



云计算开源产业联盟
OpenSource Cloud Alliance for Industry (OSCAR)

CAICT 中国信通院

多云数据存储白皮书

(2023 年)

2023年3月

参编单位

中国信息通信研究院

阿里云计算有限公司

华为技术有限公司

京东云计算有限公司

北京火山引擎科技有限公司

天翼云科技有限公司

中移（苏州）软件技术有限公司

中电信数智科技有限公司

联通数字科技有限公司

新华三技术有限公司

曙光信息产业（北京）有限公司

浪潮电子信息产业股份有限公司

联想凌拓科技有限公司

中电云数智科技有限公司

上海道客网络科技有限公司

杭州飞致云信息科技有限公司

北京佳杰云星数据科技有限公司

云宏信息科技股份有限公司

安超云软件有限公司

北京星辰天合科技股份有限公司

上海爱数信息技术股份有限公司

深圳市杉岩数据技术有限公司

深圳市泛联信息科技有限公司

华瑞指数云科技（深圳）有限公司

上海霄云信息科技有限公司

甲骨文（中国）软件系统有限公司

北京焱融科技有限公司

深圳大普微电子科技有限公司

四川启睿克科技有限公司

序 言

很高兴看到这本集合产业界众多代表编写的白皮书隆重发布。

国家对于数据要素愈加重视，势必加快企业数字化转型，推动企业上云用数赋智，政务云、行业云、国资云等快速发展，进一步催化云计算市场服务需求。数据存储是云的数据载体，构建先进可靠的多云数据存储能够促进数据要素的共享，流动，保障数据安全。

这本《白皮书》系统地梳理了多云时代的驱动力和面临的数据挑战，提出了多云数据存储的应对措施和发展趋势。构建多云统一的数据底座促进数据跨云共享，增强智能分级存储使能数据跨云流动，通过面向容器的存储创新助力应用上云改造，发展新型存算架构实现云计算基础架构的扁平分层和灵活组装，提升数据应用的效率，降低能耗，提高数据存储的安全可信能力保障数据的不丢失、不泄露。我们相信这是一次非常有价值的研究，对于相关决策部门、企业界、学术界具有很好的启示和参考意义。

数据成为企业发展的核心资源，上云用数赋智是企业数字化转型的“助推器”。期待在“政产学研用”多方力量的共同努力下，云和数据存储的生态将不断加强和优化，多云数据存储技术创新将不断深入，为数据基础设施建设提供关键和强有力支撑。

中国工程院院士



目 录

一、 多云数据存储发展背景.....	1
（一） 数字经济提速，云计算蓬勃发展.....	1
（二） 云化转型深入，企业使用多云成为新常态.....	2
（三） 数据价值凸显，数据存储成为多云关键底座.....	6
二、 多云数据存储现状.....	8
（一） 多云数据存储发展迅速，呈现三大特点.....	8
（二） 多云数据存储是上云用数的助推器.....	9
三、 企业上云用数面临的挑战.....	11
（一） 多云数据流动难.....	11
（二） 跨云数据共享难.....	13
（三） 数据安全保护难.....	13
（四） 多云应用改造难.....	14
（五） 数据应用能效低.....	14
四、 多云数据存储发展趋势.....	16
（一） 面向多云构建统一数据底座，扩大数据共享应用.....	16
（二） 数据存储智能分级能力增强，使能数据跨云流动.....	19
（三） 数据存储面向容器持续演进，助力应用云化改造.....	21
（四） 以数据为中心新型架构兴起，提升数据应用能效.....	24
（五） 边缘加速向超融合架构演化，云边协同不断深入.....	27
（六） 安全可信数据存储部署加快，保障数据可靠安全.....	30
（七） 数据存储节能技术逐渐成熟，促进绿色低碳发展.....	32
五、 总结与展望.....	36

图 目 录

图 1 2022 年全球企业多云调研情况	3
图 2 2022 年中国企业多云调研情况	3
图 3 典型企业多云应用情况示意图	4
图 4 企业多云架构数据底座示意图	6
图 5 2016-2022 年数据存储全球市场规模	8
图 6 多云数据流动面临诸多挑战	12
图 7 多云趋势下有状态应用云化改造越来越深入	14
图 8 跨云统一数据管理示意图	18
图 9 全局文件系统示意图	19
图 10 企业多云分级协同示意图	20
图 11 有状态容器数据持久化示意图	22
图 12 容器应用级高可用示意图	24
图 13 传统云计算 IT 架构与新型云计算 IT 架构示意图	25
图 14 云边缘协同示意图	28
图 15 安全可信数据存储示意图	30
图 16 绿色数据存储示意图	33
图 17 多协议互通减少数据搬迁示意图	34
图 18 多云数据存储助力企业上云用数	37

一、多云数据存储发展背景

（一）数字经济提速，云计算蓬勃发展

云计算作为信息技术发展和服务模式创新的集中体现，已成为企业及产业实施数字化转型的重要基础。自2006年云计算概念推出以来，云计算功能日趋完善，种类日趋多样，近年来，我国高度重视云计算产业发展，在上云用云政策牵引、数字经济提速等因素驱动下，我国云计算市场持续高速增长，2012年以来年均增速超过30%，成为全球增速最快的云计算市场之一，2021年市场规模已达3,229亿元。

政策牵引深度上云用云。国务院、工信部先后发布了《云计算发展三年行动计划》《推动企业上云实施指南》《“十四五”数字经济发展规划》等系列政策措施促进企业深度上云用云，实施上云用云行动，促进数字技术与实体经济深度融合，赋能传统产业转型升级。我国上云企业已累计超过 380 万家，云计算在推进经济社会数字化转型中发挥了重要作用。

数字经济提档加速。2002 年到 2011 年，中国数字经济增速低于同期 GDP 平均增速，数字经济仅仅是国民经济的一部分。2012 年以来，我国数字经济规模占 GDP 的比重不断提升，数字经济年均增速显著高于同期 GDP 平均增速，中国数字经济规模保持高速增长。近几年，我国在数字经济关键领域加快部署，“东数西算”工程加快实施，国家算力网络加快构建，以 5G、人工智能为代表的技术进步和

产品创新快速演进，加速与垂直行业深度融合，应用场景迸发，在数字产业创新能力提升带动下，产业数字化转型实现了提档加速。

（二）云化转型深入，企业使用多云成为新常态

企业云化已经成为全球共识，对于大型企业，选择云化可以进一步提升 IT 建设水平。对于中小企业，则可以实现信息化管理节约资金投入，避免陷入复杂 IT 架构建设的窘境。我国云计算应用已从互联网拓展至政务、金融、工业、医疗、交通等传统行业。互联网和信息服务业已基本实现云计算的深化应用。金融、政务、交通等云化改造能力持续加深。能源、医疗、工业等行业的核心系统的云化改造逐步向核心系统推进。随着企业云化演进的不断实践，企业在持续探索选择最佳建设模式。最初，单一云平台被企业广泛采用，但随着云平台支撑的业务类型增多，单一云平台的架构也暴露出诸多问题。一方面不同的云厂商各有各的特长，另一方面，单一云供应商也增加了企业对数据安全的担忧，因此多云战略正在逐步成为最优的企业云战略。所谓多云，是指企业使用多个异构云供应商来提供统一的计算/存储服务，以提升云基础设施能力、控制成本，架构上通常包含两个或以上公有云或者私有云。

根据 Flexera 的 2022 年云趋势报告¹，89%的企业在 IT 架构上选择多云战略，其中 80%的企业选择混合云，企业 IT 架构日益复杂化，

¹ Flexera 2022 年云趋势报告: <https://www.flexera.com/blog/cloud/cloud-computing-trends-2022-state-of-the-cloud-report/>

多云战略已经是当下大多数上云企业的选择。多云既具备公有云的灵活性、规模优势与创新能力，又可满足企业的数据安全监管及深度控制诉求，同时也满足用户不断追求更高性价比的诉求，逐步成为最优的企业云战略。

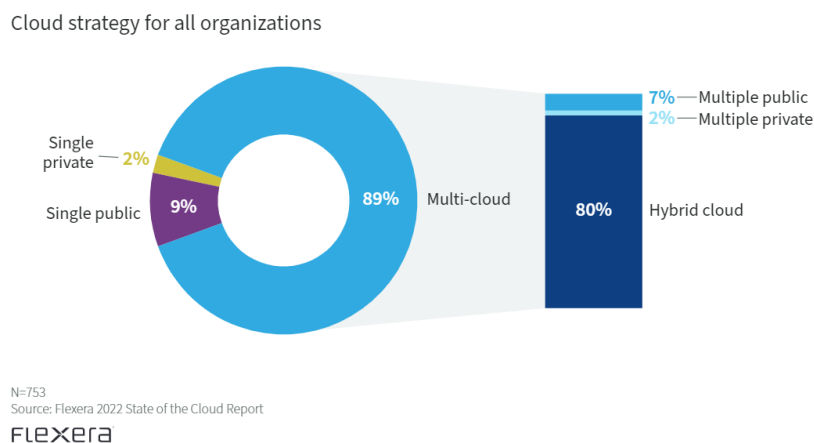


图 1 2022 年全球企业多云调研情况

根据信通院调研，在国内，已经有近九成企业选择多云，其中 63.4%企业使用混合云，25.3%的企业采用多公有云或者多私有云。76%的企业使用 2~4 种云，其中使用 3 种云的企业占比 35%，使用 4 种云的企业占比 20%。

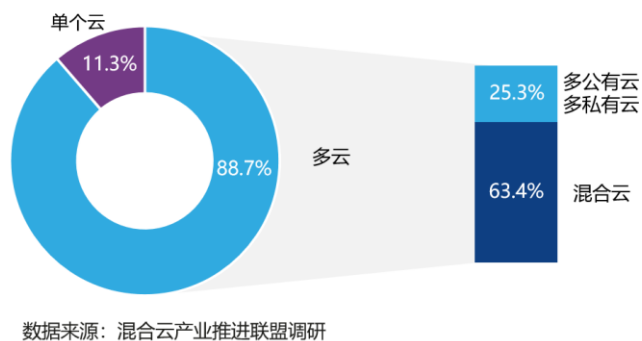


图 2 2022 年中国企业多云调研情况

比如如下典型企业，同时在使用 2~5 种云：

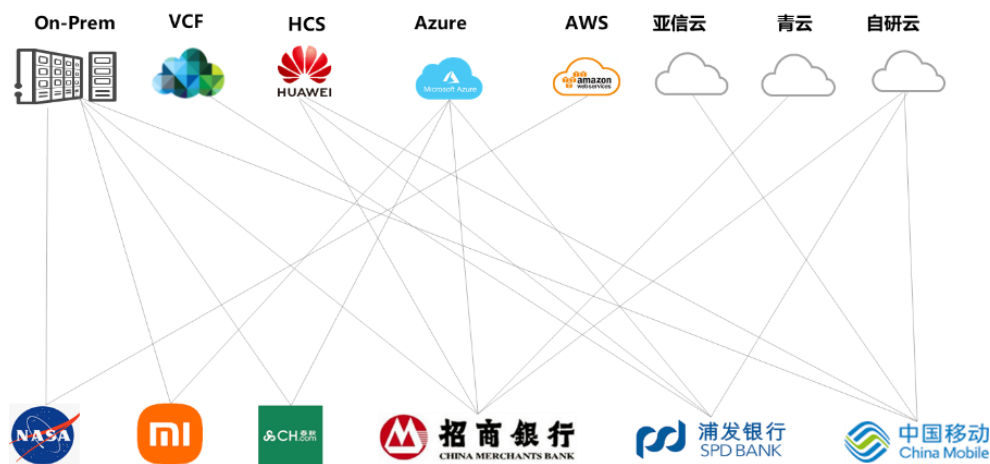


图 3 典型企业多云应用情况示意图

国内不仅民营企业积极部署多云 IT 架构，央国企近些年也加快了多云的建设，并对外赋能。据统计，到 2021 年，近 98 家中央企业已建成私有云 1053 个、公有云 238 个，业务系统上云率已经达到 52.15%。

多云可以让企业拥有最佳选择权，可以在三五个解决方案中选择最优方案，可以从根本上降低系统性风险。同时，企业可以随时做好准备，把其中一朵云的服务，迁移到另一朵云，以应对不可接受的价格上涨、业务中断等问题，如下四个因素是驱动企业选择多云的主要因素。

优势互补

多家云厂商优势互补：不同的公有云和私有云厂商都有自己在 IaaS, PaaS, SaaS 某些方面的优势，企业正在试图根据自身业务诉求，

选择将不同的业务应用部署在相应的云平台，以期获得最佳的 IT 技术组合。

公有云和企业自建数据中心优势互补：公有云资源具备良好的弹性，按需申请、释放，适合企业快速开展各类不确定和创新业务；为了保证企业数字化竞争力，企业的核心业务又要求 IT 平台的差异化能力，加上核心数据私密性和管控的诉求，企业会选择在企业数据中心自建私有云，提供类公有云的一致性服务体验。

业务与数据安全

企业考虑到业务高可用，通过应用跨云部署避免单云整体性的故障，避免单一基础设施带来的风险。另由于不同业务系统数据安全等级不同，企业不同的业务也选择使用不同的云，如计费、银行核心交易、制造执行等系统多数选择私有云，电子商务、游戏、票务等系统选择公有云或者混合云。跨国企业还需考虑各国法律合规遵从，如欧洲国家要求本地数据不能出境，只能在各个国家选择使用本地公有云。

最优成本组合

企业引入多个云服务商，通过引入竞争争取商业价值，选择最优成本组合的云服务。还可以根据业务场景，每个业务场景选择成本最优的云，获得业务与成本均衡的多云组合。

实现多元化供应

“鸡蛋不能放在同一个篮子里”，对于企业用户，特别是大型企业用户来说，把公司的“生命线”业务完全依托给单个供应商来保障是有

一定风险的。使用多云可以保证企业用户同时具备多个方案可供选择和切换，最大限度地保证了业务稳定性。

（三） 数据价值凸显，数据存储成为多云关键底座

数据作为关键生产要素的价值日益凸显，深入渗透到经济社会各领域全过程，为充分发挥我国海量数据和丰富应用场景的优势，促进云计算技术和实体经济深度融合，数据存储的重要性也日益凸显。

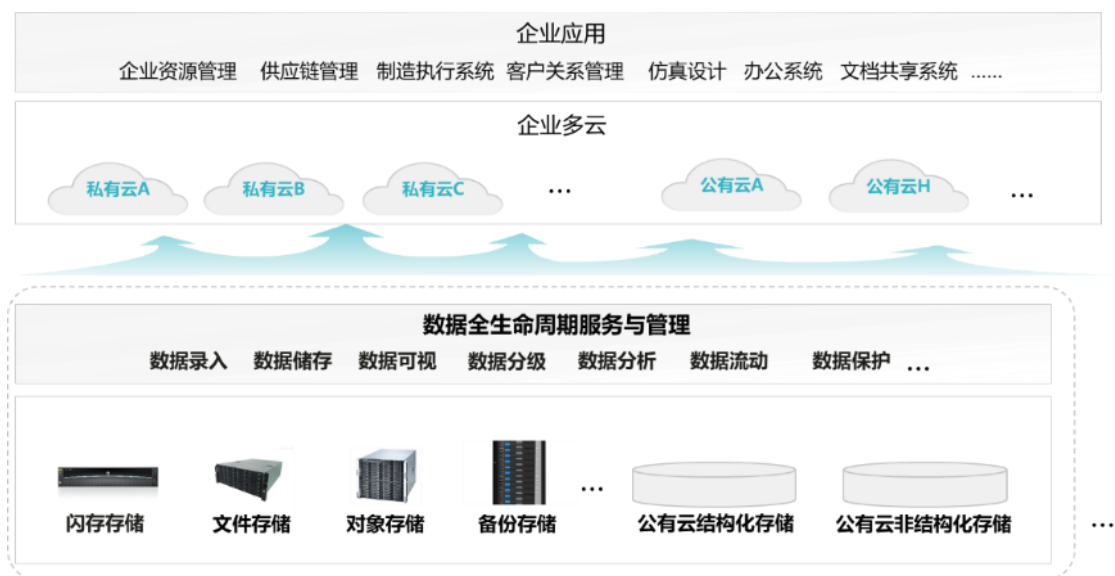


图 4 企业多云架构数据底座示意图

数据存储是云计算的重要组成，是多云架构中数据资源的载体。数据资源是企业的关键生产要素，数据的起点就是把采集的数据进行存储，从而给数据处理和应用提供原材料，支撑企业的数字化运作，产生的各类数据与物理世界完成交互后，最终处理的结果又回到数据存储进行保存，指导模型、算法的精确性和可用性提升。数据存储在多云架构中提供存储资源，供用户任意时间、任意地点通过网络存取

数据，是企业多云架构的数据底座。

数据存储正逐步成为多云架构下数据全生命周期服务与管理的平台。当今社会和企业运行已经表现出高度数据化、数字化，每一个组织、部门甚至个人都是一个数据集。数据存储正在成为这些数据集的统一管理平台，并且为社会治理、企业运营发展提供关键的技术支撑。针对多云架构的数据管理、流动、分析以及绿色低碳发展正在成为数据存储技术的重要实践。

二、多云数据存储现状

（一）多云数据存储发展迅速，呈现三大特点

数据存储是将信息以各种不同的形式存储到介质上的系统，多云数据存储即指云计算 IT 基础架构中的企业级数据存储系统。简单的说，多云数据存储就是云上数据的智能自动化“仓库”，参考国际上基于用途和应用场景的分类准则，数据存储产品分为分布式存储、集中式存储、数据保护（包含磁带存储、蓝光存储）等类别。

企业级数据存储市场发展迅速，2022 年市场规模达到 557 亿美元。云计算是数据存储最为重要的应用场景，在云计算中的数据存储规模占比超过 50%，随着企业上云进程不断加快，企业对云平台中的数据存储需求不断释放，数据存储在云计算中的市场规模正在不断提升。

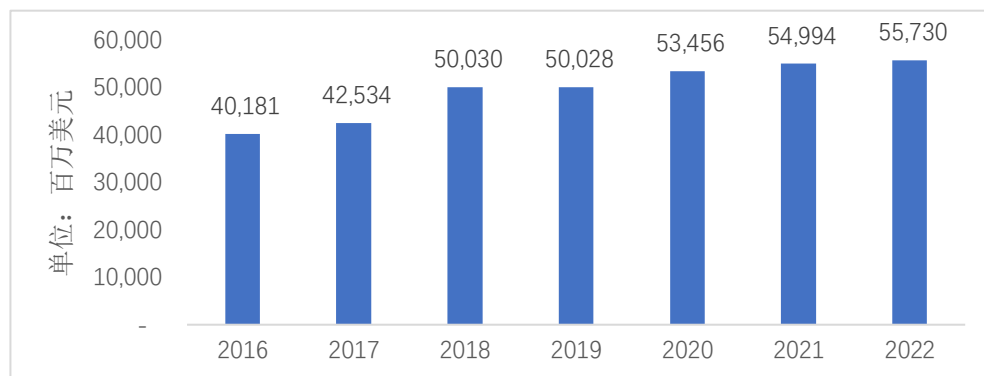


图 5 2016-2022 年数据存储全球市场规模

多云数据存储呈现三大特点，一是产业玩家众多。随着市场持续高速增长，吸引各类厂商参与，市场竞争日益激烈。公有云厂商，私有云厂商、专业存储厂商、边缘设备厂商等纷纷入场，其中，公有云

厂商为发挥规模效应，往往自研大规模集群的分布式块存储和分布式对象存储，摊薄成本。私有云厂商则更了解企业用户对规模、性能、功能的均衡要求，除了自研部分类型的数据存储，还广泛和专业存储厂商互联互通，形成合作生态。专业存储厂商数十年支撑各类企业业务，对企业核心应用有着很好的支持，通过积极与云厂商合作，助力企业应用上云改造和多云数据流动。边缘设备厂商通过一体机的形式，凭借更高集成度更易管理等优势与云厂商形成互补，加快云边协同。

二是烟囱式建设仍然是主流。伴随着多云的烟囱式建设，多云数据存储的建设也呈现烟囱架构，企业每建设一个云就对应一套独立的数据存储，但烟囱型建设的比例正在逐渐降低，根据 Flexera 的调研统计，从 2021 年的 49% 下降到 2022 年的 45%。

三是对上云应用的适配不足。企业上云已经进入深水区，企业基于构建差异化竞争力和业务安全等考虑，在定制化、成本等方面存在独特需求，如高性能计算，金融核心交易系统，对数据存储有着高性能，高可靠，丰富的企业级存储功能的要求，通用的云计算存储服务能力往往难以满足需求，造成用户上云困难。

（二）多云数据存储是上云用数的助推器

在多云环境中，如何将数据安全可靠地存储，并高效地利用，将数据化为知识，是企业上云能否成功的重要因素，也是企业实施云计算战略的重要考量。

应用上云的基础是数据上云，没有数据做为支撑，应用将无法运

行。数据从产生、储存、传输、处理、共享甚至销毁都离不开数据存储，相较于多云架构的其他各层，数据存储层在数据相关的操作上有着天然的优势。如数据流动，通过 PaaS 层进行的数据流动，需要将数据从本地数据存储中读取出来，再通过消耗应用和算力资源才能将数据流动到另一端，到另一端后，同样需要消耗应用和算力资源接收数据，并最终将数据储存到另一端的数据存储中，而通过构建良好的多云数据存储系统，数据只需要从本地的数据存储远程复制到另一端的数据存储中，少量甚至无需占用算力资源，且流动效率最高，通过设置针对性的流动策略，还能够实现数据流动，应用无感。

数据存储承载着上云用数的变革之道，解决数据挑战是多云数据存储发展的必然使命。

三、企业上云用数面临的挑战

近年来我国不断推动企业上云用数，上云是企业数字化转型、信息化发展的重大变革和必然趋势，用数是在更深层次推进大数据的融合运用，让数据转变为企业核心生产要素，进一步释放数据要素价值。面对当前多云趋势，企业上云用数面临如下五个数据挑战。

（一）多云数据流动难

多云数据流动的目的是为了应用的跨云，因此需要从数据使用的角度来考虑数据在多云间流动的一致性。例如对于数据库而言，为了确保业务的可用性和逻辑的正确性，需要根据数据库的事务特性来进行数据的跨云流动。同样，企业的业务是多样性的，同一个业务系统可能涉及到块存储、对象存储、数据库、缓存、消息队列等的各个层面，需要各种数据都能在多云间双向自由流动。此外，对于企业来说，数据流动的及时性越强，数据的价值和意义就越大。比如在进行跨云跨地域的应用多活时，数据毫秒级的快速流动能够将 RPO 从传统灾备方式的小时级降低到毫秒级。因此，多云间数据流动的一致性、多样性和及时性能够极大的提升多云的价值，丰富多云的应用场景，对于企业能否真正的用好多云至关重要。

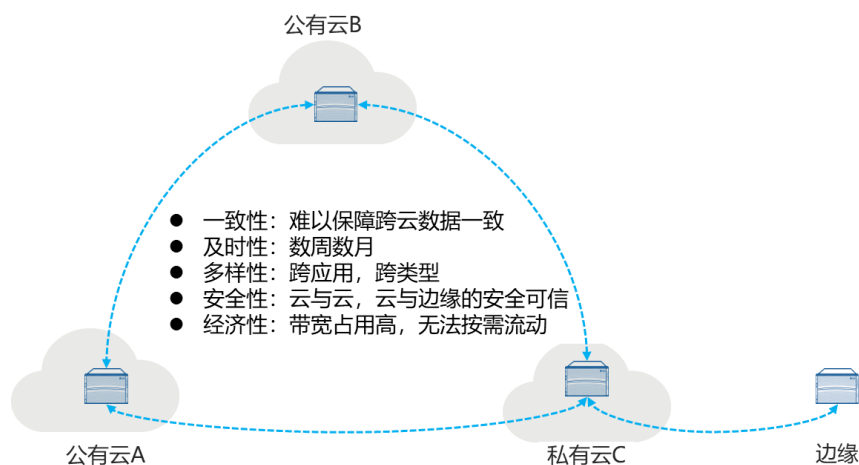


图 6 多云数据流动面临诸多挑战

数据流动往往发生在多个云之间，单个云按照各自的标准进行建设，事实上形成了数据孤岛，数据难以流动。当前，数据流动只能依靠针对特定应用的数据流动工具，难以在不同应用间实现，也难以保证数据流动后的一致性，造成数据在多个云之间完成流动并达到数据可用需要数周至数月，效率低下。跨云跨域的数据流动往往需要通过广域网，数据流动需要占用大量的网络带宽，但广域网的带宽资源是稀缺的高成本资源，需要进一步提升数据的传输效率，另传统通过点对点 VPN 来建立双向数据通道，也无法按需快速搭建。

随着企业数字化的深入，大量数据生成在端侧和边缘侧，多云环境下，企业数据的流动不仅仅需要跨云，云边端的数据流动也越来越多，但边缘侧存储空间受限、数据量大、可靠性低，并且数据传输带宽受限，网络时延不确定，云端和边端设施的数据按需协同流动就显的尤为重要，比如边缘数据分析的结果传回云端、原始数据留在本地，AI 训练在云端，训练完的算法模型按需更新到边缘。

（二）跨云数据共享难

企业采用多云建设模式之后，数据被分散在不同的云平台中，由于各家云平台的服务存在差异，生态难以互通，应用无法实现跨云的数据调用。多云间数据缺乏互操作性，每个云服务支持的数据交换标准不一样，大多数数据只是被扔到多云中，几乎没有跨云互操作性机制，导致数据无法共享。不同的云上技术堆栈采用了不同的数据平台，往往只能采用迁移数据的方式实现数据共享，虽然不同的云平台提供商提供了一些针对特定数据平台的迁移工具，但企业很难迁移格式不同且驻留在不同技术中的数据，也难以保证数据迁移后的完整一致，使得基于迁移方式的数据共享获取和维护成本很高。

（三）数据安全保护难

多云和大数据时代，数据规模持续增长，数据流转共享、融合分析成为数据运用的新常态，多云数据安全面临着开放性、规模性、分离性所带来的新挑战。一是数据承载边界被打开，数据将面临更加开放和未知的安全风险，阻碍了数据共享生态的形成；二是云数据中心存储了 PB/EB/ZB 级数据规模，数据泄露或者数据不可用、数据丢失都将会造成灾难性影响；三是云数据中心存储的海量异构的数据关系多维交叉、聚集增长，使得数据安全管控难度大大增加；四是数据的拥有权、使用权和管理权的分离，迫切需要解决可用不可见、数字主权可管可控等数据安全问题。五是云逐渐成为勒索病毒组织眼中的“数据金矿”，多云环境下，数据复杂度大大提高，如何确保数据始

终可恢复成了防勒索的最后底线。

（四）多云应用改造难



图 7 多云趋势下有状态应用云化改造越来越深入

企业业务创新持续加速，需要进一步缩短应用上市的周期，传统的虚拟化已经无法满足新型服务架构和技术选型的需要，拥抱容器成为了当今绝大部分企业的选择。应用云上开发云下部署既能利用公有云的敏捷弹性，降低开发测试成本，又能利用私有云保证数据安全和主权，由于云上云下的开发和应用架构不一致，业务割裂，无法保证应用的一致性交付，一致性运维，云上 PaaS/SaaS 服务下推到云下企业数据中心需要对企业 IT 架构进行容器化改造，有状态应用在进行云化改造时，云原生应用场景对服务的敏捷度、灵活性要求非常高，很多场景期望容器的快速启动、灵活的调度，这样即需要存储也能敏捷的根据应用容器组的变化而调整，快速的进行云盘的挂载、卸载。

（五）数据应用能效低

众所周知，云服务最大的成本就是电力，面对未来海量数据的增长，如何提升数据的应用能效将是提升数字化质量的关键。根据谷歌、

三星等研究机构 2018 年的一份研究报告²显示，当前计算体系架构下 CPU 利用率仍然偏低，计算设备约 62.7% 的能耗花费在数据搬运上，CPU 大量时间处于等待数据和数据搬运。同时云上数据存储资源利用率仍然很低，云存储常采用多副本方式保证数据可靠性，1 份数据保存 3 份副本，空间利用率仅为 33% 左右。据统计，数据中心中数据存储系统耗电量约占 IT 基础设施总耗电量的 35%，预计 2025 年数据存储的耗电量将达到 630 亿千瓦时，耗电量约占全国用电量 0.91%，数据绿色存储仍存在较大改善空间。

² 报告见：<https://dl.acm.org/doi/pdf/10.1145/3173162.3173177> Conclusion

四、多云数据存储发展趋势

为解决多云架构下的数据挑战，产学研用各界正在积极探索破局思路，推动数据存储持续创新升级，并形成了一些典型的发展趋势。

（一）面向多云构建统一数据底座，扩大数据共享应用

企业开始在多个云与 IT 平台上部署应用与数据，而在使用多云带来的弹性、敏捷的基础服务的同时，也要面对多云部署带来的数据孤岛和不同异构平台之间的冲突。为解决此问题，构建跨多云和企业 IT 的统一基础服务，将底层各种大规模云的计算、存储、网络、安全以及其他资源，抽象成统一多云的云服务底座成为产业界各方的共识。

其一、企业的多样化应用对数据存储的容量、性能、能效、安全、可靠、应用生态等有着不同的需求，多云架构需要构建足够宽的存储服务序列，并提供统一的存储资源池和存储云服务能力，才能更好的去满足企业多样化的上云场景。如云原生的 CDN 应用，要求很好的可扩展性和网络访问便捷性，大量采用对象存储；金融账务，支付交易等应用要求极低的数据存取时延，全面的容灾能力，多数采用专业的全闪存块存储；半导体领域关键的 EDA 仿真应用对小文件存取性能、数据快速检索要求很高，普遍采用专业的高性能文件存储；科学计算，人工智能等应用，对混合负载下的带宽性能、多种数据协议互通等能力有刚需，广泛采用专业的分布式并行文件存储。基于以上的因素，云平台厂商与数据存储厂商已经开始初步构建互联互通的生态，

并在政府，运营商，金融，电力等行业开始商用，如 AWS 可以与联想凌拓、戴尔科技、IBM、华为、杉岩数据等厂商的数据存储进行对接，阿里云可以与华为，新华三，浪潮等厂商的数据存储进行对接，华为云可以与 XSKY、新华三、杉岩数据等厂商的数据存储进行对接，华三云可以与华为、曙光、惠普等厂商的数据存储进行对接，浪潮云可以与 IBM、华为、XSKY 等厂商的数据存储进行对接。企业可以根据业务应用的需求，综合考虑云平台能力，数据存储能力，数据安全能力，最优成本等维度灵活的选择云平台和数据存储的组合。

云平台与数据存储的对接，一般优先采用通用的协议，如 OpenStack Cinder、Swift，VMware VAAI，Kubernetes CSI，AWS S3 等，同时在此基础上，产业界也在共同定义和扩展更多的企业级存储功能对接，例如双活、复制、快照等高阶能力。一方面满足云平台对数据存储集成性和管理性的需求，另一方面满足企业跨云一致的数据存取和数据安全体验。

其二、企业根据应用的类型、数据的安全，在企业数据中心或公有云部署不同的应用和数据，这就需要从以往的以单一云平台或单一业务为中心，转变为以数据为中心，构筑跨云跨数据中心的统一数据管理、数据流动，以实现数据存取和管理的最佳效率。统一数据管理包括存储资源的统一运维、统一数据视图、统一数据调度，数据流动包括企业自定义策略的各类数据分级存储。



图 8 跨云统一数据管理示意图

统一数据管理围绕企业多云架构的数据全生命周期，包含生产数据，容灾数据，备份数据，开发测试副本数据等各类数据，并对于热温冷数据分级调度。针对企业数据存储资源池，通过统一划分性能区间、服务类型、灾备配置、增值服务等指标，构建跨云全局一致的存储服务 SLA，并基于标准化 API 把存储资源提供给多个云支撑各类应用和数据服务，实现数据一池共享，应用多云部署。统一数据视图和统一数据调度通过全局文件系统实现，全局文件系统将所有云上存储资源整合到一起，基于统一元数据构建全局命名空间提供统一数据视图，用户看到的是单一存储空间，并通过文件、对象、HDFS 等多种标准协议，随时可视全局文件，在逻辑上整个多云数据底座成为一套存储系统，可进行数据全局调度。全局文件系统可以构建 EB 级以上的虚拟存储池，用户可以根据权限在同一个命名空间下共享访问跨云跨数据中心的数据。企业的数据管理员，可以定制统一元数据策略、检索统一元数据，随时看到各个数据中心文件状态，并对这些文件进

行管理，极大简化了多云环境下海量数据管理的复杂性。

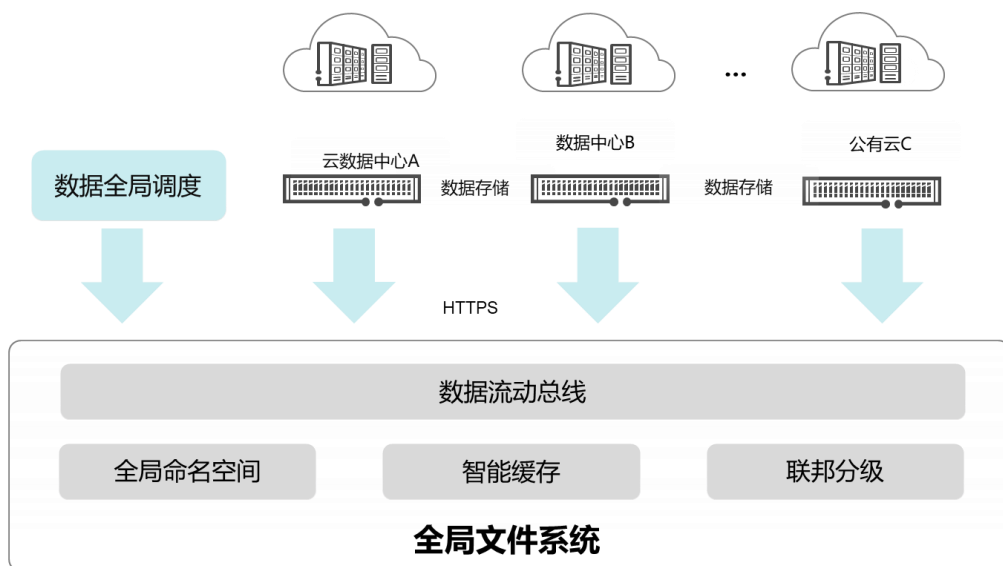


图 9 全局文件系统示意图

例如，在东数西算的场景下，用户在云数据中心 C 要进行 AI 训练，数据集产生自云数据中心 A 和云数据中心 B。通过统一元数据，C 可以看到全局的统一数据视图，并对训练数据集进行存取操作，这些数据文件实际上是全局文件系统在后台自动同步到 C 并本地缓存的，AI 训练并不感知数据流动过程，AI 训练后产生的模型参数数据自动传回到 A 和 B 进行 AI 推理，对 C 来说，只需要操作本地文件系统，数据的全局调度自动完成，达到 AI 应用无感。

国内多家金融机构已经构建面向多云的统一数据存储池，进行数据快照业务自动化发放、数据存储资源自动发放、统一数据管理，助力云化转型。

（二）数据存储智能分级能力增强，使能数据跨云流动

数据的流动性，是与生俱来的的根本属性，也是决定数据要素价

值的充要条件。如果数据没有流动，意味着数据无从应用，理论上数据也就没有价值。所谓数据要素，就是建立在流动基础上的数据。根据 IDC 的调研，85% 的大型企业管理者表示不会将所有数据只存储在一个地方，会采用更多适合自身业务的存储方法，即在多云环境中实现数据整个生命周期中的存储与流转。数据分级存储将架起企业数据中心和公有云数据中心的桥梁，是企业用户上云用数的新路径。

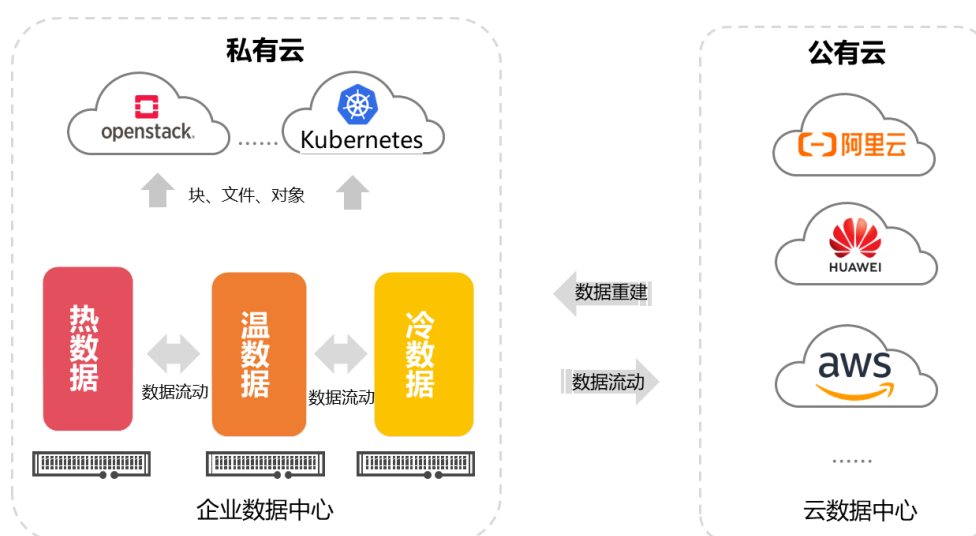


图 10 企业多云分级协同示意图

在企业数据中心、多个云之间，通过基于数据存储的分级（Tiering）、复制等技术，实现企业多云数据中心之间、不同类型数据存储之间基于策略的数据分级流动，以实现跨云数据复制，数据迁移，数据备份等。用户可以根据自身应用属性、数据属性部署一种或组合部署多种多云数据分级方案。

数据热度：用户可通过自定义策略，根据数据热度在不同的存储设备，云上云下之间按需流动，高频访问的热数据流向高性能存储提

升访问体验，低频访问的温冷数据流向容量型温冷存储降低存放成本。用户不需要对存储设备进行频繁的更换，而是将新存储接入分级存储资源池中，重要数据可根据策略迁移到新设备中，老设备可作为温冷数据存储资源池

归档：将企业数据中心或边缘的海量数据有甄别地通过分级方式归档到公有云存储或者蓝光存储，有效应对海量数据爆发式增长对存储空间的需求及数据长期留存需求。

容灾：通过快照等技术，将企业数据中心的数据以复制或拷贝方式存放至不同的云，实现多个云之间的业务连续性保障。降低数据流动成本。例如，通过数据重删或压缩等技术，进行数据缩减后再上云流动，降低对传输网络带宽的需求，从而进一步降低数据上云的总体拥有成本。通过全局文件系统和透明的分级存储技术，构建跨云全局命名空间实现多云数据的快速检索和访问，数据流动业务无感知，简化了业务系统开发，避免了大量数据流动占用网络带宽。

（三）数据存储面向容器持续演进，助力应用云化改造

云原生场景下，单个容器集群通常位于某个公有云或者私有云中，用户往往会选择多个公有云搭建自己的容器集群用于支撑数据安全不敏感且具有大量网络流量的业务或者关键业务的开发测试，而对数据安全性和服务稳定性要求较高的关键业务则需要企业数据中心的私有云环境搭建容器集群。随着以 Kubernetes 为核心的云原生技术的日益普及，企业拥有的 Kubernetes 集群的数量也呈现爆发式的增

长，海量集群的管理和应用的分发，给企业带来了不同程度的问题。业务在集群之间需要差异化配置以适应不同的云平台，各个集群间应用的统一发布、升级和流动不易。容器在设计初期，是面向无状态应用的，而随着应用改造的逐步深化，容器需要承载有状态的关键应用，其在容器化改造过程中面临着各种问题，最显著的问题是容器数据持久化的问题，Kubernetes 社区通过制定 CSI(Container Storage Interface) 标准，使得容器可以调用外置存储资源，一定程度上缓解了这一问题，更令企业头疼的是，容器欠缺支撑有状态关键应用的灾备机制，如果业务系统有节点级、集群级的高可用要求，则必须由业务系统自己实现数据层面的高可用，这极大的增加了业务系统的复杂度，成为企业应用云化改造的一大阻碍。

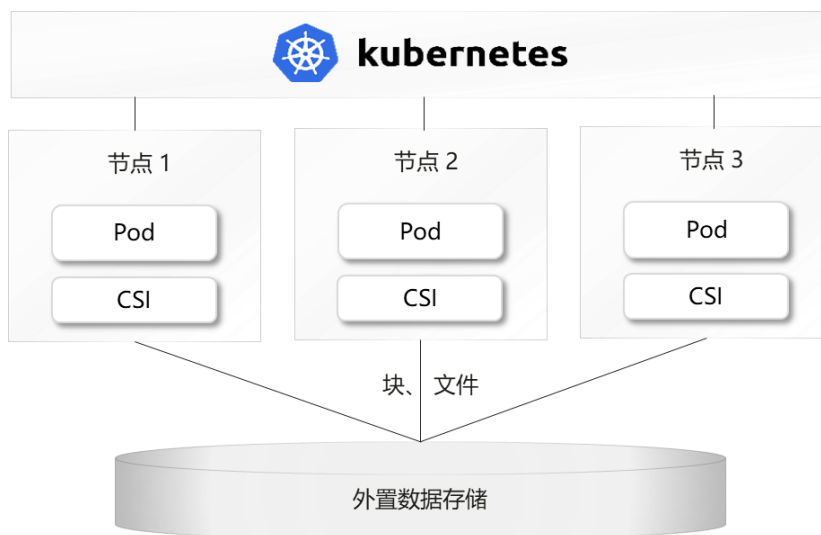


图 11 有状态容器数据持久化示意图

为解决以上的阻碍，云原生时代的存储系统，除了要满足性能、稳定性、可靠性的要求以外，还要满足业务系统对敏捷性的要求，能

够通过统一的编排系统配合业务上线，并且可以实现快速扩容。首先存储支持 CSI 规范，实现与容器的对接，在这一基础上，NAS 是最适合作为容器底座的存储，原因一，业务系统可以非常方便的挂载和卸载到 NAS 文件系统上，契合容器弹性伸缩的诉求；原因二，NAS 拥有丰富的企业级特性，如多租户、QoS、配额等，与容器面向应用级的多租户、QoS 等诉求可以配合使用。此外，NAS 还可以将文件系统开销卸载到存储上，在大并发应用下，性能不受主机操作系统开销影响。根据调研，当前 73% 的容器用户都使用 NAS 作为容器底座。数据存储可以通过 CSI 在应用侧敏捷的发放各类存储资源，并实现与容器 Pod 的自动挂载，快照、克隆等高级企业特性则大幅提升了开发人员体验，帮助企业实现 DevOps。

此外，在数据存储、CSI 和灾备插件的配合下，可以为容器提供应用级的跨云、跨数据中心容灾，使云原生关键应用拥有和传统应用一样的容灾能力，避免单节点故障、单数据中心故障和集群故障，达到 RPO=0 高可用容灾级别。同时这种高可用机制只需通过存储底层实现，应用无需为考虑高可用进行任何的改造，使企业只关注应用本身，加速企业 IT 向多云架构迈进。

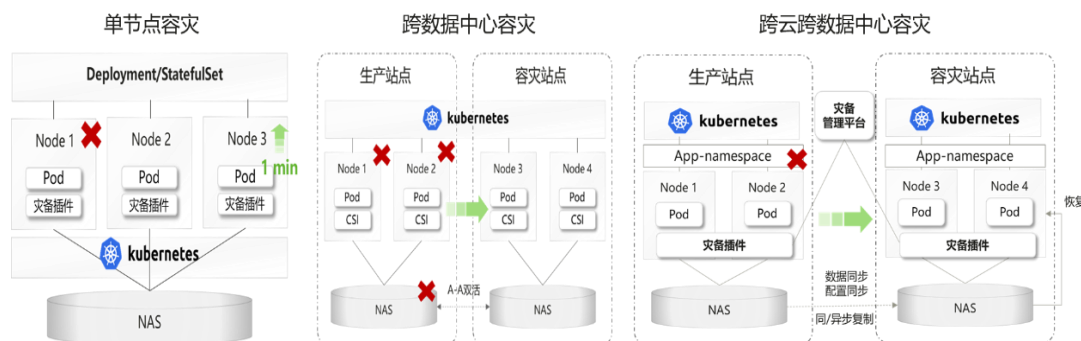


图 12 容器应用级高可用示意图

国内数个银行已经开始基于此架构实现应用上云，如四川天府银行、北京银行，为加速数字化转型，携手容器云提供商、数据存储提供商，部署云原生平台和高性能数据存储，有状态应用基于 CSI 接口可动态申请、使用底层存储资源，实现了存储资源的标准化自服务，降低应用开发者使用云资源的配置门槛，实现敏捷可靠、降本增效，为关键数据库容器化部署提供了有力支撑，助力金融应用顺利上云。

（四）以数据为中心新型架构兴起，提升数据应用能效

随着企业上云用数赋值的不断深入，基于大数据、AI 等数据密集型的应用越来越多，企业存储的数据快速增长，达到几百 PB 甚至 EB 级，传统云数据中心采用以 CPU 为中心的存算架构，所有数据都需要先被送到 CPU，CPU 的计算包含了应用的计算，通信的计算和存储的计算等，每一类计算都需要独占 CPU 资源，此外，由于 CPU 并不擅长数据处理运算，导致其能效比低下，例如，服务器内的 CPU 为处理网络及存储 IO 请求，需要消耗高 30% 的算力，因此以 CPU 为中心的架构已然成为云数据中心释放算力、存力的瓶颈，造成数据存

算性能难以线性增长，能效难以提升，能耗成倍增加。

鉴于当前以 CPU 为中心的架构所存在的局限，新型存算架构正在迅速兴起，它以数据为中心进行构建，由算力池、存储池、高通量网络等灵活组成，能提升 30%数据存算效率、降低 20%系统成本、节省 10%耗电量。

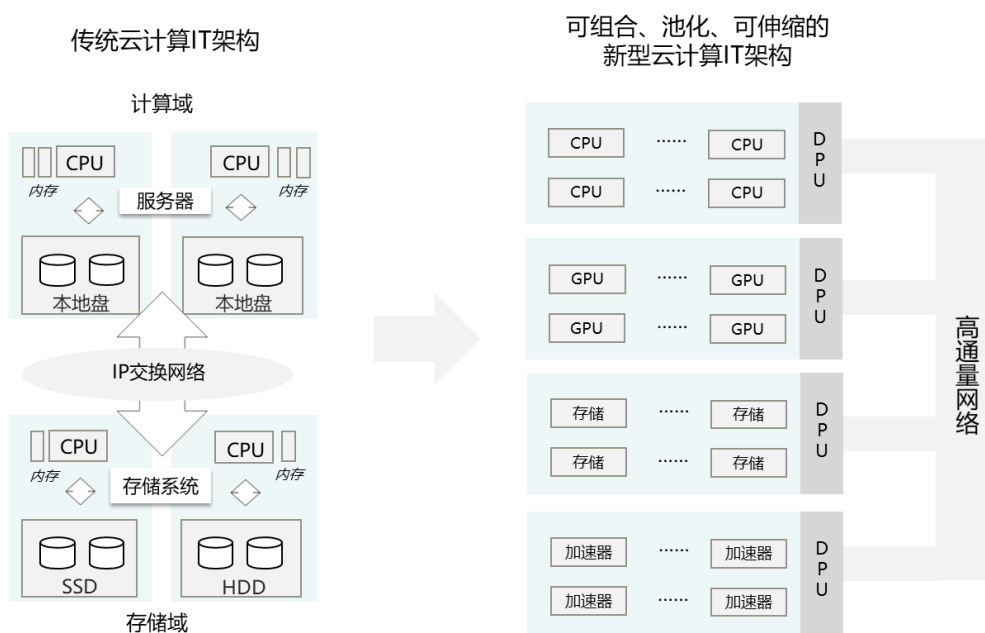


图 13 传统云计算 IT 架构与新型云计算 IT 架构示意图

算力池：

新型存算架构将数据存储和访问等操作从通用处理器剥离，卸载到以 DPU（Data Processing Unit，数据处理器）为代表的专用处理器将上，相比通用处理器，能以更低的功耗处理更多的数据，从而释放 CPU 算力、提升算力能效比。此外，这类专用处理器还具备即插即用、即换即用等优势，在硬件使用上更为灵活，在运行状态下不与业务应用发生资源争抢，保证用户业务正常运行的同时也保障了基础设施的

服务质量。引入 DPU 等专用处理器后，算力池化是必然选择，将打破传统以 CPU 为中心的存算架构，由此提升数据应用的效率。此外，特定的数据操作可由专用硬件加速器进行进一步加速，如纠删码、加密压缩、网络通信等。

存储池：

存储池主要以以太网闪存簇（Ethernet Bunch of Flash, EBOF）、以太网内存簇（Ethernet Bunch of Memory, EBOM）、以太网磁盘簇（Ethernet Bunch of Disk, EBOD）等新型盘框形态存在。RAID、纠删码等传统存储能力下沉到新型盘框中，形成“盘即存储”，对外通过 NoF 等高速共享网络提供块、文件等标准存储服务，可减少冗余资源开销比例。此外，还可通过基于专用处理器加速的场景化数据缩减技术提供更多数据可用空间。这类新型盘框对传统磁盘阵列的冗余池化技术和数据缩减技术进行了高度集约化和小型化，将服务器的本地盘和内存拉远进行池化共享，有效支撑了传统云数据中心向极简分层的新型存算架构演进。

高通量网络：

高通量网络决定了系统的响应速度以及吞吐能力，也决定了系统资源池化的能力范围，可以选用基于 CXL Fabric、NoF（NVMe over Fabric）、IP 等多种网络协议的组合进行组网。面向热数据处理，通过 NoF 高速网络实现闪存池化；面向温冷数据处理，通过 IP 高速网络实现磁盘池化；面向极热数据处理，通过 CXL Fabric 等新型内存

型网络实现内存资源池化，将网络时延降到亚微秒级别，可构建 PB 级内存资源池，为业务提供更大的共享内存空间。

以数据为中心的新型存算架构相对于传统云数据中心架构具有显著的优势。一、充分利用 DPU 等专用处理器的硬件卸载和加速能力，可缩短 IO 访问的响应时间，提升 IO 效率，以 CPU 为中心的数据处理场景里，由于数据转发都需要经过 CPU，CPU 的处理时延通常会比较高，约 30 至 40 微秒，在新型存算架构中，数据转发不再需要 CPU 的软件转发，DPU 可以直接送达数据，时延可以提升到 3 至 4 微秒，数据应用性能提升近 10 倍。二、显著提升数据应用能效，剥离了数据存储、传输和安全等非计算操作后，CPU 的算力能充分释放，资源利用率预计能提升到 60%。三、存算资源的更灵活更合理调度，它打破各类存算硬件资源的边界，组建彼此独立的硬件资源池，对 CPU、GPU、存储、内存和加速器等资源独立扩展及灵活共享，使得提供给用户侧的资源上限更大，实现能效比最优的硬件组合。

目前，产业界以数据为中心的新型存算架构创新不断。在供应端，英特尔、英伟达、阿里、华为、中科驭数等均推出专用数据处理芯片来卸载 CPU 的数据处理功能，西部数据推出 OpenFlex Data24 EBOF 盘框、华为推出 OceanDisk 智能盘框等，存储系统和介质厂商已经开始面向新型架构推出相应产品。在需求端，天翼云、移动云、中国电子云等云服务企业也已积极投入到新型存算架构的实践中。

（五）边缘加速向超融合架构演化，云边协同不断深入

边缘成为云架构下新的数据爆发源头，据 Gartner 统计预测，75% 的企业数据将会在传统数据中心或云平台之外的位置创建，且超过 50% 的数据需要在边缘侧分析、处理和储存。云边协同包括许多应用场景，比如 IoT、远程采矿、园区入口电子识别、远程智慧养殖等，其特点是数据要在边缘侧本地快速处理，并通过网络将处理结果传输到中心云，可在发生断网、网络拥堵等特殊情况下，保证业务系统的基本功能性运转，中心云作为关键数据的最终存储设施，既可保证数据的安全性，又可对所有分支边缘数据做全局处理。

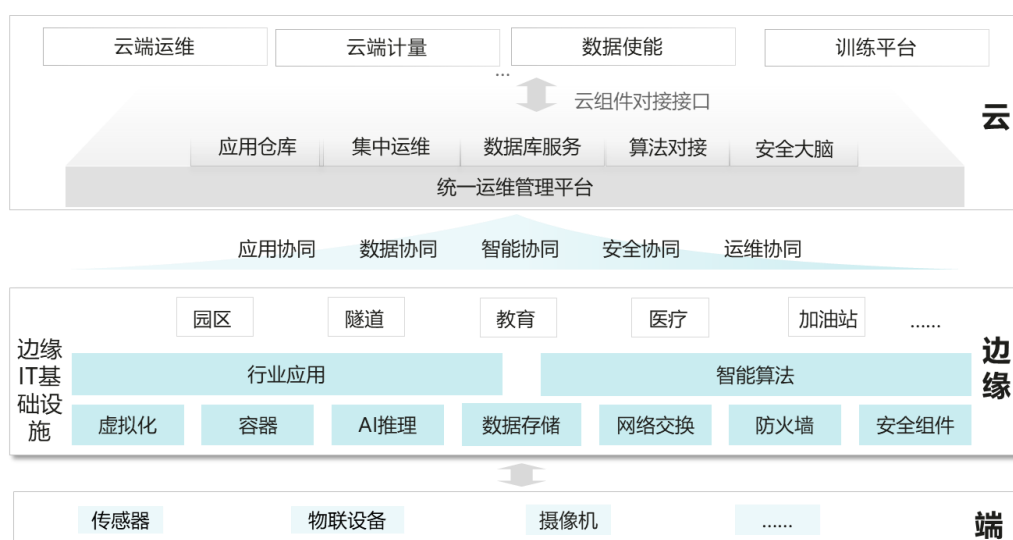


图 14 云边端协同示意图

云边协同的能力与内涵主要包括资源协同、应用协同、数据协同、智能协同、管理协同。

资源协同：主要是对边缘节点的 ICT 基础设施和各类资源的协同（包括计算、存储、网络 and 虚拟化资源），对多样的底层硬件进行抽

象，使得上层应用可以方便地使用底层硬件提供的计算存储等能力，同时依据业务的资源占用、流量、运行状态等，从全局视角，对计算资源、网络互联带宽和延时、存储资源和应用软件资源完成调度和配置，以达到全局的业务整合。

应用协同：应用在云上开发，边缘节点提供应用部署环境，通过应用商店等方式进行应用的一键式边缘部署，以实现业务快速上线，达到云上云下一致性体验。

数据协同：针对边缘计算数据量大、隐私数据要求本地处理的特性，日常在边缘侧对数据进行存储和分析整合等实时处理，然后将处理结果协同上报中心云。为了妥善保护数据，可采用认证和授权、加密、数据屏蔽、WORM 写保护等技术。对于重要数据可进行边缘节点到中心云的备份、边缘节点与边缘节点之间的容灾。

智能协同：云上开展 AI 的集中式模型训练，并将模型下发到边缘节点，边缘节点按照 AI 模型执行推理，实现分布式智能；云上训练边缘推理的方式使得 AI 算法识别响应更快，效率更高，识别场景更丰富，而且不受网络条件的影响。

管理协同：边缘节点部署偏远，且很难配备专有的运维人员，通过端到端可视化的统一管理平台是提升运维效率最有效的途径，在云上多边缘远程集中管理，能够快速识别问题所在，减少问题定界时间，实现边缘节点无人值守。同时考虑到云与边缘之间网络的不稳定性，边缘节点也需具备离线自治管理和运行的能力。

功能齐全、集成度高、宽环境适配的超融合设备成为边缘设施的首选，边缘设施规模较小，一般不超过 5 个机柜，但对 ICT 的各个模块均有需求，超融合设备集成了服务器、存储、AI、UPS、防火墙、网络等 ICT 基础设施能力，能满足边缘节点的全部功能诉求，同时出厂预置虚拟化、容器、云上云下统一管理软件、应用商店等，能够很好的实现资源协同、应用协同、数据协同、智能协同和管理协同，同时超融合设备的 2 台轻量起配，低噪，通用市电，短机身设计，可塔式独立部署等宽环境适应性也能够满足边缘节点复杂多样的环境要求。

（六）安全可信数据存储部署加快，保障数据可靠安全

不管是公有云，混合云还是私有云，一旦多云数据存储的安全防线被攻破，企业核心的数据将会被泄露。综合考虑多云时代数据存储面临的安全挑战，安全可信的数据存储从存储系统防护、数据安全防护两个层面保障多云数据安全。



图 15 安全可信数据存储示意图

存储系统防护：数据存储是企业数据的存放室，如果被攻击控制，后果不堪设想，加强存储系统的安全能力尤为重要。其一是存储系统底层安全：通过实施安全配置、安全补丁，安全启动、远程证明、认证鉴权、访问控制、web 安全防范等机制对存储系统做安全加固增强；其二是存储软件供应链安全：Black Duck 的代码库审计报告指出，60%的开源程序和代码库含有漏洞，40%以上的软件代码库包含高风险漏洞，因此需要通过开源软件管理、采购第三方件管理、开源和采购第三方件完整性校验等机制确保引入的软件包没有被植入恶意程序，且符合国家或者行业信息安全规范。其三是存储软件安全：通过建立安全可信的软件研发过程，制定安全编码规范，定期安全代码扫描等措施把安全性植入软件开发生命周期。

数据安全防护：数据的安全风险主要来自信息泄露导致的机密性丧失，篡改或者软硬件灾难导致的完整性、真实性和可用性被破坏。数据存储会围绕数据采集、传输、存储、处理、共享和销毁全生命周期构筑安全防护能力，保证数据的机密性、完整性、真实性和可用性。其一是围绕数据全生命周期，提供端到端加密能力，包括数据传输加密，落盘加密等，提升数据流动和数据储存的安全，并在数据生命周期结束时提供数据安全销毁能力；其二是针对数据完整性、真实性和可用性提供安全快照、本地备份、异地备份、双活容灾等数据容灾与备份能力，结合企业数据分类分级可实施符合政策法规要求的数据保护措施；其三是针对日益猖獗的勒索病毒攻击提供网络+存储协同防

勒索的能力，实现事前防御勒索病毒攻击、事中进行勒索病毒检测、事后保障数据可恢复。勒索病毒隐蔽性极强、变种频繁，往往使用零日漏洞、钓鱼邮件等方式入侵，单靠杀毒软件防不胜防；网络防火墙等设备以进不来为防范目标，可拦截 90%左右勒索病毒，而数据存储作为数据的最终载体，可以始终在第一时间感知勒索软件对数据的修改行为，通过侦测分析、安全副本、及时恢复，确保病毒进不来、改不了，数据可恢复，形成数据安全最后一道防线。

（七）数据存储节能技术逐渐成熟，促进绿色低碳发展

当前正处于“碳达峰、碳中和”国家战略之下，云数据中心作为数字经济时代的核心基础设施和战略资源，承载了绿色高质量发展的重要使命，在确保云数据中心业务能力满足需求同时，对云数据中心的绿色化提出了更高的要求。除了降低 PUE 之外，降低包括数据存储在内的 IT 设备能耗，对于促进云数据中心绿色节能也至关重要。

数据存储的高效节能技术包括：介质节能、高密化设计、数据融合、算法节能、风液冷设计等，都可应用于云数据中心节能提效。



图 16 绿色数据存储示意图

全闪存化存储：

闪存介质有高密度、高可靠、低延迟、低能耗等特点。随着全闪存存储的快速发展，基于闪存介质的固态硬盘数据访问比机械硬盘快 100 倍左右，吞吐量大 100 倍，单盘的 IOPS 大 1000 倍以上，电力消耗减少 70%。随着存储通信协议的发展，把存储介质变化的效率和网络带宽提升的作用充分发挥，这些变化促使了计算机系统处理能力的大幅提升，优化了数据中心算力与存力的基础能力结构，加速了数据资源的快速流通。

高密化设计：

受限于散热和空间限制，普通的通用型服务器配备硬盘的数量有限，通常为 1U10 盘、2U24 盘、4U36 盘。而专门设计的高密存储型节点，能做到 1U32 盘、2U36 盘、4U60 盘、5U80 盘、4U80 盘、5U120 盘等，密度达到传统存储服务器的 2~2.6 倍，结合存算分离架构，相对使用通用型服务器，减少了节点 CPU、内存及配套交换机，同等容

量下带来能耗节约 10%~30%。

数据融合：

随着分布式存储数据融合技术的成熟，一套分布式存储系统能够同时提供文件、对象、大数据访问能力，系统规模从 PB 级走向 EB 级，允许多云多业务共享存储系统的硬件资源，提升资源利用率。同时通过多协议融合互通能力，一份数据无须协议转换就能够被多种协议同时访问，解决业务流程中多环节多应用访问相同数据的需求，减少数据搬迁和重复存储，提升 35%的数据处理效率，降低约 20%能耗。

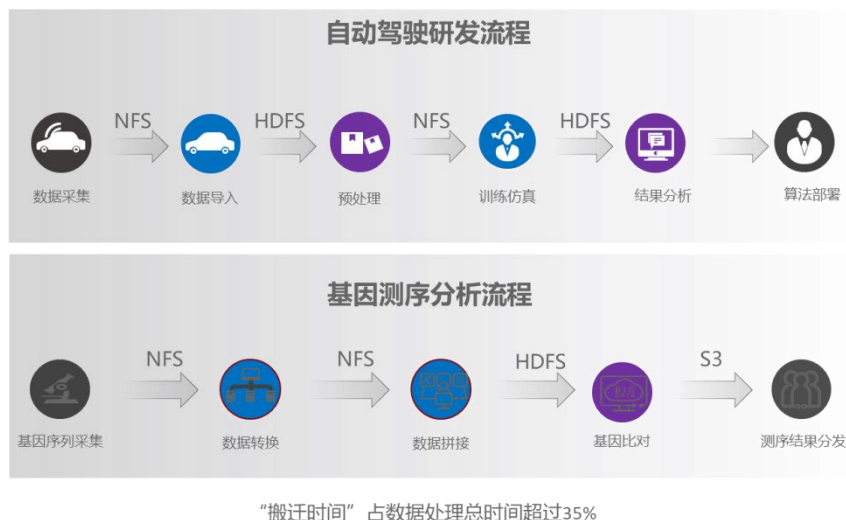


图 17 多协议互通减少数据搬迁示意图

算法节能：

数据纠删码（Erasure Code，EC）是指将要存储的数据分割成小的数据块，通过算法将数据块进行编码生成校验数据块，然后把这些数据块（包括原始数据块和校验数据块）存储在不同的存储位置，比

如不同磁盘、不同存储节点、不同存储区域等。

相较于多副本冗余策略，18+2、22+2 等大比例的数据纠删码冗余技术能显著提升空间利用率，从而降低建设成本和运行能耗。在大数据分析场景，采用存算分离架构后，利用数据纠删码（Erasure Code, EC）技术替代传统三副本数据冗余，可以把磁盘利用率从 33%提升到 91%，降低能耗 40%。

风液冷设计：

随着单器件功率密度增大，例如单 CPU 功耗高达 350W 以上，常规风冷无法满足散热要求，采用风液冷技术与数据存储结合已经成为趋势。当前主要有浸没式液冷、风液结合、水冷门等实现方式；风液结合技术中 CPU、GPU 等大功率器件采用液冷，其他器件可采用风冷/液冷，显著降低内存、HDD 等存储关键部件的工作温度，可降低风扇转速 50%左右；水冷门技术采用特殊机柜降低机柜出风口温度，从而优化整机房空调制冷循环。

五、总结与展望

云的发展伴随着数字化转型，经历了尝试、探索，现在进入了全面践行时代，企业在享受云带来的敏捷、创新以外，也要解决由此带来的一系列的挑战和痛点。绝大多数企业不能简单的抛掉已运行多年的系统，重新在云上搭建新的平台，因为这是一个多云共生的市场，需要考量技术、成本、合规、安全、环境等各个方面的要求和限制，企业需要带着多云的思维来改造、构建 IT 系统，支持业务的可持续发展。多云架构中，不管是企业自建私有云，还是使用公有云的占比是需要根据不同的企业数字化战略、法规遵从和综合成本分析进行动态调配。

可以预见，多云不会是云发展过程中的一个临时阶段，它将是一个 ICT 基础设施服务的新常态，从“上云”到“多云”，从“用云”到“用数”，云上或云下的单一数据存储模式也一去不复返。面向企业上云用数的目标，实现数据自由流动、数据高效共享，数据安全可靠，数据应用协同和数据绿色低碳，数据存储作为云的数据底座，将伴随着云的发展不断创新，进一步帮助企业释放数据要素价值。



图 18 多云数据存储助力企业上云用数

云与数据存储的互联互通至关重要。面向未来，白皮书编写成员共同发起倡议：

完善多云数据存储能力，促进数据共享、跨云自由流动

多云时代，数据流动产生价值。数据的自由流动面临的诸多挑战，最适合解决的位置就是数据存储层，需要云提供商和数据存储提供商共同创新，提供更加全面的数据服务，保障多云数据交互安全可信，消除数据自由流动的障碍，让数据产生更大的价值。

联合行业用户进行多云数据存储创新实践，助力行业推广应用

建立典型应用实践，加强行业应用案例推广，形成可推广的经验成果。与金融、政府、运营商等行业用户联合创新，组织开展优秀案例评选、行业技术交流等活动，助力多云数据存储创新在行业的推广应用。

合作构建多云数据互联互通标准，搭建产业合作平台，构建开放的数据流动环境

当前多云数据互联互通标准建设在早期阶段，建议围绕实际应用

场景，联合产、学、研、用各方，通过业界优秀实践孵化行业标准，建设高质量标准体系。搭建产业合作平台和多云生态实验室，支撑互联互通测试、评估认证等工作。

中国信息通信研究院

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-62300072

传真：010-62304980

网址：www.caict.ac.cn

