

下一代数据存储技术 研究报告 (2021 年)

中国信息通信研究院云计算与大数据研究所

2021 年 7 月

版权声明

本报告版权属于中国信息通信研究院，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：中国信息通信研究院”。违反上述声明者，本院将追究其相关法律责任。

编制说明

本报告的撰写得到了行业内许多专家的支持和帮助，他们分别来自：华为技术有限公司、阿里云计算有限公司、北京百度网讯科技有限公司、星辰天合（北京）数据科技有限公司、新华三技术有限公司、北京志凌海纳科技有限公司、北京金山云网络技术有限公司、深圳市杉岩数据技术有限公司、西安奥卡云数据技术有限公司、浪潮电子信息产业股份有限公司、英特尔（中国）有限公司。对这些专家的大力支持表示由衷的感谢！

前 言

伴随着信息技术的迅猛发展，以寄存器、缓存、主存、外存等构成的多级数据存储系统已成为现代信息系统不可或缺的底层基座。近年来，随着移动互联网和物联网的飞速发展，数据存储系统面临的业务环境呈现出两大特点：一是数据量爆发式增长，规模到达EB级、访问频率到达亿级每秒；二是现代业务负载动态变化，需要极强的弹性伸缩和智能资源管理能力。传统数据存储系统受限于介质、架构、协议等因素，其能力逐渐出现瓶颈，无法满足业务需求，这催生了下一代数据存储技术。

下一代数据存储技术指在存储介质、存储架构、存储协议、应用模式及运维模式等方面迭代创新的一系列技术的集合，总体呈现高性能、易于扩展、服务化和智能化等特点。数据存储技术革新从根本上改变着数据存储产品的形态。如何围绕下一代数据存储技术，构建一套稳定、高效、满足未来业务发展需求的数据存储系统，是企业组织夯实数据底座、挖掘数据价值、释放数据潜能的关键。

本报告聚焦于企业级存储系统，从数据存储技术面临的挑战切入，详细梳理了下一代数据存储技术，跟踪重点行业的应用动态，并展望了存储产业发展趋势。力图展现下一代数据存储技术路径及其发展前景。由于时间仓促，水平所限，错误和不足之处在所难免，欢迎各位读者批评指正，意见建议请发送至 tianjiafeng@caict.ac.cn。

目 录

| | |
|----------------------------|----|
| 一、存储产业与技术概述..... | 1 |
| (一) 存储技术发展史..... | 1 |
| (二) 存储技术分类..... | 3 |
| (三) 存储产业市场简析..... | 7 |
| (四) 存储产业痛点分析..... | 11 |
| 二、下一代数据存储技术详解..... | 13 |
| (一) 存储介质演进..... | 13 |
| (二) 存储架构演进..... | 17 |
| (三) 存储协议演进..... | 21 |
| (四) 应用模式演进..... | 24 |
| (五) 运维模式演进..... | 25 |
| 三、下一代数据存储技术赋能数字化转型..... | 27 |
| (一) 异构数据统一管理推动数据湖产品演进..... | 28 |
| (二) 多级存储介质助力实时分析能力构建..... | 29 |
| (三) 云存储备份简化数据安全实践路径..... | 30 |
| 四、下一代数据存储技术实施建议..... | 31 |
| (一) 指导原则..... | 31 |
| (二) 实施策略..... | 34 |
| 五、总结与展望..... | 36 |
| 参考文献..... | 39 |

图 目 录

| | |
|----------------------------------|----|
| 图 1 存储介质发展简史 | 1 |
| 图 2 存储技术架构 | 4 |
| 图 3 存储市场全景图 | 8 |
| 图 4 中国企业级存储市场空间 | 8 |
| 图 5 中国存储市场总览 | 9 |
| 图 6 中国存储市场介质使用情况 | 10 |
| 图 7 闪存及机械介质成本预测图 | 10 |
| 图 8 非易失性存储介质 | 15 |
| 图 9 软件定义存储逻辑架构示意图 | 18 |
| 图 10 超融合基础架构图 | 20 |
| 图 11 超融合基础架构特征 | 20 |
| 图 12 基于 FC 和 SAS 协议的经典存储网络 | 22 |
| 图 13 基于 NVMe 协议的 IP 存储网络 | 24 |

表 目 录

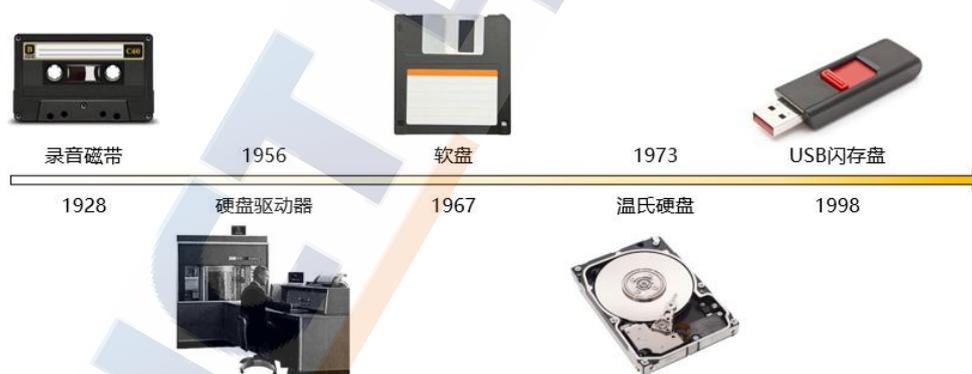
| | |
|-------------------------|----|
| 表 1 下一代数据存储技术革新方向 | 13 |
| 表 2 机械硬盘与固态硬盘性能对比 | 14 |
| 表 3 新型非易失性存储简介 | 15 |
| 表 4 不同类型软件定义存储简介 | 18 |

一、存储产业与技术

（一）存储技术发展史

数据的存储从古至今都是人类活动的重要环节。早在语言文字还没有形成时，人类就已经开始探索使用树枝和石头来记录信息。从结绳记事到刻画在岩石上的象形文字，再到甲骨文的出现，从竹简、纸张的发明，到活字印刷、打孔卡等科技进步，人类探索世界得到的信息和数据不断积累，代代传承，提升了人类认识、改造世界的能力。对信息沟通量与质的不懈追求，促使人类探寻更大容量、更高性能的存储模式，推动开发和应用更多更先进的数据存储技术，使数据更好地储存和交互，提高数据使用的便捷性与持久性。

20世纪20年代以来，伴随着电子技术的发展，存储技术进入了崭新的时代。



来源：中国信息通信研究院

图 1 存储介质发展简史

1928年，可存储模拟信号的录音磁带问世，每段磁带随着音频信号电流的强弱不同而被不同程度的磁化，从而使得声音被记录到

磁带上。1951年，磁带开始应用于计算机中，最早的磁带机可以每秒钟传输7200个字符。20世纪70年代后期出现的小型磁带盒，可记录约660KB的数据。

1956年，世界上第一个硬盘驱动器出现，应用在IBM的RAMAC305计算机中，该驱动器能存储5M的数据，传输速度为10K/S，标志着磁盘存储时代的开始。1962年，IBM发布了第一个可移动硬盘驱动器，它有六个14英寸的盘片，可存储2.6MB数据。1973年，IBM发明了温氏硬盘，其特点是工作时磁头悬浮在高速转动的盘片上方，而不与盘片直接接触，这便是现代硬盘的原型。

1967年，IBM公司推出世界上第一张软盘。随后三十年，软盘盛极一时，成为个人计算机中最早使用的可移介质。这个最初有8英寸的大家伙，可以保存80K的只读数据。四年后，可读写软盘诞生。至上世纪九十年代，软盘尺寸逐渐精简至3.5英寸，存储容量也逐步增长到250M。截止1996年，全球有多达50亿只软盘被使用。直到CD-ROM（只读光盘，Compact Disc Read-Only Memory）、USB存储设备出现后，软盘销量才开始下滑。

进入21世纪，信息爆炸导致数据量成倍增长，硬盘容量也在飙升，单盘容量已可达到TB级别。即便如此，单块磁盘所能提供的存储容量和速度已经远远无法满足实际业务需求，磁盘阵列应运而生。磁盘阵列使用独立磁盘冗余阵列技术（RAID）把相同的数据存储在多个硬盘，输入输出操作能以平衡的方式交叠进行，改善了磁盘性

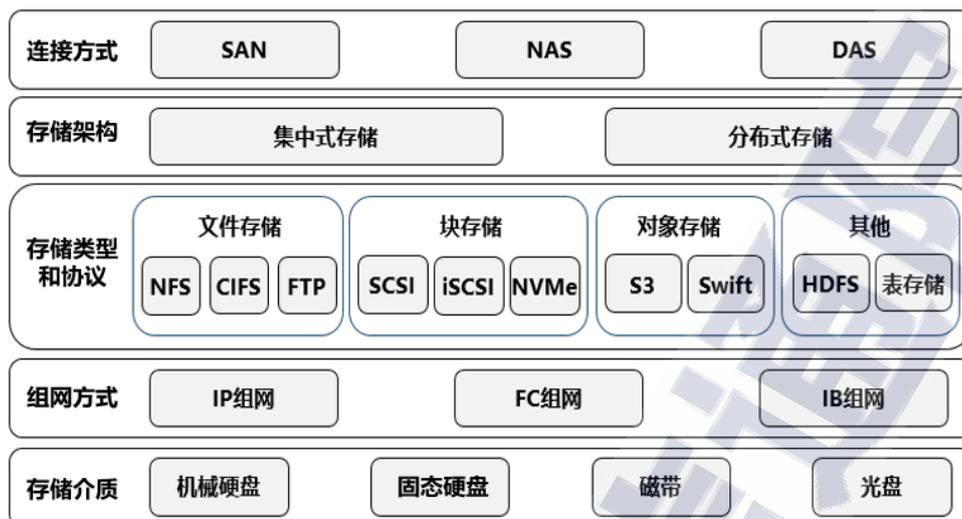
能，增加了平均故障间隔时间和容错能力。RAID作为高性能、高可靠的存储技术，已经得到非常广泛的应用。

21世纪以来，计算机存储技术飞速发展，如何快速高效的为计算机提供数据以辅助其完成运算成为存储技术新的突破口。在RAID技术实现高速大容量存储的基础上，网络存储技术的出现弱化了空间限制，使得数据的使用更加自由。网络存储将存储系统扩展到网络上，存储设备作为整个网络的一个节点存在，为其他节点提供数据访问服务。即使计算主机本身没有硬盘，仍可通过网络来存取其他存储设备上的数据。基于网络存储技术，分布式云存储、容灾备份、虚拟化和云计算等技术得以广泛应用。

数据存储，是现代信息产业架构中不可或缺的底层基座。经过百余年的发展，存储技术已经呈现出非常多的形态，且仍在不断完善和创新，以适应日益增长和不断变化的数据存储需求。

（二）存储技术架构

现有存储系统从底层到上层由存储介质、组网方式、存储协议和类型、存储架构、连接方式五个部分组成，整体架构如下图。



来源：中国信息通信研究院

图 2 存储技术架构

1. 存储介质

企业级存储中的存储介质包括机械磁盘（HDD）、固态硬盘（SSD）、磁带（TAPE）、光盘（Optical Disk）等，其中最常见的是以HDD和SSD为介质的存储系统。依据存储介质不同，存储系统可分为磁盘存储、全闪存储、混闪存储、磁带库、光盘库等。

- 磁盘存储：指全部以磁盘为永久存储介质的存储；磁盘性能一般，但价格便宜。
- 全闪存储：指全部以固态硬盘为永久存储介质的存储；全闪存储性能优异，但价格也较高。
- 混闪存储：指永久存储介质同时含有磁盘和闪存盘的存储；与全闪存储相比，混闪存储在性能和价格上进行了一定折中。

- 磁带库和光盘库：指以磁带或光盘为存储介质，由驱动器及其控制器组成的存储设备；单位存储空间价格较低，支持冷数据的长期保存，但读写性能不高。

2.组网类型

按组网方式，存储系统可分为IP（互联网协议，Internet Protocol）组网存储、FC（光纤通道，Fiber Channel）组网存储、IB（无线带宽，InfiniBand）组网存储等。

- IP 组网存储：指采用以太网技术进行组网的存储设备，常见速率包括 1Gb、10Gb、25Gb、100Gb 等；IP 组网的兼容性较好，建设成本较低。
- FC 组网存储：指采用 FC 光纤技术进行组网的存储设备，常见速率包括 8Gb、16Gb、32Gb 等；FC 组网的效率较高，但采购成本和维护难度也相对较高。
- IB 组网存储：指采用 InfiniBand 技术进行组网的存储设备，常见速率包括 40Gb、56Gb、100Gb、200Gb 等；IB 组网的延迟较低、速率较高，但采购成本相对较高，组网的扩展性也较弱。

3.存储类型

按存储类型，存储系统可分为文件存储、块存储、对象存储、其它存储等。

- 文件存储：指自身构建文件系统后，通过互通的网络提供给服务器或应用软件使用，支持数据文件读写和文件共享服务的存储设备。文件存储的常用协议包括NFS、CIFS、FTP等。
- 块存储：指将物理存储介质上的物理空间按照固定大小的块组成逻辑盘，并直接映射空间给服务器使用的存储设备。块存储的常用协议包括SCSI、iSCSI、NVMe等。
- 对象存储：指采用扁平化结构，将文件和元数据包装成对象，并抽象成网络URL（统一资源定位器，Uniform Resource Locator），通过HTTP（超文本传输协议，Hypertext Transfer Protocol）协议直接访问的存储设备。对象存储的常用协议包括S3、SWIFT等。
- 其它存储协议：包括在大数据存储中广泛使用的HDFS协议，以及表存储协议等。

4. 存储架构

按存储系统架构，存储系统可分为集中式存储和分布式存储。

- 集中式存储：指基于双控制器或多控制器架构的企业级存储系统，具有较强的纵向扩展（Scale-up）能力和一定的横向扩展（Scale-out）能力。集中式存储的特点有高可靠、高可用、高性能等。
- 分布式存储：指将商用服务器上的存储介质虚拟化成统一的存储资源池来提供存储服务。分布式存储的特点有高扩展性、低成本、易运维、和云紧密结合等。

5.连接方式

按连接方式，存储系统可分为SAN存储、NAS存储、DAS存储。

- **SAN: Storage Area Network**（存储区域网络），指通过光纤通道交换机、以太网交换机等连接设备将磁盘阵列与相关服务器连接起来的高速专用存储网络。
- **NAS: Network Attached Storage**（网络附加存储），是一种专业的网络文件存储及文件备份设备，对不同主机和应用服务器提供文件访问服务。
- **DAS: Direct Attached Storage**（直接附加存储），将存储设备通过 SCSI（小型计算机系统接口，Small Computer System Interface）接口或光纤通道直接连接到一台主机上，主机管理它本身的文件系统，不能实现与其他主机的资源共享。

各类型数据存储技术仍在不断的完善和创新，以适应日益增长和不断变化的数据存储需求，形成了多层次、广泛覆盖的产品体系结构，为用户提供了各应用场景下的存储解决方案。

（三）存储产业市场纵览

根据统计机构国际数据公司（IDC）的定义，一般意义上的企业存储市场特指的是企业级存储系统（ESS, Enterprise Storage System）市场，它在存储整体市场中的位置如下：

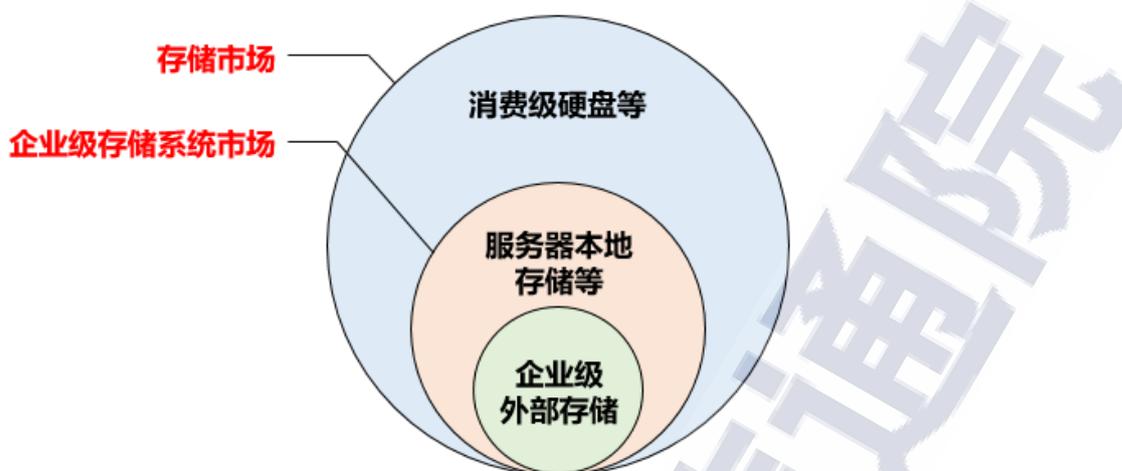
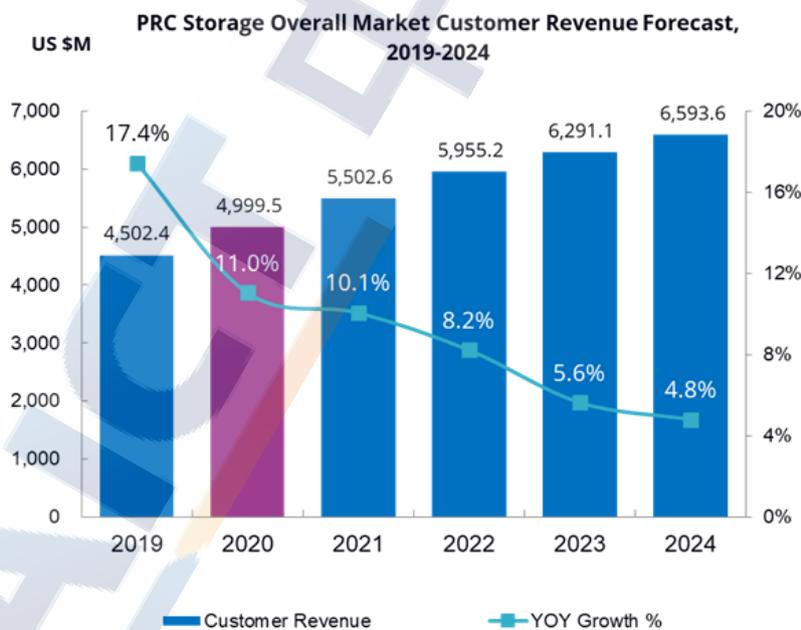


图 3 存储市场全景图

据IDC的预测，2021年中国企业级存储市场空间55亿美元，2020-2024年中国企业级存储市场将保持7.16%的年复合增长率，到2024年中国企业级存储市场空间将达到65.9亿美元。



数据来源¹：IDC，2020

图 4 中国企业级存储市场空间

¹ 数据来源：2020Q3 IDC Enterprise Storage System Market Overview。

存储市场按照存储架构可以分为传统企业级存储（TESS）、软件定义存储（SDS）、超融合基础架构（HCI）。据IDC统计，2020年Q1-Q3中国存储市场中传统企业级存储占比59.2%，仍是最主流的存储架构，SDS软件定义分布式存储占比22.2%，HCI超融合占比18.6%。SDS和HCI的增速远远超过TESS，其中SDS未来四年复合增长率12.8%，HCI未来四年复合增长率13.0%。分布式存储在政府、运营商、金融、企业、教育、医疗、能源等众多行业被广泛接受和采用。以中国移动为代表的运营商更是在2021年的集采中大量采购分布式存储，构建存储资源池，全面替代以前的中低端存储。

China Storage Market Overview (HW), 2020YTD

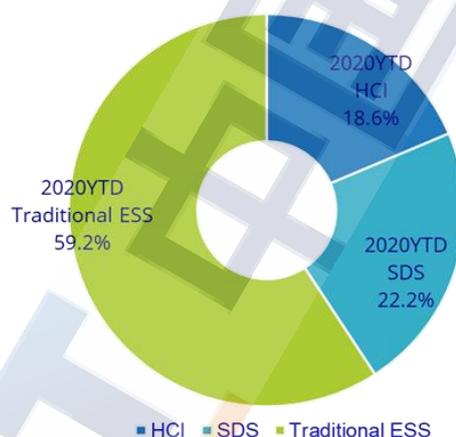
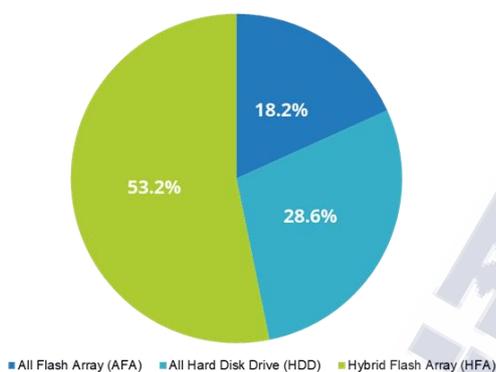
数据来源²：IDC，2020

图 5 中国存储市场总览

存储市场按照存储介质可以分为全闪存储（AFA）、混闪存储（HFA）、全机械盘存储（HDD）。2020年Q1-Q3中国存储市场中全闪存储占比18.2%，混闪存储占比28.6%，两者之和接近50%，且全闪存储的增速较快，较2019年Q1-Q3同比增长20%。

² 数据来源：2020Q3 IDC PRC Software-defined Storage and Hyperconverged System Market Overview。

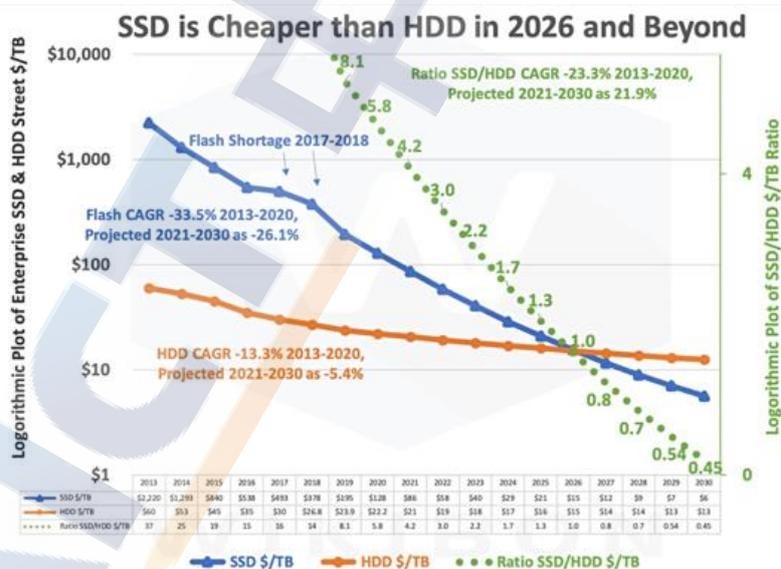
PRC Storage Market Overview by Storage Array Type, 2020YTD



数据来源³：IDC，2020

图 6 中国存储市场介质使用情况

在全球存储市场中，由于闪存成本不断下降，全闪存储份额快速增加，市场占比超过20%。根据第三方机构Wikibon的预测，2026年SSD单TB成本将低于HDD，达到15美元/TB。2025年后，HDD的出货量将每年下降27%，只应用于扩容和归档等少数场景。



数据来源：Wikibon，2021

图 7 闪存及机械介质成本预测图

³ 数据来源：2020Q3 IDC Enterprise Storage System Market Overview。

随着国内存储厂商的技术进步及国家自主创新的政策激励，国产存储厂商份额不断提升，国产化存储产品也逐渐被用户接受。在传统存储阵列领域，当前市场份额较大的厂商包括华为、新华三、浪潮、曙光等；在分布式存储领域，华为、曙光、新华三、XSKY 等厂商占有较大市场份额。而在云存储领域，阿里云国内领先的云存储厂商。

（四）存储产业痛点分析

“数据”是数据中心乃至企业最重要的资产。在数字社会，数据具有基础战略资源和关键生产要素的双重角色。作为信息化系统中的核心部分和底层基座，存储系统的构建和使用直接关系到数据这一企业核心资产的存储、使用和挖掘。当前，存储系统面临的业务环境呈现出以下两个特点。

一是数据量爆发式增长。随着移动互联网不断发展，企业的数据规模呈现爆发式增长。《IDC：2025 年中国将拥有全球最大的数据圈》研究表明，2018 年中国新增数据量为 7.6ZB，成为世界第一数据生产国；2025 年中国新增数据量将达到 48.6ZB，年平均增长率为 30%。

二是业务负载呈现动态变化。现代业务平台的负载是非线性的、动态变化的，尤其是互联网类的业务，随时可能出现业务负载的突发性变化。以 2020 年双十一为例，0 点至 0 点 30 分，天猫实时成交额突破 3723 亿。订单创建峰值为 58.3 万笔/秒，是 2009 年第一次双十一的 1457 倍。

当前存储技术存在以下四个问题，导致难以满足上述业务需求：

一是主要采用集中式架构，横向可扩展性差。传统存储由于紧耦合架构和单一的协议，导致性能和空间无法独立扩容，支撑业务快速部署上线。相较于计算虚拟化环境下秒级的虚拟机部署和回收，存储的扩容和资源分配往往需要几个小时甚至几天以上的时间，难以满足客户快速扩展的需求。如何提高存储系统的横向扩展能力，实现对资源的即时使用和释放，是未来存储技术亟待解决的问题。

二是无法统一调度，存储容量浪费。受限于架构设计，不同存储之间的资源无法进行统一调度和管理，时常出现个别节点资源存在剩余，但其他节点却出现空间不足的情况。存储整体利用率低，造成存储、计算资源浪费。大数据时代，这一问题被进一步放大，如何减少资源浪费、降低使用成本将是未来各行业用户关注的重点。

三是产品兼容交互能力差，业务复杂性变高。据《IDC FutuerScapes 2019》分析指出，未来企业 49%的应用将与其他应用有紧密关联，每个企业级应用都将与其他 4-8 个应用有频繁的交互。不同供应商或型号存储产品往往难以互通，当需要替换存储设备时，数据迁移问题无论对运维还是业务平台本身都是一项极大的挑战。

四是多种存储产品混合使用，运维难度大。目前我国各行业用户采购的存储产品往往由多家厂商供应，由于接口、协议、工具的差异，造成了诸如业务复杂性变高、运维成本上升等问题。如何打破“数据孤岛”，避免重复式存储规划和数据丢失带来的损失成为企业数字化转型过程中面临的一大挑战。

为了解决上述问题，满足各行业在使用存储系统过程中出现的新场景、新需求，下一代数据存储技术应运而生。

二、下一代数据存储技术

进入移动互联网时代，存储应用场景急剧变化，下一代数据存储技术应运而生。下一代数据存储技术主要指在存储介质、存储架构、存储协议、应用模式及运维模式等方面迭代创新的一系列技术的集合，总体呈现出高性能、易于扩展、服务化和智能化等特点。

表 1 下一代数据存储技术革新方向

| 技术分类 | 传统存储 | 下一代存储 |
|------|--------------|----------------|
| 存储介质 | 机械硬盘存储、易失型内存 | 全闪存储、非易失性内存 |
| 存储架构 | 集中式存储 | 软件定义存储、超融合基础架构 |
| 存储协议 | AHCI、SCSI 协议 | NVMe 协议 |
| 应用模式 | 本地部署 | 云服务化 |
| 运维模式 | 人工运维 | 智能化运维 |

来源：中国信息通信研究院

本章将从这五个方面对下一代数据存储技术的演进进行梳理与分析。

（一）存储介质演进

1. 全闪存储

全闪存储指全部由固态存储介质构成的独立存储阵列或设备。全闪存储的高速发展，既是新应用对性能需求驱动的结果，也是闪存技术不断创新成果。

在需求侧，当前数据爆炸式增长，云计算、物联网、大数据、

人工智能、区块链等新技术快速发展，驱动人脸识别、自动驾驶等新智能应用不断涌现，业界已进入智能驱动的新数据时代。这一时代，数据存储具有三大需求，分别是 EB 级容量、亿级 IOPS（每秒进行读写操作的次数，Input/Output Operations Per Second）和智能管理，亿级 IOPS 需求使得存储介质的变革势在必行。

在技术侧，全闪存储普遍被认为是存储行业的发展方向，其具备远高于传统磁盘存储的数据吞吐能力及更低的时延。数据显示，固态硬盘对比机械硬盘，拥有更快的读取速度、更低的功耗以及更低的故障几率，实现了对机械硬盘性能的全面超越，为底层存储介质的替换提供了客观条件。

表 2 机械硬盘与固态硬盘性能对比

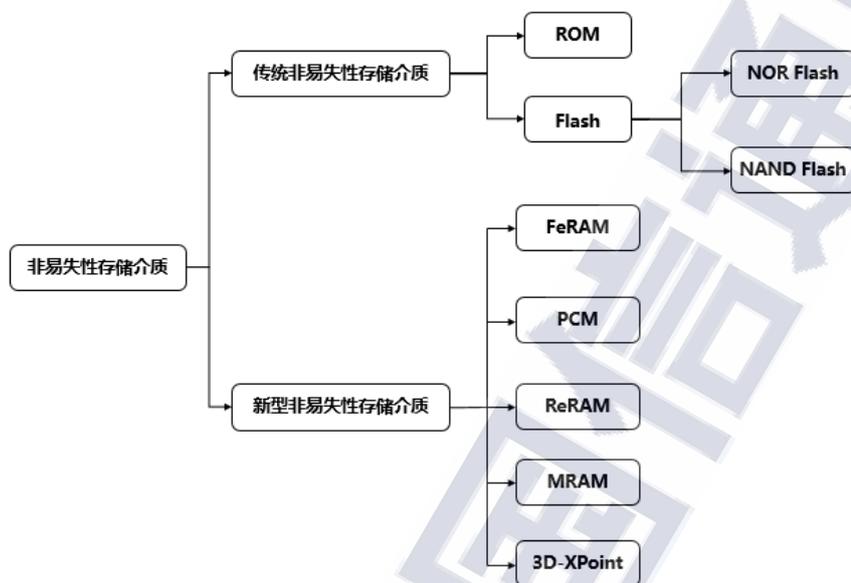
| | 机械硬盘 | 固态硬盘 |
|--------|-------|--------|
| 时延 | 2ms | 0.02ms |
| 5 年返还率 | 13.4% | 0.8% |
| 功耗 | 10w | 3w |

数据来源：华为技术有限公司

在产品侧，各大存储厂商均推出了全闪存储产品。如华为的 OceanStor Dorado 及浪潮的 HF 系列产品。以 OceanStor Dorado 为例，相较传统机械存储，在存储性能委员会（SPC）的 SPC-1 基准下，业务性能提升了 5 倍；在数据库场景下，业务性能提升了 10 倍；在虚拟桌面场景下，在 Word/PowerPoint/Excel 应用测试中，启动响应时间缩短 80%。充分体现了全闪存储产品优异的性能。

2.非易失性内存

非易失性内存（Non-Volatile Memory, NVM）是指断电后，存储的数据不会消失的存储器，是存储技术领域近十余年来最具革命性的创新。依据技术原理，非易失性内存介质可分以下几类。



来源：西安奥卡云数据科技有限公司

图 8 非易失性存储介质

为开发出比传统非易失性存储介质更高速、更低功耗、更高密度、更可靠的新型非易失性存储介质，研究者们把目光聚焦到一些特殊材料上，提出了一些存储介质模型。下表对这五种存储介质的存储原理、优缺点及发展状况进行了简要介绍。

表 3 新型非易失性存储简介

| 存储介质 | 存储原理 | 优点 | 缺点 |
|---------|-----------------------------|---------------------------|----------|
| 铁电随机存储器 | 通过铁电材料的不同极化方向来存储数据 | 读写速度,低功耗和擦写循环性能好 | 数据保持能力较差 |
| 磁性随机存储器 | 通过磁化方向的改变来存储数据,并通过磁效应来实现数据读 | 可反复擦写次数高等优点 | 难以小型化 |
| 阻变存储器 | 利用材料的电阻在电压作用下发生变化的现象来存储数据 | 擦写速度快、存储密度高、具备多值存储和三维存储潜力 | 材料耐久性较差 |

| | | | |
|-----------|---|-----------------------|------|
| 相变存储器 | 以硫属化合物为基础的相变材料在电流的焦耳热作用下，通过晶态和非晶态之间的转变来存储数据 | 重复擦写次数高、存储密度高、多值存储潜力大 | 功耗较高 |
| 3D-XPoint | 通过特定的电压差，改变存储单元中特殊材料的电阻，实现写操作 | \ | \ |

来源：中国信息通信研究院

随着存储技术的发展和人们对存储性能的不懈追求，高性能存储的探索开始向内存通道迁移。非易失性双列直插式内存模块（Non-Volatile Dual In-Line Memory Module, NVDIMM）便在这种趋势下应运而生。根据电子器件工程联合委员会标准化组织的定义，有三种 NVDIMM 的实现方式：

- （1）NVDIMM-N 指在一个模块上同时放入动态随机存取存储器（DRAM）和闪存。通过使用一个小的后备电源，为掉电时数据从 DRAM 拷贝到闪存中提供足够的电能，当电力恢复时，再重新加载到 DRAM 中。
- （2）NVDIMM-F 指使用了 DRAM 的 DDR（双倍速率，Double Data Rate）总线 Flash 闪存，一定程度上减少协议带来的延迟和开销，但只支持块寻址。
- （3）NVDIMM-P 是真正 DRAM 和 Flash 闪存的混合。它既支持块寻址，也支持类似传统 DRAM 的按字节寻址。容量可以达到类似 NAND 闪存的 TB 级，又能把延迟保持在 10^2 方纳秒级。Intel 发布的基于 3D XPoint 技术的英特尔傲腾持久化内存，可认为是 NVDIMM-P 的一种实现。

非易失性内存的出现填补了从硬盘到 DRAM 之间，存储在性能、延迟、容量成本的鸿沟，为多样化的解决方案奠定了坚实的基础。非易失性内存技术能够存储不适用于 DRAM 的庞大数据集，进行快速计算，同时与其他存储介质共同组成多级存储池，让数据更加靠近处理器，提升存储系统的整体性能表现。

（二）存储架构演进

大数据、云计算和虚拟化等技术的出现，使得传统的 IT 架构难以满足企业日益增长的数据存储需求。为应对这一挑战，软件定义存储（SDS, Software Defined Storage）和超融合基础架构（HCI, Hyper-Converged Infrastructure）应运而生，打破了传统 IT 系统复杂和繁冗的现状，优化了网络的可扩展性和管理方式。

1. 软件定义存储

企业对数据服务的需求变得更加复杂、精细和个性化，对数据存储的高可靠性、高性能、高扩展以及面向云架构的延伸能力等层面提出了更高的要求。虚拟化和云技术的发展和成熟转变了数据中心的设计、建造、管理和运维方式，这种变革使软件定义存储越来越有吸引力。

软硬解耦、易于扩展、自动化、基于策略或者应用的驱动是软件定义存储的特征。就业务应用来说，即不限制上层应用，不绑定下层硬件；除了提供块存储，也可以在同一平台提供文件、对象、HDFS 等存储服务，实现非结构化数据的协议互通；同时应具备完善的监控能力，实现应用感知。



来源：星辰天合（北京）数据科技有限公司

图 9 软件定义存储逻辑架构示意图

软件定义存储对全行业业务都具备适用性，可根据当前业务场景需要进行规划设计。

表 4 不同类型软件定义存储简介

| 类型 | 功能 | 应用场景 |
|--------|---|--|
| 块存储 | 提供块存储服务，常见多用于提供虚拟化、私有云、数据库等使用块存储资源的结构化数据业务，业务应用直接对磁盘块进行访问操作。 | 虚拟化环境：VMware、KVM、OpenStack、容器等 云环境：CloudStack、ZStack、WinCloud 等 数据库：Oracle、MySQL 等 |
| 对象存储 | 提供 S3 协议接口服务，支持通过 Restful 的访问方式进行数据的上传、下载、删除。常用于互联网、移动终端应用业务云存储，以海量非结构化类型数据为特征。 | 流媒体文件、非结构化数据、内容管理、数据备份归档、网盘、电子单据 |
| 文件存储 | 提供文件存储功能，支持 NFS/CIFS 等访问协议，常用于传统局域网共享型的业务应用，以支持多个用户对同一文件协同操作作为特征。 | 高性能计算、非结构化数据、内容管理、数据归档 |
| 统一数据平台 | 同时提供块、文件、对象等多种协议 | 全业务场景的适配，支撑企业的各种各样的应用负载 |

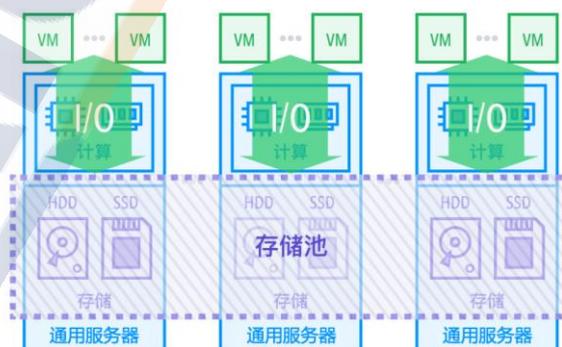
来源：星辰天合（北京）数据科技有限公司

开放化和水平扩展是软件定义存储的两大特点。开放化意味着接口标准化、服务原子化，保证客户的应用系统能够以最顺畅的方式对接基础存储设施，可微调解决方案细节，达成高质量的服务。水平扩展则是云计算弹性环境的必然要求，在移动互联网环境下，业务应用的负载量是突发式、潮汐式、难以精确预测的，应用要求存储的容量和性能都必须能够线性扩展以满足上层应用需求。

软件定义存储产品在提供高可靠和高可用服务能力的同时，集成了数据智能处理和分析能力，简化了海量数据处理所需的基础设施，帮助客户实现数据互通、资源共享、弹性扩展、多云协作，有效降低用户的使用成本。Gartner 报告指出，通过使用 XSKY 的 XEDP 数据存储平台，企业用户实现整体使用成本降低 54%。

2.超融合基础架构

超融合基础架构是一种软件定义的 IT 基础架构，它可虚拟化常见“硬件定义”系统的所有元素。超融合基础架构包含的最小集合是：虚拟化计算、虚拟存储和虚拟网络；超融合系统通常运行在标准商用服务器上。



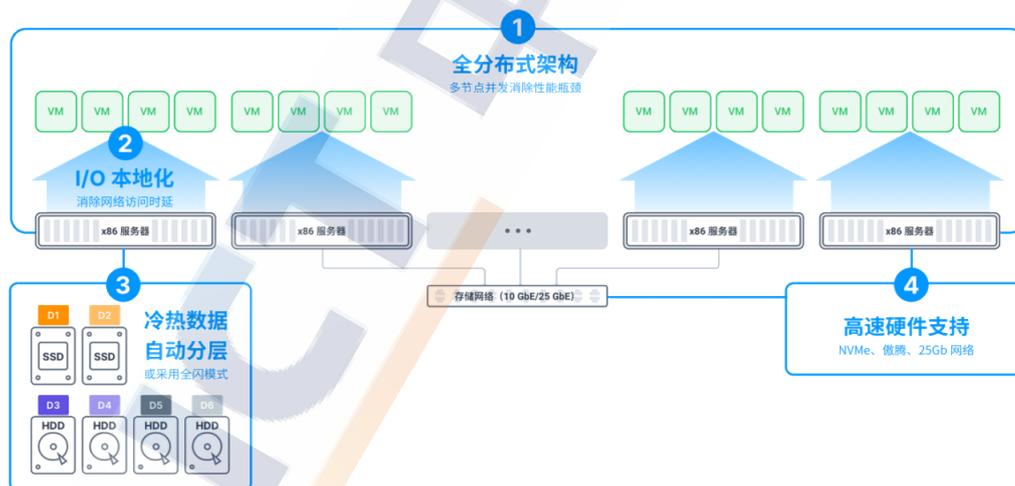
来源：北京志凌海纳科技有限公司

图 10 超融合基础架构图

超融合基础架构除对计算、存储、网络等基础元素进行虚拟化外，通常还会包含诸多 IT 架构管理功能，多个单元设备可以通过网络聚合起来，实现模块的无缝横向扩展，形成统一资源池。

超融合基础架构通过为企业客户提供一种基于通用硬件平台的计算存储融合解决方案，为用户实现可扩展的 IT 基础架构，提供高效、灵活、可靠的存储服务。对于用户来讲，超融合基础架构的主要价值如下：

以简洁的架构提供高可用方案。超融合由于其融合部署架构，可有效协调虚拟化和存储高可用联动的问题，从而以非常简洁的架构提供不同级别的高可用方案。



来源：北京志凌海纳科技有限公司

图 11 融合基础架构特征

存储系统整体性能的大幅提升。分布式架构提升了系统整体的聚合性能，可以在不改变硬件配置下进一步降低访问延迟。

扩展性大幅提升。超融合基础架构的核心分布式存储在可扩展性上有了本质的提升，包括如下特点：支持少量节点起步，支持硬件部件及节点级扩容，容量自动均衡，异构节点支持，卷级别存储策略等。

采购成本和总体拥有成本降低。在使用成本方面，服务器+超融合软件（或超融合一体机）的采购成本有大幅度的降低。除采购成本外，超融合系统在总体拥有成本上有更大的优势。SmartX 在某证券客户的支撑案例中，超融合解决方案协助客户提升 15% 的资源利用率、降低 60% 的运维工作量、同时每年的 IT 采购成本降低 50%。

（三）存储协议演进

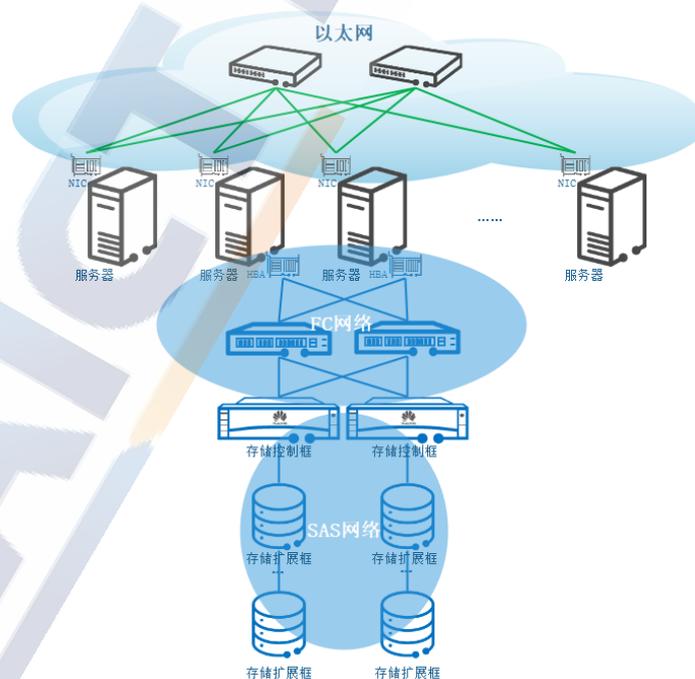
在存储系统中，HDD 磁盘和早期 SSD 磁盘的传输协议一般采用 AHCI（高级主机控制器接口，Advanced Host Controller Interface）。AHCI 为单队列模式，主机和 HDD/SSD 之间通过单队列进行数据交互。对于 HDD 这种慢速设备来说，主要瓶颈在存储设备，而非 AHCI 协议。不同于 HDD 的顺序读写特点，SSD 可以同时从多个不同位置读取数据，具有高并发性。AHCI 的单队列模式成为限制 SSD 并发性的瓶颈。随着存储介质的演进，SSD 盘的 IO 带宽越来越大，访问延时越来越低。AHCI 协议已经不能满足高性能和低延时 SSD 的需求，因此，存储系统迫切需要更快、更高效的协议和接口，NVMe（NVM Express）协议应运而生。

NVMe 协议旨在提高吞吐量和 IOPS，同时降低延迟。基于 NVMe

的驱动器可实现高达 16Gbps 的吞吐量，且当前供应商正在推动 32 Gbps 或更高的吞吐量产品的应用。在 IO 方面，许多基于 NVMe 的驱动器，其 IOPS 可以超过 50 万，部分可提供 150 万、200 万甚至 1000 万 IOPS。与此同时，延迟持续下降，许多驱动器的延迟低于 20 微秒，部分低于 10 微秒。

凭借卓越的性能，NVMe SSD 在 2018-2023 年间以 38% 的复合年增长率持续增长，2020 年占据企业级 SSD 出货量的 55% 以上，云客户、OEM 厂商和企业均将 NVMe SSD 作为存储设备的首选。

在网络协议层，30 年来，存储网络都是以 SCSI（小型计算机系统接口，Small Computer System Interface）协议为基础框架，前端传输网络层一直以 FC（光纤通道，Fiber Channel）网络为主，后端则以 SAS（串行 SCSI 技术，Serial Attached SCSI）网络为主，这构成了服务器间以 IP 为主要互联手段的 IP 存储网络。

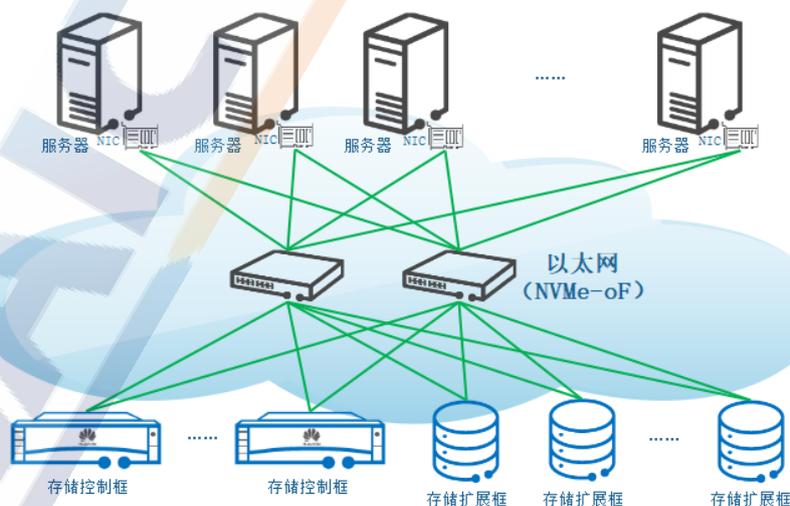


来源：华为技术有限公司

图 12 基于FC和SAS协议的经典存储网络

2010 年后，随着闪存介质的普及，SCSI 协议框架对性能的限制也越来越突显出来。NVMe 和 NVMe Over Fabric (NVMe-oF) 技术的出现打破了这些限制，面向高性能介质设计的多队列模型更能发挥闪存介质的性能。NVMe-oF 推动 IP 化、低时延化，基于 IP 网络的 NVMe-oF 技术不但使得前端存储网络可以基于 IP 直接与本地局域网连接，甚至可以直连广域网；同时，利用 NVMe-oF 技术小于 10us 的超低附加时延，使得替换后端 SAS 网络也成为了可能，整个数据中心可以基于统一的以太网来构建；一方面降低整个数据中心的建设成本，一方面降低独立存储网络的运维成本，同时有利于云及大数据应用环境下的数据共享。

华为 NoF+存储网络是 NVMe-oF 技术一个典型应用，在 OLTP/OLAP(联机事务/分析处理过程)场景，IOPS 最高提升 85%，拥塞时延最大降低 46%，端到端故障切换时间<1s，充分体现新协议对网络性能的巨大提升。



来源：华为技术有限公司

图 13 基于NVMe协议的IP存储网络

（四）应用模式演进

云存储是基于云计算相关技术延伸和发展而来的全新应用模式。云存储的内核是应用软件与存储设备相结合，通过应用软件来实现存储设备向存储服务的转变。本质上，云存储是一种服务，是由多个存储设备和服务器所构成的集合体。

通过解耦，计算与存储可以在云数据中心独立扩展，提供调度和资源共享的灵活性，提高资源使用效率，降低成本。其次，计算和存储可以更加灵活的针对不同负载进行优化。网络技术的高速发展，个位数微秒的延时、百 G 级别的带宽，使得计算和存储分离、分布式存储等架构在稳定性和性能等领域变得更加高效。

对于行业用户而言，云存储的价值主要体现在以下三个方面：

- 提供诸如块存储和文件存储等标准化的存储方式。云存储提供标准化接口，使客户能够直接迁移数据，避免应用的大量修改。
- 存储系统的服务化。在云计算时代，云存储将存储系统演化成为一种云服务，用户只需要关心自己的业务逻辑即可。如百度云 BOS 服务，相较于传统存储服务，数据处理流程的易用性提升了 95.2%，满足了一站式存储和处理的诉求。
- 存储系统的开放化。云存储服务提供众多管理与控制的 API（应用程序接口，Application Programming Interface），通过

开放接口，使得用户可以通过编程调用 API 管理与监控存储资源，实现跨平台的管理。

云存储在发展的过程中也面临着诸多挑战，为满足云时代发展的需求，新一代存储需要不断演进。未来云存储技术主要呈现出以下三点趋势：

集成设计能力不断加强。数据库、大数据处理和分析、人工智能、容器等领域具有自身的特点，存储技术针对以上典型场景也需进行相应的集成设计，结合场景的特点进行适配和优化，提高与场景结合的端到端优化能力和存储的效率。

云上大规模运维能力的不断提升。随着越来越多的企业不断上云，云服务商的运维方式也发生了巨大的变化，需要在大规模、高复杂度下，保障云服务的高可用。以阿里云对象存储 OSS 为例，在数据迁移、数据湖、数据备份及归档方面，可以提供 12 个 9 的数据持久性，99.995% 的数据可用性。

云存储产品形态不断丰富。由于数据访问方式以及业务场景的不同，云存储不断丰富自身的产品形态。阿里云、腾讯云、百度云等头部厂商均构建了对象存储、块存储、文件存储、表格存储等多种形态的产品。

（五）运维模式演进

随着数字化转型的加速，企业用户需要更加敏捷地响应快速变化的市场需求。不仅是业务模式，IT 基础架构的革新在其数字化转型中也是非常关键的一部分。现代化的应用、多数据中心、多云及

边缘等趋势在加速业务的同时，也对运营管理带来了巨大的压力，靠人力投入完全管控存储系统变得不可维系。运维团队急需新技术协助，智能运维平台应运而生。

智能运维指利用大数据和机器学习等方法提高运维的自动化、效率及故障自愈的技术，同时利用机器智能从运维数据中持续挖掘深层信息，是一种结合工程能力与算法能力的综合性科学。智能运维常用于集群自动扩缩容、服务变更、库存管理等日常管理事务以及异常定位、根因分析、系统自愈等异常处理事务。

智能运维架构主要由两部分组成，集成了众多传感器的服务器、存储设备和基于云端的智能运维平台。通过数据收集、转换及训练对基础架构中可能出现的问题做出预测并提供建议。

智能运维技术的主要特点有：

- 主动式问题处理：智能运维平台能够对简单问题自动进行处理；对于复杂的问题，能够自动开具工单，通过支持中心分派工程师主动联系用户，帮助用户发现和解决问题。
- 智能需求预测：通过人工智能、机器学习等技术，实现对存储未来容量、性能的提前预测，便于客户提前规划。
- 智能风险预防：除了预测分析外，领先的智能运维平台还能做到对风险的预防，主动优化 IT 基础架构，通过黑、白名单等功能特性预防已经发现的问题，并给予可行的建议。
- 云上管理：基于云创建的智能运维平台，可以让 IT 运营团队随时随地以任意终端访问智能运维平台，查看运行状态；云

端自动升级，在新特性上线时，不会影响用户的日常使用。

随着下一代数据存储技术的发展和成熟，性能不再是困扰用户的首要难题，各存储供应商也逐渐将注意力放到智能运维上，希望能够提供更丰富的功能、更好的使用体验。未来智能运维平台的发展呈现出以下两点趋势。

一站式分析。存储的智能运维平台将逐步成为整个 IT 基础架构运维平台的核心。通过将服务器、网络、虚拟化应用逐步接入智能运维平台，能够提供更丰富的 IT 运行状态数据，帮助智能运维平台判断故障和瓶颈。

离线智能。由于监管因素的限制，部分企事业单位无法使用基于云的智能运维平台。存储厂商开始考虑提供离线智能运维平台，通过集成智能运维平台知识库，提供基于本地的性能瓶颈分析、容量和性能的预测、硬盘故障预警等功能。

现今，智能运维已成为热门存储技术，并在实践中逐步应用，提供更好的用户体验。如新华三的 InfoSight 智能运维系统，企业用户通过使用智能运维系统，86%的问题得到自动解决；平均给出解决方案的时间小于 1 分钟；93%的问题能够自动提交工单；用户满意度高达到 98.2%。

三、下一代数据存储技术赋能数字化转型

下一代数据存储技术深刻的改变着存储产品形态，伴随着需求的不断变化，不断演进出了新特征、新指标，使得存储系统能够更好地满足全行业客户的需求。

本章将以数据湖、数据备份及实时大数据分析三个解决方案为例阐述下一代数据存储技术如何对传统的存储产品的存储、使用、备份全流程带来革新，赋能企业数字化转型。

（一）异构数据统一管理推动数据湖产品演进

随着数据量的爆发式增长，许多企业产生数据的量级由原有的 TB 级别迅速的提升到 PB 甚至 EB 级别。企业付出成本来存储这些数据的同时自然也想通过挖掘数据信息辅助商业决策，提升管理效率。

大数据经过了多年发展，存储需求的不断变化及以云存储、智能管理为代表的下一代数据存储技术的成熟，推动了数据湖的不断演进。不断变化的业务需求对数据湖提出了以下需求：

- **统一调度：**构建统一数据底座，把数据放在合适的位置上，同时提供覆盖存储网络的发放自动化、拓扑自动化和性能分析自动化服务。
- **按需流动：**基于数据冷热和应用负载分析，使得数据按需流动，满足不同生命周期阶段性能及成本诉求。
- **多云对接：**通过 API、脚本和插件等多种方式对接云管平台，确保融入客户流程不改变客户习惯。
- **存储与计算分离：**存算分离能够有效降低计算资源与存储资源扩展需求不平衡场景下的运维成本与硬件成本，成为数据湖存储的必要选项。
- **数据全周期管理：**提供基础的数据访问能力，同时根据实际

的业务与成本提供数据分层管理与归档备份的能力。

- **数据多协议支持：**数据的存入与使用需要适配数据湖领域的各种应用场景，比如支持 HDFS 协议、SQL 语义、对象存储 S3 协议等。

相较于传统解决方案，下一代数据存储技术为数据湖带来的最大改变就是企业用户无需关心存入数据的类型，系统自行选择一种最优形式进行存储。同时，智能化的统一运维管理平台使得系统能够存储海量异构数据，构筑统一的数据底座，提供统一存储访问接口，解决系统间数据孤岛、各类应用统一访问问题，真正做到“一个数据中心一套存储”。

（二）多级存储介质助力实时分析能力构建

IDC 预测，到 2021 年，全球 2000 个机构中的 60-70% 将至少有一个关键的实时型工作负载任务，如金融、医疗领域的实时交易分析场景。这些应用场景中，海量数据的分析、推理模型的演进等均需要在内存进行大量数据的实时运算，并为用户提供实时的响应服务，支撑其高效数据价值的挖掘。

以非易失性内存为代表的存储级内存（SCM）和闪存介质的出现填补了实时大数据分析解决方案缺少大容量、高性能存储介质的空白。通过由 DRAM、SCM、SSD、内存型网络构成的多级存储架构，能够为上层业务提供极致性能的服务，在多节点间实现多级介质的全局共享及大内存容量的无感知扩展；同时，SCM 介质层能够在近计算侧实现元数据/数据的高速本地化访问；最后，在整个存储

系统中实现多种介质的多级资源池并进行统一管理、调度，提供统一的访问接口，发挥企业级内存和闪存介质成本优势，达成存储系统总体拥有成本最优。

以存储级内存为中心的数据存储，将为实时的海量数据分析、实时应用等提供大容量、高 TCO 的共享内存介质层，同时具备大规模扩展、高可靠等关键特性，为上层实时分析业务提供高效、可靠的数据底座。

（三）云存储备份简化数据安全实践路径

数据备份是保证企业数据安全的重要手段，往往指为防止系统出现操作失误或系统故障导致数据丢失，而将全部或部分数据集合从应用主机的硬盘或阵列复制到其它的存储介质的过程。

随着数据量的急剧增长，企业陷入非结构化数据溢出的危险境地。问题不在于企业购置容量来存储全部数据，而是如何以低成本高效率的方式妥善管理数据，尤其是长期数据保留，以创造商业价值、释放数据潜能。如何在确保数据可访问性和安全性不受影响的同时，使企业获得更优的数据管理和可见性，是企业 IT 部门面临的主要难题。

云存储的出现为数据备份提供了新形式，通过将数据备份在云上，企业能够突破地域和设备的限制，实现对同一备份的获取。相较于传统备份，云存储的优势主要有：

首先，备份场景多样，集中管控，实现多台云主机的集中配置管控。

其次，简单易用，全自动运维。支持将备份数据托管到云上备份仓库，无需担心硬件配置、集群扩展等问题。控制台可自动推送备份代理，无需手动安装，实现全自动运维。

再次，高重删压缩比，节省成本。备份服务多采用重删、压缩技术，可有效降低云端存储空间，减少成本投入。

最后，备份数据安全可用。支持全量、增量以及日志备份，备份数据可实现快速恢复，复原时间目标（RPO）大大降低，同时凭借全自动数据加密校验等优势，让备份数据更加安全。

四、下一代数据存储技术实施建议

存储系统作为IT系统的底层基础架构，存储技术进一步发展和推广对于整个信息产业具有重大意义。在数字化转型过程中，存储系统作为底层基础架构，其改造和实施过程需被重点关注。

本章力图分析梳理存储系统改造中应当遵循的原则和策略，从而指导相关人员从业务设计和实际操作等方面完成意识的认知和过渡。

（一）指导原则

在下一代数据存储技术的规划实施过程中，应遵循如下指导原则。

1. 适度的技术先进性

衡量存储系统的技术指标包括性能、可靠性、数据效率、可扩展性等，采用下一代数据存储技术可以很好地利用技术迭代的红利

帮构建性能更快、安全可靠、空间利用率更高、可扩展性更强的存储系统，是IT技术发展的趋势。

但同时我们也应注意，新技术的出现往往会经历萌芽、增长、低谷、复苏等阶段，才会被广泛应用，过早投入新技术的使用，会产生更高的试错成本。所以，采用新技术应该适度，充分考虑自身的使用需求、技术的发展阶段、技术与周边生态方案的对接等因素。

2.可控的成本

在采用新技术的过程中，也要充分考虑成本的影响，主要是总体拥有成本（TCO, Total Cost of Ownership）。它既包括前期的采购成本，也包括后期使用过程中产生的成本，具体如下：

- 采购成本：包括硬件成本、软件成本、服务成本，以及因此产生的机房空间、组网设备、应用二次开发等成本。
- 使用成本：包括设备使用产生的人力、耗电、散热、带宽占用、网络流量等成本。

采用新技术需要综合考虑存储系统的TCO成本和由此产生的所有收益，衡量最终的投资回报率。

3.可维护性

新技术往往意味着对传统运营体系的改变，所以要关注系统的可维护性，避免产生高昂的学习成本，或系统切换产生的业务风险，具体要考虑如下因素：

- 安装部署：技术或产品的安装部署是否方便，过程需要多少人力和时间，是否需要调整现有环境，是否对人员的技术能力有要求等；
- 管理配置：技术或产品的管理配置是否简单，是否提供图形化界面，常用功能操作是否方便；
- 故障处理：出现故障时，是否能够快速定位问题并恢复业务，对人员能力的要求如何，故障处理是否需要依赖于其他环境；
- 升级扩容：需要版本升级或能力扩容时，对现有系统的影响如何，是否需要停机，停机窗口多长，对性能影响如何；

4. 兼容性

存储架构是数据存储的基础，确定存储架构时除了要考察存储的扩展性和稳定性外，要充分考察业务应用对存储兼容性的需求，其中包括存储协议的兼容性、存量设备的兼容性、存储对云平台的兼容性等，明确相关问题，需要提前评估业务应用的现状及未来的发展趋势。具体如下：

- 协议兼容。目前云存储厂商大都兼容了 AWS S3 的常用接口，以方便基于 AWS S3 协议开发的应用能够快速对接到自己的存储系统。
- 云平台兼容。现有的存储软件设计很多都是基于公有云的场景做的优化，对于私有云的场景很多地方都需要优化，要做到对于各种场景灵活可配置，而不是一套程序对应各种私有化场景。

- 存量设备的兼容性。对于已购买的存储设备，需要考虑新购存储设备是否能与旧设备兼容，或考虑其他利旧方案。例如将旧设备作为新设备的二级存储，定期将新设备中的数据备份到旧设备中。

（二）实施策略

1. 产品选型

在选型过程中首先应根据业务需求确定存储形态。如果是传统关系型数据库类应用，对数据可靠性和业务可用性要求较高，集中式存储阵列依然在此类应用场景中广泛使用。如果是 NoSQL 类数据库，或其他对于存储性能和容量扩展性要求较高的应用场景，则分布式存储架构更容易满足未来业务的增长需求。

接下来还需考虑存储的性能需求。除了存储设备本身架构设计的因素外，存储介质可能更直接影响块存储最终能够达到的性能。全闪存储节点可提供更低的时延、更高的 IOPS、更高的可靠性和性能，适合于对性能要求较高的核心业务系统。在性能需求一般的业务场景，混闪的配置能提供更优的性价比。

完成存储形态和存储介质的选型后，对于分布式架构的存储设备，通用模块的选型主要基于以下几个方面：

- 网络传输设备：存储的网络环境目前主流仍然是 10Gbps 网络，但随着终端网络带宽的不断提高，未来对存储数据在网络上传输需求的带宽也会越来越大，基于未来的发展应优先考虑相对成熟的 25Gbps、100Gbps 网络。

- **业务逻辑层设备：**此层设备主要工作是处理存储业务产品功能逻辑，主要资源消耗点在 CPU，参数上更多核心更高主频可以带来更好的处理性能。
- **元数据设备：**元数据作为整个存储系统的一个核心部分，整体性能的高低直接影响了存储的响应能力，设备选择上首先要考虑基于 NVMe 协议的全闪存用于存放海量元数据，同时较大的内存也可以进一步提升元数据处理的响应速度，CPU 只需选择主流中低端产品即可。

2. 业务规划

业务规划主要考虑容量的规划和功能的规划。

容量规划。考虑不同的场景、性能需求、容量需求需搭建的集群规模，以及后续扩容，置换的规划。更应考虑到未来业务增长，在给业务划分容量时需预留足够的空间。同时还需要综合考虑业务的需求，划分不同的存储池，做到业务隔离并防止资源争抢。

功能规划。软件功能的规划，伴随着新硬件产生和技术的进步，软件功能也要有相应的升级。

3. 部署规划

部署规划是结合业务的场景和可靠性要求等条件，对机房、机柜、交换机、网络等设备进行统筹考虑。存储服务的部署要结合实际物理条件综合考虑部署架构以实现服务的高可用，需要从分散性、高性能、扩展性及可维护性进行规划。

分散性：存储服务的核心就是数据的可靠性与服务的可用性，

基于这两点所有服务的组件部署都要考虑到多可用区（Availability Zone, AZ）的部署方式，确保所有服务集群在单个 AZ 故障时依然有集群在线提供服务，同时后端存储服务优先要考虑三 AZ 部署（典型存储服务为三副本）。

高性能：集群初建时业务体量尚未成型，此时选用大容量介质建设目标容量集群会因为介质数量太少，在突发业务压力下导致单个存储介质到达性能极限，进而影响了整个存储集群的性能，基于这种场景可以选用多个相对容量较小的存储介质提高集群的整体性能。

扩展性：设计之初各服务模块应充分考虑横向扩展能力，扩展模型是否足够简单，确保可以通过扩展满足未来业务的需求。

可维护性：传统存储运维面对万台规模的存储集群，从变更到日常故障巡检都要消耗大量的人力，运维向智能化转变是未来发展的方向，通过智能运维来大幅提高存储系统的自运维能力，降低运维成本。将运维人员从繁杂的工作中解放出来，将更多的精力集中在运维系统的开发和完善上。

五、总结与展望

数字经济时代，信息技术已经渗透到生活的方方面面，数据存储系统作为信息化系统的基础设施，构建一套稳定、高效、满足未来业务发展需求的数据存储系统将是企业和组织夯实数据底座、挖掘数据价值、释放数据潜能的关键。人工智能、大数据、5G 等新技术发展使得数据量指数级增长，数据激增带来存储计算需求的飞速

增长，为存储产业带来了新需求、新挑战和新机遇。

新机遇、新挑战不仅驱动了存储技术革新，也推动了存储产业发展，为存储产业带来了翻天覆地的变化，存储产业呈现出以下三点趋势：

一是底层软硬件平台多元化。芯片、服务器、存储介质等通用技术的发展改变了存储架构，为存储行业的发展奠定了基础。诸如全闪存储、混闪存储、NVMe 协议、软件定义存储等新技术的出现为厂商提供了多样的选择。在芯片领域也出现了众多供应商，如鲲鹏、飞腾、海光、兆芯、申威等等。存储软硬件的通用化使得存储可以选择更多的软硬件技术，充分利用新型硬件红利的时候，也让用户享有更多的选择权。

二是基础设施云化。IDC 发布的《IDC FutureScape: 全球云计算 2020 年预测——中国启示》预测，到 2021 年，中国 90% 以上的企业将依赖于本地/专属私有云、多个公有云和遗留平台的组合。在云计算基础设施中，通过虚拟化存储、计算和网络，将通用的物理资源整合池化成逻辑资源，并对外提供服务，以软件定义的方式快速、敏捷的将基础设施资源以服务的方式提供给用户，满足企业在数字化转型中对敏捷、灵活、快速、高效 IT 系统的需求。

三是上下游生态融合。未来数据存储系统必须具备高性能、高可靠性、高扩展性等多种能力。既可以满足数据库等传统核心业务需求，也可以支持大数据分析、虚拟化、容器、云平台等新业务应用。在云计算、大数据技术的驱动下，不同类型企业的应用场景不

断细分，对性能和可靠性的要求越来越高，企业用户对存储解决方案提出了全新要求。这需要供应商与产业生态深度融合，根据业务需求提供个性化的解决方案。

随着数字经济时代数据的价值被进一步重视和挖掘，及全闪存存储、企业级内存为代表的下一代数据存储技术的应用和发展，各行业对存储系统的需求不断扩大，存储产业欣欣向荣，具有非常广阔的发展空间。

同时我们也应认识到，存储产业的发展需要供给侧、需求侧的协同推进。提升企业用户对下一代数据存储技术的认知，推动下一代数据存储技术的成熟是存储产业可持续发展的关键。需要通过强化政策布局、积极组织交流研讨等手段，加强宣传，鼓励产学研合作，不断提升技术成熟度。同时以产品评估评测为抓手，梳理典型业务场景，形成技术指南和案例集，降低行业技术门槛，不断优化技术水平。

我们相信，数据存储技术作为现代信息技术的底座，对加快数字化发展，建设数字中国有着至关重要的作用。下一代数据存储技术的广泛推广和应用能够解放和发展数字化生产力，推动数字经济与实体经济深度融合，夯实数据底座，提升我国的数据治理能力，实现高质量发展。

参考文献

- [1] 2020Q3 IDC Enterprise Storage System Market Overview. IDC.
- [2] 2020Q3 IDC PRC Software-defined Storage and Hyperconverged System Market Overview. IDC.
- [3] 《IDC：2025 年中国将拥有全球最大的数据圈》. IDC.
- [4] 《IDC FutuerScapes 2019》. IDC.
- [5] 《IDC FutureScape：全球云计算 2020 年预测——中国启示》. IDC.

中国信息通信研究院 云计算与大数据研究所

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：18795986568

传真：010-62304980

网址：www.caict.ac.cn

