



Choosing Storage Systems

For MySQL

Peter Zaitsev,
CEO Percona
Percona Live MySQL Conference and Expo 2013
Santa Clara, CA
April 25, 2013



PERCONA
LIVE

Why Right Choice for Storage is Important ?

2

- ... because Wrong Choice will often bite
- Performance Problems
- Downtime
- Data Loss

Storage System Requirements

3

- Performance
- Size
- Reliability
- Additional Features

- “Random IO” is the most important
 - Especially for reads
- Look at “Latency” (response time)
- Performance at different concurrency level
 - Number of outstanding IOps

Sequential IO

- Analytical Use
- Backup/Restore
- Fast Warmup (BP preload)

- Fast acknowledgement for writes
 - Durable writes to database transactional log
- MetaData access Performance
 - On local file systems these come from cache
 - Network file system can have issues

How much do we need ?

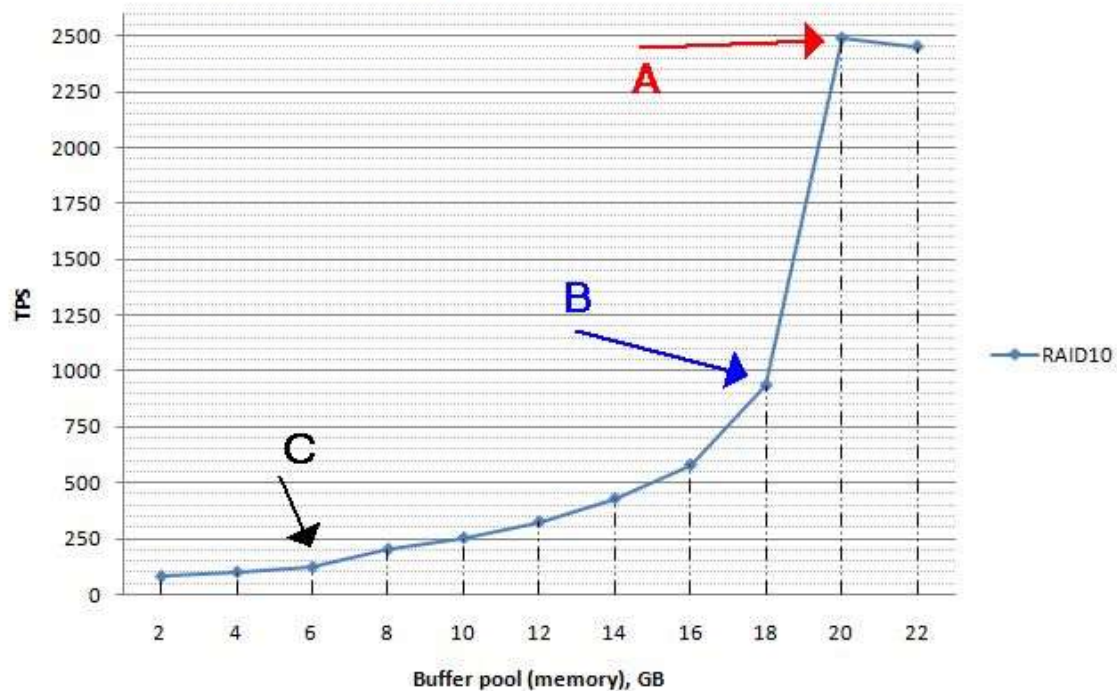
7

- Benchmark
 - How Capable is current system
 - What are the best options ?
- Extrapolate Benchmark results
 - Typical cause of mistake
- Need right Model for Extrapolation

- Often performance needs can be met by More memory or Faster storage
 - Avoid Reads
 - Reduce and Delay Writes
- Keep in Mind: Warmup Issues
 - Fast warmup in MySQL 5.6 and Percona Server

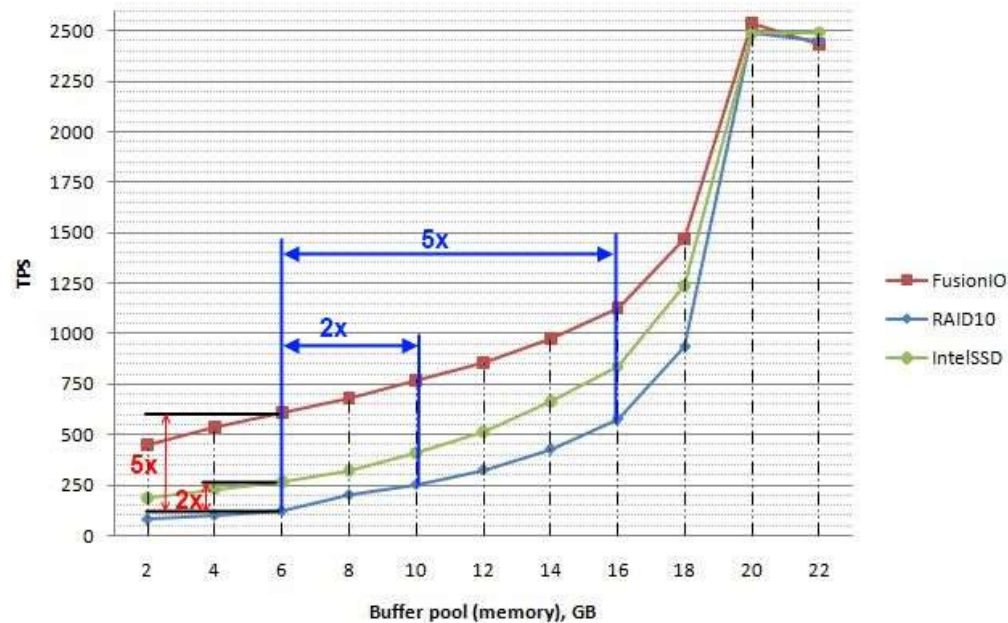
Performance vs Amount of Memory

sysbench oltp, 80mln rows (18GB data)



“Compression” of Performance Advantage

sysbench oltp, 80mln rows (18GB data)



- You need enough space to keep your database
 - Logs; Allow space for unforeseen
- Can you grow volume size online ?
- Flash storage performance is often impacted by used space
 - TRIM

- Protection from Hard Drive Failure (RAID)
 - No storage eliminate chance of data loss
- Good Flash storage is as reliable as RAID card
- MultiPath/Redundant network
- Durability Guaranty
 - No loss of acknowledged data

- Requirements range a lot
 - Just to need to place my files
 - “Enterprise Storage Features”
- Size Scaling
- Performance Scaling

- Replicated Storage
 - DR or Clustering
- Encryption
- Deduplication
- Compression
- Snapshots/Cloning
- Backup Integration

What is your environment ?

15

- Bare Metal
 - Your own Data Center or Rental
- Cloud
- Virtualized Environment/Private Cloud

Key Components to consider

16

- Hardware
- Operating System
- File System and Storage Configuration
- MySQL

HARDWARE

Conventional Hard Drives

- Great if need space 4TB costs less than \$200
- 2.5” 15k RPM fastest, very expensive
- You can go by specs. Vendors are really close
- “Desktop” hard drives might have problems with RAID

- This is where you go for performance !
 - And better power usage too
- SLC, MLC, eMLC
- SATA drives and PCI-E cards



Things to think about

20

- Amount of Writes
 - Performance and longevity factor
- Space
- Reliability
- Cost
- You might get more IO than MySQL can handle

Rapidly improving technology

- A lot of difference between vendors
- Firmware and drivers are key
 - Large difference between releases
- Hard to give specific model advice
- Evaluating with your workload is best idea

SATA flash Options

- “Enterprise” and “Laptop” Markets
- Intel
- Samsung

PCI-E Flash Cards

- FusionIO
- Virident
- Micron
- Intel

- Idea: Store most accessed data on flash
- Many Options Available
 - Hybrid Hard drives
 - RAID with SSD cache option
 - SAN w Tiered storage
 - FlashCache at OS level or File System
 - Manually place some data on Flash

- Can cause huge impact for PCI-E Performance
- Faster “Cores” beat more cores
 - MySQL in rarely needs more than 2 sockets
- Keep System balanced

- Are you fitting your data set in memory ?
 - Working set ?
- Very good investment up to a point
- Much faster even than fastest Flash
- More memory beats Faster memory

- For Flash or Conventional Drives ?
 - Flash requires Flash-aware RAID controllers
- Software RAID options (and ZFS)
- Write Cache on RAID is helpful for both
- Learning problem with “BBU” systems
 - Capacitor + Flash avoids this problem

- RAID10 most common choice
- RAID5 for less critical mostly read data
- RAID1 – operating system, logs etc
- Do not waste cache for “Read Caching”
- Read-ahead might be helpful if OS not doing it
- RAID stripe size varies
 - Benchmark if you have a chance

- Many vendors on the market
- Often outside of DBA choice at all
- Not used for Performance
 - Network latency is costly
- Fiber Channel and iSCSI common option
- Often crosses over to NAS

- NFS is most common file system
- Mind Configuration
 - Ensure all guarantees database expects are met
- Meta Data access costs can be surprise
- Fast Redundant Network
 - Bonding, 10G good for heavy loads

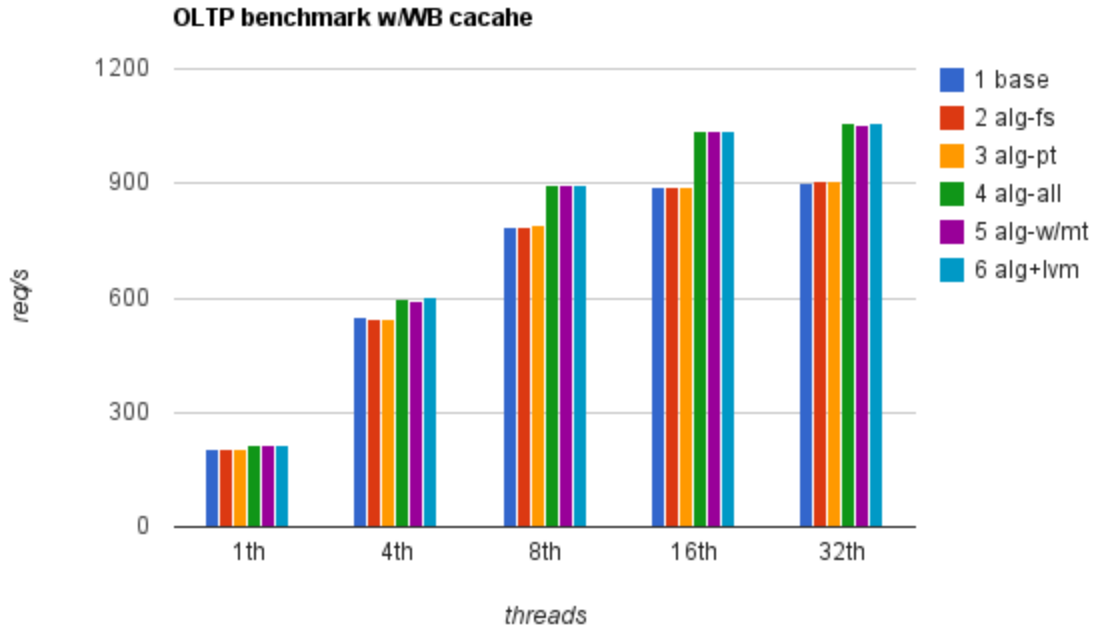
OPERATING SYSTEM AND BEYOND

- Linux most commonly used
- Recent version which supports your hardware
 - Especially Flash storage support is moving fast
- Scheduler on Linux
 - CFQ bad for some workloads
 - Check with Deadline and NOOP

- LVM
 - Great for growth, snapshots etc
 - Very low overhead (if not in snapshot mode)
- FlashCache
 - Using Flash as Cache
 - Originated from Facebook
 - Limited use

- Partition Alignment
- File System Alignment options
- Can cause significant performance difference
 - <http://bit.ly/ilfNkT>
 - <http://bit.ly/ks9trM>

Alignment Benchmarks



File System Choices

36

- EXT4
- XFS
- EXT3
- ZFS (coming up)

Things to consider

37

- Internal Concurrency
- Operational Issues
- How does it handle DIRECT IO
- Barrier on XFS
 - Might be good to disable with write cache
- Error handling: remount-ro

- Recent versions are better
 - Percona Server and MySQL 5.6 have a lot of SSD optimizations
- Check out Webinar on configuration Tuning
 - <http://bit.ly/UuNB1l>
 - Innodb Architecture Tutorial (slides will be posted)

MySQL storage tuning 101

39

- `Innodb_buffer_pool_size`
- `Innodb_log_file_size`
- `Innodb_flush_method=O_DIRECT`
- `Innodb_file_per_table`
- `Innodb_flush_log_at_trx_commit`

Virtualized Environments

- Most of the same apply
- Choices are often limited
- Watch overhead for high performance storage
- Understand what kind of features your Virtualization Technology provides for you

- Depends on the cloud provider
- Great features available
 - Use them
 - Snapshots, Backup etc

- Local storage and EBS available
- EBS performance can be unpredictable
- High Performance Options
 - Local Flash storage (fastest)
 - Provisioned IOPS for EBS

Reach me at pz@percona.com

THANK YOU !