MySQL

ORACLE

MySQL 8.0 Performance: Our Linux Stories..

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 2011 “officially” @MySQL Performance full time now
- http://dimitrik.free.fr/blog / @dimitrik_fr

IO_uring

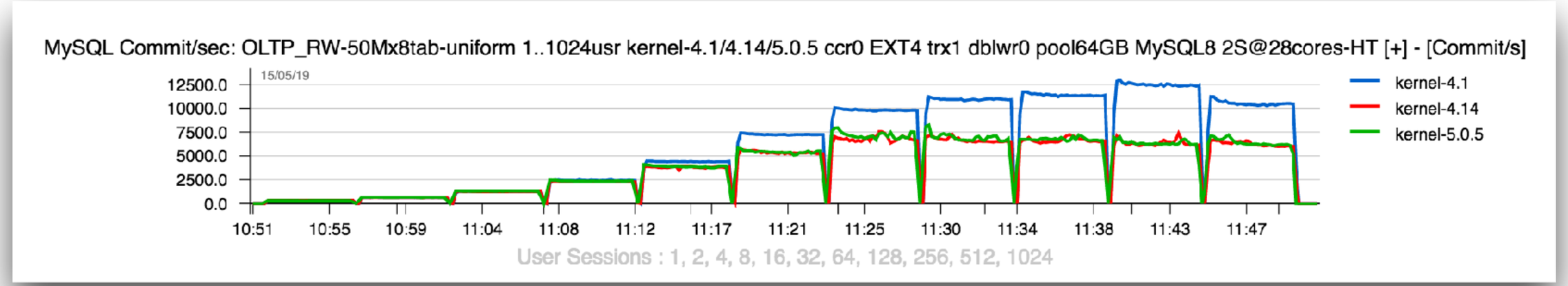
- stable enough ?
- upstream ?
- always “same or better” than current AIO ?
- any feedback ? => please, share !

O_ATOMIC

- InnoDB Double Write (DBLWR)
  - the only way (currently) to guarantee non-corrupted page writes (16K)
  - cons: x2 more IO writes than really needed..
- Fusion-io :
  - proprietary storage + kernel driver + FS => atomic IO writes
  - 6 years ago proposed kernel patch to implement O_ATOMIC feature
- Any expectations to see O_ATOMIC upstream ?
  - at least for O_DIRECT ?..
  - please, comment..

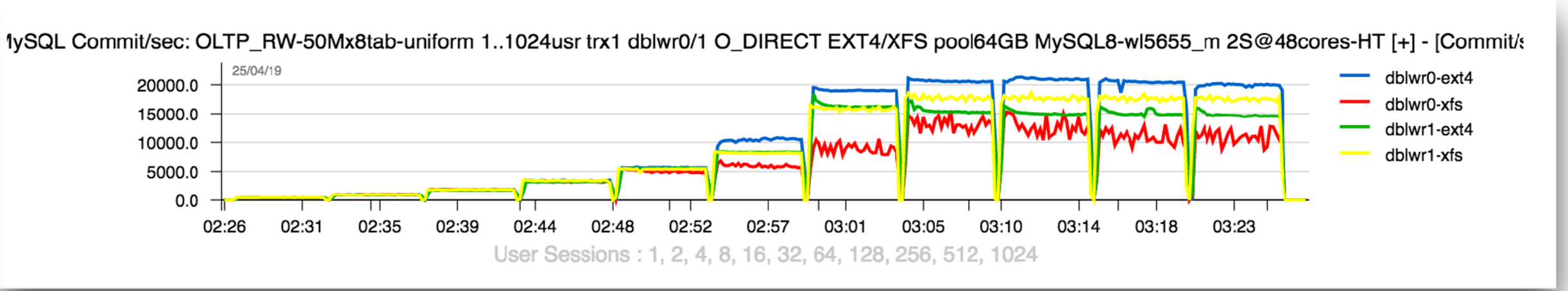
MySQL 8.0 IO-bound Workload @EXT4

- IO-bound Sysbench OLTP_RW @28cores-HT
  - ext4 and Linux kernels :



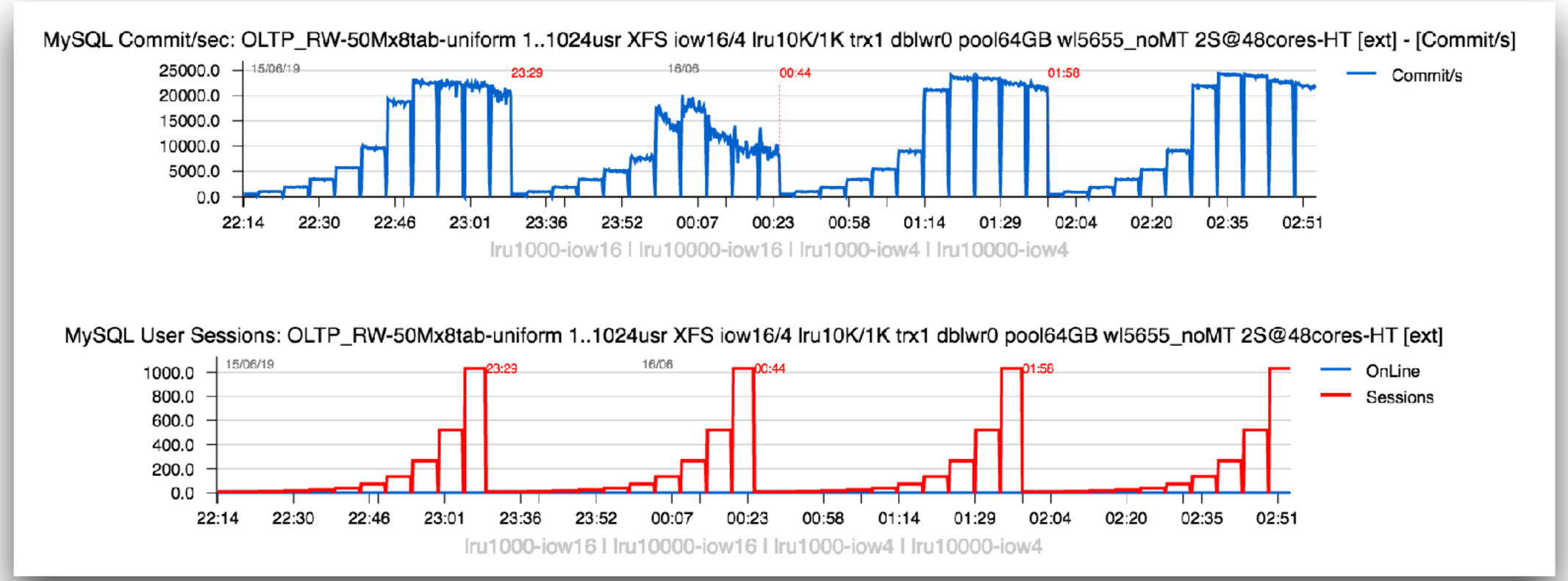
InnoDB Double-Write (DBLWR) and EXT4/XFS..

- DBLWR impact on IO-bound OLTP_RW @48cores-HT & Fast SSD :
  - dblwr=0 : ext4 is doing better than XFS
  - dblwr=1 : XFS is doing better than ext4
  - XFS : doing better with dblwr=1 than with dblwr=0 !!! => WTF ?..



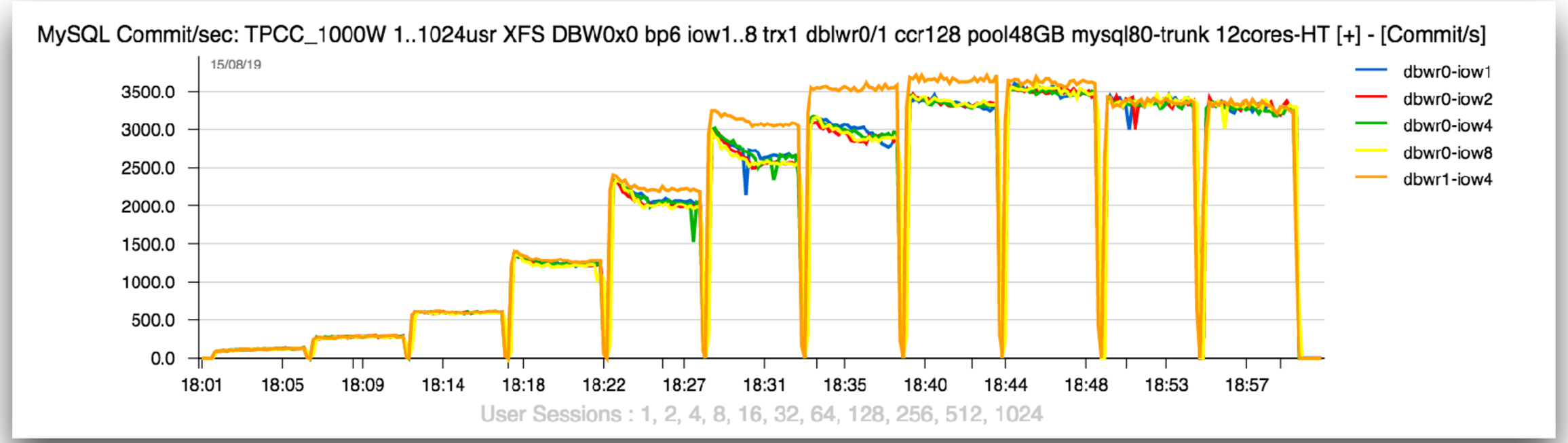
XFS

- IO-bound OLTP_RW, dblwr=0
  - Storage : Optane SSD (very fast)
  - Workaround : tuning InnoDB IO-write threads.. => why ? ;-))



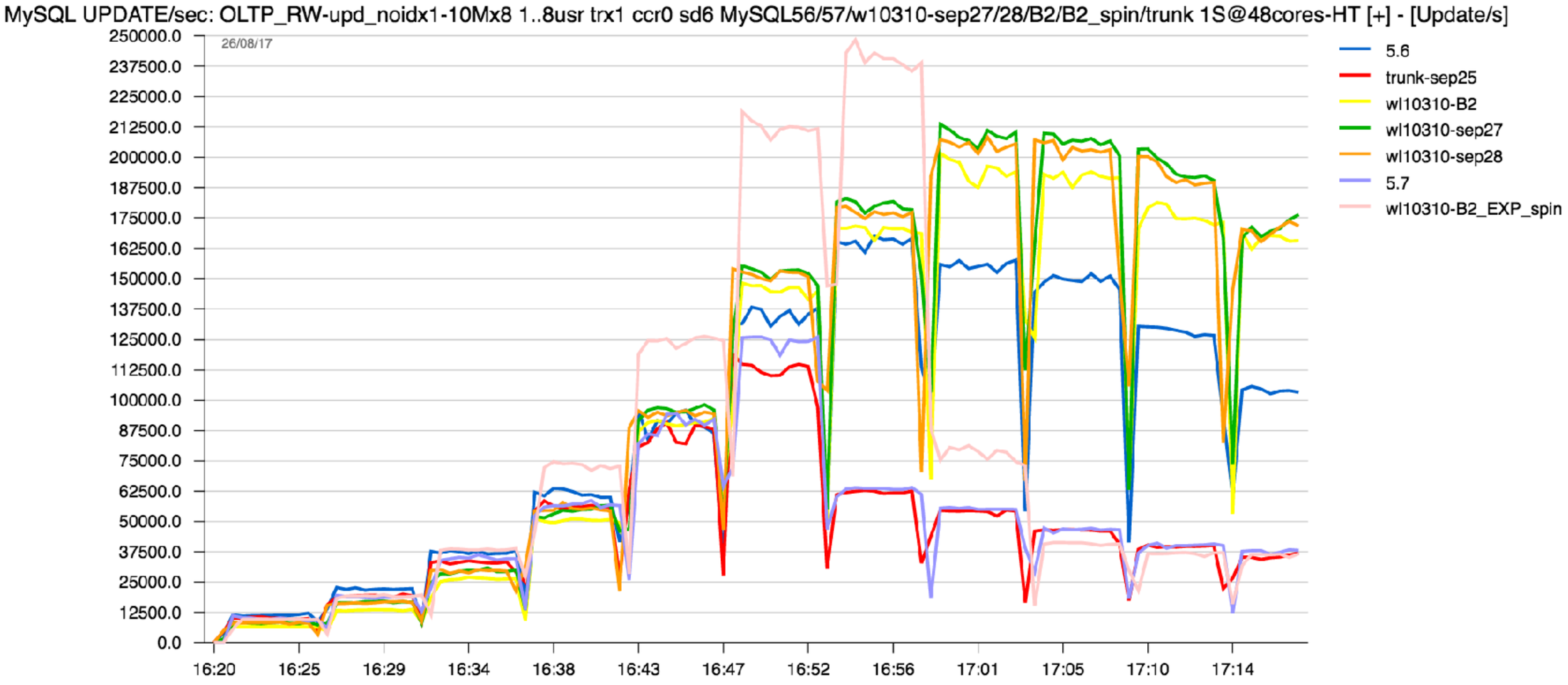
XFS

- IO-bound TPCC 1000W, dblwr=0/1
  - Storage : SATA SSD (not fast)
  - Workaround : nope.. ;-))
  - any ideas ?..



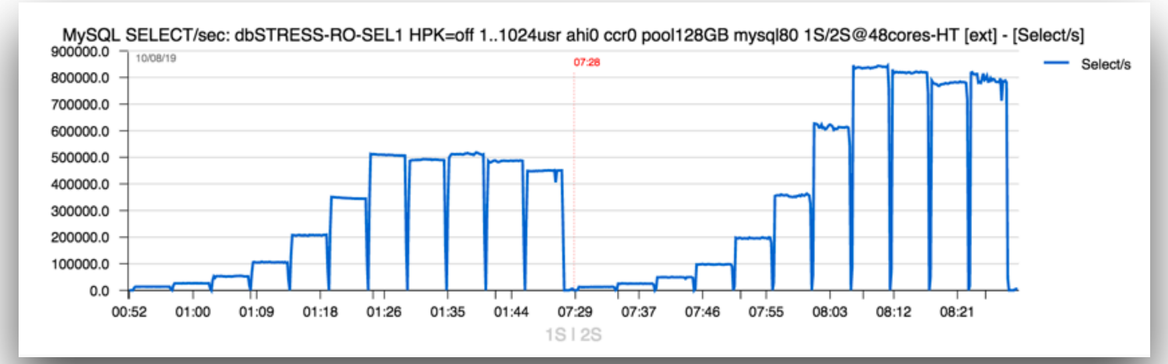
MySQL 8.0 : Re-Designed REDO

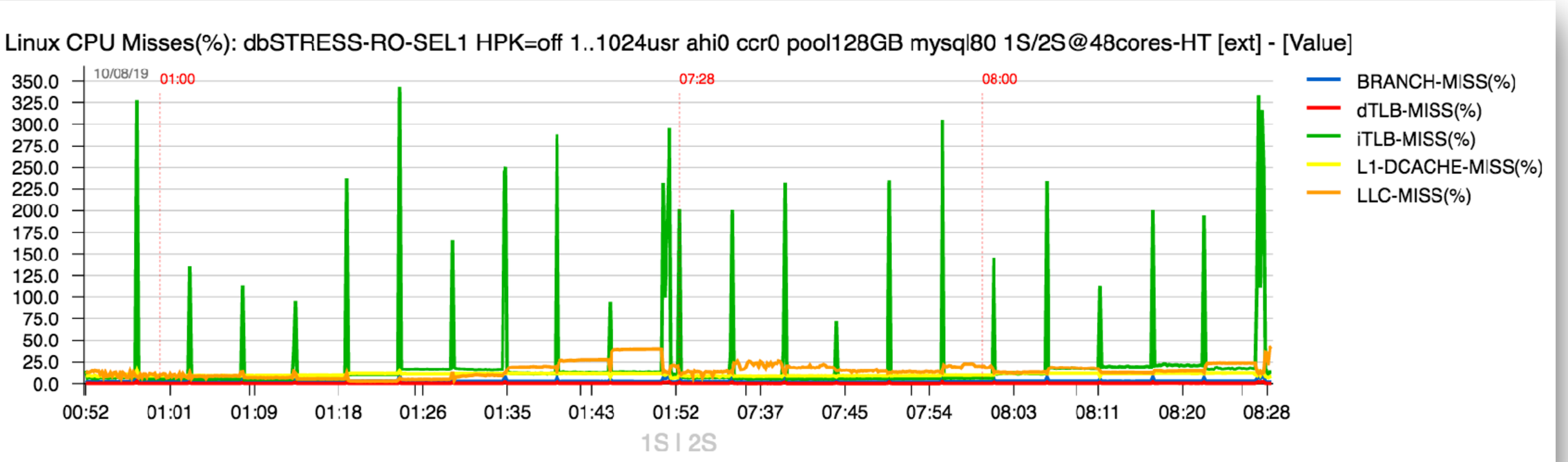
- New design tradeoffs.. — “to spin or not to spin”



dbSTRESS-RO-SEL1

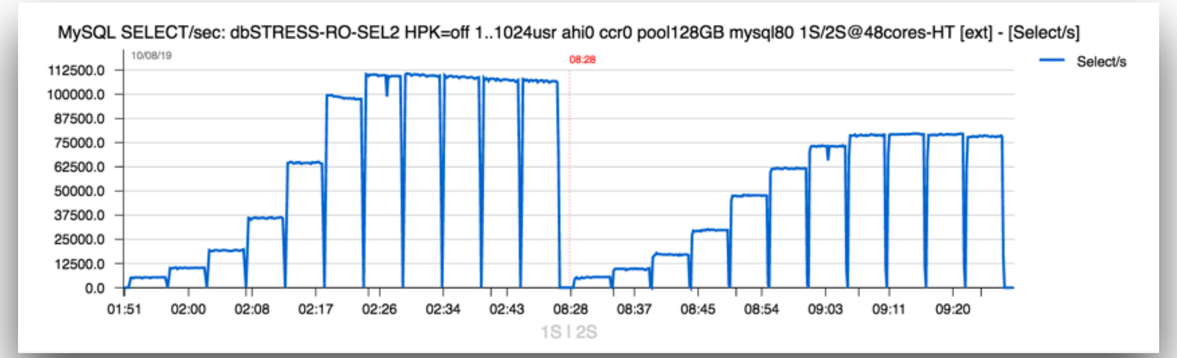
- SEL1 :
  - scaling from S1 to S2

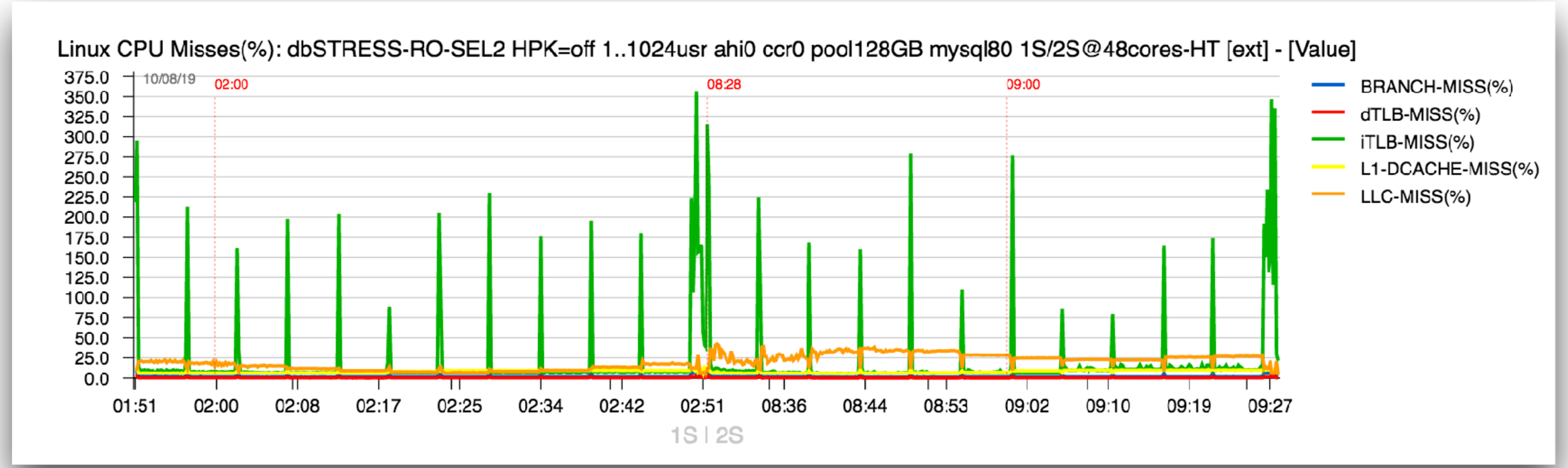




dbSTRESS-RO-SEL2

- SEL2 :
  - NOT scaling from S1 to S2

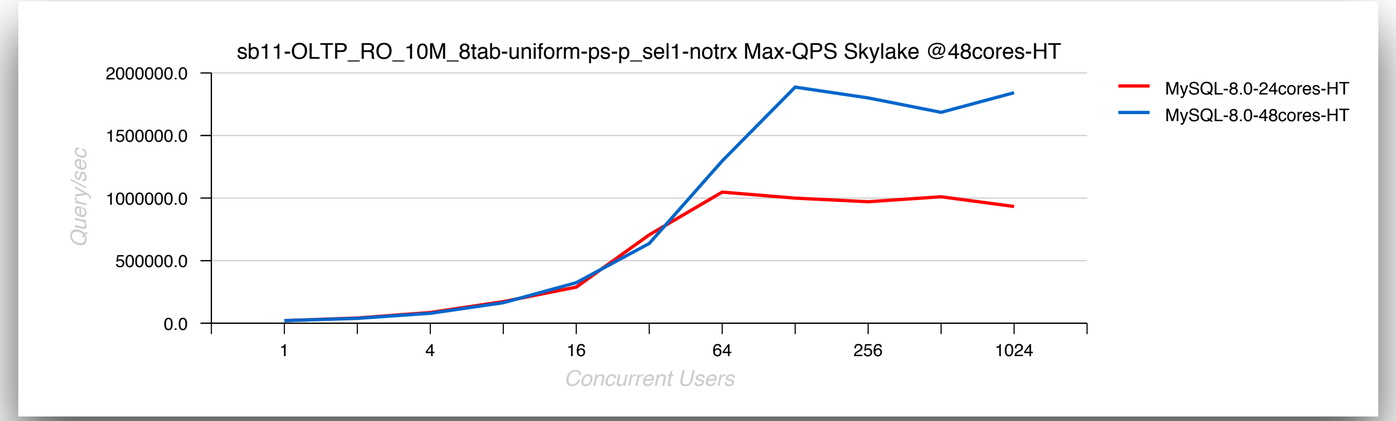


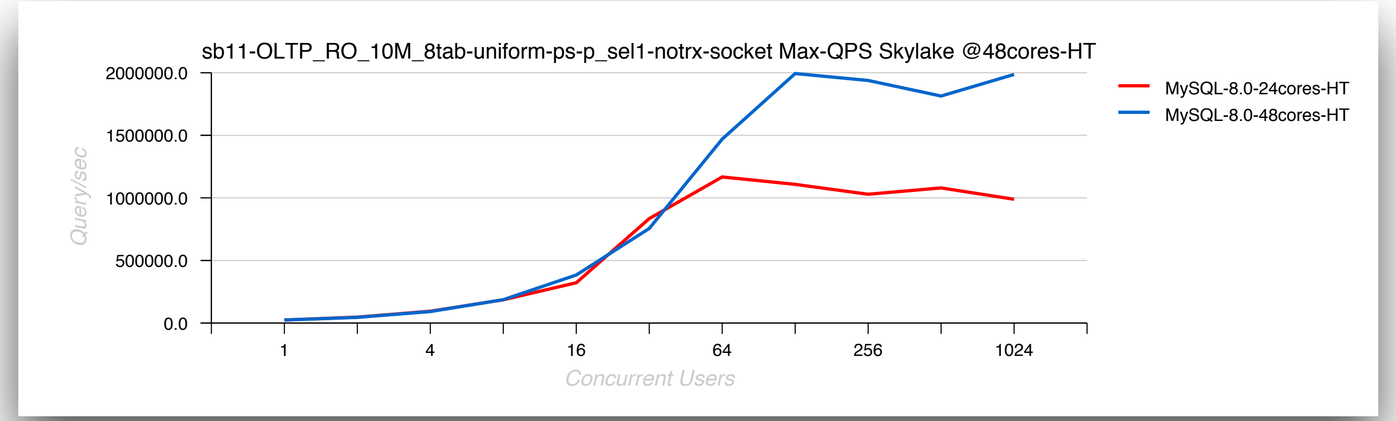


ORACLE

Point-Selects : IP port -vs- UNIX socket

- Skylake 2S, 48cores-HT :
  - up to **19%** difference !
  - HW ? OS kernel ?
  - IP stack ?
- ARM64 :
  - similar tendency..
  - so ?.. ;-))





Backlog

- Backlog :
  - OS level (IP tuning)
  - MySQL level (socket option)
- Common workload :
  - connect
  - query exec
  - disconnect
  - => massive connect / disconnect..
- Observation :
  - Backlog = 0 => up to 15% better performance..
  - any reason ?..