

## 实验室报告

### QSAN XCubeNXT “XN5124D” 整合式存储设备与东芝 22TB 近线 SAS HDD

作者：Rainer W. Kaese

业务开发高级经理, 储存装置事业部

Toshiba Electronics Europe GmbH

面对虚拟化、容器化、监控、备份等工作负载，现代企业和公司比以往任何时候都更需要大规模的多功能存储，而性能是一切的关键。为此，东芝团队在其实验室中展开一项测试，使用东芝新一代企业级容量型 MG 系列 HDD (22TB SAS 12Gb/s 型号 MG10SFA22TE)，并安装在 QSAN 旗舰产品 XN5124D 中 (图 1)。这是一款 4U/24 盘位的整合式存储设备，可以用作网络附加存储 (NAS) 来提供共享文件夹/文件系统，也可以作为

专用存储区域网络 (SAN) (如 iSCSI、光纤通道) 的块存储运行，甚至可以同时用作 NAS 和 SAN。QSAN 将其宣传为“最新一代 XCubeNXT，利用业界领先的多功能性来处理混合工作负载。”此外，该设备高可用度且安全的多协议和跨 QSAN 平台支持，可为高容量需求应用提供最佳的总拥有成本 (TCO)。



图 1: 东芝 HDD 实验室中的 QSAN XCubeNXT XN5124D

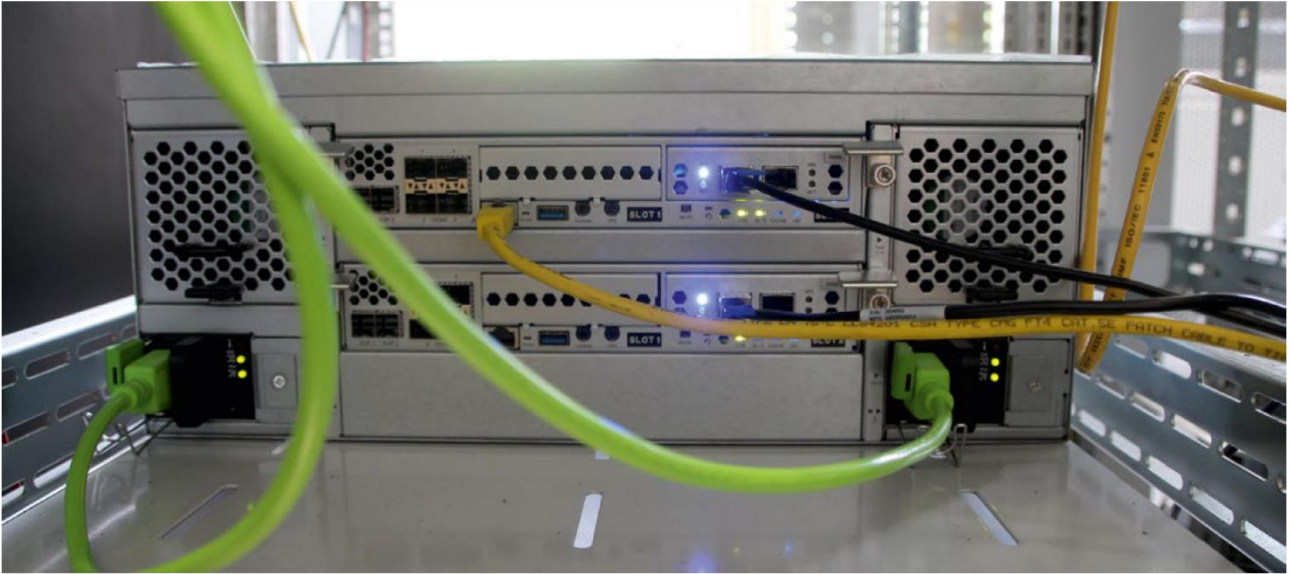


图 2：XN5124D 网络连接

## 配置和性能比较 重点关注最大性能

在之前的 24 盘位存储设备（如 QSAN XN8024D）的实验室报告中，最大性能受限于网络接口（通常为 10GbE），最高传输速度为 1200 MB/s。由于这个限制，我们目前还无法评估该装置的性能极限。实验室中的 XN5124D 型号，是第一款配备 25GbE 接口的 24 盘位存储设备，因此我们重点评估不同 HDD 配置的最高性能。

## 设置

XN5124D 配备双控制器架构，因此最适合大容量 SAS 近线 HDD。机架式，具有从网络到 HDD 访问的双路径以及双重/冗余电源，非常适合企业级应用。为了实现大规模部署，我们安装了东芝企业级容量型 MG 系列的当前热门型号，22TB SAS 12Gb/s 型号 MG10SFA22TE。

每个控制器默认配备 4 个 10GbE SFP+ 通道、一个 1GbE RJ45 管理端口和两个 SFF-8644 (mini-SAS-HD) 端口，可使用外部 JBOD 扩展容量。两个子模块插槽支

持升级不同的连接技术，例如光纤通道，或者本例中带有两个 SFP28 端口的 25GbE。我们通过 SFP28 直连铜缆将每个控制器单元的一个端口连接到应用程序服务器（图 2）。

仅供参考：我们测试的型号版本为 XN5124D-044C20，控制器固件为 4.0.2。

## HDD 配置

为了进行性能评估，我们创建了一个包含 24 个 HDD 磁盘组的池 (Pool)，这些磁盘组采用不同的 RAID 级别配置（表 1）。

**RAID6 of 24 HDDs**：一个具有双重奇偶校验 (Double Parity) 的 RAID 配置。拥有 22 个 HDD 的净容量，并且由于双重奇偶校验而具有较高的数据保护级别。如果一个驱动器发生故障，数据仍然受到第二个奇偶校验的保护。Pool 效率为 92%（24 个 HDD 中有 22 个承载用户数据），或者说 528TB 总安装容量中有 484TB 可用容量。

RAID Level	RAID6	RAID60	RAID60	RAID60	RAID10
Number of sub-arrays	1	2	4	6	12
Pool Efficiency (%)	92	83	67	50	50
Usable Capacity (共 528TB)	484	440	352	264	264

表 1：RAID 设置和存储效率

**RAID60 of 24 HDDs, 具有不同数量子阵列 (2/4/6):**

RAID60（相当于多个 RAID6 组条带化结合在一起）允许更快速地并行访问磁盘。但每个子阵列都需要两个奇偶校验单元，Pool 效率降低——阵列越多，效率越低。

**RAID10 of 24 HDDs:** 该配置由 12 个子阵列组成，每个阵列在两个磁盘上有镜像数据。这种配置在写入方面相当高效，尤其是随机写入，因为不需要计算奇偶校验信息——数据只是简单镜像而已。也因此，效率下降到 50%，这对于所有基于镜像的配置来说都是典型现象，并且与 RAID6x 类型的数据保护相比，RAID10 并不是严格的双重保护。如果一个镜像发生故障，数据仅存在于剩余的一个未受保护的镜像磁盘上。

在具有上述 RAID 级别的 Pool 中，我们安装了一些卷，其中一个 50TB 的卷始终作为 LUN 连接到 iSCSI 主机组（图 3）。此 iSCSI LUN 连接到一个应用服务器，作为一个 Windows 物理盘。我们使用“fio”基准测试评估该物理盘的性能、同时也测量了 HDD 的温度和功耗。

所有测量都是在 block volume 状态在线（即完全初始化）的情况下完成的，对于奇偶校验 RAID 级别（例如 RAID6 和 RAID60），10TB volume 容量大约需要两个小时。

设置如下（表 2）：

Pool 设置	Disk Write Cache	Enable
	Disk Read Ahead	Enable
	Disk CMD Queuing	Enable
	Disk Standby	Enable
Volume 设置	Backgr. IO Priority	Low
	Cache Mode	Write Back
	Video Edit Mode	Disable
	Read Ahead	Enable

表 2：Pool/Volume 设置

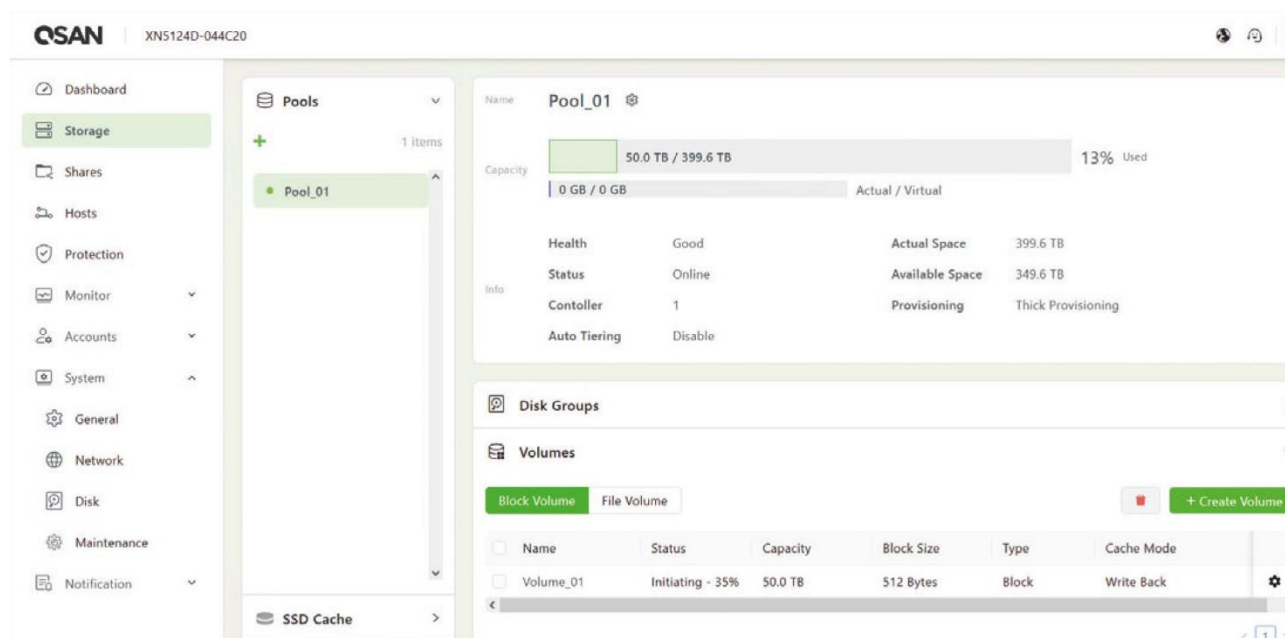


图 3：QSAN 的 QSM GUI 中的 Pool/Volume 配置

## 性能测量结果：

RAID 级别/性能	RAID6	RAID60	RAID60	RAID60	RAID10
Number of sub-arrays	1	2	4	6	12
Pool Efficiency (%)	92	83	67	50	50
Usable Capacity (out of 528TB)	484	440	352	264	264
Sequential read 1MB (MB/s)	2600	2730	2590	2750	2770
Sequential write 1MB (MB/s)	1500	1470	1500	1460	1450
Random read 4kB (IOPS)	3720	3660	3510	3340	3340
Random write 4kB (IOPS)	1600	1600	1550	2700	8160
Mix r/w 4k/64k/256k/2M (IOPS)	770	790	790	820	920

表 3：性能测量结果

顺序读取性能约为 2700 MB/s，与应用服务器的 25GbE iSCSI 连接速度相匹配，因此速度由网络决定，而不是由 RAID 级别决定。无论 RAID 级别如何，顺序写入速度都在 1500 MB/s 左右。这表明，由于 RAID6 级别具有较高的存储效率，实际上是顺序访问（存档、流媒体、视频）为主的工作负载的最佳解决方案。任何并行性更高的磁盘访问配置（如 RAID60 或 RAID10）都不会带来更高的性能，而只会减少可用的存储容量。

即使对于随机读取（对于 AI/视频分析类型的应用程序很重要），RAID6 表现出最高的 IOPS 性能，因为在这种配置下，数据分布在最多数量的有效磁盘上（24 个中的 22 个）。

对于以随机写入为主的工作负载（例如，活跃目录/数据库和电子邮件服务/虚拟化），具有更多子阵列的 RAID60 并行结构可提供更高的性能。对于 RAID10 也是如此，没有奇偶校验操作，只是简单的数据镜像，性能甚至能高达 8k IOPS。因此，如果 HDD 大容量存储的写入性能至关重要，则 RAID10 是最佳配置。

虽然 RAID10 的存储效率较低，但其保护级别不如 RAID6 和 RAID60 高 - 因为发生错误的两个故障 HDD 如果位于同一个镜像，那么就会导致数据丢失，而在 RAID6/60 中，即使任意两个驱动器发生故障，也不会导致丢失数据。

此外，对于上述随机类型工作负载，SSD 缓存有机会提供显著的性能优势。但由于提升的程度高度取决于工作负载的实际情况，因此这个选项不在本实验室评估的范围内。

## 功率和温度测量

Maximum Power at Startup:	480 W
Power under full read/write workload:	480 W
Power when unit is idle (no read/write):	330 W
Standby Power (unit off):	10 W

RAID 级别/功耗	RAID6	RAID10
Sequential read 1MB (W)	410	410
Sequential write 1MB (W)	400	400
Random read 4kB (W)	400	420
Random write 4kB (W)	420	380
Mix r/w 4k/64k/256k/2M (W)	410	390

表 4：功耗测量

该型号配备了东芝 MG10SFA22TE 类型的高容量 HDD，为降低 IT 和数据中心存储功耗做出了重要且重大的贡献，测得的最高功耗值不到 500W，对于高达半 PB 的净存储，平均功耗为 400W，包含高速网络接口，整体工作功耗不到 1kW/PB，这是一个非常优秀的数值。

实验室环境温度:	24°C
HDD 最低温度 (SMART 值):	36°C
HDD 最高温度 (SMART 值):	42°C (HDDs in full read/write activity)

东芝 MG 系列 HDD 的内部温度最高可至 60°C，可见 42°C 的最高温度远远没有达到极限值。但长期可靠性在平均温度或恒定温度超过 42°C 时往往会下降。因此，为了获得尽可能高的长期可靠性，保持环境温度在 24°C，以确保在较长时间内 HDD 温度不会超过 42°C。



## 测量脚本

使用 **fiio** 进行测量。列出指令以供参考：

```
fiio --filename=\\.\Physicaldrive1 --direct=1 --rw=read --bs=1m --iodepth=128 --time_based
--runtime=300 --group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=4
--norandommap --randrepeat=0 --output=seqread.txt

fiio --filename=\\.\Physicaldrive1 --direct=1 --rw=write --bs=1m --iodepth=128 --time_based
--runtime=300 --group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=4
--norandommap --randrepeat=0 --output=seqwrite.txt

fiio --filename=\\.\Physicaldrive1 --direct=1 --rw=randread --bs=4k --iodepth=128 --time_based
--runtime=300 --group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64
--norandommap --randrepeat=0 --output=randread.txt

fiio --filename=\\.\Physicaldrive1 --direct=1 --rw=randwrite --bs=4k --iodepth=128 --time_based
--runtime=300 --group_reporting --name=job1 --ioengine=windowsaio --thread --numjobs=64
--norandommap --randrepeat=0 --output=randwrite.txt

fiio --filename=\\.\Physicaldrive1 --direct=1 --rw=randrw --bssplit=4k/20:64k/50:256k/20:2M/10
--iodepth=128 --time_based --runtime=300 --group_reporting --name=job1 --ioengine=windowsaio
--thread --numjobs=64 --norandommap --randrepeat=0 --output=mixed.txt
```

## 结论

经测试的 QSAN XN5124D 整合式存储系统提供大容量、高性能、高度可用和可靠的企业级数据存储，完全可以满足开篇描述的工作负载（例如虚拟化、容器化、监控和备份）的严苛容量和性能需求。

使用 24 块东芝 22TB 企业级 SAS HDD，原始容量为 528TB，根据配置不同，净容量从 264TB (RAID10) 到 484TB (RAID6) 不等。

即使没有任何 SSD 缓存，接近 3GB/s 的顺序读取和高达 1.5GB/s 的顺序写入性能也是非常出色的数据，与 25GbE 网络连接的理论极限值相符。建议采用 RAID6 配置，因为它兼具高顺序性能和双重奇偶校验保护，并实现了高达 92% 的存储效率。随机读取性能约为 3000 IOPS，写入性能峰值高达 8000 IOPS（在 RAID10 配置下）。

满载下的最大功耗不超过 400W（每 TB 净容量功耗不到 1W，包括扩展卡、控制卡、25GbE SFP+ 和 RJ 网络接口——非常出色的能效值!）。

适当且高效的散热确保 HDD 温度保持相对较低，温差较小，这有助于延长旋转式磁盘的使用寿命并降低故障率。而且所有这些优异性能都是在确保高水平的数据保护和低拥有成本的前提下实现的，这对于企业级应用来说也是一个重要的考量因素。

## 致谢合作伙伴

我们要感谢我们的合作伙伴 QSAN ([www.qsan.com](http://www.qsan.com)) 提供 XN5124D 测试机，也感谢他们在设置和性能测试方面给予的大力协助和合作。此次合作是价值链合作的优秀范例，为 QSAN 和东芝客户提供了如何设计经过验证的存储解决方案的宝贵指导。

# TOSHIBA

---

**Toshiba Electronic Devices & Storage Corporation**

东芝电子元件及存储装置株式会社

<https://toshiba.semicon-storage.com/cn/storage>

Copyright © 2024 Toshiba Electronics Europe GmbH. All rights reserved.  
Product specifications, configurations, prices and component / options availability  
are all subject to change without notice. Product design, specifications and colours  
are subject to change without notice and may vary from those shown. Errors and  
omissions excepted.

发布时间：2024年7月