

现代数据中心不使用PCIe SSD的几大理由 ?

阳学仕, PhD
宝存科技CEO



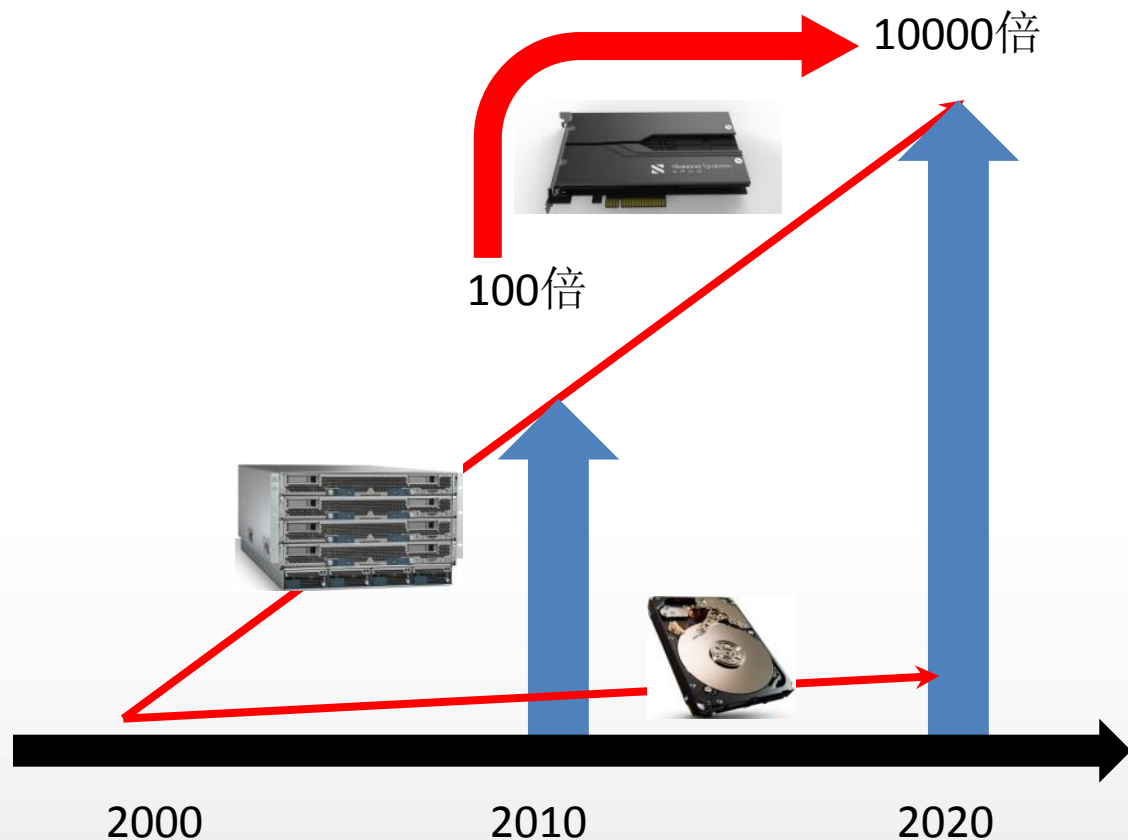


提纲

- 数据中心架构的变迁 – 闪存革命
- PCIe 固态闪存卡简介
- 使用PCIe固态闪存卡的顾虑
 - 寿命
 - 稳定性
 - 可靠性
 - 成本
- PCIe Flash存储的优势



基于传统磁硬盘（HDD）存储的IO和读写延迟限制



- 单台服务器性能十年100倍
- 机械硬盘性能十年2倍
- 固态存储技术基于全半导体存储，与CPU/DRAM同步
- SSD随机读写性能优异，延迟极低，稳定可靠



架构变迁 (1) -分

- 个人PC阶段

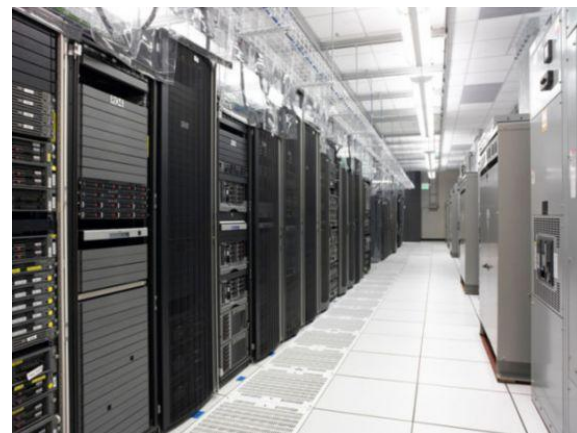
10' IO/每秒





架构变迁（2）-合

- CPU及互联网技术的发展
→ 存储IO要求
→ 异构数据中心



10万 IO/每秒





架构变迁（3）-分

- 闪存的出现大幅度提升单机IO
- 分布式架构主导
- 同构数据中心

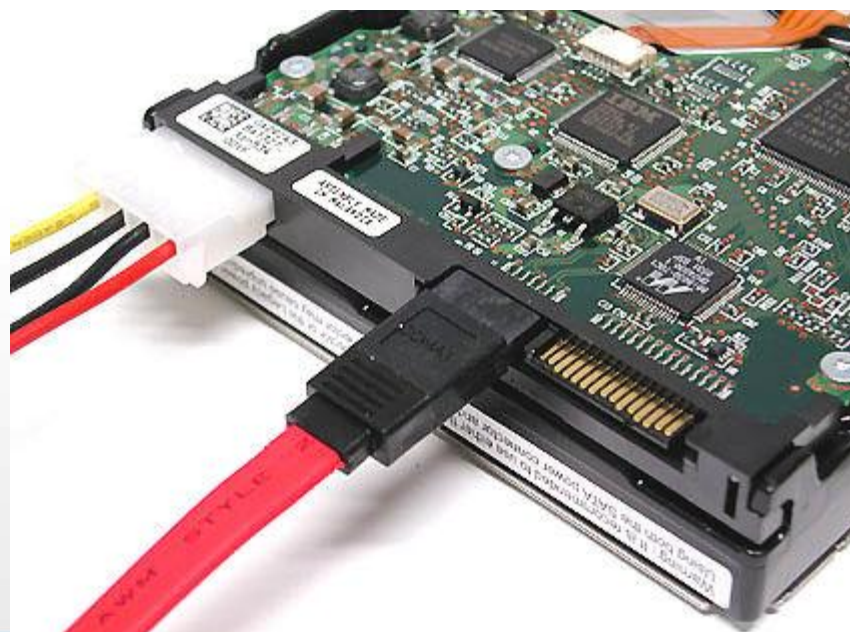
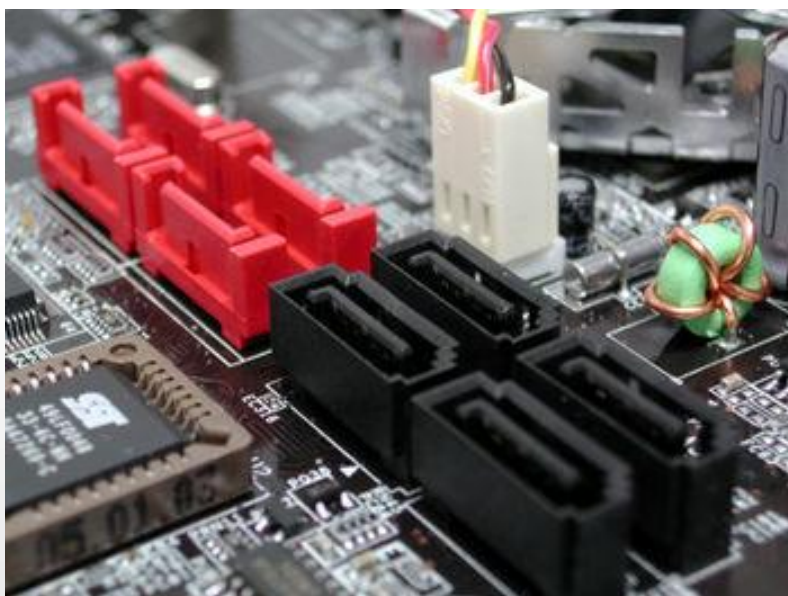
100万 IO/每秒





SATA和PCIe接口

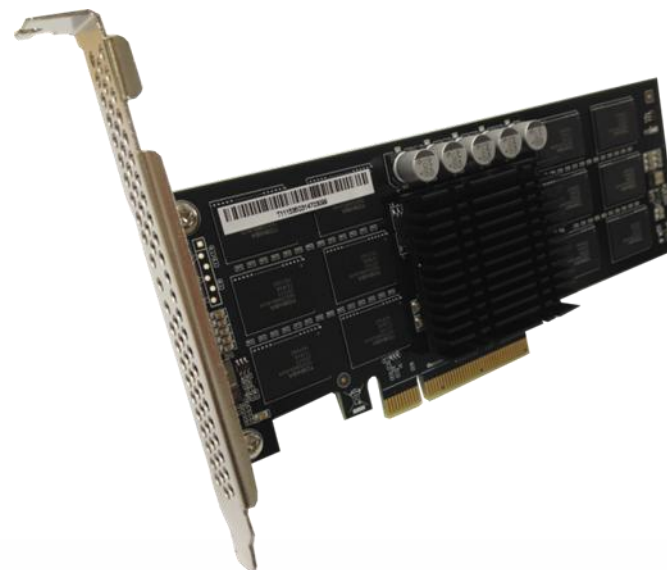
- SATA接口
 - SATA2.0: 3Gb/s
 - SATA 3.0: 6Gb/s





SATA和PCIe接口

- PCIe 接口



PCIe Architecture	Raw Bit Rate	Interconnect Bandwidth	Bandwidth Per Lane Per Direction	Total Bandwidth for x16 Link
PCIe 1.X	2.5GT/s	2Gbps	~250MB/s	~8GB/s
PCIe 2.X	5.0GT/s	4Gbps	~500MB/s	~16GB/s
PCIe 3.X	8.0GT/s	8Gbps	~1GB/s	~32GB/s



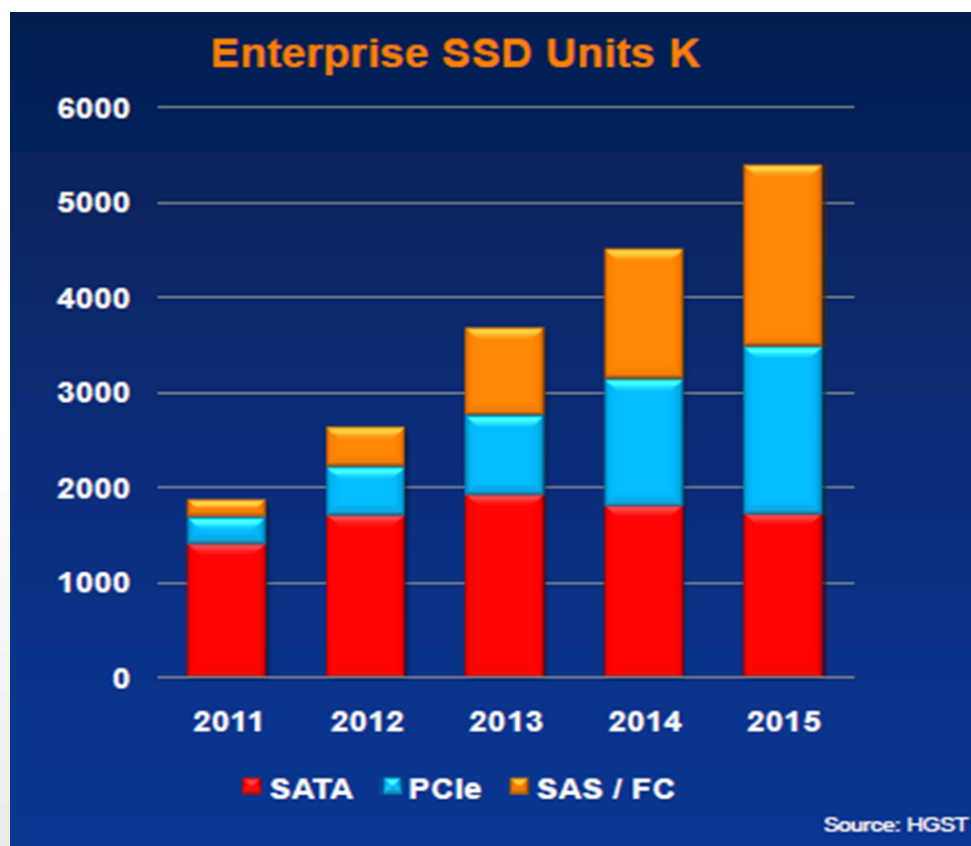
PCIe接口优势

- 最低时延（微秒级）
- 高带宽, 最高达 32GB/s （PCIe 3.0x16）
- 灵活可扩展性（x1, x2, x4, x8, x16）



PCIe 接口的Flash

- SATA 向PCIe接口转化, Enterprise

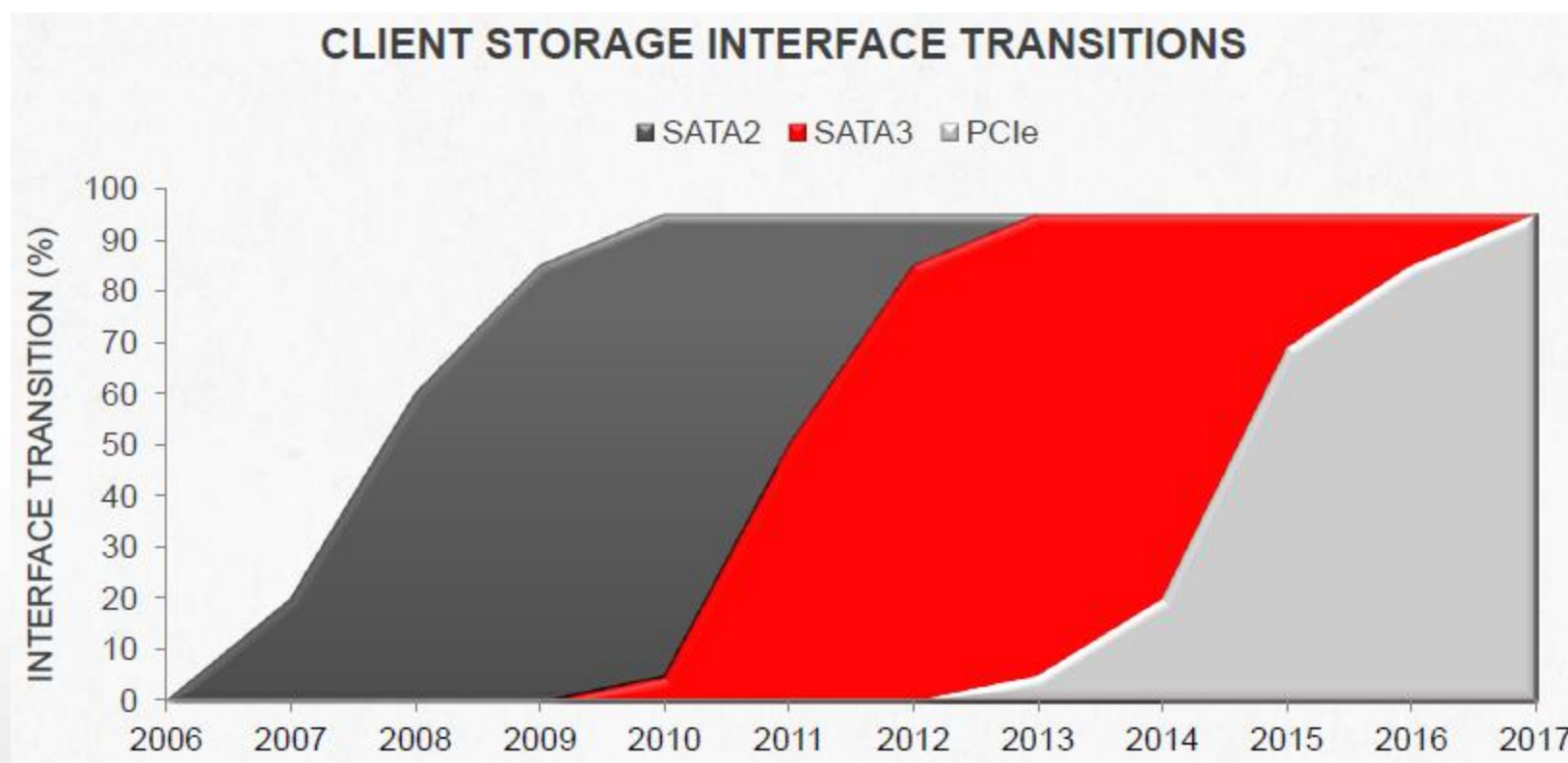


Source: HGST



PCIe 接口的Flash

- SATA 向PCIe接口转化, Client



Source: Marvell



不同存储设备比较 – HDD, SATA SSD, PCIe Flash





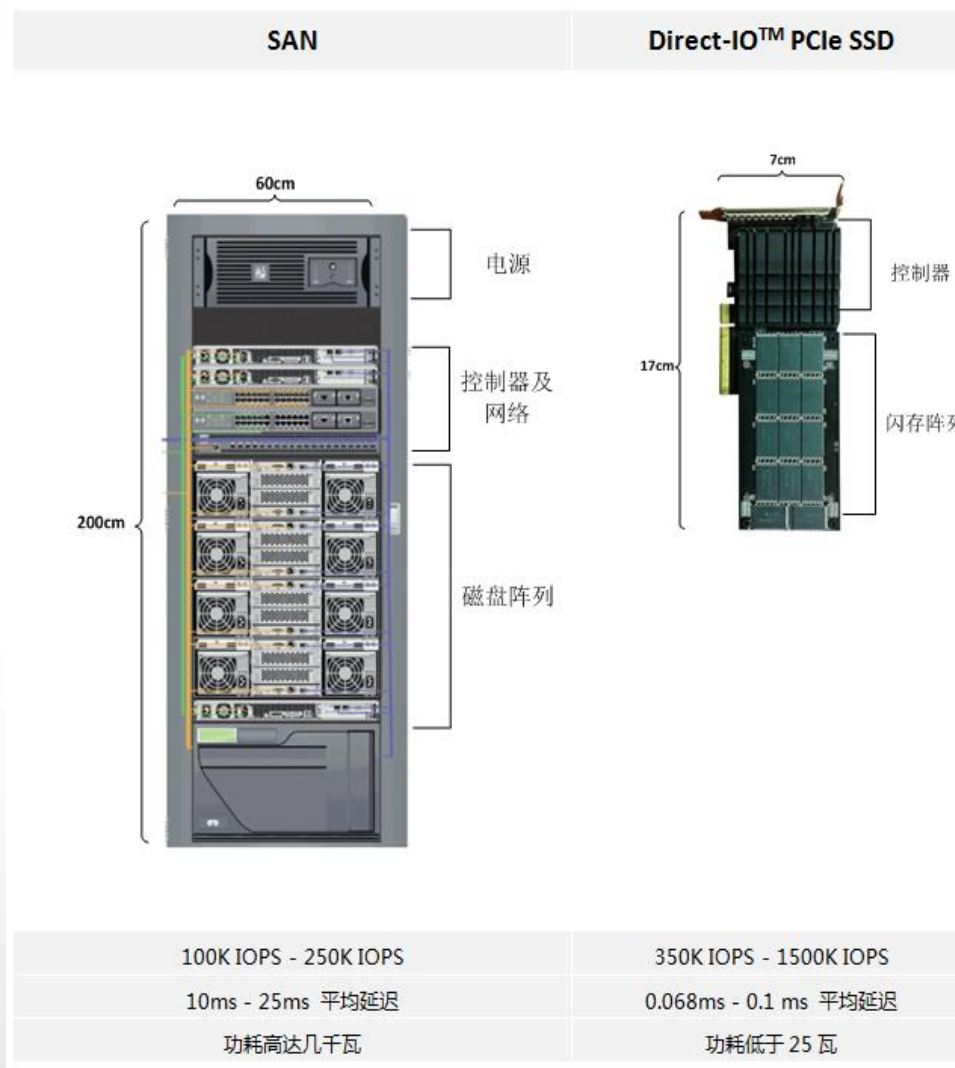
PCIe Flash更优的解决方案

PCIe Flash完全超越SAN的IO性能

读写延迟降低 **100x**

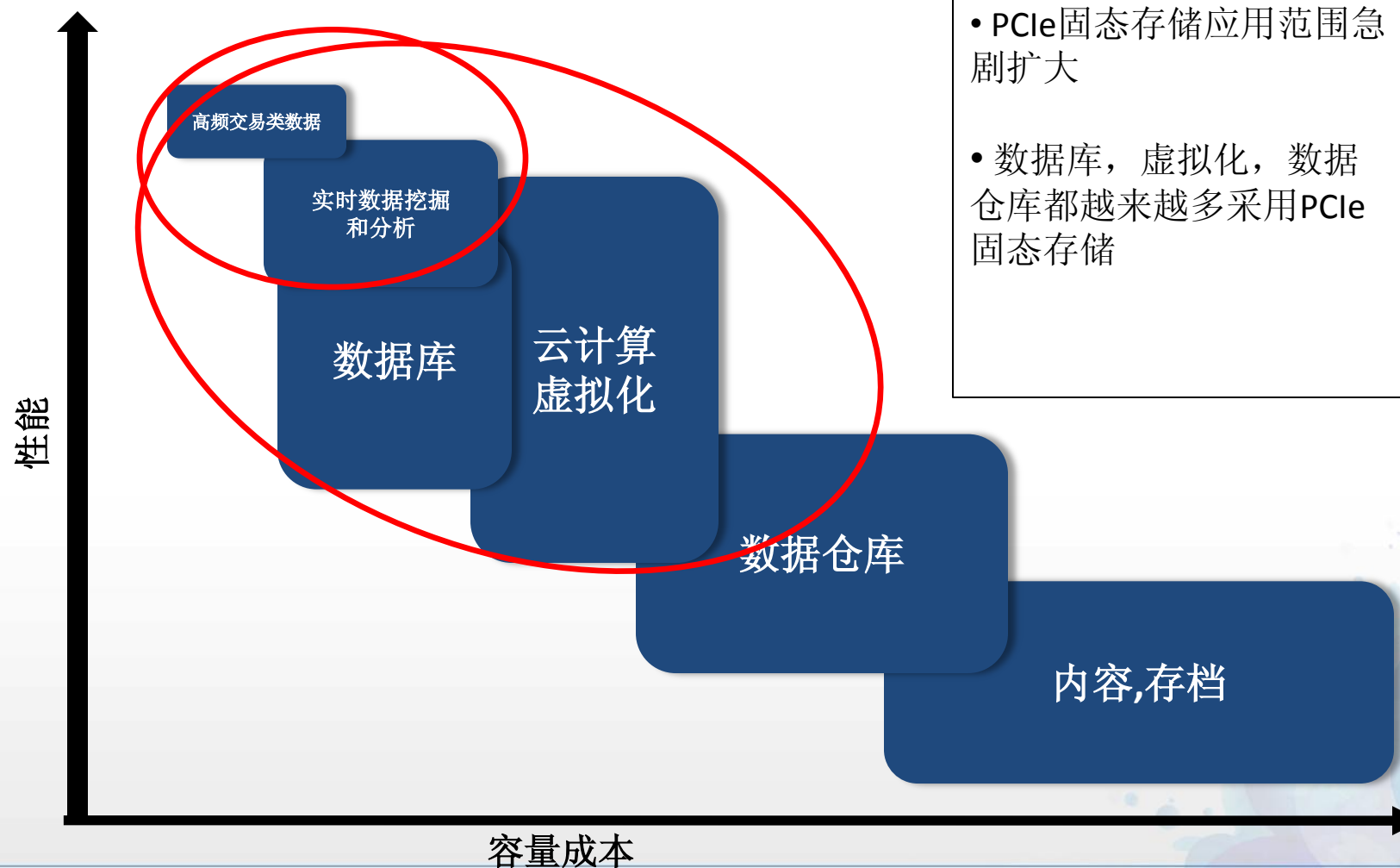
采购成本降低 **10x**

耗电量减少 **40x**





应用领域





PCIe Flash加速应用

数据库



虚拟化



搜索



大数据分析



高性能计算

AUTODESK



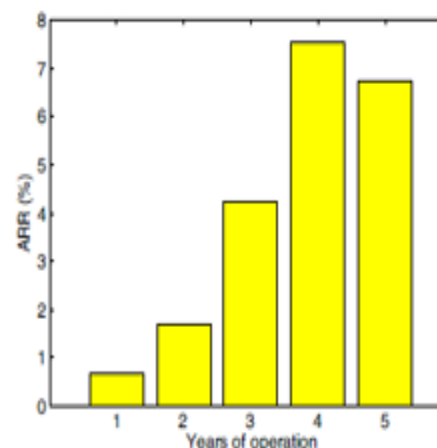
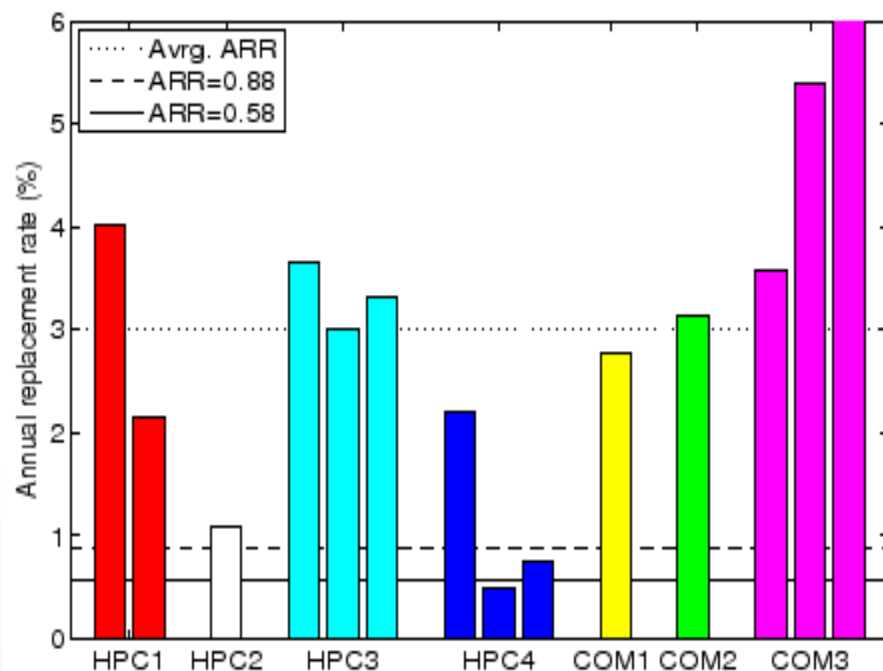


提纲

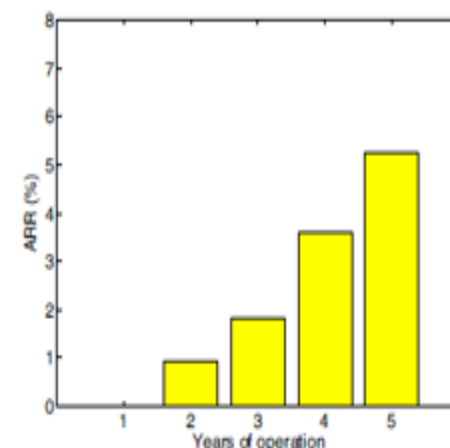
- 数据中心架构的变迁 – 闪存革命
- PCIe 固态闪存卡简介
- 使用**PCIe**固态闪存卡的顾虑
 - 寿命
 - 稳定性
 - 可靠性
 - 成本
- PCIe Flash存储的优势



HDD的寿命



HPC1 (compute nodes)





HPC1 (filesystem nodes)

实际失效率：Datasheet的2x – 30x



可用寿命

	15K SAS HDD	PCIe Flash 3.2TB
		
稳态8KB IOPS	400	60K
可写入数据量/天	280GB	16TB @ 5DWPD



PCIe Flash寿命

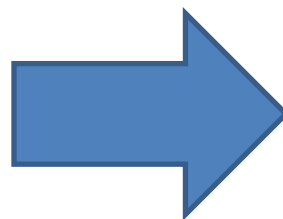
- 10x HDD RAID
 - 4K IOPS@8KB
 - 2.8TB/天写入寿命，3年3PB数据
- PCIe Flash
 - Up to 60K IOPS@8KB
 - 16TB/天写入寿命，3年17.5PB数据
 - 如果允许2.8TB/天写入，Flash使用寿命为17年



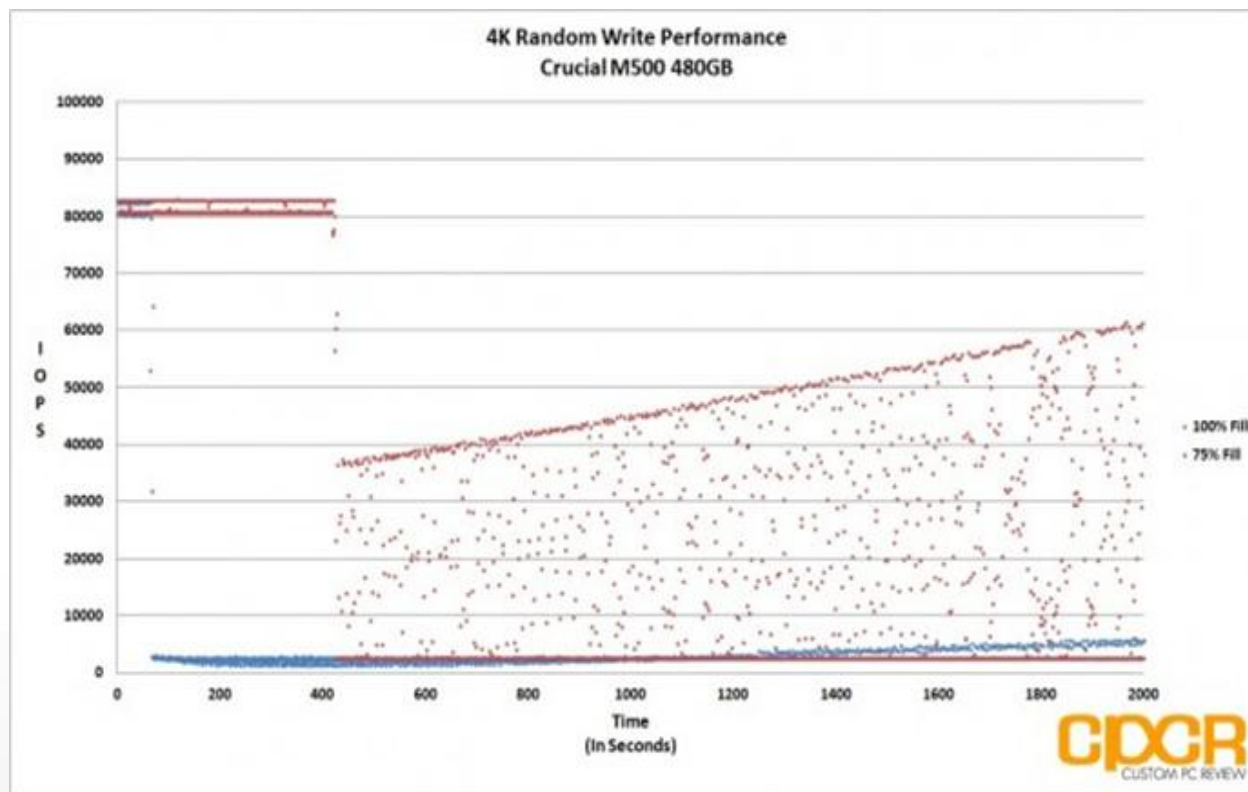
PCIe Flash寿命



6x 实际使用寿命



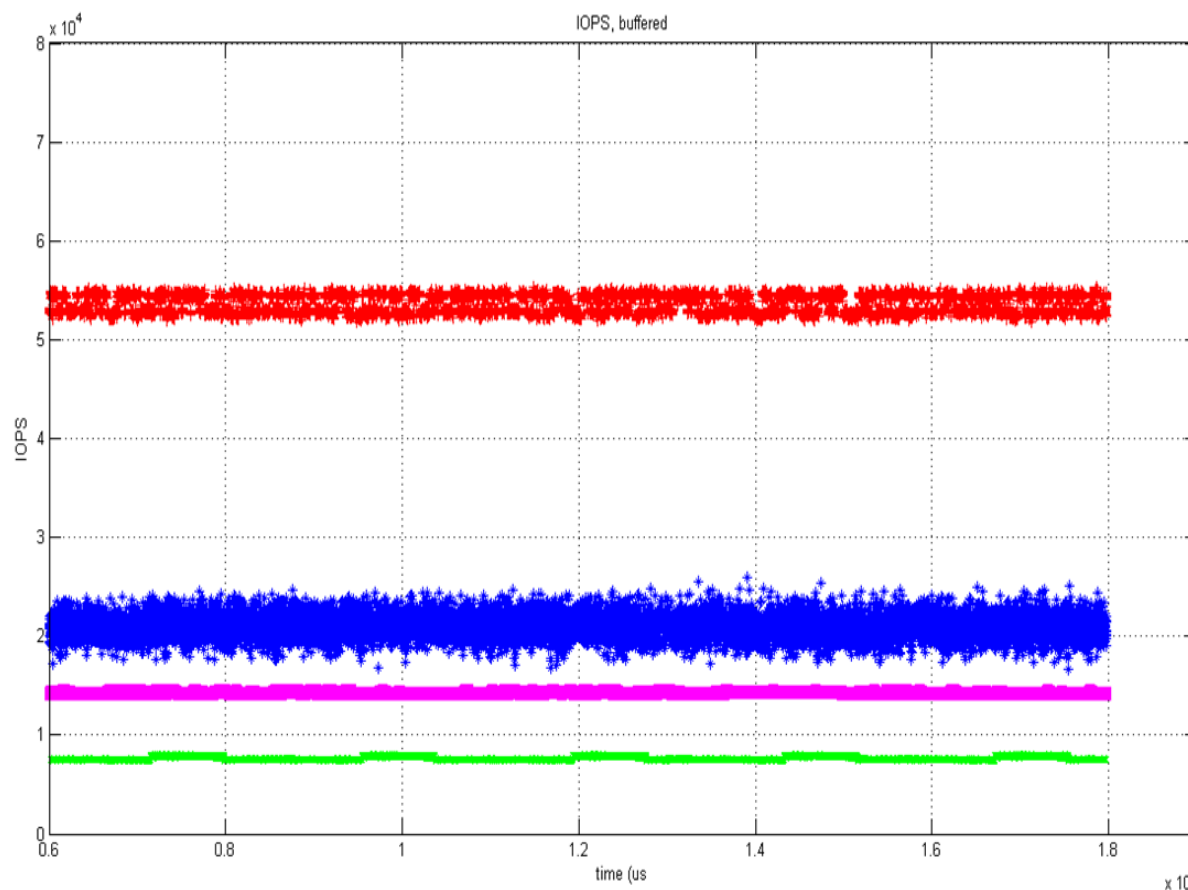
SATA SSD系统的稳定性



- 由于控制器原因，可能造成SSD性能波动



PCIe SSD的性能稳定一致性



- 4K随机读, 64K顺序读, 4K随机写, 64K顺序写

PCIe SSD可靠性和安全性



- 智能闪存转换层 – Smart Flash Translation Layer

- 智能热、冷数据动态跟踪
- 动态垃圾回收和磨损平衡，写放大因子最小化
- 双重数据保护机制
- 最大化NAND闪存的寿命

- 完善的容错数据保护机制

- 高达 40bit/1KB ECC或更高
- 读写，擦除出错处理及数据保护
- 页面，坏块出错处理及数据保护

- 内置RAID机制

- RAID-(N+1)冗余阵列
- 动态，可配置冗余度
- 进一步防止数据丢失

Lun 0	Lun 1	Lun 2	Lun 3
Data block	Data block	Data block	Parity block
Data block	Data block	Parity block	Data block
Data block	Parity block	Data block	Data block
Parity block	Data block	Data block	Data block

- 端到端数据保护

- 企业级端到端数据链路保护
- 多重数据完整性及正确性校验

- 过热保护机制

- 防止系统过热对系统造成不可恢复损伤

- 掉电数据保护

- 完善的突发掉电数据保护机制
- 防止系统不正常关机的数据完整性和安全性

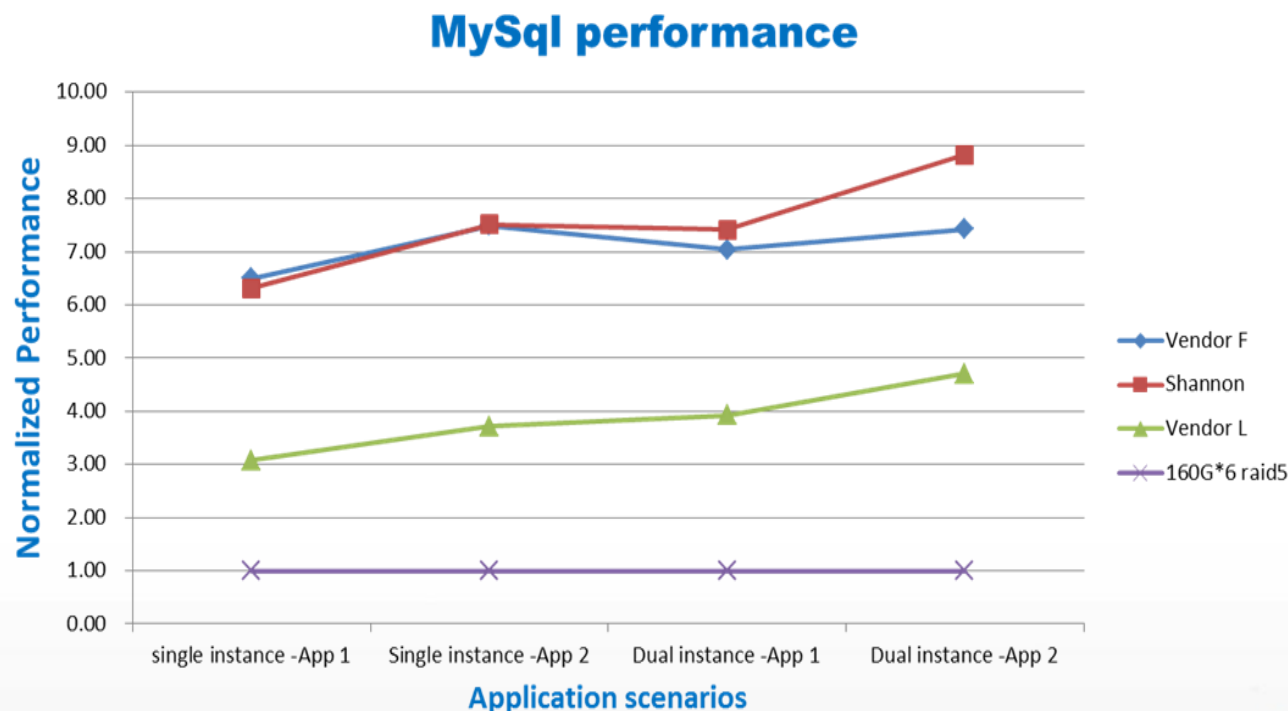


衡量成本的维度

	每IOPS的成本 (USD)
Direct-IO PCIe Flash	<0.01
SATA SSD	0.03 - 0.05
HDD (7200RPM)	0.9
HDD (10K RPM)	0.84
HDD(15K RPM)	0.78



PCIe Flash vs RAID SATA SSD



- 10x 业务处理能力提升
- Server + HBA 及运维成本考虑



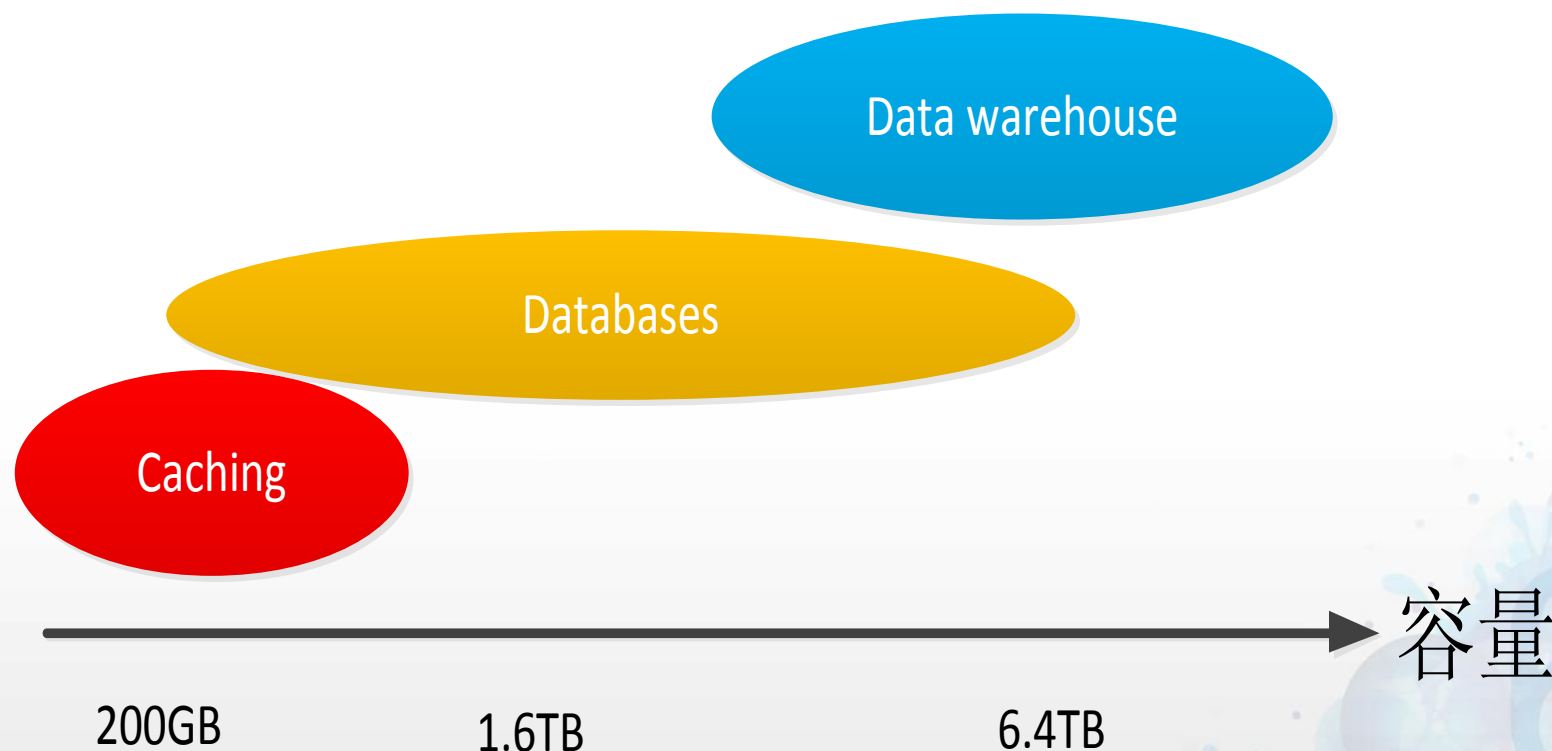
提纲

- 数据中心架构的变迁 – 闪存革命
- PCIe 固态闪存卡简介
- 使用PCIe固态闪存卡的顾虑
 - 寿命
 - 稳定性
 - 可靠性
 - 成本
- PCIe Flash存储的优势



Shannon PCIe Flash

- 200GB 至6.4TB容量





PCIe Flash物理形态

- PCIe卡





Shannon 8639 2.5寸盘

- Native PCIe接口
- 支持热拔插
- 专利cross盘的RAID技术



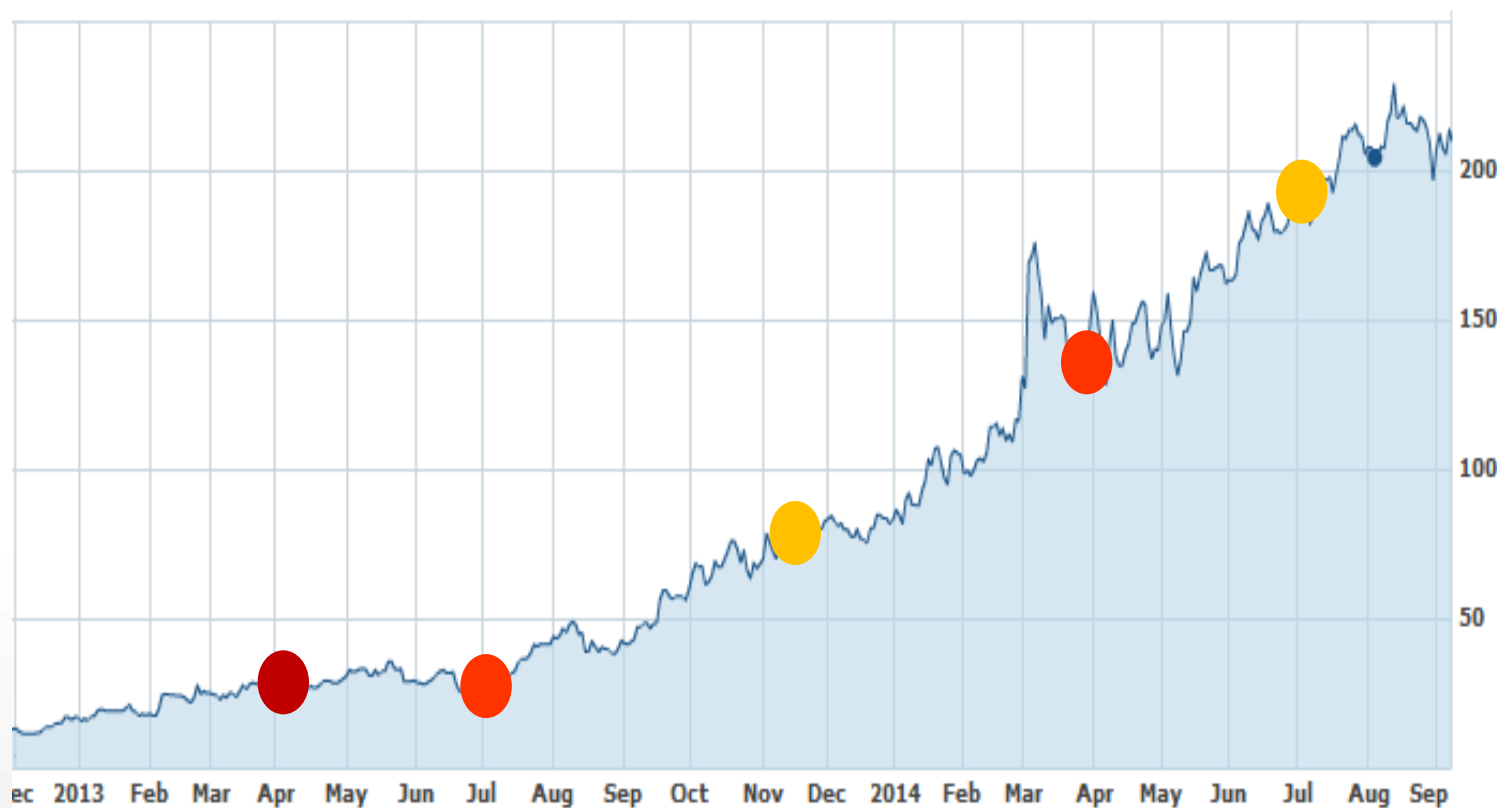


PCIe Flash使用者

1. Facebook
2. Apple
3. ...



Shannon PCIe Flash支撑业务发展



部署时间点

联系我们

地址：上海市杨浦区大连路588号宝地广场A座305室

电话：021-55580181

邮箱：contact@shannon-sys.com

官方网站：www.shannon-sys.com

Weibo: @宝存科技

WeChat: Shannon-Systems



Q&A

THANKS

SequeMedia
盛拓传媒

IT168.com
www.it168.com

ChinaUnix

ITPUB