2022闪存市场全景白皮书

2022 Memory panorama white paper

百易传媒(DOIT)

百易存储研究院

2022年7月

PREFACE

前言

回顾一下，2020年的时候看闪存市场，看到了什么？

在**《2020闪存市场全景白皮书》**中，我们关注到了闪存存储介质从2D-3D的变化，关注到3D堆叠设计、层数之争带来的成本降低，关注到TLC-QLC-PLC的变化；与此同时，螳螂捕蝉，黄雀在后，3D Xpoint给SSD市场带来新的变数，中国元素给市场带来的影响预测。在2020年前后，从闪存存储介质到闪存盘设计，性能和功耗还是设计上的限制和天条、紧箍咒，要性能，要控制功耗，几乎是二选一的选择，鱼和熊掌不可兼得；类似，要不要控制器，所谓Host-Base、Device-Base之争，当时还是市场关注的焦点。

NVMe SSD和PCIe x.0是几年来存储系统技术发展的两条线，市场在不断演进和变化，新技术的采用不断提高存储系统的延迟、IOPS性能表现，NVMe SSD的价值在于时延，PCIe x.0的作用在于SSD和CPU连接的带宽，从PCIe 3.0、到4.0、5.0以及未来的6.0，带宽会以成倍的速度提升，也许是因为疫情的缘故，支持PCIe 4.0的SSD产品市场有所放缓。

在**《2020闪存市场全景白皮书》**中，准确预测到分布式全闪存的方向，当时，尚无存储系统供应商跟进分布式全闪存的概念，2年后分布式全闪存成为了市场现实。当时，看到了NVMe-OF、SCM技术趋势，如今，这两个技术仍然演进推广中，市场普及和进展稍显缓慢。

以上基本上就是**《2020闪存市场全景白皮书》**的主要内容。

现在问题来了！

当时光来到2022年的今天，当我们重新聚焦闪存市场，2022年闪存市场有哪些变化和最新进展呢？

如今的闪存市场，看上去风平浪静，波澜不惊，新的技术和新的话题没有什么，但实际上已经是暗潮涌动。也因为如此，需要着墨的内容并不多，为此**《2020闪存市场全景白皮书》**仅仅是在此前的基础上进行比较的修订。

《2022闪存市场全景白皮书》在编纂的过程中得到了新华三、浪潮、华为、Hitachi Vantara、联想凌拓、宏杉科技、XSKY、同有、宝存、中存超为、川源等企业专家的鼎力支持（排名不分先后），在此表示由衷的感谢。

百易存储研究院

2022/7月

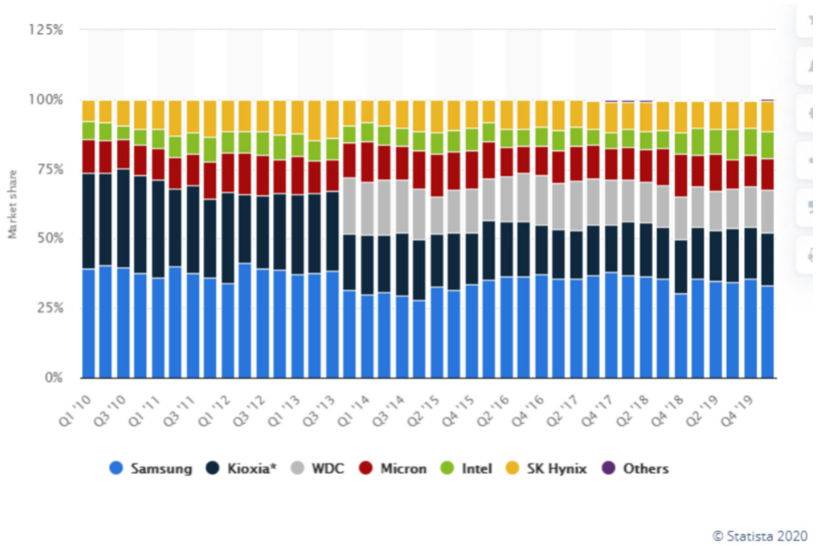
CONTENTS

目录

01

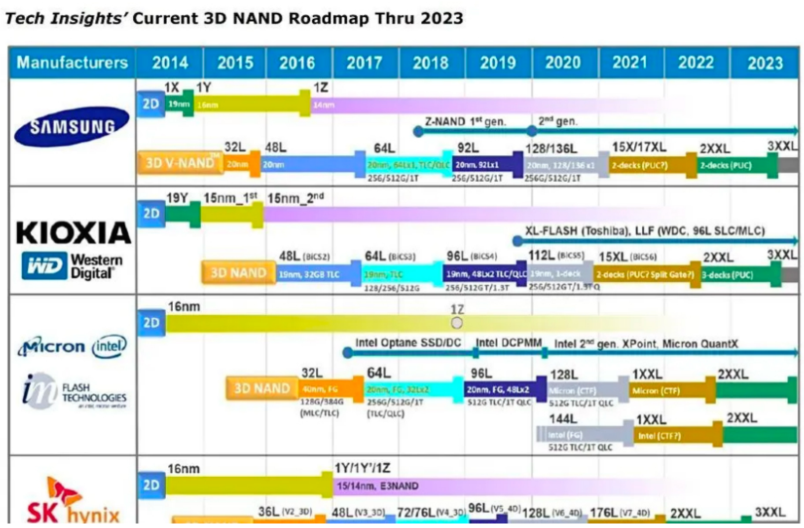
闪存介质层

1.1 NAND 层数进化，成本降低不如预期



最早的 3D NAND 要追溯到 2007 年东芝的一次发布，而真正将 3D NAND 商业化应用的其实是三星在 2013 年的发布，从 2013 年起，3D NAND 作为最具突破性的技术开始推向市场，揭开了闪存介质容量提升的新时代。

从 2D 到 3D 的变化，为闪存容量提升带来新变数，从 2D 平房到 3D 楼房的演进下，层的概念成为介质进化的新标准。也是从 3D NAND 开始，制程工艺的演进显得并不那么重要了。



3D NAND 路线图（图源：Tech Insights）

从最初的 32 层开始，到 64 层大约经过了三四年的时间，2019 年前后，96 层 NAND 开始出现，2020 年进入 200+ 层时代，从 128 层开始，3D NAND 告别了 32 整数倍的演进，开始出现 162、176 层的差别。

高层堆叠的 NAND 技术推动了容量发展和闪存普及的步伐，让闪存在更多场景得到应用，在一定程度上，是 3D NAND 的出现拉开了闪存普及的大幕。3D NAND 颗粒层数不断推高，GB/$ 成本势必会不断降低，缩短闪存 SSD 和磁盘的差价，但从市场的演进看，闪存 SSD 成本是在逐步降低，但是速度不如预期，究其原因和产能、良品率等很多因素有关。如今，闪存 SSD 产能被少数品牌把持，受产能限制，大幅度降价未必符合这些厂商的利益，从这个意义上，中国元素的崛起，符合全球市场的利益。

3D NAND 的堆叠技术方案有两大派系，一种是浮栅技术（Floating gate），另外一种是电荷撷取（Charge Trap）技术。

浮栅技术的上下单元之间是分离的，而电荷撷取技术上下层之间是连通的，两者都是几十层水平时，差异不大，当变为几百层之后，采用电荷撷取技术的 NAND 上下电荷之间会发生串扰，造成数据不可靠。而浮栅技术虽然工艺更复杂一些，但是在数据保留方面更可靠一些，浮栅技术在层数发展上更有前景。

1.2 QLC 不是主流 SSD 被写穿需预防

2014 年开始，TLC NAND 开始在市场上大范围出现，所谓 TLC 是指每个单元（Cell）中存储 3 个 Bit，当时的主流还是 MLC（每个单元中存储 2 个 Bit），再往前看还有 SLC（每

Will TLC-Based SSDs Prevail?

	Client			Enterprise		
	SLC	MLC	TLC	SLC	MLC	TLC
2005	Stock	Never	Never	Never	Never	Never
2006	Stock	Never	Never	Never	Never	Never
2007	Stock	Never	Never	Never	Never	Never
2008	Stock	Never	Never	Rare	Never	Never
2009	Stock	Rare	Never	Rare	Never	Never
2010	Stock	Rare	Never	Stock	Rare	Never
2011	Rare	Stock	Never	Stock	Rare	Never
2012	Rare	Stock	Rare	Stock	Stock	Never
2013	Rare	Stock	Rare	Stock	Stock	Never

个单元中存储 1 个 Bit），随着每个单元存储的数据增多，容量得到提升，作为代价的是，性能和耐久性都有所损失。

从 MLC 到 TLC 转变过程中，人们最大的关注点还在于介质耐久性的降低，随着时间的推移，技术上的进步最终掩盖了介质上的差距，过去八年以来，TLC 逐渐成为主流，从 2018 年 QLC 开始出现，介质耐久性又一次降低。

有数据显示：从 SLC-MLC-TLC-QLC，其可擦写寿命 P/E（Program/Erase cycle，一次完整 NAND 全盘写入）从 10 万次、3000~1 万次、300~1000 次、100~150 次一路走低，编程复杂性的提升，读写速度，特别写速度不断降低。

特别是 QLC，耐久度已经降低到 100~150 次，对于消费类用户来说，高清视频等存储的需求，主要集中在大容量、低成本，写满一次磁盘需要较长时间，频繁擦写的情况并不多见，因此 P/E 100 是可以满足需求的。对于 QLC 大容量 SSD 盘来说，应该没有足够的数据，将全盘写满 100 次，且写满 100 次也需要很长的时间。

在消费级市场，很少听说 SSD 被写穿现象。

但在企业级市场就不一样了，不仅数据频繁采写，企业级 SSD 对于数据的可靠性、安全性要求更高，不能够容忍任何的数据丢失。如今，企业级市场，出现了 SSD 被写穿的现象。这个现象背后，应该有类似 QLC 耐久度不高的问题，也有软件、调度、监管和控制的问题。针对 NAND 颗粒耐久度，在 SSD 盘固件设计上，就有磨损均衡设计，尽可能将数据均匀散步在各个颗粒，避免局部热点的出现。此外，外部存储系统在设计上，会使用多个 SSD 盘，用户的数据也会被条带化，均匀分布到各个 SSD 盘，一来通过并行存取提高性能，二来提高可靠性，防止局部热点所导致的 SSD 被写穿。

总而言之，无论 SSD 盘，还是存储系统，有很多的技术手段来预防局部热点，防止 SSD 被写穿。人尽其才，物尽其用。

当企业级用户出现 SSD 被写穿的时，还是应该认真分析问题的缘由，是否为了追求价格忽略了应用场景和耐久度的问题。总的来看，QLC 适合一次写入、多次读取的应用场景，并不适合频繁的写入，这也是 QLC 没有成为主流存储的原因。

另外需要引起关注的：最新技术表明，无论是 TLC 还是 QLC，在可擦写寿命 P/E 上都取得了明显的技术进步，分别从 300~1000 次、100~150 次，提高到了 3000~7000 次、1000~3000 次，虽然不了解具体做了哪些技术改进，但是可以肯定的是，PE 的提升就为企业级应用提供了良好的保障，与此同时，提高 GB/$，降低闪存介质的使用门槛。如今，在很多的用户业务场景，如容灾备份等也开始接受全闪存配置，这应该是技术进步带来的成果。

1.3 全新存储介质 3D XPoint

2015 年，兼具高性能和非易失性两种特性的 3D Xpoint 问世，介绍资料显示，其速度比 NAND 闪存快 1000 倍以上，耐久性也远胜于 NAND 介质，可满足企业对高性能、高耐用性、低成本固态盘解决方案的需求。

3D XPoint 的工作原理与 NAND 存在着根本性的不同。NAND 工作原理是将电子来回移动于晶体管中的一个独立部分——称之为“浮空栅”来表示二进制中的 0 和 1，这种技术带来的问题是它不能随时重写单个比特单元的数据，需要采用块擦除，然后全部进行重写；3D Xpoint 则是改变材料本身的性质，通过一个高电阻的电信号来代表 1，一个低电阻信号来代表 0，这意味着每个存储单元可以被单独寻址，从根本上加快了执行效率。由于 3D XPoint 将材料熔化或凝固来存储数据，因此比 NAND 耐用性提高了数百倍。

3D XPoint 作为一种创新介质，区别于 RAM 内存和 NAND 存储芯片，性能和成本上的表现介于两者之间，有许多观点认为，RAM 内存与 NAND 芯片之间的差距比较大，导致整体架构上的数据流动受限，而 3D XPoint 则填补了中间的空白地带。

作为 3D Xpoint 的开发者，美光和英特尔都将其进行了商业化应用，在实际使用中，3D Xpoint 技术可以改善个人计算的体验，还可以用来打造高性能、高容量的储存解决方案，与内存（DRAM）相比密度高出 10 倍，而成本又远低于 DRAM 内存，这意味着可以输出更高的存储容量和更快的访问速度。

1.4 3D 堆叠必争制高点

全球 3D NAND 市场竞争是一场技术赛跑，其中，堆叠层数被视为一个制高点。进入 2022 年，堆叠层数已经来到 200+ 的门槛，已经宣布的有 162、176 层产品量产，长江存储 128 层实现量产，500+、1000+ 的方案也在规划中。目前市场上有两种堆叠方式：字符串堆叠（ String stack）和 单一堆叠（Single stack），各有利弊。

单一堆叠（Single stack）相当于原生垂直堆叠 1xx+，仅一道工序，有成本优势。但是制程复杂度随层数攀升不断加大，蚀刻和沉积最终可能导致裸片良品率下降，包括交叠层厚薄不均、蚀刻不彻底（打孔未到达底部）、弯曲、扭曲及线宽变化，其间电子垂直通过的迁移率也在随层数增加而更加不可控。

而字符串堆叠（ String stack，目前主要指双层堆叠），制程复杂度相对较低，只要拥有 48 层或者 64 层技术，那么实现 96（2*48）层、128（2*64）层不是梦，还降低了打孔的蚀刻深度，良品率不会受到影响，但需要两道工序，会增加 30% + 的成本。

3D Layer Cake

SK hynix		SK hynix Solidigm		Western Digital/Kioxia		Samsung		Micron		YMTC	
Generation	Layers	Generation	Layers	Generation	Layers	Generation	Layers	Generation	Layers	Generation	Layers
V3	48	Gen 1	32	BiCS 2	48	V3	48	Gen 1	32		
V4	72	Gen 2	64	BiCS 3	64	V4	64	Gen 2	64	Xtacking 1	64
V5	96	Gen 3	96	BiCS 4	96	V5 (May 2018)	9x	Gen 3	96		
V6 (Q4, 2019)	128	Gen 4 (Q4 2020)	144	BiCS 5 (Q1,2020)	112	V6 (June 2019)	128	Gen 4 (Q4, 2019)	128	Xtacking 2	128
V7 (2021)	176	Gen 5? (2023)	196?	BiCS 6	162	V7 (April 2021)	176	Gen 5 (Nov 2020)	176		
V7(2023?)	238			BiCS 7	212	V8 (July 2021	200	Gen 6 (End 2022)	232		
				BiCS 8	?	V9	3xx	Gen 7	2yy		
V? (2025)	500+					V10		Gen 8	3xx		
V? (2030)	800+					V11	5xx	Gen 9	4xx		

值得一提的是长江存储的 Xtacking 技术，这是一种晶圆键合（Wafer bonding）而非堆栈方式。一片晶圆上加工负责数据 I/O 及记忆单元操作的外围电路，另一片晶圆上加工存储单元，最终通过数百万根金属 VIA(垂直互联通道) 将二者键合接通电路。

目前，字符串堆叠方式在市场上较为普遍。

1.5 GB/$ 的考验

GB/$ 将是 NAND 跨越磁盘成为主流存储介质最直接的考验，为此，技术上一是堆叠晶圆的层数；二是增加单位单元存储比特位数量。层数增加，意味着制程复杂性和处理时间依次增加，裸片良率可能大大下降继而导致产量下降。

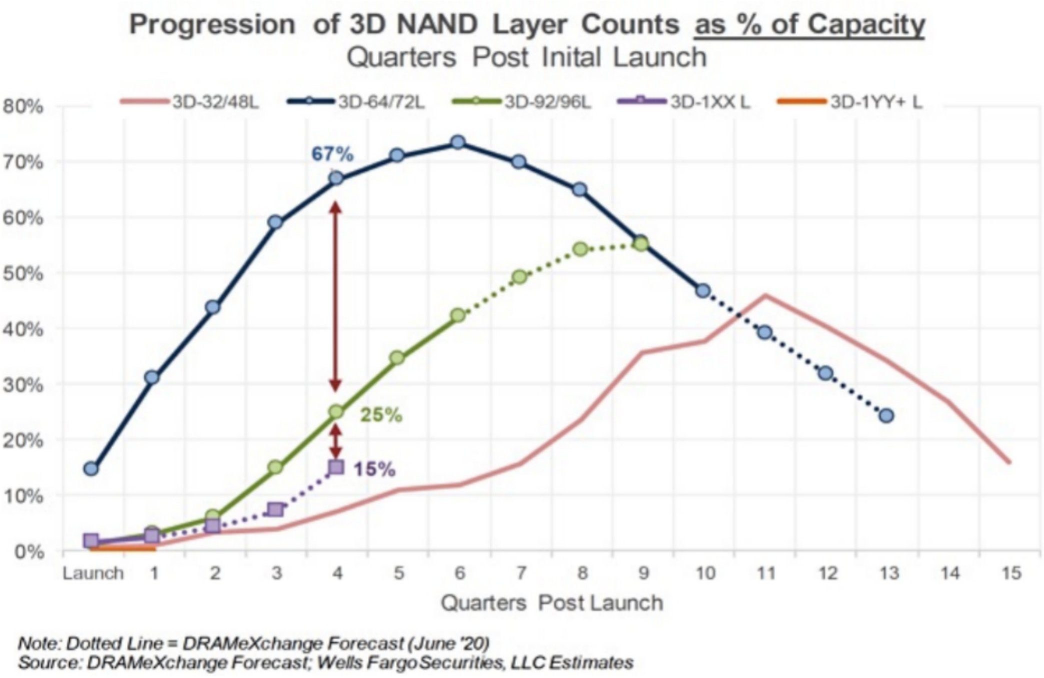
国际分析机构富国银行的分析师 Joe Quatrochi 提到了晶圆上蚀刻通道的纵横比问题，预计 96 层设备的纵横比约为 70：1（对应 64 层是 60：1），纵横比会随着层数增加而提升。3D NAND 的垂直堆叠属性逐渐依赖蚀刻工艺的精度平衡更高的纵横比，此外，通过沉积实现晶圆上薄膜的一致性也更加困难。

128 层单层蚀刻时间约是 96 层蚀刻时间的 2 倍。想象一下自己的机器用之前两倍的时间制造晶圆，通过增加层数让每个裸片容量增加约 30%，实际发售容量是否真的增加了？为了增加产能还要开设或者更新更多的生产线乃至最后建立新的晶圆厂，这部分的投资数额通常要以百亿美元来计算。

从 SLC 到 MLC（每单元存储 2 个比特位），100% 容量增加，到 TLC 是 50% 容量增加。但转向 QLC 就意味着 33% 的容量提高，到 PLC（每单元 5 个比特位）则是容量增加 25%，下次可能只是 20% 增长。此外，从 TLC 到 QLC 再到 PLC 耐久性越来越差，为此 SSD 要做超额配置（Over-provisioning ），通过预留空间 (extra cells) 替换故障的单元，耐用性越低，超额配置就得更多。

为了用上更高级的 NAND 技术，有许多介质以外的许多成本。比如，需要通过重构标准 SSD 外形尺寸增加 NAND 裸片数来提升可用物理空间，需要通过控制器误差校验技术减少错误，减少随机写入并且做更多超额配置（OP）来延长正常使用寿命。

考虑到以上各种 NAND 裸片问题，厂商为了产能与利润的平衡要做出自己的决策，于是，便导致 SSD 容量增速放缓，下图是富国银行通过层数呈现的市场发货量随时间发生的变化。



与 64/72 层 NAND 相比，64 层 3D NAND 在发布 4 个季度后占当时闪存发货量的 67%，而 96 层 NAND 达到了这样的发货量需要更长的时间。100 层以上 NAND 第一次迭代同样需要更长的时间。所以说，SSD 容量增速放缓。

1.6 中国元素

截止 2021 年，中国大陆在全球晶圆产能中的份额达到 16%，仅次于韩国和台湾地区。这个数字在 2011 年仅为 9%。较低的建设和运营成本是中国近期晶圆产量增长的主要原因。SIA 和波士顿咨询集团于 2021 年发布的一份报告显示，在中国建造和运营晶圆厂的成本低于任何其他国家。

在过去两年里，中国大陆的晶圆产能几乎每年要增长 1 个百分点，产能扩张速度比其他任何地方都快。截至 2021 年底，全球 IC 晶圆产能达到每月 2160 万片 200 毫米当量晶圆。中国大陆的晶圆厂在其中拥有处理约 350 万片晶圆的产能。

但根据 Knometa Research 的报告，中国大陆约有一半的晶圆产能，都归属其在大陆以外地区的母公司，包括台积电、联电、三星和 SK 海力士等。

中国作为全球第二大经济体国家，所取得的进步和成就世所瞩目；经济实力的增长，需要中国对于全球经济发展和科技进步有所贡献。从技术发展看，全球 NAND 制造正在从 2D 走向 3D，对于半导体制程工艺技术的要求有所降低，所使用的是落后一代半左右的技术，这也是目前中国所掌握的技术。因此无论是自身的实力，还是人类社会发展前所未有的需求，中国都有能力也有义务作出自己的贡献。

长江存储早在 2020 年中就已筹建存储基地项目二期，目前正在加速推进，预计未来两年就能投产，其二期项目的月产能为 20 万片，一期项目为 10 万片，如此两年多之后 NAND flash 芯片产能将增加两倍。

产能方面，长鑫也不甘落后。公开资料显示，长鑫在 2020 年、2021 年分别实现了 4.5 万片晶圆 / 月、6 万片晶圆 / 月的目标，2022 年的产能目标是 12 万片晶圆 / 月，未来的产能目标是 30 万片晶圆 / 月。

截止 2021 年，全球 DRAM 总产能大约是 150 万片／月，预测到 2022 年底，会提升至 160 万片／月左右，届时长鑫存储的 12 万片产能将获得全球市场份额的 8% 左右。

中国存储力量的快速崛起，对于全球市场而言，将是一幅美好的画卷。全球闪存市场会因为中国变得更加美丽。

02

闪存控制器

2.1 OpenChannel 的话题

所谓 Open-Channel SSD 指的是 FTL（Flash translation layer）不在 device 端，而是由 Host 的操作系统维护的 SSD。

Linux 内核从 4.4 开始已经支持 Open-Channel SSD 了， 并提供了符合 NVMe 协议的抽象层——lightNVM。

Host-Base 还是 Device-Base 之争，一度是控制器领域的焦点话题，技术核心在于是否需要本地处理器的问题，Host-Base 不需要本地处理器，主要借助主机 CPU 的资源。

Host-Base 还是 Device-Base 两条技术路线看上去针锋相对，没有妥协的余地。但是从发展的历程来看也并非如此。以互联网企业为代表，应用层希望更多了解底层资源的管理和调度。以预留空间（OP）为例，原本是预留资源，用于提升 SSD 可靠性和稳定性。

NAND 闪存颗粒故障率呈现一种双曲线的模型，刚开始使用的故障率高，而后会进入稳定器，然而，随着大规模使用，质量不好颗粒会出现损坏，导致故障率上升。借助 OP 设计，可以将质量不高的 NAND 颗粒进行替换，故障率降低，SSD 进入稳定工作状态。OP 预留容量过大造成浪费，过小又无法满足可靠性的需求。

针对 SSD 盘的这些特性，互联网厂商希望能够从应用的整体系统性的考虑可靠性的问题，扬长避短，用于追求系统资源的最大化。

由于没有本地处理器设计，Host-Base 的 SSD 成本更低，也更加节能。但是并非所有用户都具备互联网厂商的技术能力，很多成熟的应用解决方案也未必考虑到了底层设备的特点，它们更需要一种通用的产品设计，在兼容性、继承性上，Device-Base SSD 也更加符合这种特点。

从历史的发展进程看，Host-Base 还是 Device-Base 彼此竞争，但是也彼此借鉴，取长补短，又呈现出一种融合发展的趋势，也渐渐为二者之争画上了休止符。如今，

Host-Base 还是 Device-Base 已经成为历史，很少有人提及，成为一段尘封往事。

如今，OpenChannel PCIe 3.0 产品已大规模商用；PCIe 4.0、NVMe、ZNS 的 Open Channel SSD 已有陆续推出，如今，OpenChannel SSD 超过 80%服务于国内互联网企业，涵盖互联网、服务器、政企、运营商等领域。

对于 OpenChannel SSD 的未来发展，需要引起关注的是 ZNS（ZonedStorage，分区存储技术），从一定程度上来说，二者殊途同归。通过对数据存储、管理方式的重新组织，最大发挥底层存储介质的特性，给数据存储带来效率。二者的结合，需要引起高度的重视。

2.2 高性能还是低功耗？

对于 SSD 而言，由于 25w 功耗上限的制约，处理器选择将决定 SSD 盘的性能以及功耗，目前 SSD 盘有不同功耗模式，存储系统可以选择，可以自动调节性能区间，但是 25w 功耗上限总会是一个紧箍咒。在新的 EDSFF 规范中，这个上限将被打破。

SSD 控制器多采用 ASIC 设计，选择的重点是性能和功耗的平衡。但处理器的功耗确定下来之后，SSD 盘的性能也就大体确定了。

SSD 盘的功能特性，如磨损均和策略、加密 / 解密算法、ECC 校验等是由固件来决定的，每个系统盘设计，所采用的策略不同，SSD 盘所表现出来的功能和特性也不同。

这些不同的功能特性在 SSD 盘使用初期并不易察觉。以磨损均衡和垃圾回收策略为例，只有在使用一段时间之后，数据量达到一定程度之后，它们对 SSD 盘的影响才能够看出来，例如应用系统性能出现抖动，可靠性下降等。

这些都需要通过测试，认真细致加以观察和了解才能发现。

不同品牌的 NAND 颗粒差别非常大，好的控制器应该懂得扬长避短，各取所长。也因为这些差异性的存在，SSD 盘，特别是企业级 SSD 盘并不能够简单等同于“标准工业化”产品，这是需要存储阵列厂商和服务器存储厂商需要重视的。

2.3 分区存储技术的最新发展

有数据表明， 2018 年全球产生的数据量是 32ZB，预计 2023 年会达到 103ZB。但是其中，2018 年被保存下来数据只有 5ZB，占比不足 16%；预计 2023 年 12ZB，占比 12%，这意味着会有越来越多的数据没有办法保留下来，很重要的一个原因就是存储能力增长没有办法追赶上数据的增长。

为了追逐更大容量的存储，以及 GB/$ 成本，无论是 SSD 盘，还是传统 HDD 磁盘都在努力，其中 SMR HDD 磁盘、QLC 颗粒 SSD 盘是成本最低的产品，是最新科技的体现。但是 SMR、QLC 也都面临写入限制的问题，从而制约了产品技术的普及。

分区存储技术（ZonedStorage）就是在这样的背景下应运而生的，简单说，这是一套存储资源命名空间分区的技术。由西部数据提出并开源，得到了 SUSE、ATTO、Microchip、Broadcom 等伙伴的支持。

最了解 SMR、QLC 的莫过于西部数据这样的厂商，分区存储技术就是基于对底层技术的深刻理解，试图从系统层面，改变数据的读写的顺序和规则，从这个角度来说，分区存储技术是一种突破和前瞻。

但是分区存储技术是否能够如愿却并不取决于技术，而是取决于生态和商务环境，

取决于分区存储技术是否能够得到系统厂商、平台和应用软件的厂商的支持，这是一场深刻而重大的改变，也因为改变之重大和影响，产业生态要取得共识并不容易。

分区存储技术发布于 2019 年，如今 2 年多时间过去了，分区存储技术的应用如何？

我们在前面提到 SSD 被写穿的问题，也说到了 QLC 耐久度只有 100~150 次，说到了 SMR 磁盘驱动器在数据临时存储缓存区域写满时引发的性能缓慢问题，这就需要分区存储扬长避短。

2022 年 3 月，三星和西部数据签署一份谅解备忘录（MOU），携手开发下一代数据放置、处理和结构（D2PF）存储技术标准，三星和西部数据启动了一项围绕分区存储设备的计划，其中包括了分区命名空间（ZNS）、SSD 和 SMR（叠瓦式磁记录）HDD。通过 SNIA（存储网络行业协会）和 Linux 基金会等组织，三星和西部数据将为下一代分区存储技术定义高级模型和框架。

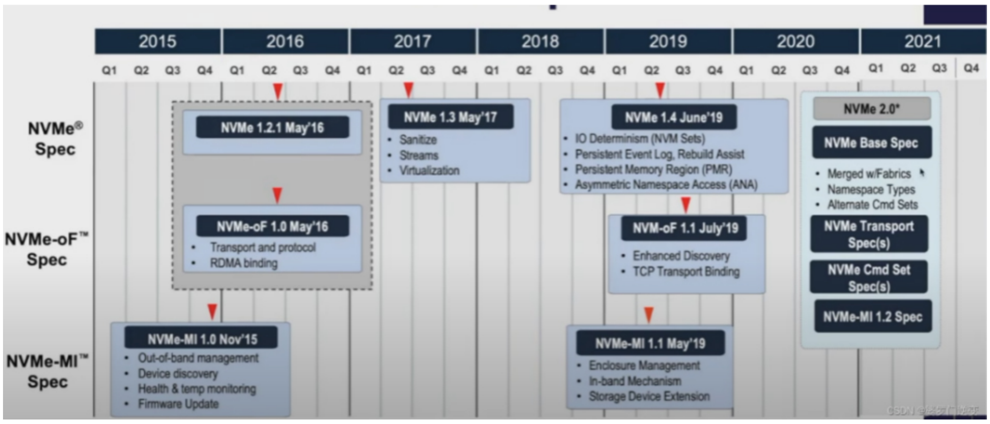
为了实现开放和可扩展的数据中心架构，双方还成立了 Zoned Storage TWG（技术工作组），目前已经在定义和指定分区存储设备的通用用例，以及主机 / 设备架构和编程模型。ZNS 技术允许根据使用和访问频率对数据进行分组，并按顺序存储在固态硬盘内的独立区域中，从而避免了移动和重新排列数据，能显著减少写入操作的数量，降低写入放大系数（WAF）。SSD 需要分配一定的空间作为预留空间（OP），以有效管理、改善闪存芯片的性能，而采用 ZNS 技术的固态硬盘则不需要这么做，可以让其可用容量更大。ZNS SSD 可以降低成本和功耗，并简化使用新的 NAND 架构，比如 QLC 3D NAND。

如今，三星和西部数据都发布了 ZNS SSD，但使用还不普遍，可能的原因是 ZNS 存储技术平稳互操作生态系统仍未形成，也没有得到存储系统厂商的响应。

但是无论如何，作为一种具有战略性全局观念的技术，分区存储技术应该得到重视和关注。

2.4 磁盘和 NVMe

在我们的认知中，NVMe 是专门针对闪存 SSD 而设计，目的是为减少存储协议自身带来的的开销，是对原有 AHCI 存储协议的优化，去掉了原有协议中有关磁盘部分的有关内容。



但是令人意外的是 NVMe 2.0 引入了对旋转存储介质的支持，也就是 HDD 机械硬盘（磁盘）。

如今 7200 转 / 秒的磁盘并不能跑满 SATA 3.0 的 6Gb/S 的带宽性能，但磁盘技术也在不断发展之中，希捷、西部数据都推出了多磁臂的技术，以希捷 Exos 2X14 企业级 Mach.2 磁盘为例，其 IO 性能可以达到 524MB/s 的速度，性能接近低端 SATA SSD。

为了发挥多磁臂磁盘的性能，业内早就在推进磁盘支持 NVMe 的进程，统一数据存储通信协议。在去年的 OCP 虚拟峰会上，希捷就对当前磁盘现状和支持 NVMe 的好处进行了介绍，也向 NVMe 委员会提出了相关的技术建议。希捷认为，NVMe 的 HDD 很可能会在两年之内实现。

NVMe 2.0 将这一愿景再度拉近了一些，要知道 SATA 3.0 这一接口已经十数年没有更新了。彻底抛弃这一接口将为诸多应用提供更多空间，留给那些更加高效的接口。

存储系统厂商 Infinidat 也支持 NVMe 用于磁盘。Infinidat 的技术策略是用磁盘

加上 DRAM 的组合方式，然后在中间放置一层薄薄的闪存层，据说性能表现不逊于全闪存存储，Infinidat 宣布存储阵列 InfiniBox 支持 NVMe/TCP 协议。

从根本上来讲，当 NVMe 把磁盘直接连到 PCIe 总线之后，磁盘能拉近与 CPU、GPU、内存、SSD、网络交换机等这些部件的距离，数据流动会更顺畅。

03

闪存设备 SSD 盘

3.1 企业级 SSD 与消费级 SSD 的区分

固态硬盘普通的分级方式可分为企业级和消费级。

消费级 SSD，目前已经成为品牌电脑的标准配置，也可以通过单独购买添加，消

费级 SSD 和磁盘并存，能满足大部分用户的普遍需求，无需过度配置。

与消费级 SSD 相比，企业级 SSD 对稳定性、可靠性要求更高，数据作为企业用户的核心资产，不容有任何的闪失，此外，涉及 7 * 24 小时的频繁读写操作，其面临环境更加严苛，可以说，企业级 SSD 和消费级 SSD 完全不是同一级别的产品。

企业级 SSD 可以按照高性能、高容量、读取敏感、写入敏感等多种场景细分，稳定性不仅单指盘的可靠性，也包括性能一致性表现。除此之外还会有一些特定行业专用产品，如工业用小容量 SLC 固态硬盘 (通常是以存储芯片形式出现)、车载硬盘。他们的特征大多是要求宽温适应 (-40~85℃)、抗震动和高耐久度长期应用。相比家用固态硬盘 (工作温度 0~70℃) 来说，它们才是真正的耐用品。

事实上，针对各种固态硬盘的分级方法都有一定的局限性，某些名义上是消费级产品，但由于固件强壮性高同时也可被用于部分业级应用当中，某些名义上是零售消费级产品，但其实它是客户级系列变换而来，继承了高稳定性的血统。

这里从性能、可靠性、耐久性三个方面分析企业级 SSD 和消费级 SSD 的不同。

性能方面：在许多场景下，系统要求必须以最少的延迟量访问储存设备，即使有大规模并发访问也有稳定延迟表现，不能对响应时间造成影响。这意味着企业级 SSD 不仅要持续地提供服务，针对访问前数秒中的最高性能而且还针对运用更大的预留空间 (OP) 加以优化，在较长时段内连续提供保持稳定状态的性能；写入过程中产生累计功能性错误的统计值要求不大于 3%。而消费级只需要响应客户端使用程序仅涉及单个用户或使用程序访问，在任何用户或系统操作的最短和最长响应时间之间具有更高的容许值。

可靠性方面：企业级 SSD 的要求是 7*24 小时工作，环境温度不高于 55 摄氏度。一般来说，主控芯片结温过高会影响其寿命，而颗粒温度会影响其断电保持数据的时间。通常，消费级 SSD 每天最多 8 小时，环境温度不高于 40 度。

在断电数据保持时间方面，企业级 SSD 的要求 40 度的条件下不少于 3 个月，而消费级 SSD 的要求是不高于 30 度时不少于 1 年。

耐久性方面: 由于NAND 闪存颗粒的寿命问题,SSD 设备中的存储能力在不断下降,直至 NAND 闪存不再能够可靠地储存数据，此时，这类闪存应当从用户可寻址的储存池中除去，该逻辑地址移至闪存阵列上的新物理地址。

3.2 企业级 SSD PCIe 4.0 新时代

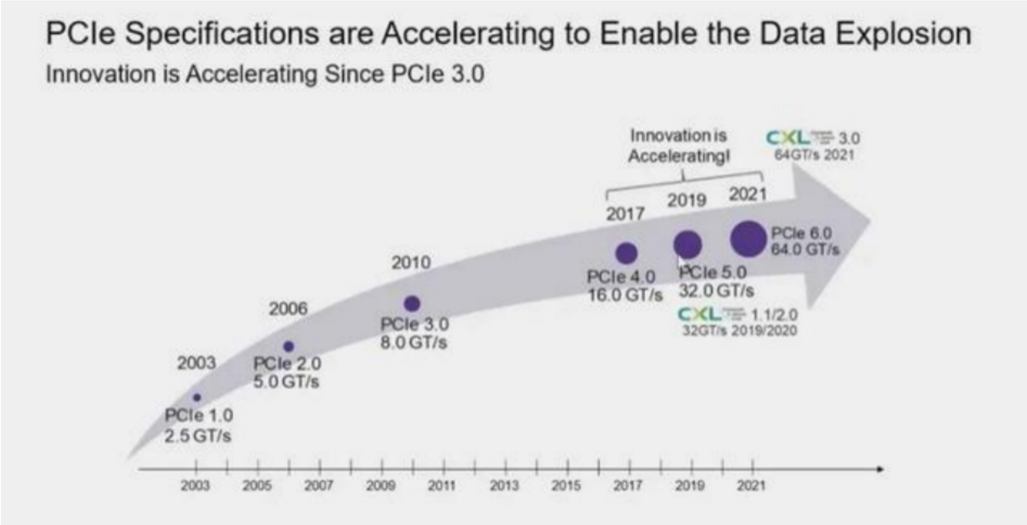
PCI-Express (peripheral component interconnect express，快捷外设互联标准)，是一种高速串行计算机扩展总线标准，它原来的名称为“3GIO”，是由英特尔在 2004 年提出的，旨在替代旧的 PCI、PCI-X 和 AGP 总线标准。如果将一台完整的 PC 比作一个人的话，CPU 是人的大脑，而 PCIe 就是人的神经中枢，它负责内部数据信息的交换。

PCIe 和 PCI (peripheral component interconnect，外设组件互连标准) 都是计算机总线的一种。总线（Bus）是计算机各种功能部件之间传送信息的公共通信干线。总线分为并行总线和串行总线，PCI 和 PCIe 都是总线的一种标准。

PCI 总线是并行总线，PCIe 是比 PCI 总线更快的一种串行总线；PCIe 通道包括两条差分信号，一条用来接收数据，另一条用来传输数据，其接口根据总线位宽不同而有所差异，一个 PCIe 连接可以被配置成 x1、 x2、 x4、 x8、 x12、 x16 和 x32 的数据带宽。其中 x2 速度是 x1 两倍，x4 是 x2 两倍，以此类推。

PCIe 属于高速串行点对点双通道高带宽传输，所连接的设备分配独享通道带宽，不共享总线带宽，支持主动电源管理、错误报告、端对端的可靠性传输、热插拔以及服务质量 (QOS) 等功能，数据传输速率高，发展潜力巨大。

PCIe 1.0 标准于 2003 年问世，在三年之后推出了 PCIe 2.0，在 4 年后的 2010 年 PCIe 3.0 问世。2017 年 10 月，姗姗来迟的 PCIe 4.0 仍然采用前代的 128b/130b 标准编码，但频率和带宽都直接翻番，分别达到了 16.0GHz 和 64GB/s。即将亮相的 PCIe 5.0 预计带宽和速度会继续翻倍。



PCIe Architecture	Raw Bit Rate	Interconnect Bandwidth	Bandwidth Lane Direction	Total Bandwidth for x16 link
PCIe 1.1	2.5 GT/s	2 Gb/s	~250 MB/s	~8 GB/s
PCIe 2.0	5.0 GT/s	4 Gb/s	~500 MB/s	~16 GB/s
PCIe 3.0	8.0 GT/s	8 Gb/s	~1 GB/s	~32 GB/s
PCIe 4.0	16.0 GT/s	16 Gb/s	~2 GB/s	~64 GB/s

基于 PCIe4.0 高速带宽的 NVMe 控制芯片、SSD 存储产品陆续问世，从 2019 年后半年开始，市场上陆续出现了许多 PCIe4.0 SSD，IBM Power 9 和 AMD 的处理器都已经开始支持 PCIe 4.0，而英特尔方面预计很快也将推出 PCIe 4.0 的 x86 处理器，2020 年将全面开启 PCIe 4.0 时代。

PCIe 的不断提速，对于数据中心、工控等各个领域影响非同一般，但在功能实现上需要考虑的问题也很多，包括散热、供电等方面的考虑。

目前市场上，企业级 SSD 已经迎来了 PCIe 4.0 的时代。

3.3 端到端 NVMe

对于需要极高性能和非常低延迟的工作负载，端到端 NVMe 是目前极致的选择。

为了实现端到端 NVMe 的需求，不仅需要内部 NVMe 连接的存储系统（这也是为什么高端存储开始选用 NVMe SSD 配置的原因），而且还需要对外支持 NVMe 连接，

如网络交换机，无论是基于光纤通道还是基于 IP，都将需要支持 NVMe，而两家主要的 FC 存储基础设施提供商现在正在提供这种支持， 任何支持 RDMA 的以太网交换机都将支持 NVMe。

从 IT 角度来看，确保交换机基础设施同时支持 NVMe 和传统 AHCI (SCSI 或 iSCSI) 协议非常重要，因为大多数环境不会一次切换到 NVMe-oF。要通过 NVMe 连接到存储系统的服务器需要具有支持 NVMe 的网卡，如今大多数网卡都具备这种能力。

一旦所有这些要求都得到满足，IT 规划人员就需要创建一条 NVMe-oF 唯一路径，从 NVMe 服务器通过交换机到存储， 在相同的逻辑网络路径上混用 AHCI 和 NVMe-F，可能会强制网络处理所有 SCSI 流量，这会影响性能。

实际上，网络可能要求 NVMe 流量位于其自己的逻辑路径上。

结合端到端 NVMe 的趋势，市场上有闪存存储发展三部曲的说法。

第一部曲是集中式全闪存阵列，它还是一个集中式全闪阵列，有控制器，有全闪存盘，基于专有的算法，类似分区存储技术的采用，发挥闪存颗粒性能的同时，保护其寿命。

第二部曲是扩展式、低时延的架构，基于 NVMe-oF 的存储网关，把存储性能和存储容量的扩展做解耦，通过 RDMA 网络实现极致性能和性价比。由于 RDMA 的出现，存储访问远端磁盘和本地盘已无区别，在这种情况下，集中式存储的控制器逐步退化成网关（算力与存力的解耦），慢慢的，甚至演变成板卡上的智能 SPU。

可以预计，一旦闪存二部曲实现，市场上现有的软件定义存储和机械硬盘就完美谢幕了。

第三部曲是一个更远大的目标，如今分布式存储非常热，但没有解决好的一个问题是容量扩展后的性能问题，其性能较之集中式全闪存是有差距的。第三部曲是一个高性能分布式存储解决方案，提供一个极致的性能和性价比。

3.4 傲腾需要区分为 SSD、Memory

作为市场上最早商用的 SCM 解决方案，英特尔傲腾 SSD 凭借介质上的优势，在延迟以及稳定性方面的表现碾压传统 NAND SSD，特别是在多用户、高并发访问 OLTP 的场景下，能持续稳定地输出非常低的延迟表现。介质层的优势将性能基线拉到了一个非常高的水平，这是通过软件层次上的创新无法达到的。

与此同时，傲腾凭借其高性能表现和非易失性特点，还打造了补充性的内存方案，由于其性能接近内存，傲腾持久内存可作为内存，与真正的 DRAM 内存混搭使用，可将 DRAM 作为其缓存，打造低成本的内存资源池，在单位空间内提供更大的内存密度。

英特尔 SCM 方案走向市场之初，有人觉得这一方案有些多余，但随之而来的是，市场上也出现了许多同类型的解决方案，包括三星的 Z-NAND，还有西部数据的 XL-Flash 都是提供介于 DRAM 内存和 SSD 之间的新的存储层级，弥补两者之间的差距，这也直接反映出此类方案确实有实际需求。

3.5 Z-NAND 与 XL-Flash 没有更新的进展

三星 Z-NAND 又称 Low Latency V-NAND，与铠侠（原东芝）XL-Flash 和 3D Xpoint 一样，也都旨在填补（访问速度和延迟在）DRAM 和 NAND 之间的空白区域。

三星 Z-NAND。2016 年，三星发布第一代 Z-NAND SSD 产品，基于自身改良的 V-NAND 设计和 SLC 芯片，具有独特的电路设计和控制器以加速性能，其页面大小为 2-4 KB，传统 NAND 页面大小为 8-16 KB 有可以快速读写更小的数据块，降低延迟。第二代 Z-NAND SSD 有 SLC 和 MLC 两种模式，容量从 800GB 到 4TB 不等。

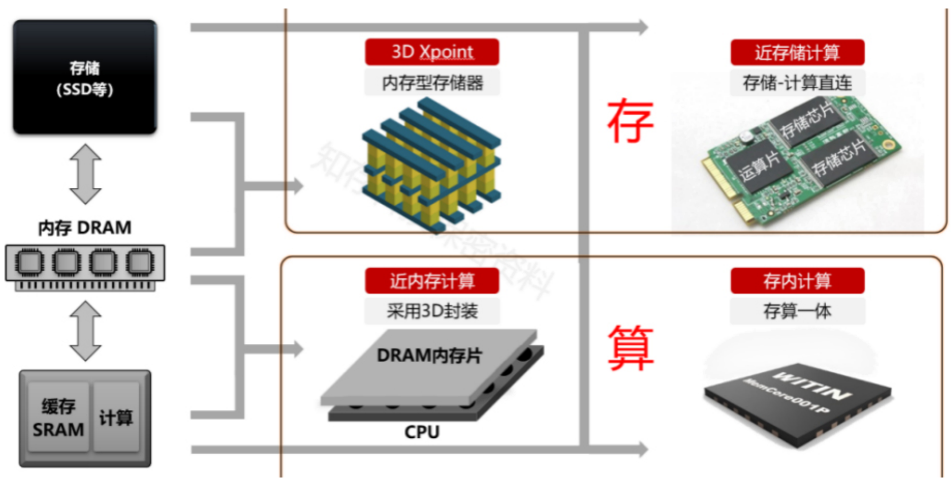
铠侠 XL-Flash。基于 BiCS 架构打造的 SLC NAND 闪存，通过立体堆叠提升存储密度。但与传统 3D 闪存不同，它面向高性能应用，具体体现是高带宽和低延迟。高带宽指更接近 DRAM 内存的读写速度。低延迟则是提升自身 4K 随机读写能力，提高响应速度。

XL-Flash 最初以 SSD 形式发售，但铠侠希望其最终可扩展为 DRAM 总线上的存储设备。

铠侠的 XL-Flash 和三星 Z-NAND 将存储分成更小的阵列块，配置更多的感应放大器（sense amps）支持数据的并行访问，加快响应速度。但 XL-Flash 有比 NAND 更大的裸片，因此价格更昂贵。

3.6 计算型 SSD 和存算一体化

在 AI、大数据、内容分发、数据库、机器学习等新兴应用的推动下，人们开始希望计算能更靠近传统存储设备和系统，于是就有了 Computional SSD（计算型 SSD）与 Processing In Memory（存算一体化），Processing In Memory（存算一体化）受限于内存的限制，两者都是拉近了计算和存储的距离。



Processing In Memory（存算一体化）无法对大量数据进行处理，相对而言，关注度更高的是计算型存储，计算型存储指的是一些数据处理的任务交给存储层来做，而不是把数据挪到主内存里用 host 的 CPU 来处理。

计算型存储背后反映出，人们对于将数据从存储的地方迁移到计算的地方时的考虑，这一过程耗费时间，效率比较低。在 AI 和数据分析领域，数据量都非常大，这时，将数据从产生的地方迁移的话，产生的延迟对于延迟敏感型应用来说是不可接受的。

计算型存储在 SSD 里加入用于处理数据的 ARM、FPGA 或者 ASIC 芯片，可以实现现场数据处理，处理的时效性大为提升。

计算型 SSD 也好，存算一体化也好，要解决的还是数据处理性能的问题，要么解决数据传输的问题，要么就地处理数据，就地处理数据，要么依靠本地计算，要么依靠存算一体的芯片，这也是为什么存算一体芯片也被称为 AI 芯片的原因，在 AI 应用的场景被广泛讨论，可以说，这里要解决的是计算性能的问题，并不是传统意义上的存储。如果以 SCM 为例，它可以被视为高端存储的有机组成部分，被划入存储的范畴。但是与此同时，SCM 并不承担主存储的任务，海量数据仍然要依靠 SSD、HDD。

04

闪存存储系统

4.1 SCM：全闪存存储的重要组成部分

数据的飞速增长以及包括大数据、AI 在内的数据分析应用的发展，极大地消耗着硬件性能，更快速处理器和网络传统对于存储 I/O 已经远远不够，如果为了提升性能而大量采用 DRAM 时，系统成本会非常高。

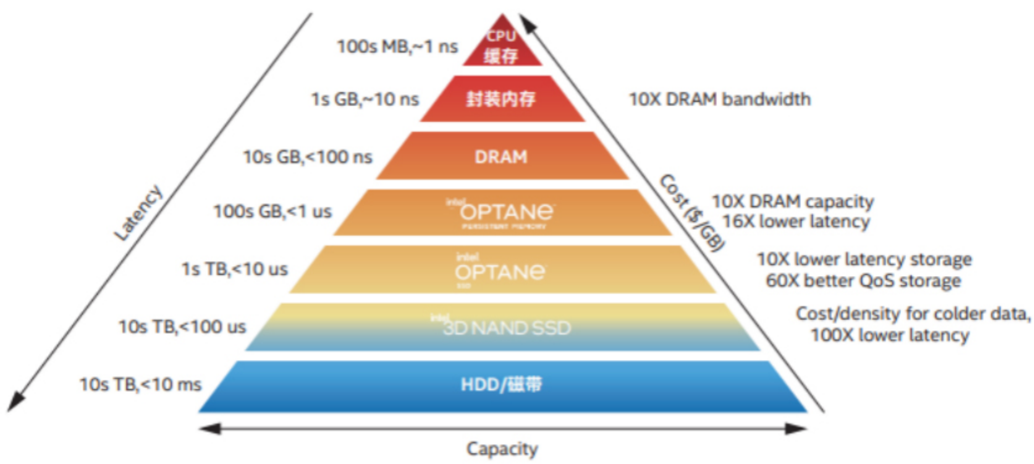
基于 NAND 的全闪存存储主要提供了高性能、高可靠、高效率存储的问题，适用于数据分析、数字成像、VDI、数据库应用、金融交易系统等应用场景，而在需要大量存储容量的场景中，考虑到成本，磁盘仍是主流的选择。

对于全闪存阵列性能的追求永不止步，百尺竿头更进一步，SCM（Storage Class Memory，存储级内存），也称为持久型存储器 (Persistent Memory），其核心作用是解决处理器到存储过程中的性能瓶颈。

SCM 在服务器和存储两方面都有应用，在服务器应用中， SCM 是作为较低成本的内存来使用，在许多内存密集型应用中，如内存数据库，SCM 可以提供更低些的成本。在存储应用中，SCM 用于加快控制器（机头）和后端盘阵之间的性能，SCM 也是作为缓存（Cache）来使用，作为元数据等极热数据的存储层，在性能上会有更高表现。

SCM 的读写速度介于 NAND 闪存和 DRAM 内存之间，成本也是介于 DRAM 和 NAND 闪存之间，时至今日，它最大的价值在于低延迟表现上，而且是在非常高的写入负载压力下，也有非常低的延迟表现，这是 NAND 介质本身完全达不到的。

SCM 是分层存储技术思想的延伸。



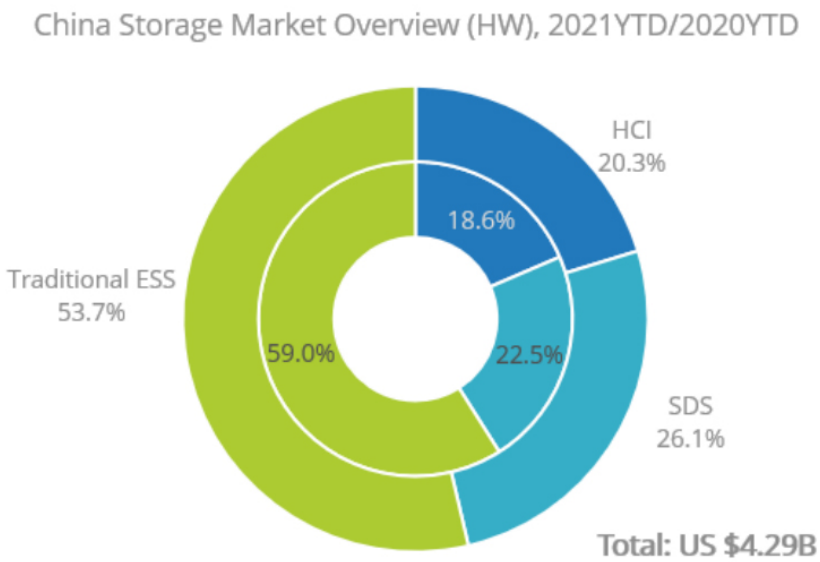
在存储产品的演进中，也是因为 SSD 产品出现，催生了分布式存储技术的诞生，其中，SSD 担负的作用就是热数据存储。随着 SSD 技术的普遍，以及极致存储性能的追求，SCM 很好弥补了 SSD 和内存之间所存在性能差距。

4.2 分布式全闪存阵列

分布式全闪存从概念上很容易理解，分布式存储架构，满配 SSD 的全闪存配置，命名分布式全闪存是为了与全闪存阵列进行区分。

分布式存储也被解读为软件定义存储，但是没有严格定义，没有达成普遍共识。普通的理解是：以 x86 服务器为硬件，借助软件定义的存储资源池，对外提供存储服务，可以是块存储、文件存储，也可以是 NFS 和对象，总之对外提供多协议共享存储服务。但基于 x86 服务器硬件的设定也在被打破，专用定制的硬件设备也已经进入市场，也可以设计为全闪存硬件的配置。（参见：《2022 分布式存储市场调研报告》）

不仅考虑到硬件，控制管理软件也有很多优化的空间。以 Linux 为代表，操作系统主要用于协调硬件和应用软件的资源，也并非针对存储而设计，更没有充分考虑 SSD 的特点，也会带来不必要的损耗。有鉴于此，一种以裸金属 + 软件的存储系统设计应运而生，也可以采用全闪存的配置。



分布式全闪存与分布式存储相伴相生，是分布式存储的一个配置分支。按照 IDC 给出的统计数据表明，2021 年中国存储市场规模为 42.9 亿美元，其中，软件定义存储 14 亿美元，占比 26.1%; 超融合为 12 亿美元，占比 20.3%。按照百易存储研究院《2022 分布式存储市场调研报告》的调研，分布式存储市场规模最高可以达到 1000 亿人民币，从中不难看出，分布式存储已经跨越了市场“戏红人不红”的阶段，迎来市场快速增长期，这完全符合数据市场爆发式增长，以及数据创新应用的需求。

水涨船高，随着分布式存储在行业用户业务中扮演更加重要的角色，性能瓶颈的问题日益凸显，这必将为分布式全闪存营造良好外部市场环境和氛围，前景谨慎乐观、值得预期。

4.3 NVMe-OF

对于 NVMe over Fabrics（NVMe-oF）的讨论由来已久，NVMe-oF 让系统对于 NVM 介质的访问不再局限于使用 PCIe，而是可以在包括 TCP 以太网、FC 网络上以及 RDMA 上实现对于 NVM 的访问，可以在存储与存储之间，存储和服务器间建立更快和更高效的连接，取代原来存储系统当中 iSCSI 的地位。

最重要的是，NVMe-oF 提供了额外的数据中心和网络整合。NVMe-oF 能让外部存储的延迟直逼直连存储的延迟，使整个体系的并行性能大为提高，避免了系统 IO 瓶颈。

NVMe-oF 对于新一代闪存存储系统的设计至也关重要，现有的大多数全闪存阵列架构本质上还是基于磁盘系统来设计的，它并不能发挥 NVMe SSD 的全部性能，随着全闪存对 NVMe 的持续挖掘，NVMe 的全闪存存储系统将带来极高的性能提升，NVMe 是存储系统厂商竞争的一大要点。

目前而言，NVMe-oF 无论是标准还是存储厂商的应用实现都处于发展的早期阶段，只有少部分厂商有在 NVMe-oF 方面进行探索，市场上对于 NVMe-oF 的期待很高，但对于 NVMe-oF 这种攻坚可能求取极致性能的做法业内其实也有争议。

Infinidat、StorONE 和 VAST Data 纷纷指出即使用上了 NVMe-oF，对于全闪阵列发展的帮助也会非常有限，不看好 NVMe-oF 全闪。这三家公司的技术理念各有不同，对如何经济高效地满足存储容量和性能的需求，各家反应也不尽相同。

Infinidat 认为 PB 级容量全闪存阵列造价过高，于是，主张用近线硬盘做容量层存储，用缓存引擎将数据预取到内存里，在两者间加入 SSD 作为中间层，在大量数据存储场景下，也能保持不亚于全闪存的性能表现。

StorONE 的创新主要集中在软件架构上，能利用包括磁盘，SSD 以及 3D Xpoint 在内的各种存储介质，最终将大容量与高性能两个特性进行结合。

VAST Data 的产品以高效地数据压缩和纠删码技术见长。这种能够降低数据写入到 SSD 次数的方式意味着其阵列成本和耐用性比传统全闪存阵列要高。这样用户就能把大量的数据存储在闪存上，以磁盘的价格享受闪存的性能。

截止目前为止，这三家的发展思路还没有对主流的存储厂商带来什么影响，主流的存储厂商没有做出相应的反应。这也反映出闪存市场的多元化。

4.4 40+w SSD 的新时代

“不在沉默中爆发，就在沉默中灭亡”，这句话出自《记念刘和珍君》一文，鲁迅先生要表达的是对社会现状的选择。我们引用这句话，是想提醒大家：闪存市场看上去波澜不惊，实际上已是暗流涌动。对此，要么视而不见，沉默中被动；要么未雨绸缪，运筹帷幄，积极应对，等待新一轮爆发时变革的到来。

先生说的是人，我们说的是闪存技术变化。

最大的变化来自 SSD 外形规格的变化。EDSFF 提供了一系列新的外形尺寸动态，另起炉灶，将在容量、可扩展性、性能、可维护性、可管理性、散热和电源管理方面带来一系列变化。

简单说，SSD 将突破现有 2.5 英寸、3.5 英寸的变化，另起炉灶。实际上，从磁盘

到闪存、从 AHCI 到 NVMe、从 SAS/SATA 到 U.2，从介质、协议、接口，盘的变化一直在持续演进中，如今轮到了外观规格尺寸。

规格尺寸变了意味着什么?

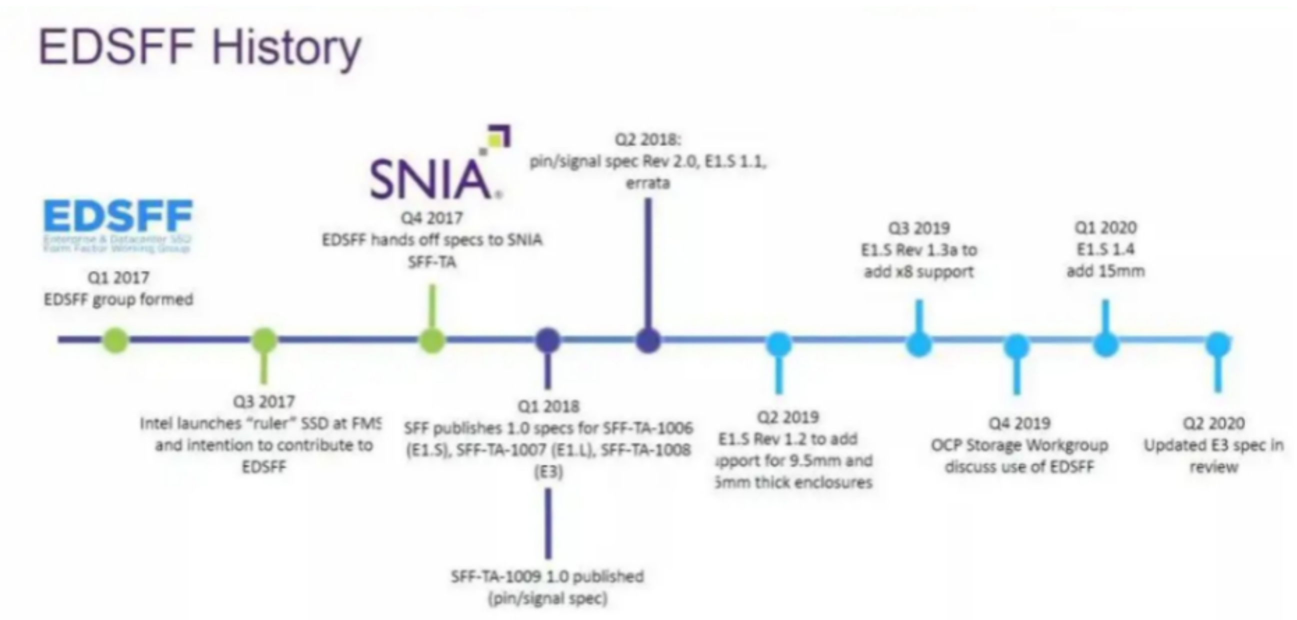
所谓量身定制，相比 2.5 英寸、3.5 英寸，新的 EDSFF 规格，更能够发挥 NAND 闪存介质的能力。

以功耗限制为例，原有极限功耗是 25w，这个天花板的存在，就决定了 SSD 的性能和容量。如今，EDSFF 将功耗提高到 40w、70w，如虎添翼。

就如同装修，EDSFF 规格如同“全屋定制 “，必然会更贴心，更温暖。

但是就像亚马逊森林蝴蝶的翅膀，小小的 EDSFF 规格定制，必然会掀起存储的滔天巨浪。从供电到散热，服务器需要全新的设计，期待着新一代服务器的全新亮相。

EDSFF 概念：EDSFF（Enterprise & Data Center SSD Form Factor，企业和数据中心固态硬盘规格），基于英特尔 Ruler SSD 标准，由 15 家公司共同制定，旨在解决数据中心存储问题。现在由 SNIA 作为 SFF 技术附属技术工作组（SFF TA TWG）的一部分进行维护。

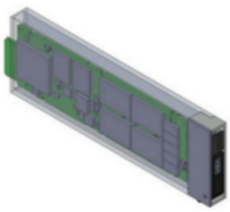


如今，EDSFF 分为适配 U1 服务器的 E1 和适配 U2 服务器的 E3 两种外形尺寸。其中，E1 又分为 E1.L 和 E1.S，其 L 代表“long，长“；S 代表 “short，短”。

E1.L 有 9.5mm 或 18mm 两种宽度，分别支持高达 25W 或 40W 的功率。

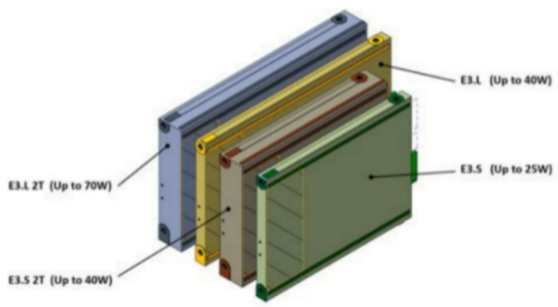
Type	Width	Length	Thickness
E1.L 9.5mm	up to 25W - 38.4mm	318.75mm	9.5mm
E1.L 18mm	up to 40W - 38.4mm	318.75mm	18mm

Type	Width	Length	Thickness
E1.S 5.9mm	31.5mm	111.49mm	5.9mm
E1.S 8mm heat spreader	31.5mm	111.49mm	8.01mm
E1.S Symmetric Enclosure	33.75mm	118.75mm	9.5mm
E1.S Asymmetric Enclosure	33.75mm	118.75mm	15mm
E1.S Asymmetric Enclosure	33.75mm	118.75mm	25mm



Type	Height	Length	Width	Max Power
E3.S	76mm	112.75mm	7.5mm	25W
E3.S 2T	76mm	112.75mm	16.8mm	40W
E3.L	76mm	142.2mm	7.5mm	40W
E3.L 2T	76mm	142.2mm	16.8mm	70W

EDSFF E3.S and E3.L



E1.S 有 5.9mm、8.01mm、9.5mm、15mm 和 25mm 等 5 种规格设计。相比 E1.L，E1.S 是用宽度、厚度来取代长度。后三种规格可以满足 PCIe 5 SSD 散热的需求。EDSFF E3：和 EDSFF E1 一样，E3 也有 E3.L 和 E3.S 两种规格，因为厚度不同，有分为 1T 单宽、2T 双宽 2 种规格。

现有的实践案例中，Microsoft Azure 在其 Olympus FX-16 1U JBOF 中，使用了 16 个 18mm E1.L SSD，每个 SSD 16TB(总共 256TB)，前面板上也有配有 PCIe 接口，对外连接 9.5mm E1.L SSD，其存储密度提升了 2 倍，但前提是解决好有关功耗和散热等问题。

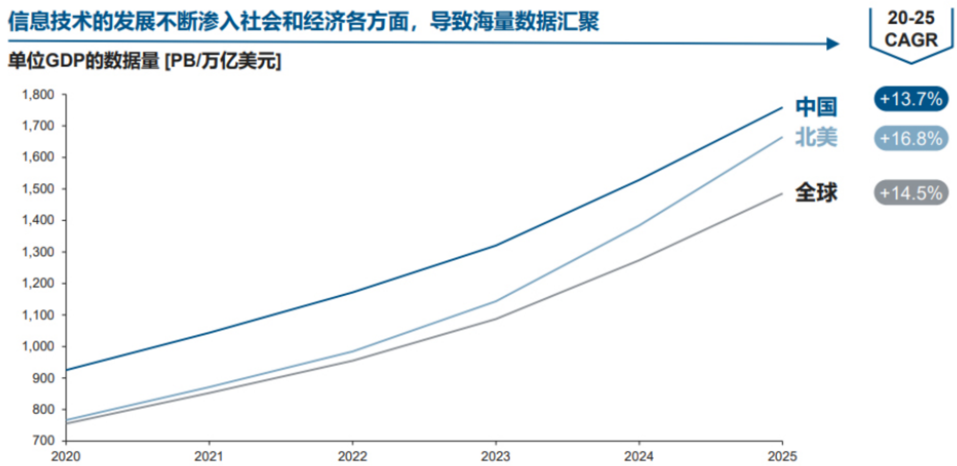
有预测显示：2025 年 EDSFF 将占总服务器容量的一半左右。

山雨欲来风满楼，一个崭新的新时代即将到来！

4.5 数据存力和闪存

数据存力是今年 4 月 20 日，华为、罗兰贝格联合撰写的《数据存力，高质量发展的数字基石》白皮书所提出的核心概念。

作为一个重要的衡量指标，数据存力揭示了数据存储和国民经济发展之间的关系，通过数据存力的研究，揭示了未来促进中国经济持续发展的办法，其作用和重要性不言而喻。



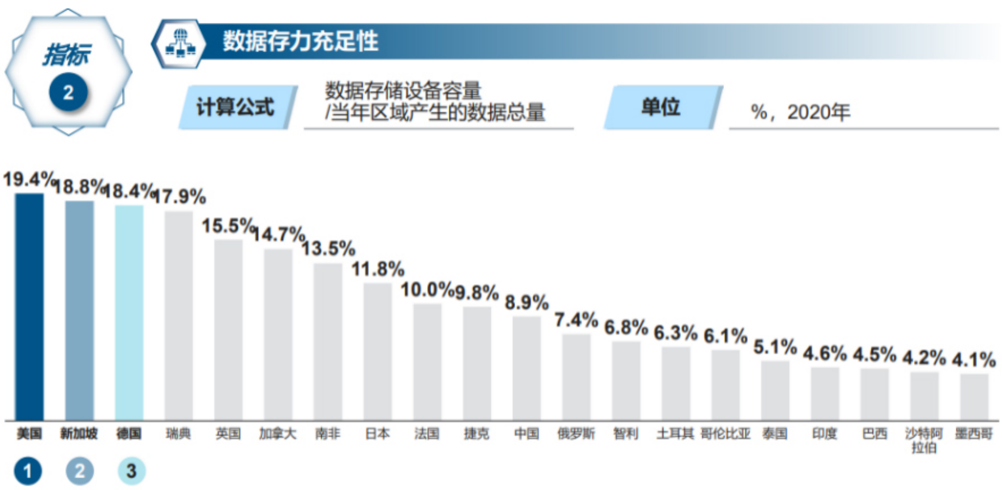
资料来源：IDC；国际货币基金组织

这张图揭示了中国经济和数据增长的正比关系，中国数据增长领先北美和全球。

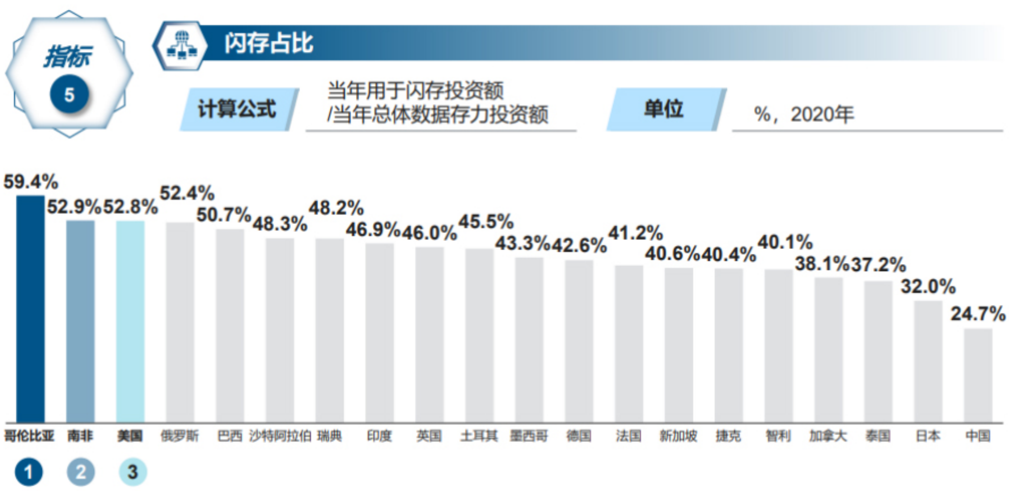
但在数据存力的指标方面，中国与美国相比差距明显。简单说，我们有全球第一数据，但是缺乏数据存储和处理的能力。

数据存力的差距会成为制约中国经济发展的首要因素，这个结论应该不存在争议，因为这是由数据经济的本质来决定的，对于中国来说，提高数据经济的能力和水平是核心的问题，而数据存力会是一个可供量化的衡量指标。原因也很简单，数据经济的核心技术支撑有两个：一个是存储能力，一个是数据处理能力。

数据存力就是数据存储、数据处理的量化衡量指标。



资料来源：罗兰贝格分析



资料来源：罗兰贝格分析

闪存占比是衡量数据处理能力的重要指标，中国差距更加明显。

作为数据存力的核心指标，闪存占比至关重要，应该成为衡量各个省市、地方政府政绩发展的技术指标。所谓“要想富，先修路”，这与数据存力、闪存占比是异曲同工。

4.6 数据备份拥抱全闪存阵列

如今，数据备份也开始拥抱全闪存阵列，“昔日王谢堂前燕，费用寻常百姓家“，可谓沧海桑田，日新月异。有性能需求的场景，就有全闪存阵列， 数据备份也不例外，没有什么好奇怪的。如今数据爆炸式增长，核心业务系统性能提高，容灾备份水涨船高是自然而然的事情，闪存价格开始走低也发挥了推波助澜的作用。

尽管如此，全闪存的价格还是偏高，用户需要一定的承受力。

数据备份时间窗口不足的问题早就有所体现，由此也催生了快照、增量技术的产生，让问题得到环节，如今，核心业务系统数据量激增，备份时间窗口压力加大，选择全闪存阵列缓解压力是水到渠成的事情。

重视容灾备份，但是眼光不应仅限于此。容灾备份所牵涉的更是数据分析应用的大问题。

数据应用例来划分为 OLTP 和 OLAP，前者往往是关键业务系统，涉及关键的业务交易，对稳定性、可靠性和可用性有极高的需求，不容任何闪失；但是后者往往是辅助决策、保障系统，尽管非常重要，但并非不可或缺。

有些用户没有使用备份系统，仅仅存储阵列自身的数据保护能力，不是他们不了解数据备份，而是他们结合自身的现状做出的不得已选择。可以说，能够使用全闪存做为备份存储的用户，还是比较有实力的用户。

OLAP 的应用往往不会直接使用生产数据，防止给生产系统带来干扰。此前，OLAP 常常通过数据复制来满足研发、测试以及数据分析的需求。但是这些复制数据

往往缺乏行之有效的集中管理，在安全性、存储的效率都存在很多的问题，此外，复制的数据实时性缺失，会导致研发、测试或者数据分析失准，不能准确反映业务发展的现状。

针对容灾、OLAP 等数据缺乏集中有效管理的现状，第二存储应运而生，以 CDM 为核心，为研发、测试、容灾、数据分析等应用提供集中数据保障，在确保安全的基础上，提高水平的数据存储管理和服务。

如今，很多创新业务应用完全建立在数据实时分析、处理的基础上，OLAP 不仅不可或缺，同时要求性能和效率作为保障，在这样的背景下，全闪存阵列将大有用武之地，这也是超越全闪存阵列数据备份的额外话题。如今容灾备份不再孤独，而与第二存储相伴相生，焦不离孟，孟不离焦，这是目前用户应有的视野和认知。

4.7 高性能计算和 AI 存储方案

高性能计算被喻为科技皇冠中的明珠，被视为衡量一个国家科技水平高低的重要指标。高性能计算系统也是一个相对封闭的市场，大的系统项目常常由品牌供应商把持。

但是如今，高性能计算市场也呈现出来了存算分离的趋势，计算集群、高性能存储采购，以增加适应性和灵活性。

在高性能存储的设计上，以并行文件系统 + 全闪存阵列的设计方式，目前受到市场的关注和欢迎，从而为全闪存阵列应用开辟了新的战场。

在 AI 应用方面，以 Colossal-AI 为代表，AI 算法已经突破 180 亿参数的规模的，尽管 Colossal-AI 让用户可以利用消费级显卡（GPU）多维并行操作，但涉及到 GPU 之间的数据交换，其性能仍然没有办法满足需求。AI 大模型训练，导致对于算力提升非常快，对存储 IO 性能的需求水涨船高，对高性能 SSD 需求不断提升。

如今 GPU 不低，导致 AI 训练成本居高不小，如果能够提高计算的效率，对于控制成本、提高性能的意义重大，从技术上说，无论 CPU、GPU、DPU、SSD 在设计上

都有大量缓存加持，基于缓存的数据一致性访问就成为了新的诉求，如今，通过 Genz 的并入，CXL 标准备受关注，有望成为数据访问新的方式，技术的发展也会被包括闪存 SSD 设计在内的一系列的变化。

技术的改进和更新必将给市场、应用带来的新的变化。山雨欲来风满楼， CXL、PCIe 5 扑面而来，对此应有足够重视和关注。

05

总结概述 结束语

无论是从 IDC 以及 Gartner 的数据，还是与许多企业用户接触的过程中，都能看到闪存市场的勃勃生机；闪存存储的市场份额仍保持高速增长，在存储门类当中，闪存存储是增长最快的，闪存已经是主流存储的选择。

下一阶段，介质层面上，随着 QLC 和 PLC 的发展，随着更高层的 NAND 的研发，闪存在成本上有更大的优化空间，虽然演进的相对速度会慢一些，但在技术发展以及市场需求推动下，这一趋势不可逆转。

与之对应的，闪存存储控制芯片，闪存存储设备以及闪存存储系统都会随着介质演进迎来新的发展空间，如何在保证性能的前提下，安全可靠用上更具竞争力的介质，而且不断丰富功能特性，用上 NVMe 之类的技术，将会是所有存储厂商的必答题。

在企业级市场，2020 年，随着新基建带来的数据中心投资热潮，国内数据中心投资增长，为企业级 SSD 的市场发展提供了广阔的发展空间。在消费级市场，SSD 在大范围替代磁盘存储，各种智能设备的采用将进一步刺激闪存在嵌入式设备市场中的发展。

06

附录，企业案例推荐

RD8000

极致性能 | 冗余设计 | 安全可靠

柏科介绍

柏科数据技术（深圳）股份有限公司（以下简称“柏科数据”)，于 2007 年 5 月成立，是国家高新技术企业。通过了 ISO9001 质量管理体系等近十项专业认证，并拥有省级工程技术研究中心和院士专家工作站等产学研用合作平台。公司以“Bring Life to Data”为核心理念，以用户至上为宗旨，致力于在数据全生命周期管理中为客户数字化业务提供全面产品、解决方案及专业服务。公司重视技术的自主研发，投入资金已超亿元，获有百余项软著及专利授权。

近年来，公司把国产化和自主可控作为发展战略方向，专设自主可控事业部，推进全系列产品国产化工作。从关键部件到操作系统均采用国产化生态配套产品，推出了集中式统一存储、全闪存存储等多款自研产品。公司已有多个产品通过国家信创产品入围测试。未来，柏科数据将继续加大资金投入，助力我国信创发展。

柏科简介

15年成功经验
2007 年公司成立，专注于存储、备份容灾、融合计算领域

3000+用户
服务国内外科研院所，政府和医疗等行业用户

8大分支机构
北京、上海、深圳、重庆
长沙、杭州、武汉、合肥

3大研发中心
杭州、深圳、武汉

200+员工
核心技术人员超过 100 人
市级以上技术领军 2 人

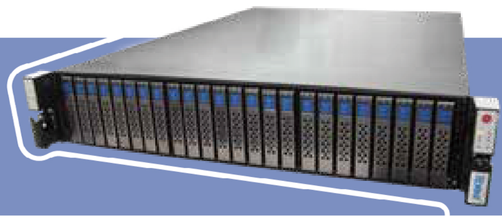
20+ 体系认证

50+ 发明专利

110+ 软件著作权

RORKE

柏科数据RD8000是面向企业级市场，为满足超高性能需求而推出的闪存阵列。RD8000采用优化的闪存存储架构设计，配合领先的缓存管理和读写I/O调度算法，提供远超传统阵列的读写性能。从而满足高性能计算、虚拟化、非线编、大型数据库等核心业务对性能的需求。



产品特点

灵活应用

多样性的选择，优化现有业务的特定需求

绿色节能

先进的节能技术，全面降低业务运营成本

便捷管理

图形化管理工具，提高企业运维效率

自动化服务质量

根据业务需要控制应用程序优先级

稳定可靠

企业级闪存技术，确保业务的连续性

卓越性能

低延迟、高 IOPS，满足核心业务对性能的极致需求

全面保护

多种保护功能，避免核心数据丢失的潜在风险

产品规格

产品型号	RD 8100	RD 8300	RD 8500
产品定位	全闪（SAS SSD）	全闪（SAS SSD）	全闪（NVMe SSD）
产品结构	双控制器		
处理器	双节点四路处理器		
缓存	128GB	256GB-1TB	512GB-1TB
NAS 功能	支持		
存储协议	FC、iSCSI、NFS、CIFS		
前端端口类型	8/16Gb FC、1/10GE	8/16/32Gb FC、1/10/25GE	8/16/32Gb FC、1/10/25GE
后端端口类型	4 个 SAS3.0 物理复合端口共支持 16*12Gbps 扩展		
RAID 级别	0、1、5、6、10、50、60		
资源效率提升	控制器负载均衡、空间精简、在线数据压缩、在线数据重删、智能卷组统一管理、IO 资源智能动态分配、传输带宽智能动态分配		
数据保护功能	智能快照、智能克隆、远程复制、智能 HA 高可用		

MacroSANMach系列存储

端到端超高性能 宏杉科技NVMe全闪存存储解决方案

随着互联网金融、数字银行、自动驾驶、高端制造等新型应用的不断涌现，数据量呈现爆发式的增长，需要性能卓越的存储系统为新型应用持续增加的算力需求提供数据存储支撑，这对存储产品的性能及创新提出了更高的要求。

MacroSAN Mach系列是宏杉科技推出的基于NVMe协议的全闪存阵列，采用NVMe SSD作为存储介质。Mach系列产品针对NVMe协议采用了全新的硬件架构和软件算法，全方位地释放闪存颗粒的性能。

宏杉科技MS7000G2-Mach

极致可靠的性能

完全针对闪存的传输协议

Mach的闪存介质与CPU采用NVMe协议进行数据交换，使用PCIe总线直接进行通信，解决通道瓶颈问题，不再有繁杂的协议转换，大大提高了传输效率和延迟。

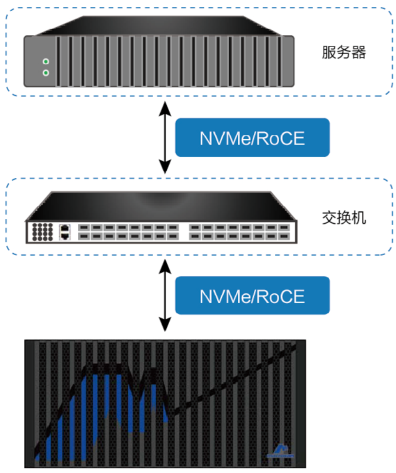
矩阵式全交换硬盘通道设计

MS7000G2-Mach系列产品双控可插50块NVMe SSD，每块NVMe SSD全部配置2个x2PCI-e3.0的接口，分别和两个控制器直连进行数据交换，每块SSD均有自己的独立数据通道，不需多块硬盘复用数据通道。MS7000G2-Mach系列产品双控制器可提供1600Gbps的后端总带宽，可以真正做到释放后端闪存介质的性能，解决后端链路瓶颈。

端到端的NVMe架构

宏杉Mach系列产品是针对NVMe协议量身打造的极致性能全闪存产品，其前端网络支持业界领先的NVMe over Fabric技术，硬盘通道设计上采用矩阵式全交换架构，有效降低I/O处理时延，充分释放闪存性能。

宏杉Mach系列产品实现了端到端的NVMe架构，可满足金融、交通、政府等行业的核心业务对于极致性能、超低时延的要求。



MS7000G2-Mach

产品规格

产品型号	MS7770G2-Mach	MS7370G2-Mach
控制器扩展	支持2-48个控制器横向扩展	
缓存（每双控）	3TB	
闪存盘配置硬盘数（每双控）	50	
最大主机接入数	65536	
前端端口类型	8/16/32Gb/s FC、10/25/40/100Gb/s iSCSI、16/32Gb NVMe over FC、25/100GbNVMe over RoCE	
RAID级别及热备特性	RAID/CRAID（CRAID3.0）0、1、3、4、5、6、10、50、60、X0等，支持专用热备、全局热备、空闲硬盘热备	
操作系统支持	AIX、HP-UX、Solaris、Windows、Linux等	
虚拟化平台支持	VMware、Citrix、Hyper-V、OpenStack、KVM、XEN等	
基础管理软件	MacroSAN管理套件，含基本存储管理、RAID、系统监控、日志及告警等功能	
高级特性	自动精简配置、智能分层存储、无中断数据迁移（NDM）、性能监控、数据快照、数据复制、数据 镜像、对称双活、存储异构虚拟化、服务质量控制（QoS）、多租户、在线全局重删、在线压缩、云网盘等	
NAS特性	支持CIFS、NFS、HTTP、FTP等协议	

超高性能NVMe端到端全闪存方案

宏杉科技基于MS7000G2-Mach全闪存存储、无损以太网交换机，构建了端到端超高性能NVMe全闪存存储解决方案，搭建高性能、高可靠、易管理的数据中心架构，提供基于iNOF的端到端NVMe over RoCE存储解决方案，实现以太网存储的零丢包，高性能，秒切换，易配置。为数据中心核心业务提供高性能、低延时的端到端NVMe存储平台，实现对FC的替代，助力用户构建高性能全IP数据中心。

宏杉基于NVMe over Fabric的网络控制优化技术iNOF进行自主研发，在业界率先实现NVMe-oF环境下的存储双活、存储多路径访问等技术，实现关键业务数据跨数据中心的实时保护，数据零丢失，保障核心生产业务实时在线，为MS7000G2-Mach系列存储赋予了“即插即用”、“故障快速感知”等类FC SAN高级特性：

即插即用

当服务器与存储设备接入RoCE-SAN网络时，可实现网络自动发现设备，并通知服务器自动建立与存储设备的连接关系，在开局、扩容和维护（故障更换）的场景，提供便捷高效的运维体验，提升运维效率。

故障快速感知

面临网络故障时，交换机可快速检测故障状态，并及时传递给相关业务服务器，以便业务快速切换到冗余路径，从而实现在网络故障时业务秒级切换，而用户无感知，可有效应对链路故障、PFC（基于优先级的流量控制）风暴等场景。

MS7000G2-Mach系列存储已与主流的操作系统、服务器、无损以太网交换机完成兼容性适配，能够从应用端到存储端打造全NVMe的全闪存存储系统，保障整个数据中心的稳定运行，并提供高性能的数据存储服务，满足用户多样性的业务需求。

杭州宏杉科技股份有限公司 MacroSAN Technologies Co., Ltd.
地址：杭州市滨江区阡陌路482号智慧e谷大厦A座10-12层 Tel: 400-650-5527 Fax: 0571-28182001

芯盛智能，固态存储中国芯

存储作为数据的载体，是基础设施的重要组成部分，实现存储的自主建设是构建现代化基础设施体系的前提和必要条件。芯盛智能科技有限公司作为国家集成电路产业投资基金注资的数据存储与安全的集成电路设计企业。芯盛智能致力于固态存储技术的自主研发，倡导“固态存储中国芯”理念，持续走在技术和市场前沿。

芯盛智能现有员工 500 余人，其中 70% 以上为研发人员，在北京、上海、成都、济南、长沙、常州等地设有分子公司及研发中心。公司自成立以来，始终坚持自主创新理念，推出多款存储控制器芯片、安全加密芯片，固态存储产品及数据安全解决方案，产品覆盖数据中心、边缘中心、工业控制、消费类终端和车载电子等多各行业，广泛用于党政、金融、电力、轨道交通、平安城市等领域。

公司持续加大研发投入，承担了多项省、市、区科研项目，形成多项拥有自主知识产权的关键核心技术，被评为高新技术企业、专精特新企业，省级民营科技企业和年度质量信用 A 级企业等。

芯盛智能秉持“一体双翼，双驱动”的市场战略，即公司集“芯片、固件、供应链、解决方案”为一体，在技术创新、与市场应用创新的双驱动下，为中国数字经济下数据的存储和安全保驾护航。

芯盛智能凭借雄厚的技术研发实力，在 2022 全球闪存峰会上，发布了全球首款基于 RISC-V 架构的 12nm 双模高端固态存储控制器芯片——行者。

行者芯片使用了 RISC-V 开源架构进行设计研发，采用 12nm 制程工艺，其中 XT8210 系列，具备 PCIe4.0x4 高速接口，支持 NVMe1.4 协议，顺序读写速度高达 7000/6000MBps，4K 随机读写高达 1000k/900kIOPS，最大可支持 16TB 容量；XT8210 内置芯盛智能可靠性引擎和安全引擎，通过国密局商用密码二级认证和中国信息安全测评中心 EAL4 安全认证。XT8210 不仅拥有高性能还同时兼顾高可靠和高安全的特性，性能领先于国际同类产品，它的发布实现了 PCIe 的国内全自主模式，填补了国内空白。另外，行者芯片也具备 SATA 3.0 企业级高稳定接口，此型号为 XT6130。XT6130 同样为 12nm 制程工艺，顺序读写为 560/520MBps，4K 随机读写高达 100k/90kIOPS，最大支持 8TB 容量；XT6130 同样内置芯盛智能可靠性引擎和安全引擎，通过国密局商用密码二级认证和中国信息安全测评中心 EAL4 安全认证。XT6130 的诞生打破了国产企业级市场被寡头垄断的现象，为国产企业级存储拿下一席之地。行者芯片的发布，彰显芯盛智能为实现固态存储 100% 国产化而持之以恒踏步向前的信心和决心。

与此同时，芯盛智能还发布了基于行者芯片的企业级 SATA3.0 固态硬盘 SS3000 系列，其采用长江存储 Xtacking® 3D TLC 颗粒，顺序读写速率为 560/520MBps，4K 随机读写高达 95000/40000IOPS，具备高安全、高可靠、高性能、全国产等特性。

另外，芯盛智能PCIe4.0接口的企业级固态硬盘SP3000系列预计今年第四季全面上市。SP3000 系列搭载行者芯片 PCIe4.0 控制器芯片，采用长江存储 Xtacking® 3D TLC 高可靠颗粒，U.2 15mm 的外观形态，顺序读写高达 7000/5000MBps，4K 随机读写为 900k/60k IOPS，性能领先国内同类产品；另外 SP3000 支持 PLP、E2E、RAID+ 等企业级特性。

企业级SATA SSD-SS3000系列

高安全　高可靠　全国产

产品与型号	SS3000系列					
	SS3000S327				SS3000S3N6	
容量GB	480	960	1920	3840	480	960
形态	2.5''（7mmT）				M.2(2280)	
接口	SATA 6.0Gbps					
颗粒	KIOKIA或YMTC 3D TLC					
128KB顺序读	最大560MB/s					
128KB顺序写	最大520MB/s					
4K随机读(稳态)	最大95K IOPS					
4K随机写(稳态)	最大40K IOPS					
功耗	最大功耗≤4.0W;空闲功耗≤1.0W					
MTBF	≥200万小时@55°C					
UBER	≤10^{-17}					
DWPD@5年	≥2.5					
温度	工作温度：0°C~70°C；存储温度：-40°C~+85°C					
振动	3.13Grms(5Hz~800Hz)					
冲击	1000G at 0.5ms, 3 axis					
特性	增强PLP、E2E、RAID+、ECC、4K LDPC、高温保护、TRIM、在线升级、SMART、AES256、SM2/3/4等					

除了企业级产品，芯盛智能还有桌面级产品、工业级产品、安全级产品、嵌入式产品以及 CUBEXStor 电商、渠道类产品。

以 EP2000 为代表的桌面级产品，搭载芯盛智能自研主控 XT8111，采用 3D TLC 颗粒，具备 PCIe3.0x4 高速接口，顺序读写最高 2500/2100MBps，具有 256GB、512GB、1TB 三种容量，由于性价比较高，主要应用于党政办公及信创通用类 PC 机、笔记本等。

以 DP3000 为代表的工业级产品，搭载行者芯片的 PCIe4.0 控制器芯片 XT8210，采用 3D TLC 颗粒，U.2 15mm 和 M.2 2280 形态，灵活满足工业市场的应用需求；DP3000 顺序读写为 7200/7000MBps，工作温度可满足 -40° C ~ +85° C，支持 PLP、E2E 及定制化服务，主要应用于工业计算机、工业自动化一体机、工业存储设备等。

以 MP2500 为代表的安全加密产品，搭载芯盛智能自研 XT8111V00 主控，采

用 3D TLC 颗粒，具备 PCIe3.0x4 高速接口，顺序读写最高 2100/1600MBps，具有 256GB、512GB、1TB 三种容量，通过商用密码 SM4、国密一级认证，全国产化平台兼容，主要应用于加密一体机、网安及重点加密场景。

以 E110 为代表的嵌入式产品线系列，精选长江存储原厂 NAND Flash，采用全国产化设计，BAG 封装，内置自主知识产权 LDPC 纠错算法，在保证国产安全的同时读写速度高达 310/247MBps，容量从 4GB ~ 128GB，满足不同用户的多样化需求，广泛应用于党政办公、金融、轨道交通、水利等各行各业。

芯盛智能的技术优势不仅体现在芯片的创新性和固态存储产品的差异性上，还在推动行业标准的编制、产业白皮书发展的修订中。在本次 2022 全球闪存峰会上，中国计算机行业协会携手芯盛智能等国内头部固态存储企业及华中科技大学、山东大学、北京理工大学等知名院校联合撰写的《全球闪存存储技术及产业发展研究报告》重磅发布，此报告作为闪存存储领域最权威的报告，一经发布便得到与会人员的青睐。不仅如此，芯盛智能还参与国标级《信息技术 固态盘分级技术规范》的编写工作，致力于推动国产存储自主生态向高水平、绿色、安全和健康的状态发展。

宝存科技

上海宝存信息科技有限公司 (Shannon Systems）是一家致力于开发高性能企业级 SSD 固态存储解决方案的高新技术公司，2011 年由硅谷资深技术团队回国创办。公司自成立以来，提供的标准化 NVMe SSD 和定制化 Open Channel SSD 企业级固态存储产品具有高性能的 IOPS 和更加稳定优异的 QOS 表现，持续帮助企业用户优化 IT 系统架构和性能，降低采购成本和维护费用，已经在互联网、电商、运营商、能源企业等行业领域超多一线大厂企业客户中广泛应用。

2015 年，公司被全球较大的 NAND 闪存主控芯片供应商慧荣科技 (Silicon Motion Technology Corp.,NasdaqGS:SIMO) 全资收购，加速了 Shannon Systems 企业级产品自主研发和创新的发展道路。未来，宝存科技企业级 SSD 解决产品方案会紧追适配服务器规模部署的节奏，通过不断的探索提升产品性能和品质以满足消费者的更高需求。

研发及销售技术支持中心分布

研发中心：上海、杭州、苏州、台北、新竹

销售与技术支持：北京、上海、广州、深圳、杭州、苏州

行业领先历程

2014 美国全球闪存峰会上推出业界首款单卡容量 6.4TB PCIe SSD

2016 中国互联网技术大会发布业界首款单卡容量 12.8TB PCIe SSD

2018 发布业界首款商用企业级 Open Channel SSD

2021 推出了新一代 Zoned NameSpace SSD (ZNS SSD)；发布了传输速度更快的企业级 PCIE NVMe Gen4 SSD 产品—SP4E/SP4X 系列

核心优势

稳定长期的主控供应：全面采用 SMI 主控

NAND 适配及迭代管理经验丰富：NAND Flash 的使能和管理技术具有扎实的成功案例

一线大厂规模部署经验：SSD 解决方案已被大厂企业客户采用超过十年

全套 SSD 解决方案提供：主控，FW 到 QA 一站解决

应用定制化领域的成功案例：满足开发产品各方面的高要求

可提供弹性化的本土服务：依据客户需求规划解决方案

产品简介

NVMe SSD（Gen4/5 Series）

数据高可靠性：端到端的数据安全保护；MTBF 超过 2 百万小时；内部多条带 RAID

遵循行业规范：NVMe 1.4 标准；灵活的功耗管理；可配置扇区字节数

多平台兼容：x86（Intel、AMD、兆芯、海光）；ARM（鲲鹏、飞腾）；MIPS（龙芯）

Shannon SP4 Series NVMe SSDs

高带宽，低延迟，独有智能算法；端到端数据保护机制，兼容最新 OCP spec Telemetry 标准

PCIe Gen4 x4

容　量：SP4E（1.92/3.84/7.68TB）SP4X（1.6/3.2/6.4TB）

顺序读 / 写：高达 6,300/3,500 MB/s

随机读 / 写 IOPS：高达 900/340 K

7:3 R/W 混合 IOPS：500K+

4KB 最大稳态延迟 <10ms

闪存支持：3D eTLC

Shannon SP5 Series NVMe SSDs

卓越的 QoS 控制，领先的控制器架构，可在 NVMe 标准和 NAND 上轻松过渡；具有竞争力的质量、稳定性和可靠性

PCIe Gen5 x4

容　量：SP5-E（3.2/6.4/12.8/25.6TB）SP5-X（3.84/7.68/15.36/30.72TB）

丰富的功能集：NVMe 2.0 和 OCP Cloud SSD Spec 2.0

顺序读：高达 14GB/s

随机读：高达 3M IOPS

闪存支持：3D eTLC 多达 6 个平面

时间计划：Q1/2023; CS:Q3/2023

Open Channel SSD

NameSpace & IO 隔离：保障不同逻辑卷 QoS；灵活管理设备和分配性能；方便易用的命令行配置工具

原子写：保障数据一致性；降低应用响应延迟；延长闪存寿命多达一倍

用户态 OCS 合作开发简化堆栈：缩短存储路径；基于 OCS 1.2 接口规范

Zoned Namespaces SSD

提升寿命：降低写入放大率；提高 SSD 的写入耐久度

降低成本：SSD 寿命提升，并支持 QLC NAND，有效降低成本

降低延迟：通过 ZONE 的隔离，可以显著的降低写延迟，提高整体性能

标准化接口：可实现强大的软件和硬件生态

DapuStor 大普微

深圳大普微电子科技有限公司（DapuStor）成立于 2016 年，是国内领先的 SSD 主控芯片设计和智能企业级 SSD 定制专家；拥有国际一流的研发实力，超过 200 名团队成员，具备从芯片设计到存储产品量产交付的全栈能力。 公司已累计获得十数亿元融资，旗下企业级智能固态硬盘、数据存储处理器芯片及边缘计算相关产品，已广泛用于国内外主流服务器、运营商、互联网数据中心市场。

DapuStor 作为高新技术企业，率先于业内提出存储“DPU”概念，以推动中国“存算一体”与“智能存储”产业发展为己任。公司拥有全球先进的核心技术，已申请国内外发明专利超过 200 项。

Q1 200+
专利技术
主要技术 :PIS(Processing In Storage)，存储固件，存储芯片，智能存储 ...

Q2 四大
研发中心
深圳、苏州（芯片）、成都、南京

Q3 五大
产品线
海神系列、嵘神系列、Xlenstor 系列、蛟容系列、DPU 存储芯片系列

Q4 240+
团队成员
本科 54%　硕士 36%　博士 2%

DapuStor产品组合及优势

SCM （Storage Class Memory）SSD — Xlenstor系列
全球领先 低延迟 高耐用

PCIe Gen4 企业级主控芯片 — DPU600
国内首家 成功量产

PCIe Gen4 NVMe 企业级SSD — Roealsen系列
国产领先 自主可控 规模出货

PCIe Gen3 NVMe 企业级SSD — Haishen系列
能效比之王 经全球市场认可

DapuStor Roealsen5 PCIe 4.0 NVMe SSD

1 第三代固件架构
固件稳定，相比上一代多达200余项优化

2 PCIe Gen4.0，NVMe1.4a，双端口
支持最新的NVMe协议，为企业运维带来更多便利

3 领先性能
带宽：7400/7000 MB/s
IOPS(稳态)：1750/550K
低延时： 65/10 μs

4 容量丰富
1.6TB-15.36TB 容量大范围可选

5 自研控制器
DapuStor DPU600
12mm FinFet智能存储SoC

6 领先的NAND技术
采用KIOXIA最新的BiCS 112L
3D Enterprise TLC NAND Flash

DapuStor Xlenstor2 PCIe 4.0 NVMe SCM

1 超高性能
读写带宽：7400/7060 MB/s
随机读写：1750/1340K IOPS

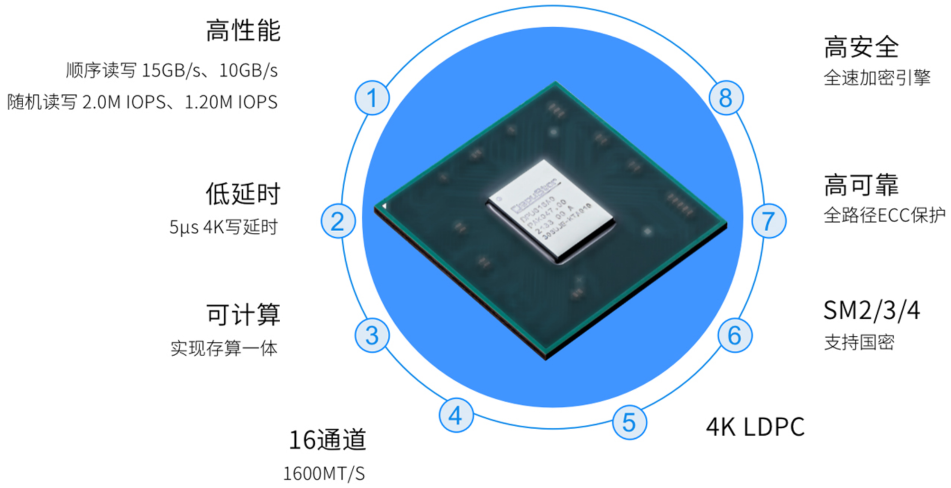
2 极低延时
4K读写延时:17/5 μs

3 企业级特性按需定制
单/双端口、VSS、Multinamespace、NVMeMI等

4 SCM级别SLC NAND
采用KIOXIA最新的XL-FLASH

5 超长使用寿命
60-120 DWPD

DapuStor PCIe 4.0 主控芯片 DPU600



DapuStor 产品概述

产品名称	容量	形态	接口	生产状态
Haishen3	1-8TB	U.2/AIC/E1.S	PCIe 3.0 x4, NVMe 1.3	量产
Roealsen5	2-16TB	U.2	PCIe 4.0 x4, NVMe 1.4a	量产
Xlenstor Gen2	0.4/0.8/1.6/3.2TB	U.2	PCIe 4.0 x4, NVMe 1.4a	量产
Xlenstor	0.75/0.8/1.6TB	U.2/AIC	PCIe 3.0 x4, NVMe 1.3	量产

欢迎关注微信

回复*白皮书*下载原文