
MySQL™ Connect

MySQL Performance: Benchmarks, Tuning and “Best” Practices..

Dimitri KRAVTCHUK
MySQL Performance Architect @Oracle

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract.
It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Are you Dimitri?..

- Yes, it's me :-)
- Hello from Paris! ;-)
- Passionated by Systems and Databases Performance
- Previous 15 years @Sun Benchmark Center
- Started working on MySQL Performance since v3.23
- But during all that time just for fun only ;-)
- Since last years officially @MySQL Performance full time now
- http://dimitrik.free.fr/blog / @dimitrik_fr

Agenda

- Overview of MySQL Performance
- Workload oriented tuning and MySQL Internals
- Performance improvements in MySQL 5.6 & Benchmark results
- Pending issues..
- Progress in MySQL 5.7 Performance
- Q & A

Why MySQL Performance ?...

Why benchmarking MySQL?..

- Any solution may look “good enough”...

ORACLE

Why benchmarking MySQL?..

- Until it did not reach its limit..

ORACLE

Why benchmarking MySQL?..

- And even improved solution may not resist to increasing load..

Why benchmarking MySQL?..

- And reach a similar limit..

Why benchmarking MySQL?..

- A good benchmark testing may help you understand ahead the resistance of your solution to incoming potential problems ;-)

Why benchmarking MySQL?..

- But keep it in mind:
  - Even a very powerful solution but leaved in wrong hands may still be easily broken!... :-)

The Main MySQL Performance Tuning
#1 Best Practice is... ???..

The Main MySQL Performance Tuning
#1 Best Practice is... ???..

USE YOUR BRAIN !!! :-)

The Main MySQL Performance Tuning
#1 Best Practice is... ???..

USE YOUR BRAIN !!! :-)

IN FACT
THIS IS THE
MAIN SLIDE! ;-))

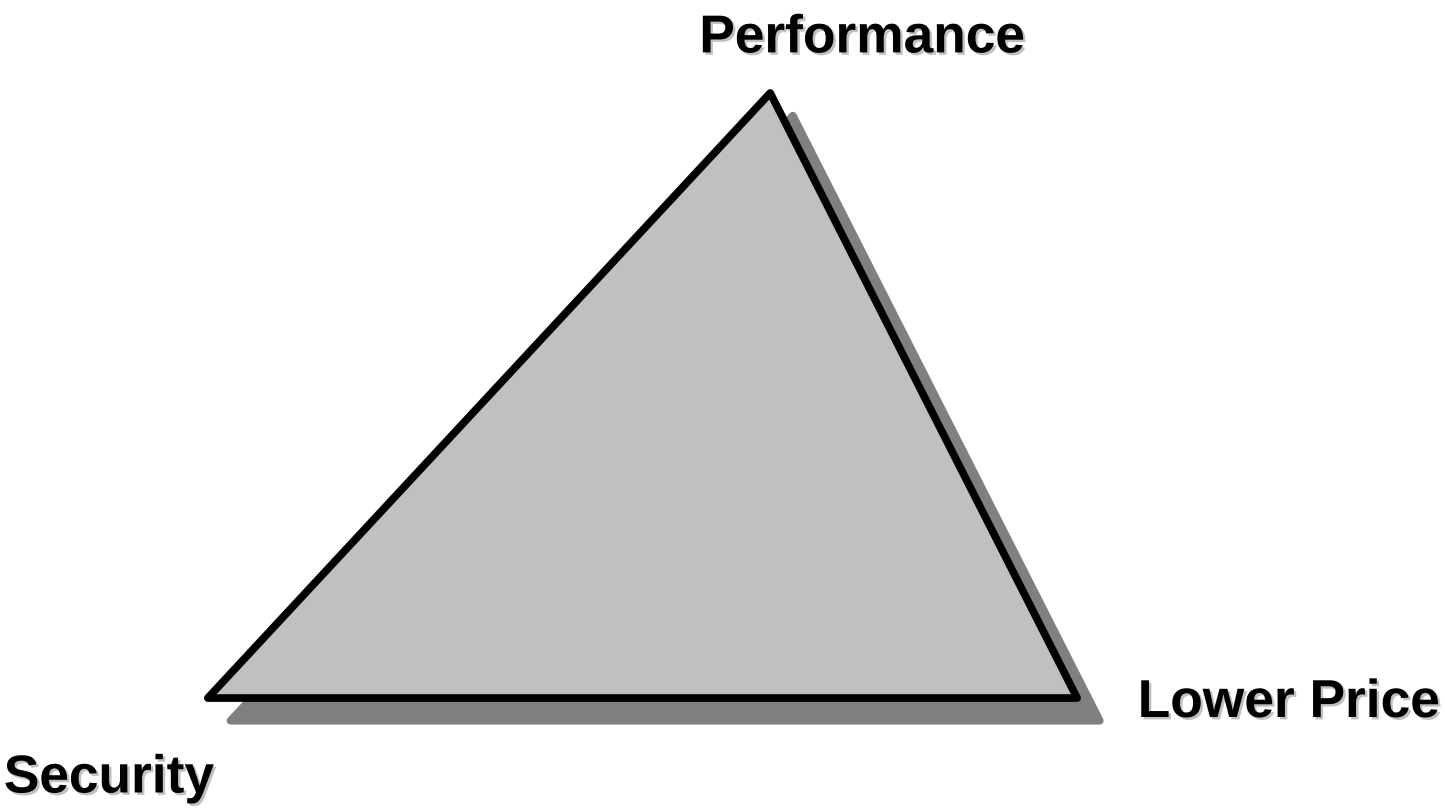
ORACLE

Think “Database Performance” from the beginning!

- Server:
  - Having faster CPU is still better! 32 cores is good enough ;-)
  - OS is important! - Linux, Solaris, etc.. (and Windows too!)
  - Right malloc() lib!! (Linux: jemalloc, Solaris: libumem)
- Storage:
  - Don't use slow disks! (except if this is a test validation goal :-))
  - SSD helping random access! (index/data) more and more cheaper
  - FS is important! - ZFS, UFS, QFS, VxFS, EXT3, EXT4, XFS, etc..
  - O_DIRECT or not O_DIRECT, AIO or not AIO, and be aware of bugs! ;-)
  - Do some generic I/O tests first (Sysbench, IObench, iozone, etc.)
- Don't forget network !! :-) (faster is better, 10Gbit is great!)

ORACLE

Seek for your best option..



Know your platform limits / “features”..

- My backup is finished on Linux faster than on Solaris same HW
  - Be sure first there is **really** no more I/O activity once backup is “finished”
  - Keep in mind Linux buffering..
- Linux distro: MySQL Performance has x4 regression! Fix it!
  - How did you see it? – Our QA test is taking x4 times more time..
  - Which engine? – InnoDB..
  - What is innodb_flush_log_at_trx_commit value? – set to 1.. why?
  - Tried innodb_flush_log_at_trx_commit=2 ?.. – Oh! You fixed it!! Thanks!!
  - Wait, what did you “improve” recently in distro? – FS flushing, why?..
  - Well, the test in fact is proving that you did not “sync” on every fsync() before, that's all.. But now in your FS flushing you get it fixed ;-)

Advise: Benchmark your platform / prototype!

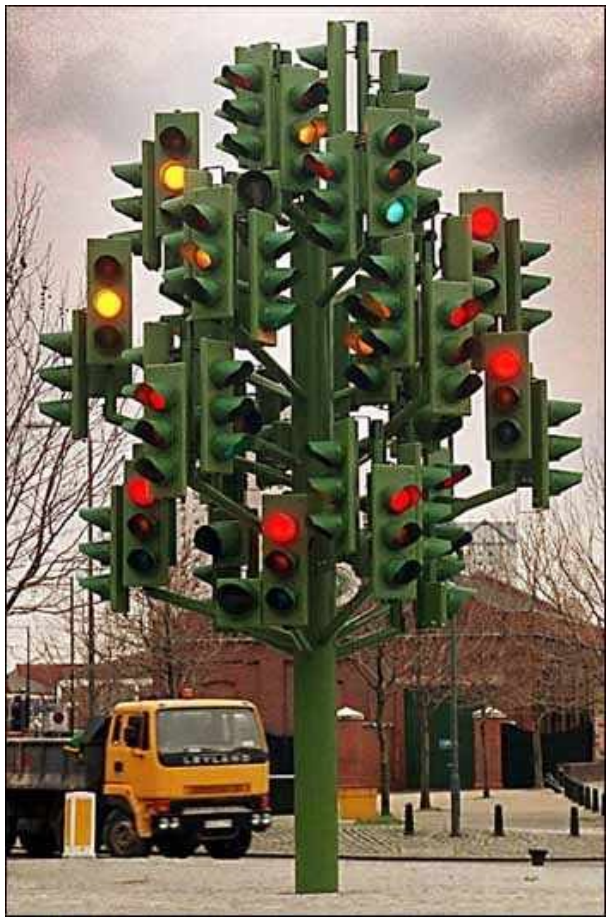
- Have a clear goal!
  - Otherwise: I've obtained all these results, and now... so what?..
- Want to simulate your production workload?..
  - Then just simulate it! (many SW available, not always OSS/free)
  - Hard to simulate? - adapt some generic tests
- Want to know capacity limits of a given platform?
  - Still try to focus on the test which are most significant for you!
- Want just to validate config settings impacts?
  - Focus on tests which are potentially depending on these settings
  - Or any, if the goal to prove there are not depending ;-)
- Well, just keep thinking about what you're doing ;-)

Popular “Generic” Test Workloads @MySQL

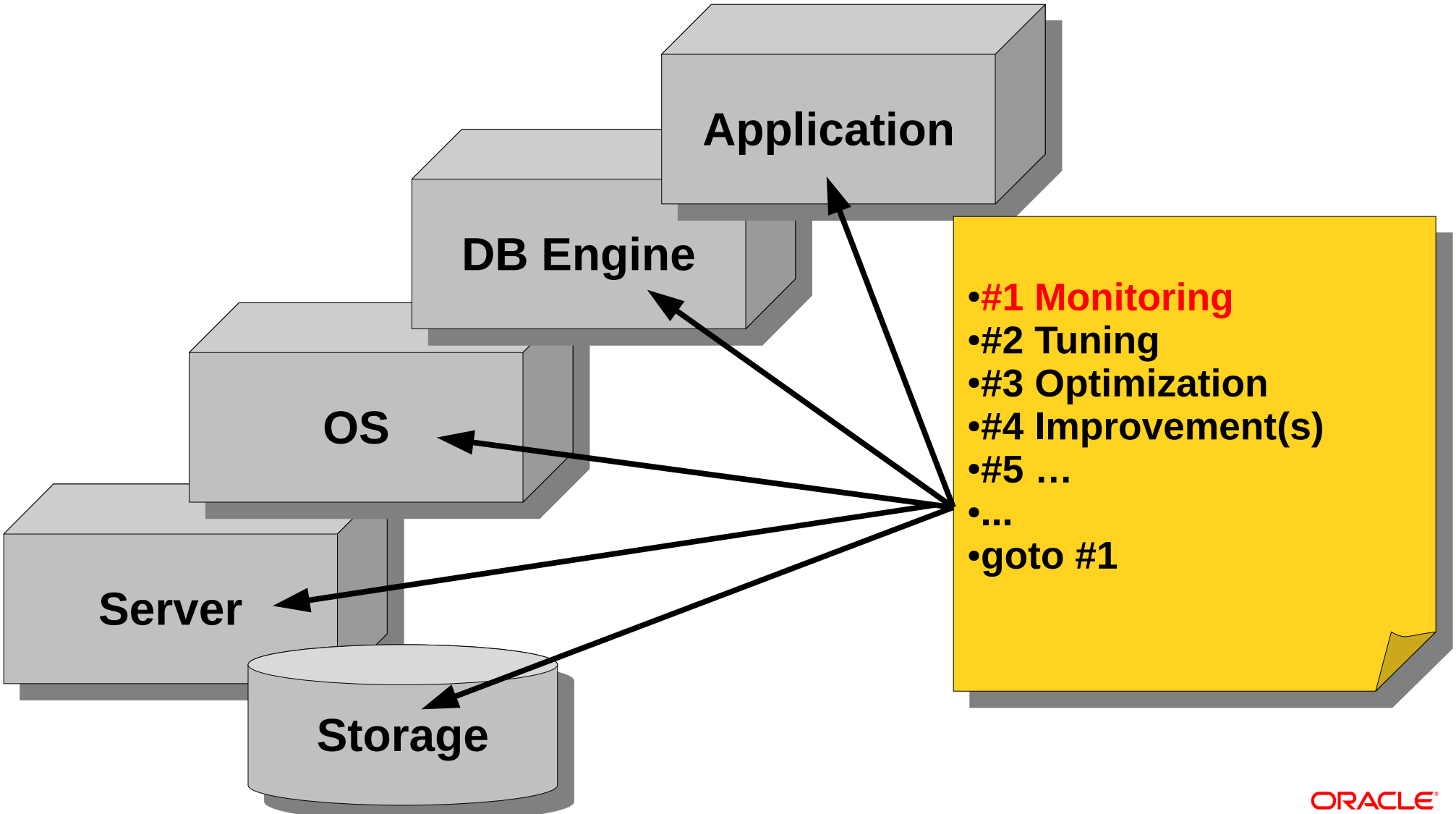
- Sysbench
  - OLTP, RO/RW, 1-table, since v0.5 N-table, lots load options, deadlocks
- DBT2 / TPCC-like
  - OLTP, RW, very complex, growing db, no options, deadlocks
  - In fact using mostly only 2 tables! (thanks Performance Schema ;-))
- dbSTRESS
  - OLTP, RO/RW, several tables, one most hot, configurable, no deadlocks
- linkbench (Facebook)
  - OLTP, RW, very intensive
- DBT3
  - DWH, complex heavy query, loved by Optimizer Team ;-)

Test Workload

- Before to do something complex...
  - Be sure first you're comfortable with “basic” operations!
  - Single table?
  - Many tables?
  - Short queries?
  - Long queries?
- Remember: any complex load just represents a mix of simple operations..
  - So, start from as simple as possible..
  - And then increase complexity progressively..

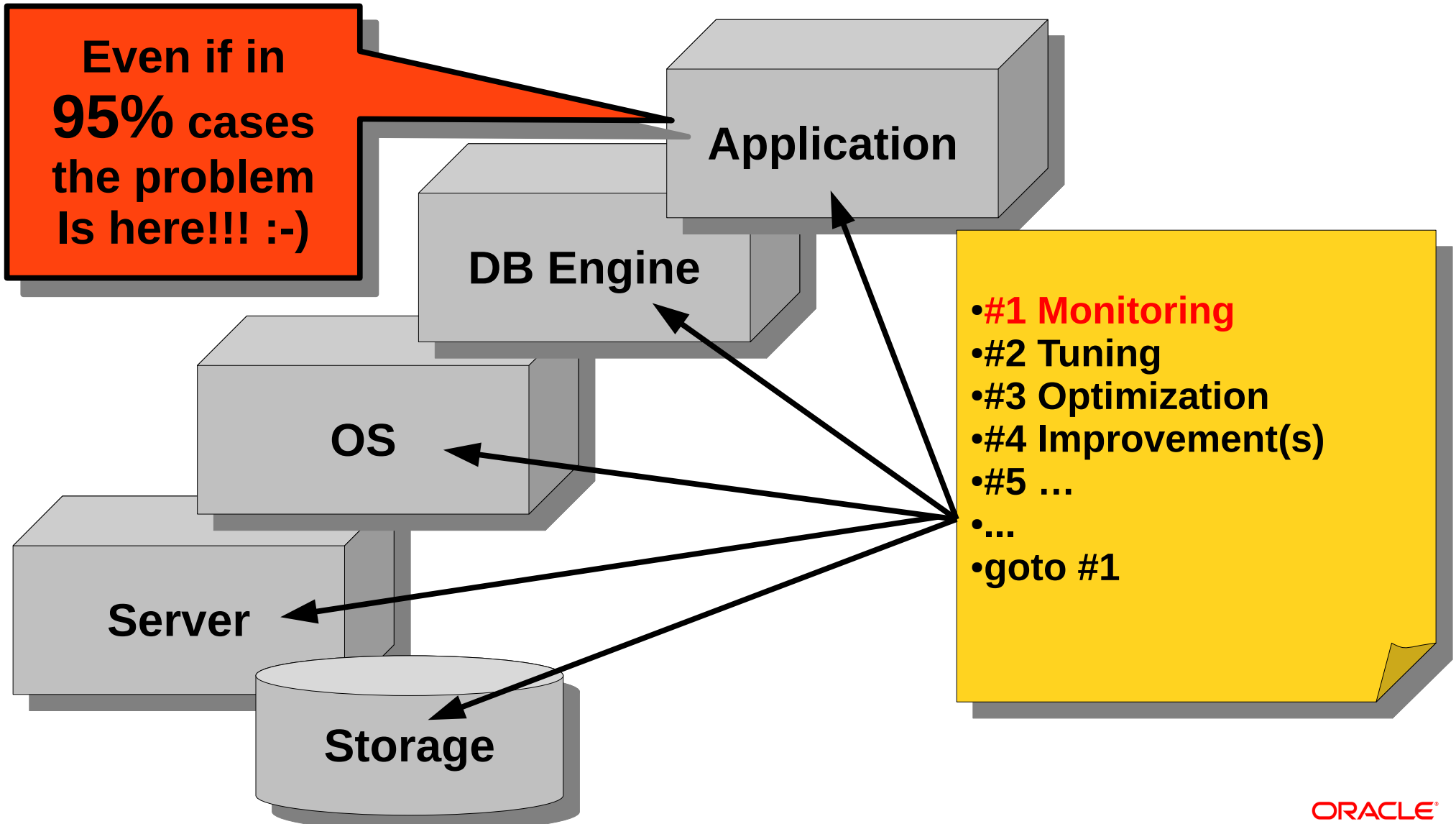


The Infinitive Loop of Database Tuning...



ORACLE

The Infinitive Loop of Database Tuning...



ORACLE

MySQL Enterprise Monitor

- Fantastic tool!
  - Did you already try it?.. Did you see it live?..

ORACLE MySQL Enterprise Monitor

22 22 0 248 0 admin

Dashboards Events Query Analyzer Reports & Graphs Configuration Refresh: Off

Group Overview: All

Database Statistics

Database Availability

Day	100%
Week	100%
Month	100%

Connections - All MySQL Instances

Connections 200 100 0 12:45 13:00 13:15 13:30

Total (SUM) Running (SUM)

Database Activity - All MySQL Instances

Statements ... 400 200 0 12:45 13:00 13:15 13:30

Select (SUM) Insert (SUM) Update (SUM) Replace (SUM) Delete (SUM) Call (SUM)

Query Response Time Index Sep 16, 2013 1:38:04 pm

QR Ti 1 0.5 0 12:45 13:00 13:15 13:30

optimal level acceptable level unacceptable level

qrti

Current Problem MySQL Instances

Show / hide columns

ID	Status	Emergency	Critical	Warning
bur05:33030	Up	0	2	11
tyr55:33300	Up	0	2	13
tyr58:3399	Up	0	1	17
tyr52:33030	Up	0	1	12

Showing 1 to 4 of 4 entries

Current Problem Hosts

Show / hide columns

ID	Status	Emergency	Critical	Warning
bur05	Up	0	1	0

Showing 1 to 1 of 1 entries

Emergency & Critical Events

Show 5 entries Show / hide columns First Previous 1 2 Next Last

Subject	Topic	Time	Actions
bur05, MEM Built-in Agent	Agent CPU Usage Excessi...	about a minute ago	✖
bur05, bur05:33030	Table Cache Not Optimal	about a minute ago	✖
tyr52, tyr52:33030	Table Cache Not Optimal	2 minutes ago	✖
bur05, bur05:33030	Attempted Connections T...	3 minutes ago	✖
tyr58, tyr58:3399	Table Cache Not Optimal	4 minutes ago	✖

Showing 1 to 5 of 7 entries First Previous 1 2 Next Last

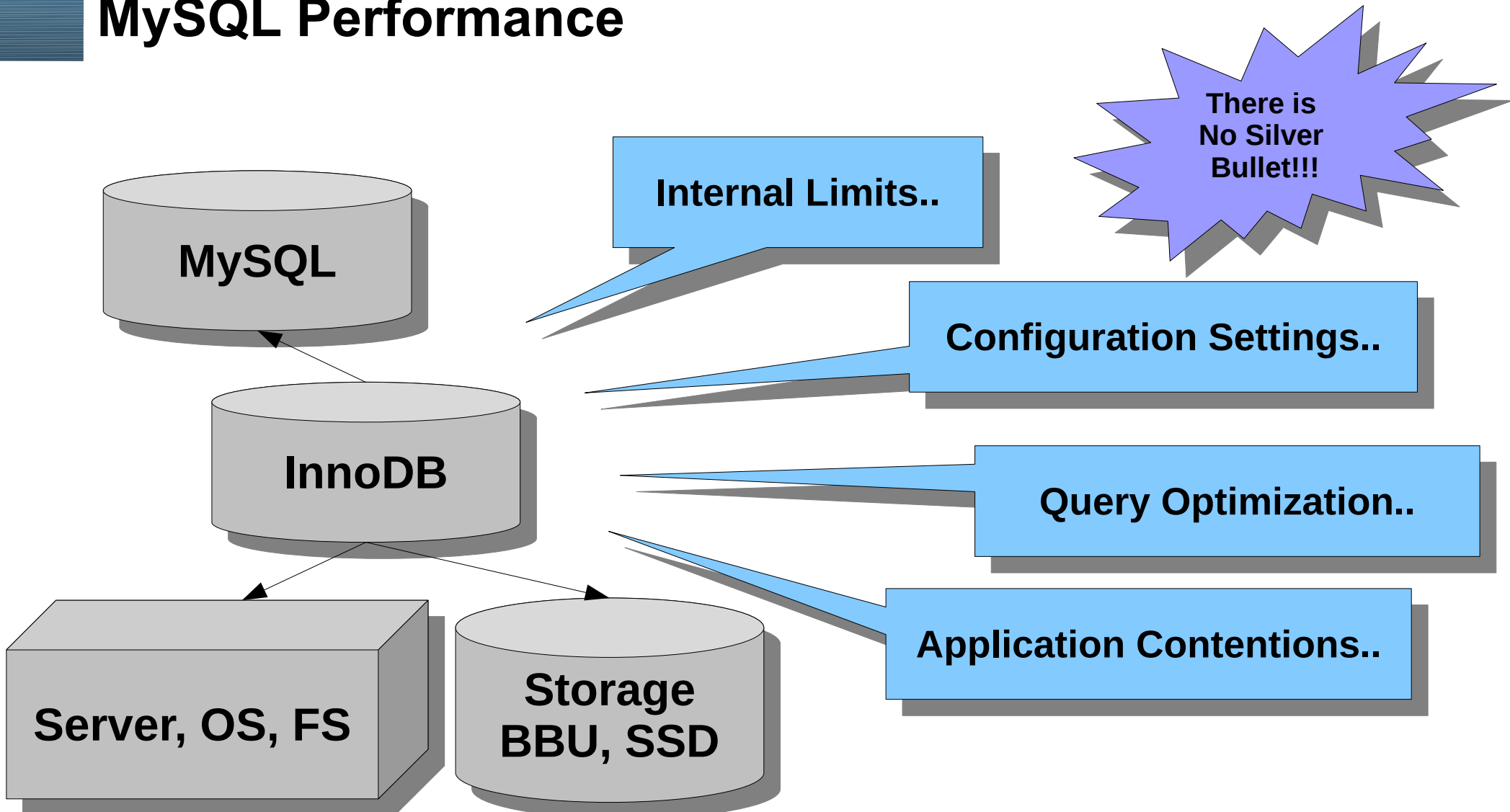
3.0.2.7154 - bur05 (10.172.161.65) - Sep 16, 2013 1:38:02 pm (Up Since: 1 day, 18 hours ago) - About

Other Monitoring Tools

- Cacti
- Zabbix
- Nagios
- Etc.....

- ***dim_STAT***
  - well, I'm using this one, sorry ;-)
  - all graphs within presentation were made with it
  - details are in the end of presentation..

MySQL Performance



Basic Tuning

- Understanding HW platform limits
  - helps you to deploy your MySQL Server in the most optimal way..
- Understanding MySQL Server internals
  - helps you to configure your database settings in the most optimal way..
  - use the best adapted Storage Engine
- Understanding of your Workload
  - helps you to tune the whole solution in the most optimal way ;-)
  - 20% of known issues covering 80% of observed problems..
  - So, adapt some best practices from the beginning..

Storage Engines: use InnoDB & MySQL 5.6 ;-)

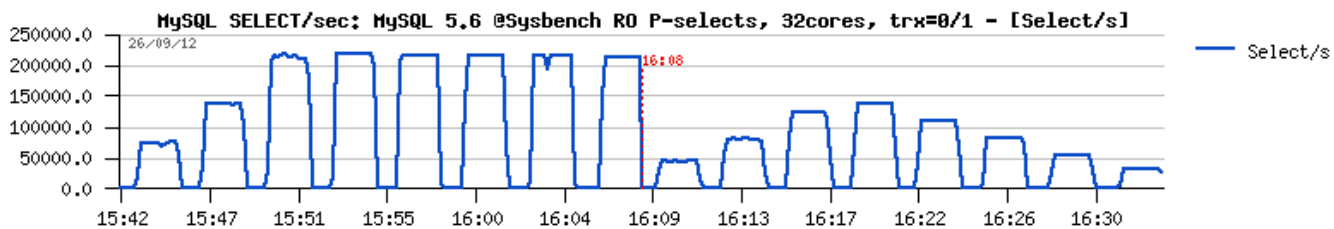
- MyISAM
  - Table level locking for everything, cache(s) for indexes only
  - Easily spending 50% of time on syscalls to read() data..
  - Many global locks, no transactions, no recovery..
- InnoDB
  - Row locking, transactions for everything, Buffer Pool
  - Double write buffer? Checksums?
  - innodb_flush_log_at_trx_commit= 1 / 2 ??
  - **many huge improvements since MySQL 5.6!**
- Binlog / Replication
  - Sync? - **binlog group commit** is since MySQL 5.6

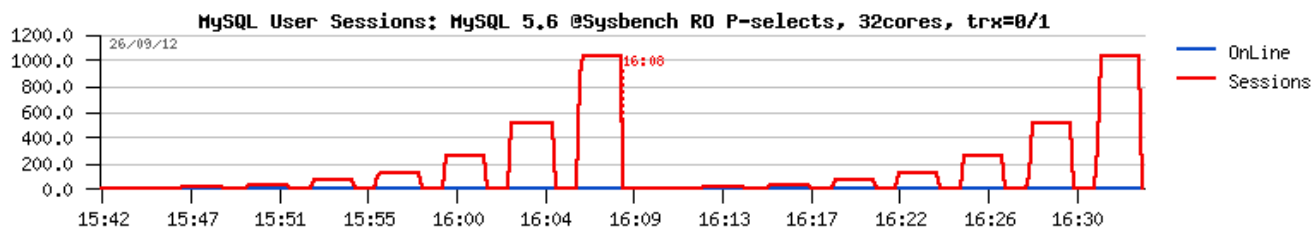
Workload: Read-Only oriented

- Bigger Buffer Pool (BP) is better
  - BP < dataset = IO-bound
- TRX list (kernel_mutex, since 5.6: trx_sys mutex)
- Read view
- Auto-commit or transactions?..
  - Grouping **many** queries within a single transaction may also largely reduce MDL locking, but still keep them **short** ! (check with PFS)
- Prepared statements
  - Observed 10% performance improvement in 5.6 (while Parser time is not more than 3% according to profiler)..
- Read-Only transactions!

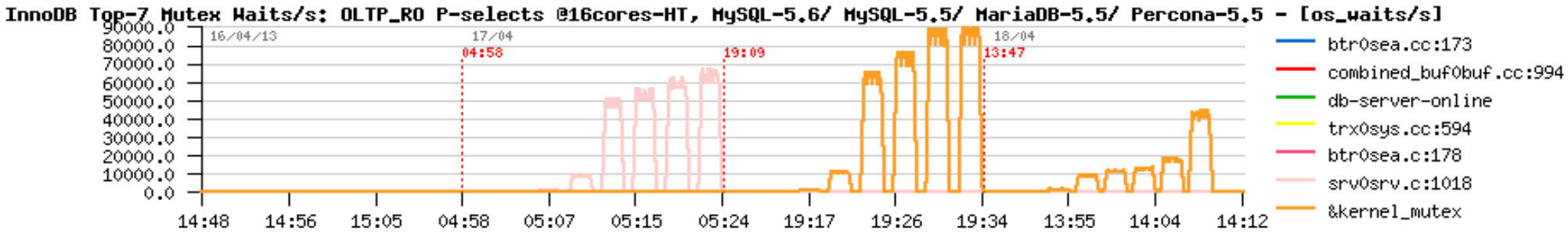
InnoDB: Read-Only Transactions in 5.6

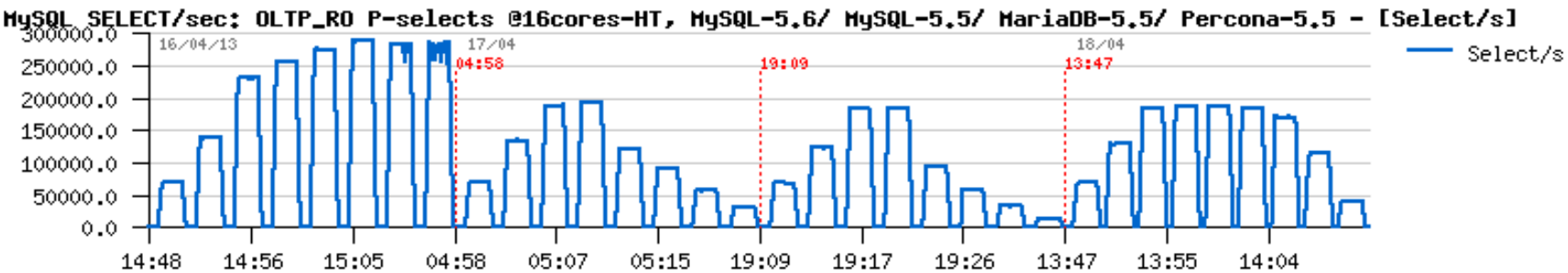
- Sysbench OLTP_RO Point-Selects:
  - Concurrent user sessions: 1, 2, 4 .. 1024
  - Using of transactions in sysbench = 0 / 1

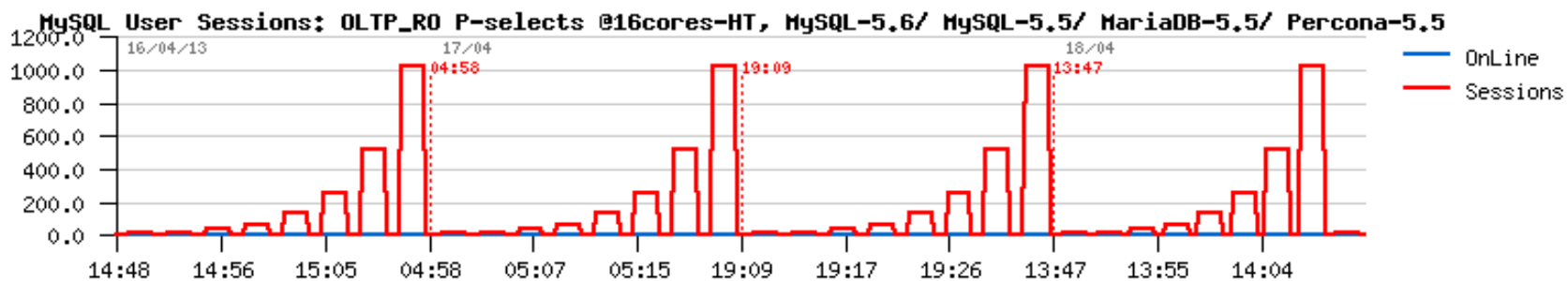




InnoDB: Read-Only Transactions in 5.6 (Apr.2013)



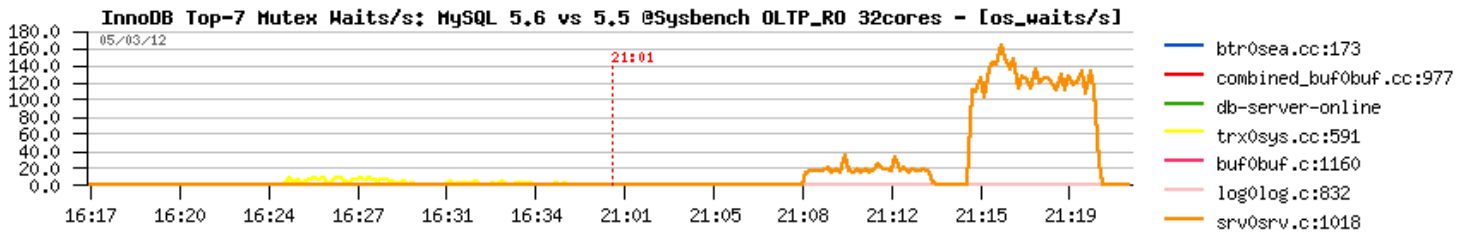


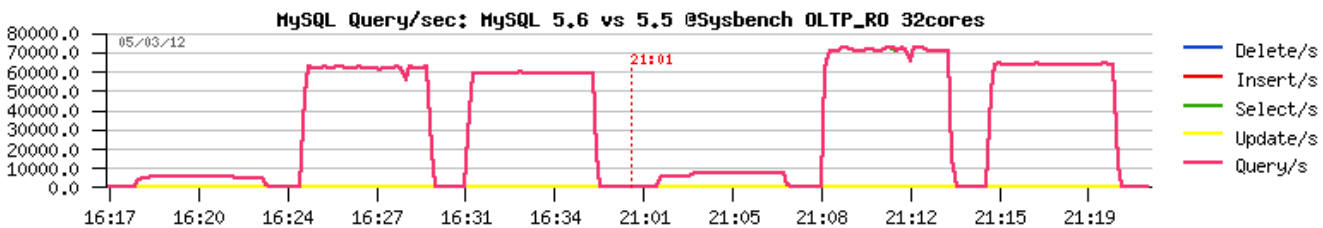


ORACLE

InnoDB : false sharing of cache-line = true killer

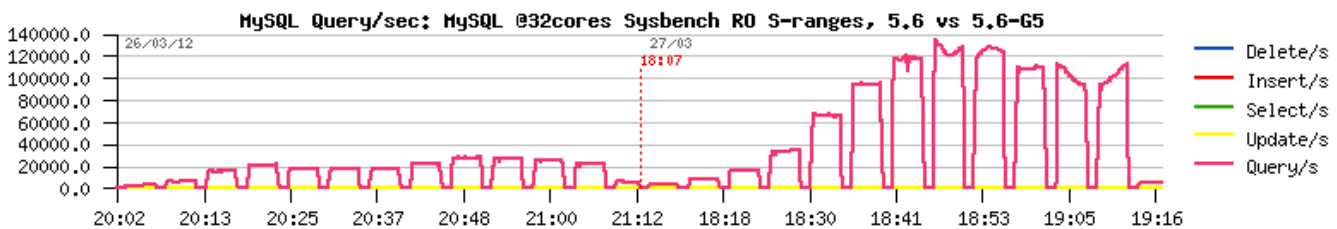
- RO or RW Workloads
  - Same symptoms in 5.5 & 5.6 : no QPS improvement between 16 and 32 user sessions:

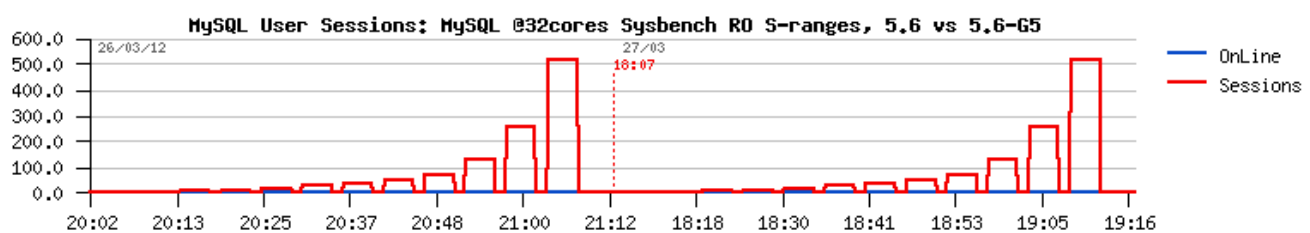




InnoDB : false sharing of cache-line fixed!

- RO or RW Workloads
  - “G5” patch! :-)
  - Over **x2(!)** times better on Sysbench OLTP_RO,
  - **x6(!)** times better on SIMPLE-Ranges!
  - NOTE: the fix is not applicable on 5.5..





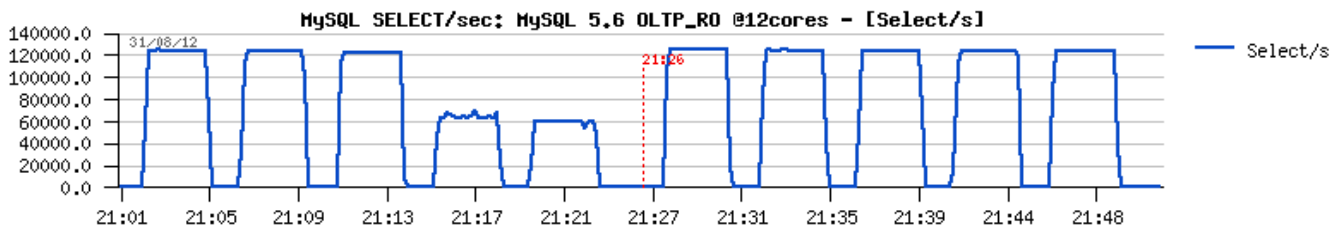
ORACLE

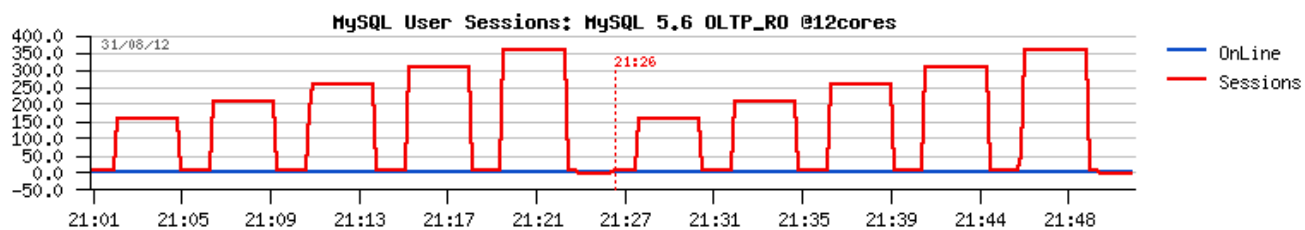
MySQL Internals: “killer” LOCK_open mutex

- MySQL 5.5 and before:
  - Keep “table_open_cache” setting big enough!
  - Monitor global status for '%opened%'
  - Once this contention become the most hot – well, time to upgrade to 5.6 ;-))
- Since MySQL 5.6:
  - Fixed: several table open cache instances
  - But it doesn't mean you can use a small “table_open_cache” either ;-)
  - Monitor PFS Waits!
  - Monitor “table_open_cache%” status variables!
  - Keep “table_open_cache_instances” at least bigger than 1

MySQL 5.6 Internals : low table_open_cache

- MySQL 5.6 :
  - Not big enough “table_open_cache” setting

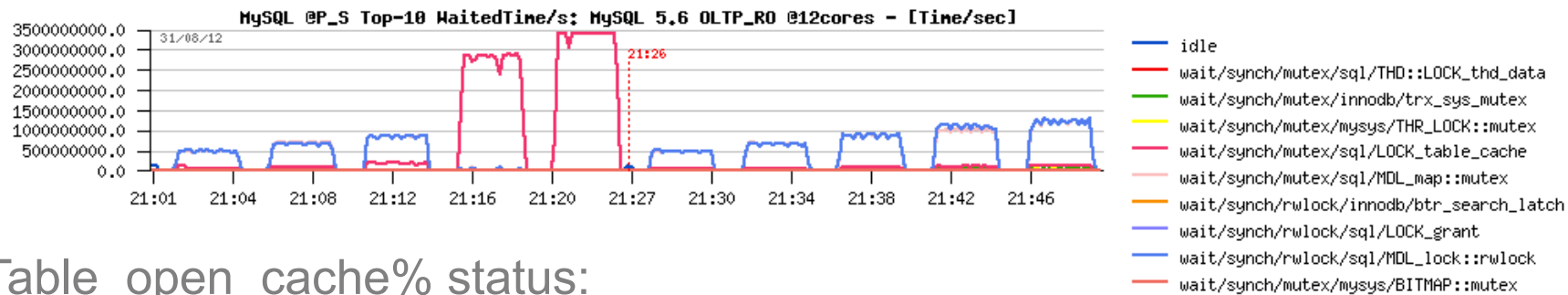




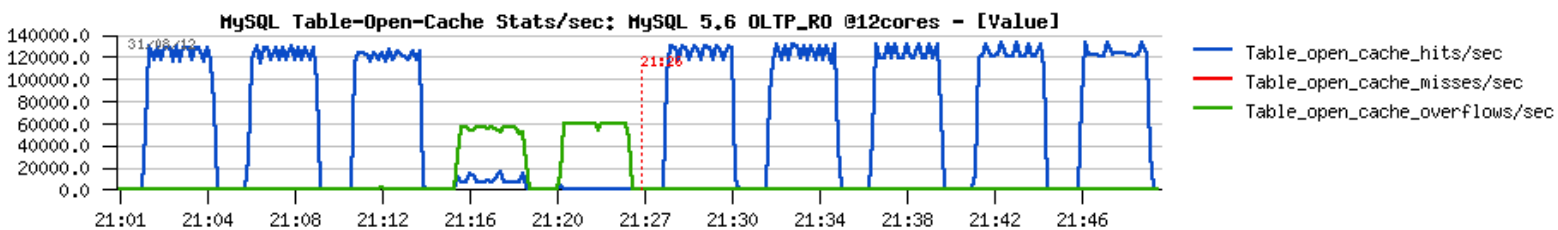
ORACLE

MySQL 5.6 Internals : low table_open_cache (2)

- MySQL 5.6 :
  - Not big enough “table_open_cache” setting
  - PFS Waits monitoring: LOCK_table_cache become the most hot:



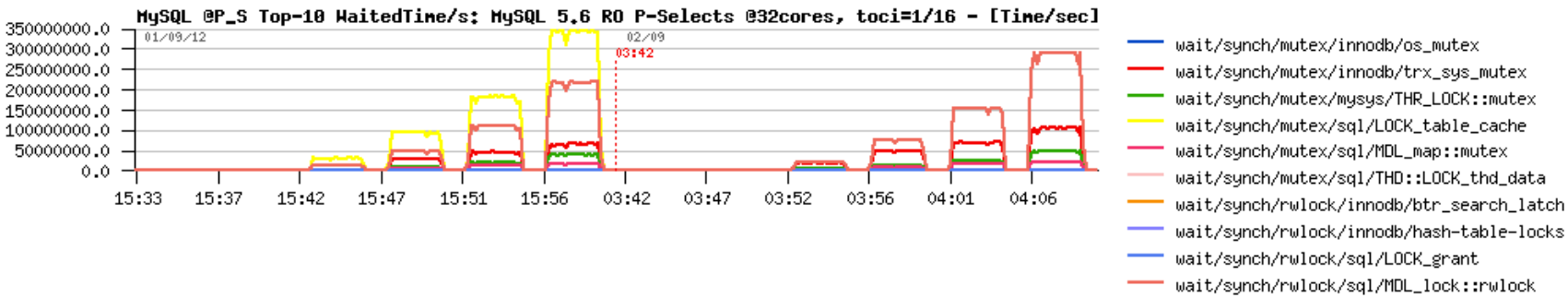
- Table_open_cache% status:

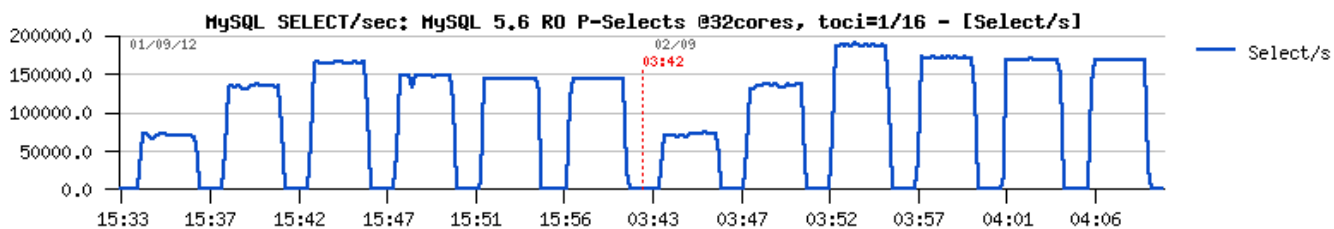


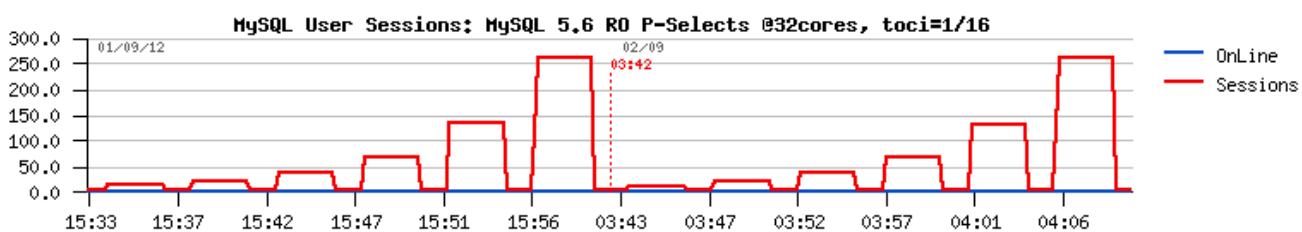
ORACLE

MySQL 5.6 Internals : table_open_cache_instances

- MySQL 5.6 :
  - When LOCK_table_cache wait is on top, the gain is usually well visible:







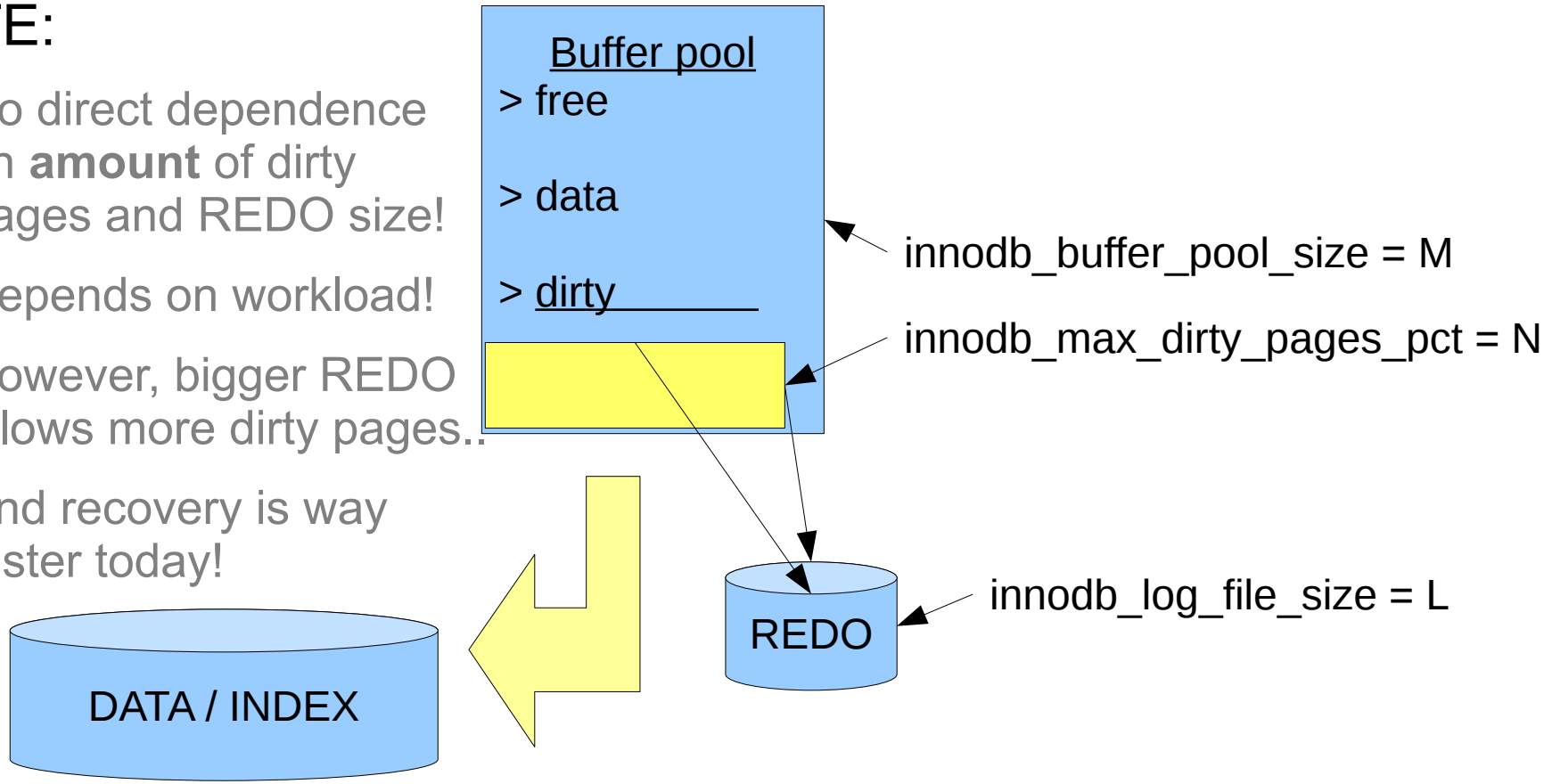
ORACLE

Workload: Read+Write

- RW activity
  - Updates only? Insert? Delete? R/W %ratio?
- Bigger Buffer Pool (BP) is still better
  - BP < dataset = IO-bound Reads(!) or R+W
  - BP > dataset = CPU-bound or IO-bound Writes(!)
- REDO size matters a lot! (up to 2TB in 5.6)
- Adaptive Flushing matters a lot!
- LRU flushing matters a lot as well!
- Tip: Neighbor Pages flushing = off / on

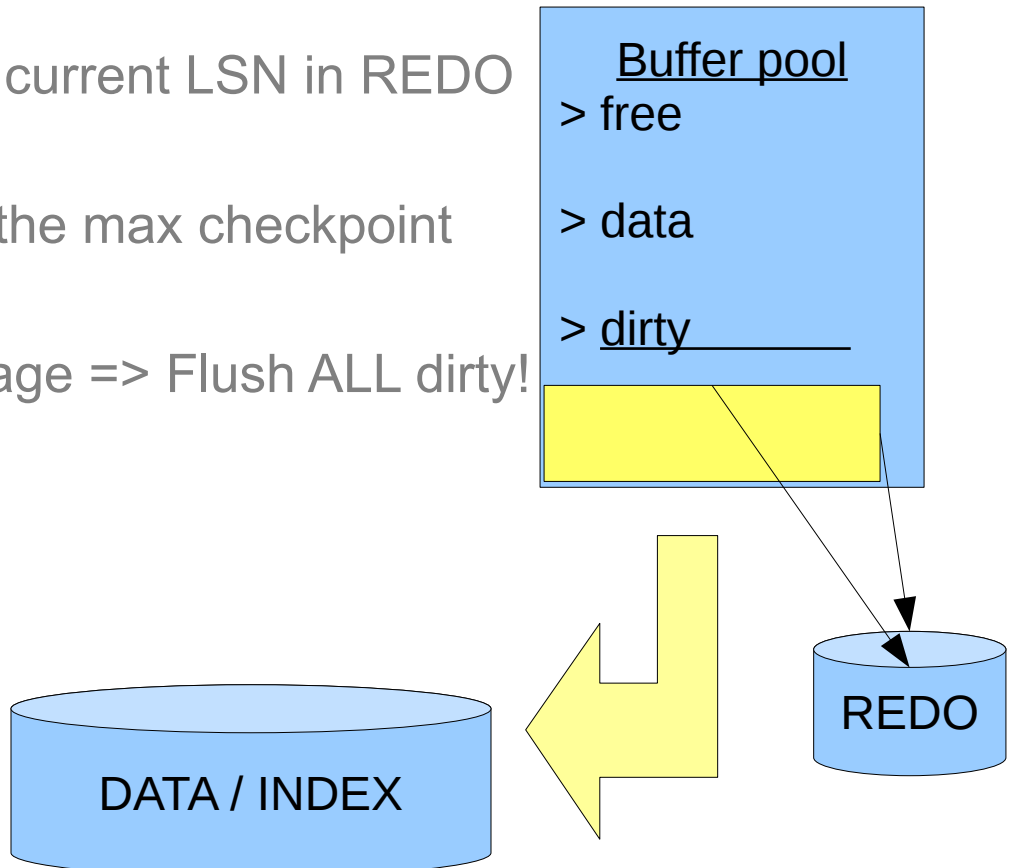
InnoDB: Dirty pages, Flushing

- Direct dependence on REDO log size
- NOTE:
  - No direct dependence on **amount** of dirty pages and REDO size!
  - Depends on workload!
  - However, bigger REDO allows more dirty pages..
  - And recovery is way faster today!



InnoDB: REDO log constraints

- REDO log constraints: (Always monitor Checkpoint Age!!!)
  - Cyclic, need free space
  - Checkpoint age: diff between the current LSN in REDO and the oldest dirty page LSN
  - Checkpoint age cannot out-pass the max checkpoint age (redo log size)
  - If Checkpoint age >= 7/8 of Max age => Flush ALL dirty!
  - => AKA “furious flushing”...
- Adaptive Flushing:
  - Keep REDO under Max age
  - Respecting IO capacity limit

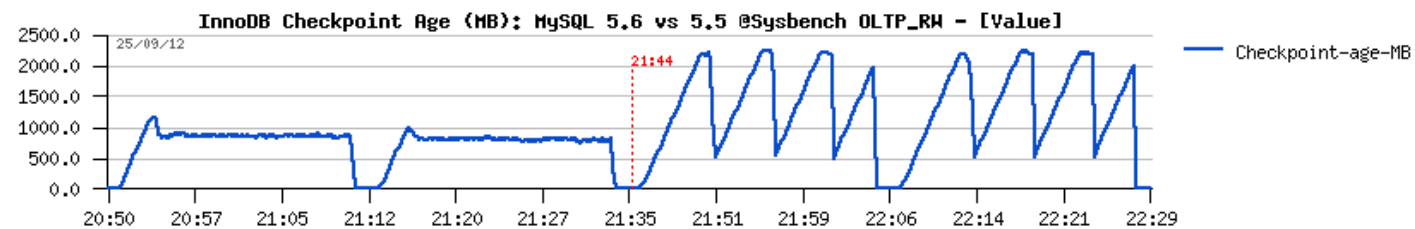


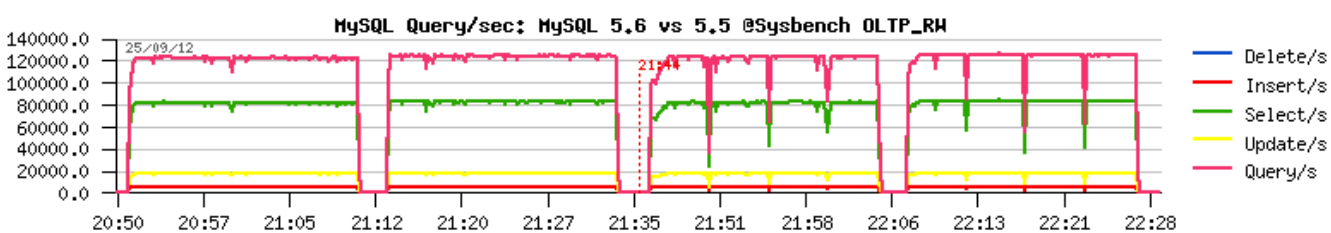
InnoDB: Adaptive Flushing

- MySQL 5.5:
  - Estimation based
  - Sometimes works ;-)
- MySQL 5.6 :
  - Based on REDO write rate + I/O capacity Max
  - Involving batch flushing with N pages to flush (progressive, depending on REDO %free) + page age limit (according REDO rate)
- Tuning:
  - innodb_io_capacity / innofb_io_capacity_max
  - innodb_adaptive_flushing_lwm / innodb_max_dirty_pages_pct_lwm
  - ALL are dynamic!
  - Monitor Checkpoint Age..

Adaptive Flushing: MySQL 5.6 vs 5.5

- OLTP_RW Workload:
  - Same IO capacity
  - Different logic..

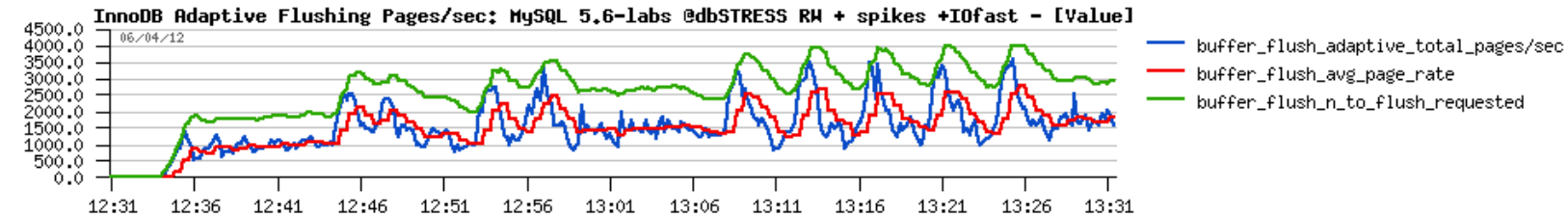


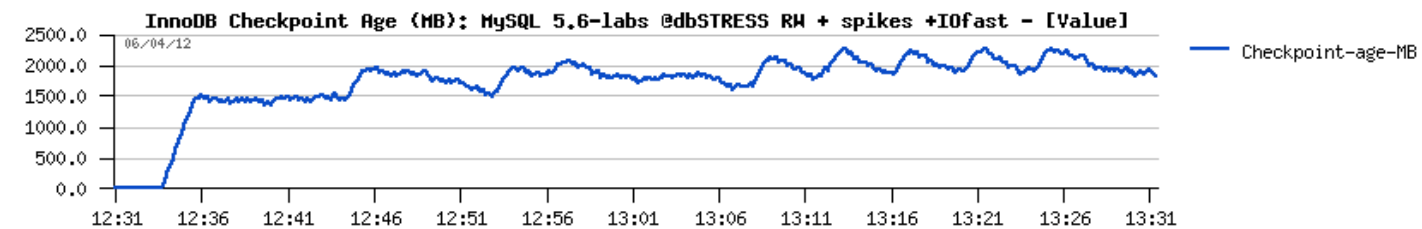


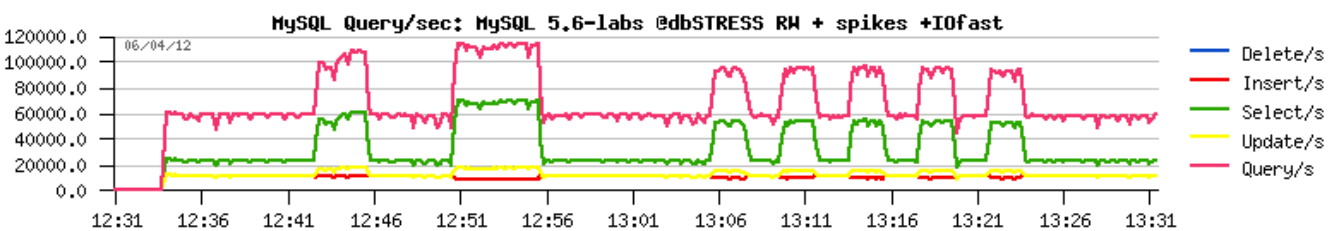
ORACLE

InnoDB : Resisting to activity spikes in 5.6

- dbSTRESS R+W with spikes







ORACLE

User Concurrency scenarios

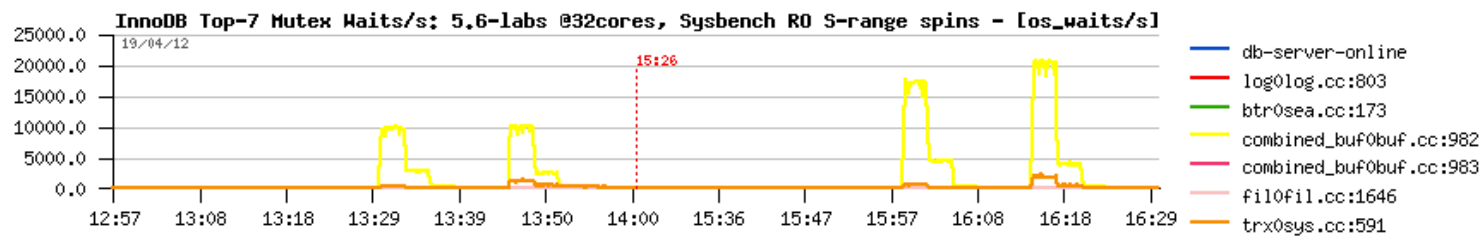
- Single user?..
  - With a bigger code path today 5.6 simply cannot be faster than 5.5
  - But then, why you're not considering Query Cache? ;-)
- More users?..
  - Up to 8-16 concurrent users all internal contention are not yet hot
  - So, 5.6 will not be better yet..
- More than 16 users?..
  - Then you'll feel a real difference, but if you have at least 16cores ;-)
  - Or if you have really a lot of concurrent users
- But don't forget other 5.6 improvements either!
  - On-line DDL, Binlog group commit, Memcached, etc..

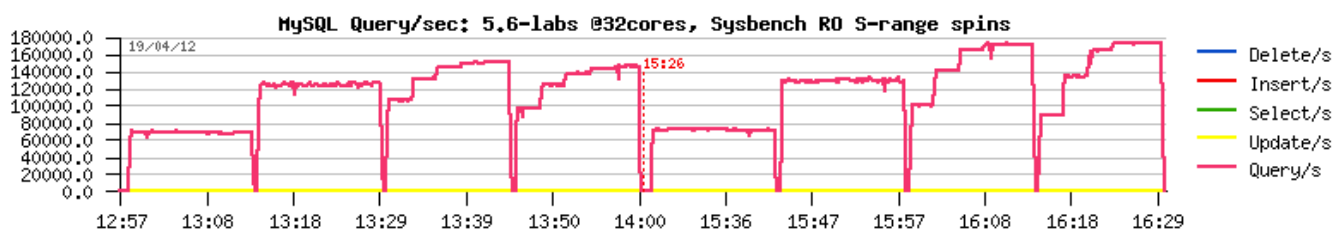
High Concurrency Tuning

- If bottleneck is due a concurrent access on the same data (due application design) – ask dev team to re-design ;-)
- If bottleneck is due MySQL/InnoDB internal contentions, then:
  - If you cannot avoid it, then at least don't let them grow ;-)
  - Try to increase InnoDB spin wait delay (dynamic)
  - Try innodb_thread_concurrency=N (dynamic)
  - CPU taskset / prcset (Linux / Solaris, both dynamic)
  - Thread Pool
  - NOTE: things with contentions may radically change since 5.7, so stay tuned ;-)

InnoDB Spin Wait Delay

- RO/RW Workloads:
  - With more CPU cores internal contentions become more hot..
  - Bind mysqld to less cores helps, but the goal is to use more cores ;-)
  - Using innodb_thread_concurrency is not helping here anymore..
  - So, **innodb_spin_wait_delay** is entering in the game:

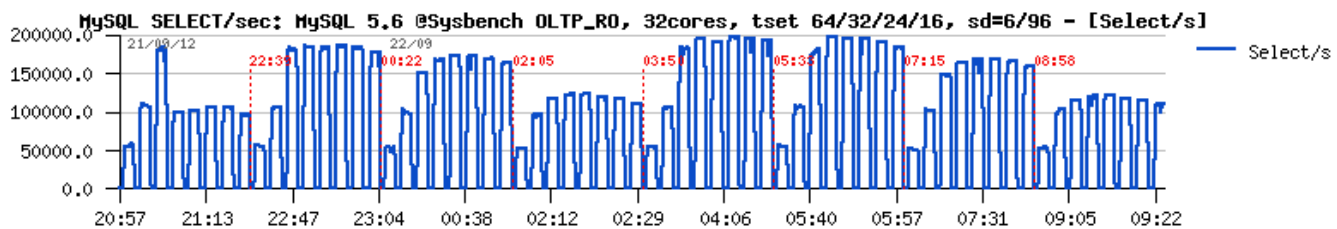


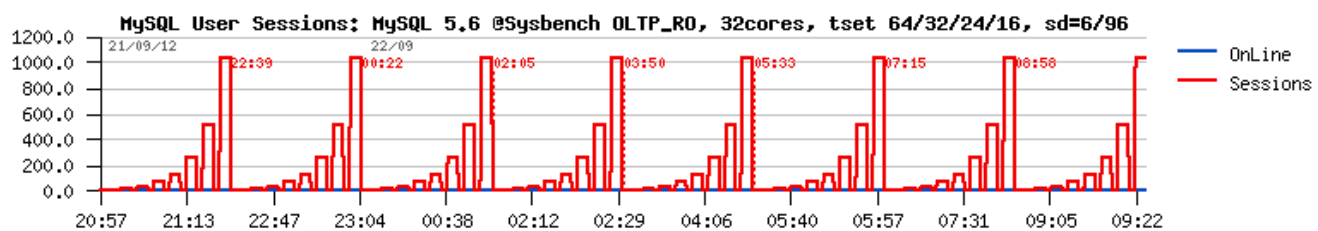


ORACLE

Tune InnoDB Spin Wait Delay

- Notes :
  - is the max random delay on “sleep” within a spin loop in wait for lock..
  - Ideally should be auto.. while the same tuning works for 5.5 as well ;-)
  - General rule: default is 6, may need an increase with more cores
  - Test: 32-HT/ 32/ 24/ 16cores, spin delay = 6 / 96 :





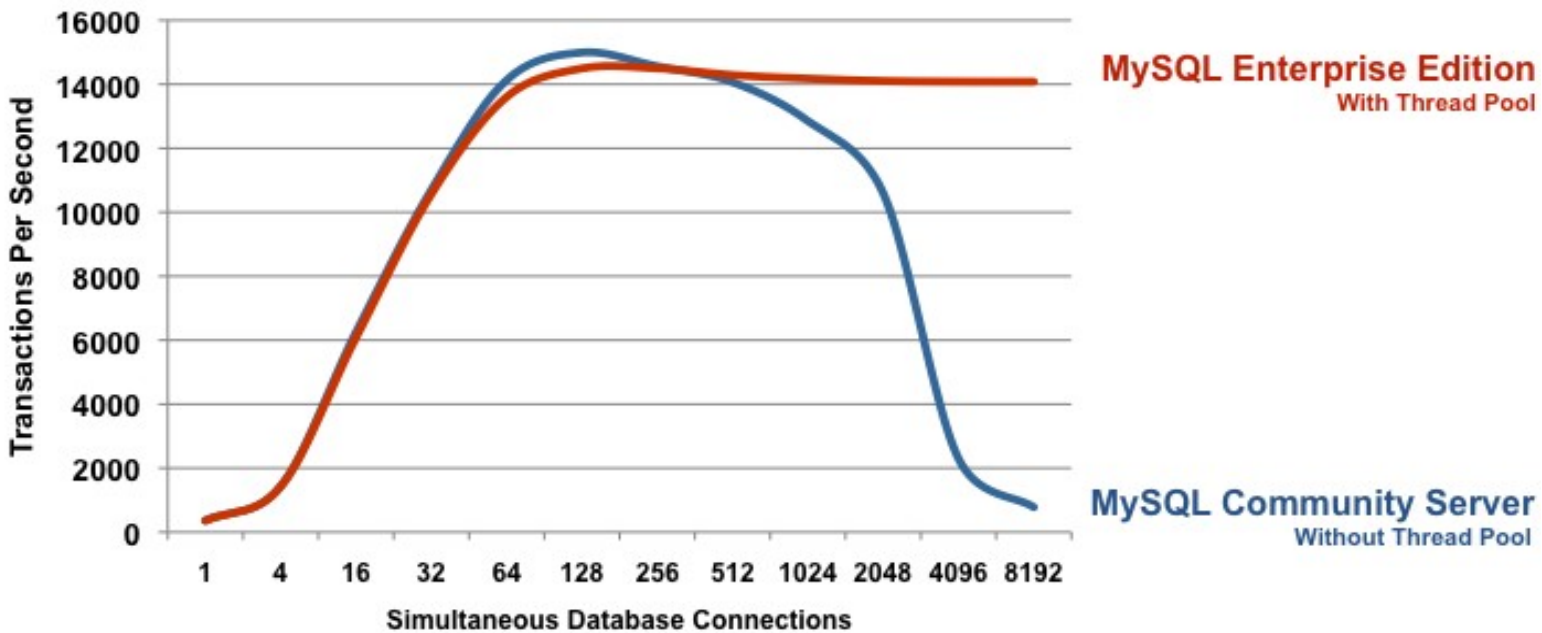
ORACLE

Thread Pool @MySQL

- None of these solutions will help to increase performance!
  - it'll just help to keep the peak level constant (and you yet need to discover on which level of concurrency you're reaching your peak ;-))
- ThreadPool in MySQL 5.5 and 5.6 is aware if I/O are involved!
  - So, better than innodb thread concurrency setting or taskset
  - May still require spin wait delay tuning!
  - The must for high concurrency loads!
  - May still start to show a difference since 32-128 concurrent users! (all depends on workload)..
  - Keep in mind that OS scheduler is not aware how to manage user threads most optimally, but ThreadPool does ;-)

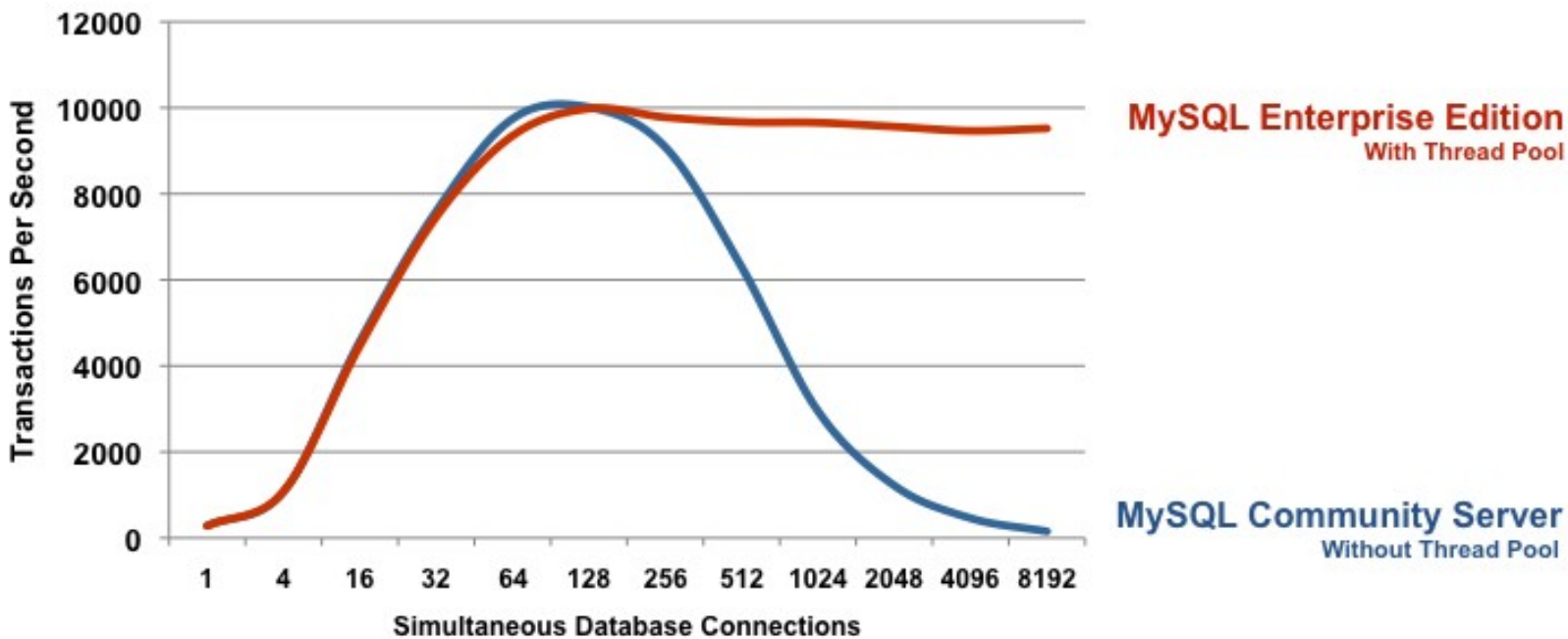
Thread Pool in MySQL 5.6

- OLTP_RO:

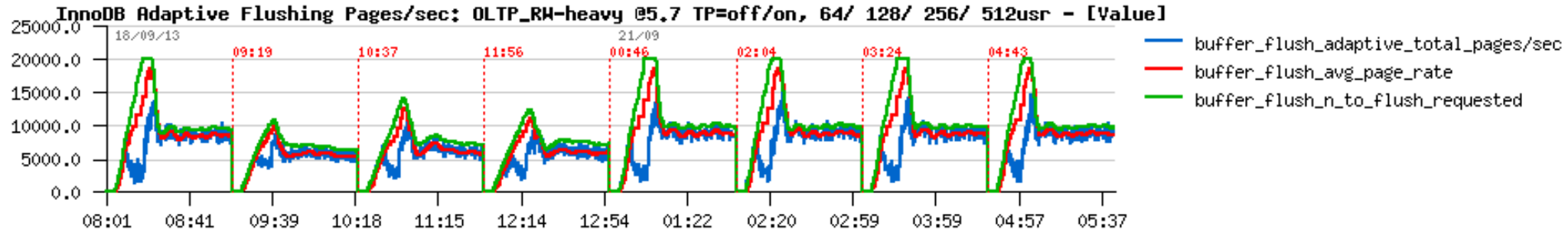


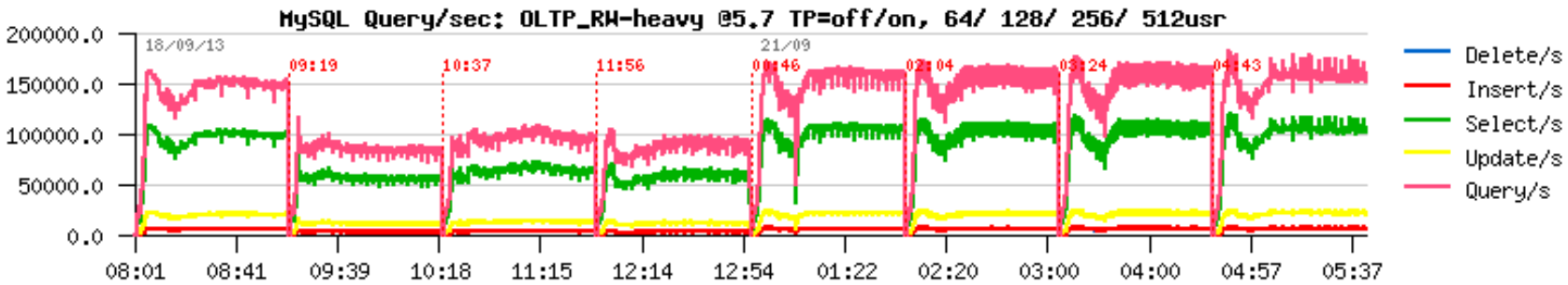
Thread Pool in MySQL 5.6

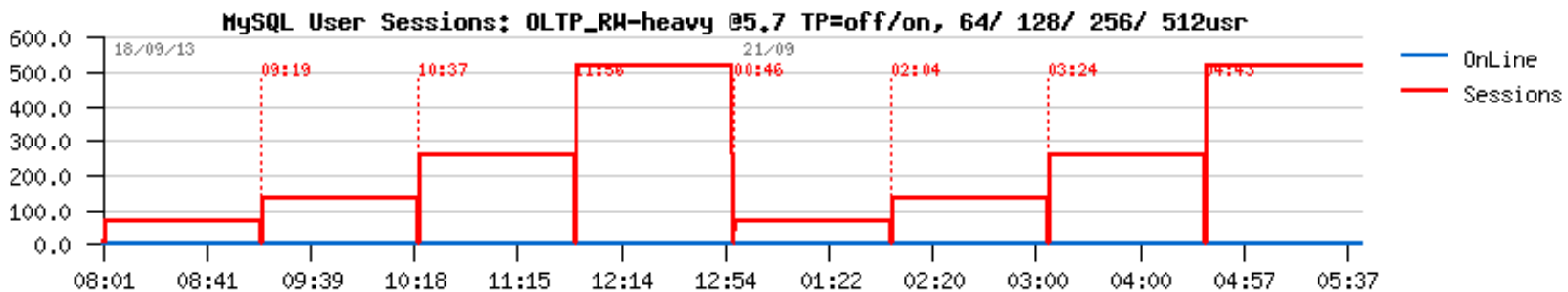
- OLTP_RW:



Thread Pool in MySQL 5.7 @Heavy OLTP_RW







ORACLE

InnoDB High Concurrency: AHI

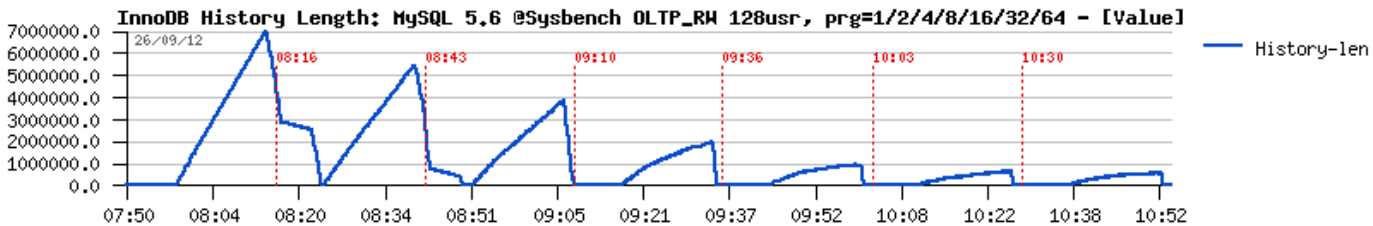
- Adaptive Hash Index (AHI)
  - Helps a lot on Read-Only workloads
  - In fact it helps always until itself become not actively modified
  - AHI contention is seen as its btr_search_latch RW-lock contetnion
  - So, on Read+Write become a huge bottleneck..
  - In many cases on RW the result is better with AHI=off..
  - NOTE: there is still a big mystery around AHI when it's having btr_search_latch contention even when there is no changes at all (pure RO in memory).. - expected to be fixed in 5.7 ;-)

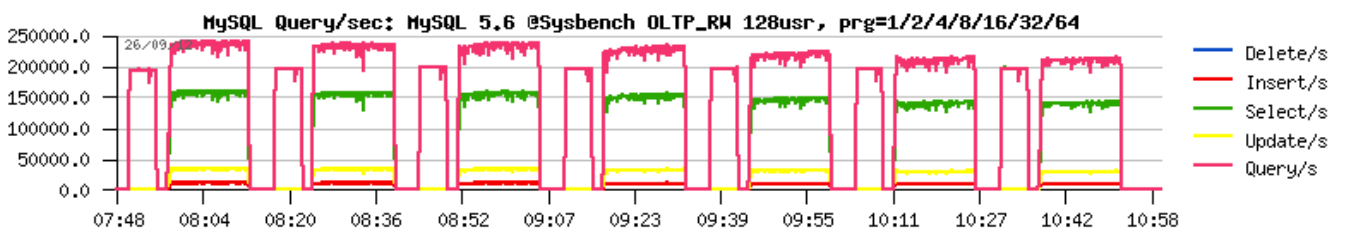
InnoDB Purge

- Purging (similar to Garbage Collecting)
  - Since MySQL 5.5: purge thread
  - Since MySQL 5.6: purge thread(s) (up to 32)
- Having Purge following workload is very important!
  - Ex.: On aggressive RW got 400GB of undo records within few hours(!)
  - Then it took days to reach zero in History Length..
- The main problem is the past – how to dose it?..
  - Since 5.6: with many threads, Purge become auto-stable itself
  - Still missing a dynamic config option to say how many purge threads to run in parallel right now (but it'll be fixed soon ;-))

InnoDB : Purge improvement in 5.6

- Several Purge Threads :
  - NOTE: activation is auto-magical (I'm serious ;-))

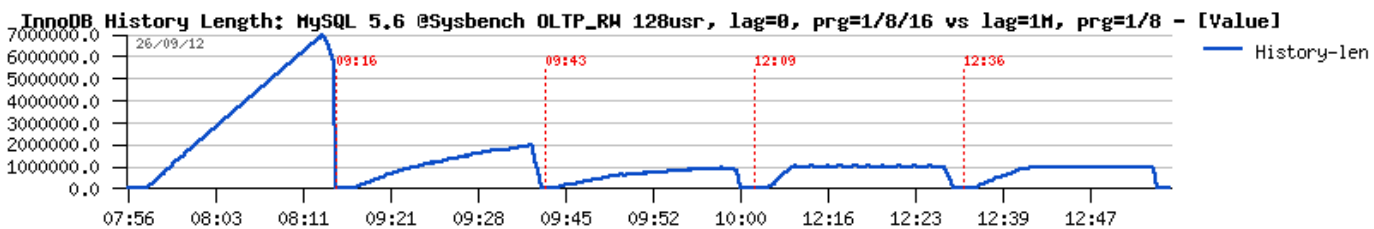


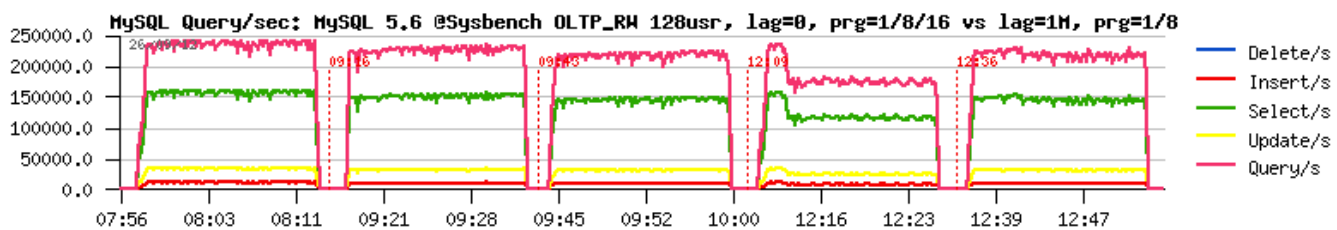


ORACLE

InnoDB : Purge improvement in 5.6

- Fixed max purge lag code!
  - innodb_max_purge_lag
  - innodb_max_purge_lag_delay <= configurable!
- Setting innodb_max_purge_lag=1M:



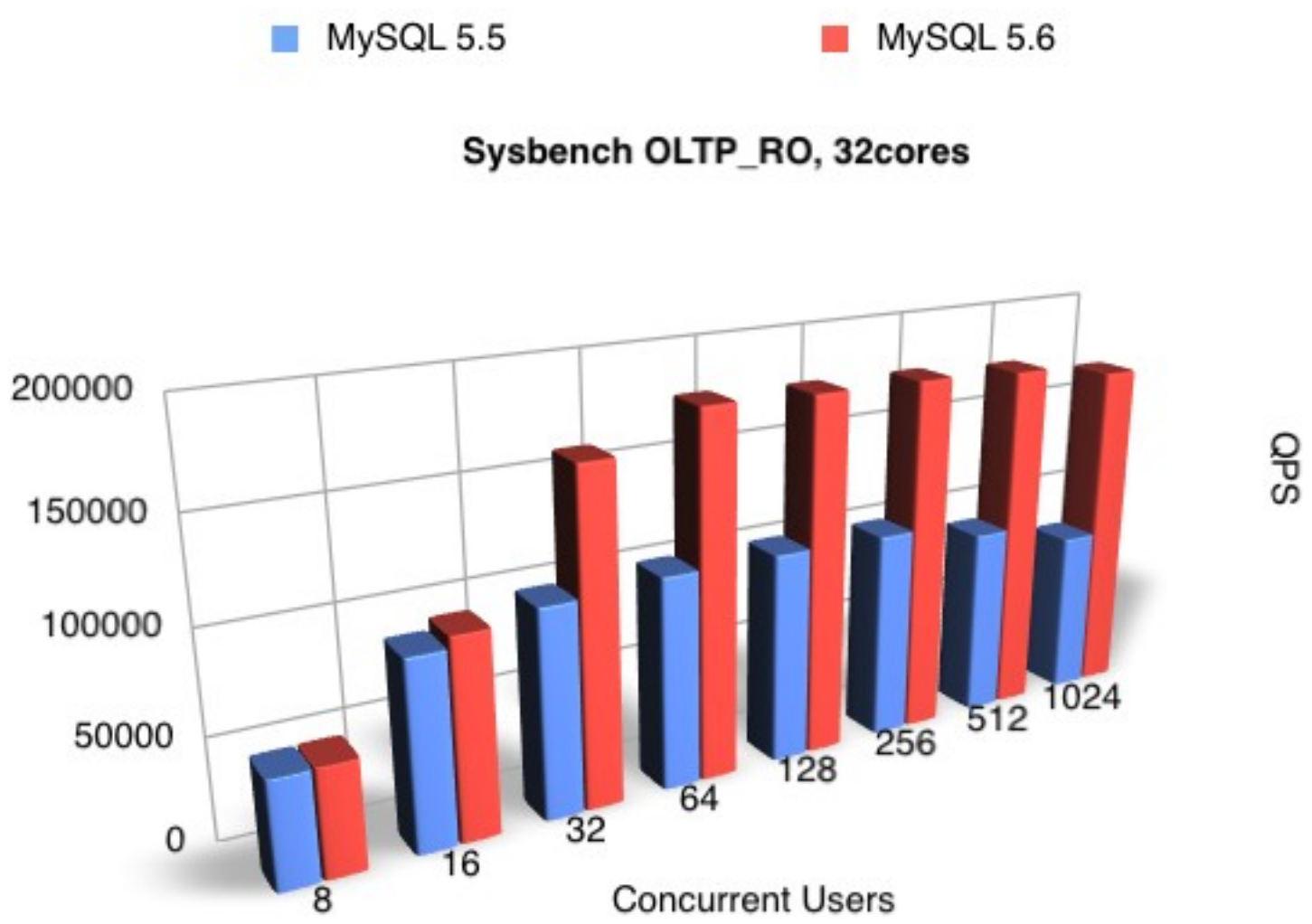


ORACLE

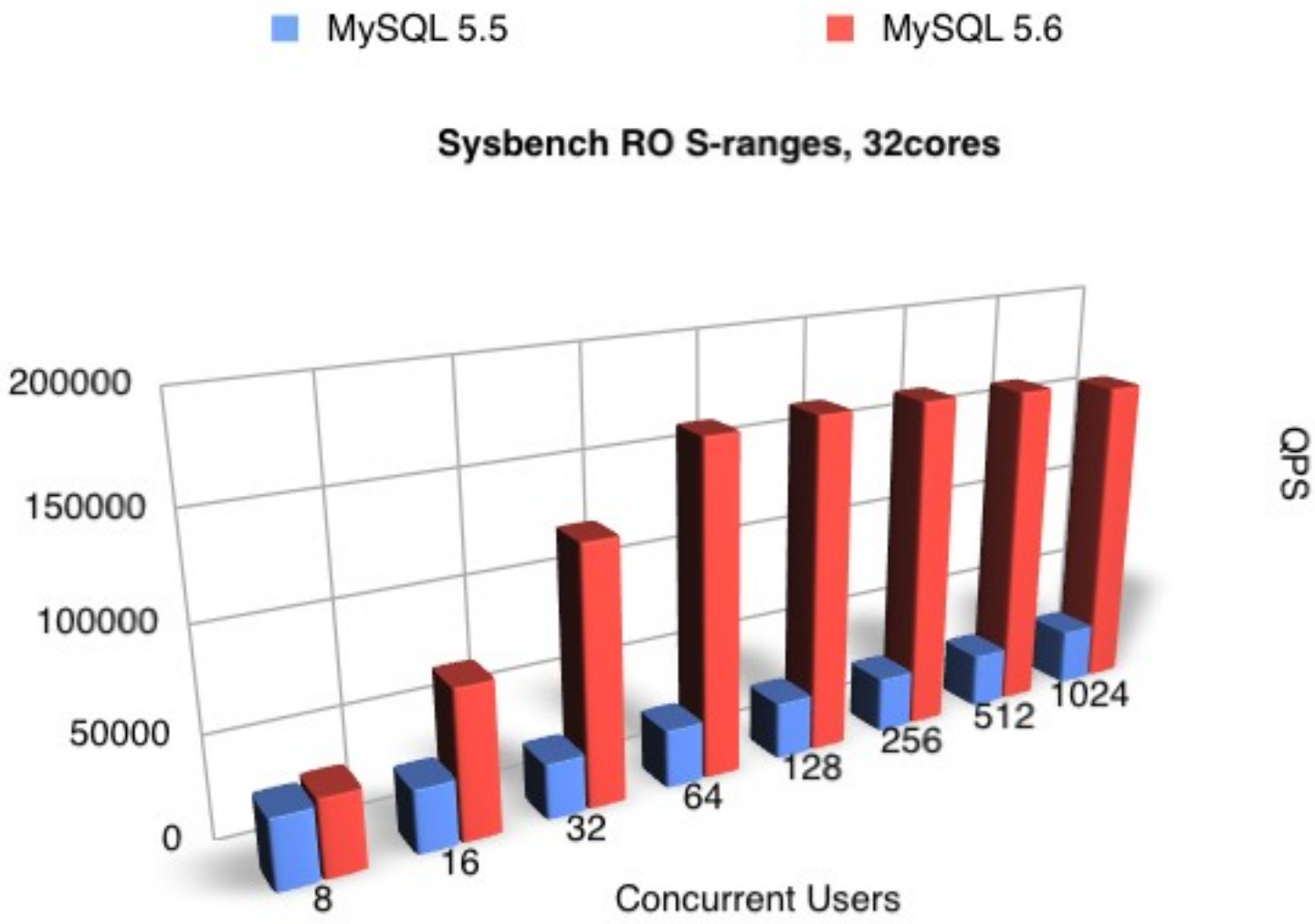
Testing Apples-to-Apples...

- Comparing MySQL 5.6 vs 5.5 :
  - Don't have G5: dead..
  - Don't have open table cache instances: bad..
  - Don't have improved Adaptive Flushing; bad..
  - Don't have fixed Purge & Lag: danger!..
  - Don't have binlog group commit and use binlog: dead..
  - Etc. etc. etc.
  - NOTE: some “improvement” are also fixes which are making stuff working properly, but coming with additional overhead (like Purge)..
  - NOTE: when comparing 5.6 and 5.5 keep in mind that Performance Schema is enabled by default in 5.6, and not in 5.5, so think to disable it in both (as 5.5 also has a way less PFS instrumentation)..

Hope you did not miss ;-)

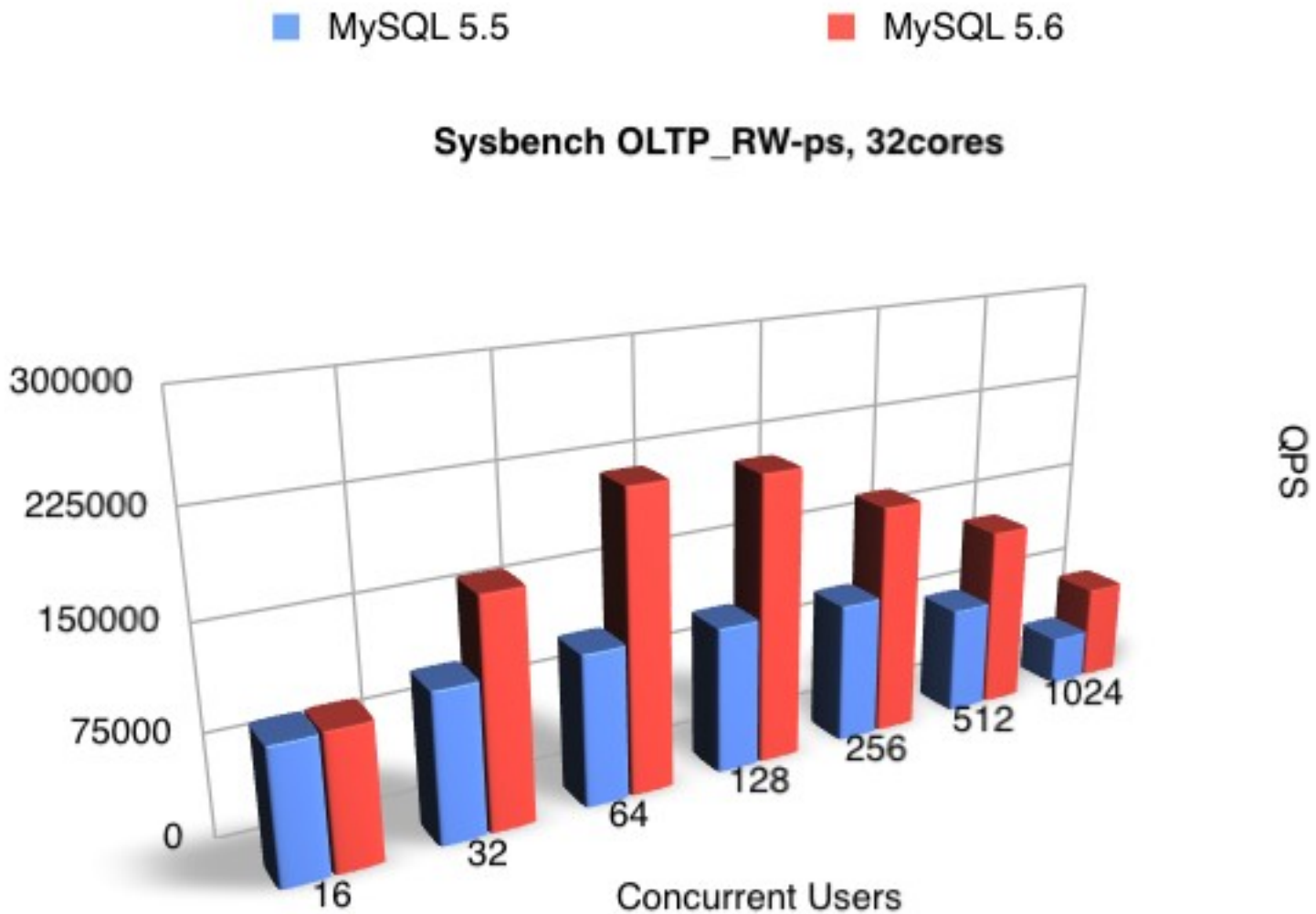


Hope you did not miss ;-) (2)

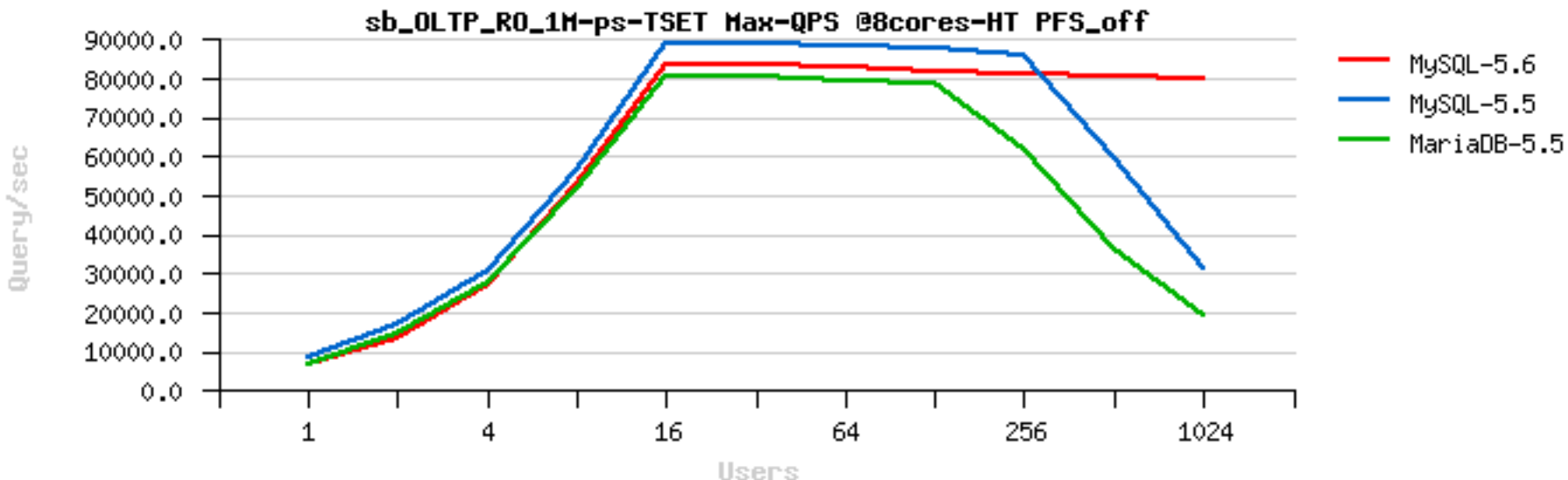


ORACLE

Hope you did not miss ;-) (3)

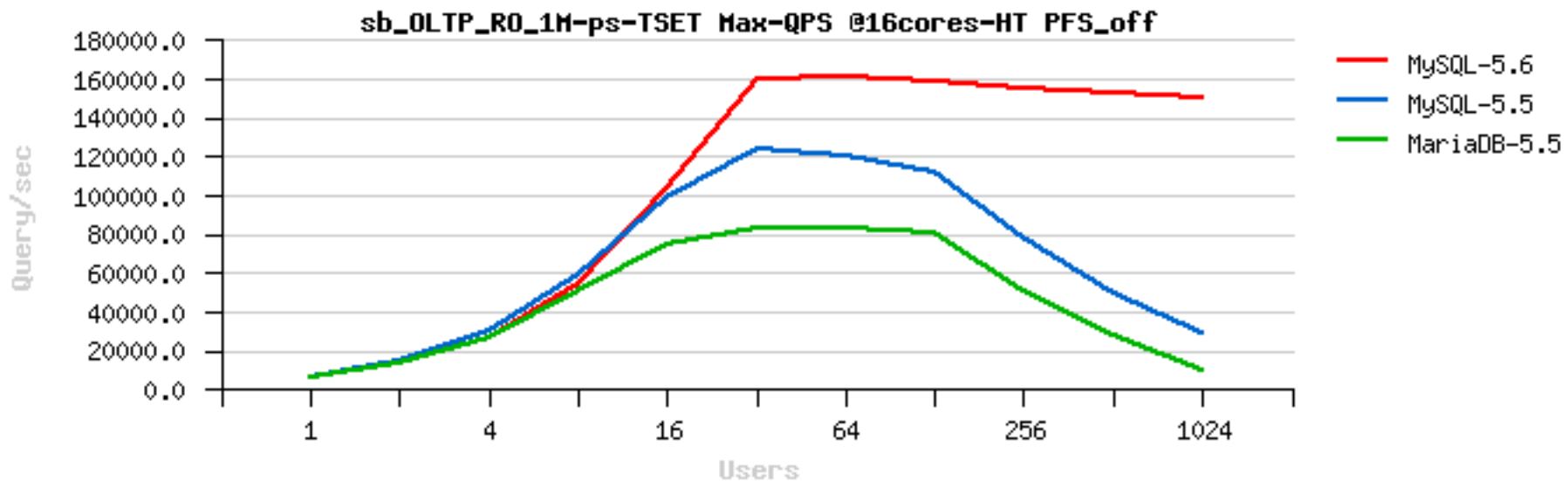


Sysbench OLTP_RO @8cores-HT (Apr.2013)



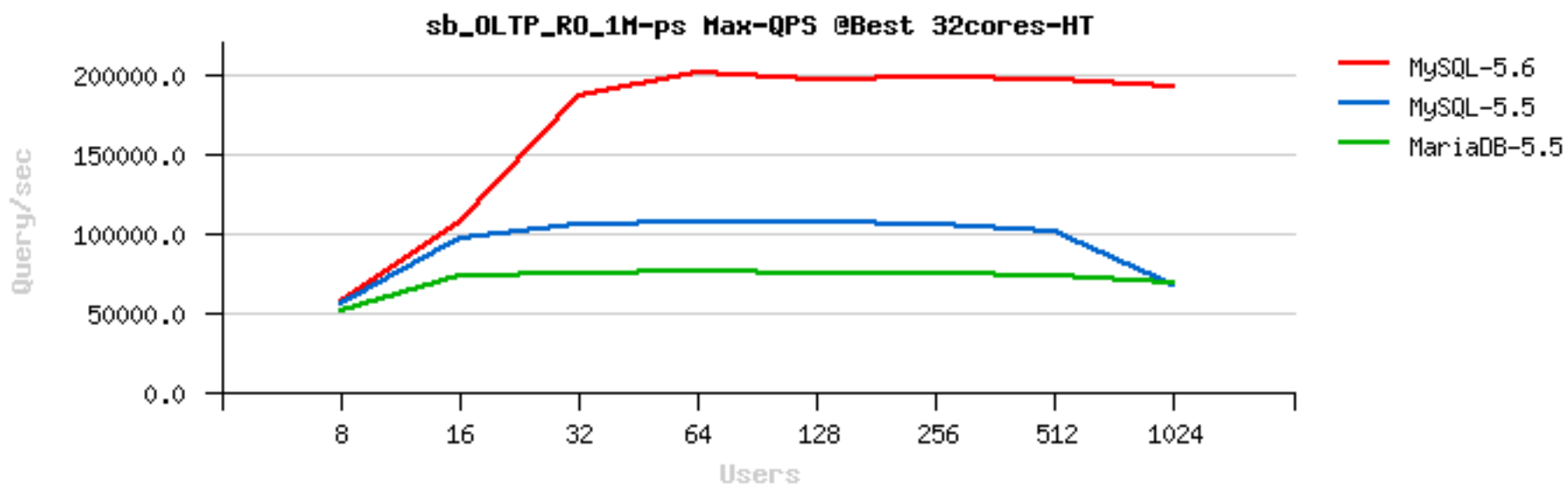
ORACLE

Sysbench OLTP_RO @16cores-HT (Apr.2013)



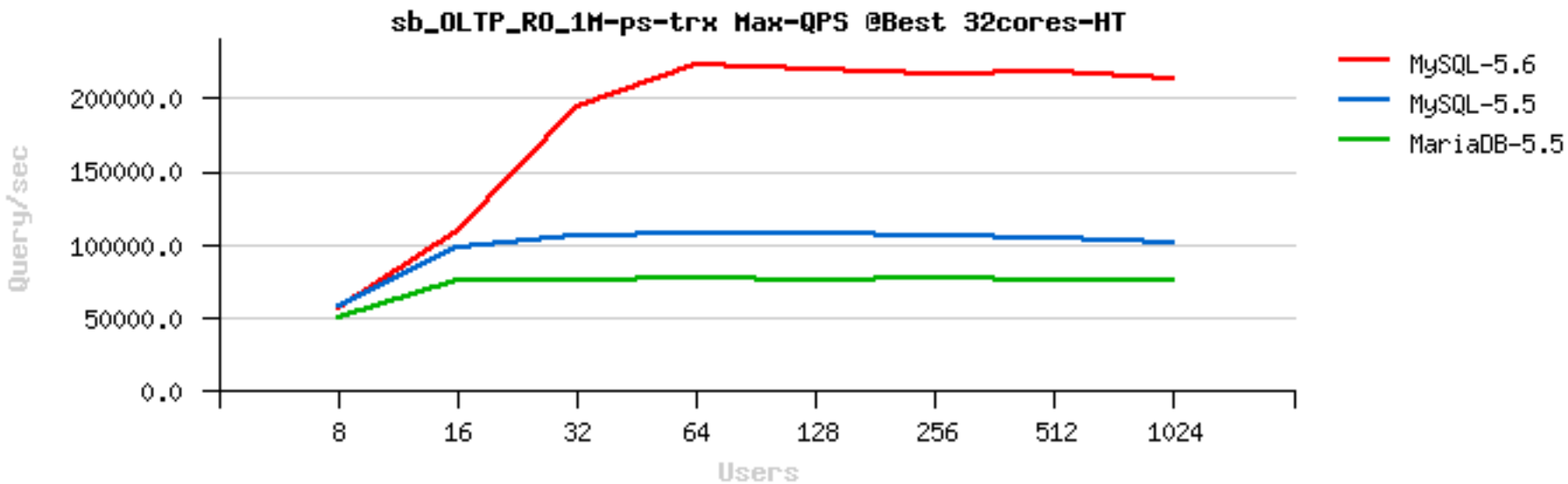
ORACLE

Sysbench OLTP_RO @32cores-HT (Apr.2013)



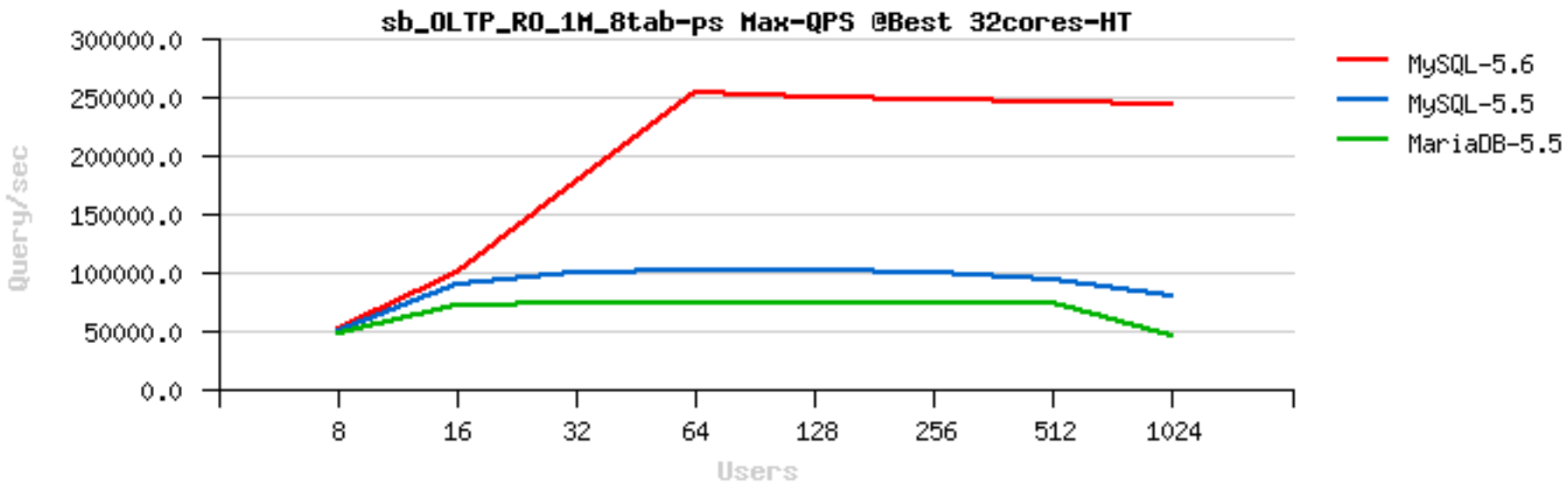
ORACLE

Sysbench OLTP_RO-trx @32cores-HT (Apr.2013)



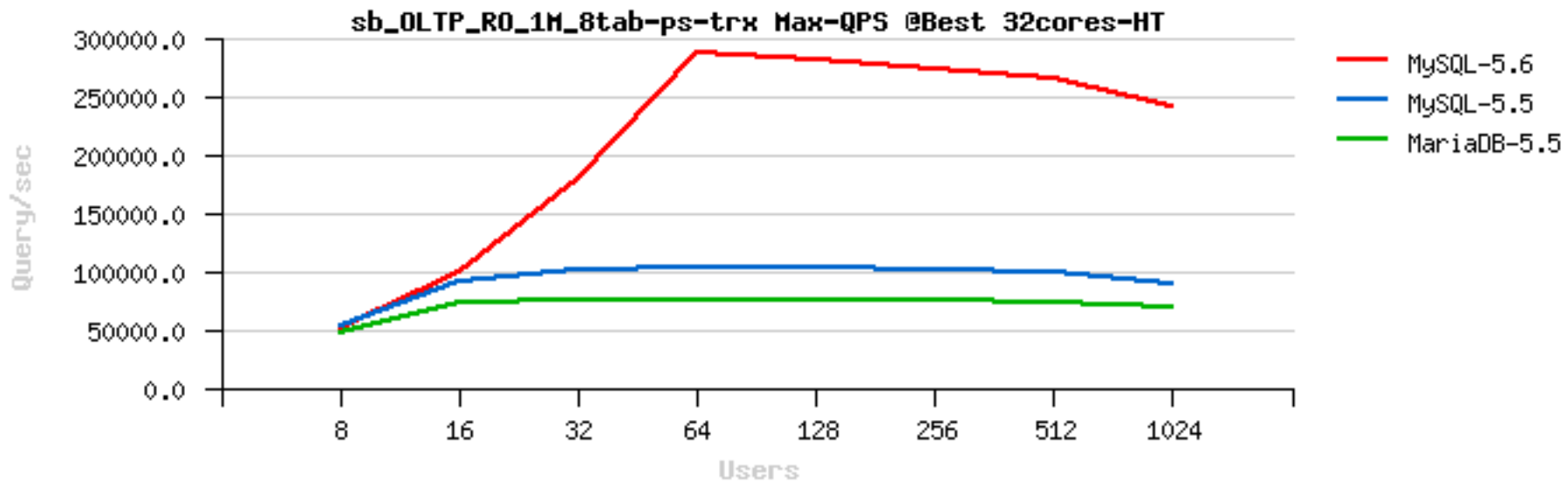
ORACLE

Sysbench OLTP_RO 8-tab @32cores-HT (Apr.2013)



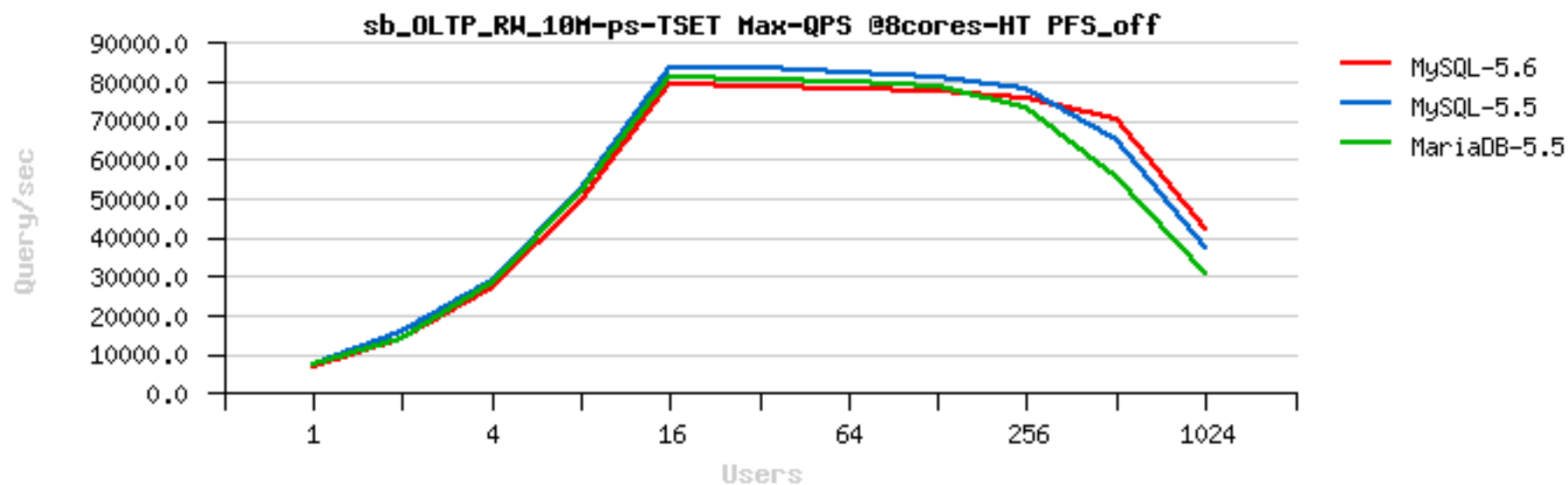
ORACLE

Sysbench OLTP_RO-trx 8-tab @32cores-HT (Apr.2013)



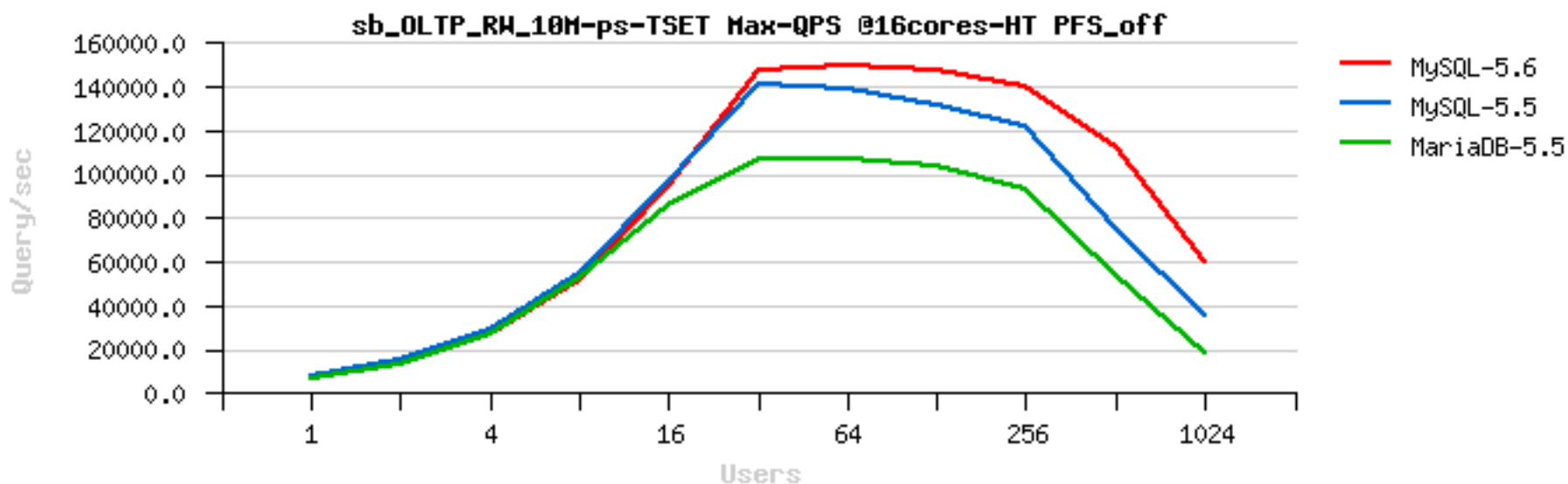
ORACLE

Sysbench OLTP_RW @8cores-HT (Apr.2013)



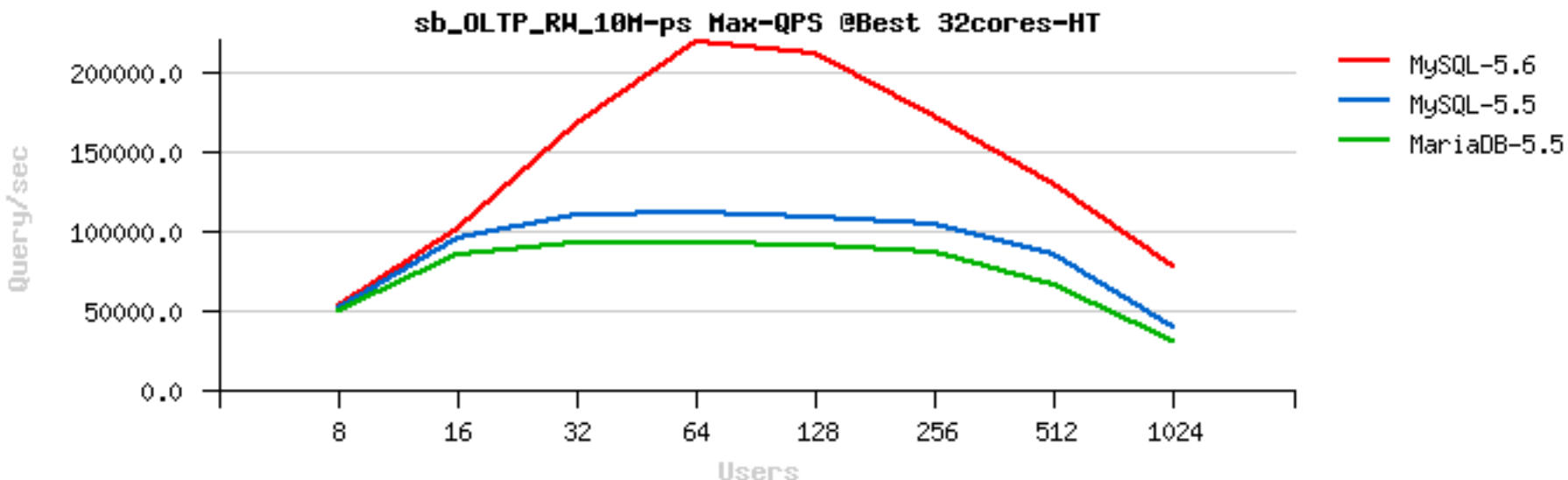
ORACLE

Sysbench OLTP_RW @16cores-HT (Apr.2013)



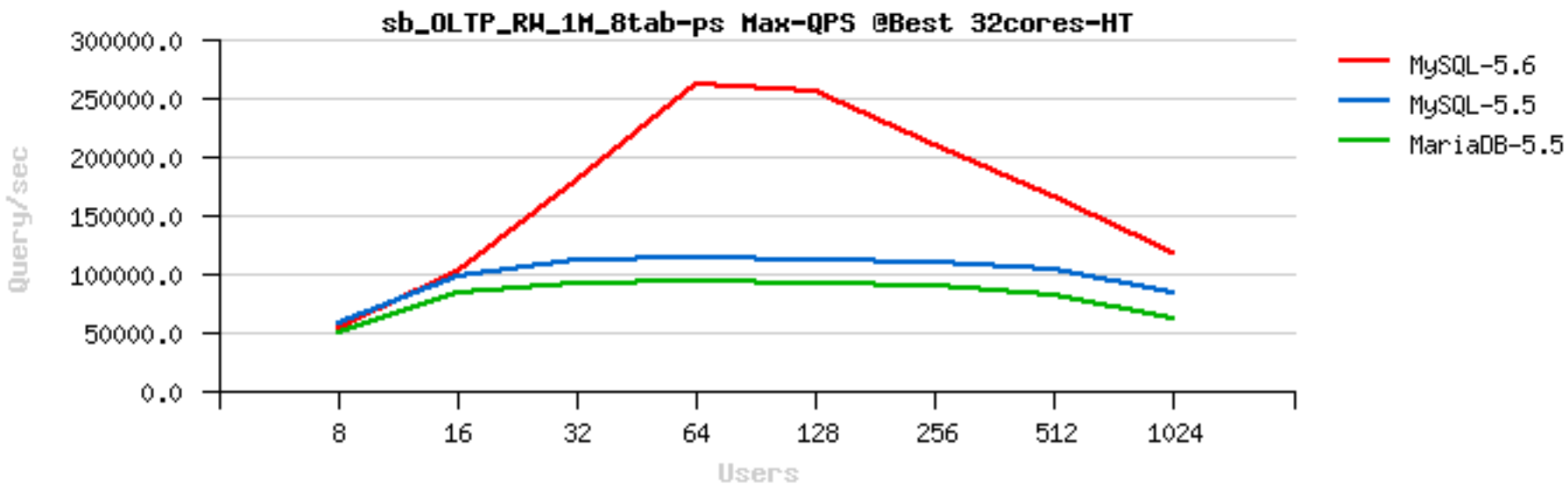
ORACLE

Sysbench OLTP_RW @32cores-HT (Apr.2013)



ORACLE

Sysbench OLTP_RW 8-tab @32cores-HT (Apr.2013)



ORACLE

Things are changing constantly, stay tuned ;-)

- MySQL/InnoDB Scalability:
  - 2007 : up to 2CPU...
  - 2008 : up to 4CPU cores
  - 2009 : up to 16CPU cores (+Sun)
  - 2010 : up to 32CPU cores (+Oracle)
  - 2012 : up to 48CPU cores..
  - 2014 : …?? ;-)
  - NOTE: on the same HW performance is better from version to version!
- InnoDB today:
  - At least x4-8 times better performance than 2-3 years ago ;-)
  - Capable of over ~~100K~~ ~~300K~~ **500K** QPS(!) + FTS & Memcached

ORACLE

MySQL 5.6: Pending issues

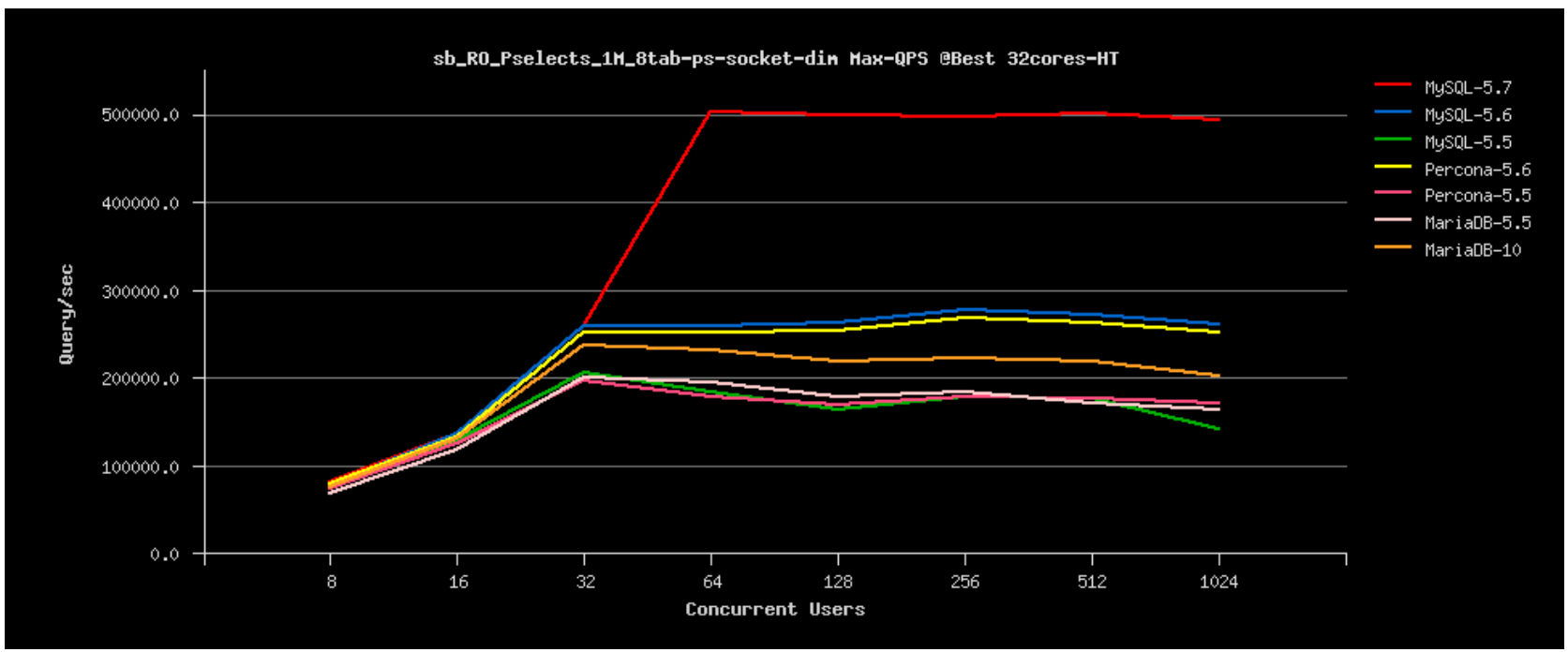
- Index lock..
- Lock_sys contention..
- Trx_sys contention..
- MDL scalability..
- Flushing limits..
- LRU flushing..
- Design bug on block locking.. (was here from the beginning)
- Not able yet to use 100% I/O capacity on a powerful storage..
- “Mysterious” contentions on dbSTRESS..
- etc..

MySQL 5.7: Work in progress.. ;-)

- Index lock.. <== fixed !
- Lock_sys contention.. <== lowered !
- Trx_sys contention.. <== improved a lot !!!
- MDL scalability.. <== in progress..
- Flushing limits.. <== in progress..
- LRU flushing.. <== in progress..
- Design bug on block locking.. (was here from the beginning)
- Not able yet to use 100% I/O capacity on a powerful storage..
- “Mysterious” contentions on dbSTRESS..
- Etc.. <== well, ALL in progress / investigation ;-)

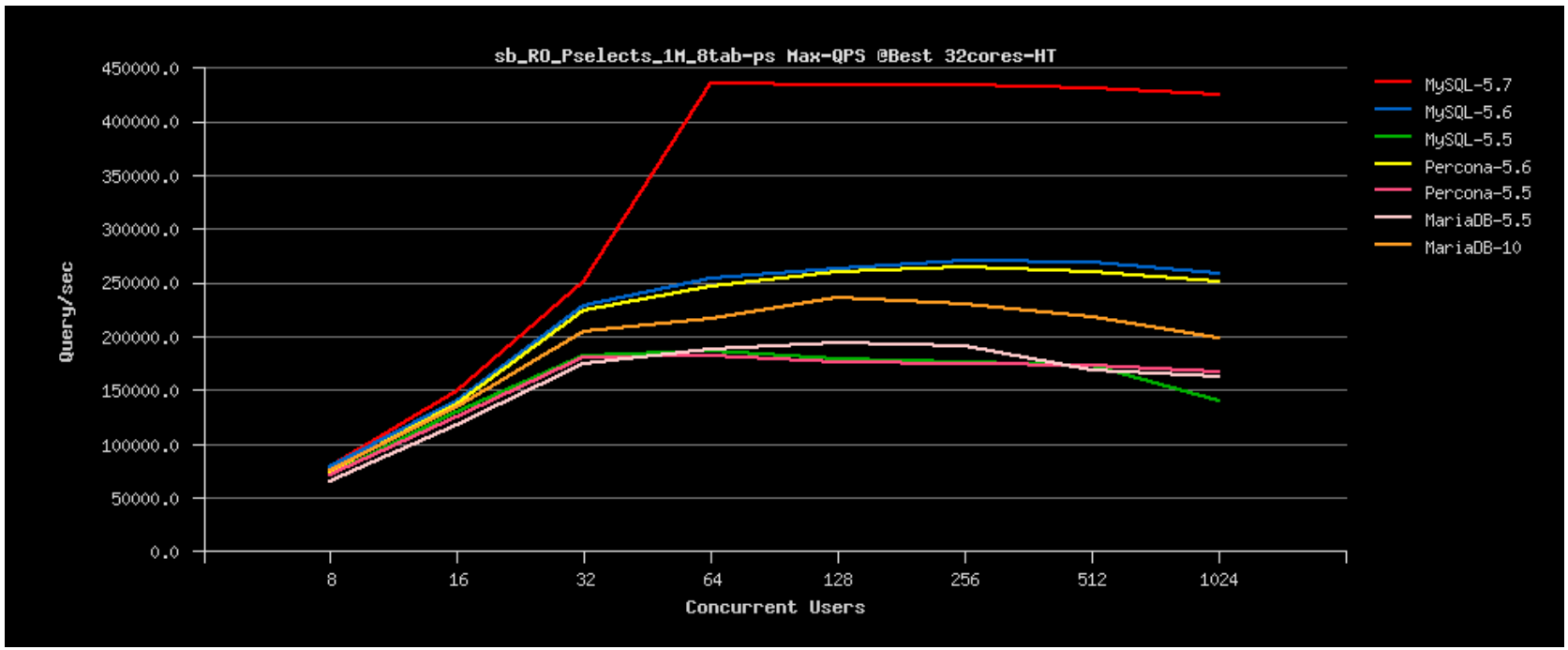
MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RO Point-Selects 8-tables: **500K** QPS !!!
  - UNIX socket, sysbench 0.4.8 (older, using less CPU)



MySQL 5.7: DMR2 (Sep.2013)

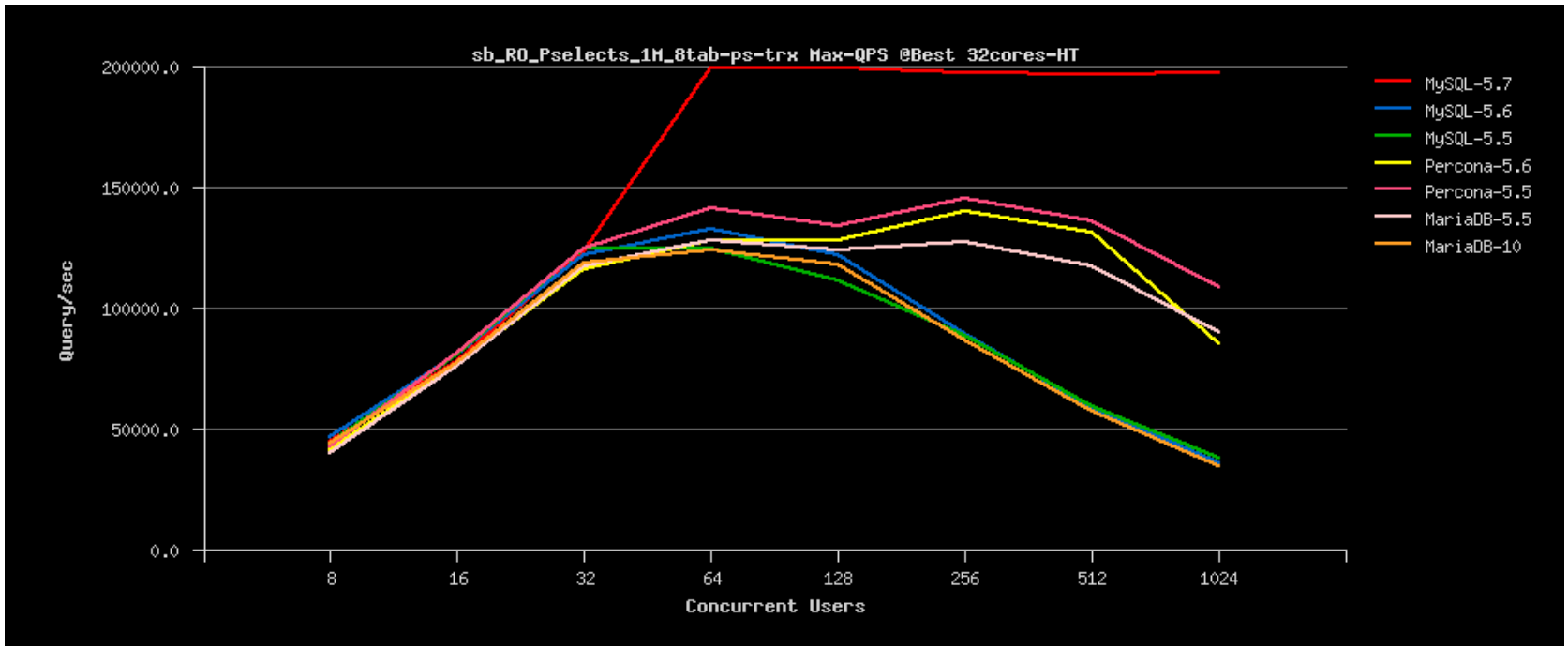
- OLTP_RO Point-Selects 8-tables: **440K** QPS
  - IP port, sysbench 0.4.13 (“common”, using more CPU)



ORACLE

MySQL 5.7: DMR2 (Sep.2013)

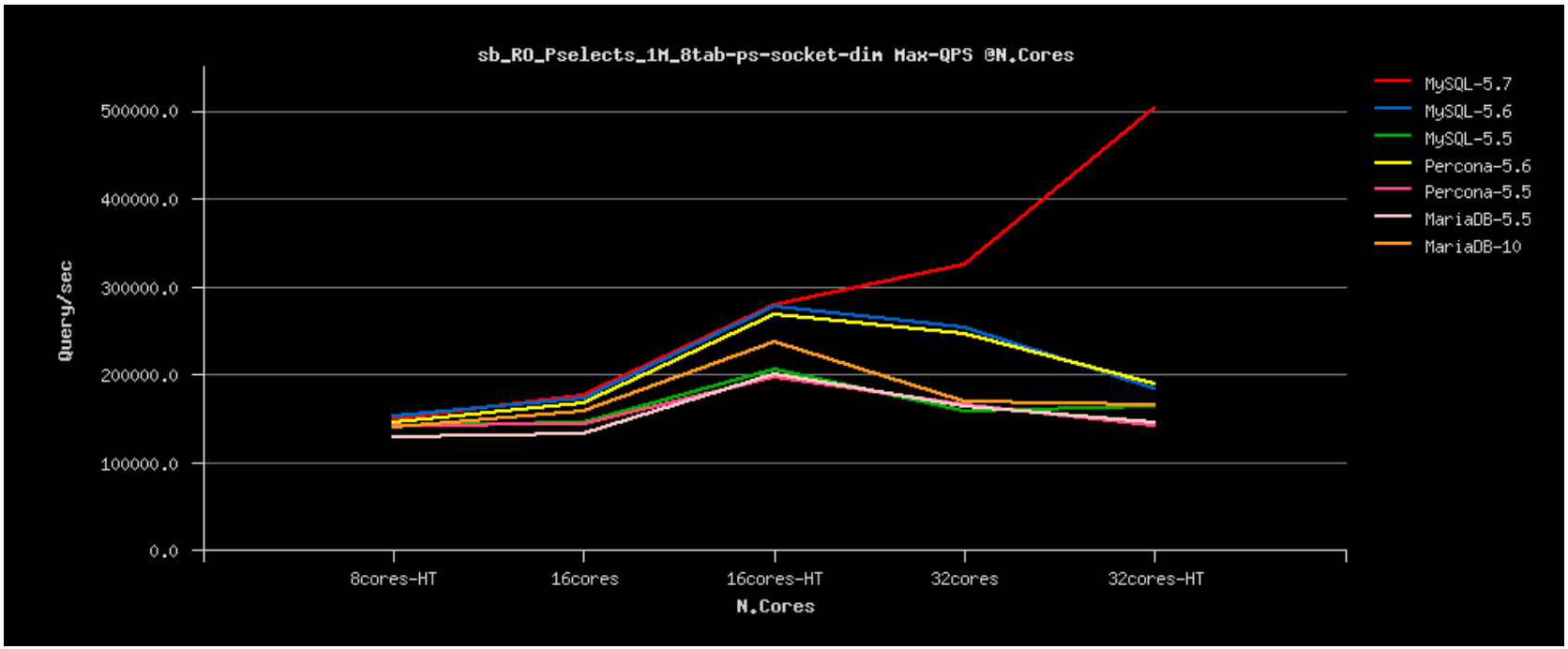
- OLTP_RO Point-Selects-TRX 8-tables: **200K** QPS
  - IP port, sysbench 0.4.13 (“common”, using more CPU)



ORACLE

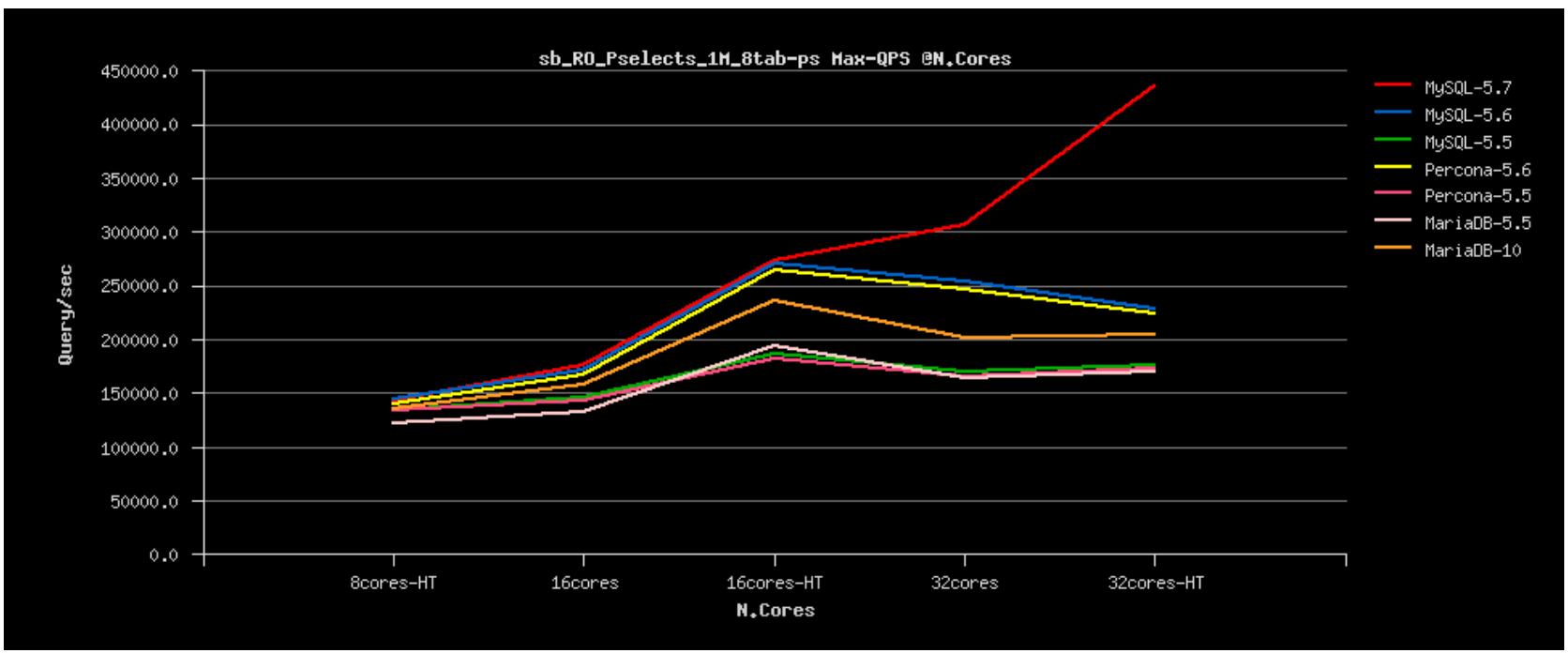
MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RO Point-Selects 8-tables: Scalability..
  - UNIX socket, sysbench 0.4.8 (older, using less CPU)



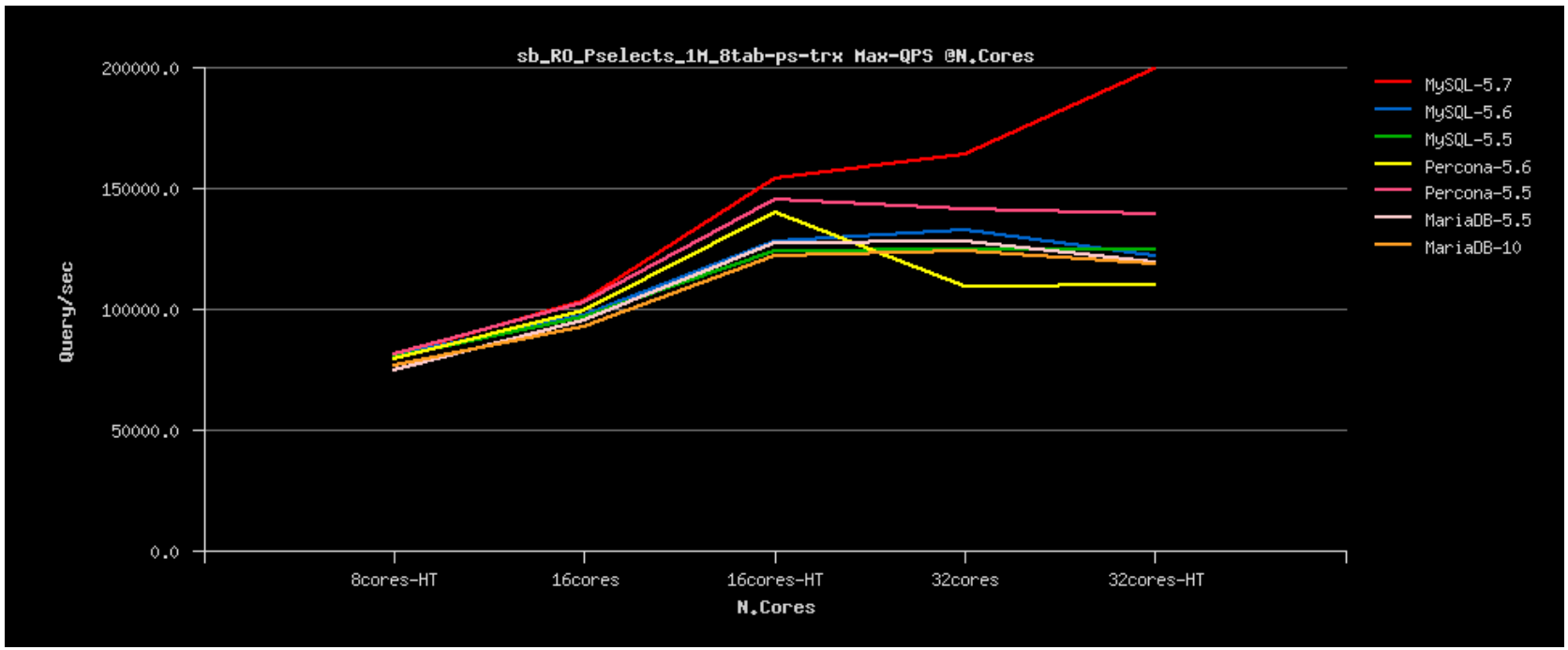
MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RO Point-Selects 8-tables: Scalability..
  - IP port, sysbench 0.4.13 (using more CPU)



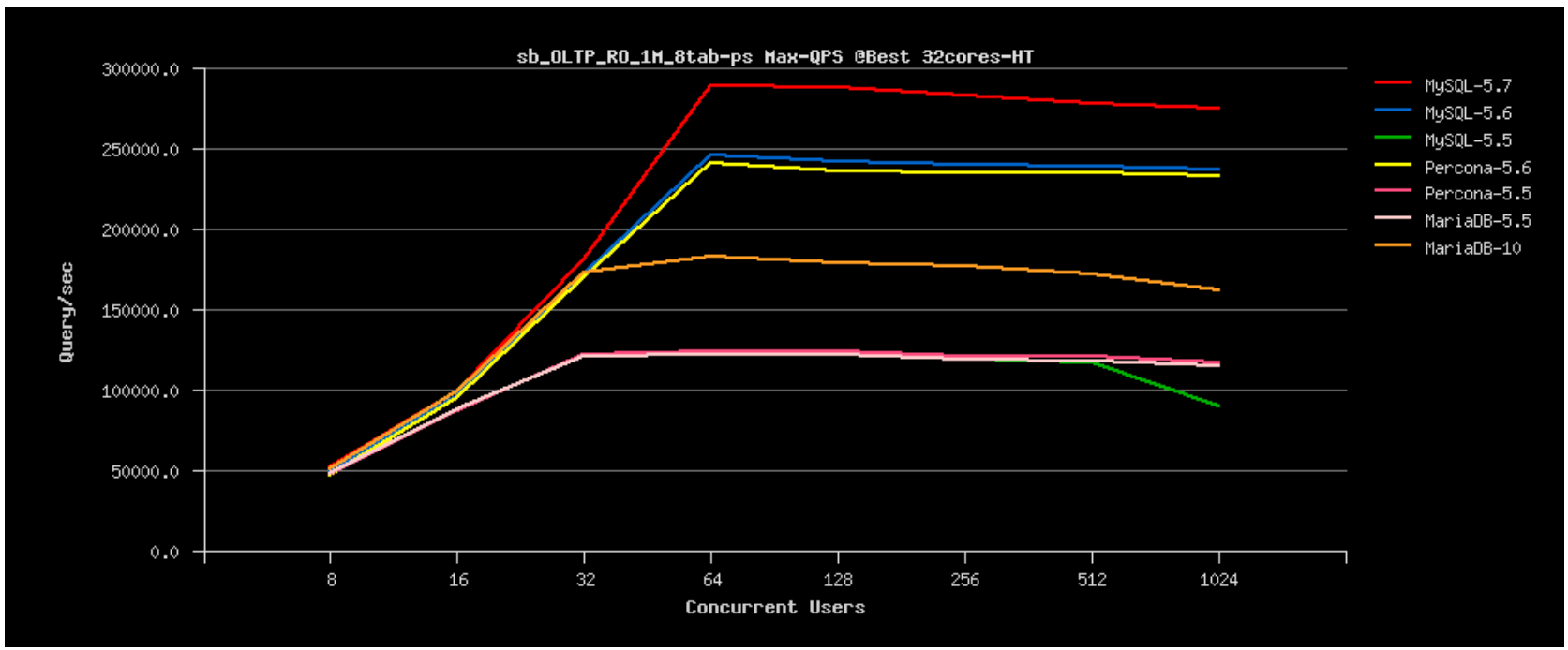
MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RO Point-Selects-TRX 8-tables: Scalability..
  - IP port, sysbench 0.4.13 (using more CPU)



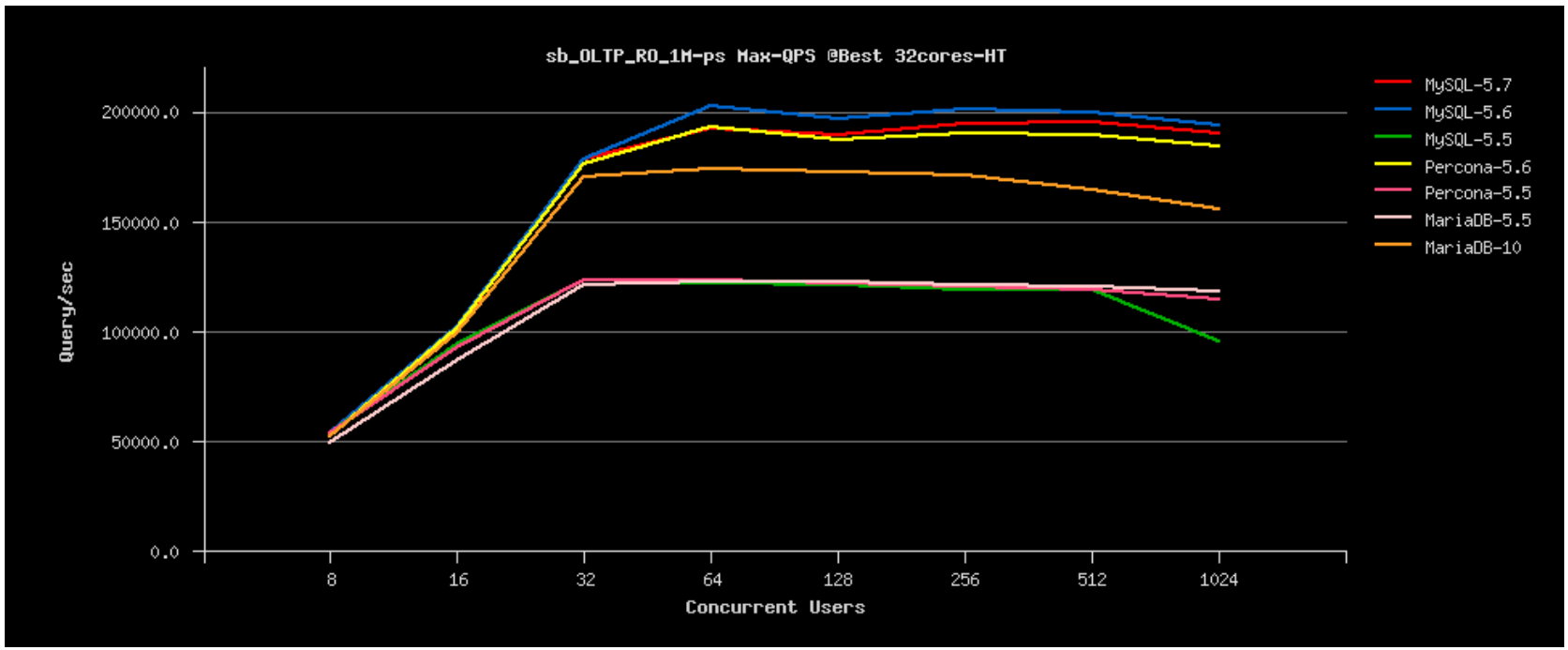
MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RO 8-tables: **280K** QPS
  - IP port, sysbench 0.4.13 (“common”, using more CPU)



MySQL 5.7: DMR2 (Sep.2013)

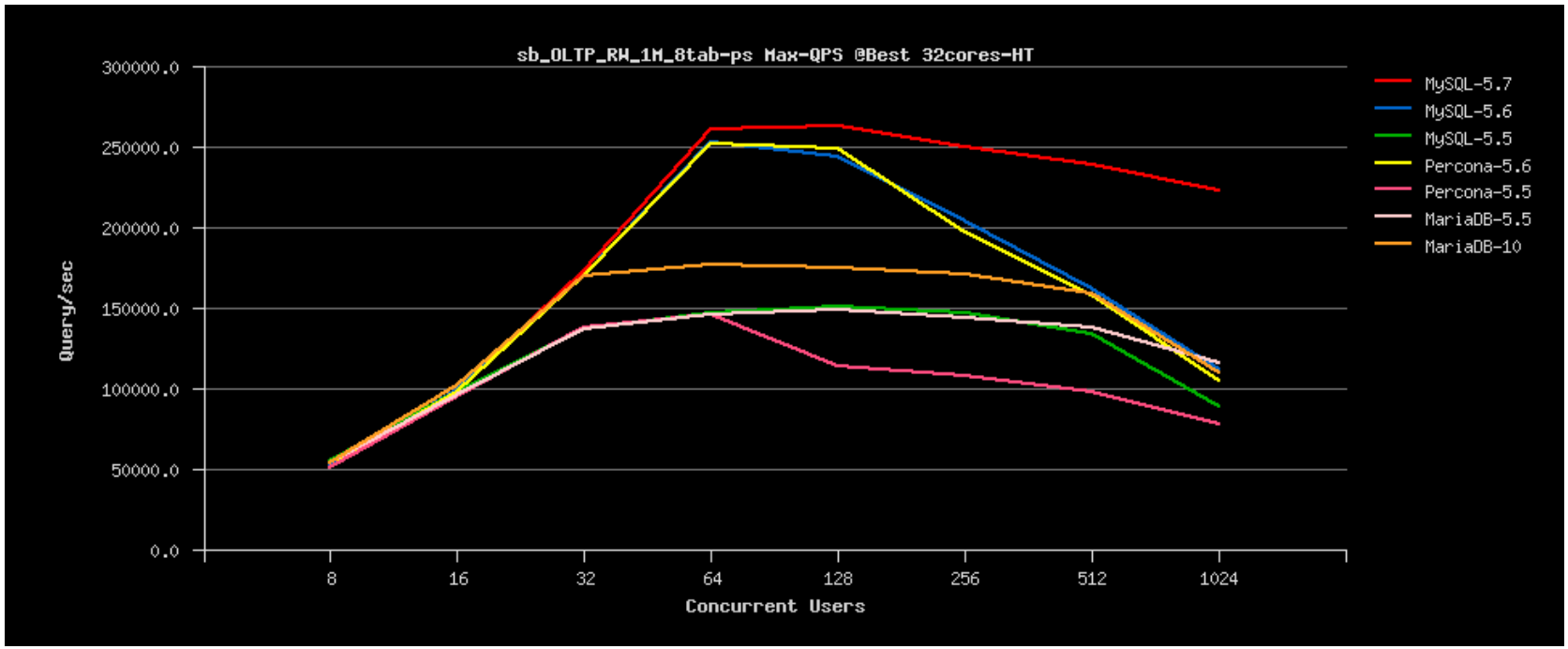
- OLTP_RO 1-table: lower than 5.6...
  - Due higher MDL contentions, work in progress..



ORACLE

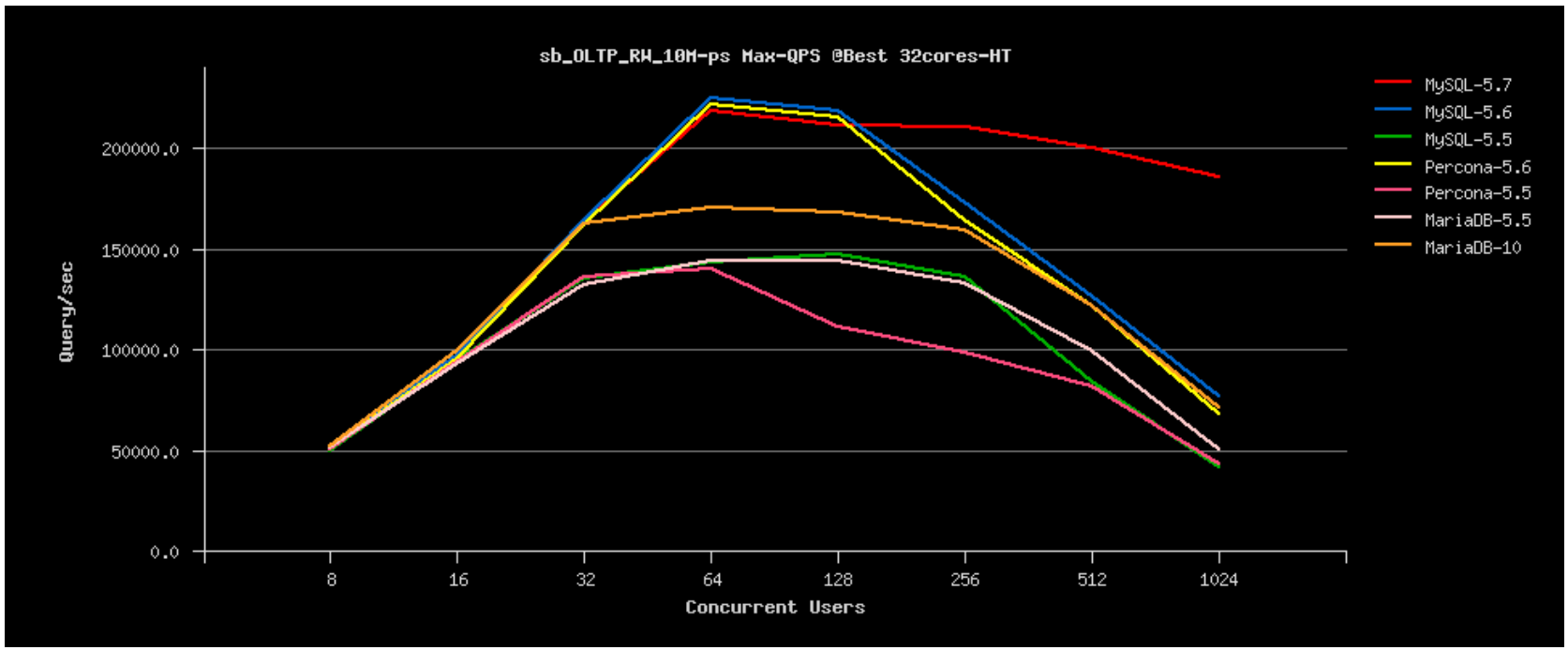
MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RW 8-tables: **265K** QPS
  - IP port, sysbench 0.4.13 (“common”, using more CPU)



MySQL 5.7: DMR2 (Sep.2013)

- OLTP_RW 1-table: lower than 5.6
  - MDL contention, work in progress..



ORACLE

Few words about *dim_STAT* (as you'll ask ;-))

- All graphs are built with ***dim_STAT*** (http://dimitrik.free.fr)
  - All System load stats (CPU, I/O, Network, RAM, Processes,...)
  - Manly for Solaris & Linux, but any other UNIX too :-)
  - Add-Ons for Oracle, MySQL, PostgreSQL, Java, etc.
  - MySQL Add-Ons:
    - mysqlSTAT : all available data from “show status”
    - mysqlLOAD : compact data, multi-host monitoring oriented
    - mysqlWAITS : top wait events from Performance SCHEMA
    - InnodbSTAT : most important data from “show innodb status”
    - innodbMUTEX : monitoring InnoDB mutex waits
    - innodbMETRICS : all counters from the METRICS table
  - And any other you want to add! :-)

THANK YOU !!!

- All details about presented materials you may find on:
  - http://dimitrik.free.fr - dim_STAT, Benchmark Reports
  - http://dimitrik.free.fr/blog - Articles about MySQL Performance