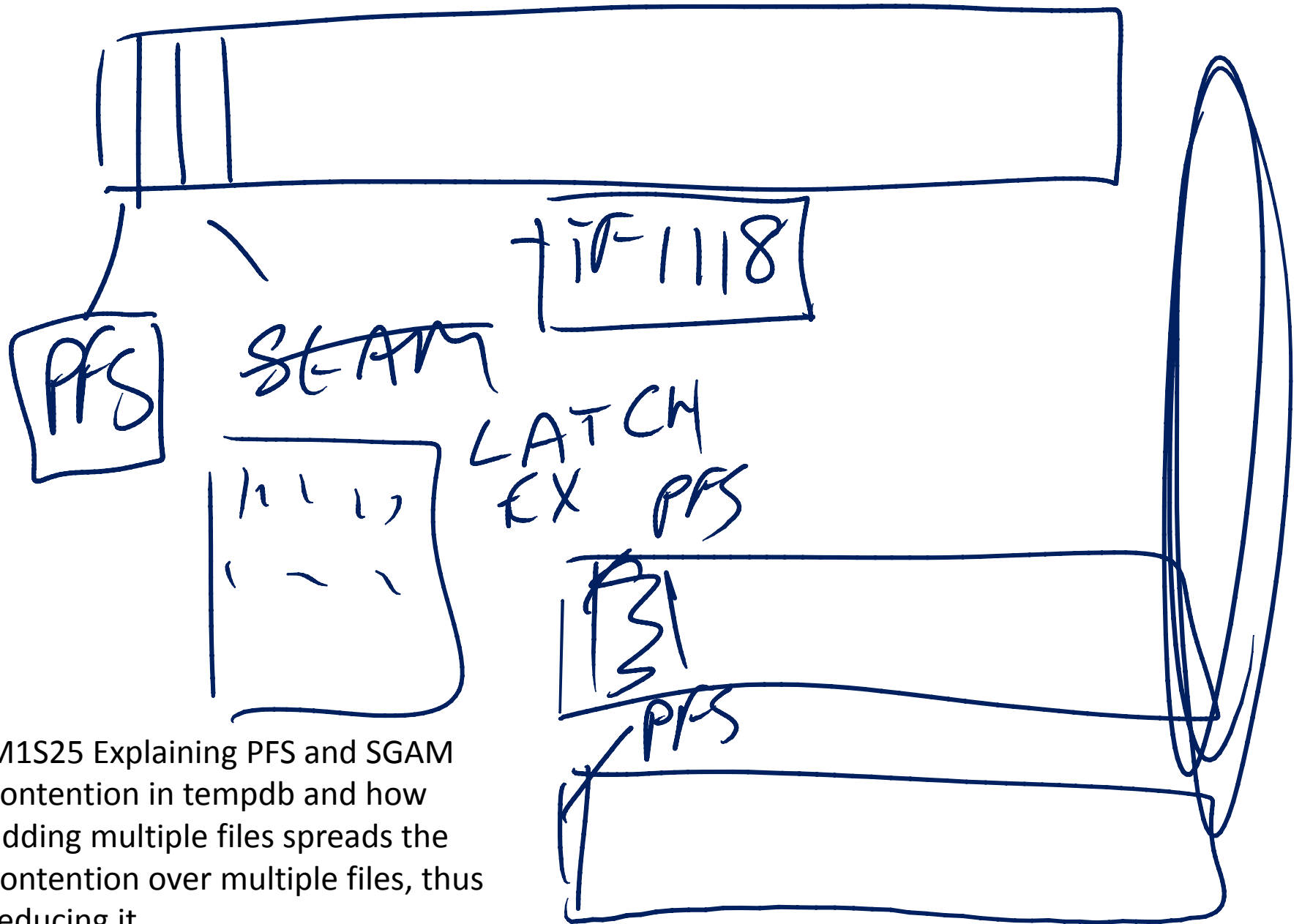
LRU-K, K=2



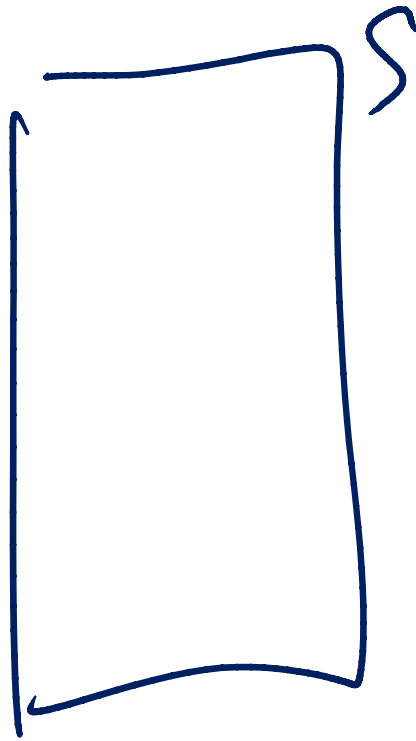
M1S6: Buffer pool implements a least-recently-used algorithm, to determine which pages are dropped when there's memory pressure. When memory pressure, lazy writer sweeps through buffers to find LRU ones. Disfavoring is discussed [here](#).



M1S13-14 Showing what happens when a transaction commits and the final log block for it is flushed to disk, when sync DBM or AG is involved. The local I/O will complete first but the transaction commit cannot complete until the mirror acknowledges the remote I/O. It's trace flag 3499, not 3471, that prevents excessive page flushes on the mirror.



M1S25 Explaining PFS and SGAM contention in tempdb and how adding multiple files spreads the contention over multiple files, thus reducing it.



RES	G.L.	P.Q
-----	------	-----

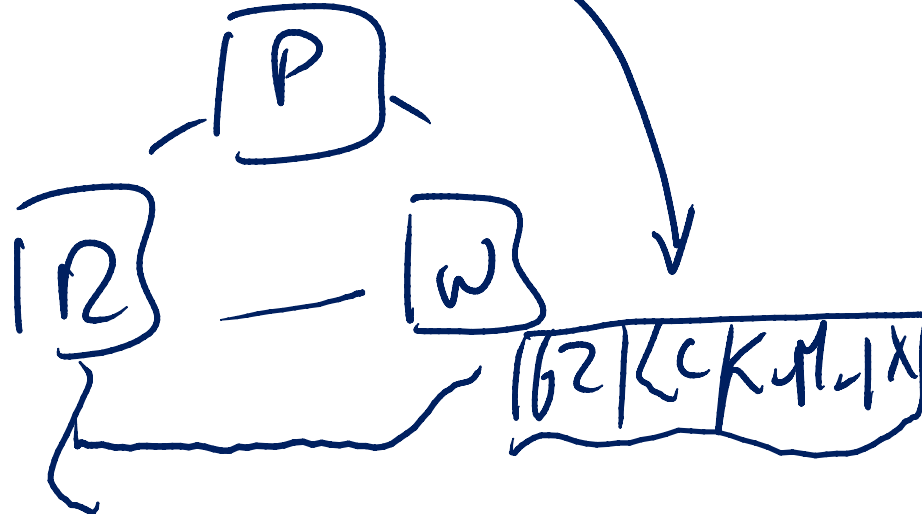
GRANTED
LIST

61	X
62	X

PENDING
QUEUE

62	X
64	X

M7: see next slide



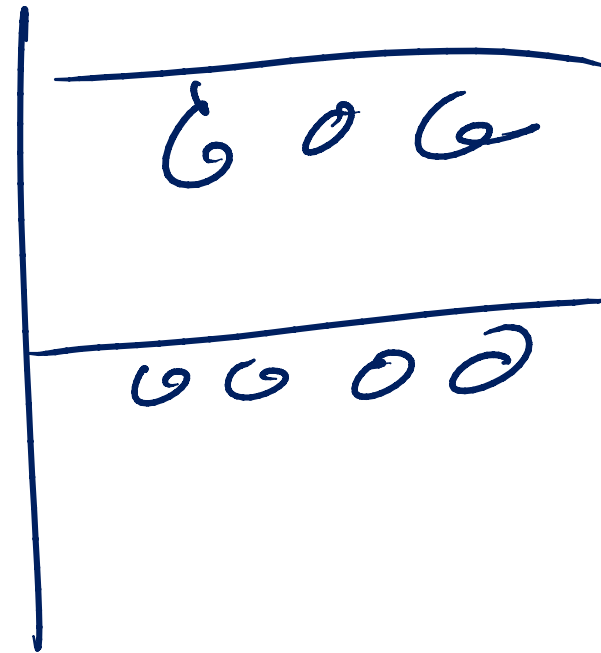
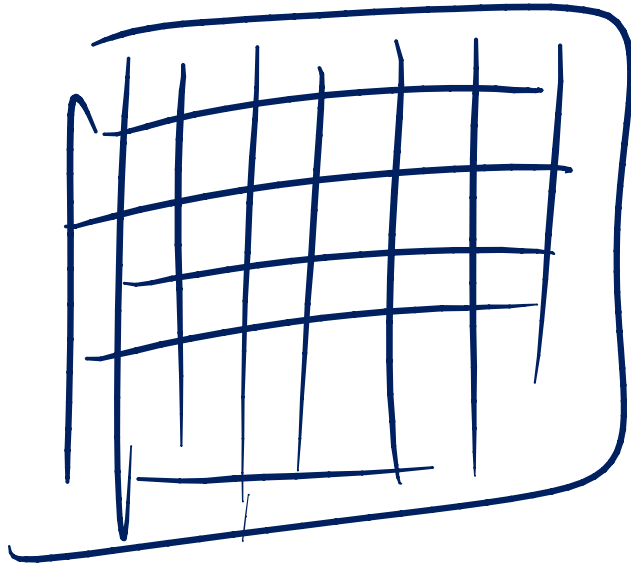
M7 Explaining how the lock pending queue works.

At time:

X: Thread 61 holds the lock in S mode

X+1: Thread 62 tries to acquire the lock in IX and can't. It registers a callback routine for itself on the pending queue in the lock value block. Then it suspends itself, moves off the CPU and onto the waiter list.

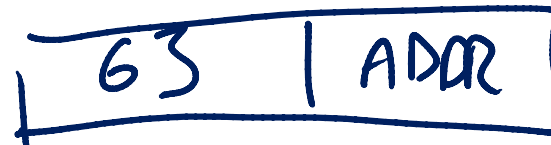
X+2: Thread 61 releases the lock. It checks the pending queue for locks that can now be granted ownership of the lock (taking into account any other threads still holding the lock). In this case, thread 62 is granted the lock in IX mode. Thread 62's callback routine is called, signaling it and moving it to the runnable queue on its scheduler.



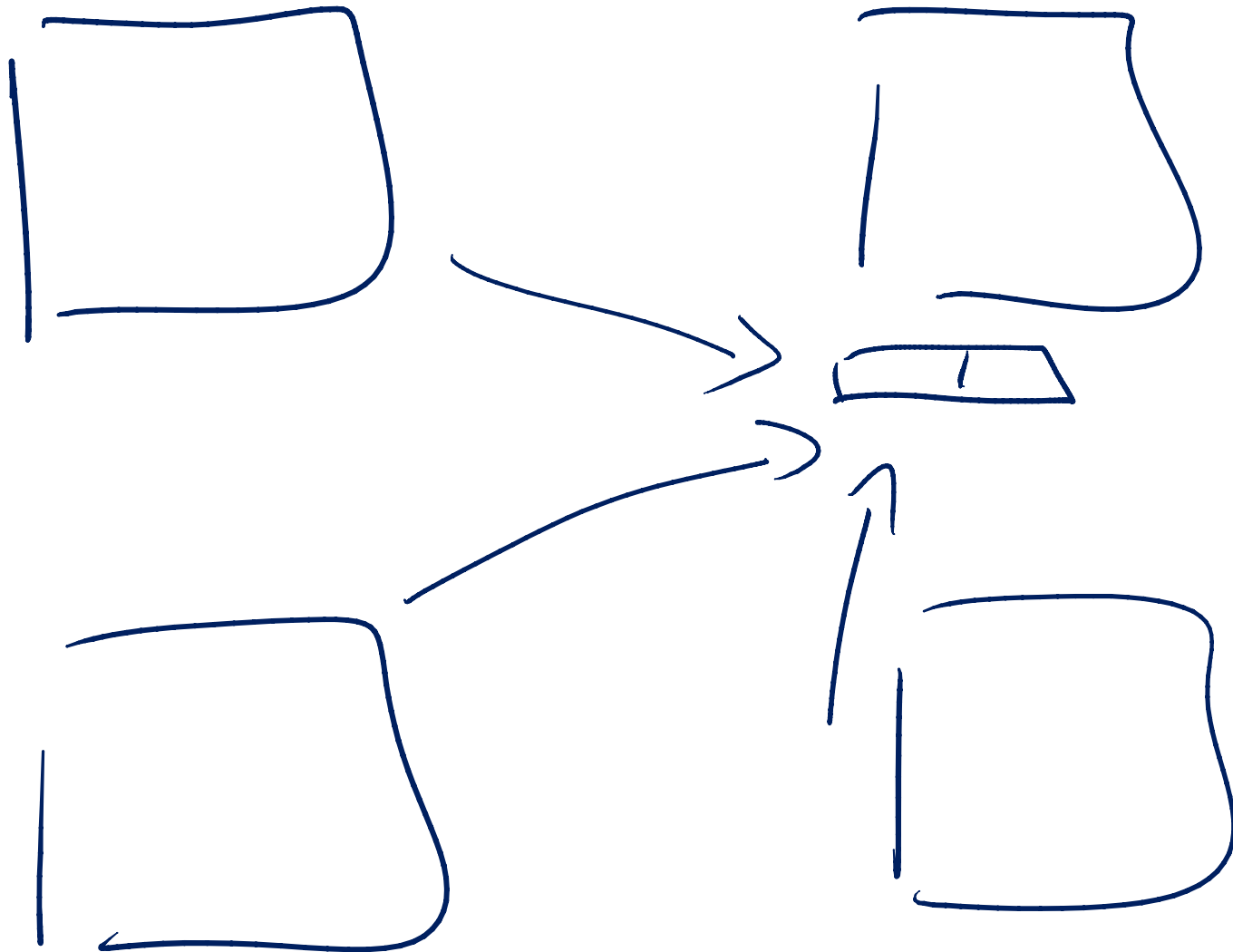
I/O COMPLETION
ROUTINE



ASYNC I/O

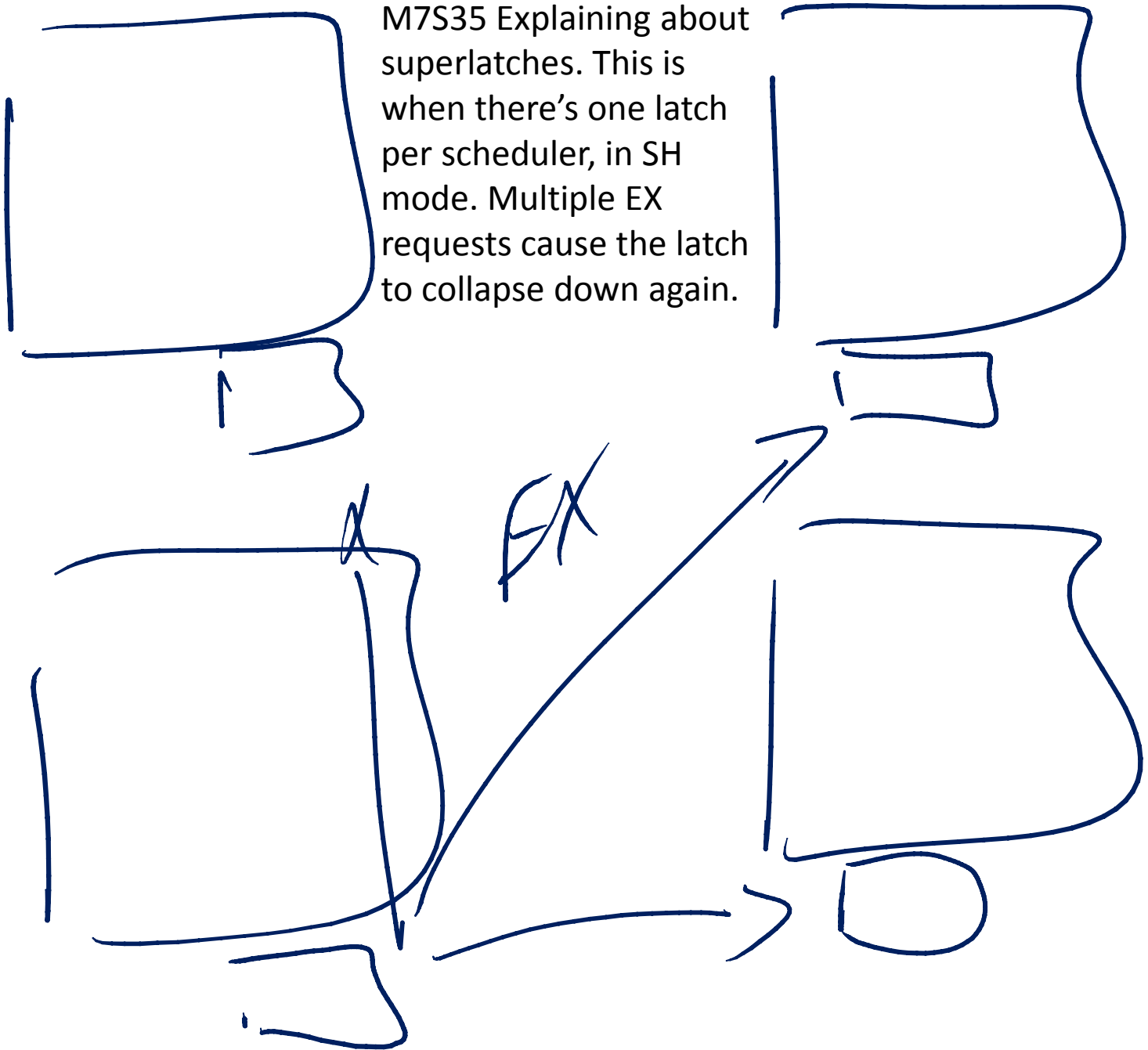


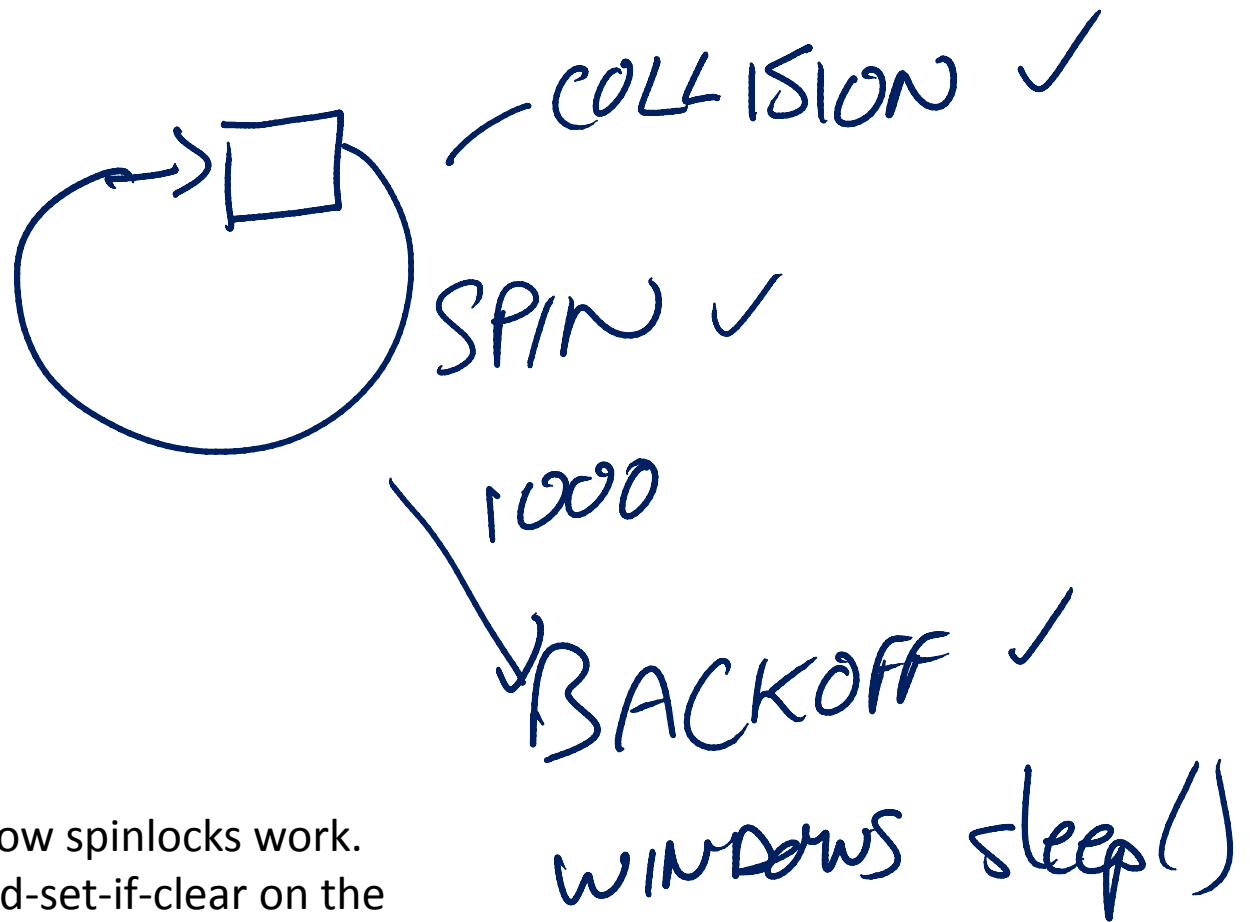
M7: A similar mechanism works for when a page must be read from disk.



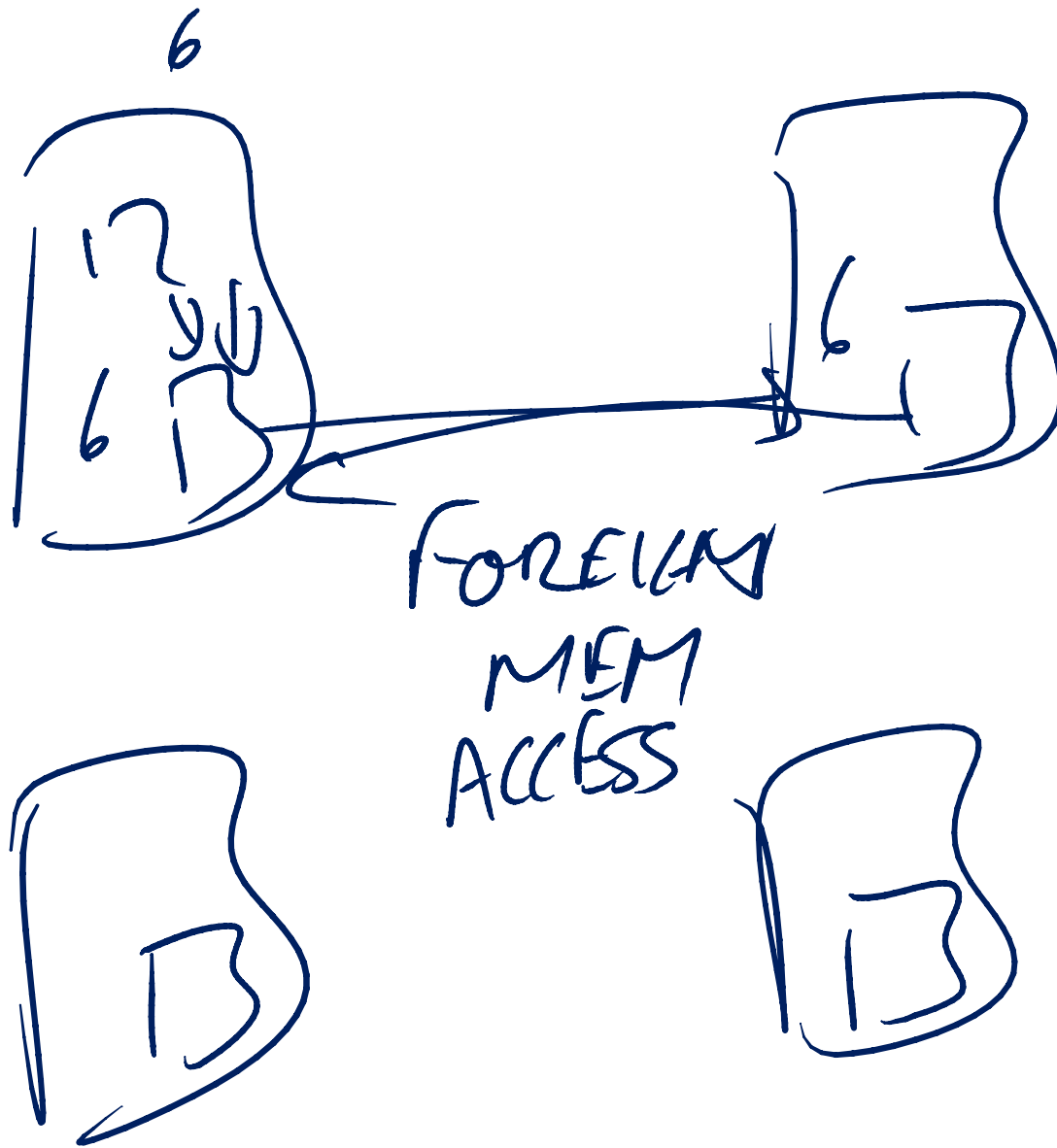
M7S35 Explaining about
superlatches. This is when
there's only one latch...

M7S35 Explaining about
superlatches. This is
when there's one latch
per scheduler, in SH
mode. Multiple EX
requests cause the latch
to collapse down again.





M7S38 Explaining how spinlocks work.
Code does a test-and-set-if-clear on the
memory that implements the spinlock. If
it can't get it, it spins (tries again) up to
1000 times and then backs off.



See next slide...

8

IX

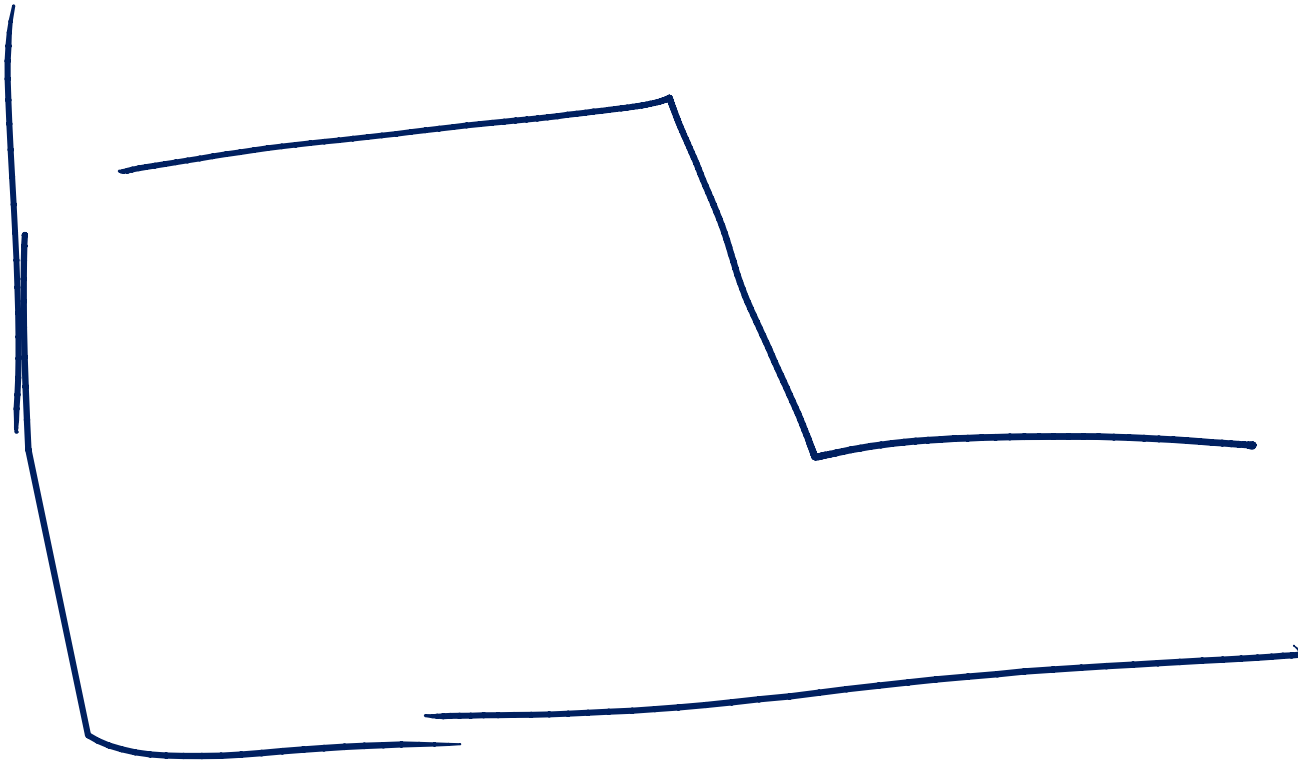
X	←
X	←
X	←
X	←
X	←
X	←
X	←

IDENTITY
TABLE IX

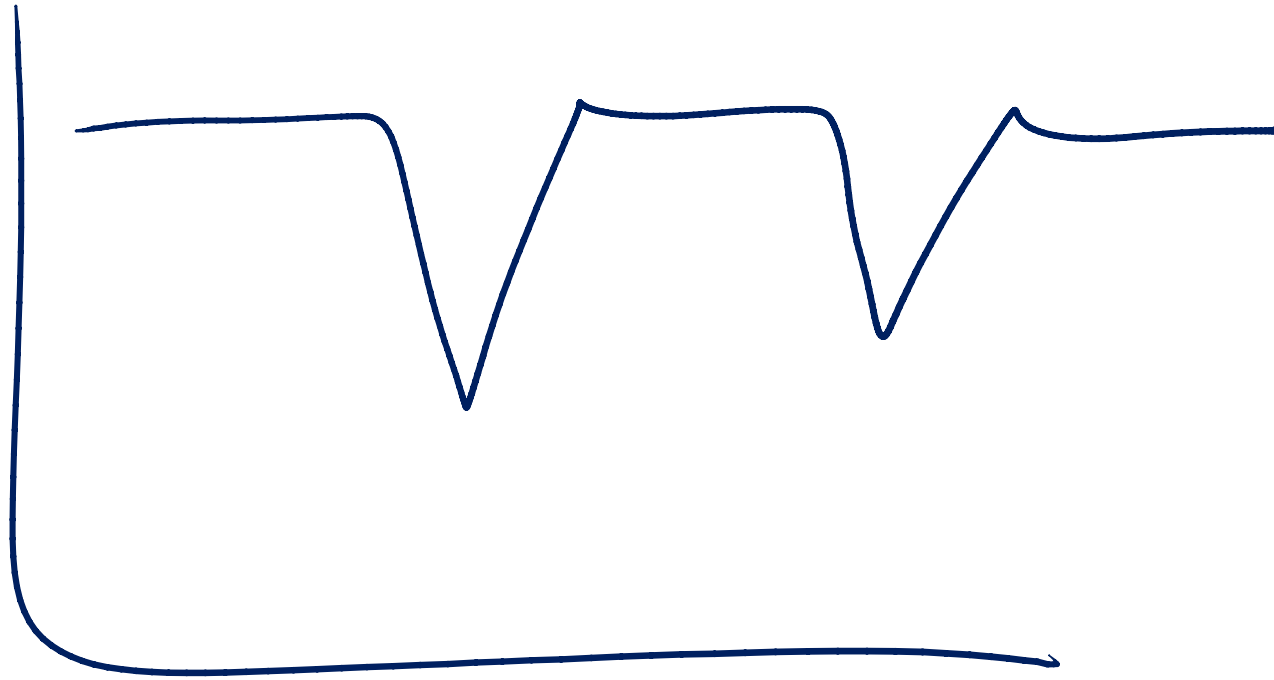
EX LATCH

M7S48 Explaining that to access a page to do an insert, multiple concurrent threads can hold non-conflicting row X locks, but they all need to acquire the EX latch on the page, which becomes a bottleneck. If the clustered index has an int/bigint identity, and the row size is small enough, it could result in an insert hotspot.

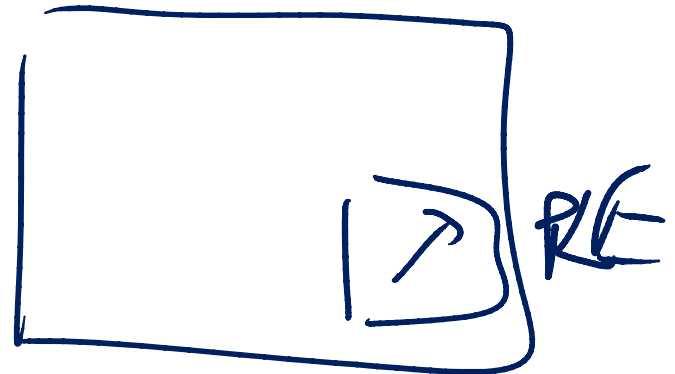
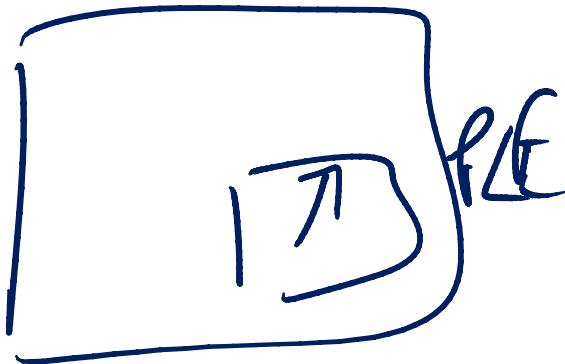
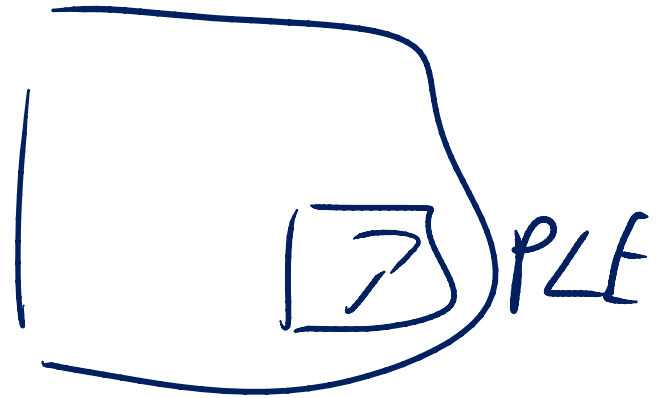
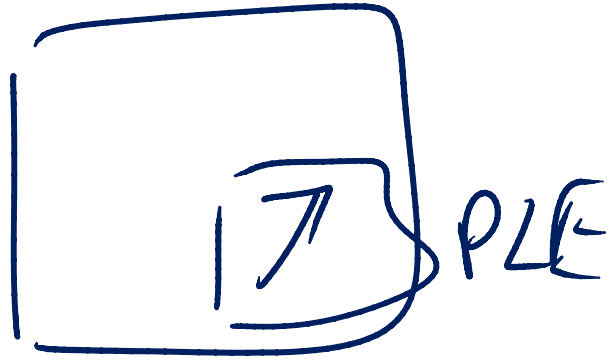
$$\underline{PLE = 300}$$



PLE tangent: There is no one-size-fits-all PLE threshold. When your PLE drops from your norm, and stays low, you have a problem.



PLE tangent: When PLE does this, it's normal.



PLE tangent: On NUMA, you need to watch each Buffer Node PLE, not the Buffer Manager PLE. See <http://www.sqlskills.com/blogs/paul/page-life-expectancy-isnt-what-you-think/>