



PERCONA
Performance Consulting Experts

Choosing Hardware and Operating Systems for MySQL

Apr 15, 2009

O'Reilly MySQL Conference
and Expo

Santa Clara, CA

by Peter Zaitsev, Percona Inc

We will speak about

- Choosing Hardware
- Choosing Operating System
- File System and Operating System tuning

Hardware

- Platform
- CPU
- Memory
- Storage
- Network

How to Chose Hardware

- **Evaluation**
 - Available to larger customer
 - Requires engineering effort and systematic testing
 - Way to go for large hardware purchases
- **Expert Advice Based**
 - Based on best practices and common sense
 - Will not offer perfect fit but get close
 - Something to do for smaller purchases

General Considerations

- Avoid Hardware “zoo”
 - Standardize on limited number of configurations
- Beware of Hardware Bugs
 - Every server in your large order may be affected
 - Do stress testing before putting in production
 - Gradual hardware upgrade often good idea
- Your Team Experience Matters
 - Hardware is typically commodity
 - Details – Support, Drivers, BIOS settings are different
 - Consider both experience and bias

Many Special Considerations

- Remote Management
- Support
- Space and Power usage
- Reliability
- Extensibility
- Equipment life time
- Requirements differ depending on your application

Platform Choice

- X86-64 based all around
 - Intel leading, though some AMD use
- Number and type of CPU support
- Type and Amount of Supported memory
 - 18 slots (3*6 min)
 - Large memory modules are very expensive
 - Check how many of them can be used at the time
 - CISCO Servers – 48 Slots
 - 384GB using 8GB DIMM slots.
- 2 Socket systems typically good enough.

CPU Choice

- Workload Dependent
- CPU single core performance is very important
 - Response time for your queries
 - Replication is single thread
- Number of cores depends on workload
 - MySQL Scalability Also varies
- Nehalem based CPUs
 - Up to 6 cores, 12 threads (Westmere)
- Cache Size
- Idle cores are typical in CPU bound workloads

Memory

- A lot of Performance problem is solved if data fits in memory.
- Many applications look to fit all data/working set in memory
- More Memory is More than saving IO
 - A lot of CPU Is saved when data is in memory too
- How much “cheap” memory the box supports ?
- Tradeoff Memory vs Fast Storage

Storage Choices

- Directly Attach Storage
- Flash
- SAN
- NAS

Hard Disk Basics

- Slow and Large
 - 3.5” 7200RPM SATA - 2TB+
 - Optimized for storage volume
- Fast and Small
 - 2.5” 10-15K RPM 300-600GB
 - Optimized for IOPS
- OLTP databases mainly need IOPS
- Large Slow hard drives for
 - Data Analytics; Archiving

RAID Basics

- RAID0
 - Stripping – typically bad idea
- RAID1
 - Mirroring. OS; Log Files
- RAID10
 - Good General choice
- RAID5
 - Read mostly workload. Beware recovery time'
- RAID6, RAID50, RAID-Z
 - Being creative

Directly Attached Storage

- Classical Choice for MySQL
- Simple and Inexpensive
- Good Latency for Hard Drives
 - Drives as close to CPU as it gets
- Inexpensive RAID BBU
 - Battery backed up Cache
 - Very important for Transactional Performance
- Can have 24drives (more with shelves)
 - Typically more than enough for MySQL

Flash

- SSD
 - SATA/SAS Interface. 64-160G in Size
 - 35K read Iops; 3.5K write IOPs less than 100 μ s latency
- PCI Express Card (Fusion IO)
 - 320GB per card; 100K IOPS less than 50 μ s latency
- Often provides more IO than MySQL Can handle
- Space typically the issue
- Integrational Challenges
 - Durability; Lifetime; OS&Database optimization

SAN

- Enterprise Approach from Big Databases
- Fiber Channel and iSCSI
- Data management issues
- High Availability Scenarios
- Expensive per size and IOS
- Can offer many IOs
 - Though rarely as much as Flash
- Higher Latency than Directly Attached Storage
- Rarely a Performance Solution

NAS

- Filesystem level mount
- NFS is typical
- Rare choice for Primary Database
- Can be used for Backup storage etc.

Network

- 1Gb is the common baseline
 - Beware of 100Mb cross connect in hosted environments
- Trunking/Bonding
 - Improved Throughput and High Availability
- 10Gb
 - Even better latency
 - Can be pushing throughput in some cases
- Beware Operations bottleneck
 - What if you need to restore 10 slaves from backup at the same time ?

Operating System

Operating System, File System, IO Scheduler

Operating System Choice

- Linux
 - The most popular – de facto standard.
- Solaris
 - Picked up pace with Sun. ZFS with SSDs
- Windows
 - Typical platform for web development
 - Is NOT typical platform for large scale deployment
- FreeBSD etc
 - Rarely used. Less experience, higher problem potential

Linux Distribution

- **CentOS/RHEL**
 - The Most popular
 - May want to upgrade MySQL to recent version
- **Ubuntu/Debian**
 - A lot of successful experience too
- **Fedora**
 - Be careful in production. Too frequent updates

32bit or 64bit

- 32bit is Dead !
 - At least on server side
- 64bit capable CPU
 - Plus 64bit OS
 - Plus 64bit Binaries
 - Must have all 3 to enjoy 64bit advantages

LVM

- Great choice on Linux
- Get some SAN features on Direct Attach Storage
- Online Backups for All Storage Engines
- Online Storage Extension
- Beware of snapshot overhead for write performance

Filesystems

- EXT3 is common default
 - Has inode level locking; slow deletes
- EXT4
 - Rather new. Does not scale too well
- XFS
 - Good Choice on Linux
- ZFS
 - Good for use with SSDs/Flash. Allows using Flash as a cache.

IO Scheduler

- Performance is rather workload specific
 - **cat /sys/block/hda/queue/scheduler**
- Names are not everything
 - There are significant differences in schedulers over time
- “beware of cfq”
 - Can cause serialization in certain workloads
- Noop
 - Good choice with SSDs, Smart RAID
- Deadline
 - Can be decent choice

Mount Options

- Defaults are usually pretty good
- Changes for specific workloads
- “Noatime”
 - Helps when there are a lot of tables, especially MyISAM

Other Tunings

- Avoid running out of memory
 - Having swap file is good. Swapping is bad
- **echo 0 > /proc/sys/vm/swappiness**
- innodb_flush_method=O_DIRECT
 - Avoid double buffering and IO pressure
- SELinux has performance impact
 - Consider disabling it

Thanks for Coming

- Questions ? Followup ?
 - pz@percona.com
- Yes, we do **MySQL and Web Scaling Consulting**
 - <http://www.percona.com>
- Check out our book
 - Complete rewrite of 1st edition

