



MAKLEE

software engineering
solutions

Implementing Oracle Rdb Row Caches

Norman Lastovica
Executive Vice President
Maklee Engineering
norman.lastovica@maklee.com

Agenda

- › Row Cache Review
- › Deciding What To Cache
- › Configuring Caches
- › Results



Row Cache Review

- › Copies of database rows in memory
- › Locking & IO Reduction
 - Fetch/Modify cached rows with no database IO or page locks
 - Doesn't help sequential scans – can even hurt
- › Requires database open on single computer



Deciding What To Cache

- › Often follows “80/20” rules of thumb
- › Start with “Busy” sorted indexes & “Hot” tables
- › Snapshots in cache for tables/indexes with high update activity
- › How much memory is available; how bad is IO and page locking?



Snapshots In Cache

- › Specify count of snapshot slots
 - May be less or more than the cache slot count
- › Start by evaluating relationship between storage area size and snapshot storage area size
- › Performance penalty if too small



Sweeping

- › Write some of modified rows from cache to database
- › Triggered upon cache full
- › Optionally via periodic timer



Checkpointing

- › Write all modified rows from cache to disk
 - Suggested to backing store with snapshots in cache, to database otherwise
- › RCS checkpoint timer
- › Per-process - evaluated at end of transaction
 - Fast Commit timer – 5 to 10 minutes?
 - Fast Commit transactions – 1000?
 - Fast Commit AIJ growth – 500,000 blocks?



Checkpointing

- › RMU /CHECKPOINT
 - Global checkpoint
- › RMU /SERVER RECORD_CACHE CHECKPOINT
 - /WAIT
 - /LOG
- › Checkpoint at database close, backup, & verify



CASE STUDY



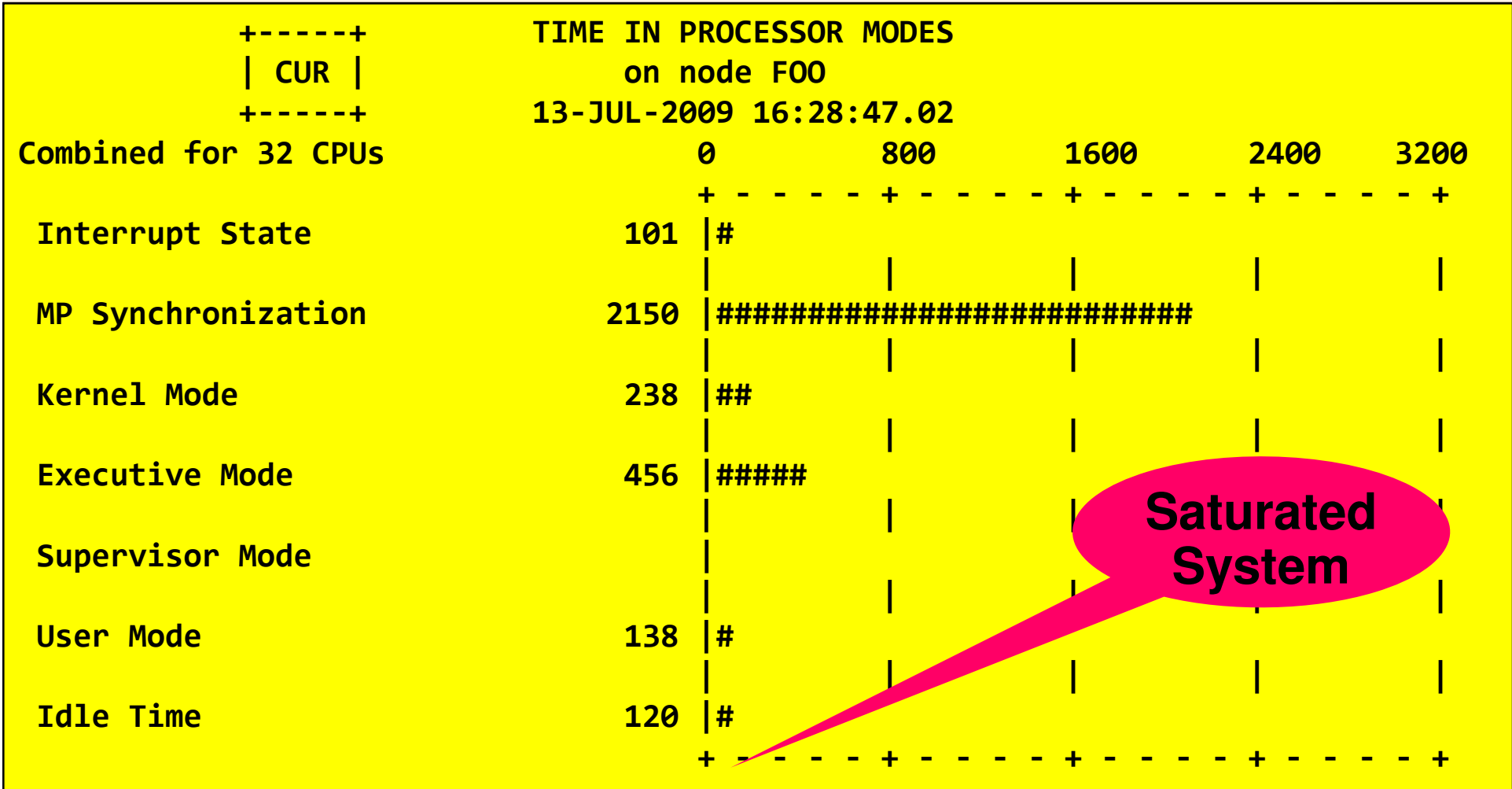


Case Study

- › Production system
- › “It is slow”
- › “Backups are slow”
- › Cluster of 164 rx8640 with high-end storage



21 of 32 CPUs in MPSYCH



170,000+ Locking Operations Per Second



LOCK MANAGEMENT STATISTICS

on node F00

13-JUL-2009 16:41:10.72

	CUR	AVE	MIN	MAX
New ENQ Rate	73395.00	73395.00	73395.00	73395.00
Converted ENQ Rate	23417.00	23417.00	23417.00	23417.00
DEQ Rate	69951.66	69951.66	69951.66	69951.66
Blocking AST Rate	1103.00	1103.00	1103.00	1103.00
ENQs Forced To Wait Rate	846.66	846.66	846.66	846.66
ENQs Not Queued Rate	3723.66	3723.66	3723.66	3723.66
Deadlock Search Rate	0.00	0.00	0.00	0.00
Deadlock Find Rate	0.00	0.00	0.00	0.00
Total Locks	509849.00	509849.00	509849.00	509849.00
Total Resources	271984.00	271984.00	271984.00	271984.00



120 TPS

110 I/O Per Transaction

Node: F00 (1/1/2) Oracle Rdb V7.2-350 Perf. Monitor 15-JUL-2009 17:13:38.85
Rate: 3.00 Seconds Summary IO Statistics Elapsed: 00:41:48.89
Page: 1 of 1 DSA35:[000000.DATABASE]DB.RDB;4 Mode: Online

statistic.....	rate.per.second.....			total.....	average.....
name.....	max.....	cur.....	avg.....	count.....	per.trans....
transactions	785	82	117.7	295280	1.0
verb successes	257550	6984	12916.6	32401684	109.7
verb failures	1	0	0.0	125	0.0
synch data reads	129529	10710	9721.2	24385844	82.5
synch data writes	1229	42	127.2	319143	1.0
asynch data reads	46100	2749	3298.1	8273497	28.0
asynch data writes	1801	104	184.8	463779	1.5
RUJ file reads	6	0	0.1	327	0.0
RUJ file writes	155	14	16.4	41308	0.1
AIJ file reads	215	0	1.2	3191	0.0
AIJ file writes	133	11	16.9	42528	0.1
root file reads	147	0	0.3	902	0.0
root file writes	1551	154	222.7	558997	1.8



Top Indexes

Node: F00 (1/1/2) Oracle Rdb V7.2-350 Perf. Monitor 15-JUL-2009 17:12:07.19
Rate: 3.00 Seconds Logical Area Overview (Btree Indexes) Elapsed: 00:40:17.24
Page: 1 of 57 DSA35:[DATABASE]DB.RDB;4 Mode: Online

```
-----  
Logical.Area.Name..... leaf fetches leaf inserts leaf removal discarded  
XPK_NR_ACC_ANALYTICAL2          54724472          32           0           0  
SI_NR_ACCOUNT_ANALYT_BANK       14281695          32           0           0  
XPK_ORGSTRUCTURE                 11775100           2           0           0  
SI_KR_RADM_ROLE                  10598169           0           0           0  
SI_NR_ACC_ANALYT_NUM             5102910           32           0           0  
XPK_KBK_ADM_RECEIVER            4884833            0           0           0  
XPKCD_CLIENT_INFO                3644873           30           0           0  
SI_ANK1_1                        3626054           53           0           0  
I_H_CRED                         3344220            0           0           0  
SI_NR_DOC_DATE2                  3222071           6808         4587         0  
SI_NR_DOC_COMPLEX_DATE           3087472           8442         4005         0  
SI_NR_KRED_REQUEST_PHYS          2713654            3           0           0  
SIR_ACCOUNT_ANALYTICAL_1        2468185           32           0           0
```



SETUP FOR CACHING



Configure Database

```
ALTER DATABASE FILENAME FOO  
  NUMBER OF CLUSTER NODES IS 1  
  RESERVE <n> CACHE SLOTS  
  ROW CACHE IS ENABLED (...)  
  JOURNALING IS ENABLED  
  (FAST COMMIT IS ENABLED ...)
```



Configuring Checkpointing

```
ALTER DATABASE FILENAME FOO
  JOURNALING IS ENABLED (
    FAST COMMIT IS ENABLED (
      CHECKPOINT EVERY <t> TRANSACTIONS,
      CHECKPOINT TIMED EVERY <s> SECONDS,
      CHECKPOINT INTERVAL IS <b> BLOCKS));
```

```
ALTER DATABASE FILENAME MF_PERSONNEL
  ROW CACHE IS ENABLED (
    CHECKPOINT TIMED EVERY <m> SECONDS,
    SWEEP INTERVAL IS <n> SECONDS);
```



Configuring Database/System

- › Grant VMS\$MEMORY_RESIDENT_USER to users who open databases or create caches
 - When SHARED MEMORY IS PROCESS RESIDENT
- › Consider RDM\$BUGCHECK_IGNORE_FLAGS
 - LGR



Configuring Caches

- › Slot size
- › Slot count
- › Snapshots in cache



Row / Node Size

- > Alternately let Rdb figure it out for you

```
$ RMU /SHOW AIP /BRIEF FOO EMPID
```

```
*-----*  
* Logical Area Name                LArea PArea   Len Type  
*-----*  
EMPID                               158   13   430 UNKNOWN
```

Be careful of non-ranked sorted indexes allowing duplicates



Index Node Count

```
$ RMU /ANALYZE /INDEX FOO
```

```
Index PLUGH for relation XYZZY duplicates allowed  
Max Level: 8, Nodes: 9399698,  
Used/Avail: 2237181111/3741079804 (59%),  
Keys: 117076163,  
Records: 107604684
```





An Aside – How Would You Re-Create Indexes?

1. Drop existing index
2. Create new index

-OR-

1. Create new index
2. Drop existing index

-OR-

1. Drop existing index
2. Disconnect/Reconnect
3. Create new index



Build Cache Creation Spreadsheet

- › One row per alter cache command
- › Easy to cut-n-paste
- › A sufficiently fancy spreadsheet or command procedure could do slot count analysis from output of RMU /ANALYZE /INDEX



Build Cache Creation Spreadsheet

ALTER DATABASE FILE X\$ alter cache	SI_NR_ACCOUNT_ANALYT_BANK	CACHE SIZE IS	75000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	XPB_NR_ACC_ANALYTICAL2	CACHE SIZE IS	75000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SIR_DOC_CARTOTEC_HIST_1	CACHE SIZE IS	30000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	XPB_ORGSTRUCTURE	CACHE SIZE IS	8193	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	XPB_NR_DOC_COMPLEX1	CACHE SIZE IS	10421241	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_NR_ACC_ANALYT_NUM	CACHE SIZE IS	75000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SIR_ACCOUNT_ANALYTICAL_1	CACHE SIZE IS	12000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_NR_DOC_DATE2	CACHE SIZE IS	7670707	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	XPB_NR_DOC	CACHE SIZE IS	3097377	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	I_R_LS_NUMBER_2	CACHE SIZE IS	600000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	I_R_LS_NUMBER_1	CACHE SIZE IS	200000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_LS_2	CACHE SIZE IS	300000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_KR_RADM_ROLE	CACHE SIZE IS	200	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_NR_KRED_REQUEST_PHYS	CACHE SIZE IS	5000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_ANK1_1	CACHE SIZE IS	75000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_ILS_1	CACHE SIZE IS	8324497	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	KR_DOC_COMPLEX_CLSB_DC	CACHE SIZE IS	150000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_R_ZCH_TOPSPIS_PRIM	CACHE SIZE IS	5000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_NR_DOC_COMPLEX	CACHE SIZE IS	20934080	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT
ALTER DATABASE FILE X\$ alter cache	SI_NR_BANDEROL_K8	CACHE SIZE IS	130000	rows	ROW LENGTH IS	430	BYTES CHECKPOINT UPDAT



Create Cache

```
ALTER DATA FILE FOO
  ADD CACHE PLUGH
    CACHE SIZE IS 9999999 ROWS
    ROW LENGTH IS 430 BYTES
    SHARED MEMORY IS PROCESS RESIDENT
    CHECKPOINT UPDATED ROWS TO DATABASE
    ROW SNAPSHOT IS ENABLED (
      CACHE SIZE IS 99999 ROWS)
```



How Much To Cache

- › Use your memory
- › Points of diminishing return
 - Fix the worst offenders
- › RMU /DUMP /HEADER
 - “Shared memory section requirement”



Ease of Use

- > **RMU /SET ROW_CACHE**
 - **/ENABLE**
 - **/DISABLE**
 - **/SHARED_MEMORY**

- > **RMU /CLOSE /WAIT**
 - If required, use **/ABORT=FORCEX** of user processes

- > **RMU /OPEN /WAIT**

- > **RMU /CHECKPOINT**



RESULTS





Results

- › Look at both “whole day” and “interval” statistics
- › There is always more tuning possible

Accept that some days you are the
pigeon, and some days you are
the statue



Excellent Hit Rates ... A Few Problems

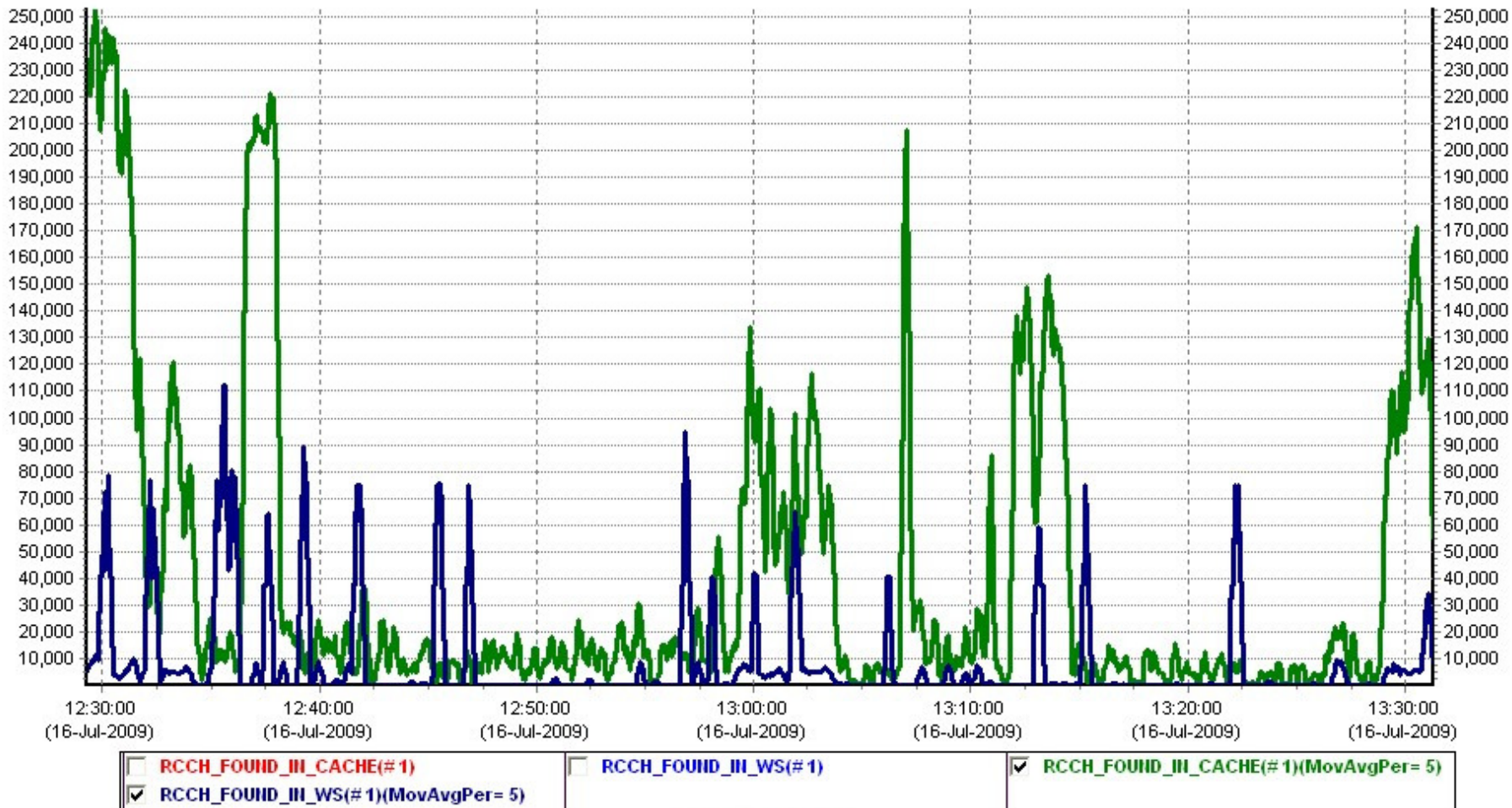
Node: F00 (1/1/1) Oracle Rdb V7.2-350 Perf. Monitor 17-JUL-2009 10:16:31.18
 Rate: 3.00 Seconds Row Cache Overview (Unsorted) Elapsed: 02:50:21.66
 Page: 1 of 1 DSA35:[VOX_DB]VOX.RDB;1 Mode: Online

Cache.Name.....	#Searches	Hit%	Full%	#Inserts	#Wrap	#Slots	Len
TWISTY_PASSAGE	250887097	99.9	72.3	54276	0	75000	432
SI_NR_ACC_ANALYT_NUM	58139058	99.9	61.0	45777	0	75000	432
XPK_KBK_ADM_REC	9841561	99.9	55.5	285	0	513	432
PLOVER	768137386	99.9	76.1	57100	0	75000	432
SI_NR_ACC_AN	2075325	99.8	33.9	2782	0	8193	432
SI_NR_ACC_AN_GRO	5536387	99.0	73.6	55227	0	75000	432
SI_NR_RED_REQ	20579202	99.9	41.1	2058	0	5000	432
XPK_ORGST	57380427	99.9	60.0	4920	0	8193	432
SI_KR_RADM	24873333	99.9	37.5	75	0	200	432
XPKCD_INFO	2263936	99.8	87.5	3585	0	4097	432
END_OF_ROAD	2266	74.8	0.0	570	0	600000	432
SI_ANK1_1	12551786	99.5	70.3	52755	0	75000	432
SI_CR_CB	1000646	99.7	71.0	2910	0	4097	432
DUSTY	106262174	0.0	100.0	106234285	47	3	432





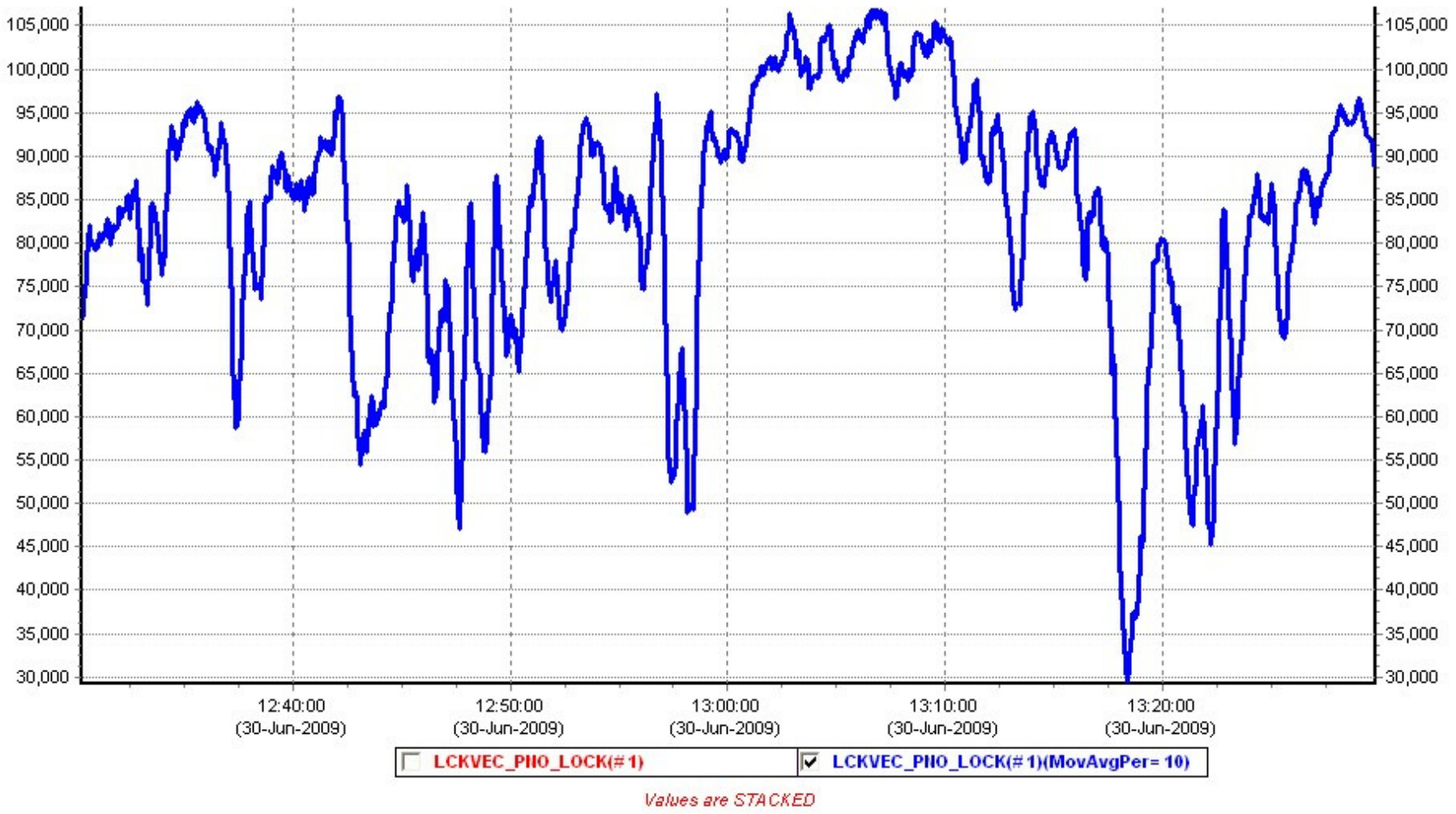
250,000 Cache Hits Per Second



Values are STACKED

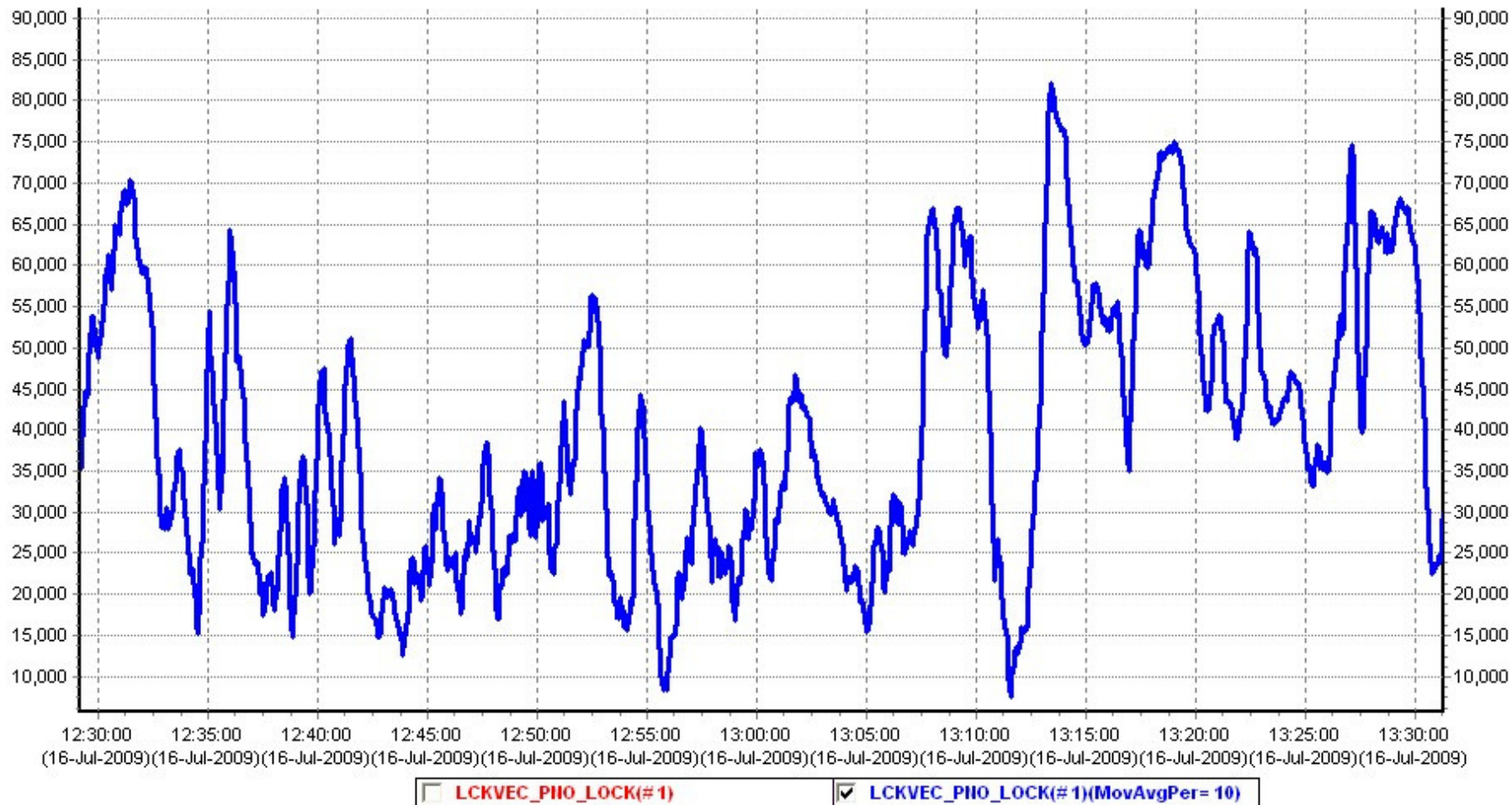


Month Before: Page Locks Acquire Per Second

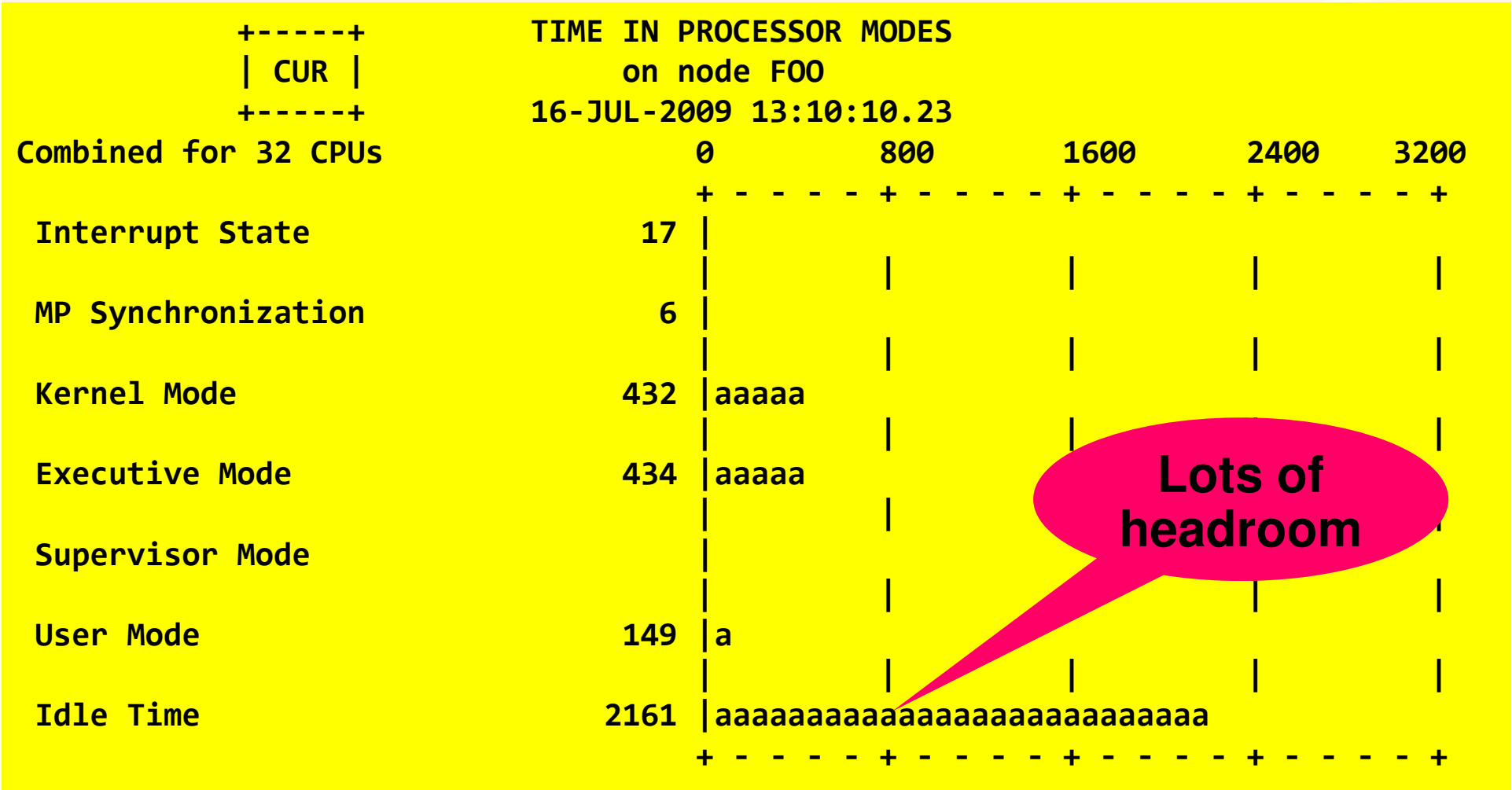




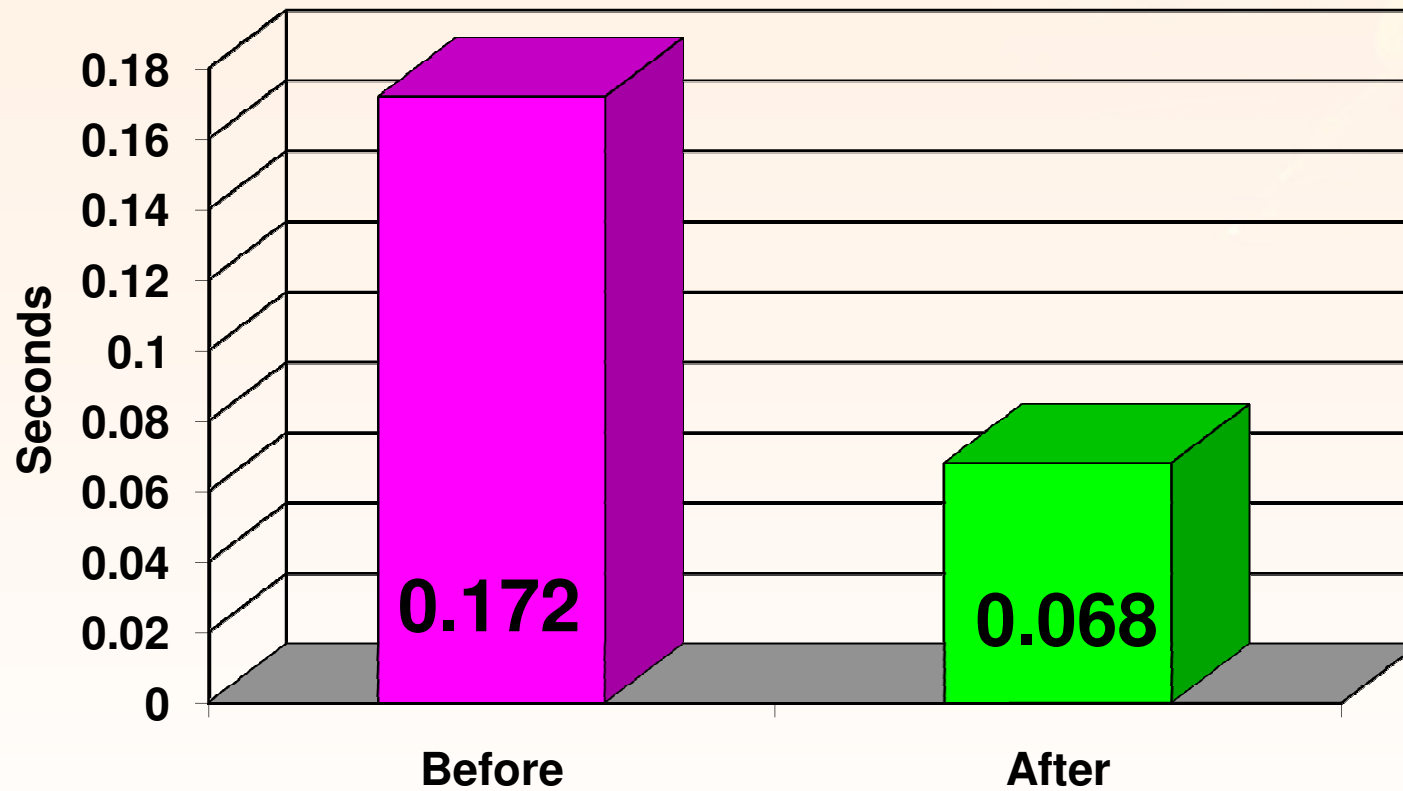
After: Page Locks Acquire Per Second



21 Idle CPUs



Average Transaction Duration



Test, Analyze, Change, Repeat

- › Find bottlenecks and work on them
 - Application
 - Database
 - System



Evaluate

- › See how the caches are doing
- › Do not look at every problem as a nail

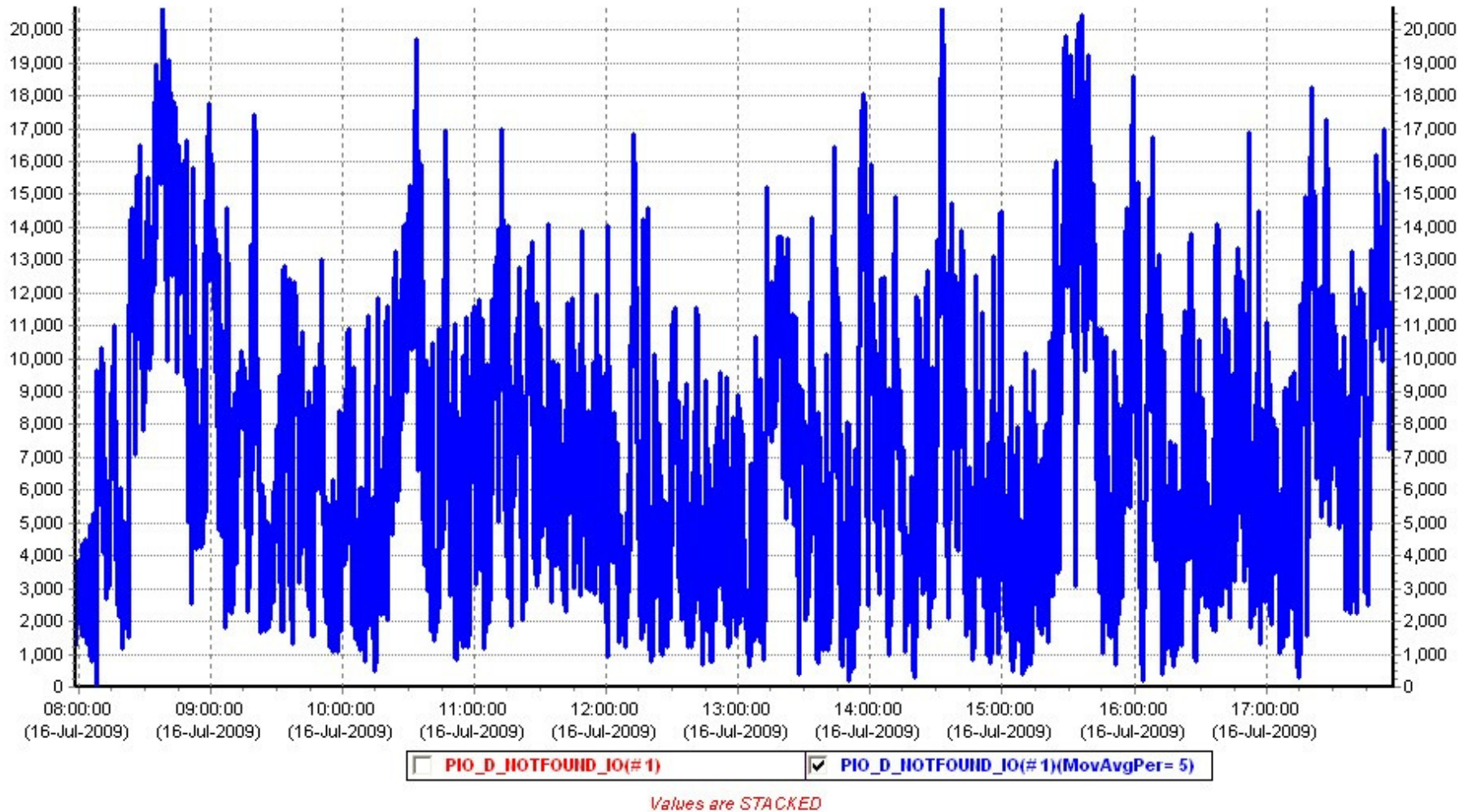


Final Thoughts

- › Row Cache & Hot Standby database
- › Locking & IO reduction
- › CPU consumption increases due to reduced waiting
- › Re-open after “node failure”



Plenty of Work Always Remains



If at first you don't succeed, you're about average