

中国分布式存储市场研究报告(2023)



行业报告资源群



《微信扫码 长期有策

- 1. 进群福利:进群即极万份行业研究、管理方案及其他学习贡源,直接打包下载
- 2. 行研構施: 每日分享6+份行业精施报告及3个行业主题研究资料
- 2. 查询报告: 行研资料免费帮助查询下载
- 4. 严禁广告:本群仅限行业报告交流,禁止一切无关衡量

CONTENTS 目 录

报告概览	01
第一章	
厘清概念,把握分布式存储的核心要点	
一、分布式存储是什么	04
二、分布式存储的分类	06
三、分布式存储的优势	08
四、分布式存储的应用领域	09
第二章	
量化研究,明确分布式存储的发展现状	
一、2020-2022年中国分布式存储市场规模	14
二、2020-2022年中国分布式存储市场结构	15
三、2022年中国分布式存储主要场景市场	17
四、2022年中国分布式存储区域市场	18

CONTENTS 目 录

第三章				
瞄准企业,洞察分布式存储的厂商动态				
一、2022年中国分布式存储市场重大事件	20			
二、2022年中国分布式存储厂商竞争力情况	21			
三、2022年中国分布式存储部分厂商发展情况	23			
四、2022年中国分布式存储部分细分市场情况	24			
第四章				
谋篇布局, 预测分布式存储的未来发展				
一、2023-2025年中国分布式存储市场规模预测	27			
二、驱动及阻碍因素分析				
三、分布式存储的发展趋势				
第五章				
建言献策,助力分布式存储的全面跃升	34			
企业: 根植应用个性化需求				
投资机构:关注自研路线厂商				

政府及第三方协会: 细化多维度标准

报告概览

1010 1010

分布式存储与集中式 存储相互配合,满足 了多个场景用户的个 性化需求



2022年,分布式存储 在云原生、高性能计 算(HPC)等领域表 现突出



2020-2022年,中国 分布式存储市场规模 快速发展;2022年文 件存储占比最大 分布式存储指基于分布式架构,通过软硬件协同,依托高效网络连接多个节点来实现存储功能的IT产品和服务。在传统存储面临拓展性有限、结构化/非结构化数据兼容性不佳、I/O性能与成本不经济、可靠性不高和面向新应用(如容器)支持能力有待提升、用户体验不佳等问题的背景下,分布式存储一定程度上与传统存储实现了较好配合

相比传统的集中式存储,分布式存储在以下领域表现 突出:

云原生:指"生在云上、长在云上"的政企业务场景, 典型的有容器和微服务等

高性能计算(HPC): 支撑每秒百亿亿次浮点计算 (EFLOPS)场景的分布式存储

融媒体:指文件大且I/O延迟要求低的元宇宙、虚拟数字人和虚拟现实等场景

智能备份: 指视频监控、医疗影像和智能制造等场景

从规模看,近三年中国分布式存储市场规模均保持30%以上的高速增长,2022年市场规模达104.2亿元 从结构看,2022年,文件存储占比最高,一体机、分 布式混闪、虚拟化融合在各自领域占比较高 从场景看,HPC在2022年中国分布式存储主要场景市

从场景看,HPC在2022年中国分布式存储主要场景市 场中居于重要地位

从区域看,2022年华北、华东地区占据半壁江山,京 津冀、长三角市场领跑

报告概览



中国分布式存储厂商 加快构筑自身竞争力 2022年中国分布式存储市场中,

领导者有华为、中科曙光、浪潮、新华三 **挑战者**有XSKY、 SmartX、 NetApp、 Dell&EMC、焱融 科技

跟随者有同有、杉岩和中国电子云

可期待者有深信服和联想

中科曙光在分布式存储**教育科研和气象**市场份额居首位,

SmartX在分布式块存储金融市场居首位



预计到2025年,中国 式存储市场规模有望 达到211.4亿元 驱动因素: 数据爆发式增长、可靠性要求提升、硬

件性能提升、云原生加速业务创新

阻碍因素:资金人才支撑不足、标准滞后缺失

主要发展趋势:分布式块、文件和对象存储快速发

展,分布式全闪存适用范围扩展,在云环境、云原

生场景普及, 自研线路具有较好发展前景



企业、投资机构和政 府部门需形成合力 企业: 根植应用个性化需求,根据不同应用对存储性

能的要求,提供高适配性产品

投资机构:关注自研路线厂商,建议关注具备面向应

用需求的架构自研、技术自研的厂商

政府部门:细化多维度标准,统筹产学研用各方力量,

推动相关评测认证体系的推广宣贯和评测结果的发布



厘清概念,

把握分布式存储的核心要点



一、分布式存储是什么

随着数字中国加速落地,数据作为数字经济的核心部分,具有基础战略资源和关键生产要素的双重作用,对于建设网络强国、科技强国、质量强国都具有重要意义。

在新一代信息技术不断迭代升级的背景下,全球数据量呈现爆发式增长。作为全球数据生产大国,中国数据量也面临急剧增长的态势。数据量的急剧扩展,愈发凸显了数据存储及管理的重要性。数据存储作为数据管理的必要手段,其发展迎来了新机遇。

传统存储面临拓展性有限、结构化/非结构化数据兼容性不佳、数据读写任务(I/O)性能与成本不经济、可靠性不高和面向新应用(如容器)支持能力有待提升等问题,用户体验不佳。在此背景下,分布式存储近年来迅速发展,一定程度上与传统存储实现了较好配合,满足了多个场景用户的个性化需求。

分布式存储指基于分布式架构,通过软硬件协同,依托 高效网络连接多个节点来实现存储功能的IT产品和服务。

一一 分布式架构:与集中式架构不同,分布式架构没有负责所有业务的中心节点,而是由多个可相互通信、协调的普通节点构成集群,对外提供服务。

图 1 集中式架构与分布式架构



表 1 常见的分布式存储软硬件

The state of the s							
类型	名称	特点					
硬件	CPU	相比服务器的CPU,侧重计算数据存放位置、备份功能					
	缓存	常规操作无需更多内存,但在执行恢复操作时,内存负载大					
	硬盘/闪存盘	考虑需求和成本,配置硬盘和闪存盘的数量					
	网络	每个存储节点至少2个千兆网卡; PB级集群则多为一个10G网卡					
软件	操作系统	调度硬件资源构成灵活的存储资源池					
	服务软件	提供副本、纠删码(EC)、远程复制、缓存加速等功能					
	其他软件	资源、硬件状态监控;简化配置、在线扩容等功能					

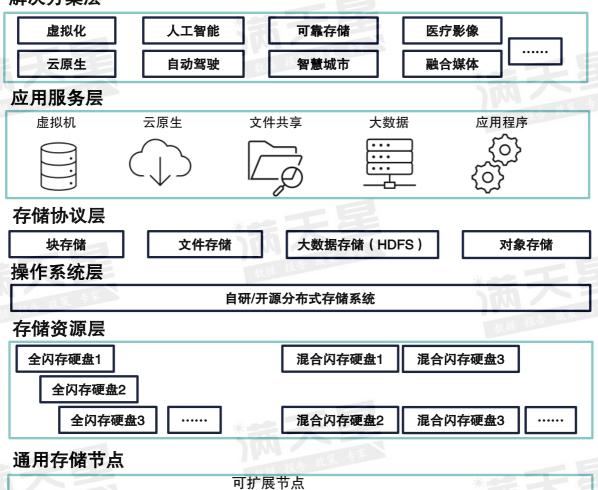
高效网络:并不是传统的TCP/IP网络,而是采用存算分离的设计思路,不通过CPU,直接由内存传输到智能网卡完成I/O的远程直接地址访问网络 (RDMA)。

图 2 分布式存储架构图

解决方案层

存储节点1

存储节点2



存储节点3

存储节点4

二、分布式存储的分类

根据场景特点和需求,2022年中国分布式存储产品共有四种主要的分类方式:

按存储对象,可分为分布式块存储、分布式文件存储、分布式对象存储、分布式统一存储:

按产品形态,可分为一体机、纯硬件和纯软件;

按存储介质,可分为分布式全闪存、分布式混闪;

按部署方式,可分为虚拟化融合方式、容器融合方式和分离方式。

(一)按存储对象分类

分布式块存储:指将各存储节点的硬盘、闪存盘等存储硬件资源通过划分逻辑卷(LVM)、创建独立冗余存储阵列(Raid)、逻辑分区等方式,由物理存储资源转化为面向需求的逻辑块,提高使用效率。

分布式文件存储:指基于文件系统对各存储节点的数据进行存储,形成目录、子目录、文件,常见的分布式文件存储应用有Ceph、HDFS、GFS、 FastDFS、GridFS、mogileFS、TFS等。

图 4 分布式文件存储示意图



分布式对象存储:指各存储节点由标识符、数据和元数据的对象数据构成。其中标识符在该存储系统中唯一,用于区分不同存储区域;元数据将提取数据特征,便于快速检索。分布式对象存储相比块存储和文件存储,数据查找和提取的效率大大提升,适用于文本、音频、视频等非结构化数据。

分布式统一存储:指各存储节点同时支持块、文件和对象三大类型数据,满足虚拟化、云平台和容器平台等新兴存储需求,为用户提供服务典型应用和计算平台的统一存储资源池。

(二)按产品形态分类

一体机:从设备层面将各节点存储资源进行融合优化,同时软硬件为同一厂商生产、一体交付,适配度高,可有效缓解单个节点或设备短板导致整体系统功能受影响等问题。扩展能力较强,且可在线增加节点,对前段业务完全透明。不足之处是价格偏高。

纯硬件:指交付形态为磁盘阵列、闪存盘集群等硬件资源的分布式存储产品,适配的软件分为自研和开源二次开发两种路线。纯硬件交付模式成本高,但可靠性较高,适用于重要敏感数据存储场景。

纯软件:指交付形态为定制化应用软件、平台授权码等的分布式存储产品,一般应用于优化存储硬件的场景,如老旧数据中心改扩建等。纯软件交付定制化空间大、成本低、交付周期短,但扩容及存储硬件兼容性问题一定程度上也会影响运行效率。

(三)按存储介质分类

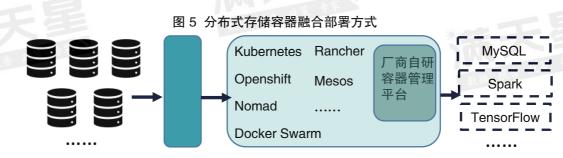
分布式全闪存:指各存储节点完全由固态硬盘(SSD)构成,主流的适配接口为 NVNe,每秒读写次数(IOPS)可跃升至百万级别,相比传统的机械硬盘(HDD) 提升了近千倍。缺点是价格高昂,且受限于系统总线协议和其他部件,分布式全闪存 性能较难完全发挥。

分布式混闪:指各存储节点由SSD、HDD等构成。尽管性能不及分布式全闪存,但可针对场景需求进行个性化定制,从而最大程度上均衡成本和性能,是目前主流的分布式存储产品。

(四)按部署方式分类

虚拟化融合:指在分布式存储(块存储为主)的架构基础上,利用服务器虚拟化的隔离机制,实现存储和服务器虚拟化在同一硬件节点上的部署。这种架构的优势在于整体架构更为简单,并且节省了硬件成本,缺点是计算和存储需要同时扩展,不适用于计算和存储应用不均衡的场景。

容器融合: 面对容器化持久化存储日益增长的需求和Kubernetes特有的管理架构,专门针对此类场景的分布式存储产品也开始涌现。容器融合的分布式存储产品对 K8s 集群内的存储资源进行整合与管理,不仅可以通过与容器融合部署降低成本,简化系统架构,还可以无缝融入 K8s 原生的开发和运维体系,更加符合 K8s 运维团队的使用习惯。



分布式存储物理资源 容器插件

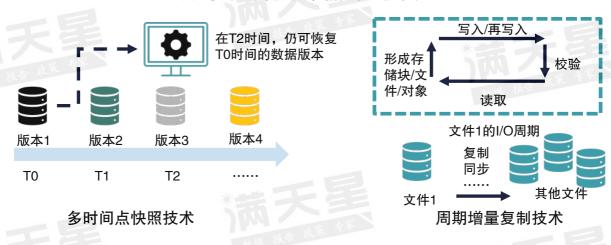
开源/自研容器管理平台 基于容器的各类应用

分离: 指各存储节点资源与应用分离,虽然架构更加复杂并需要更多的硬件节点,但 方案更为灵活,并适用更多场景,尤其适合大容量的数据存储,以及从裸金属、虚拟 化到容器的不同计算节点的混合资源池。

三、分布式存储的优势

高可靠性:分布式存储的数据采取存放在多个存储节点中的全冗余部署,通过多时间点快照、周期增量复制两大核心技术,可实现在一定时间间隔内,对各版本数据的保存,并且同时进行恢复,可帮助分析和研究,避免类似灾难的再次发生。

图 6 多时间点快照、周期增量复制技术原理



高效作业:面对可划分为若干个并行运行的子任务的存储任务,分布式存储可将这些子任务分散到不同的存储节点上,使其同时运行作业,从而提高效率。另外,分布式存储系统具有任务负载平衡功能,如果某个存储节点的负载过重,则可把其中一些作业迁移到其他节点去执行,从而减轻该节点的负载,实现整体效率的提升。

高可扩展性:分布式存储横向连接了多个存储节点,可按存储容量和性能的需要,灵活横向拓展(Scale-out)新节点。新存储节点和原系统连接到同一个网络,同时业务的连续性基本不受影响,可基本实现分布式存储系统的总容量和性能无感线性扩展。新存储节点的资源将由分布式存储操作系统管理,进行分配或取消,原有数据可通过复制、镜像、同步等方式,实现在新节点的迁移。

高质量产业生态:分布式存储有开源和自研两种路线,前者是在开源系统基础上进行二次开发,自研则是厂商根据自身技术积累、主要客户群体需求自主研发的分布式存储系统。开源系统的优势在于便捷可用、成熟度高、成本可控,但在定制化服务上不及自研线路厂商。自研路线技术门槛普遍较高,服务能力强,并且自持技术专利可实现迭代,但有待更多案例实践加速其产品和服务的成熟。

系统	Ceph		Swift	HDFS	GlusterFS		
存储方式	统一存储(块、文 件、对象)		对象	文件	统一存储(块、 文件、对象)		
节点间通 信协议	TCP		TCP	TCP	TCP/RDMA		
在线扩容	支持 数据 报告 政策						
冗余备份	支持						
易用性	一键安装,官方文档支持						
单点故障	不存在	不存在		存在	不存在		
适用场景	统一存储 (块、文件、 对象)	对象存储		文件存储	统一存储(块、 文件、对象)		

表 2 2022年热门开源分布式存储系统

四、分布式存储的典型应用场景

2022年,中国分布式存储充分发挥上述优势,与传统的集中式存储相比,在云原生、高性能计算(HPC)、融媒体、智能备份和分布式数据库等领域表现突出。

云原生: 指"生在云上、长在云上"的政企业务场景,典型的有容器和微服务等; 高性能计算(HPC): 支撑气象、基因测序等每秒百亿亿次浮点计算(EFLOPS)场景的分布式存储,通常以PB,乃至EB计量;

融媒体: 指文件大且I/O延迟要求低的元宇宙、虚拟数字人和虚拟现实等场景; 智能备份: 指视频监控、医疗影像和智能制造等场景;

分布式数据库:指具备分布式事务处理能力、可平滑扩展、分布于计算机网络且逻辑上统一的数据库,分布式存储是分布式数据库的核心技术之一。

云原生



相比虚拟机存在占用空间大、运行速度慢、成本高等不足,容器技术更轻量化,成本开销更小,随着Docker、Kubernetes等技术的成熟,容器成为了2022年云原生主流技术手段之一。

当前云原生的特点为同一应用多租户并行,且和其他应用运行在同一个空间,因此需要分布式存储为有状态的容器应用提供弹性可靠的存储资源。



微服务将容器化的资源按需拆分为多个专门服务于应用的载体,在性能和 功耗之间动态平衡,推动云计算进入更为定制化和高效率的阶段。

- 微服务将复杂的应用分解成小而专一、耦合度低并且高度自治的一组服务,每个服务都是很小的应用,需要分布式存储提供持久化存储 微服务之间通过轻量级的通信机制实现彼此之间的互通互联,与分布式

高性能计算(HPC)

HPC指计算能力达到EFLOPS的计算任务, 超算是其典型代表。一般来说, HPC具 有以下几类特点:

- ✓ 采用分布式架构,将多台计算、存储设备聚合在一起并行工作
- ✓ 多协议并存,异构计算、异构存储的现象较为普遍
- ✓ 处理的多为PB乃至EB级别的I/O需求,数据量大且温冷热数据管理要求严格



近年来,台风、洪水、沙尘暴等对经济的危害越来越大,重要会议、重大赛 事等活动举办前,气象预测也发挥着重要的作用。

- f.象预测主要通过数值预测模型实现,已从一个简单的正解方程发展到数 万参数求解的复杂问题,需要分布式存储的各个节点记录中间值
- 目前,气候模式系统将预报提前为60天,下一阶段要实现"无缝隙"气候 预测,产生的海量数据需要分布式存储有效存放、高速提取、定期更新



基因测序应用从科研服务的基因图谱、遗传多样性,到医疗服务的疾病早 筛和治疗都在大幅拓展,对提升医疗智慧水平意义重大。

- 式存储提供PB级别存储空间扩容的性价比更高基因测序过程是一个超高并发的文件转换过程,分布式存储可以较好完



自动驾驶数据存储包含了采集、标注等环节,并且只有通过这些环节形成

- 个模型,才能确保自动驾驶的落地。 L3及以上的自动驾驶汽车对数据的存储需求巨大,一天生成的数据量可 达5TB到8TB,需要分布式存储强大的扩容能力加以支持
- 自动驾驶功能的最终形成需要经过大量的训练,分布式存储可通过多个 存储节点同步向中控系统发包,提高GPU访问数据的效率



2022年,以ChatGPT为代表的大型语言模型爆发,带动AI初步实现了从感

- 知智能到认知智能的新突破,在内容创作上显示出了其价值。
 ✓ 多模态 AI 场景参数规模通常可达到数千亿乃至万亿,带来的数据存储任务巨大,且灾备成本高昂,而分布式存储可有效应对这一问题
- Al训练和版本迭代依赖海量结构化、非结构化数据存储能力,分布式文 件存储具备数据读写和元数据查询功能,适配度较高

融媒体指借助虚拟现实、超高清视频等新兴技术,实现各类媒介载体互补融合的新型 媒体。融媒体的数据存储具有以下共性特点:

- ✓ 对于热门媒体内容,访问密集并发,I/O的速度和延迟都有较高要求
- ✓ 对于不常访问的媒体内容,长期存储的需求显著,并且普遍存储容量较大
- ✓ 媒体内容属于数据资产,具有较高的经济和社会价值,灾备要求高



元宇宙搭建了与物理世界平行的数字世界,离不开存储、网络传输等技术。

分布式存储是支撑数据成为元宇宙资产的关键手段之一。 ✓ 元宇宙的部分数据具有隐私性,分布式存储可保护数据不被篡改,同时将 数据分散至各个存储节点上,多方授权后,才能拿到密钥提取相关数据 元宇宙的资产需要各方共识,而当前实现的途径是分布式存储,各存储节



虚拟数字人指借助虚拟现实技术制作数字化外形的虚拟人物,具备人的外 观、性别和性格,同时拥有一定识别外界环境、并能与人交流互动的能力。

- 据需要有序分散在不同节点,与分布式存储契合度较高
- 在直播等交互性强的环节中,虚拟数字人将产生海量数据,并且数据需要实时流转,分布式存储能够提供海量空间、低延时的I/O能力



虚拟现实以VR、AR为代表,其沉浸式内容相比智能手机、PC等规模更大。 实时交互的要求更高

- 虚拟现实的"身临其境"将扩大用户视场角、提高分辨率,产生PB级数据,需要分布式存储应对大量数据存放的挑战
- 速网络助力实现

智能备份

在传统备份的基础上,智能备份具有高清化、智能化、网络化、数字化等优势,是 保证数据可追溯、提升作业效果、明晰责任边界的重要手段。具有代表性的智能备 份包括:

√ 视频监控:对于热门媒体内容,访问密集并发,I/O的速度和延迟都有较高要求 ✓ 医疗影像:对于不常访问的媒体内容,长期存储的需求显著,且普遍存储容量大 ✓ 智能制造: 媒体内容属于数据资产, 具有较高的经济和社会价值, 灾备要求高



"智慧城市""智慧交通"等工程深入推进,在维护社会治安、 保证生产安全、提升管理水平等起到<mark>了重要作用。</mark>

- 单一的监控数据采集、回放转变为综合智能系统,离不开分布式存储实 现有效的数据存放和实时的数据回传
- 视频监控高清化将带来EB级累积视频数据,加上目标识别、事件检测、 车流分析的需求,需要分布式存储助力扩容和高效I/O



CT、CR、US等设备产生的数字化医学图像信息在现代医疗中处于重要位 医疗图像存储与传输系统(PACS)的数据存储瓶颈亟待解决。

- 三甲医院的年累积数据量通常在数百TB甚至PB,住院和门急诊的电子病 历要求保留时问分别不少于30年、15年,需要具有分布式存储提供海量 存储空间
- 分布式存储也可以满足大中型医院不同科室同时阅片,以及多种医疗设 备同时存储PACS文件的需求



生产环节的海量非结构化数据以图片、文本等数据为主,用于智能制造系

- 亿级别的文件数量、PB级的存储容量、图片像素高,分布式存储能够按
- 需高效满足扩容需求 工业级可靠性要求,分布式存储支持不同存储节点为系统的智能分析处 理提供数据,并将分析结果融入对应的生产环节

分布式数据库



分布式数据库将数据存放在多个局部数据库、由不同的系统进行管理,可

- 实现不用的通信网络连接、多个节点调取使用。
 实现不同的通信网络连接、多个节点调取使用。

 〈 分布式存储的多个节点并行作业、高可靠性和多协议互通,能够较好地响应分布式数据库需求,实现其独立、透明
 〈 分布式存储可以把分布式数据库的数据复制到其他节点,从而允许应用
- 程序高效使用复制到其他节点的数据,在本地或其他物理地址完成操作













一、2020-2022年中国分布式存储市场规模

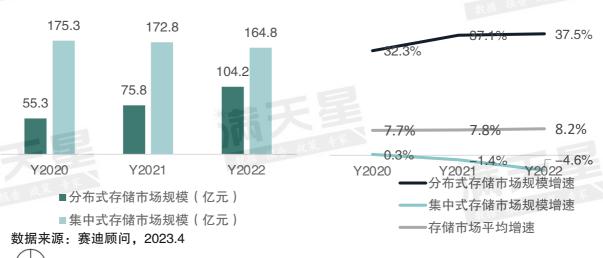
2020-2022年,受数据要素价值持续释放、分布式架构场景加快普及和疫情推动线上经济发展、丰富数字治理实践等因素影响,中国分布式存储市场规模快速发展。

"十四五"时期是迎接数字时代、激活数据生产要素潜能的关键时期,推进网络强国建设,加快建设数字经济、数字社会、数字政府,以数字化转型整体驱动生产方式、生活方式和治理方式变革,产生了海量数据有效存储、去中心化存储等需求,是中国分布式存储市场快速发展的根本原因。

自动驾驶、智慧交通、智能制造等新应用场景大量涌现,这些新应用、新场景在基础设施层普遍基于分布式架构,并且融合了大数据、AI、物联网等技术,其普及将持续产生PB级、EB级海量数据,适合采用分布式存储来存储和处理数据。

新冠疫情对中国分布式存储市场产生了深远影响:一方面,疫情推动线上办公、 学习规模化,带来海量数据快速流转,增加了分布式存储需求;另一方面,疫情 防控更新了数字治理理念,丰富其模式和实践,为疫情后依托数据进行基层治理 和公共服务提供了有效路径,从而持续提升分布式存储市场规模。

图 7 2020-2022年中国分布式存储市场规模及增速

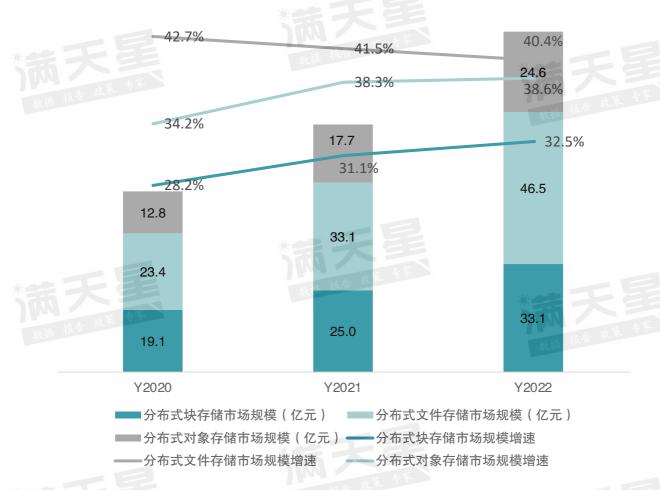


比重快速提升: 从市场规模来看,2020-2022年中国分布式存储的市场规模占比由24%增长到38.7%,尽管相比集中式存储仍有一定差距,但重要性进一步提升

增速表现亮眼:从市场规模增速来看,分布式存储的市场规模增速远高于存储系统市场的平均水平,成为拉动市场增长的重要力量。2020-2022年分布式存储的市场规模增速远高于集中式存储

二、2020-2022年中国分布式存储市场结构

2020-2022年,分布式文件存储凭借其对各类应用的强大兼容性,优势地位逐渐明显。分布式块存储的稳定性较高,而分布式对象存储的成本偏高,有待进一步优化。 图 8 2020-2022年中国分布式块、文件、对象存储市场规模及增速



注:分布式统一存储产品的市场份额已均摊到分布式块、文件和对象存储的市场份额中数据来源:赛迪顾问,2023.4

相比单一硬件和软件形态产品,分布式存储一体机的市场接受度更高,原因在于其交付及维护效率高、性能稳定和可靠性等优势突出,2020-2022年的市场占比均超过85%。

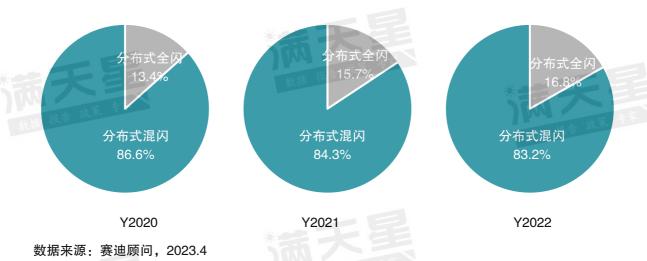
图 9 2020-2022年中国分布式存储一体机、纯硬件、纯软件市场规模占比



数据来源: 赛迪顾问, 2023.4

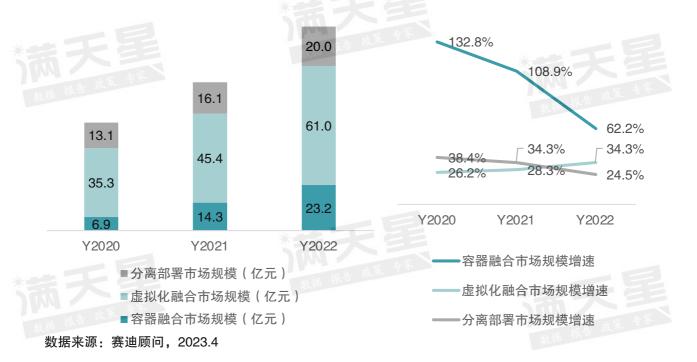
2020-2022年,分布式混闪市场规模占比均较大,受限于成本偏高和适用场景较少等因素,分布式全闪存的比重较低。

图 10 2020-2022年中国分布式存储全闪、混闪市场规模占比



2020-2022年,容器融合的分布式存储部署方式快速发展,成为与虚拟化融合部署互相配合、彼此互补的重要手段。分离部署方式将应用和存储资源隔开,虽然可以避免系统繁忙时负载不均衡,但一个节点存储需求骤升容易导致整个系统不稳定,存在一定的风险,实践中部署偏少。

图 11 2020-2022年中国分布式存储容器融合、虚拟化融合和分离部署的市场规模及增速

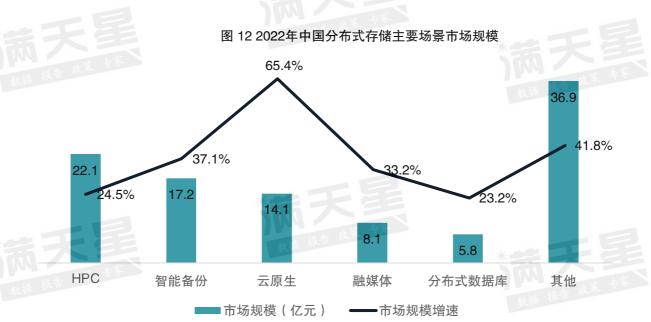


三、2022年中国分布式存储主要场景市场

整体上看,HPC在2022年中国分布式存储主要场景市场中居于重要地位。除了气象预测、基因测序、自动驾驶和AIGC外,HPC还有能源勘探、卫星遥感、各学科数值计算等领域,分布式存储的需求量大,增速保持较高水平。

除此之外,云原生是2022年增速最快的行业市场,原因是云原生企业依托分布式存储 开展业务,而后续业务的拓展也将通过扩展分布式存储的存储节点实现,需求持续增加。随着云原生不断普及,对应的分布式存储需求将越来越大。

除图12提到的几大典型应用场景外,分布式存储还广泛应用在金融、保险、政务等领域,是应对大规模扩容、运营成本均衡、灾备冗余等难题的有力措施,未来有望保持高速增长态势。

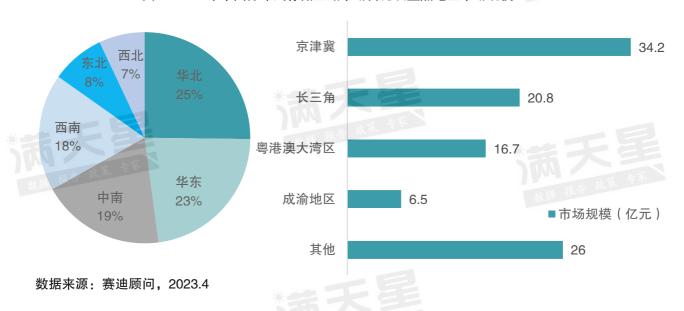


数据来源: 赛迪顾问, 2023.4

四、 2022年中国分布式存储区域市场

从区域市场来看,2022年华北、华东地区占据中国分布式存储市场的半壁江山。一方面,华北、华东地区数字经济发展成果显著,对分布式存储需求量较大;另一方面,华北的京津冀、华东的长三角是分布式存储的重点地区市场,一定程度上对两地区市场的表现起到了带动作用。

图 13 2022年中国分布式存储区域市场占比及重点地区市场规模





一、2022年中国分布式存储市场重大事件

2022年,分布式存储在处理高效海量数据的作用突出,成为"东数西算"工程正式全面启动的重要技术之一;同时保障业务稳定性的优势进一步被行业肯定,是应对突发事件的重要解决方案;随着IT厂商加快云原生实践步伐,以Ceph为代表的分布式存储开源社区的影响力越来越大,而部分厂商采用自研分布式存储产品以支持Kubernetes等工具的效果也较为显著。

○ 分布式存储在"东数西算"中广泛应用

2022年2月,国家发改委、中央网信办、工业和信息化部、国家能源局联合印发通知,同意在京津冀、长三角、粤港澳大湾区、成渝、内蒙古、贵州、甘肃、宁夏8地启动建设国家算力枢纽节点,并设立10个国家数据中心集群,标志着"东数西算"工程正式全面启动。无论是在东部城市数据中心托管的"热数据",还是迁移到西部算力枢纽节点的"冷数据",首当其冲面临的就是数据存储问题,如何实现海量数据的跨区域传输是重中之重。分布式存储将数据分散存储在不同区域的多个存储节点上,对外提供一个虚拟的海量存储资源池,具有性价比高、低功耗、高可靠、通用、免维护等优势,已广泛应用于具有海量数据存储需求的"东数西算"各枢纽节点。

分布式存储保障业务稳定性的优势进一步被行业肯定

2022年12月18日,阿里云的香港机房节点发生故障,导致云服务器ECS、云数据库、存储等产品,以及全球加速、NAT网关、VPN网关等云网产品一度无法正常访问使用。受此影响,受到监管及自身对数据安全和业务连续性的要求较高的主体,已开始加快推动多中心分布式架构场景由"同城灾备中心"向"两地三中心双活"架构(即生产数据中心、同城灾备中心、异地灾备中心)演进,随着业务的发展,将来可能演进成"多地多中心分布式多活"架构。分布式存储技术能够最大程度保障业务稳定性,实现数据资产无间断管理,激发数据潜力,成为提升自身整体竞争力的重要手段。

Ceph开源社区运维团队转移到IBM,对开源的中国分布

式存储厂商产生深远影响

2022年10月,全球知名分布式存储开源社区维护团队,红帽的Ceph存储团队转移 到IBM。根据红帽官方公开信息,团队转移后,项目的管理模式和操作保持不变, Ceph保持100%开源,IBM则将以上游优先的方式进行支持,手段包括但不限于 IBM将承担红帽对Ceph基金会的赞助,并帮助支持Ceph的上游测试实验室。这一 消息对于中国走开源路线的分布式存储厂商具有较大影响:一方面, Ceph运维团 队转移到IBM后,研发重点极有可能加快向IBM Spectrum Scale等IBM主要的云产 品转移,从而减少对混合云原生等场景的支持,给分布式存储厂商的二次开发新增 一定难度和压力;另一方面,随着IBM在Ceph话语权越来越大,深度使用Ceph功 能的分布式存储厂商可能面临数据安全等问题。

自研分布式存储产品对Kubernetes等云原生工具的支持 更好。未来有望加快普及

2022年1月,云原生计算基金会(CNCF)发布云原生应用报告,指出Kubernetes等 云原生工具在数据存储、DevOps、app测试以及许多其他领域功能已经基本成熟, 与传统云计算技术互补配合。然而,主流的开源实用程序在容器编排、服务发现、 负载均衡等任务中,与Kubernetes等工具的配合存在一定不足,而厂商自研的分布 式存储产品, 在原生备份架构、自动化数据备份计划、顺序恢复等功能中具有更好 的表现。随着容器编写越发依赖Kubernetes等工具,应用程序对存储的I/O性能、灾 备能力等需求不断升级、自研分布式存储产品将持续扩大其在云原生企业的服务范 围,进一步加快普及。

2022年中国分布式存储厂商竞争力情况

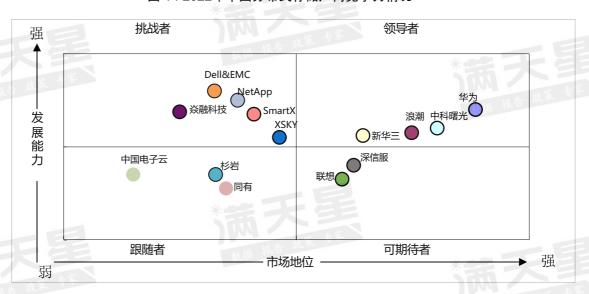


图 14 2022年中国分布式存储厂商竞争力情况

数据来源: 赛迪顾问, 2023.4

2022年中国分布式存储市场领导者有华为、中科曙光、浪潮、新华三。



INSPUT 浪潮 H3C 华为、浪潮、新华三凭借自身完善的存储产品体系,在分 布式存储市场上优势显著



中科曙光ParaStor分布式系统产品基于全自研技术,提供了文件、对 象、块接口的统一存储服务,在海量数据存储、高扩展性、高可靠性、 高性能和方便管理表现较好, 市场份额领先

2022年中国分布式存储市场挑战者有XSKY、 SmartX、 NetApp、 Dell&EMC、焱融科技。



XSKY帮助政企客户通过分布式存储实现海量数据的"汇-存-算-管-用".助力数据全生命周期管理、充分挖掘数据价值



SmartX自主研发的分布式块存储ZBS能够以超融合、存算分离以及容 smartx 器融合等多种产品形态交付和部署,凭借领先的性能、稳定性与硬件 开放性, 在金融、制造、医疗等行业客户生产环境中广泛部署



D≪LLEMC NetApp和Dell&EMC在分布式存储深耕多年,在高速I/O、海量存储、 高效灾备等具有较强实力



焱融科技自研的通用分布式文件系统YRCloudFile 既具备传统并行存 储的高性能,又可作为企业关键业务和新兴业务需求的存储系统。在 AI、智能汽车、高性能计算等领域具有较强实力,行业竞争力突出

2022年中国分布式存储市场跟随者有同有、杉岩和中国电子云,可期待者有深信服和 联想。



同有分布式存储基于去中心化的全对称、云原生的线性扩展分布式 存储架构,企业级NVMe SSD 等硬件平台性能较好



杉岩自主研发企业级分布式存储引擎,采用预集成交付方式,具有 安全可靠、开放生态的优势



中国电子云分布式存储以软件为核心,可同时提供多类型存储服务, 满足用户各种业务产生的结构化和非结构化数据的海量存储需求



深信服分布式海量统一存储,可有效应对海量数据爆发增长带来的挑 战,可向上层业务提供多类存储协议接口、支持EB级数据量



联想高性能分布式存储系统采用了云原生的数据管理设计思路,在 超大规模文件存储应用环境中, 应对数据管理的各项挑战

三、2022年中国分布式存储部分厂商发展情况

Sugon

曙光信息产业股份有限公司

曙光ParaStor分布式存储是以软硬一体化形式交付,通过软硬协同实现端到端的高可靠、高性能和高扩展及一体化的运维能力,同时支持文件、对象、块多种协议,消除存储设备多样性及管理复杂性,实现了海量异构数据资源的融合,高效助力企业的数字化转型。



在某气象项目中,曙光自研ParaStor分布式存储产品得到大规模应用,其高性能、高可靠等特点,基于每个存储节点都可以独立地提供数据访问能力,系统分析处理能力提升了至少30%,保障计算平台在数值天气预报中发挥更大算力。多年来,曙光ParaStor在该气象项目中搭建了近300 PB的存储资源池,支持数十万作业,不断提升气象业务、服务和管理智能化水平



曙光自研分布式存储系统ParaStor为众多科研机构构建的大数据平台,实现数据统一存储与共享,高效支持科研数据的分析、深度挖掘与应用,高效解决了海量科研数据存储场景高聚合带宽、小文件存储场景高I/O的需求。曙光ParaStorEB级的扩展能力表现较好,性能与容量的线性增长,满足其对存储系统数据处理能力和扩展性的要求



北京焱融科技有限公司

焱融科技自研的分布式文件存储系统YRCloudFile,支持 POSIX / SMB / NFS 等文件协议接口,可满足业务发展对存储容量和性能同步提升的需求。作为企业级存储,提供DataLoad、冷热数据自动分层、大数据支持等数据全生命周期管理功能和方案;支持200Gb/400Gb Infiniband 网络及NVIDIA GPUDirect,产品性能行业领先。



在人工智能领域,焱融YRCloudFile高性能分布式文件存储以优异性能,高效解决国内某知名AI技术创新型公司小文件读取缓慢、存储和Kubernetes平台需无缝对接、训练数据在容器中的持久化存储及高速吞吐等难题,YRCloudFile所具备的高带宽、低延时的特性可较好满足GPU等计算服务器的计算效率达到饱和的需求,可实现近千台高性能计算服务器稳定运行



在智能汽车领域,焱融科技提供的"混合云存储"解决方案为全球 某知名无人驾驶企业构建混合云存储体系,完成业内领先的边缘、 数据中心、混合云存储架构,实现数据在本地及多云间流转,各业 务阶段的高性能存储,IT基础架构随业务发展具备性能和容量按需 扩展的能力



焱融全闪存储一体机F8000X助力某全球500强制造企业完成首个超大规模 AI 存储集群。焱融科技提供存储性能与建设成本兼顾的方案满足AI 计算场景对存储的极高IOPS和带宽、极低延时需求,性能和容量可随业务发展同步扩展,单次训练时间由一周缩短至两天,训练精度大幅提升

Smartx MAKE IT SIMPLE

北京志凌海纳科技有限公司 (SmartX)

SmartX分布式块存储ZBS完全自主研发,采用全分布式架构,能够以超融合、存算分离以及容器融合等多种产品形态交付和部署,不仅核心稳定,同时支持双活、异步复制等企业级特性,并通过RDMA、NVMe-oF等技术释放软硬件性能潜力,在金融、医疗和制造等关键业务场景表现优异。



制造场景中,某全球知名电子科技制造服务商通过超融合的方式部署 SmartX的分布式块存储,承载国内、海外等各分公司生产内网的 MES、ERP、产线综合管理等业务系统,在实现对关键业务高可用 支撑的同时,降低采购与试用成本,灵活选择硬件配置,同时系统上 线快、运维简单,可对各地集群进行统一的可视化管理。



银行场景中,某大型国有银行通过超融合的方式部署 SmartX分布式块存储,并在虚拟化层采用SmartX 原生虚拟化ELF替换VMware—ESXi虚拟化,构建轻量、开放的自主路线的云产品,对专属云技术路线进行补充,并且技术路线可快速复制到分行业务场景中,助力总行、分行的云化和其他IT应用创新转型



证券场景中,某大型券商以分离部署的架构打造了可承载投研平台、私有云平台的分布式块存储集群,构建数据中心的"底座",助力业务发展,同时借助分布式存储高并发、弹性扩展和易维护等特性,减少硬件绑定,降低总体成本和运维复杂度,提升系统灵活性与适应性、提升服务器资源利用率

四、2022年中国分布式存储部分细分市场情况

教育科研市场对分布式存储的需求除了海量数据I/O和温冷热数据管理外,还要求支撑线上课程、教务系统、科研成果共享等多样化应用。2022年中国分布式存储教育科研市场中,中科曙光ParaStor存储系统通过全栈自主研发的NVMe全闪存优化、混合负载承载等技术,有效响应了教育科研用户需求,凭借26.1%的市场份额占据市场首位。华为、浪潮分列第二、三位。除此之外,中科曙光ParaStor在保障高端气象数值平台发挥作用上广受用户认可,占据2022年中国气象市场首位。

图 15 2022年中国分布式存储教育科研市场厂商市场份额



数据来源:赛迪顾问,2023.4

2022年中国分布式块存储金融市场用户较为看重产品稳定性、企业级存储特性、小I/O 高并发和低延迟性能, SmartX的ZBS产品受到青睐, 在稳定性、性能及开放性方面具备独特优势, 并提供了多种的部署形态, 2022年在分布式块存储金融市场占比达21.3%, 市场份额最高。华为和新华三分别以19.1%和17%紧随其后。

图 16 2022年中国分布式块存储金融市场厂商市场份额



数据来源: 赛迪顾问, 2023.4

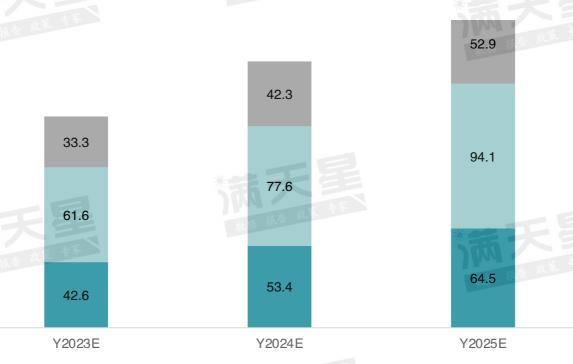




一、2023-2025年中国分布式存储市场规模预测

2023-2025年,数字中国建设进入重要时期,分布式存`储的高扩展性、高效作业、高可靠性等优势将越发显著,市场规模持续增长,预计到2025年有望达到211.4亿元。





- ■分布式块存储市场规模(亿元) ■分布式文件存储市场规模(亿元)
- ■分布式对象存储市场规模(亿元)

注:分布式统一存储产品的市场份额已均摊到分布式块、文件和对象存储的市场份额中

数据来源: 赛迪顾问, 2023.4

未来三年,分布式存储的主要应用场景需求个性化、高性能化要求显著。云原生场景中,持久化存储将成为容器化场景大规模落地的关键,Kubernetes等容器场景需要与之真正匹配的运维和管理模式的分布式存储产品;HPC的需求将主要体现为可扩展性,EB级数据的高效I/O和成本控制一定程度上决定了厂商的竞争力;融媒体场景的需求将随着元宇宙、虚拟数字人的落地和普及相应提升,低时延的特性将提振分布式全闪存产品的市场需求。与此同时,分布式存储也将拓展其在多元场景的应用。

到2025年,云原生、HPC和融媒体的市场规模有望达到44.0亿元、32.3亿元和23.7亿元,领跑中国分布式存储市场。

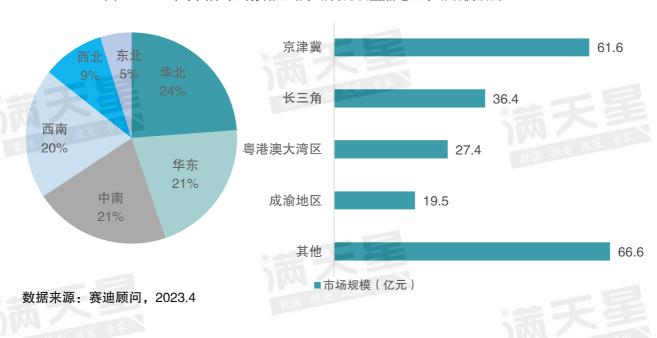


数据来源: 赛迪顾问, 2023.4

随着"东数西算"工程持续推进,东西部地区算力、存力协同发展态势将越发显著,东部地区保持市场优势地位的同时,西部地区支撑东部地区的效果有望提升,对分布式存储的需求也将相应提升。到2025年,预计西北、西南地区市场份额占比之和将接近30%。

从重点地区市场规模来看,京津冀、长三角和粤港澳大湾区将继续领跑,到2025年,分布式存储市场规模将分别达到61.6亿元、36.4亿元和27.4亿元。成渝地区作为西部地区数字经济发展重点区域,对分布式存储的需求较大,2025年有望达到19.5亿元。

图 19 2025年中国分布式存储区域市场占比及重点地区市场规模预测



二、驱动及阻碍因素分析



数据爆发式增长、业务新型需求有待高质量响应

移动互联网、大数据、云计算等技术发展使数据量呈现爆炸式增长。2022年,中国数据产量达8100EB,同比增长22.7%,占全球数据总量的10.5%,排在第二位^{*}。面对庞大的数据量,政企业务需求也发生变化,对存储的I/O提出了更高要求。集中式存储具有较强的稳定性,但在需求的响应速度和数据吞吐质量上难以满足业务快速部署需求。

集中式存储过于依赖硬件,需根据设备的扩容能力等性能响应需求,而分布式存储可按需求进行动态配置,实现存储与应用的紧密的联系,提供更高的敏捷性,以需求或管理模式驱动的高级功能匹配政企业务需求,数据请求直接传达至管理软件,再下放到对应请求分配的各存储节点资源,人工干预较少,从而较好支持大量并发的存储需求。



可靠性要求提升, 存储可持续性和持久性越发重要

现行存储建设模式下,软硬件种类繁多,供应商各异,不同设备构成的存储基础设施易形成异构孤岛,数据无法流通,难以对存储节点内的基础设施进行统一的监控、管理、报告和远程访问,同时数据的价值不断提升,一旦发生宕机故障,灾备恢复的压力和成本极高,可持续存储和持久存储需求迫切。

分布式存储把所有的数据分散保存在不同的存储节点中,同时基于副本、纠删码(EC)、数据均衡、容灾、故障检测和恢复等技术,有效保证数据可靠性,实现磁盘、机柜甚至是数据中心的数据冗余,从最大程度上解决数据孤岛等问题,保障前端业务的连续性和可持久性。



各类硬件性能提升,为分布式存储快速普及提供基础

SSD内置在存储设备内部,可在低于毫秒的时间内对任意位置的存储单元完成 I/O操作,全闪/混闪分布式存储配置了SSD,可达到机械硬盘的几十甚至上千倍的IOPS;目前各大存储厂商对X86架构的计算硬件适配较好,存算标准趋于标准化、通用化,为分布式存储的普及打下基础;分布式存储借助于节点之间的有效通信,实现数据缓存、同步复制来确保冗余性,而RDMA 高速网络技术已经基本成熟,

*数据来源:《数字中国发展报告(2022年)》,2023.4

其发展对跨存储节点间的延迟缩短起到了重要作用,一定程度上也为分布式存储的 快速普及提供了有利条件。



云原生加速业务创新,容器部署及高效使用成为热点

传统IT面临数据爆炸式增长、业务种类增加、应用复杂性提高、软件迭代速度加快等冲击和挑战,容器为其提供了有效解决方案,成为云原生场景下业务创新的关键点。

分布式存储能够将各存储节点资源无缝对接Kubernetes等容器工具,同时支撑多容器并行,支撑MySQL、Redis、Cassandra等数据库业务、各类中间件和微服务平台高效I/O。在支撑容器部署过程中,硬件成本偏低、可提供多类型的管理策略、大幅缩减运维成本等优势突出,也为业务持续演进提供了有力支撑。



资金、人才支撑不足制约了自研分布式存储产品的发展

由于分布式存储产品研发、试用需要较长周期,资金投入大且见效慢,自主产品研发的企业资金实力要求较高,加上国内投资机构缺少对纯技术投资的传统,以及中国存储行业起步较晚导致的经验丰富、技术能力强的专业技术人才和管理人才较为缺乏,导致目前中国独立研发的分布式存储产品较少,多数产品基于开源框架、面向应用进行二次开发,在技术上与Dell EMC的Isilon、NetApp等国外厂商的产品具有一定差距,一定程度上影响了中国自主研发的分布式存储产品发展。



标准滞后和缺失影响了分布式存储市场良性发展

中国分布式存储产品和服务存在统一的技术参数、收费价格、服务评判等标准滞后和缺失问题,行业缺乏规范性的统一指引,制约了行业发展。与此同时,中国分布式存储市场蓬勃发展,驱动大批企业投身其中,但不同厂商的技术实力参差不齐,产品也难以评估和筛选,一定程度上延缓了新技术的推广和普及。另外,由于分布式存储涉及硬件购置和安装,以及软件的对应部署,更换设备和服务的供应商将产生较高的经济和时间成本,因此首次购买极为重要,对分布式存储关键技术服务和应用的标准需求迫切。目前标准不统一的问题导致厂商和用户之间存在巨大的信息鸿沟,市场难以健康良性发展。

三、分布式存储的发展趋势

(一)分布式块、文件和对象存储将在各自适用的业务场景保 持快速发展

分布式块存储凭借自身直接读写存储节点的某段地址即可访问、较高效率的资源调取等优势,将在分布式数据库等领域加速普及;分布式文件存储延时虽高于块存储,但在NAS虚拟化处理非结构化数据上表现突出,将在人工智能、自动驾驶等结构/非结构数据训练的领域作用显著;分布式对象存储兼顾高速、共享、智能,与容器等适配性较好,打包交付和扩展能力较强,虽存在部署成本高等不足,但在Hadoop、Spark等大数据分析平台将保持自身优势、广泛应用。

(二)分布式全闪存成本有望进一步下降,其适用范围将大大 扩展

分布式全闪存具有远高于机械硬盘的IOPS性能、低延迟、低能耗的优势,但SSD的单位容量成本仍然是机械硬盘的数倍,阻碍了分布式全闪存在众多应用场景下的推广与使用。随着双重RAID技术、短码纠删码、在线数据压缩以及数据去重技术的成熟与应用,分布式全闪存成本将显著下降,存储空间使用效率不断提升。推动分布式数据库、云平台、海量文件系统、大数据分析的多应用场景闪存化,提升用户体验的同时,也可控制能耗。

(三)分布式存储将在云环境下为用户提供高效存储服务

当前,分布式存储通过各存储节点协同形成存储资源池,以弹性云硬盘的形式提供给各类应用的虚拟机等载体使用。随着各类应用类型不断增多,在性能成本均衡的个性化要求下,由于云环境下的业务与分布式存储弹性扩展、存算分离、存储类型可选、EC纠删可靠和高并发等便捷场景服务等优势结合度较高,分布式存储将在应用上线初期的渠道类应用、各类外围应用、灰度发布应用、容器技术应用、桌面以及开发测试环境中扩展应用,提供高效、便捷、经济的存储服务。

(四)分布式存储将成为云原生场景中的主要存储方式

云原生是云计算及其应用的关键技术趋势,对存储的需求是公有云/私有云/混合云等场景扩展性强、应用程序编程接口(API)驱动、对K8s等工具友好等。分布式存储将文件、块和对象按特点引入到容器集群,与其他正在使用存储资源的应用程序和服务一起无缝运行。通过这种方式,云原生集群可在公有云和本地部署中自给自足并且具备可移植性,从而较好地响应可扩展、API驱动需求。与此同时,分布式存储没有中心节点,允许各存储节点使用统一界面管理存储,大大提升用户体验。

(五)分布式存储将助力AI大模型加快落地

分布式存储依托可靠性、高效作业、可扩展性等优势,在助力AI大模型等新技术和新应用落地等方面将发挥重要作用。分布式存储扩容能力强大,可支撑AI大模型参数数量的提升、训练质量的提高,助力降低开发门槛、高效利用算力资源、提升预测效果,如在医疗领域通过学习医生记录数据,提供辅助的护理建议总结,从而改善医生的工作流程;而在金融行业,可通过训练投融资数据,优化投资组合管理等。

(六)存储安全问题日益凸显, 自研产品前景更广

依托Ceph、GFS等开源架构进行二次开发的分布式存储产品,在面对复杂任务的响应能力不足、开源许可存在不确定性等不足越发明显。未来,分布式存储产品与应用的深度绑定将成为衡量产品水平的关键指标,即自研不意味着必须"从无到有"、创造全新的架构和技术,而是深度理解应用需求、参考大规模应用验证的企业级产品特性,并结合自身技术优势,面向垂直应用自主设计的产品,此类产品能够更好地响应应用需求,具有广阔的发展空间。

行业报告资源群



- 1. 进群福利:进群即领万份行业研究、管理方案及其他学习责要,互接打包下载
- 2. 行研構造: 每日分享6+份行业精造报告及3个行业主题研究员料
- 2. 查询报告: 行研资料免费帮助查询下载
- 4. 严禁广告: 本群仅限行业报告交流, 禁止一切无关他是



厂商: 根植应用个性化需求

改变以通用的分布式产品适配各类应用需求的产品研发思路,重视快速发展的各行业应用,根据应用个性化需求,提供高适配性产品。如除了应用需要的I/O效率、可靠性、扩展性等性能外,还要将成本、用户专业程度、当地碳排放要求等列入产品设计的考虑范围;按应用I/O需求,合理搭配SSD与机械硬盘的比例,在保证产品性能的前提下尽可能控制成本;根据用户数据安全等级需求,提供对应的2副本4节点、3副本5节6点和4副本6节点冗余策略等。

投资机构:关注自研路线厂商

建议投资机构关注具备面向应用需求的架构自研、技术自研厂商,研判其在元数据节点集群化、虚拟目录、数据分布算法等分布式存储核心技术的成熟度,以及在基因分析、医疗影像、影视渲染及HPC等拥有海量数据的场景的表现。同时,分布式存储产品自研对团队能力要求较高,建议重点关注团队核心成员的学历背景、就职经历、发明专利申报等信息,特别是是否有曾在云计算、人工智能、软件开发等知名企业就职经历,以及行业既有分布式存储部署的项目经验。

政府及第三方机构: 细化多维度标准

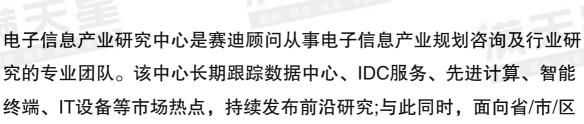
统筹产学研用各方力量,研究出台分布式存储产品功能、性能、 兼容性、扩展性、可靠性、安全、运维、硬件、网络等方面的技术标准和测试方法,以及面向新型数据中心、绿色数据中心等标杆示范的分布式存储应用标准。推动相关评测认证体系的推广宣 贯和评测结果的发布,为分布式存储用户建立起对不同厂商、不同场景和不同路线的产品的直观依据,从而助力中国分布式存储 市场的良性发展,也为有关新技术测试迭代提供必要基础。





— 股票代码: HK02176

思 维 创 造 世 界



(县)政府和产业园区提供电子信息产业专项规划、工业和信息化综合 规划、电子信息产业园区规划等顶层设计服务

赛迪顾问股份有限公司

北京市海淀区紫竹院路66号

邮编: 100048

邮箱: yuanyu@ccidconsulting.com

联系电话: (010) 8855 9095

(+86) 18810576574

www.ccidconsulting.com

www.mtx.cn