PERCONA
Performance Consulting Experts

MySQL and SSD: Usage Patterns

Date, time, place:	MySQL Conference & Expo 2011 12-Apr-2011
Reporter:	Vadim Tkachenko Co-founder, CTO, Percona Inc

- You can get up to 7x gain running MySQL on SSD
  - Even 20x with some tricks

In this talk

- What is best setup of MySQL to get most benefit from SSD / Flash

What's inside

- MySQL and SSD: basics
- MySQL and SSD: advanced schemas

Types of SSD

- SATA
- PCI-E
- SAN

Types of SSD

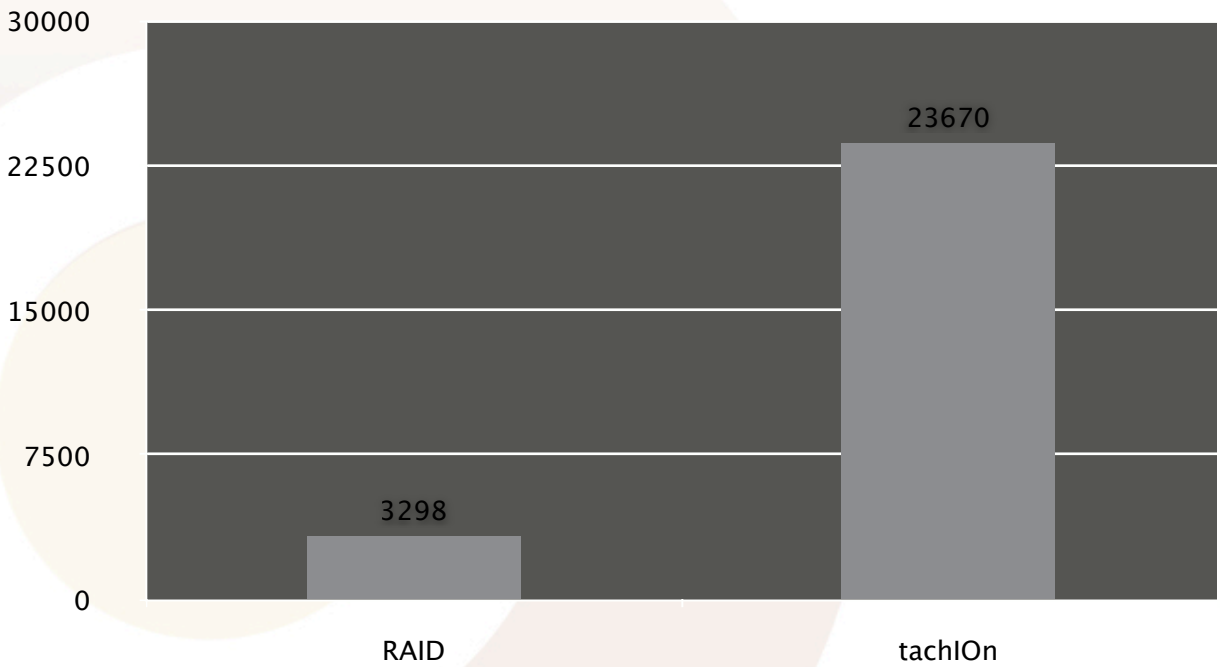
- SATA
  - 200-500MB/sec
  - Intel X25-M/E, OCZ, Unigen
- PCI-E
  - Over 1GB/sec, 70.000 req/sec, under 1ms response time
  - Virident, FusionIO
- SAN
  - Violin memory

What is this for MySQL ?

- 1GB/sec – 70,000 req/sec
  - A lot, but MySQL can't use that all

MySQL basic usage

- Put all data (ibdata1, ib_log, tables.ibd) on SSD
  - 5-7x difference

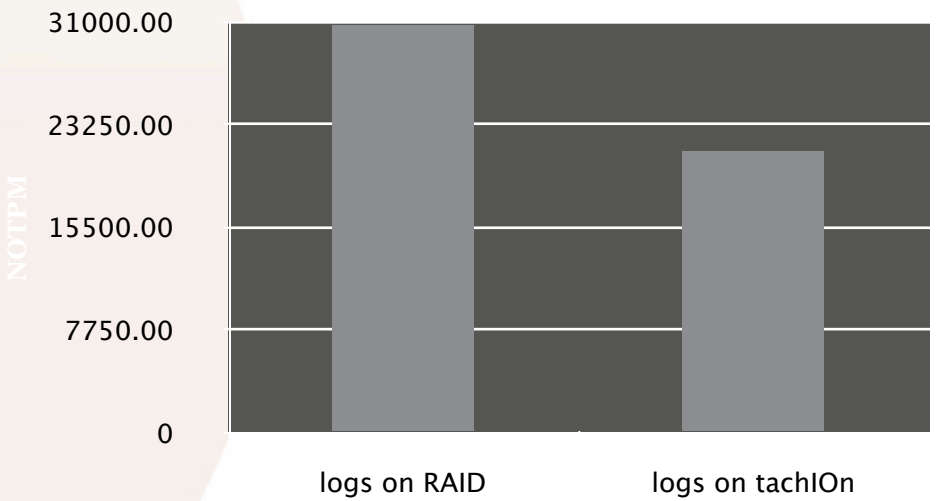


MySQL basic usage

- Boring, some recommendation:
  - XFS, better with 4k blocks
    - Mkfs.xfs -s size=4096
    - Mount –o nobarrier
  - Multiple threads
  - Percona Server or InnoDB-plugin or MySQL 5.5
- Still uses about 5,000 req/sec, ~200MB/sec

Can we do better ?

- Single threaded sequential stuff
  - InnoDB transactional logs with fsyncs
  - Binary logs
  - Doublewrite buffer (with whole ibdata)
- RAID with BBU good place for them
  - Up to 45% improvement



PERCONA Performance Consulting Experts

What is in Percona Server for it ?

- innodb_flush_neighbor_pages= ON | OFF
- innodb_log_block_size = 512 | 4096
- innodb_page_size = 4K | 8K | 16K
  - Use carefully
- innodb_doublewrite_file
- Innodb_adaptive_checkpoint=keep_average
- innodb_log_file_size > 4GB

What if ?

- Data is too large
  - Or SSD too small

Naive solution

- Most data on regular disks
- Put “hot” ibd files on SSD
  - Symlinks from data directory

Naive solution

- It works
- But pain to manage and till first “ALTER TABLE”
- Facebook has patch for it

Advanced solution: caching

- Data stored on regular disks
  - Caching data on SSD
- Flashcache
  - Open source
  - Developed and maintained by Facebook, deployed in production
- DirectCache
  - Proprietary solution from FusionIO

FlashCache

- Shows good results
- Stable work in production deployments
- Not much user friendly

FlashCache details

- Write-through and write-back modes
- FIFO and LRU block management
- You need to compile kernel module by yourself
- The same choice for ibdata1/ib_logs location

ZFS

- ZFS supports SSD caching mode
- Linux native port available

Interesting trick

- For temporary tables on disks
- --tmpdir=/mnt/ssd

Still not enough space?

- Several SSD / Flash cards in the server

Combining cards

- SATA
  - Choice good RAID controller, maybe challenge
- PCI-E
  - Software RAID
    - Usually comes as stripping (RAID0)
    - Reliability ?
    - Mirroring does not work really well

What about promised 20x ?

- Single MySQL instance is not able to utilize all IO provided by card

Several instances

- Solution obvious:
  - Several instances

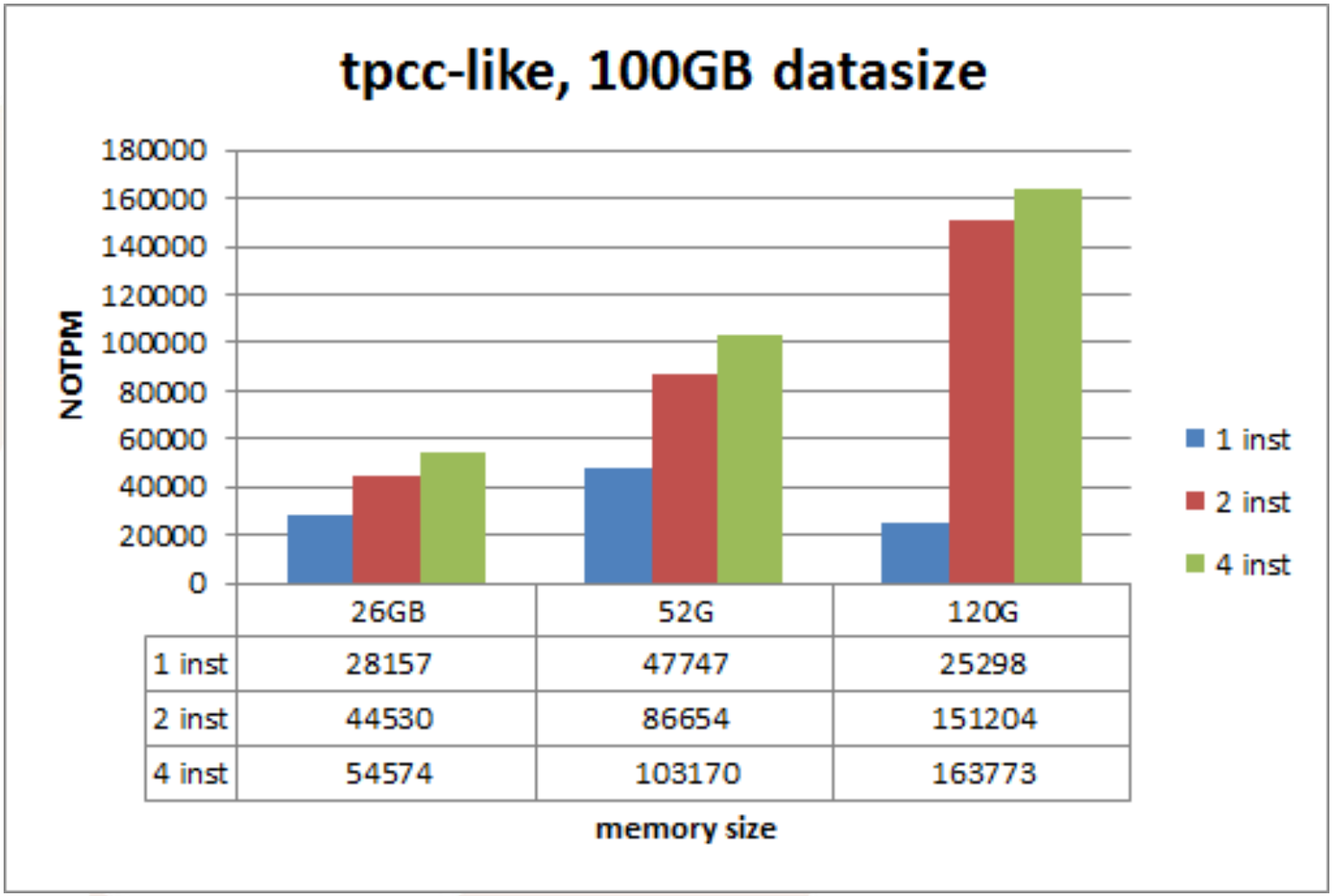
Experiment

- Dell PowerEdge R815
- 4 physical AMD CPUs / 48 cores
- 144GB of RAM
- Virident tachIOn 200GB card
- 48 user connections

Experiment

- Dell PowerEdge R815
- 4 physical AMD CPUs / 48 cores
- 144GB of RAM
- Virident tachIOn 200GB card
- Tpcc-mysql workload
  - 48 user connections

Results



	26GB	52G	120G
1 inst	28157	47747	25298
2 inst	44530	86654	151204
4 inst	54574	103170	163773

Results conclusions

- With 120GB memory single instance result worse then with 26GB
  - InnoDB contentions problems again
- Two instances allows to improve 1.5x-6x

Multi-instance

- I do not like it
  - Management complexity
  - Good scripts solve it
- 2 instances seems reasonable
- SAN-like Flash-arrays

The end

- Slides will be online. http://www.percona.com/about-us/presentations/
- vadim@percona.com
- Your questions ?
- We are hiring!