



2022 分布式存储市场调研报告

百易存储研究院
中国计算机学会信息存储专委会
中国计算机行业协会信息存储与安全专委会
DOIT 传媒

2022 年 4 月 14 日

目录

一、分布式存储的定义与分类 P1

1. 定义
2. 分类

二、分布式存储与其他相关技术的关系 P3

1. 与集中式存储的关系
2. 分布式存储、软件定义存储和云存储的关系
3. 超融合市场现状
4. 超融合一体机还是存算分离？
5. 从超融合到混合云？
6. 纯软件还是软硬一体机？
7. x86 还是专用设备？
8. 分布式存储是否适用于“关键业务”应用？
9. IPFS 分布式存储与去中心化
10. 对象存储的话题

三、分布式存储市场状况 P12

1. 分布式存储市场规模
2. 分布式存储市场英雄榜

四、分布式存储行业典型的应用场景 P14

1. 电信
2. 汽车制造
3. 医疗
4. 基因测序
5. 高性能计算
6. 卫星遥感
7. 高校
8. 金融
10. 交通
11. 证券
12. 智慧城市
13. 基金

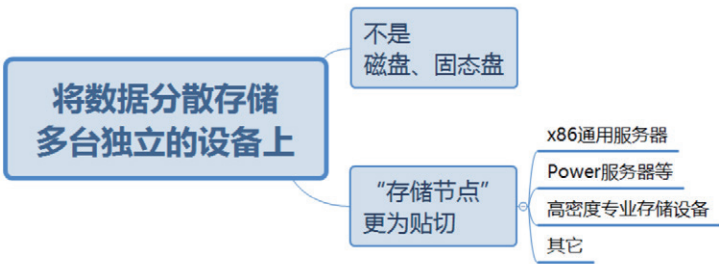
五、总结 P28

一 . 分布式存储的定义与分类

1. 定义

分布式存储发展至今，市场上并没有一个能够被广泛接受并引用的定义。

对于分布式存储，百度百科给出的定义是：为了简化用户端的使用，提供了一个分布式缓存系统来提供对此分布式存储系统的访问接口以及本地数据缓冲以降低网络压力。分布式存储系统是将数据分散存储在多台独立的设备上。传统的网络存储系统采用集中的存储服务器存放所有数据，存储服务器成为系统性能的瓶颈，也是可靠性和安全性的焦点，不能满足大规模存储应用的需要。分布式网络存储系统采用可扩展的系统结构，利用多台存储服务器分担存储负荷，利用位置服务器定位存储信息，它不但提高了系统的可靠性、可用性和存取效率，还易于扩展。

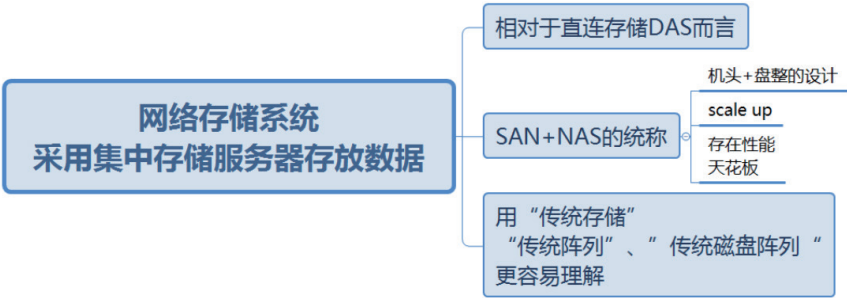


这个定义或者解释，还非常拗口和难以理解。有几个地方需要说明和解释。

首先：分布式存储系统是将数据分散存储在多台独立的设备上，这里的“多台独立的设备”，不要理解成为磁盘或者 SSD，可以理解为 x86 通用服务器、Power 服务器等，也可以理解为专门设计和优化的专用存储设备。其中，以 x86 通用服务器最为普遍。为了帮助用户理解分布式存储，百易存储研究院建议用“存储节点”来代替“多台独立的设备”，这里的存储节点，可以是 x86 服务器（或者称为存储服务器、服务器存储）、Power 服务器，也可以是专用存储设备，甚至使用传统阵列，其中，最为普遍的还是 x86 通用服务器，其市场较为成熟，非常容易获得，产品性价比非常具有竞争力。

其次，“传统的网络存储系统采用集中的存储服务器存放所有数据，存储服务器成为系统性能的瓶颈”的表述中，网络存储系统的表述多用于与直连存储（DAS）的区分，主要突出其能够通过网络被多个应用系统共享，网络存储系统又可以划分为：存储区域网络（SAN）和网络连接存储（NAS），是 SAN、NAS 两种存储系统的统称。

这里所说的“存储服务器”主要指的是阵列控制器（也称为机头）+ 盘阵的设计方式，以往多采用专用集成电路芯片（ASIC）设计控制器，经历了双控到多控的变化，控制器与控制器之间采用专用网络连接，控制器与磁盘之间通过专用后端网络连接，从而实现了连接的全冗余架构设计，具有非常高的可靠性和稳定性。也因为如此，系统成本居高不下。



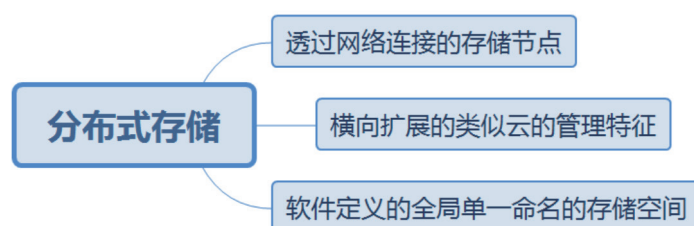
在系统的扩展性上，可以通过盘阵中盘的数量进行调节，系统具有一定的弹性，所谓 Scale Up，也就是纵向扩展，但受制于机头的性能，整个系统仍然存在性能天花板，用户只能借助入门级存储、中端存储、高端存储的升级来解决问题。

根据以上的判断和理解，百易存储研究院认为用传统存储、传统阵列或者传统磁盘阵列的表述更为便于理解。

最后，“分布式网络存储系统采用可扩展的系统结构，利用多台存储服务器分担存储负荷，利用位置服务器定位存储信息，它不但提高了系统的可靠性、可用性和存取效率，还易于扩展”的表述中，分布式网络存储系统就是我们要说的分布式存储；可扩展的系统架构更多是指可以通过增加或者删减存储节点所带来的横向扩展（Scale Out）的架构，存储系统的性能可以随着节点的增加或者减少呈现动态变化的特征。这里提到的“位置服务器更好地理解应该是分布式存储中的全局统一命名空间的元数据管理的部分。

通过上述分析，百易存储研究院认为有关分布式存储的定义应该关注几个重要的特征：一是横向扩展，一通过网络连接的存储节点，以及软件统一调度管理全局统一命名空间的存储资源池，这应该符合分布式存储的发展规律。仅仅强调软件定义，或者 x86 通用服务器；强调云的弹性管理的特点，都不能够完全准确的描述分布式存储系统，也很难取得市场的广泛共识。根据以上的判断和理解，百易存储研究院认为用传统存储、传统阵列或者传统磁盘阵列的表述更为便于理解。

最后，“分布式网络存储系统采用可扩展的系统结构，利用多台存储服务器分担存储负荷，利用位置服务器定位存储信息，它不但提高了系统的可靠性、可用性和存取效率，还易于扩展”的表述中，分布式网络存储系统就是我们要说的分布式存储；可扩展的系统架构更多是指可以通过增加或者删减存储节点所带来的横向扩展（Scale Out）的架构，存储系统的性能可以随着节点的增加或者减少呈现动态变化的特征。这里提到的“位置服务器更好地理解应该是分布式存储中的全局统一命名空间的元数据管理的部分。



2. 分类

传统存储按照功能可以分为 SAN 和 NAS，分别用于块数据、文件数据的存储支持。随着技术的不断发展和用户需求的变化，也出现了统一存储和融合存储的需求，在存储系统的设计上，也出现了软件、硬件分离的设计需求。

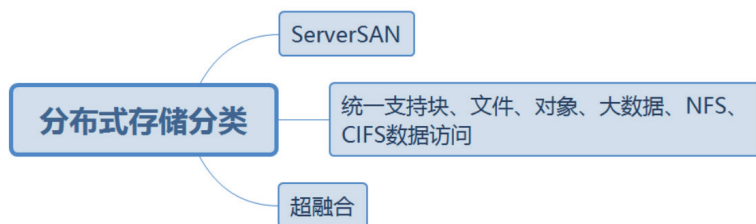
对于新分布式存储系统而言，最初设计需求就是替代传统 SAN 存储，并用 ServerSAN 以示区分。但是相比于传统存储的全冗余连接设计，分布式存储系统中的存储节点可靠性相对不高，为此，分布式存储系统通过多副本、纠删码的方式来提升系统的可靠性，为了提升这个存储系统的性能，分布式存储系统需要存储节点配备 SSD，充分利用 SSD 高 IO 的特性。

与此同时，分布式存储横向扩展的特性，非常适用于海量数据存储，特别是非结构化的文件系统。一个分布式存储系统往往同时满足块、文件、对象、大数据，以及 CIFS、NFS 等的访问需求。因此，分布式存储很难按照 SAN、NAS、对象进行分类。

对于分布式存储系统，根据是否对外提供应用计算能力，区分为分布式存储和超融合。对于超融合系统而言，最主要的技术突破就体现在存储上，也因为如此，才会把超融合纳入到分布式存储的分类中。

在超融合的产品形态上，最初是以一体机的形态呈现，但随着市场的发展，超融合也提供纯存储的形态，支持将超融合一体机作为单一存储系统的场景。

分布式存储、超融合也是相互渗透、融合发展的，并没有严格意义上的形态区分，这一点应该引起大家的重视。



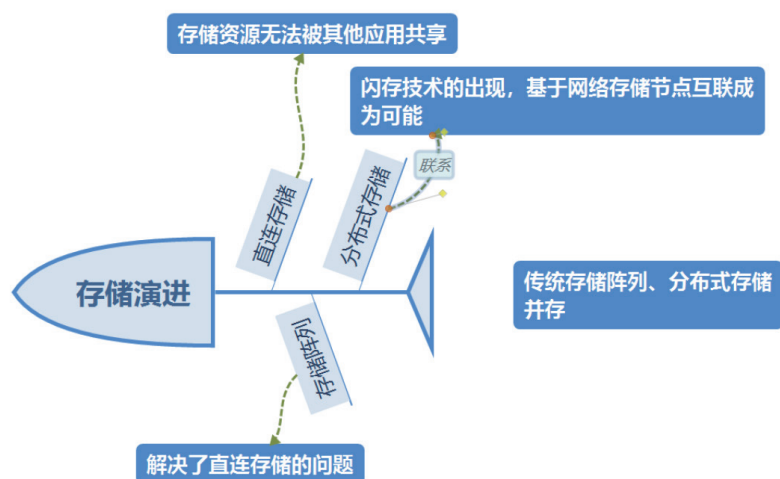
二．分布式存储与其他相关技术的关系

1. 与集中式存储的关系

从直连存储到传统磁盘阵列，再到分布式存储的出现，从演进过程来看，传统磁盘阵列要解决的是直连存储无法被其他服务器应用共享的问题，传统磁盘阵列可以被多个应用网络共享。

分布式存储的出现，主要源于 CPU 算力的提升和 SSD 固态盘的普及，闪存技术使得 IO 性能显著提升（5 倍左右），这就为基于网络的分布式存储创造了条件，如果没有闪存技术，磁盘 IO 性能是无法满足分布式存储的需求。

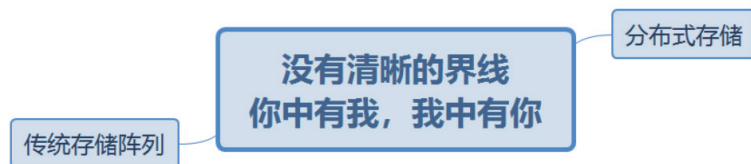
对于用户来说，事实上就存在着两种存储方式的选择：传统磁盘阵列后来发展到全闪存阵列，与分布式存储后来发展到分布式全闪存，两种存储形式并存。



两种存储形式并存无疑增加了用户选型的复杂性，那么在两种存储形式之间，有没有一个相对侧重呢？从应用的角度来看，有稳态和敏态之分，其中，数据库、ERP 等应用多属于关键业务应用，特别强调可靠性、稳定性，比较吻合传统存储阵列的特征；相比互联网应用大多属于敏态应用，强调敏捷、弹性，这与分布式存储横向扩展的特征比较吻合。

按照稳态、敏态来选型存储的类型，并不是绝对的。分布式存储也可以使用在数据库、ERP 等应用场景，提供块存储的支持。传统存储阵列中，NAS 应用也是针对文件存储的需要，如今，数据爆炸性增长，主要来自文件等非结构化数据，而分布式存储横向扩展的特征，也完全符合这个趋势。为了提升数据检索的效率，对象存储在互联网行业得到了普遍使用。

按照性能、可靠性等指标，也不适于划分传统存储阵列和分布式存储应用，原因很简单，无论是 IOPS，还是延迟，都没有一个适用的指标把二者区别开来。全闪存阵列性能指标，分布式存储并非不能够实现。可靠性方面，传统存储阵列的全冗余设计，的确提供了以盘为单位的数据保护，具有很高的可靠性、稳定保障，但是随着网络技术的不断进步，全闪存阵列等高端存储也摒弃了传统后端存储连接的方式，也开始借助以太网、Infiniband 网络对盘阵进行连接，出现了融合发展的趋势；另外，分布式存储也借助多副本、纠删码等机制对于数据进行保护，当节点或者单盘出现损坏的时候，可以借助全部节点的计算和存储的资源，群策群力对数据进行恢复，确保数据的可靠、不丢失，确保系统性能的稳定。



在传统存储阵列、分布式存储之间并不存在一个清晰、明显的界线。你中有我，我中有你。

2. 分布式存储，软件定义存储和云存储的关系？

日新月异，新技术和概念迭出，这是信息科技（IT）最基本特点，产品供应商为了突出品牌的领导地位，有意无意会选择一些新的概念。分布式存储也是如此，有的厂商选择分布式存储，也有厂商选择软件定义存储，云厂商对外更愿意使用云存储，三者之间会有什么不同吗？

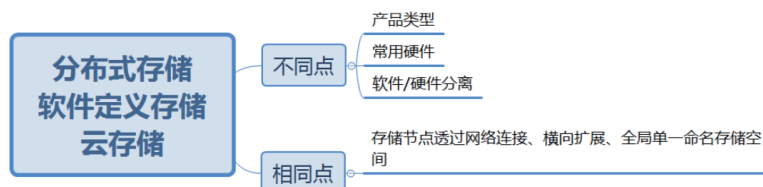
我们通过调研问卷获知了许多网友对问题的看法，从数据统计来看，85% 的用户选择了分布式存储、软件定义存储和云存储在概念上有差别，但是区别不大。

85% 受访者认为： 分布式存储、软件定义存储、云存储在概念上存在差别			
	分布式存储	软件定义存储	云存储
双层纸箱	软硬件一体	纯软件 软件 + 推荐硬件	云服务方式提供 ToC：云盘 ToB： NAS、SAN
常用硬件类似	专用硬件	x86 通用服务器	不详
软硬件分离	通常支持	支持	不支持

从市场产品的形式看，分布式存储大多以硬件加软件一体化的形式来呈现，用户可以选择配置，但是无法选择硬件产品的品牌。与之相比，软件定义存储更加强调纯软件的方式，会为用户提供支持的硬件列表；而云存储更多是以一种云服务形式来交付，其中，被消费者最为熟悉的云存储服务是所谓“云盘”，是一种 to C 性质的服务，主要提供网上的存储空间。

“云盘”在企业级应用场景中也有应用，特别是跨国企业不同国家和地区之间的数据存储和共享。以往，这些跨国企业会租用专线，透过 VPN 传输和共享数据，但是带宽的利用率不高，在性能上也存在差距。通过使用“网盘”的 CDN 加速能力，数据在节点之间自动复制和同步明显加速，如此一来，就减少了企业对广域网连接带宽的需求，实现数据本地上传，就近访问和使用；此外，网盘提供的目录功能，也便于用户组织管理和检索数据。对于跨国企业而言，“云盘”提供了类似 NAS 的服务和功能，如今，企业出海已经成为趋势和潮流，具有全球布局能力的“云盘”服务日益受到跨国企业的欢迎。

尽管在概念上各有侧重，着重点也不尽相同，百易存储研究院的观点认为三者之间的区别是不大的，特别在技术的本质上没有什么不同。对于消费者而言，不用从概念或者表述上来理解彼此之间的差异，而是应该从所解决问题的角度来理解产品。尽管表述不同，但是它们有共同点，都具有透过网络连接、可横向扩展、支持全局单一命名存储空间特征，从这个意义上说，他们所使用的技术是一致的，适用的场景也是非常接近的。一句话，分布式存储、软件定义存储以及云存储，没有一个是采用存储阵列技术来实现的，这是理解技术的最重要出发点。



一句话，分布式存储、软件定义存储以及云存储，没有一个是采用存储阵列技术来实现的，这是理解技术的最重要出发点。

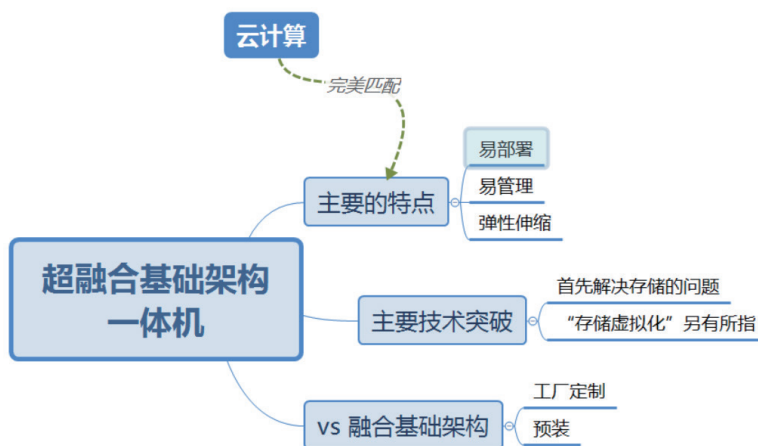
3. 融合市场现状

继软件定义存储之后，一种被称为超融合的系统快速发展，超融合的全称是“超融合基础架构（Hyper-Converged Infrastructure, HCI）”，它集计算、存储和网络于一身，在简化部署、管理的同时，还可以透过增加或者减少节点，追求计算和存储资源横向扩展，具有弹性伸缩的能力。

对于超融合基础架构而言，计算资源弹性伸缩完全可以借助服务器虚拟化技术实现，当时虚拟化技术已经非常成熟，比较难以实现的是存储资源的弹性调度和伸缩，也可以理解为存储虚拟化，但是从技术的演进发展来看，当时存储虚拟化的主要用途在于实现不同品牌存储阵列的集中管理和调度，以 IBM 为首，提供多种存储虚拟化的设备，连接不同时期、不同品牌的存储阵列产品。所以当初存储虚拟化强调的并不是存储资源池的概念，但是随着软件定义存储技术的发展，存储节点通过网络连接、横向扩展、全局单一命名存储空间技术成为了现实，也为超融合基础架构创造了条件，出现了超融合一体机的产品形式。

也因为超融合一体机主要突破的是存储资源池的技术，所以很多分析机构将超融合一体机归类到存储的范畴，成为软件定义存储的一个分支。对外提供存储资源服务的是软件定义存储，与此同时，利用存储节点自身的计算能力，对于上层应用提供支撑的就是超融合，这成为了当时一种普遍的共识。

超融合一体机易部署、易管理、弹性伸缩，与同样流行的云计算的特点完全吻合，因此当时也有观点认为超融合更像是云，并以此与软件定义存储进行区分。最初超融合一体机并不支持作为存储系统来使用，强调兼容计算，但是随着市场的发展，也开始支持纯存储的单一应用形式，这就是后话了。



最初，超融合系统的分析、比较的对象并不是软件定义存储，而是一种被称为融合基础架构（Converged Infrastructure）的系统，后者是一个工厂定制，整机柜交付的产品设备形式，通过这种方式，简化系统的部署和管理，实现系统的快速上线。融合基础架构（Converged Infrastructure）的价值在于系统集成服务，与超融合系统强调的资源池建设，特别是存储资源池的实现，还是有本质区别的。简单说，融合基础架构（Converged Infrastructure）中所集成的存储，完全有可能是传统存储阵列，不是超融合所强调的软件定义存储。这是二者的主要差别。

4. 超融合一体机还是存算分离？

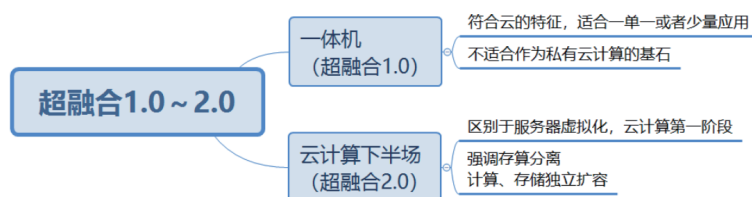
超融合基础架构最初是以一体机的形式交付的，具有易部署、易管理和弹性伸缩的云计算服务的特征，这也是超融合一体机也被称为云的原因，从单一应用的角度，通过超融合一体机来构建的应用，就相当于云的应用。超融合一体机也可以用来支持多个应用，前提是，其计算资源能够满足多业务应用的需求。

超融合一体机云化的特征完美切合中小企业技术能力有限，管理和运维能力不足的问题，超融合一体机的出现，毫无疑问降低了用户业务应用云化的门槛，在市场得到了普遍使用和欢迎。

超融合一体机可以帮助业务实现云化，但是无法从根本上替代云计算，换句话说，超融合一体机不能够作为企业私有云建设的基石。原因很简单，云计算基础设施需要集中承载企业的所有业务，动辄成百上千，超融合一体机可以承载少量的应用，如果作为企业云计算基础设施来使用，难免出现小马拉大车的局面，勉为其难。

云计算基础设施承载的业务应用更加复杂，类型也更加多样，有计算密集型和存储密集型应用的区分，以节点为单元横向扩展的超融合一体机的方式，相比就不够灵活。云计算所追求的计算、存储资源分离的形式，可以根据不同应用的需要，灵活进行扩充，计算不足扩计算，存储不足扩存储。

云计算应用最初的重点是计算资源的虚拟化，这也是为什么服务器虚拟化被称为云操作系统的原因。云计算发展到存算分离阶段，为了和已有产品方案进行区分，有厂商提出了“云计算发展下半场”的概念，并借用了超融合 2.0 的概念，赋予了超融合崭新的含义。相比于超融合 1.0 的超融合一体机，超融合 2.0 在市场上远没有得到普及和流行，很多人的认知还停留在于超融合 1.0 的阶段。



5. 从超融合到混合云？

企业私有云建设的首要问题是计算资源池化，服务器虚拟化技术满足资源动态调度的需求，这也是为什么服务器虚拟化被称为云操作系统的原因。在众多私有云开源解决方案中，OpenStack 最终脱颖而出，OpenStack 最初提供的主要功能主要集中在计算资源的调度、监控、协同和管理。

云计算下半场强调存算分离，并借用了超融合 2.0 的概念。

云计算的发展经历了从 IaaS、PaaS、SaaS 到公有云、私有云和混合云的变化，在国内，用 OpenStack 构建私有云得到用户的普遍认同，国外更加青睐使用公有云。与公有云相比，私有云更能满足数据合规的要求，但在规模和成本上存在差距，混合云就成为国内用户的目标，相比之下，国外更侧重多云的管理。

如何构建混合云？帮助用户实现业务目标？

对于混合云应用而言，以 OpenStack、vSphere、Hyper-V 为主构建的私有云，与公有云在本质上是异构系统，要实现混合应用是一件非常困难的事情。

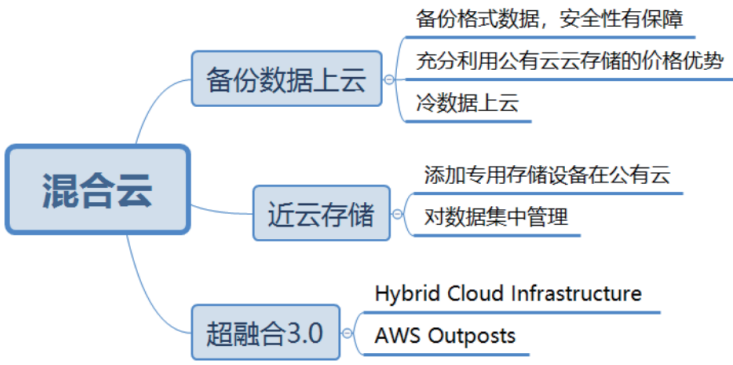
分步骤、分阶段实施混合云就成为了现实的目标。

作为混合云应用的第一步，数据上云被提了出来，最初是备份数据上云，鉴于备份数据的特殊格式，不易造成数据泄露，且可以很好利用云存储成本低的特点，类似冷数据上云就成为了最初的应用。在这个基础之上，合理利用公有云的计算资源，就可以实现业务应用的容灾。

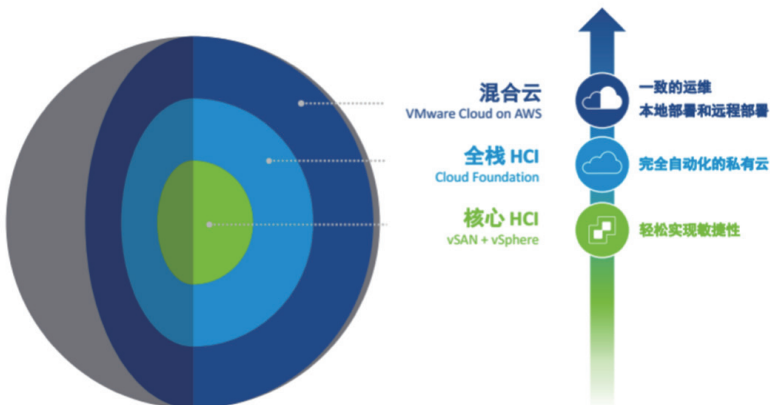
数据上云的核心是数据管理，这是容灾备份厂商的专长，也是数据存储厂商的强项。数据存储的核心，条带化、RAID 也好，本质就是数据管理的问题。在数据管理的基础上，数据存储厂商推出了类似近云存储的解决方案。

与此同时，通过与 AWS、Google 等公有云厂商合作，采用 IBM、VMware 所提供的混合云解决方案，用户可以使用所熟悉的技术，在私有云、公有云之间灵活使用资源。另外一方面，公有云厂商也面向大型行业企业用户提供了专属云解决方案，其本质也是一种同构、非异构环境的解决方案。

数据与混合云应用需求，为数据存储厂商提供新的用武之地，他们也从存储厂商上升为混合多云解决方案的供应商。为了区分，这里也借用了超融合，提出了超融合 3.0，HCI 也从 Hyper-Converged Infrastructure（超融合基础设施）升级为 Hybrid Cloud Infrastructure（混合云基础设施），实现了概念到内容的升华。



但是对于很多对技术缺乏了解的用户而言，从一体机到混合多云的跨度有点大，让他们很难跟上技术发展的步伐。



6. 纯软件还是软硬一体机？

分布式存储的核心是软件，印象中，纯软件方案具有更好的适用性，也增加了硬件选择的灵活性。但在市场实践中，用户往往会经历一个看似不合逻辑的心路历程，被称为“Crazy illogic Circle”。

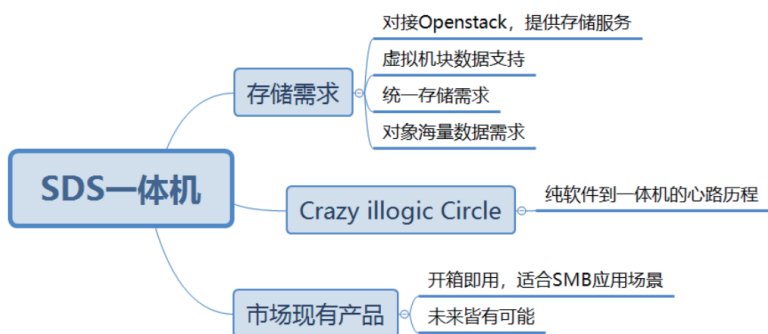


我需要一个纯软件方案，我要能够更好利用工业标准硬件，
 你们作为软件方案供应商，有哪些硬件推荐？
 你们测试过哪些硬件？
 如果硬件出现问题，联络谁？
 你们可以提供硬件并提供支持吗？
 你们可以提供一体化软硬件产品吗？

Crazy illogic Circle 让用户从纯软件的一个需求，成功蜕变为软硬件一体化的需求。这里需要关注一个概念：SDS 一体机。

什么是 SDS 一体机？与分布式存储、软件定义存储、云存储有什么区别和联系？

还是坚持百易存储研究院的一个原则：不拘泥于定义和概念，而是从实际上应用需求出发，SDS 一体机所能够解决的具体问题。



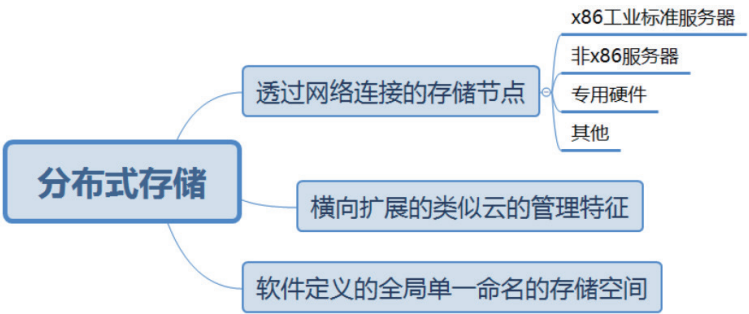
从应用角度，可以划分为需求如下：为 OpenStack 云提供后端存储、为虚拟化提供所需的块数据存储服务、面向传统架构的统一存储、以及面向海量数据的对象存储。

目前 SDS 一体机强调的是开箱即用，能够提供 iSCSI、FC、NFS、CIFS、FTP 等全协议支持的支持，显然这比较适合中小企业使用。未来会不会有其他形式的 SDS 一体机出现，应该说一切皆有可能。

7.x86 还是专用设备？

这个问题产生还是来自于分布式存储没有普遍接受的定义，在实践中就出现了认知偏差，类似盲人摸象，每个人都从各自角度阐述对问题的理解。

实际上，x86 还是专用设备在市场上是并存的，此外还有其他的形式。如果从利用的角度，传统存储阵列是不是也可以作为节点存在呢？所以，问题的分析并不是 x86 vs 专用设备，不同的方案特点不一样，需要用户根据自己的需要灵活加以选择。



性能上的高低上下，可以透过 PoC 测试来解决。可靠性和运维的问题，需要专业人员结合实践给出专业的判断。

对于用户而言，重要的不是站队，重要的是视野。

8. 分布式存储是否适用于“关键业务”应用？

分布式存储是否可以应用在关键业务应用的场景？市场上之所以有这样的疑问，主要的原因在于缺乏分布式存储的实践。与集中式存储阵列相比，分布式存储还属于新生事物，没有更多的经验作为参考。

对于新生事物的应用，有一个逐步探索和实践的过程，用户不会贸然从关键业务入手，往往会选择边缘业务，不断积累实践应用的经验，这是一个审慎稳妥的做法。

关键业务应用对于可靠性、稳定性有着极致的需求，强调业务应用的稳定和连续，所以能够用于关键业务应用的产品，在可靠性、稳定性方面采取了非常多的措施，以存储阵列为例，从前端网络连接、控制器、到后端的磁盘连接，全部采用冗余设计，确保系统没有任何单一故障点，这也是存储阵列价格昂贵的主要原因，但是这种极致的可靠性、稳定性能够确保关键业务应用的需求。

与之相比，分布式存储阵列也采取了非常多措施来确保系统的可靠性、稳定性，确存储数据不丢失，利用 x86 通用服务器的成本优势，分布式存储采用多副本、纠删码的技术确保数据的稳定、可靠性和安全，多副本数据被分散到不同的节点、机柜、机架，确保不会因为单一硬件故障，造成数据的丢失和不可逆。但出现单一节点故障之后，分布式存储系统会调用所有的节点，参与数据的恢复，可以确保系统始终处于多副本或者纠删码的保护之下。

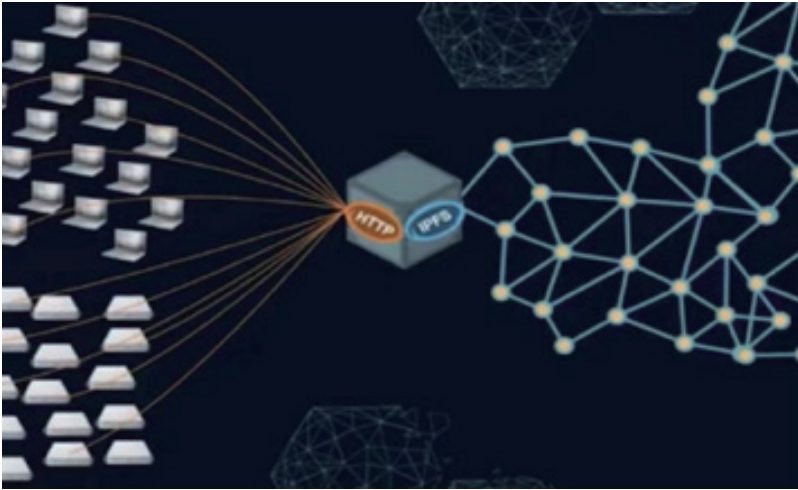
如今，x86 通用服务器等节点的性能和质量不断提升，已经具备运行关键业务应用的能力，在很多计算应用的场景已经得到了证实，这也为分布式存储奠定了稳定、坚实的基础。此外，分布式存储横向扩展的能力更加有助于用户应对海量数据存储带来的挑战。

如今，云原生成为了应用软件发展的主要趋势，透过容器 + 微服务化设计，云原生应用在适应容器的调度、故障隔离的能力极大增强，也更加能够适应迭代更新的需要，云原生应用的发展也降低了对于单一硬件可靠性的依赖，这种发展趋势，也有助于分布式存储的推广和应用。

9.IPFS 分布式存储与去中心化？

IPFS（Inter Planetary File System，星际文件系统）存储的出现赋予了分布式存储新的内涵，同时也加剧了市场上分布式存储概念上的混乱。

IPFS 是一个分布式 web，点对点超媒体协议。IPFS 协议的目标是取代现有的互联网协议 HTTP，IPFS 独有的存储和云计算技术，另辟蹊径，也为企业级业务应用提供了新的选择。



IPFS 存储是对市场现有存储产品和方案的颠覆性重建，它将 IPFS 存储协议与区块链无缝集成。使用时，IPFS 存储将需要存储的数据文件分片，然后分片存储在各个节点上，如此，数据就被复制了 N+1 份并存放在不同的地区，如果某个地区的数据因人或自然灾害被毁，通过其他地区的数据备份照样能完整恢复所有数据。

区块链技术和 IPFS 存储的完美结合，无论是加密，还是防篡改都有助于保护用户数据的隐私和数据安全性。目前 IPFS 仍处于初步发展阶段，更多的应用还是内置经济激励项目 Filecoin 挖矿；与此同时，Filecoin 也提供了海量存储资源，在 \$/GB 成本上占据优势，对于现有各种存储应用形成了降维打击。

对于现有行业 / 企业用户而言，IPFS 存储或称 Filecoin 无疑具有巨大诱惑，如果它们能够提供 SAN、NAS 或者对象的支持，就会对现有的方案带来冲击。现有行业企业应用如何使用 IPFS 存储，还存在一定的技术障碍，需要一定的技术突破。但是无论如何，IPFS 存储 \$/GB 的巨大优势，充满了无限遐想的空间。

IPFS 存储带来的最大问题是其分布式存储的称谓，容易与现有产品混淆，带来概念上的混乱。IPFS 存储与本报告所说的分布式存储，核心的技术都是去中心化，所不同的是 IPFS 存储主要去的是 HTTP 网络访问的集中化，分布式存储所去的是存储控制器的集中化，二者还是有所区别。对此，应该有一个清晰的界定。

从这个思路出发，百易存储研究院建议：用去中心化存储、或者 Web 3.0 存储来代指 IPFS 存储，而将分布式存储称谓保留给现有存储。



10. 对象存储的话题

对象存储，也叫做基于对象的存储，就像文件一样，对象包含数据，但是和文件不同的是，对象在一个层结构中不会再有层级结构。每个对象都在一个被称作存储池的扁平地址空间的同一级别里，一个对象不会属于另一个对象的下一级。

以上是百度百科给出的定义，相对还是比较费解的。

简单说，对象存储在互联网领域应用比较广泛，如 S3、Swift，主要用于存储海量非结构化数据。传统行业 / 企业应用场景大多使用 SAN、NAS，其中 SAN 用于存储结构化数据，NAS 用于存储各种文件数据，它会按照目录结构的方式组织数据，便于文件的访问和查找，类似个人电脑 C 盘的目录和文件夹应用，但企业级 NAS 目录更加复杂，在一些工业应用场景，这种目录结构嵌套会有 5 ~ 6 层，100 多个层级。当目录结构变大之后，就需要经过多次点击才能够检索到所需要的数据文件，检索速度变差，效率降低。

与 NAS 数据组织方式不同，对象存储的采用了扁平的数据组织结构，文件（切片）被直接存储在根目录的“桶”中，相当于将文件全部存储在 C 盘根目录下，这从根本上规避了维护庞大的文件目录树。

抛弃了目录树的方式，结构变得简单。但是文件如何检索呢？为了便于检索和查找，在对象存储的使用过程中，就需要对每个对象（文件切片）进行标注，生成元数据，每个元数据都对应、匹配一个 Index 或者 Key，按照 Key-Value 的方式存储和检索数据。

对象存储中的数据，透过 REST API 接口，可通过 HTTP 直接访问和使用。互联网服务中的网盘应用，大多是采用对象存储的方式，对于相同的数据文件，直接返回 Index 或者 Key，并不需要重复存储数据，这也是所谓“秒传”的功能。在对象存储的应用过程中，常常会使用到 Key-Value 数据库、Elasticsearch 数据库，主要作用就是数据文件的查找和检索。

当存储目录架构变简单之后，数据检索的复杂性就会增加，需要使用到 Key-Value、Elasticsearch 全文检索的能力。

对象存储很容易实现横向扩展，最多可以支持 4096 个节点，轻松实现 PB 级海量数据存储。目前已经投入使用的对象存储，已经可以达到 960 节点规模。

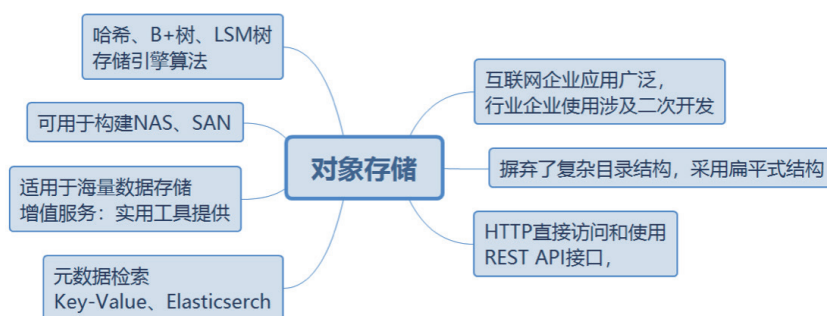
如今，存储面临的主要挑战来自数据量的快速发展，互联网技术应用缩短了企业和消费者的距离，消费者在网上的一举一动，都可以成为企业了解用户、了解市场的依据，产品技术创新和应用，完全基于数据分析，建立在数据分析的基础之上。由此也带来了海量数据存储的需求，并且以非结构化数据为主。这个存储需求与对象存储的特征高度匹配，这也是为什么对象存储在互联网率先得到应用的原因。

传统企业正在云化的过程中，数字化转型、互联网成为企业信息化强有力的工具，必然也会带来对象存储的需求，这个将是大势所趋。与此同时，传统企业有自己熟悉的行为和规范，习惯使用 NAS 存储的方式，要让他们从 NAS 转型到对象存储，会涉及到应用的少量二次开发，这会成为传统企业应用对象存储的瓶颈。

如今，对象存储的主流供应商，在功能上，除了提供最基本的对象存储支持之外，也结合实际应用提供了行之有效的工具，如图形格式转换、人脸识别、视频格式转换、OCR 等，这些增值功能和服务，受到了用户的欢迎，部分拉低了对象存储的使用门槛。

在这里，百易存储研究院特别提醒用户，对象存储作为一种基础的数据组织和存储方式，是一种最底层的基础设施，它与 NAS、SAN 存储之间的关系，并不是一种竞争的关系。

对象存储、NAS、SAN 不是排他的关系。透过构建目录服务，NAS 完成可以构建在对象存储的基础上。在一些典型的应用场景中，如电子邮件，完全有可能是对象存储和 NAS 存储、SAN 存储的混合体，这个是应该引起高度关注的。



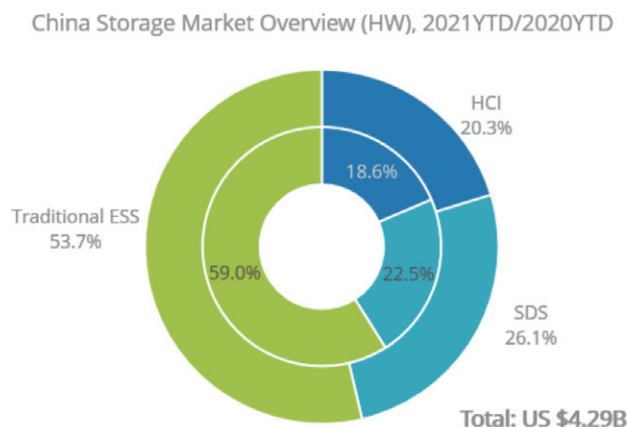
配合本次分布式存储的调研报告，调查问卷有一项关于存储引擎的内容，列举了哈希存储引擎、B+ 树存储引擎、LSM 树存储引擎等三种形式。百易存储研究院的分析师虽然对存储领域有一定的研究，但还有很多的细节搞不清楚，设置这样的内容，我们是希望提醒用户，B+ 树、LSM 树在数据库技术中经常被提及的技术，如今出现在存储引擎中，说明这都是一种底层的结构，影响的都是数据检索的效率，会在一定程度上，决定存储系统的性能表现。

对于性能，最好的做法是 PoC 测试，最能够反映实际应用的性能表现。但是与此同时，既然存在这样的技术发展和流派，也可以增加视野，帮助进行理解和分类。

三 . 分布式存储市场状况

1. 分布式存储市场规模

中国分布式存储市场规模有多大？这是一个很难准确回答的问题，一来分布式存储市场参与的厂商很多，其中很多厂商不是上市企业，没有办法获得销售数据，二来分布式存储没有一个市场可以普遍接纳的定义和分类，会导致数据统计口径混乱，例如



分布式存储是否包括去中心化存储、是否包括超融合、是否包括云存储、内涵与界定的不同，市场规模自然也不一样。

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%。

本报告中的分布式存储包括 SDS、HCI 两个部分，总体市场规模 26 亿美元，占比 46.3%。配合本次调查报告的调查问卷中，受访者也对市场进行了估算，结果从 10 亿美元～ 1000 亿人民币不等；超融合市场规模从 30 亿美元～ 300 亿人民币不等；云存储市场的规模 50 亿（美元）～ 1800 亿（人民币）规模不等。差距和出入非常大，体现了不同消费者的不同理解。

目前百易存储研究院没有官方的渠道和数据对存储市场规模进行统计，但是可以肯定，随着数据量的快速增长，数据存储市场会呈现一种快速增长的趋势，IDC 给出的预计，市场增长速率为 54.2%，未来发展不会低于这个增速。

在目前的情况下，分布式市场的规模也是仁者见仁、智者见智，但是无论如何，市场发展的趋势是明确和清晰的，未来市场

潜力巨大。

2. 分布式存储市场英雄榜

对于市场而言，2022 分布式存储企业图鉴的意义在于用户可以按图索骥，查找自己需要的产品解决方案线索。



2022 分布式存储企业图鉴中的企业可以分为几个类型：

1. 综合品牌产品供应商

以戴尔、浪潮、新华三、华为、联想凌拓、IBM、Red Hat、曙光等厂商为代表。这类厂商具有非常强的综合实力以及品牌影响力。它们不仅仅提供分布式存储，也能够提供计算、网络和软件服务，各个产品之间彼此呼应，市场相互渗透，拥有非常强大的市场服务能力，综合实力保持领先。

2. 专业存储产品供应商

一寸长一寸强，术业有专攻。作为存储专业领域的垂直供应商，这类厂商在产品技术上保持领先，专业技术服务能力非常强，在很多时候，专业存储厂商也为品牌厂商提供存储的 OEM 服务。专业存储产品供应商以国外的 EMC（被戴尔收购）、HDS（更名为 Hitachi Vantara）、NetApp 为代表，国内以同有、宏杉为代表，这些厂商至少有 20 年以上的发展历史，以存储阵列见长，为了能够兼顾到分布式存储的市场，国际厂商多采用并购的方式，国内厂商主要依靠自研的力量。

3. 创新存储厂商

数据作为行业企业最为重要的资产，用户对于存储可靠性的要求往往会超过计算，不容有任何闪失，也因为如此，创新品牌不易被用户市场接受。但是分布式存储技术的出现，改变了这样一种局面，特别是开源技术，也在一定程度上拉低了存储产品的门槛。从 ServerSAN 开始，创新存储市场极为活跃。

在资本市场的鼎力支撑下，创新存储市场如雨后春笋，快速发展。国内 XSKY、Smart X、杉岩数据、焱融科技、霄云、鹏云、青云、华瑞指数云、瑞驰、云信达、柏科、国鑫、大道云行等表现得比较活跃；国外市场以 Pure Storage、DDN、Cloudian、Scality、Swiftstack、ZADARA、Caringo 为代表，它们当中有些也已经进入中国市场。

4. 云存储服务供应商

云计算改变了传统产业的现状，从购买产品转移到购买服务，屏蔽了底层负载的技术和管理。

受云计算服务的影响，存储也正在经历从产品到服务的转型，这也是云存储的由来。云存储最初主要面向 ToC 的市场，其中以云盘作为常见，这也是至今仍然有很多人将云存储称为云盘的原因。

但是随着市场的发展，以阿里、腾讯、百度、网易、华为云、青云、华云、金山云、京东云为代表，它们纷纷推出了面向行业的存储定制服务，提供了“专属云”的产品和服务，如此就把云存储的市场从 ToC 向 ToB 进行延伸。

云厂商提供的云存储方案和以产品为核心的解决方案有所不同，这些方案更好满足了创新企业的需求，但是与传统产业的结合，还需要磨合。

四． 分布式存储行业典型的应用场景

在这个章节，我们特别征集了分布式存储在一些典型行业应用场景中的应用案例。透过这些案例，我们希望能够帮助大家深入了解分布式存储的技术，打消类似“分布式存储是否适合关键应用”的疑虑，举一反三，扩大分布式存储技术的应用。

电信行业

典型应用场景：SIM 卡认证和私有云计费业务

应用案例：

在移动时代，手机号就如同人们的“第二身份证号”，有了它我们才能在现代社会里“吃喝玩乐、出行办事”。只有当手机号经过电信运营商的认证，与我们个人身份相匹配，才能接入网正常通讯，享受到运营商和服务商提供的各种服务。

中国移动为全国 9.61 亿户手机用户提供基本通讯和各类数据服务。同时，移动云平台也为大量企业用户提供了类型广泛、规模庞大的云服务，为了更好地为广泛移动手机用户提供服务，同时优化自身云平台运营，中国移动计划对其块存储资源池进行更新和升级。

中国移动块存储平台承载着的大量业务应用：网络云应用场景下的 SIM 卡认证、网络认证、虚拟化网元管理平台；IT 资源池下的私有云计费业务，如内容计费、NGBOSS、BBOSS、网状网、计费账务中心；以及集团和各省的自有业务：移动商城、一级客服系统、安全监测平台、国际漫游计费业务；这些服务都是移动的核心关键应用，一旦出错，不仅给移动的日常运营造成难题，同时，也会严重影响每个人的使用体验。

这些业务类型都属于数据库应用，底层通常采用块存储模式为上层应用提供数据服务。中国移动原有的存储系统存在成本高、性能差、可扩展性差、容量不足等问题，在移动基础架构全面云化的大背景下，存储系统的可管理性、可集成性较差，无法满足云时代中国移动对海量存储的需求。

为此，中国移动需要打造一个高性能、高可靠、云化、敏捷、智能的块存储数据承载平台。

中国移动启动了 2021-2022 年分布式块存储新建部分集采，采购规模总数为 2100 套，其中，紫光股份旗下新华三集团 H3C UniStor X10516 G3 分布式块存储产品以优异测试成绩中标，获得 20% 份额，中标总金额达 1.15 亿元。

中国移动本次采购的分布式块存储系统，将会承载大量核心关键的移动业务数据，既包括热数据，也包括海量冷数据，为中国移动提供个人通讯服务、以及企业级云服务提供计费、认证、运维管理等关键运营活动提供底层数据平台；同时，还需要与

Openstack 云平台进行对接，与 Hadoop 生态做全面结合，以服务中国移动“上云”进程。

作为承载关键业务运营数据的块存储平台，中国移动对存储的首要要求是可靠性和性能，这是业务安全保障的前提；其次是存储系统的可管理性和可集成性，移动业务复杂多样，可管理性和可集成性会严重影响存储的部署上线及后续的运维工作；最后，产品能够对接多种版本 Openstack，能够和 Hadoop 生态进行协作。

为了全面考量参与竞标的存储系统，中国移动围绕性能、可靠性和管理功能等关键指标设置了共计 39 个测试用例，分别包括 OLTP 模型读写性能和 OLAP 模型读写性能两项关键性能测试用例以及卷功能、快照功能测试、QoS 功能等多项可靠性相关测试；存储池管理、管理维护测试等管理功能测试；以及产品常规检测、功耗测试等常规指标测试，对存储系统各个方面进行全面和严苛的测试评估。

新华三 UniStor X10000 自研分布式存储交出了一份亮眼的成绩单：在 OLTP 模型读写性能测试和 OLAP 模型读写性能测试两项重要的性能测试中，X10516 G3 分布式块存储系统解决方案分别以 18000ops 和 25200ops 的优异性能获得满分成绩，同时功能测试 24 项全部通过，在可靠性测试、管理维护测试、功耗测试等各个方面的测试用例中，均取得了优秀的测试成绩，最终赢得



了客户认可。

科研行业

典型应用场景：脑科学研究 RUSH 数据存储

应用案例：

脑神经科学的一个中心目标是揭示神经环路图谱，目前所有针对神经环路的成像手段都有很大局限性，显微技术受数据通量制约，无法获得观测对象的活体全脑高分辨率动态成像数据。清华大学戴琼海院士团队成功研发出新型超宽视场高分辨率实时显微成像仪器（RUSH）突破了低数据通量瓶颈，实现 1 厘米 × 1.2 厘米超宽视场下 800 纳米分辨率的实时动态成像，实现了兼顾“全

局形态”和“细节特征”的多尺度观测。

RUSH 前端有 28 台高速连拍相机，每秒连续拍摄 30 张（24MB/ 张），即 28 台相机每秒约有 20.16GB 数据的采集，单次实验长达 3 天，要求数据平台满足高速相机 72 小时连续采集需求。

RUSH 高通量采集业务，多维、多尺度、高分辨率持续拍摄，记录医疗实验的观察结果，核心数据为医疗研究使用的高清影像；存储确保 28 台高速相机所采集的数据可以及时、稳定、可靠地存储和访问

浪潮通过 40 个分布式存储 AS13000 节点搭建一个存储集群，配置纠删 6+2，提供可用容量超过 5PB；

恒定码流，单路采集数据流维稳，开启特定缓存模式，实现 28 路 750MB 码流同时写入，集群带宽不低于 20.16GB/s；持续写入 72 小时模拟测试工具不丢帧；

海量文件高速删除，周期性多对象并行删除，模块化松耦合设计，保障数据删除时恒定 IO 下降不超过 30%；

部署智能管理软件 InView，对数据中心的设备进行管理，通过智能算法对实验平台的数据基础设施进行性能、容量、故障进行预测，确保业务不中断。

高性能稳定的数据平台为实验的数据采集提供了保障，恒定数据流写入性能提升 40%；通过对对象聚合，数据对象聚合后落盘，提升单次落盘的对象大小，降低落盘操作次数，聚合带宽可达 45GB/s 以上，确保了图像收集和存储顺利进行；

海量文件删除优化，批量删除过期数据的同时减少数据删除对上层业务的影响，保障数据删除时恒定 IO 下降不超过 30%，



为客户实现快速作业轮转。

汽车制造

典型场景：汽车制造大数据平台

应用案例：

威马汽车引入 Dell EMC 基于 PowerScale 非结构化数据解决方案构建大数据平台，数据涵盖用户的全生命周期。从用户车辆交付开始，威马汽车将新零售进行延伸，全面激活智慧出行生态圈，通过对用户及产品数据的搜集与分析，以数据驱动智能硬件的发展。

基于 PowerScale，不断生长的大数据将直接提高产品的开发、制造与服务的效率，为产品的开发提供理性的数据支撑，精准计算排产计划，保障供应，促进 C2M 客制化生产效率持续提升。结合用户生活区域与生活方式的划分，威马汽车在充电桩网络、新 4S 网络与智行合伙人布局、智能互联生活、乃至出行合作伙伴的协同上，驱动体系性的优化，开创出诸如智能移动空间等与用



户共创共赢的全新服务模式。

Dell 的 PowerScale 常年占据 Gartner 魔力象限领导者第一的位置，无论是解决方案完整性还是方案成熟度都是业内首屈一指的。其可以助力企业构建分层数据湖，满足企业海量非结构化数据就地存储、分析和利用，全面释放企业数据价值。

PowerScale 的分层数据湖解决方案具备以下几大特点：

一数多用：同一份文件的存储、查询和利用可以在同一个地方进行。避免数据多个地方拷贝，保证数据完整性和一致性，同时因为使用同一份数据，因此也可以避免拷贝数据带来的时延，从而保证了分析和使用的时效性。PowerScale 支持 8 种访问协议，满足不同服务器、客户机、数据分析平台、云原生应用的访问。

自由伸缩：一方面，数据湖需要能够根据工作负载的变化而动态伸缩性能和容量，另一方面，大数据分析平台需要存算分离，以利于独立调整算力和空间的需求。

跨代混搭：当非结构化数据达到一定的规模，升级换代、动态按需扩展就成了非常关键而重要的挑战。若干 PB 的数据迁移到新设备，需要严格的数据迁移的方案和计划，需要合适的停机窗口。PowerScale 完美解决这一挑战。不同代、不同类型的节点可以混搭成同一集群。数据根据性能需求存放在不同速度的节点。新旧节点也可以并存，从而把升级换代透明化，完美解决相对应的数据迁移的挑战。

文件防勒索：越来越多的企业意识到除了防病毒、防非法访问外，新的安全挑战防勒索也变得至关重要。PowerScale 提供了三种级别（好、更好、最好）的文件防勒索解决方案。

有效容量：虽然非结构化数据包含许多的影视音频等不可压缩的文件，但因为可能存在多副本，所以数据重删是存在可能的。在许多用例中，非结构化数据的数据消重还是有显著效果的。PowerScale 支持多种数据消重能力，能够大力提升存储的有效容量。在不少的用例上都能给客户带来成本的节约，提升投资回报。

数据生命周期管理：PowerScale 分层数据湖平台能够根据数据的生命周期对性能、容量的需求自动配置最合适的存储空间，以匹配数据的价值体现。

医疗行业

医疗典型场景 1：新院区开业，无缝迁移系统

应用案例

2019 年，VMware 与复旦大学附属肿瘤医院合作，基于超融合的私有云平台，助力医院构建现代化数据中心。复旦大学附属肿瘤医院是国内知名的全能三级甲等肿瘤专科医院，其信息化建设处于国内领先地位。2019 年复旦大学附属肿瘤医院浦东新院区开业，为保证信息化建设水平一致，新院区需完全按照总院的统一体系进行规划运行。在建设过程中，医院主要面临系统无缝迁移，

时间紧迫和场地有限的问题。

为帮助医院新院区顺利开业，VMware 使用 vSphere、vSAN、NSX 等多项行业领先产品及服务，用了不到三个月的时间为院方架构出一套高效的桌面化系统，确保了医院准时开业以及顺利运营。同时，VMware vSAN 分布式存储技术帮助新老院区业务应用无缝打通，确保医院各应用系统的高可用性。VMware 制定两地多中心应用级容灾方案，帮助院方形成全面的防护体系，提升了数据中心的安全性。在 VMware 的助力下，复旦大学附属肿瘤医院高效解决其痛点，获得显著收益。

VMware VSAN 架构技术具有高性能、可伸缩性、高兼容性、分布性等优良特性。由 vSAN 提供技术支持的集成式超融合基础架构 (HCI)，与 vSphere 相结合，使客户可以借助单个集成平台管理计算和存储。借助集成解决方案，提高业务敏捷性，加快运维速度并降低成本，为实现面向未来的 HCI 和混合云提供更简单的途径。

医疗典型场景 2：医疗 PACS 影像存储

应用案例：

医院的医疗 PACS 数据量预计每年增长 15%，五年翻一番，并呈加速增长的态势，三甲综合性医院或胸科、肺科、骨科等三甲专科医院，一年的新增拍片量在 50TB ~ 60TB。根据国家电子病历保存相关规定，医疗机构保管保存医疗影像数据时间要求不少于 15 年。

无论从业务发展还是合规要求，医疗影像数据的存储都对存储系统提出了需求。

目前医院对医疗 PACS 数据普遍采用传统阵列存储 FC SAN 或 NAS，采用在线、近线、离线的三级存储架构。PACS 影像的典型特征是大部分文件都是小文件，其中 MR 文件平均大小为 60KB 左右，CT 文件平均大小为 300KB 左右，也是小文件。长期以来，小文件存储都是存储系统面临的挑战：小文件读写性能低；当存储的文件数量增多时性能会不断下降。PACS 图像调阅的速度最快仅为每秒 80 幅左右。典型的 MR 检查，平均每次检查产生约 3000~5000 张小图片，调阅图片需要数十秒以上。

霄云碧海分布式存储相对于传统阵列存储，在大规模在线扩展、百亿级小文件管理、高性能数据吞吐、敏捷化运维管理、TCO 成本优化等方面具有明显的技术优势，可以有效解决医疗 PACS 目前在存储上遇到的性能、数据孤岛、运维困难、成本较高等问题，在医联体或大型医疗机构 PACS 系统这样超大数据量且高并发调取、运维管理要求敏捷化的领域，成为存储架构的最佳选择。

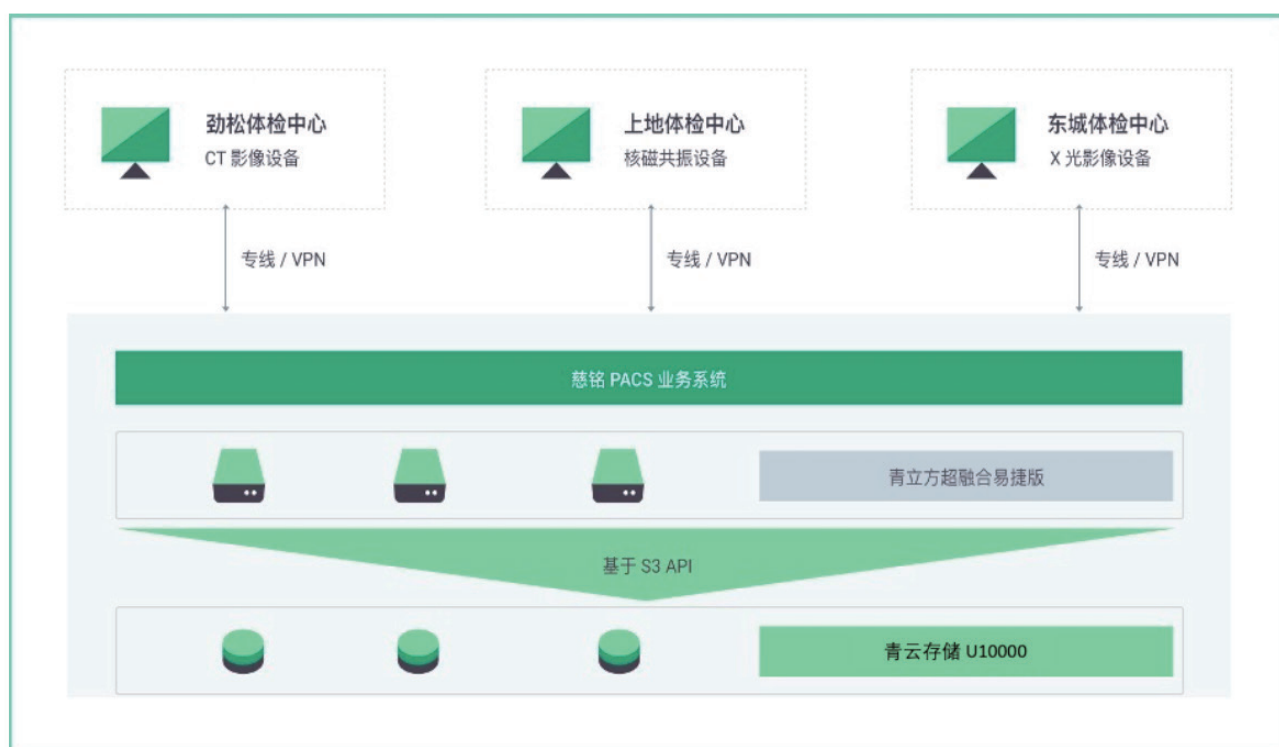
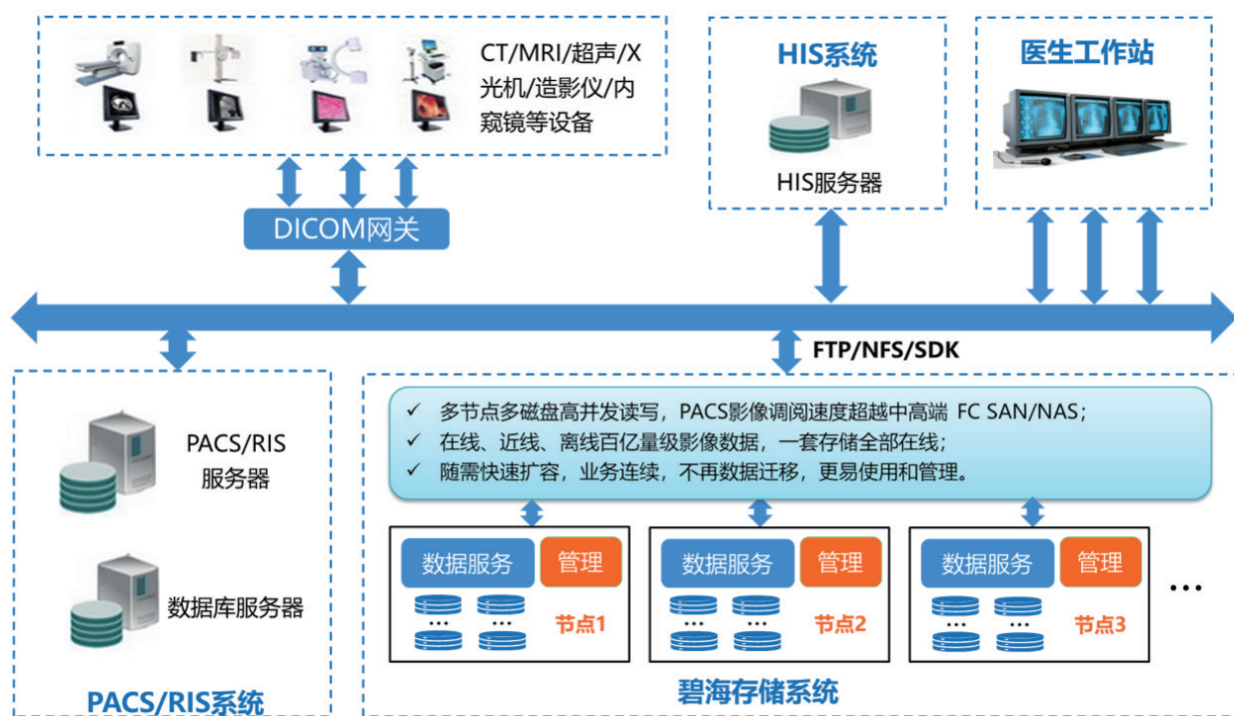
在上海市胸科医院大力支持下，碧海分布式存储经过长达一年的验证和试运行后，成功上线。通过对比医院现有中高端阵列存储的影像调阅速度，验证了碧海分布式存储在 PACS 影像调阅速度上有了 2 倍以上的提升。业务高峰时 500 多台工作站并发调阅，调阅性能依然相当稳定，显著提升患者就医体验，提高医生诊断效率，一直困扰医院的 PACS 影像存储性能瓶颈由此迎刃而解。

医疗典型场景 3：统一存储，集中管理分院数据

应用案例：

作为国内专业健康体检机构，慈铭健康进入了快速发展时期，现有的 IT 业务模式已无法适应集团业务发展的需求。随着业务拓展和体检中心分院的快速扩张，每天有大量 PACS 影像数据产生，存储和应用这些数据成为制约其发展的难题。加并且，各分院影像系统均为独立建设，数据本地存储，且采用传统离线阅片方式，各中心数据共享和应用融合无法实现，信息孤岛的存在直接导致阅片效率极低。加之，体检数据涉及用户隐私，如果存储在公有云必须脱敏，慈铭倾向于将其迁回到私有化环境。

青云存储 U10000 一体机以业界领先的分布式技术为核心，为慈铭提供医疗影像存储平台。U10000 存储系统通过 S3 API 对接 PACS 影像业务系统，各个体检中心通过专线将前端服务器临时存储的 PACS 数据上传至总部数据中心，消除了“信息孤岛”，实现信息资源共享数据，以满足各分院对数据服务及时访问、及时调阅、及时处理的远程服务需求。U10000 的分布式架构可以实现敏捷扩容，能够满足不断增长的数据存储需求，私有化部署完全解决行业合规，多重安全机制保障数据安全合规；同时，在长期成本和运维成本上，对象存储私有化部署相比公有云方案拥有更低成本。慈铭 PACS 系统通过 AWS S3 接口对接 U10000，实现 DICOM 影像高效安全的存储调阅，实现集团内各分院 CT 检查患者的统一管理、统一存储和统一服务。并且 U10000 具有对象、文件和大数据多种接口，可以支持慈铭体检未来 AI 会诊、健康分析等多种应用的顺利使用，并且可以将长期储存的数据压缩至成



本更为低廉的蓝光存储，实现存储成本最优。

慈铭体检集团借助青云存储 U10000 一体机解决方案构建的影像云中心，让众多分院能够借助集团集中、优质的资源，提升医疗影像远程诊断、审核、会诊服务的能力。通过云端阅片，不仅提高了阅片的工作效率，还实现了集团内各分院 CT 检查患者的统一管理、统一存储和统一服务，为未来集团智慧医疗的建设奠定了基础。

基因测序

典型场景：生物医学大数据应用

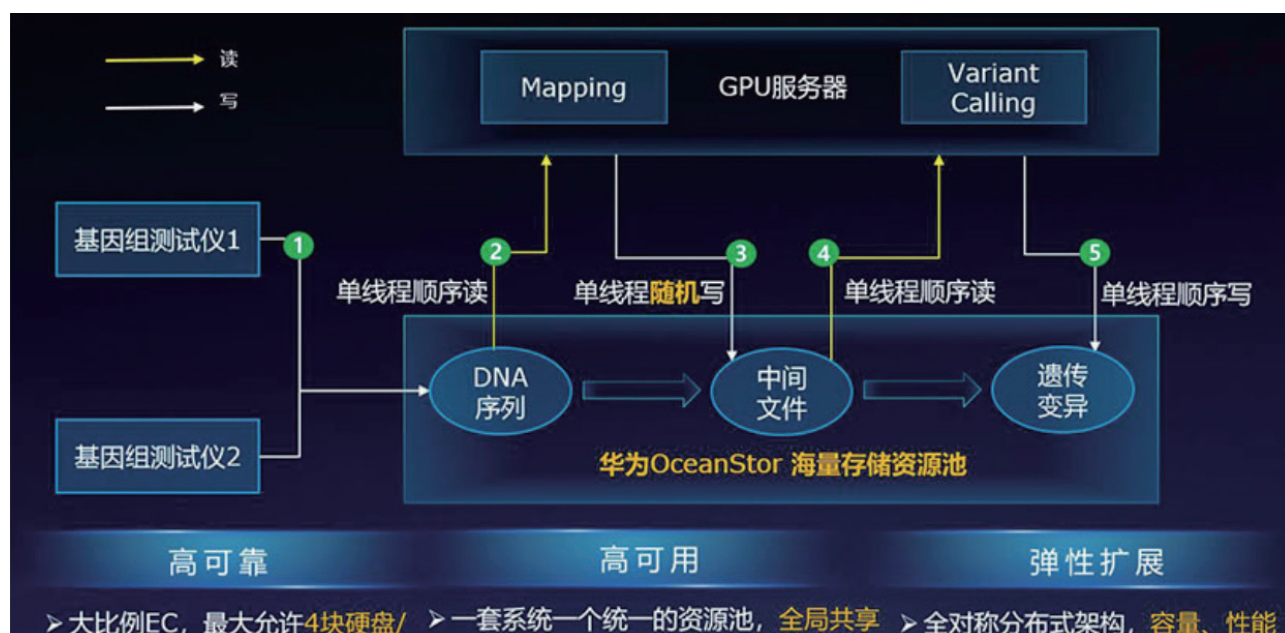
应用案例：

多组学数据分析是精准医疗与医疗大数据的基础。基因测序从血液或唾液中分析测定基因全序列，分为三个环节：提取、分析和解读。其中分析阶段包含文件格式转换、解压缩、基因拼接、比对、排序、去重、变异检测以及联合基因型检测，该阶段取决于生信分析系统的性能，是基因测序 HPC 解决方案的主要发力点。

基因测序产生的数据量都是 TB 级别，例如一台华大智造 MGI 的 DNBSEQ-T7 测序仪生产量：4.5Tb/24h，6Tb/30h。满负荷下，一年能产生 1.7PB 左右的数据量，加之生物信息分析过程一般会产生原始数据量 5 倍左右的中间文件及结果，因此要支撑一台 DNBSEQ-T7 一年的数据产出存储及分析，大约需要 8.5PB 有效存储空间。另外如何低成本地实现基因数据长时间存储，在线、离线、归档数据自动管理也是客观需求。

根据研究目的不同，不同的专用软件，不同计算实例会有 I/O 密集型、CPU 密集型、内存密集型等需求，因此海量基因数据的分析与挖掘需要流式处理，需要高性能的 GPU、CPU 异构计算集群才能满足需求；而且在后期基因序列分析时，往往会进行序列对比，这就要求要将海量数据一次性的载入到内存并加以处理，而不适宜分批导入，这就对大内存有了非常严苛的容量要求。

海量数据的传输上集群本身会对网络和带宽造成极大压力，另外计算过程中也会涉及数据的高速共享与读写检索，对存储和设备 IO 带宽要求高，至少需要存储系统单线程具备 6GB/S 以上的吞吐能力，且对实时性要求非常高，否则可能导致数据不完整。



2021 年 9 月，四川大学华西医院、华为技术有限公司、赛乐基因科技（北京）有限公司三方联合发布多组学数据加速分析平台，采用华为公司在高性能数据存储和先进基因数据管理系统上的技术积累，以及赛乐基因在 GPU 加速生物信息数据分析领域的行业经验。

在基因序列分析时，往往会进行序列对比，这就要求将海量数据一次性的载入到内存并加以处理，对存储单流带宽有了非常严苛的要求。华为提供更高性能全分布式存储底座。相比现有存储，全新存储平台将单流读带宽提升了 2 倍，单流写带宽提升了 4 倍，

4 节点即可提供读 30GB/s、写 25GB/s 的聚合带宽能力，高效支撑了多组学联合创新平台的性能提速。

高性能计算

典型场景：并行文件系统应用

应用案例：

在高性能计算场景中，并行文件系统能够很好地支持并行计算的大规模数据读写，其性能决定数据访问进程的运行效率和速度。要充分释放高性能计算的算力潜能，需要并行文件系统具备高效、稳定、可靠、海量等特性。宏杉科技的 MOFS 并行文件系统解决方案，为高性能计算应用提供全新的支撑。

MOFS 是宏杉科技针对海量数据存储应用推出的大规模通用并行文件集群系统，具备高性能、高可靠、高可用和高 IO 吞吐能力。可为用户的业务应用提供全局统一的系统映像和完全 POSIX 兼容的数据访问接口，从而为高性能计算集群提供数以万计的客户端系统支持、PB 级存储空间和每秒数百 GB 的吞吐量等动力。

宏杉科技 MOFS 并行文件系统设计研发的高性能集群并行存储技术，通过数据的前端切片和并行传输，实现多线程并发读写，提升数据并行处理效率。在全局统一的系统映像下，通过统一的接口实现数据的互通访问，提高数据分析效率。同时，MOFS 并行文件系统内嵌宏杉科技独有的负载均衡技术，将数据读写压力分布到服务器的各个硬盘上，形成一个高性能的共享存储池，为前端应用提供高带宽读写访问服务，实现每秒数百 GB 的吞吐量。目前，在宏杉科技的众多高性能计算实践案例中，MOFS 并行文件系统可提供达到 300GB/s 以上的聚合带宽，其中 8PB 存储系统甚至可以提供 150GB/s 的聚合带宽，为用户提供匹配算力的极致带宽性能。

宏杉科技 MOFS 并行文件系统提供完全 POSIX 语义兼容的文件访问接口客户端程序，为用户的业务应用提供丰富的数据访问接口支持。用户可如同使用本地硬盘一般，在 Windows、Linux 和 Mac OS 操作系统上运行业务应用，并使用 MOFS 进行 PB 级数据存储。此外，MOFS 并行文件系统完全兼容现有标准应用，无需二次开发，简便易用。

硬件支持方面，宏杉 MOFS 并行文件系统解决方案中还支持 x86 和 AMD 等多元化架构，支持包括以太网、InfiniBand 等规格网络接口，形成超算所需的存储集群。同时，MOFS 并行文件系统可实现对计算、存储、互联等多种功能的支撑，是组建高性能计算集群必不可少的设备。

卫星遥感

典型场景：卫星遥感数据存储

应用案例：

新疆二测院是全国首批的甲级测绘单位，主要承担国家、自治区重大测绘工程及项目的设计与实施，其卫星遥感数据处理系统用于存放、处理各类卫星接收的数据、研究过程中产生的数据以及成果数据，其中最重要的一环就是数据存储。新疆二测院卫星遥感数据处理系统从存储容量、系统性能、链路带宽到整体架构亟需一个能够提供超高读写性能的卫星遥感数据处理系统。

同有科技自主研发的核心技术，打破层层技术壁垒，使得同有科技的 NCS 10000 分布式存储系统的单点读、写性能最高可达 6GB/s，成就“卓越性能”。同有科技为新疆二测院设计的遥感数据快速处理系统，以分布式存储系统、高速数据交换网络、高性能工作站为核心，能够充分发挥设备及系统的综合性能，同时满足数据处理速度和处理数量方面的要求。

同有科技 NCS 10000 分布式存储系统不仅为新疆二测院提供了单点读写的“卓越性能”，而且还可以实现单个文件最大支持 TB 级处理，最大可支持数千客户端同时访问的高并发能力、提供扩展至 EB 级别的海量空间、还能够进行数据全生命周期管理，帮助新疆二测院实现与现有系统无缝对接，未来升级无缝扩容。

高校

典型场景：校园云课堂系统

应用案例：

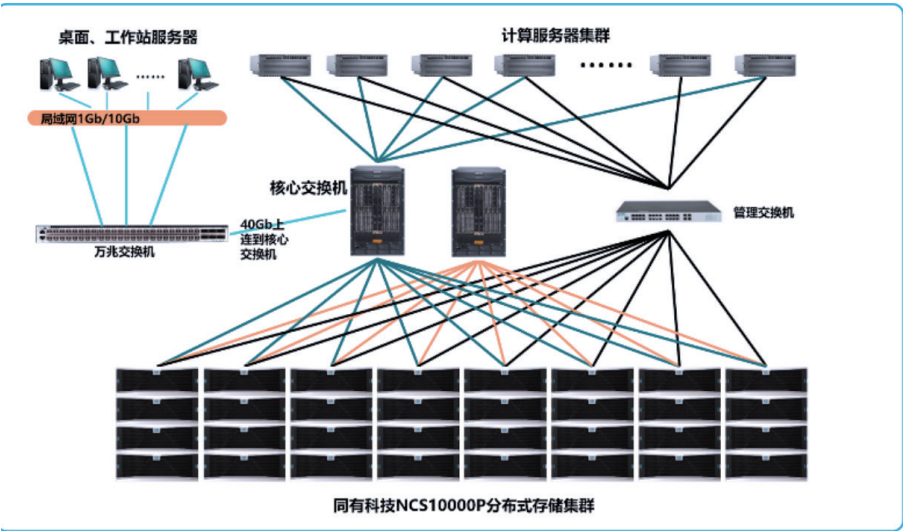
华南师范大学网络中心采用了 Nutanix Files 企业文件存储解决了云教学平台小文件访问频繁的挑战。云课堂的多数文件是 10M 以内小文件，视频教学素材多以微视频为主。网络中心曾考虑采用传统 NAS 文件存储，但测试后发现传统 NAS 系统并不适合教学平台海量小文件存储的需求，且无法应对存储资源快速增长的弹性需要。

Nutanix Files 专为处理数以十亿计的文件和成千上万的用户会话而打造，随着环境的扩展可以向 Nutanix Files 集群添加更多 vCPU 和 / 或内存，一键实现 Nutanix Files 集群的纵向扩展，也可以通过添加更多 Nutanix Files 虚拟机来实现横向扩展。

经过测试，Nutanix Files 性能可满足云课堂要求，配置和操作界面简单方便，结合 Nutanix 自带的 Files Analysis 组件可以满足对文件访问进行分析、审计、报表等管理需求。

武汉大学中南医院利用 Nutanix 超融合构建私有云，通过容灾功能实现多数据中心的容灾。此外，中南医院也通过分布式存储 Nutanix Files 进行文件存储。虽然医院数据量巨大，但两个站点的数据复制都可以在一小时内高效完成。

此外，借助 Nutanix Files，系统可自动检测异常的文件操作，并轻松实现文件授权、更改及删除等操作的监控，并且可以在



分钟内快速实现文件的恢复。

Nutanix 分布式存储的优势一是统一存储，通过一个解决方案满足企业不同的存储需求，这包括虚拟机存储（用于超融合）、块存储（用于数据库集群）、文件存储（NAS/NFS/SMB 存储支持，用于非结构数据）、对象存储（海量文件存储）、二级存储（与备份软件集成）。二是强大的企业级功能，数据存储除了存之外，还需要实现数据安全和数据的全生命周期管理。如来自内部威胁检测（恶意修改和删除等）、勒索软件的风险、数据加密的保护，包括数据自动分层、数据老化分析、用户和存储使用分析在内的数据全生命周期管理。三是容器存储的支持，Nutanix 提供针对容器的存储需求，满足持久化存储的要求。因此，Nutanix 支持各类容器平台的存储接口。

金融

典型场景 1：农信省级金融行业云平台

应用案例：

XSKY 星辰天合的软件定义存储产品和解决方案支撑着行业云、私有云、桌面云、数据库资源池、海量媒体数据、影像数据、

智能制造数据、Hadoop 大数据等不同类型的应用场景。

以 XSKY 星辰天合在金融领域的标杆客户——江苏农信为例，其目前采用 XSKY 星辰天合 SDS + 混合云的架构打造省级行业云平台，支撑 60 多家下属农商行业务稳定运行，整体成本降低了 50%，有效实现了数据统一管理和调配，推进全省下属银行共同实现数字化发展。

江苏省农村信用社联合社先后部署了 XSKY 星辰天合的 XEDP、XEBS、XEOS 等全线产品，部署存储节点数量超过 300 个，其混合云平台底层全面对接 XSKY 星辰天合分布式存储系统，数据空间超过 20PB。其中块存储对接 OpenStack 云平台，用于承载各农商行的业务应用系统，还对接 Kubernetes 容器云平台，用于承载微服务（容器应用商店、容器云测试平台）；对象存储则主要对接影像平台，用于承载海量非结构化业务数据；对象存储同时也对接了 Hadoop 大数据平台，用于承载分析类业务。

从最开始的合作到完成 90% 以上业务数据的覆盖，XSKY 星辰天合伴随江苏农信新业务建设一路走来。在云平台建设，既有影像云、大数据平台、容器迁移、行业云服务支撑等一系列项目中，采用 XSKY 星辰天合软件定义存储 + 混合云的架构统筹统建、统一容灾规划、统一安全策略、统一运维方式，有效降低了综合建设和运维成本，同时充分保障了全省各农商行业务的高可靠稳定运行和快速服务响应能力。

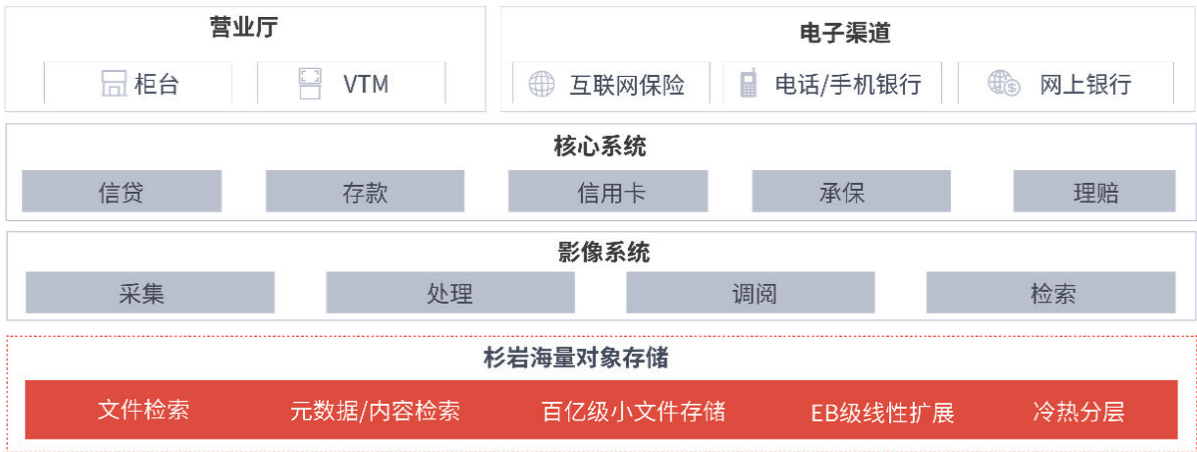
XSKY 星辰天合的 SDS 软件定义存储方案现已成为江苏农信行业云的统一数据管理底座，覆盖整个江苏地区几十家小型农商行、农村信用社的业务应用，计算节点多达 600 个，云主机数量过万，支撑 60 多家下属农商行和 3270 多家营业网点的业务稳定运行。与过去“分而治之”的数据存储架构相比，统一数据平台方案将 TCO 降低了 50%。

典型场景 2：金融影像数据存储

应用案例：

随着业务办理移动化、远程化、渠道多样化的发展以及监管要求，需要保存的影像、音视频、票据等非结构化数据与日俱增，传统技术架构已无法支撑，需要引入新技术以满足金融科技对海量数据的存储和管理需求。

杉岩数据金融影像解决方案不仅可以存储和管理客服中心语音、金融双录数据、电话销售中心、视频远程开户等系统产生的海量非结构化数据，而且可以进一步对不同数据分层存储和挖掘价值信息，在降低管理成本的同时助推科技创新。其弹性的分布式存储集群，彻底解决非结构化数据的存储问题；数据分层存储，支持整合 NAS 设备，数据自动归档迁移；多活数据中心容灾，



数据持续保护保障业务安全；支持对象元数据和数据检索，实现内容感知和大数据挖掘。

杉岩 MOS 海量对象存储系统（以下简称杉岩 MOS）是面向企业级海量非结构化数据的全分布式存储产品，针对 EB 级数据容量和百亿级文件数量的视频、影像和文档等非结构化数据，产品支持跨地域的数据中心主备、双活以及多站点容灾部署，提供高可靠和高可用的服务能力。杉岩 MOS 具有大容量、高性能和易扩展等优势，适用于归档与查询（金融影像、智能制造 AOI 影像、

医疗影像、政企电子文档）、备份、大数据分析、AI 训练和检索等应用场景。

交通

典型场景：突破 L4 级自动驾驶存储性能的瓶颈

应用案例：

随着越来越多的雷达、摄像头等传感器被部署在车辆里，各个环节的工作量与日俱增，尤其是高性能自动驾驶汽车对数据的存储需求更是巨大，每天每车生成的数据量可达到 3-8 TB。因此，越来越多的自动驾驶企业开始考虑如何高效、稳定地保证自动驾驶过程中收集到的大量数据，并快速形成自动驾驶的计算模型。

作为一家专注于软件定义存储技术的企业，焱融科技自 YRCloudFile 发布以来，长期深耕自动驾驶领域，在行业内积累了众多项目及实战经验，凭借高性能、高可用、高扩展、云上部署等特性，满足自动驾驶企业在各种场景下的需求。

以国内某专注于研发和应用 L4 级自动驾驶技术、聚焦自动驾驶出行和驾驶同城货运两大场景的自动驾驶公司为例，焱融科技在了解该公司驾驶训练场景以后，针对其自动驾驶训练数据集进行了一系列分析，总结出其训练数据具备海量数据文件、小文件难治理、读多写少的特征，从元数据处理能力、多级智能缓存、智能分层等方面，提供了一系列高性能、高可用、高扩展的存储方案。

首先是通过元数据处理能力，打造处理海量数据文件的基石。焱融 YRCloudFile 主要采用静态子树 + 目录 Hash 两者结合的方式搭建可水平扩展设计的 MDS 架构，这种架构方式有两种好处，一是实现了元数据的分布存储，通过扩展元数据节点，即可支持百亿级别的文件数量，二是在一定程度上，保证了元数据的检索性能，减少在多个节点上进行元数据检索和操作。

其次是通过多级智能缓存功能，提升整体性能。焱融 YRCloudFile 多级智能缓存能够在客户端缓存过程中，由内存缓存 + GPU 服务器本地 SSD 缓存组成，并可以指定缓存大小和位置。另外，训练程序会先从客户端内存缓存中加载，未命中则从客户端服务器 SSD 加载，不命中最后从文件系统集群中加载。

最后，利用智能分层功能，帮助数据流动做决策。焱融 YRCloudFile 通过高性能文件存储 + 低成本对象存储的组合，可以有效实现热数据依然为人工智能等新兴业务提供高性能访问的特性，让冷数据可以保存至低成本的对象存储中。

通过焱融科技 YRCloudFile，该 L4 级自动驾驶公司突破了存储性能的瓶颈，完美对接容器服务，完成数据跟随服务。另外，通过快速定位 PV 热点，该公司实现呈现 Pod、PV、PVC 实时监控与关联关系等独创性的管理功能，提升容器的管理效率。

证券

典型场景：云原生数据存储服务平台构建

应用案例：

鹏云网络 ZettaStor DBS 是一款可横向扩展的分布式块存储系统，可为数据中心提供统一存储解决方案，兼容主流云计算平台。该产品以软件定义和分布式技术，把大量经济型标准服务器的内置存储整合成为一个大规模、高弹性、高可用、高可靠的存储资源池。一套系统可向上层应用，同时提供块和文件存储服务，具备丰富的企业级功能特性，满足企业对结构化和非结构化数据存储的需求。

ZettaStor DBS 可广泛适用于电信、金融、医疗、教育、科研等各行业私有云和数据中心，无缝对接各类典型应用系统，无需任何变更，如：虚拟化 / 容器、数据库、海量数据存储、桌面虚拟化。

软件定义架构分布式存储是容器环境的理想选择，“标准化服务器 + 通用 IP 网络 + 软件定义”是云计算数据中心的主流建设模式，软件定义存储的架构与之相匹配，使得数据中心能够以统一技术路线进行整体建设规划以及运维管理，并利于后续的技术升级换代。只需添加服务器节点即可按需扩展，采用标准服务器设备，缩短采购和建设周期，使得系统规模能够以最小成本和最快速度精准建设与扩容，敏捷应对业务侧对存储资源的需求。数据分布式存储，数据访问分布式处理，增加服务器节点可实现容

量和性能的同步增长。可支持超大集群规模，实现海量数据存储规模和高并发 IO 性能，远超传统存储。

华泰证券基于 ZettaStor DBS 分布式块存储系统构建云原生数据存储服务平台；通过标准 CSI 接口与 Kubernetes 平台进行无缝对接；采用全闪存配置，为 MySQL 等数据库、Kafka 等中间件平台提供高并发低延时的数据访问性能。

从华泰证券的实践来看，ZettaStor DBS 分布式块存储系统技术路线与云原生平台相一致，满足系统架构先进性及不断演进的要求；其高可用、高可靠，满足金融行业严格要求，为云原生平台承载关键业务铺平道路；跟随容器生命周期的全自动存储资源调配，快速敏捷，加速业务上线；深度集成，提供容器环境对存储资源利用的统一视图，简化管理。

智慧城市

典型场景：海量数据存储应用

应用案例：

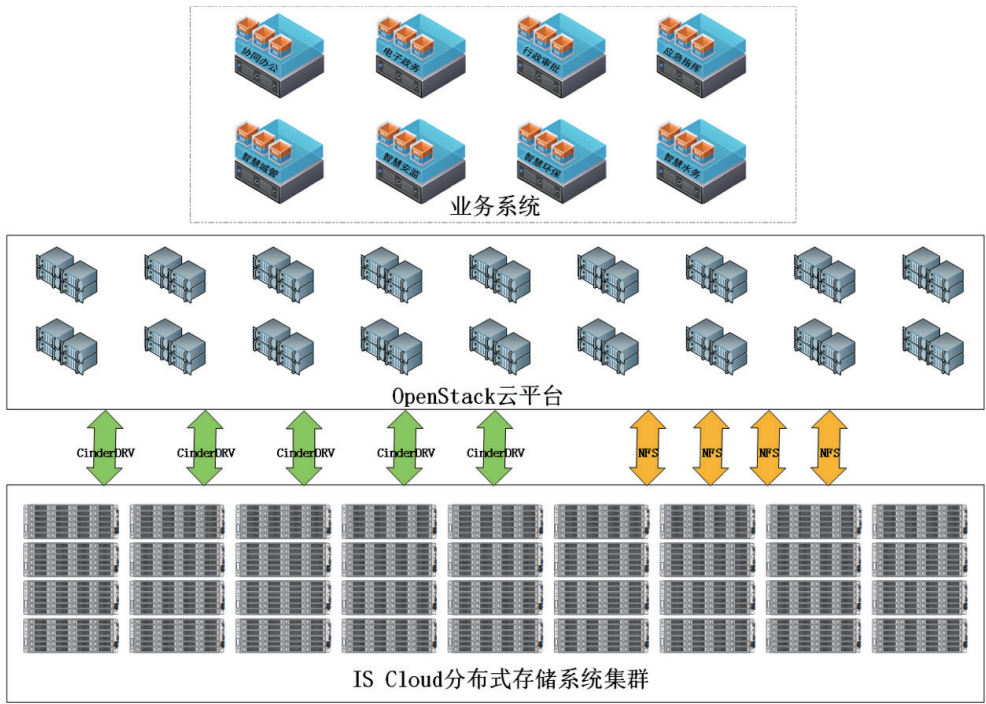
以智慧政务、智慧安全和智慧交通为代表信息化基础运营，特别是政府在云计算数据资源中心、协同办公平台、电子政务平台、网上行政审批、城市应急指挥、城市公共安全等城市智能交通等一站式、一体化业务应用支撑，提升了城市基础服务方面的管理能力。如何满足不同系统的数据多样性存储、不同业务的数据灵活使用、海量数据快速存取、海量数据存储可用性以及大规模存储设备的运维安全，对分布式存储系统提出了新的需求。

IS Cloud 分布式存储系统是由柏科数据技术（深圳）股份有限公司自 2010 年自主研发的海量数据存储系统产品。针对存储需求多样化、数据容量海量化、数据安全复杂化、存储设备规模化等应用特点的智慧城市建设需求，提供极致的存储性能、丰富的数据访问方式、可信的数据存储安全、系统安全以及良好的运维监控工具，智慧城市建设中作为数据基座的强力支撑。

由柏科 2018 年承建的某市智慧城市项目中的分布式存储项目，共计已上线 29PB 的存储容量。其中涉及政府政务云平台、治安视频监控、交通卡扣监控、交换违法图片、人脸识别等多部门、多业务系统的分布式存储系统建设。系统运行至今没有发生性能衰减、数据安全、系统安全以运维故障等情况。

智慧政务板块：

该业务模块由 33 台分布式存储节点组成 9.2PB 容量空间，为 OpenStack 云平台通过 CinderDrv 驱动提供分布式块存储空间；同时，通过 NFS 协议为 OA 办公系统提供文件存储空间。配置开启 SSDCache 功能为分布式块应用提供高速 IO 访问途径，通过系统分区技术配置 3 个故障域隔离与机柜级数据安全保护。合理、安全的对系统进行性能、数据安全方面的保护。



智慧公安板块：

该业务模块由 72 台分布式存储节点组成 20PB 存储空间，为业务系统提供 4:2 纠删码数据保护，同时，通过分区隔离技术配置 6 个故障与机柜级数据安全保护，以达到存储池级别的磁盘、节点和机柜三级数据安全保护。

数据协议方面，提供 iSCSI 协议为 1500 路高清视频监控提供数据读写链路；提供 CIFS 协议为人脸识别业务提供在小图片读写，同时，开启小文件加速技术，保障了每天千万左右的图片小文件正常的读写需求。

根据项目需求，在保证分布式存储正常运行所需要的硬件处理资源后，开启节点虚拟化支持，将业务平台的视频流媒体服务软件集成到虚拟机环境，节约大量的物理服务器资源。

基金

典型场景：海量数据存储应用

应用案例：

近几年，英大基金的业务规模从 18 亿增长至 1000 亿，增速高达 54 倍，位居全行业第一。在经营规模不断扩展、业务品种不断上线的同时，英大基金创新性地通过超融合升级生产数据中心和灾备数据中心的 IT 基础架构，有效支撑从一般生产到核心生产在内的各类业务，在保证用户体验和内部业务需求的情况下，提供了相比传统架构更低的总拥有成本和更高的敏捷性。

业务的快速发展给 IT 基础架构带来了很多挑战和需求。英大基金需要上线非常多的业务品种和模块，这要求 IT 提供快速的响应和支持，包括系统的快速分发、采购、部署上线。

以前英大基金采用传统的三层架构，其部署实施的速度跟不上业务发展的情况。同时，英大基金 IT 部门的人数非常少，只有几个人，而英大基金的发展目标是要迅速达到 1000 亿的规模。千亿规模的基金公司是需要有 10 名以上的 IT 人员。

英大基金选择了 SmartX 志凌海纳提供的超融合架构产品解决方案，因为它能够简化运维，让英大基金以最少的人员投入支撑业务的快速上线，并保障业务的安全稳定运行。英大基金采用超融合架构以后，把小机的业务全部都迁移到了超融合架构上。现在英大基金所有的业务系统，包括核心的数据库应用全部运行在超融合架构上，同时英大基金的容灾也都是在超融合架构上。超融合架构的性能效率与小机相比，其实一点不差，甚至还会更优。

目前基金公司普遍运行的等保二级系统，包括投资即 O3、TA、估值、清算、网上交易这五大核心系统，以及它们使用的 Oracle 数据库、上海报盘的 SQL Server、辅助周边系统用的 MySQL，都运行在超融合架构上。三年多来，没有任何一次的意外宕机的事件发生。

英大基金在采用 SmartX 超融合架构之前，机柜数量大概是 30 多个，在采用超融合架构之后，机柜数量减少到只有 3 个，整个机柜的能耗降低了大概 80%。同时，机房的设备也大幅减少，搬迁工作变得更加便利。

石油石化

典型场景：业务突发高峰压力及数据管理

应用案例：

作为全国首家获得成品油出口资质和配额的民企，浙石化是全国唯一一家可出口成品油的民营炼化企业。浙石化采用联想凌拓提供的联想 ThinkSystem DM3000H 混合闪存阵列打造基于全虚拟化环境下的核心数据管理系统支持平台，实现业务无中断运行，从容应对业务突发高峰压力及数据管理挑战。

因浙石化工程体量较大，在建设过程中产生大量的存档资料。这部分数据直接放在 Azure Stack 中，会占用大量的存储资源。

通过外接存储，把这部分资料挂载到存储中以减少云平台存储资源占用。

联想凌拓提供的联想 DM3000H 混合闪存阵列凭借其高性能、高扩展性、优化混合云部署的能力，以及支持无中断运营、简化 IT 运维和复杂度等功能，对浙石化的数据管理系统进行了全方位的升级。一方面，DM3000H 混合闪存阵列配置 12 块 10T 硬盘，可提供超过 80T 的可用空间，能够满足未来至少 3 年的业务数据存储需求。另一方面，凭借其无缝扩展和支持软件升级的特性，DM3000H 混合闪存阵列能够帮助浙石化更加轻松地应对当前及未来不断增长的数据存储空间扩展要求，打造适应未来需求的



IT 基础架构。

随着业务的增长，浙石化每月新增的上百万级别的非结构化数据，且都以小文件为主，联想凌拓提供的联想 DM3000H 混合闪存阵列可提供高达 80T 的可用空间，且凭借其强大的扩展特性，最大支持 24 控集群扩展，在保障其现有的业务数据存储空间的同时，也兼顾未来业务发展的数据增长趋势。

在性能方面，DM3000H 混合闪存阵列的 WAFL 文件系统能够为海量小文件类应用提供支撑能力，其 2T 的 FlashCache 可充分加速业务处理速度，在不降低可用性的情况下，延迟低于 1 毫秒，充分满足了浙石化对于数据管理平台高性能的需求。

在可用性方面，DM3000H 混合闪存阵列凭借高度可靠的硬件、创新软件和先进服务分析，可提供 99.9999% 或更高的可用性。其软件和固件更新、硬件维修和更换、负载均衡和技术刷新都可以实时执行，无需计划内停机时间，保证数据中心无中断运营、无忧更新，在运行日常工作负载的同时也无需担心维护窗口。其集成的数据保护技术也可以保护浙石化的数据，并实现加速恢复，与领先的备份应用程序集成，让数据管理变得更加简单。

此外，在混合云部署方面，DM3000H 混合闪存阵列通过 NFS 挂载到浙石化的由 Azure Stack 搭建的私有云平台上，凭借其 ONTAP 操作系统内置的 FabricPool 技术，将原有非结构化数据卸载到 DM3000H 混合闪存阵列上，同时也把测试环境搭建在 DM3000H 上使其承担 OA、测试等近线应用。此外，FabricPool 技术还可以对不活跃的数据进行分层，对应用程序的高性能 SSD 容量加以透明回收，将冷数据放置于成本较低的存储中，实现更加灵活高效的数据管理。

5. 总结

为了配合《2022 分布式存储市场调研报告》，百易存储研究院面向行业市场进行了调研，共回收有效调查问卷 3218 份，涵盖了行业企业用户和存储相关从业者。在此基础之上，百易存储研究院联合中国计算机学会信息存储专委会、中国计算机行业协会信息存储与安全专委会的专家对分布式存储技术应用的相关问题进行了深入研究，其中涉及注分布式存储系统的数据分布算法的选项，涉及哈希、一致性 Hash、带负载上限的一致性 Hash、带虚拟节点的一致性 Hash、分片等相对比较技术的内容。

如果从业务创新应用的角度出发，应该将注意力更多集中在应用，而不是存储技术细节，但是另外一方面，这些不同的技术也是针对业务实践中暴露出来的问题应运而生，代表着应用管理的先进水平，其中很多的问题会随着数据量的增加，以及硬件故障的增多逐步暴露出来，因此对于技术的关注还是非常必要的。

人的认知往往跟不上技术的发展，在一定程度的混淆。

《2022 分布式存储市场调研报告》结合专家们多年的工作经验，希望能够给行业企业的用户实践提供参考和依据，推动分布式存储技术的应用和发展。

鉴于存储技术的复杂性，有关错误在所难免，借此机会，我们也希望业内专家、广大读者的批评和指正！欢迎致电百易存储研究院，随时给我们留言。



DOIT 企业微信

