

# 利用新硬件提升数据库性能

淘宝核心系统数据库组 余锋

<http://yufeng.info>

新浪微博：@淘宝褚霸

2012-07-07

- 数据库软硬件发展趋势
- CPU
- 内存
- 磁盘
- 网络

数据库**百万TPS**不再稀罕！



MemSQL 1.2 million  
inserts/second on a 64-  
core, 1/2 TB of RAM  
machine



MySQL Cluster 7.2 achieves  
4.3BN reads per minute with  
30 data nodes

2-socket servers using X5670  
with Infiniband interconnect  
and 48GB of memory per  
machine

- 8 Xeon 7540 CPU, 96逻辑CPU
- 512 GB DDR3
- 600G SSD \*12
- 万兆网卡

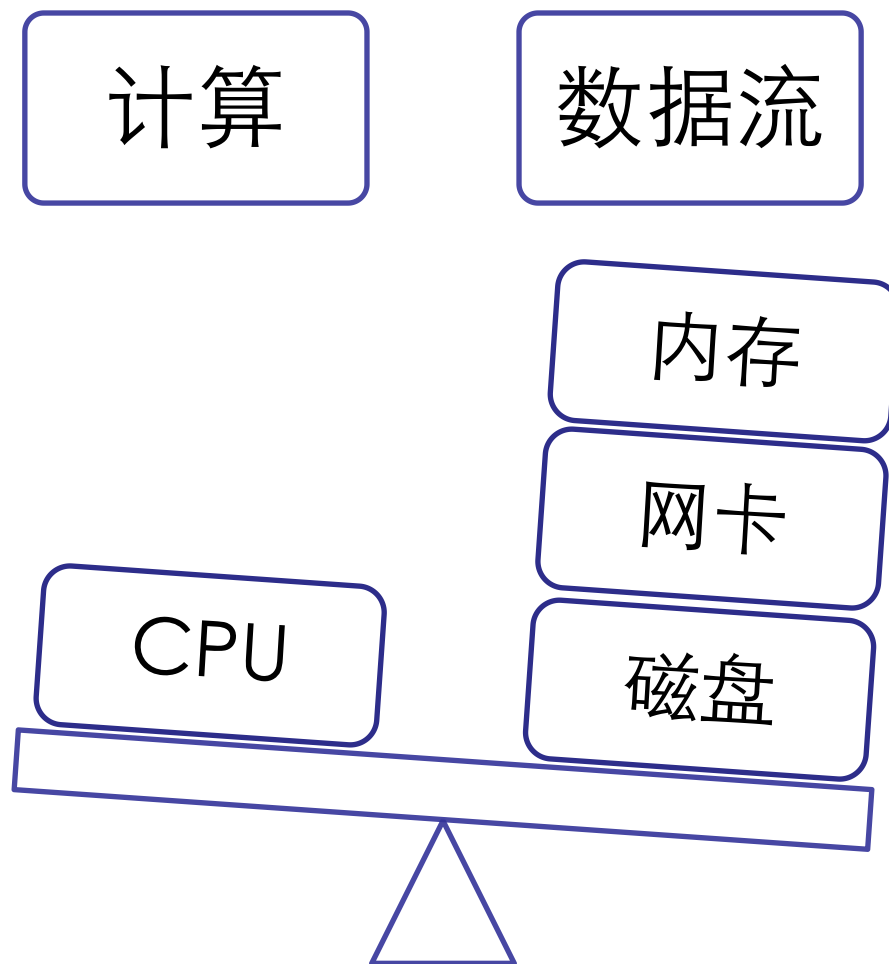


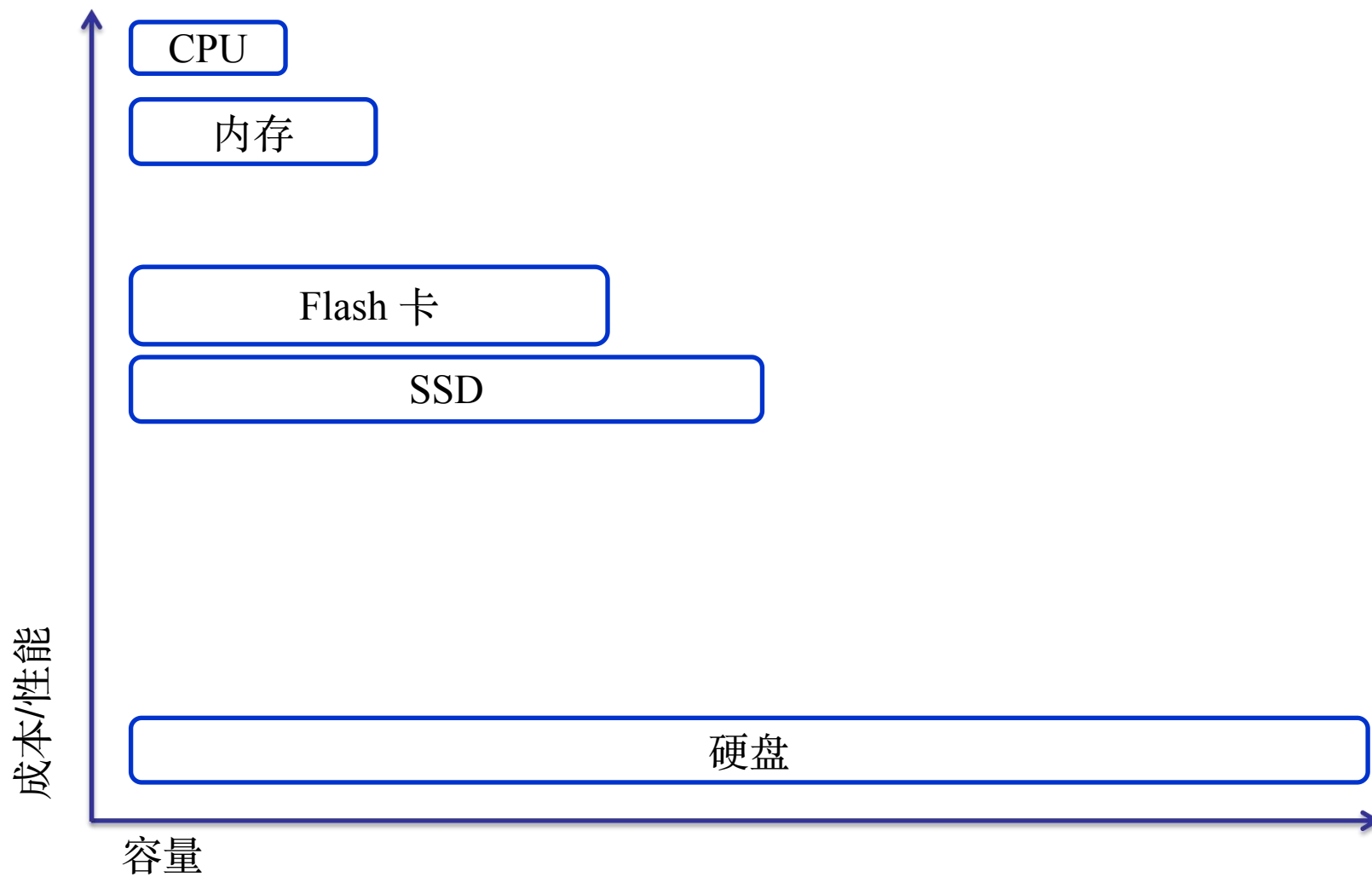
文艺青年的装备

- 2 E5-2420 CPU, 24逻辑CPU
- 96GB DDR3
- 600G SSD \*8
- 千兆网卡



普通青年的装备





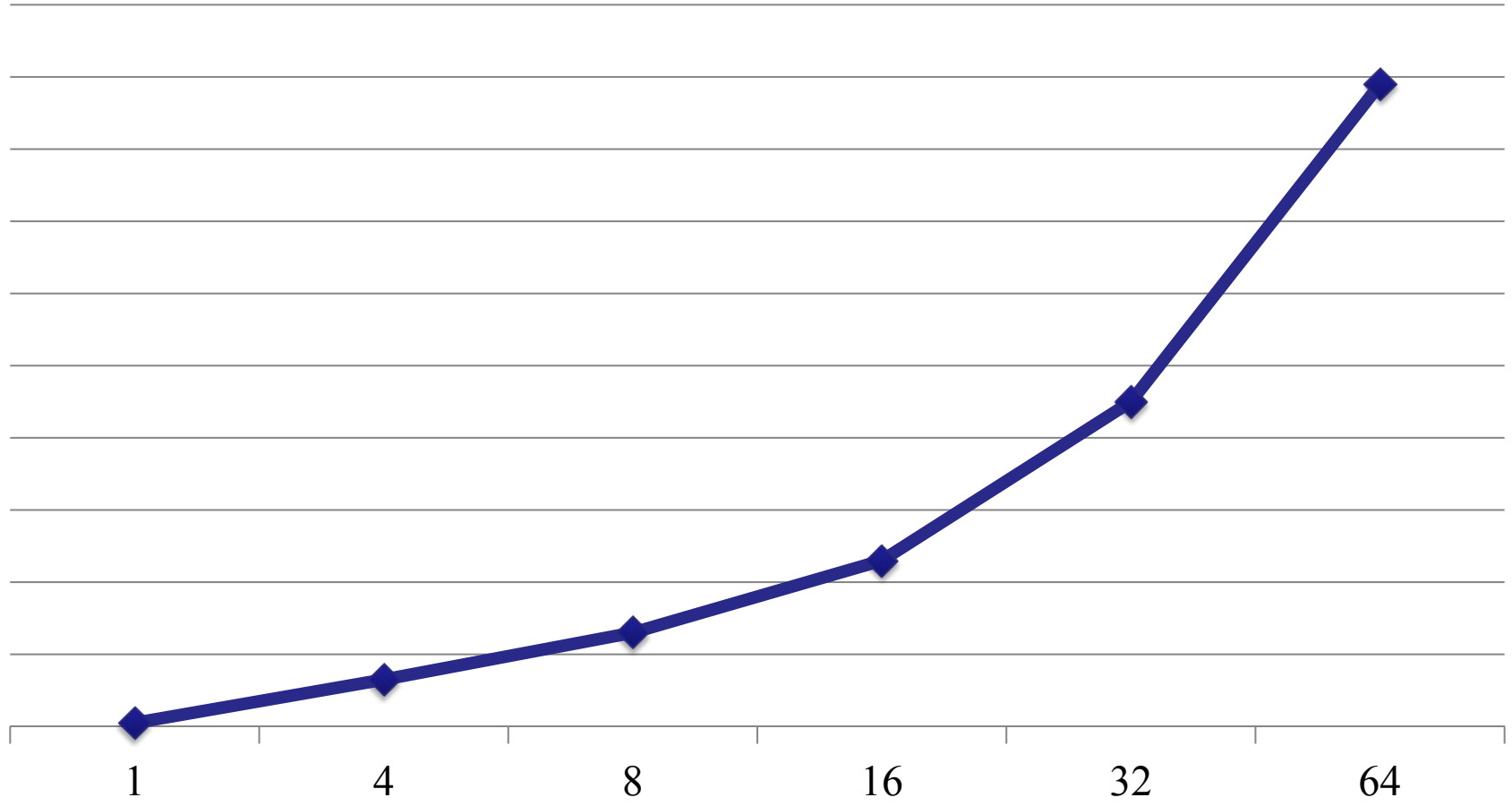


L1 cache reference	0.5 ns
Branch mispredict	5 ns
L2 cache reference	7 ns
Mutex lock/unlock	25 ns
Main memory reference	100 ns
Compress 1K bytes with Zippy	3,000 ns
Send 2K bytes over 1 Gbps network	20,000 ns
Read 1 MB sequentially from memory	250,000 ns
Round trip within same datacenter	500,000 ns
Disk seek	10,000,000 ns
Read 1 MB sequentially from disk	20,000,000 ns
Send packet CA->Netherlands->CA	150,000,000 ns

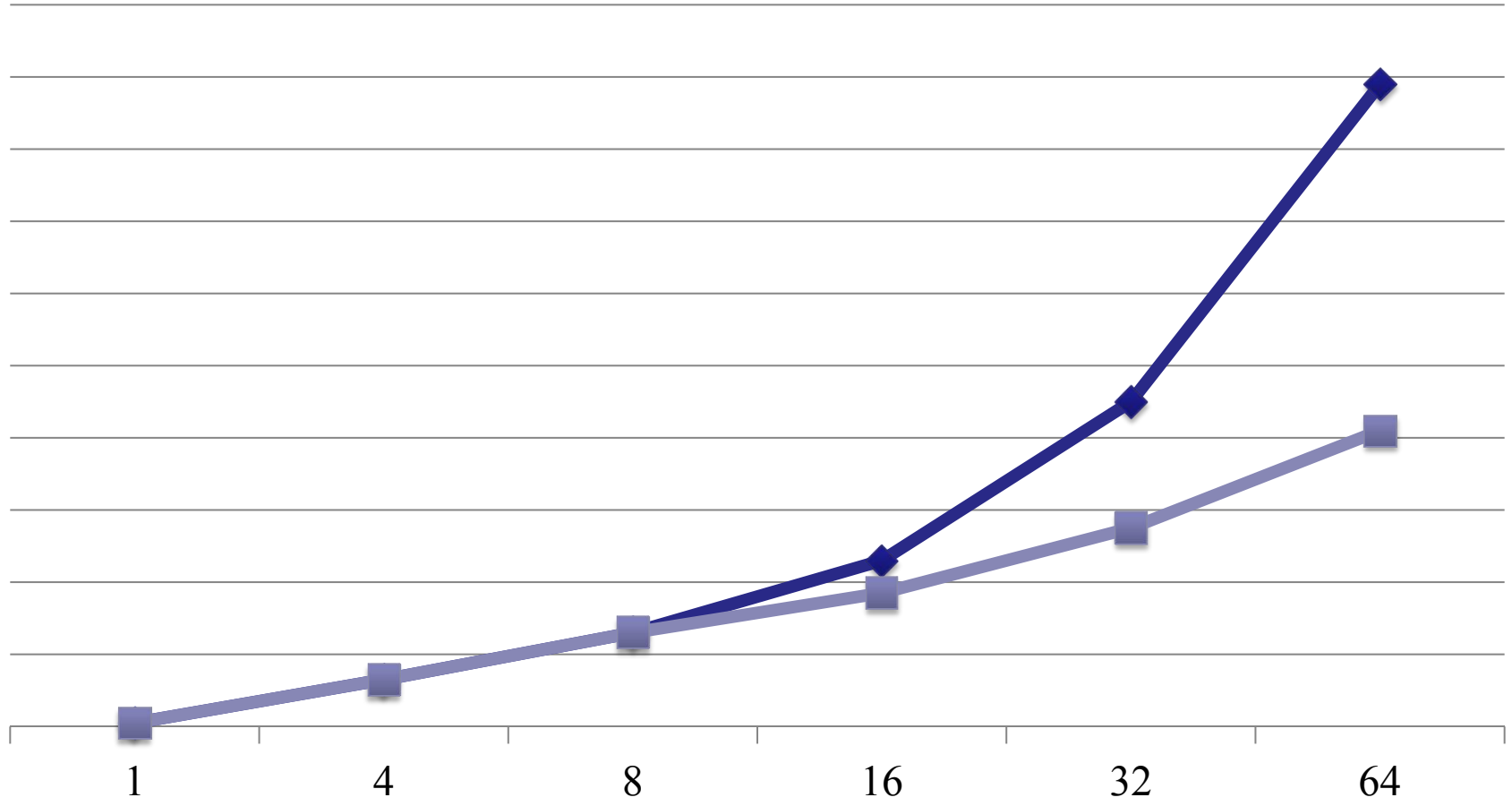


- 数据库软硬件发展趋势
- CPU
- 内存
- 磁盘
- 网络

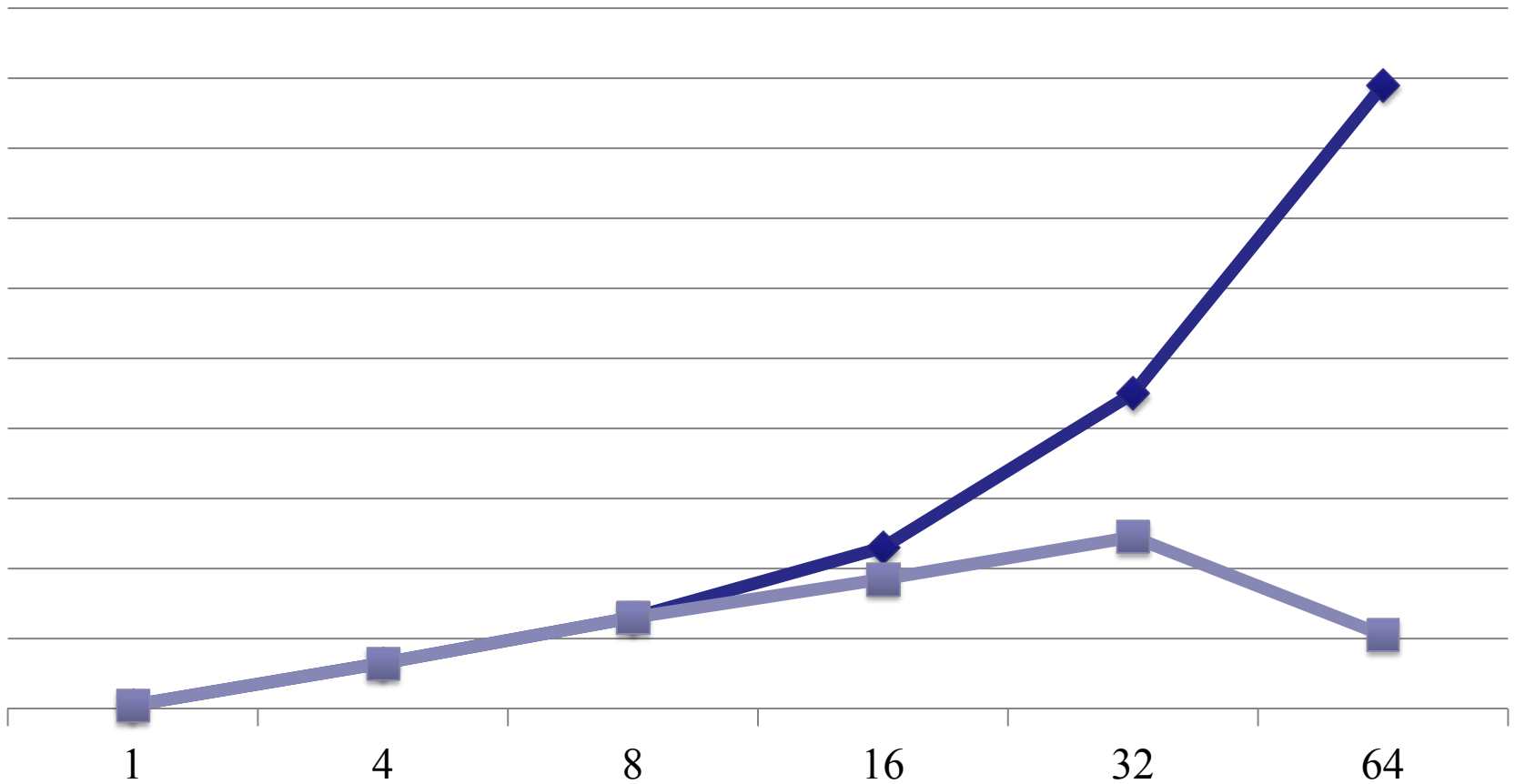
# CPU Scalability



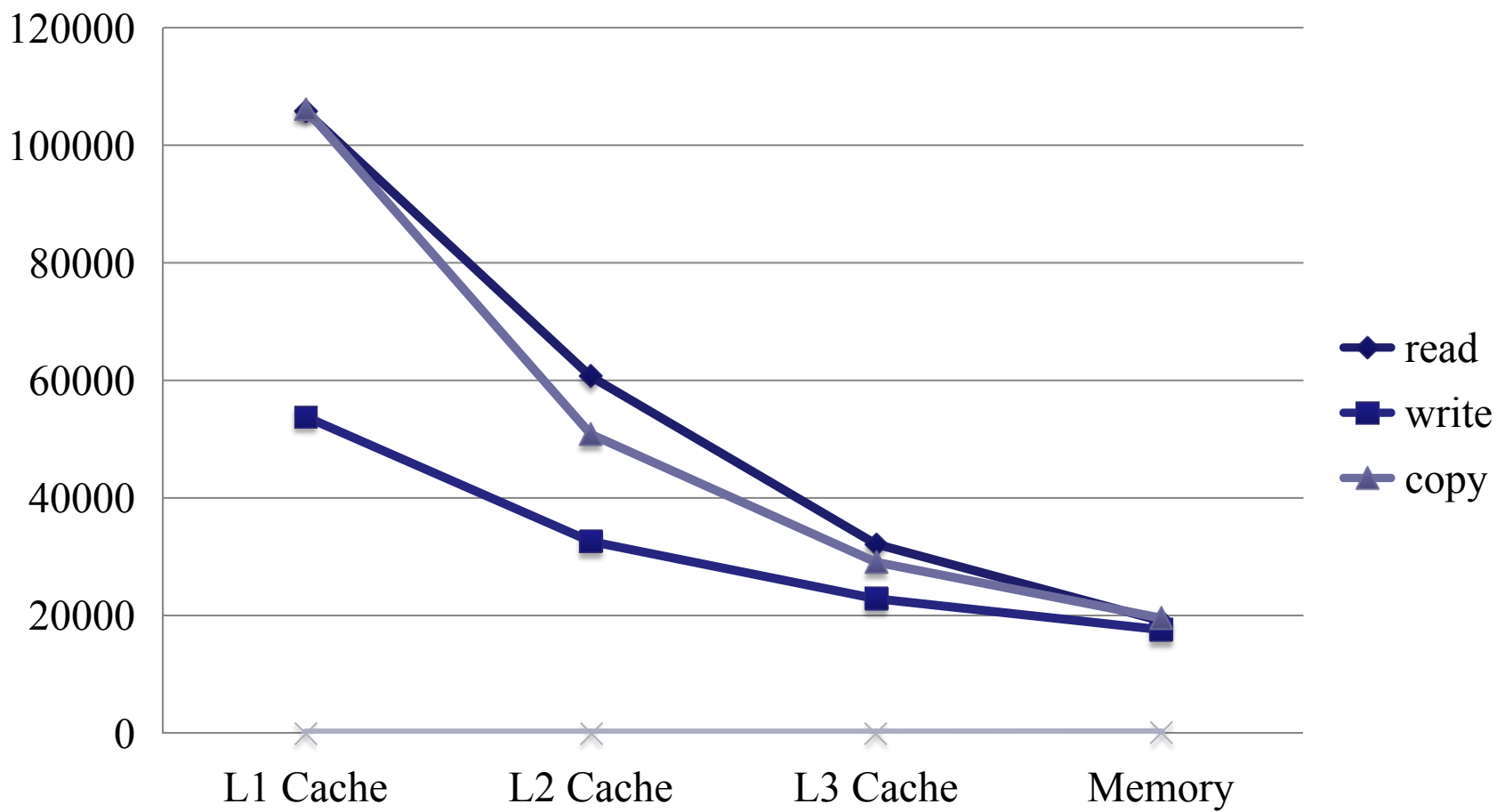
# CPU Scalability



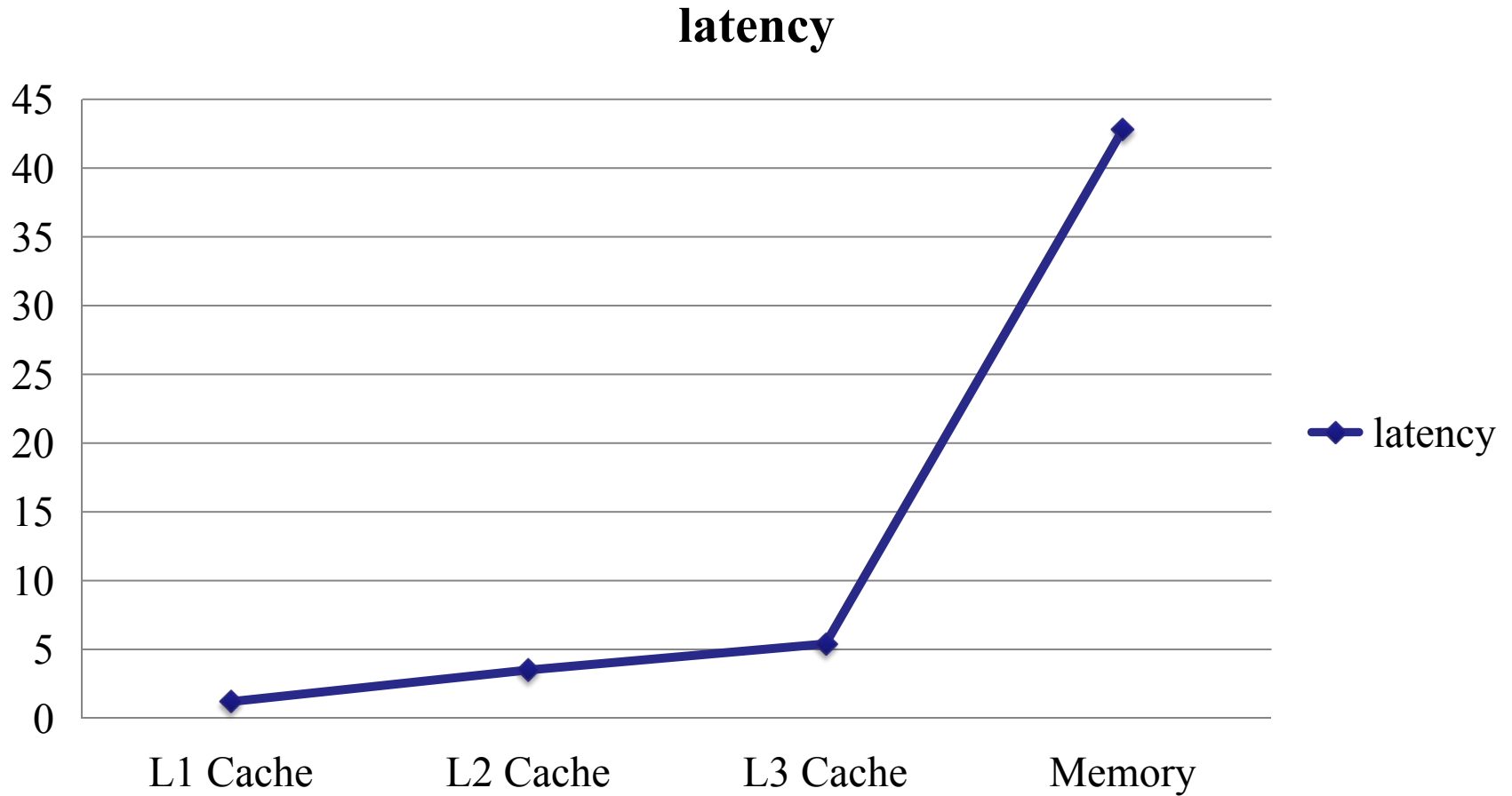
# CPU Scalability



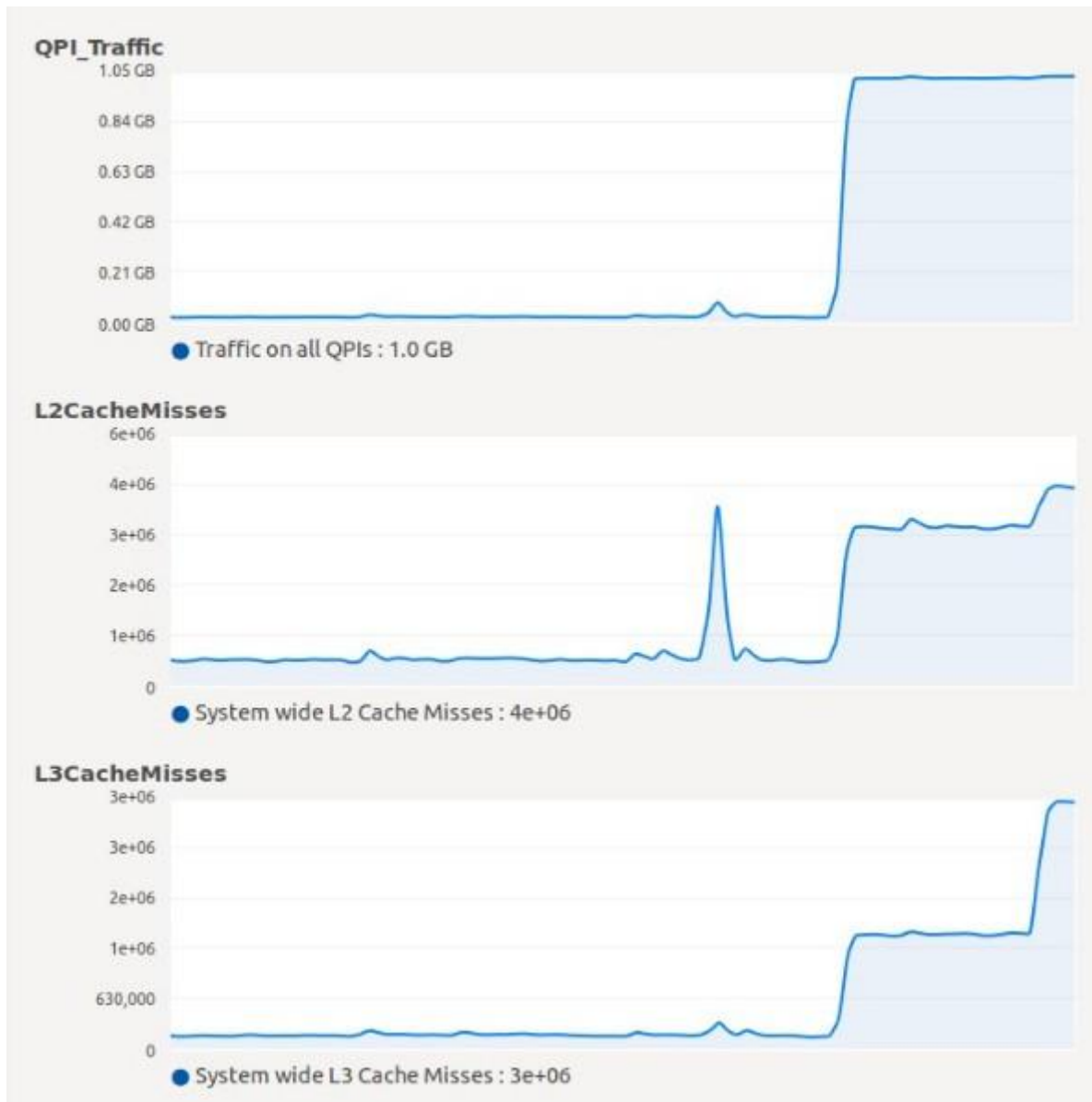
# Cache和主存吞吐量



# Cache和主存延时

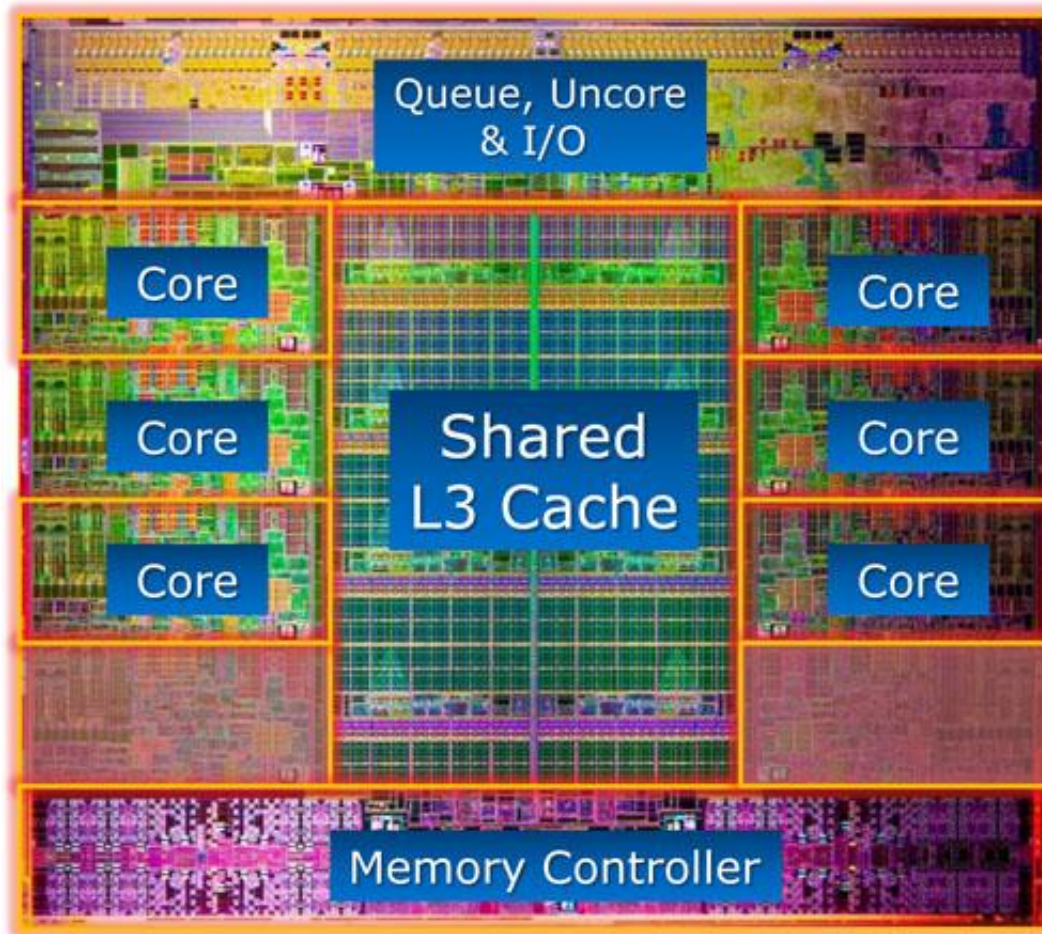


# 多核性能恶化原因





# Sandy Bridge-E



- 6个CORE，计算力强
- 4个DDR内存通道，2个QPI互联，内存带宽足
- 内置PCI-E 通道，IO能力强
- 更大的L3，Cache更高效

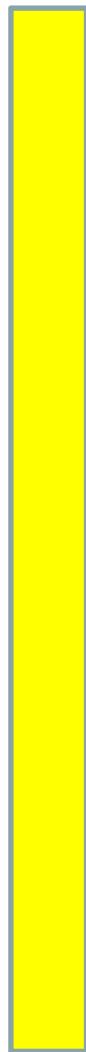
- 数据库软硬件发展趋势
- CPU
- 内存
- 磁盘
- 网络

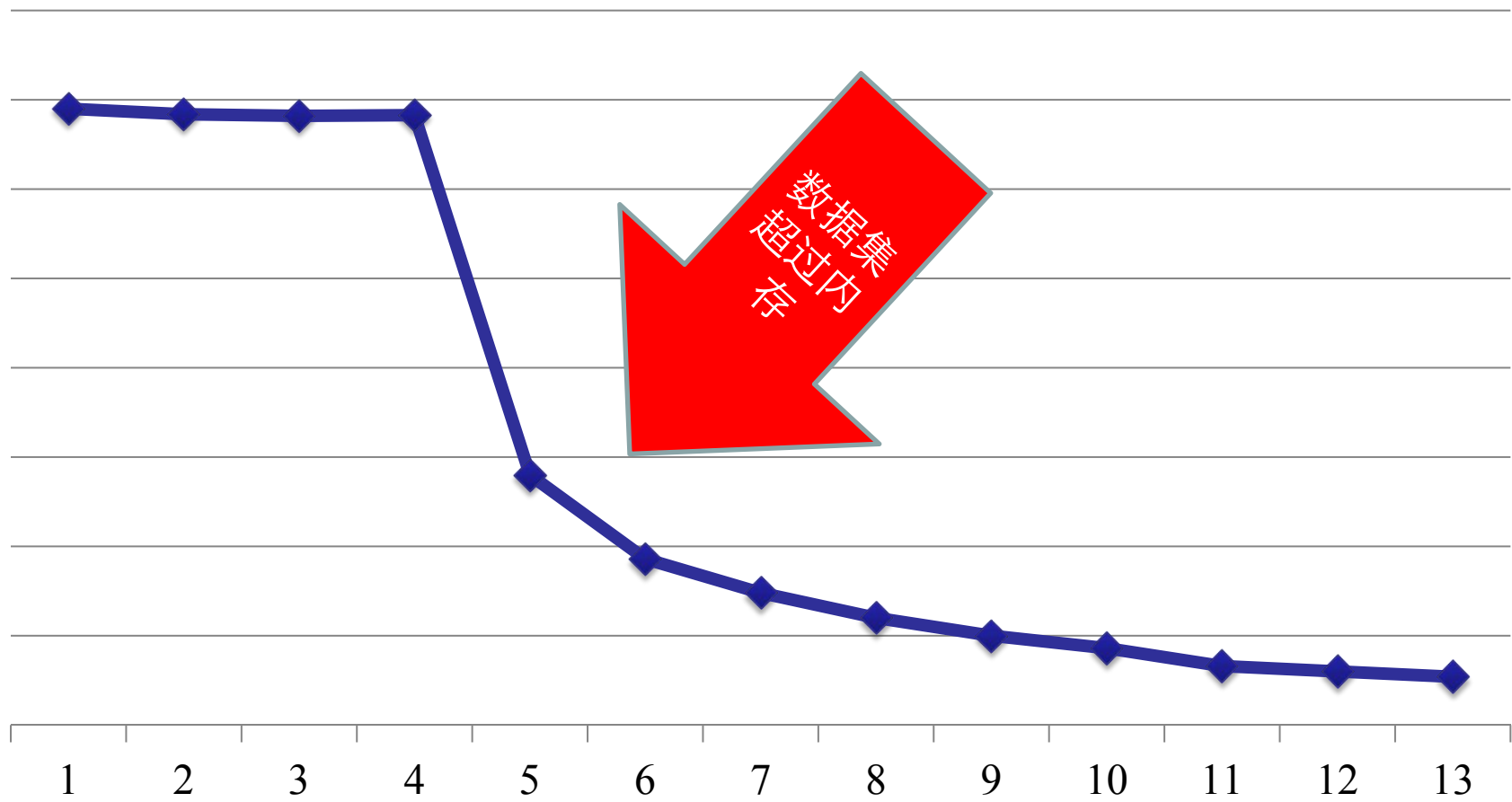
# 内存和外存的差距

毫秒



纳秒





- 来者不拒，越多越好
- 成本考虑，装下最热数据集
- 百G以上不奇怪

- 数据库软硬件发展趋势
- CPU
- 内存
- 磁盘
- 网络

- 日志文件顺序IO，落地为要
- 引擎尽最大努力把脏数据转变成顺序IO
  - 引擎不同，数据结构不同，差距很大
- 历史原因，传统数据库基于IO设计，最大内存也避免不了IO





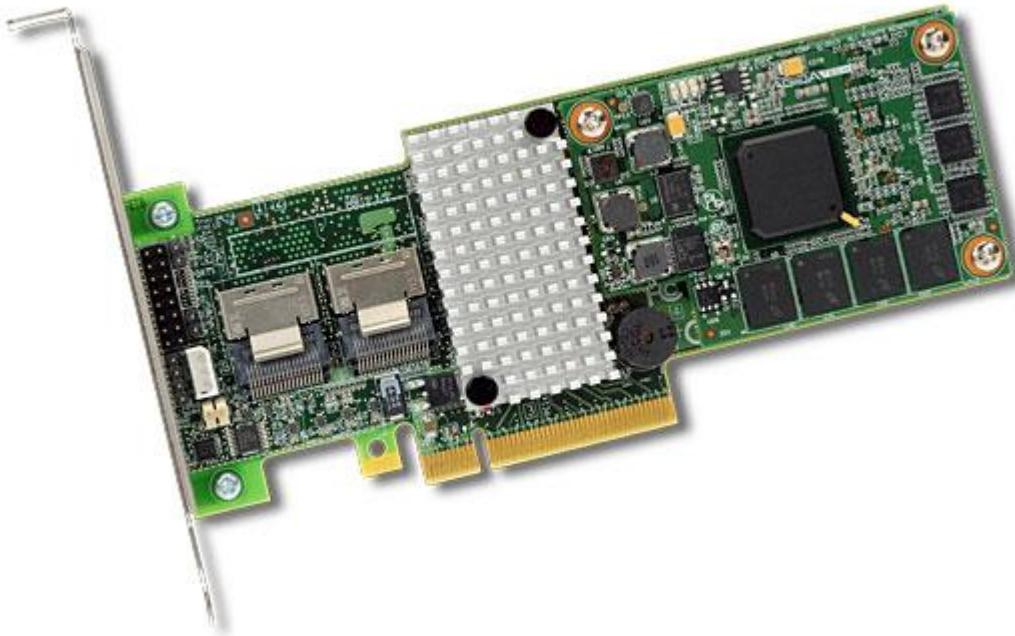
非易失内存

Flash

硬盘

# Raid卡

- PCIe 2.0x8
- Support Up to 128 SATA Devices
- Dual Core ROC
- 1GB cache

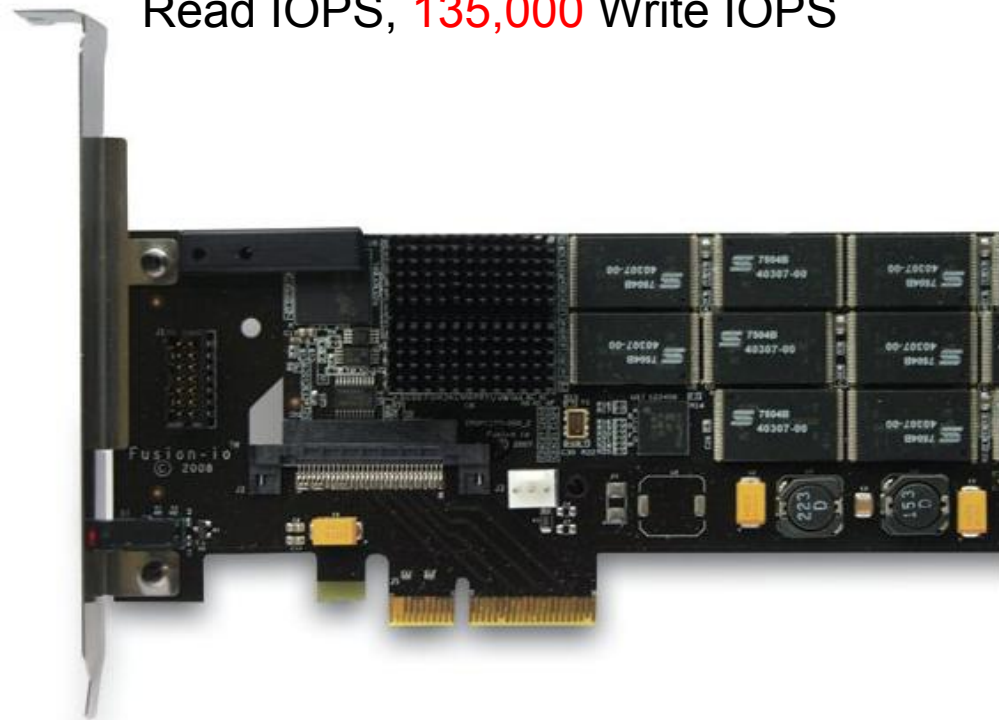




# Flash卡



PCIe 2.0x8  
850 MB/s (4KB)  
220,000 IOPS (4KB)



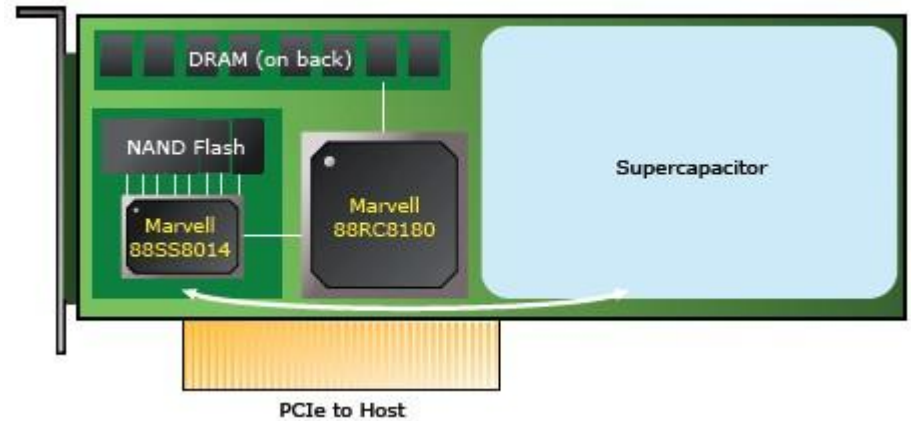
PCIe 2.0x4  
ioDrive IOPS: with Flash 140,000  
Read IOPS, 135,000 Write IOPS

# 非易失内存

PCIe 1.1x4

4K Block Writes: 165,000 IOPS

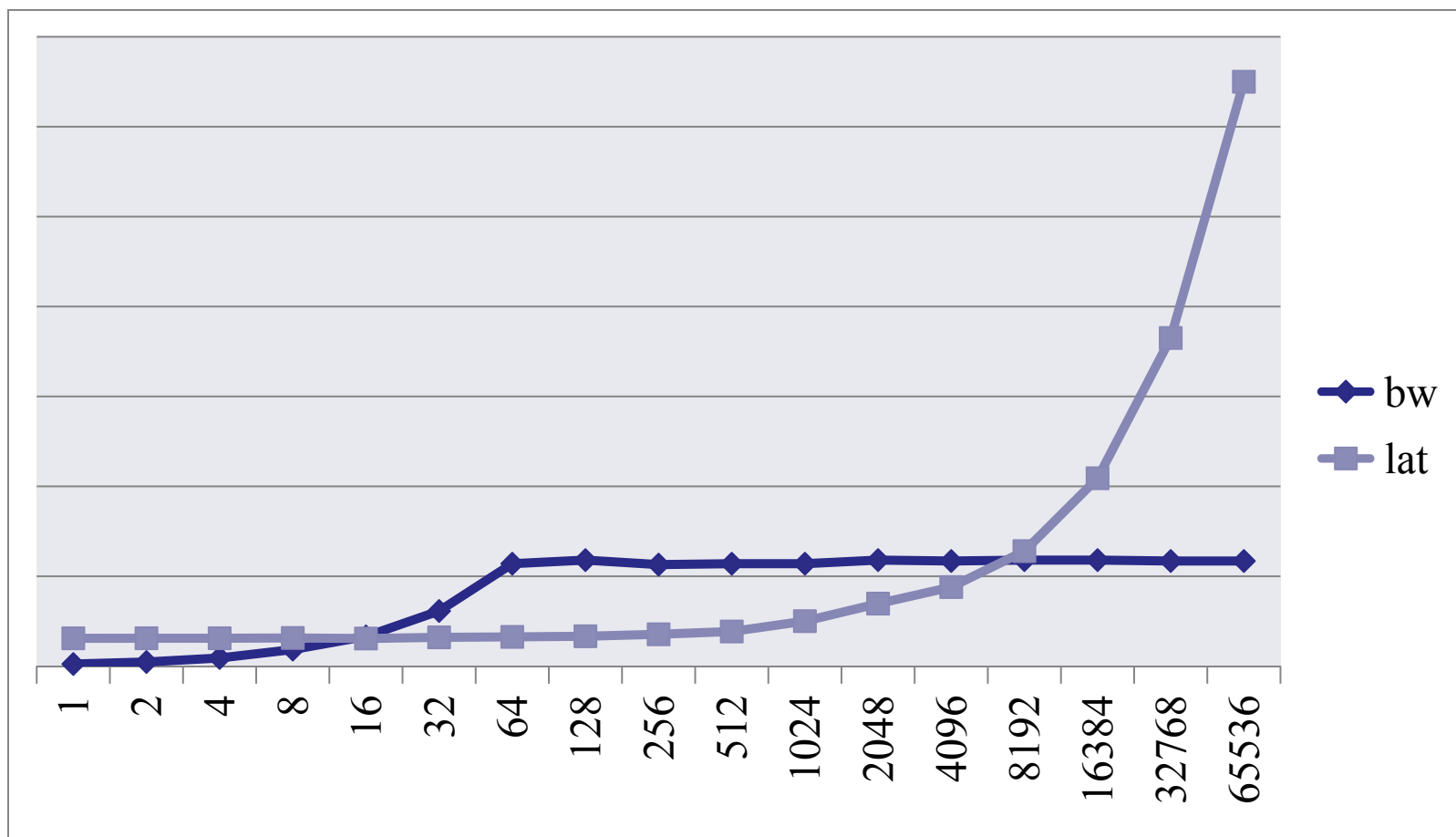
4K Block Reads: 185,000 IOPS



DDR3 Non-Volatile DIMM 8G

- 数据库软硬件发展趋势
- CPU
- 内存
- 磁盘
- 网络

# 千兆网卡性能表现



- 网卡Bonding
  - 更大吞吐量
- 万兆网卡
  - 百万以上PPS
  - CPU负担更小
  - 更小延时



# 谢谢大家！

部分图片粘贴自Google搜索，谢谢Google！  
部分数据参考自Percona,谢谢那些可爱的人！