



宝德 GS9300 系列 CStore 分布式存储

技术白皮书

发布日期

2024-11

目 录

1 产品概述	5
1.1 数字经济时代存储革新迫在眉睫	5
1.2 产品简介	7
1.2.1 核心特性	7
1.2.2 产品架构	8
1.2.3 融合基础架构	8
1.2.4 网络逻辑架构	10
1.2.5 部署架构	12
1.2.6 使用引导	14
1.2.7 硬件规格	14
2 集群基础服务	16
2.1 资源分区	16
2.2 快速重构	18
2.3 业务高可靠	18
2.4 数据冗余	19
2.4.1 多副本	19
2.4.2 纠删码	21
2.5 数据一致性	23
2.6 缓存加速	24
2.7 数据分层	24
2.8 集群扩容	25

2.9 多协议栈网络	25
2.10 快照	26
2.11 抗静默错误	27
3 产品功能描述	29
3.1 登录与注销	29
3.2 概览	30
3.2.1 资源情况	30
3.2.2 性能情况	31
3.2.3 实时告警信息	32
3.3 资源中心	32
3.3.1 节点管理	32
3.3.2 集群管理	33
3.3.3 存储池管理	34
3.3.4 硬盘管理	37
3.4 存储服务	38
3.4.1 块存储服务	38
3.4.2 文件存储	42
3.4.3 对象存储	48
3.5 监控中心	50
3.5.1 基础资源监控	51
3.5.2 块存储服务监控	51
3.5.3 文件存储监控	52
3.5.4 对象存储监控	52
3.6 告警中心	53
3.6.1 告警中心概述	53
3.6.2 告警设置	53

3.6.3 告警记录	57
3.7 运维管理	57
3.7.1 日志管理	57
3.8 系统设置	58
3.8.1 权限管理	58
4 产品规格表	60

1 产品概述

在今天的数字化时代，云计算、互联网、社交媒体以及大数据的发展，使得数据量呈现爆炸式增长。传统存储在应对这些 PB 级海量数据需求时，其技术架构面临着诸多挑战，已经很难满足不断增长的数据需求，包括超大规模的横向扩展、越来越高的性能要求、数据长期存储的可靠性、统一资源池的管理、更低的 TCO 总体拥有成本等。

1.1 数字经济时代存储革新迫在眉睫

数字经济时代，数据已成为基本生产要素，存储作为数据的核心基础设施，其重要性日渐凸显。随着企业数字化转型的不断加速，数据规模越来越大，数据价值挖掘越来越深。

海量数据爆发，存储降本成为刚需。全球数据规模呈现井喷式增长，中国将成为全球最大数据圈。海量数据中，语音、图片、视频等非结构化数据约占 80%，与结构化数据相比，非结构化数据增长迅速、管理困难。与此同时，随着日益严格的数据监管要求，数据长期保存成为刚需，进一步推高存储成本，存储降本成为企业 IT 基础设施建设的重要议题。

智能化应用涌现，亟需存储性能提升。以 GPU 为核心的智能算力已成为我国主要算力基础，支撑了自动驾驶、AI 训练、基因测序等大量智能应用场景。智能应用高性能计算、海量数据处理的典型特征，对存储性能提出了更高的要求，除要求单点存储性能表现优秀外，还要求存储同时具备高带宽、高 IOPS、低延时的混合负载能力。传统存储高带宽与高 IOPS 通常不可兼得，难以应对复杂智能应用模型的存

储需求。

市场环境复杂多变，呼唤敏捷灵活的存储架构。宏观经济环境的不确定因素增多，消费者需求变化迅速，复杂多变的市场环境导致突发业务需求大幅增加，企业创新业务探索也对 IT 敏捷性提出了更高的要求，亟需可灵活扩展的存储架构。然而，传统存储系统容量与性能扩展存在明显上限，普遍采用存算一体的架构，计算与存储按比例配置，难以灵活扩展。

多数据集群，要求数据互通、统一运维。传统存储协议类型单一，无法同时满足块、对象、文件、大数据等多样性数据的存取。为应对海量、多样化的数据存取需求，企业往往需要建立多个数据集群。不同数据集群独立建设，数据孤岛严重。此外，由于不同数据集群管理要求存在差异，往往需要设立多个运维团队，或一个运维团队需同时管理多套系统，运维难度及成本大幅增加。

数据成为核心资产，稳定安全上升为核心战略。大规模数据体量和高并发业务访问下，传统存储在支撑业务持续稳定运行上，显得愈发吃力。与此同时，复杂的互联网环境也增加了数据被攻击、泄露的安全隐患。为提升数据安全与业务稳定性，各国均在加强灾备建设，然而与发达国家相比，中国在灾备领域的投入严重不足，不足发达国家的 1/4。

面对数字经济时代存储成本、效率、风险、敏捷的四重挑战，以 SAN、NAS 等集中式存储为代表的传统存储，在成本、性能、管理便捷性、扩展灵活性上显得愈发捉襟见肘。随着虚拟化、云计算等技术的出现，以及 X86、ARM 等基础硬件架构的成熟，分布式存储应运而生。其低成本、高性能、管理便捷、扩展灵活、稳定、安全的特点，可完美适应数字经济时代大规模存储的需求，并使其在各行各业得到快速推广，成为存储行业未来发展的重要趋势。

1.2 产品简介

宝德 CStore 基于云原生理念自主研发的新一代分布式存储系统,采用分布式存储架构统一了主流的五大存储引擎:文件、块、对象存储、并行文件存储、大数据存储;满足一套集群兼容所有存储,同一块磁盘上可以有文件、对象、块、大数据的任何一种数据,又满足各行业对云盘高性能、低延时的业务诉求,打破了原有存算一体的技术限制,打造支持共享存储的数据库、中间件等业务,赋能存算分离技术架构,进而为客户节省整体基础设施成本达 30%。

1.2.1 核心特性

极致安全稳定--多场景长期打磨: CStore 存储广泛在京东公有云/专有云/京东主站/京东金融使用,覆盖复杂广泛的核心业务场景;支持国密数据加密,数据安全稳定获得长期严格检验

五合一统一存储: 基于同一存储引擎底座构建,在一个存储集群里同时提供“块、对象、文件、大数据、并行文件”五种存储产品;存储资源可以充分利用,也使数据统一管理、数据互通成为可能

全自研可控+信创: 非基于任何开源存储软件改造,自主可控,强信创友好;与飞腾/鲲鹏/海光平台信创硬件完全适配,硬件能力充分发挥,相同硬件配置下的性能指标达到行业领先

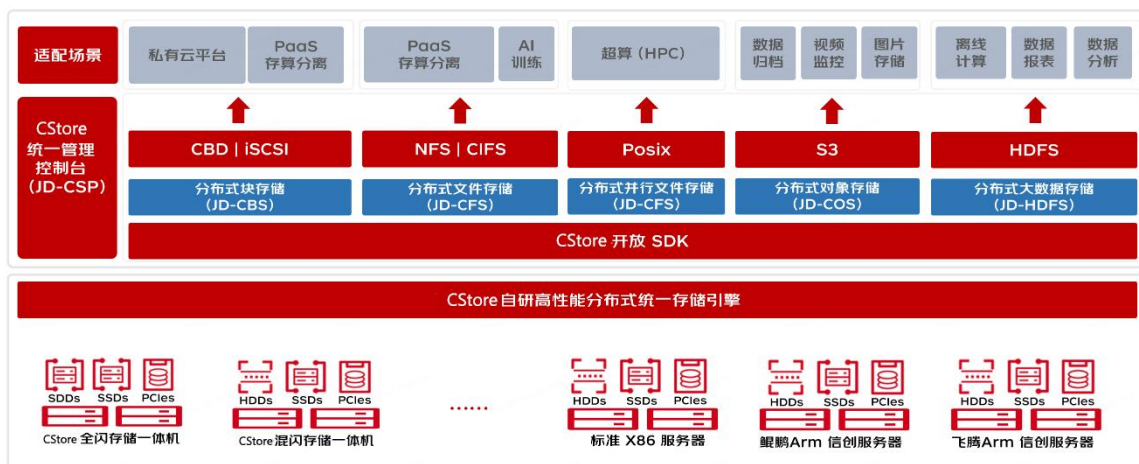
高性能: 作为可任意扩展的分布式存储,其性能接近本地 NVMe 磁盘;应用在存算分离、AI 训练、超算等场景下,可以为客户带来整体业务大幅提效和降本收益

低成本/高可用/高可靠: 低至 1.1x 副本的生产级存储,磁盘利用率最高近 90%,极大降低客户数据长期存储的成本;受益于新存储架构设计,做到 100% 写高可用,

其上承载的客户业务稳定性更强

1.2.2 产品架构

CStore 存储软件采用全分布式的架构：基础硬件层、存储软件服务、管控服务、增值服务等模块组成，这种架构为存储系统的可靠性、可用性、自动运维、高性能等方面提供了有力保证。其系统架构组成如下图所示：



1.2.3 融合基础架构

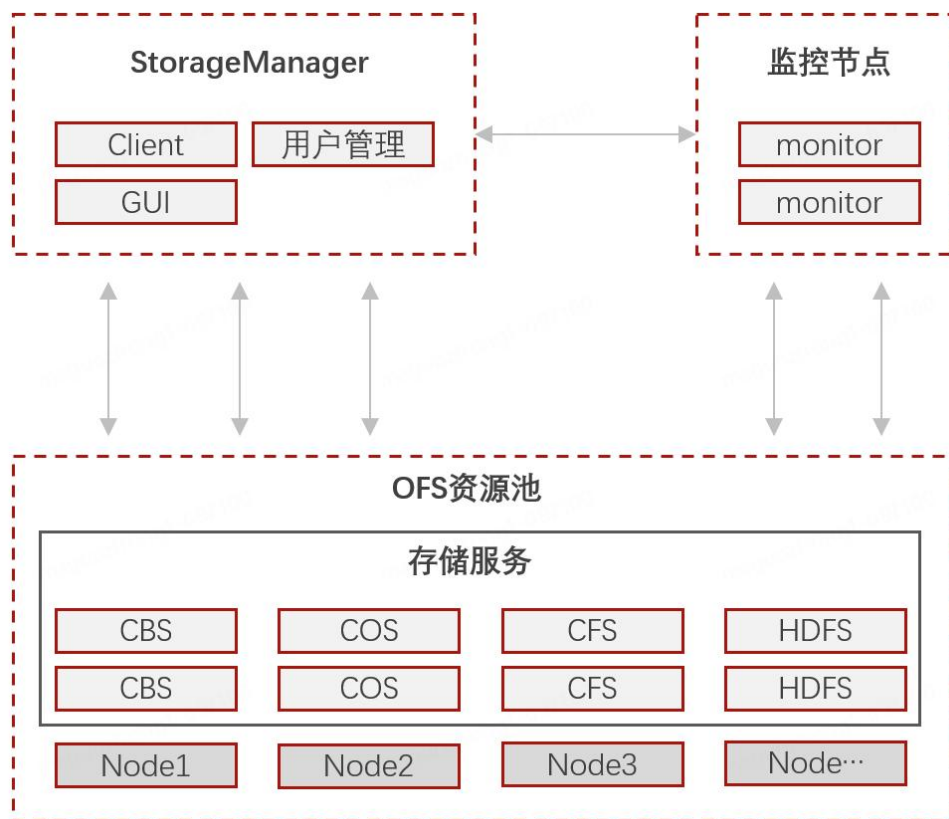
CStore 将 HDD、SSD 等硬件存储介质通过分布式技术组织成各类型大规模存储资源池，为上层应用和客户端提供业界标准接口，消除传统数据中心多类型存储系统烟囱式构建形成资源孤岛、硬件资源利用不均问题。

块存储服务：CStore 通过超低延迟 RDMA 网络，用户态 NVME、SPDK 全新存储架构，数据读写延迟降低至 100μs 以内，支持广泛的虚拟化平台及数据库应用，提供高性能与高扩展能力，满足云资源池、桌面云、开发测试云及数据库等场景存储需求。

对象存储服务：提供 Amazon S3 标准 API，满足云备份、云归档及云存储服务运营场景需求。

文件存储服务：提供 NFS 标准接口，以卓越性能、大规模横向扩展能力和海量

小文件为用户提供非结构化数据共享存储资源，应用于视频/音频、高性能计算、视频监控等多业务场景。



上图中，CStore 逻辑上可分为四部分：StorageManager、监控节点、ofs 资源池、存储服务。在实际部署中，这些逻辑组件可灵活部署，也就是说既可以部署在相同的物理服务器上，也可以根据性能和可靠性等方面的考虑，部署在不同的硬件设备上。下面对每一部分作简要说明。

(1) StorageManager

StorageManager 指的是 CStore 管控平台，包含对应的 GUI 界面以及 SDK，客户集成 CStore 管控平台在其运管平台，通过如下方式实现：

一是通过菜单的方式纳入 CStore 管控，即在客户运管平台提供 CStore 的菜单入口进行跳转，二是可以通过微前端方式，第一种应该相对改动小，易接入

(2) Monitor

Monitor 是集群监控节点。Monitor 持有 Cluster Map 信息。CStore Cluster Map 包括 ofs 集群以及存储服务系统本身的逻辑状态和存储策略的数据表示。

(3) ofs 资源池

Ofs 作为 CStore 统一存储引擎，对集群、存储池、节点以及硬盘做统一管理，在实际应用中，通常将每块硬盘（SSD 或 HDD）对应不同的逻辑资源池，ofs 负责所有的资源调度以及 MDS 信息管理、协议转换的 server 等，ofs 负责维护物理资源的更新状态以及系统的状态变化，与其他存储服务系统共同完成数据的存储和维护，与 client 通信完成各种数据操作等等。

(4) 存储服务

CBS，指的是块存储服务，可随机读写，类似物理硬盘，支持虚拟块设备协议接口或 NVMe 协议接口；CFS，指的是文件存储，通过 POSIX 接口并发读取相同文件，随机读写；COS，指的是对象存储，一个低成本的海量共享存储空间，适合存储写入后较少修改的数据；HDFS，指的是大数据存储，提供标准的 HDFS 访问协议，主要解决大数据场景下海量数据存储和数据分析，为用户提供高可用，高可靠，高性能、无限扩展、分层命名空间的分布式文件系统；

1.2.4 网络逻辑架构

CStore 网络主要包含前端网络、后端网络及管理网络。

类别	描述	网络协议类型
存储前端网络	在计算节点部署 UFS-Client，客户端与存储集群之前的数据通信	TCP/IP 协议 RDMA RoCE 协议 RDMA IB 协议
存储后端网络	存储节点之前的数据通信，用于	TCP/IP 协议

	副本同步，数据均衡、数据修复等	RDMA RoCE 协议 RDMA IB 协议
管理网络	CStore 存储系统的管理与运维	TCP/IP 协议
BMC 网络	远程带外硬件管理功能	TCP/IP 协议



● 节点机柜数量规划

三副本资源池，推荐 4 节点起配 (3+1)

EC (N+M) 资源池，推荐节点数量 $\geq N+M+1$ 节点 (比如 4+2+1)

推荐三副本机柜起配 4 个 (3+1)，EC 则为 $N+M+1$

注：请确认单机柜支持的最大功耗，机柜空间，并建议预留 1U 空间确保散热，

不同机柜摆放不同 TOR

- 组网规划

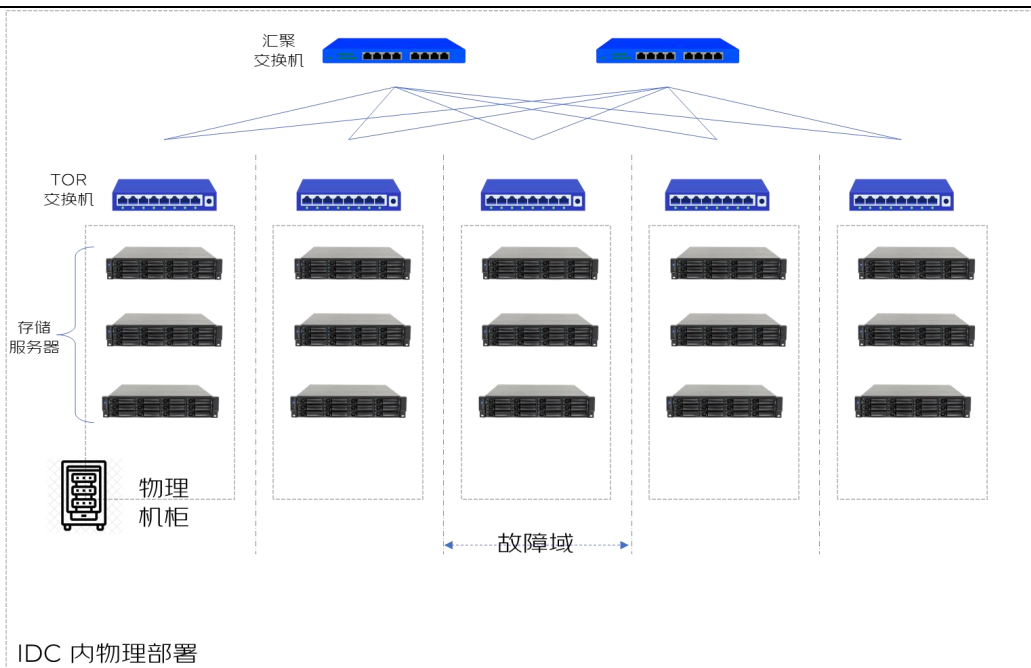
支持存储前后端共享网络和存储前后端分离网络，建议存储前端网络（Public net）、数据复制网络（Private net）以及管理网络（IPMI net）独立部署，分别采用不同的物理端口，除了管理网络外，网络端口速率不低于 10GE，对性能要求高的场景建议存储网络使用 25GE/100GE 或者 IB 组网。存储网口模式可以根据实际网络情况进行组 bond 提高性能、可靠性。

如图所示三副本组网拓扑和 EC 组网拓扑结构，以最小规模示例，机柜作为故障域，三副本推荐最少四台机器，分别部署在四个不同机柜上，每个机柜一个 TOR（T0 SW1），存储前端网络连接计算集群，存储数据复制网络连接存储集群，管理网络独立部署。EC 资源池则是 N+M+1 的最小故障域，根据具体服务器数量进行相应摆放。

1.2.5 部署架构

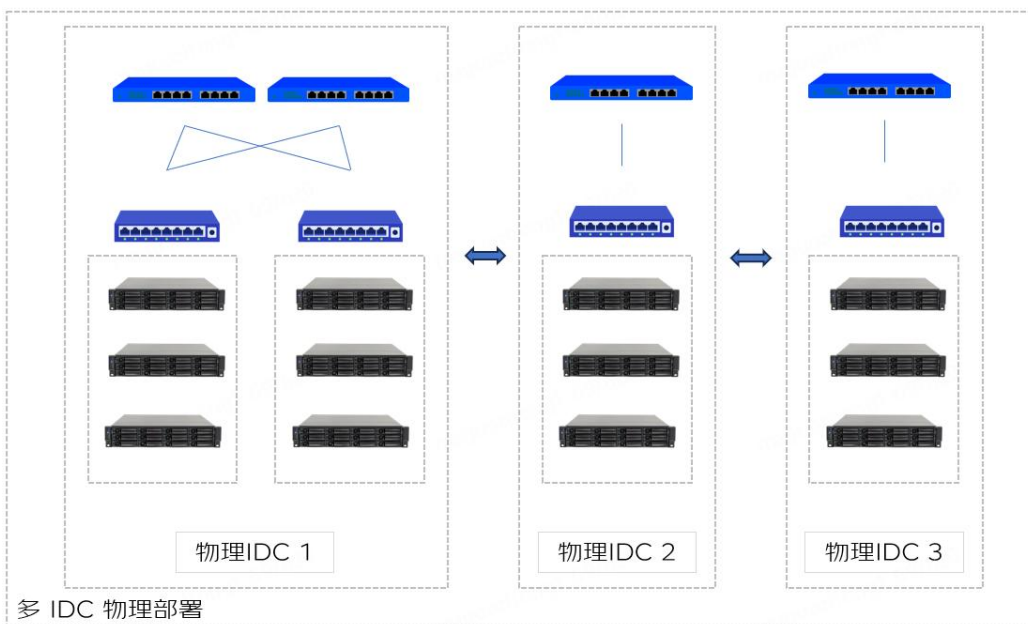
单机房内高可用部署

在 IDC 内部署，根据机柜、存储服务器、TOR 交换机等物理设备划分故障域。最佳实践是以 TOR 为基础进行高可用故障域划分。副本位于不同的故障域下，单盘、单机、单个 TOR 故障不会影响整体可用性



多机房高可用部署

对于非延迟敏感型业务，可以将故障域扩展到同城多机房。多机机房之间由高带宽的专线连接。在多机房的物理部署下，每个数据副本可存放在不同机房内，出现机房级故障时，也确保整体服务的高可用



1.2.6 使用引导

CStore 提供两个软件安装包，一是存储服务的应用安装包，即块、文件、对象或者大数据存储的服务安装包，二是在存储服务应用安装包环境部署完成后，提供 StorageManager 管理软件针对存储服务进行管理；

- 管控前端软件支持

操作系统	Windows 7 以上
浏览器	Chrome
	Firefox
	IE

- 软件安装包采用微服务形态提供

微服务名称	功能描述
StorageManager	统一管理存储软件服务以及基础资源管理，支持用户权限的设置
Ofs	针对物理节点进行管理，集群以及存储的划分，节点操作系统需要 CentOS 8.2
Cbs	块存储服务
Cfs	文件存储服务
Cos	对象存储服务

1.2.7 硬件规格

产品型号	全闪型	均衡型		容量型
系统架构	分布式架构			
每机箱最大容量	128TB	144TB	216TB	640TB
每机箱高度	2U	2U	2U	4U

每机箱节点数	1	1	1	1
单盘容量	7.68	12	18	20
每节点最大主存 盘数	16	12	12	32
每节点处理器	2 颗鲲鹏 920 处理器 或 2 颗 x86 架构处理 器或 2 颗海光架构处 理器	2 颗鲲鹏 920 处理器 或 2 颗 x86 架构处理 器或 2 颗海光架构 处理器	2 颗鲲鹏 920 处理 器或 2 颗 x86 架构 处理器或 2 颗海光 架构处理器	2 颗鲲鹏 920 处理 器或 2 颗 x86 架构 处理器或 2 颗海光 架构处理器
每节点最大内存	128 GB、256 GB	256、512GB	256、512GB	256、512 GB
每节点最大缓存	不涉及	4 个 NVMe SSD	4 个 NVMe SSD	4 个 NVMe SSD
每节点系统盘	2 个 480GB/960GB SATA SSD	2 个 480GB/960GB SATA SSD	2 个 480GB/960GB SATA SSD	2 个 480GB/960GB SATA SSD
数据盘类型	NVMe SSD	3.5 英寸 HDD	3.5 英寸 HDD	3.5 英寸 HDD
前端业务网络	>10GE、25GE 或 100GE TCP/IP >25 GE 或 100 GE RoCE	>10GE、25GE 或 100GE TCP/IP >25 GE 或 100 GE RoCE	>10GE、25GE 或 100GE TCP/IP >25 GE 或 100 GE RoCE	>10GE、25GE 或 100GE TCP/IP >25 GE 或 100 GE RoCE
存储互联网络	>10GE、25GE 或 100GE RoCE	>10GE、25GE 或 100GE RoCE	>10GE、25GE 或 100GE RoCE	>10GE、25GE 或 100GE RoCE

2 集群基础服务

2.1 资源分区

CStore 产品为统一存储软件，保存到集群中的所有数据会平均分布在所有的节点上。在集群规模过大时，为了能够对不同类型存储数据进行隔离管理，同时也为了提高系统可靠性，CStore 引入了分区特性。

分区特性的引入可以满足用户如下需求：

(1) 数据隔离：在集群同时支持块存储、对象存储情况下，不同的类型存储使用不同的物理服务器来保存数据，以达到数据隔离的效果

(2) 存储池性能划分：在块存储的场景下，不同的存储池使用不同类型的硬盘作为后端存储，以满足用户使用不同性能块存储的需求，比如一个存储池使用 SSD 硬盘作为后端存储以提供高 IOPS，另一个存储池使用普通 HDD 硬盘作为后端存储等。

(3) 无数据迁移的新业务添加：在集群现有业务正常运行的情况下新增存储节点用作其它的业务，不造成数据迁移

概念和原理

(1) 存储池

存储池是集群中以节点为单位组成的存储区域单元，不同存储池中包含若干台集群内节点，为用户提供不同的存储业务服务。通过存储池可以满足不同存储业务数据间的隔离，以及用户使用不同性能的存储需求。

在集群系统中，为了便于管理，会在存储池内建立多个机架（单个存储池机架数最多 32 个，所有节点池机架之和最多也为 32 个），属于该存储池的节点被添加

到存储池内其中一台机架上进行统一管理，每个机架只能属于一个存储池，每台节点只能属于一个机架。

(2) 硬盘

硬盘是集群节点以硬盘 为单位组成的存储单元，不同硬盘类型为用户提供不同的存储业务服务。通过硬盘池也可以满足不同存储业务数据间的隔离，以及用户使用不同性能的存储需求。

(3) 故障域

故障域是集群数据分片分布的最小单元。由于划分了故障域一个数据的几个副本或分片会分布在几个不同的故障域内，同一个故障域只允许存在一份数据分片，当集群出现故障时，可以提高集群的可靠性。

故障域类型为机架：

■ **机架：** 集群内的一个机架就是一个故障域，一个数据的几个副本或分片会分布在不同的机架里，同一份数据分片只允许存在同一个机架上。

推荐集群规模较大、机架比较多时选择。

(4) 部署方式

CStore 可以按照不同策略部署存储池，具体如下：

为了应对不同的业务场景，在部署集群时，提供了多种硬盘池部署方式：

■ **闪存加速池：** 硬盘池内只能使用 SSD 作为数据盘以及读/写缓存加速盘。

■ **均衡性存储池：** 使用 HDD 作为数据盘，SSD 作为读写缓存加速盘或者热数据盘。

■ **容量型存储池：** 硬盘池使用 HDD 作为数据盘。

可以针对用户不同的业务性能需求选择不同的部署方式，从而提供不同的性能

存储区域。可将节点上的 SSD 高性能资源单独划分到快池，来满足用户高性能的业务需求，将普通 HDD 硬盘划分到慢池为对性能要求不高的业务提供服务。该部署方式还同时实现了用户在同一分区内业务的隔离。

2.2 快速重构

CStore 分布式存储软件支持数据自动快速重构，硬盘故障或节点故障时，全资源池内自动并行重建，无需单独的热备盘支持，可极大提高重构速度，最快可达到每 TB 重构 30 分钟完成；快速重构主要是为了满足分布式集群数据负载均衡和集群高可靠性的要求。当数据冗余度下降或者数据分布不平衡的时候，分布式系统会在特定范围内发生数据的冗余度恢复的功能。

CStore 集群在两种情况下会发生重构：

- 1) 有新节点或新硬盘加入集群时会发生数据重构；
- 2) 集群中已有节点或者硬盘发生故障的时候会发生数据重构；

如数据为三副本配置，以硬盘 A 的数据分布为例，硬盘 A、B、C、D、E 组成一个存储池，A 硬盘上的数据副本分布到了 B、C、D、E，当硬盘 A 故障时，硬盘 B、C、D、E 都会参与数据的重构，这样每块盘占用的带宽以及需要重构的数据就会被分布到所有剩余硬盘上，能够极大的提高系统重构速度。

2.3 业务高可靠

(1) 简介

在使用存储资源时，如用户通过 iSCSI 连接到集群中的某个存储节点，该存储节点将为用户提供集群存储业务服务，但当 iSCSI 连接的存储节点发生故障后，整个存储集群将不可用，最终会导致用户的业务中断。CStore 的业务高可用特性可以

解决这一单点故障问题，启用业务高可用特性后，提供服务的存储节点发生故障后，服务能自动切换到备用节点，保证业务的连续性。

(2) 原理

将 CStore 集群内可以承担存储业务服务的一组节点划分在一起，组成一个高可用组。高可用组由一台 Master 节点和多台 Backup 节点组成，对外相当于一台虚拟存储业务服务节点。该虚拟存储业务服务节点具有一个虚拟 IP 地址，称为高可用 IP 地址。客户端只需要知道这台虚拟存储业务服务节点的高可用 IP 地址即可，客户端在使用该高可用 IP 地址访问虚拟存储业务服务节点时，将由该高可用组内的某一成员节点（Master 节点）为用户提供存储业务服务。

2.4 数据冗余

随着数据规模的增长和数据类型的多样化，单节点或者单盘故障概率会显著提升，从而使得用户对于海量数据的可靠性和安全性有了更高的要求，如何最大限度的保证用户数据的安全和可恢复，成为各存储厂商产品实力比拼的关键要素。

CStore 分布式存储所采用的比较成熟的冗余策略包括多副本和纠删码。

2.4.1 多副本

- 多副本概念

多副本是常用的一种数据冗余存储方式，通过复制的方式将原始数据分片的多分冗余数据保存在不同的存储位置上，如磁盘、节点等。从而提高了数据存储的可靠性和读取效率。

- 多副本实现机制

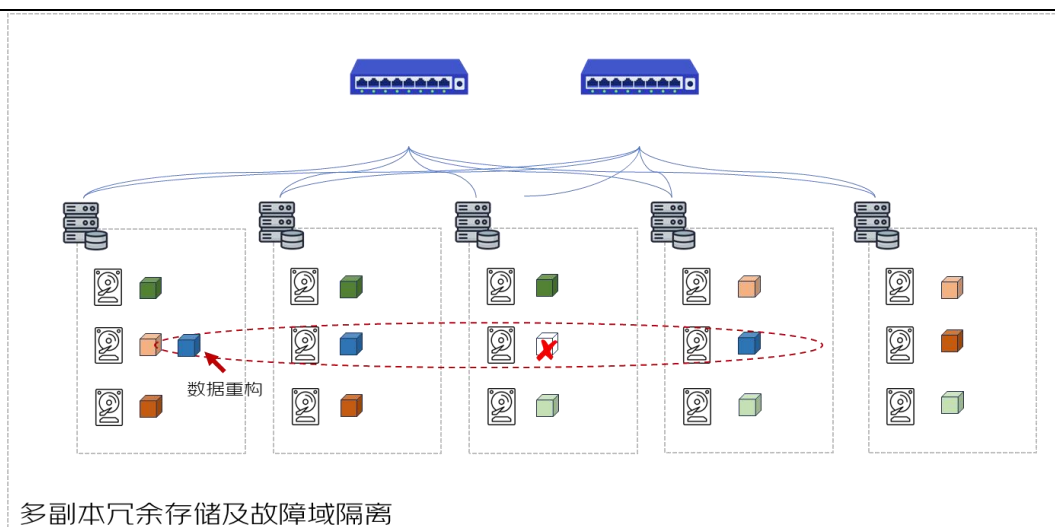
用户数据的不同副本放置到多个不同的磁盘，具体放置到哪些磁盘由数据放置

算法决定。如果某个磁盘发生了损坏，系统为保证副本个数，会将损坏磁盘所包含的用户数据利用其他磁盘的数据副本复制到其他可用磁盘（可能不止一个，不同系统有不同算法）。

CStore 支持 2-6 个副本数量的设置，副本数量决定了集群的最少节点数；在三副本冗余策略下，坏一副本，传统存储机制，剩余两副本可能有一副本数据不全，但 CStore 存储剩下两副本皆有完整数据，读写仍高可用，当剩下一副本时，传统存储机制，已无法读写，达到不可用状态，但 CStore 存储仍保持写高可用，读可用不选举，新数据写到健康的副本上，保持 100%写高可用，最后一副本数据一定正确。

副本数量	可用空间	可故障节点数	可故障硬盘数
2 副本	50%	1	同一节点上所有硬盘 不同节点上最多 1 块硬盘
3 副本	33.33%	2	两节点上所有硬盘 不同节点上最多 2 块硬盘
4 副本	25%	3	三节点上所有硬盘 不同节点上最多 3 块硬盘
5 副本	20%	4	四节点上所有硬盘 不同节点上最多 4 块硬盘
6 副本	16.67%	5	五节点上所有硬盘 不同节点上最多 5 块硬盘

当系统有故障发生时，部分副本丢失，系统基于故障域隔离，重新选择健康的磁盘并重构数据，重新形成多副本冗余。数据修复会设置对应 QoS，保障做迁移修复的数据搬移不会影响到前端实时数据业务



2.4.2 纠删码

● 纠删码概念

纠删码是指将存储数据分割成片段，把冗余数据块扩展、编码，并将其存储在不同的位置，比如磁盘、存储节点等。

同单纯的副本方式相比，使用纠删码在同样的硬盘或节点配置下可以为用户提供更多的有效存储空间，磁盘利用率更高。但由于系统需要更多的时间和处理能力来做恢复数据的计算，对 CPU、内存等硬件资源的消耗更多。所以集群整体读写性能 (OPS/Throughput) 会比副本方式略低。

● 纠删码实现机制

对于纠删码，CStore 的实现采用 N+M 模式，N 为原始数据分块个数，M 为冗余数据分块的个数。即系统收到数据后将一份数据分割、计算，在原有的数据基础上获得校验数据块，然后将原始数据块和校验数据块写到不同的设备中。在读取数据时，系统随机读取其中 N 个数据块，然后编码恢复出完整数据。

纠删码方式下，最多允许 M 个硬盘或节点故障，此时数据仍然可以正常访问，并自动恢复。除了数据的可靠性，其 Monitor 节点也是分布式部署的，同样不存在单点故障问题。这样在分布式存储系统中，不论是数据还是元数据，都不存在单点

故障，并达到极高的可靠性。

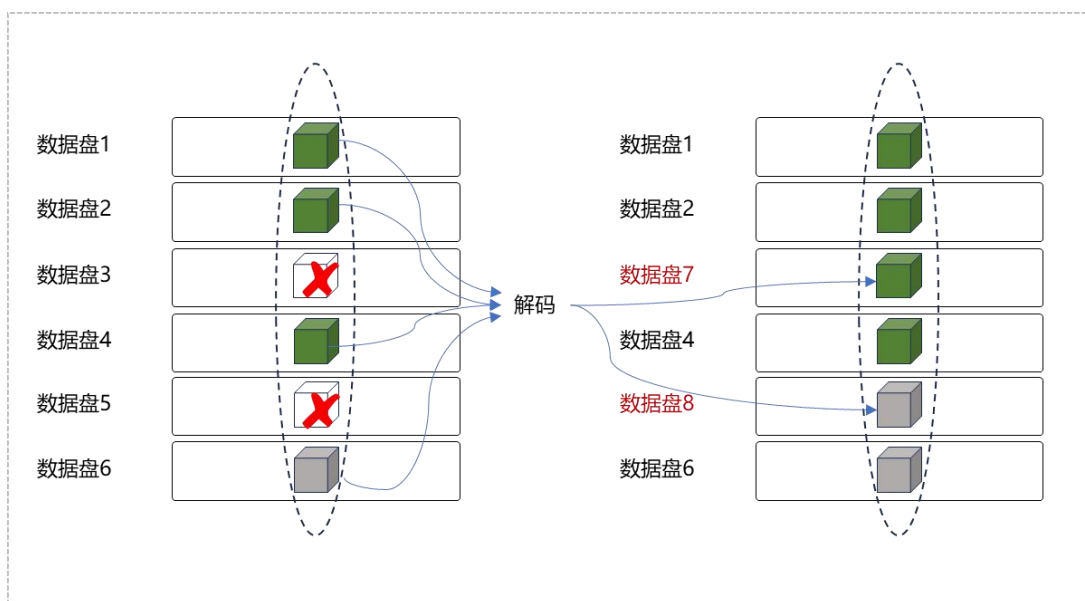
CStore 支持 N+2 到 N+4 的不同纠删码设置，主要规格如下：

类型	纠删码策略	对应副本	可用空间	可故障节点数	可故障硬盘数
N+2	4+2	1.5	66.67%	2	2
	6+2	1.33	75.00%	2	2
	8+2	1.25	80.00%	2	2
	10+2	1.2	83.33%	2	2
	12+2	1.17	85.71%	2	2
	14+2	1.14	87.50%	2	2
	16+2	1.13	88.89%	2	2
	18+2	1.11	90.00%	2	2
	20+2	1.10	90.91%	2	2
	22+2	1.09	91.67%	2	2
N+3	6+3	1.50	66.67%	3	3
	8+3	1.38	72.73%	3	3
	10+3	1.30	76.92%	3	3
	12+3	1.25	80.00%	3	3
	14+3	1.21	82.35%	3	3
	16+3	1.19	84.21%	3	3
	18+3	1.17	85.71%	3	3
	20+3	1.15	86.96%	3	3
N+4	6+4	1.67	60.00%	4	4
	8+4	1.50	66.67%	4	4
	10+4	1.40	71.43%	4	4
	12+4	1.33	75.00%	4	4
	14+4	1.29	77.78%	4	4
	16+4	1.25	80.00%	4	4
	18+4	1.22	81.82%	4	4
	20+4	1.20	83.33%	4	4

● 纠删码数据恢复

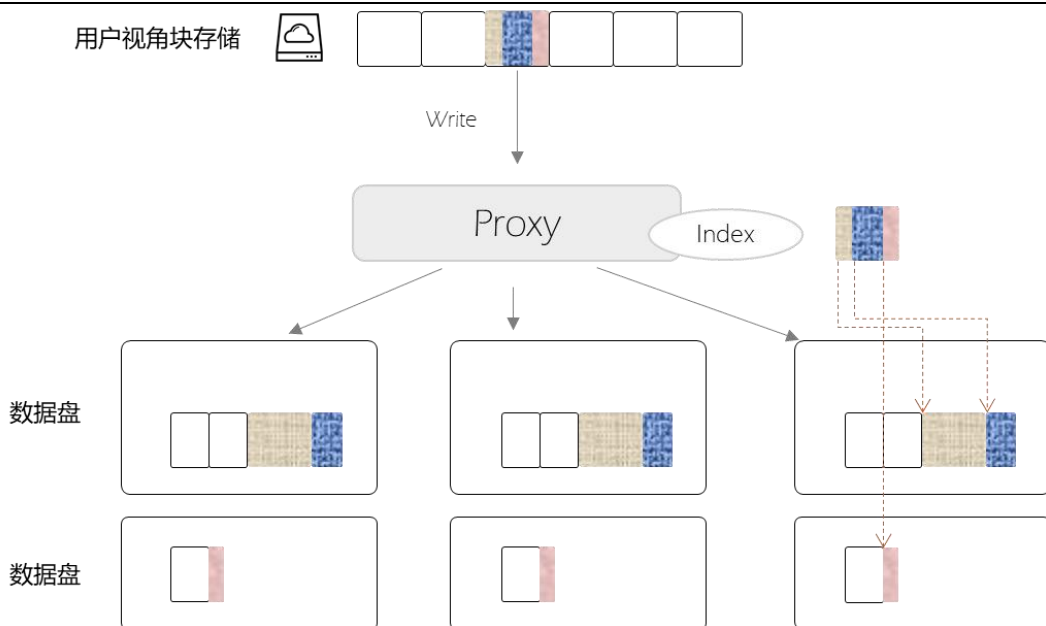
当系统有故障发生时，部分分片可能丢失，只要丢失的分片数目不超过 M 就可

进行正常数据读写。通过数据恢复算法，丢失的条带可从剩余条带中计算得到



2.5 数据一致性

在 CStore 系统中，一方面要保证元数据的一致性；另一方面要保证用户数据的一致性。CStore 高容错分布式一致性协议，数据可写在任意磁盘上，只要集群中存在 3 个可用磁盘即可保障服务可用性；数据三副本强一致写入，且不会因故障降级为两副本写入保障数据的可靠性，CStore 创新性追加写一致性，采用无 Leader 非仲裁协议，故障时无选举时间，不会因 Leader 发生故障时，选举新的 Leader 过程会有 10-20s 左右，敏感业务会受损，甚至数据有丢失的风险；

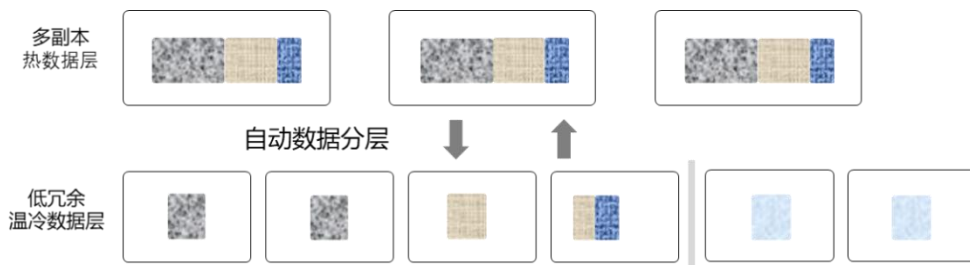


2.6 缓存加速

CStore 存储产品采用多级缓存技术逐层对 IO 进行聚合加速，提升系统整体 IO 性能。在 SSD 与 HDD 混配场景下，构建内存、SSD、HDD 的多层缓存机制，提升整体性能。在全闪存配置场景下，构建内存、SSD 的多层缓存机制，以获取极致性能表现。

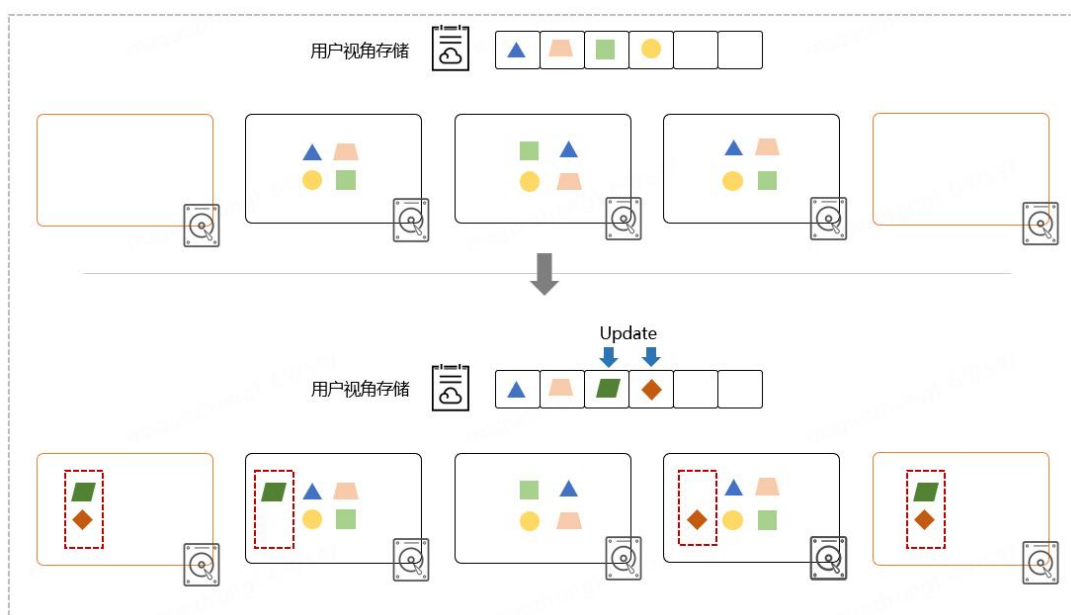
2.7 数据分层

CStore 通过独特的冷热数据分层机制，将热数据保存在内存和 SSD 组成的多级缓存分层中，极大的提升了数据命中的概率，提升读取效率。



2.8 集群扩容

CStore 架构下，所有数据都采用追加写模式写入，集群内增加服务器扩容时，不强制数据迁移再平衡。扩容后，新写入集群中的数据，会优先选择落入新扩容的节点。创建写入数据 Extent 时，根据磁盘负载、磁盘水位、健康状态分配磁盘，磁盘整体利用率均衡。

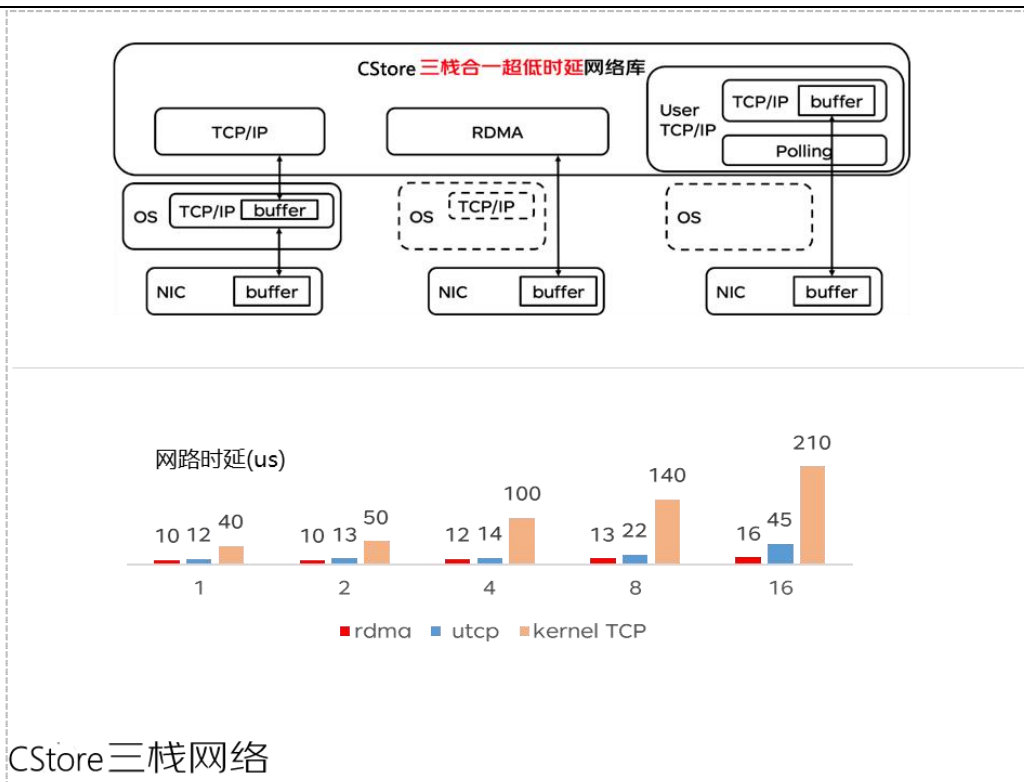


2.9 多协议栈网络

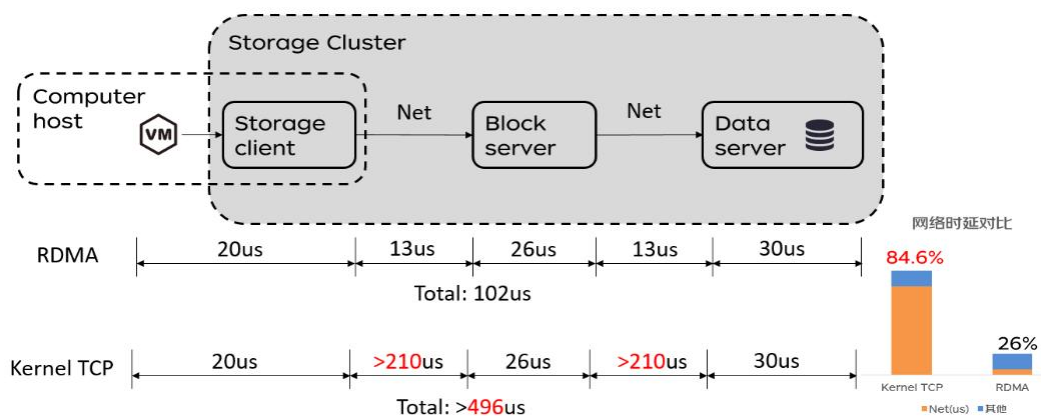
为了满足部分业务和数据的高性能存储需求，在 CStore 网络协议栈组件上，对 RDMA / DPDK / 传统 TCP 三种技术栈合一，兼容不同的 IDC 环境（高性能场景推荐使用 RDMA）。

技术上 CStore 通过超低延时 RDMA 网络及软件架构优化，带负载压力实测结果，在 4K 随机写 IOPS 达到 100 万时，平均延迟仍在 100μs 内，IO 性能接近本地盘。

另外，全数据路径采用全异步 + 非阻塞 + 无锁 + 零拷贝 全面的编程模型，最大限度利用硬件资源，性能损耗降到最小；

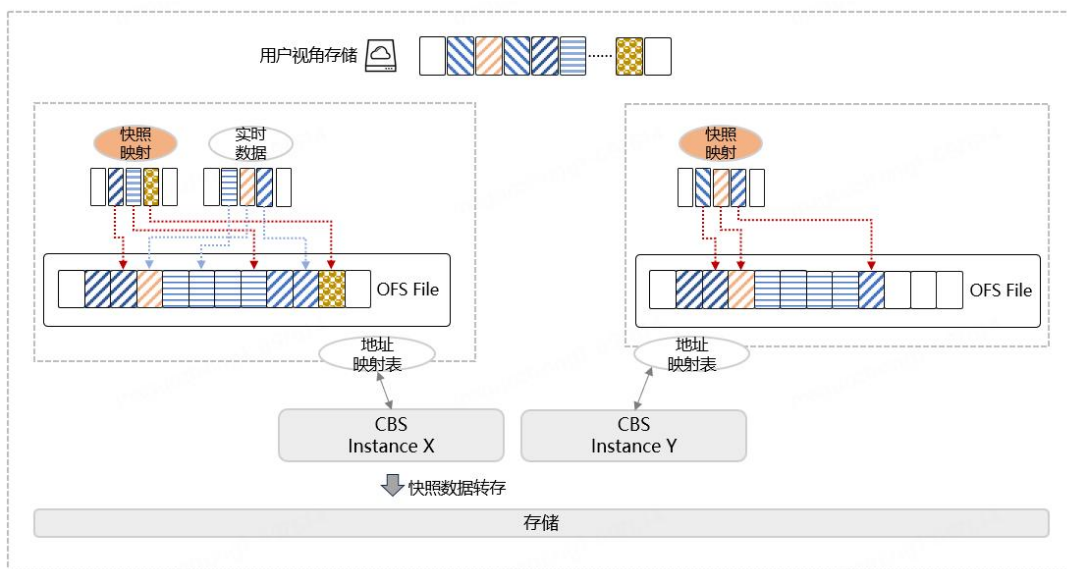


CStoreRDMA+存储 的工程实践：包分拆 + 流控 + 数据零拷贝。每一跳网络传输的延迟降低到 10us 左右，极大降低端到端数据读写请求延迟



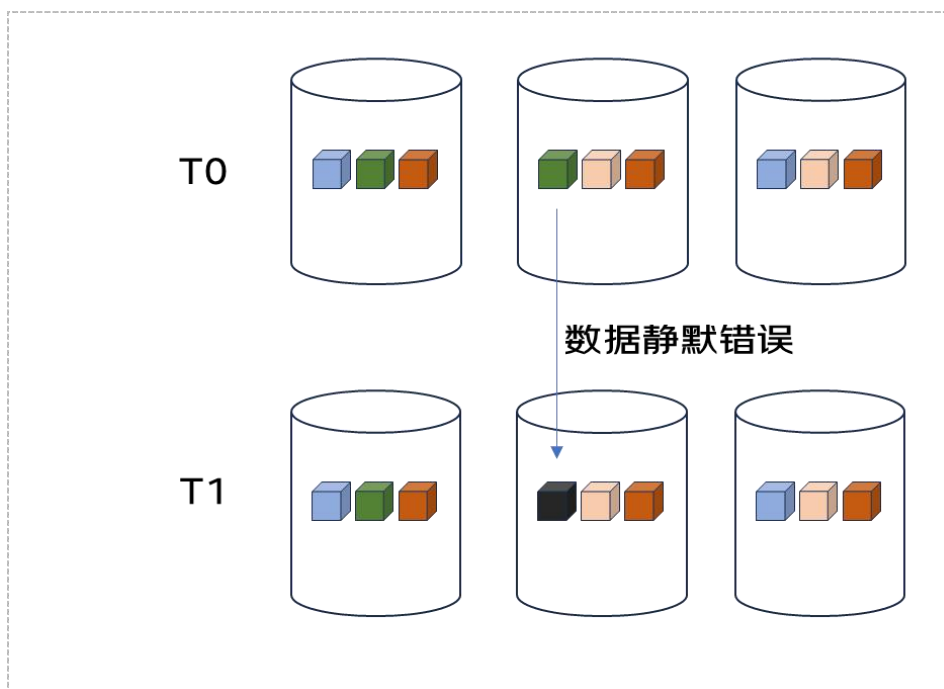
2.10 快照

CStore 块存储上提供了快照功能，将用户的存储实例的当前映射表进行保存，形成数据快照映射。对于有快照引用的数据块进行引用计数，防止数据被回收。快照数据会转存到存储上，按需进行长期保存。



2.11 抗静默错误

在数据存储过程中，数据在存储系统内部传输经过了多个部件、多种传输通道和复杂的软件处理，其中任意一个错误都可能会导致数据错误。如果这种错误无法被立即检测出来，而是在后续访问数据过程中才发现数据已经出错，会严重影响上层业务。由于静默数据破坏无法实时检测出来，导致被破坏的数据恢复难度很大，甚至不可恢复。



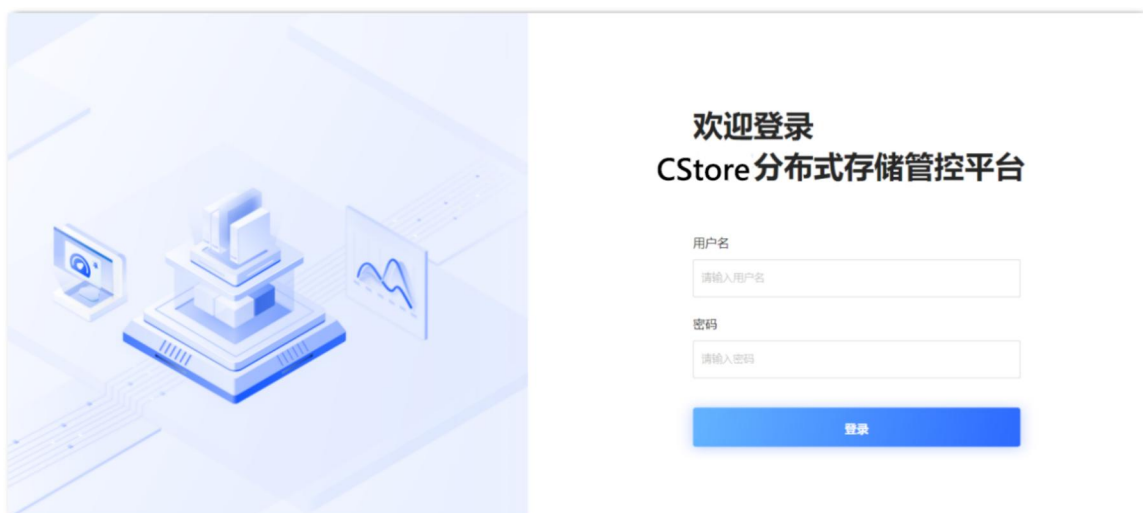
产生静默数据破坏的原因主要有：硬件故障（内存、CPU、硬盘、网络链路），软件 bug（产品软件、操作系统、应用程序等）

CStore 存储系统提供 IO 级端到端的数据完整性保护方案，能够有效检测跳变、读写偏等各种静默数据破坏场景，当检测到数据静默破坏后会实时对数据进行纠错自愈，避免数据损坏扩散。当磁盘静默发生时，数据服务对本分层追加的数据校验和校验失败，会放弃该数据块，由其它副本重新读取。无论在实时 IO 还是后台周期性 IO 识别到静默数据破坏时，均会触发自动的用户无感知的损坏数据纠错自愈机制。利用本系统内其他节点上存储介质上的数据对错误数据进行修复。

3 产品功能描述

3.1 登录与注销

用户安装完 CStore 独立管控平台后，在浏览器中输入“ <https://ip>” 进入系统登录界面，IP 为 CStore 管控软件的登陆地址；输入用户名和密码则进入系统；



说明：

- 系统登录只能单用户单客户端登录，不能单用户多客户端登录
- 系统登录前，若管理软件做过版本升级或者改动，且之前某浏览器登录过该集群，建议清空浏览器缓存后在登录
- 系统登录前，若重新配置过管理界面登录地址，需要情况该浏览器缓存后在登录
- 用户首次登录管控界面，提示用户需要进行初始密码的修改，新旧密码不能相同，新密码不能与最近一次的历史密码相同

密码修改

* 旧密码

新密码

确认密码

👁

👁

提交

重置

返回

3.2 概览

主要描述以集群为维度的资源、性能情况，并同步实时告警信息流，用户可在此页面获取到整体资源容量和使用量情况以及及时查看需要重点关注的异常信息；



3.2.1 资源情况

已使用容量：集群下所有硬盘被用户已读写使用的容量

总容量：集群下所有硬盘可被用户使用的总容量，含已使用+未使用的实际物理总容量

资源使用率：已使用容量/总容量*100%，取整数

磁盘繁忙个数：util 负载超过 80%的磁盘个数，值为整数

● 集群状态

健康：根据所有进程监控，如所有的进程告警级别不为危险/严重的，那集群集群为健康

告警：进程级别为危险/严重的，则为告警

恢复中：当告警发生后，系统已启动自修复功则为恢复中

更新中：当触发软件部署下发升级等配置更新内容时，则状态为更新中

● 数据状态

健康：根据磁盘的繁忙程度，即所有盘都到了 80%以上的 util 值即认为不健康，低于此则为健康

告警：磁盘 util 参数均在 80%以上则是告警状态

3.2.2 性能情况

带宽：展示最近 24h 的带宽使用情况，区分读写维度；并根据集群维度设置的阈值做对比；带宽采用单位时间内带宽之和，即每个节点读/写（出口/入口）带宽之和，阈值的计算逻辑是，带宽总的出口或者入口（根据网卡计算）值*经验安全值（可暂时设定 70%），如 10G 的网卡，那换算成带宽应该是 $10/8=1.25\text{GB/s}$ ，阈值应该为 $1.25*70\%=0.875\text{GB/s}$

容量：集群范围内，单位时间内的所有已使用容量之和，显示最近 7 天的，按照天的维度观察

当天以前的已使用容量统计按照每天 24 点的已使用容量计为历史值，如今天为 2 月 2 日，那 2 月 1 日的已使用是 2 月 1 日 24 点统计的已使用量，2 月 2 日已使用量为当前查询时间点的已使用量

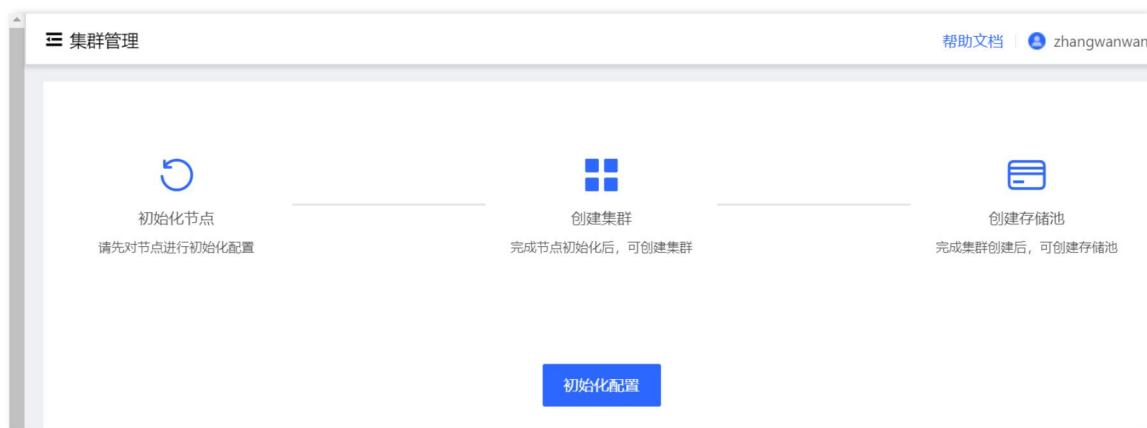
3.2.3 实时告警信息

- 告警区分级别为：危险、严重、警告、提醒（按照从高到低）；分别展示危险、严重、警告的个数
- 只展示最近 24h 内、且根据告警级别展示最近 5 条告警信息，如查看更多则点击“更多”，进入告警记录页面

3.3 资源中心

CStore 资源池的构建以硬盘为管理节点、存储池和集群，每一个资源池可提供数据&元数据存储的功能，资源池的创建前置条件需要有至少三台节点；

资源中心创建的步骤为先添加节点，添加节点的过程即部署 CStore 基础应用程序，如统一引擎系统，第二步创建集群，最后在创建存储池



3.3.1 节点管理

3.3.1.1. 节点初始化

登录系统后，直接进入节点初始化配置，初始化节点，节点至少需要 3 个，输入节点信息后即可部署基础的配置服务，并展示节点部署的进展和结果，包含检查的任务总数，成本、失败以及正在执行的个数，检查内容情况



3.3.1.2. 节点信息

添加节点后会在节点列表中展示界面名称、节点 IP、节点健康与否的状态、节点容量以及 CPU 利用率、内存利用率等相关的服务器性能指标，并可对该节点进行启停操作



3.3.2 集群管理

3.3.2.1. 集群列表创建

包含集群名称，节点选择，该节点主要存储引擎管控以及依赖的中间件等部署资源，该集群主要是 ofs 集群，作为存储服务如块、文件的基础资源池



3.3.2.2. 集群列表显示

创建集群后在集群列表中查看所有的集群信息，包括集群下的存储池个数，集群健康与否，是否有告警以及集群的总容量和已使用容量等



3.3.3 存储池管理

3.3.3.1. 创建存储池

存储池名称：名称不能重复；

所属集群：选择该存储池所在集群名称

存储类型：该存储池存储为 SSD，HDD

保护策略：副本和 EC，如果选择副本，则副本策略为三副本和二副本，如果选择 EC，则 EC 策略包含数据块和校验块的数据选择+自定义，默认是 4+2；10+3；

28+4 的选择，自定义的话则数据块和校验块由客户自行填充；采用 2 副本搭建起存储资源池则需要 3 节点，如是 3 副本至少四节点；如采用副本机制，推荐使用 3 副本，，保持较好的性价比以及可靠性；如对存储效率有比较高的要求，也可以采用纠删码技术；推进使用如下，也可以自定义设置

N+2	N+3	N+4
4+2	6+3	6+4
6+2	8+3	8+4
8+2	10+3	10+4
10+2	12+3	12+4
12+2	14+3	14+4
14+2	16+3	16+4
16+2	18+3	18+4
18+2	20+3	20+4
20+2		
22+2		

节点：如 2 副本则至少 3 台，如三副本，则至少 4 台，可多选；节点以 IP 为选择项

硬盘：根据节点的选择，罗列出该节点上对应的硬盘，硬盘以盘符(节点 IP)的格式选择

添加节点：选择一组节点+硬盘后，可在继续添加一组节点+硬盘



存储池在部署时会同步状态以及结果，如更新失败则重新部署



3.3.3.2. 查看存储池列表

可以查看存储池名称、盘类型、冗余策略、状态（健康、告警、恢复中、更新中）、故障域（目前支持 rack, ）、已使用容量/总容量、节点个数（已使用个数/总个数）、硬盘个数（已使用个数/总个数）、操作



3.3.3.3. 存储池操作

支持扩容&缩容，扩容针对节点和硬盘做添加，添加后弹出“执行结果”，确认后则完成；

缩容存储池，针对该存储池现有的节点和硬盘做缩容，可整个节点缩容，选择后弹出“请确认以下节点或者硬盘需要移除该资源池，移除后将不能被使用，且节点以及硬盘上现有数据将删除不在保留”，客户点击确认后则“执行结果”，确认执行结果后则完成

3.3.4 硬盘管理

3.3.4.1. 查看硬盘

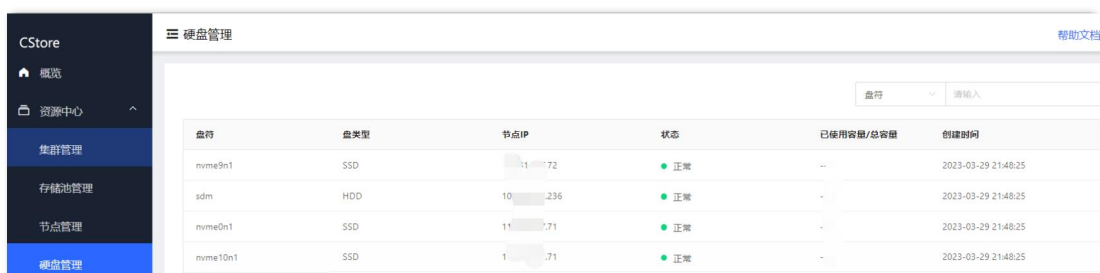
盘符：采用盘符（节点 IP）的格式来命名

盘类型：HDD、SSD

节点 IP：所在节点 IP 地址

状态：健康、告警、更新中、恢复中、停用

已使用容量/总容量：该盘已使用容量以及总容量情况



盘符	盘类型	节点IP	状态	已使用容量/总容量	创建时间
myme9n1	SSD	11.11.72	正常	~	2023-03-29 21:48:25
sdim	HDD	10.11.236	正常	~	2023-03-29 21:48:25
myme0n1	SSD	11.11.71	正常	~	2023-03-29 21:48:25
myme10n1	SSD	11.11.71	正常	~	2023-03-29 21:48:25

3.3.4.2. 操作

启用：节点初始化状态后默认为启用，云硬盘也即为可用，可对硬盘进行操作挂载等下发配置

停用：节点停用，表示节点上硬盘不可用状态，不能对其进行挂载配置下发操

作；也可单独对某硬盘停用，但需要谨慎确认硬盘上是否还有数据在运行，如停用则不能对盘进行读写数据

3.4 存储服务

3.4.1 块存储服务

3.4.1.1. 块存储集群

创建块存储集群

块存储平台集群是针对块服务做的资源池规划，是基于 ofs 基础资源池之上搭建的存储服务系统；

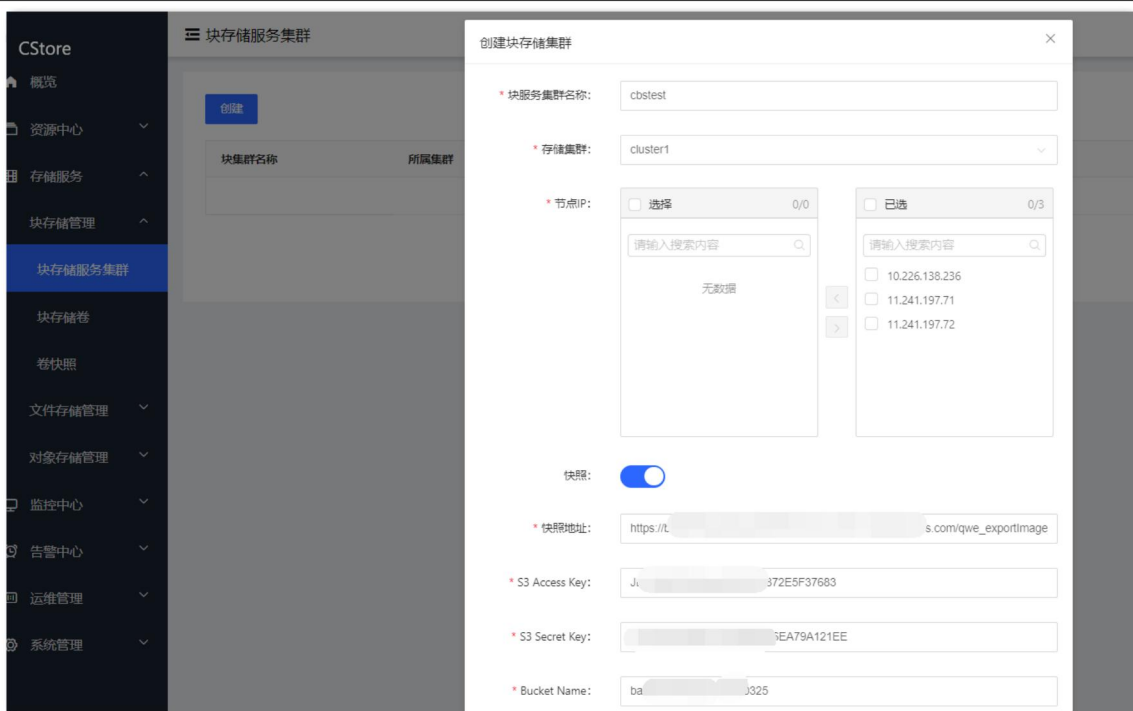
块服务平台集群名称：名称不能重复；

所属集群：即选择块服务平台集群搭建在哪个 ofs 集群中

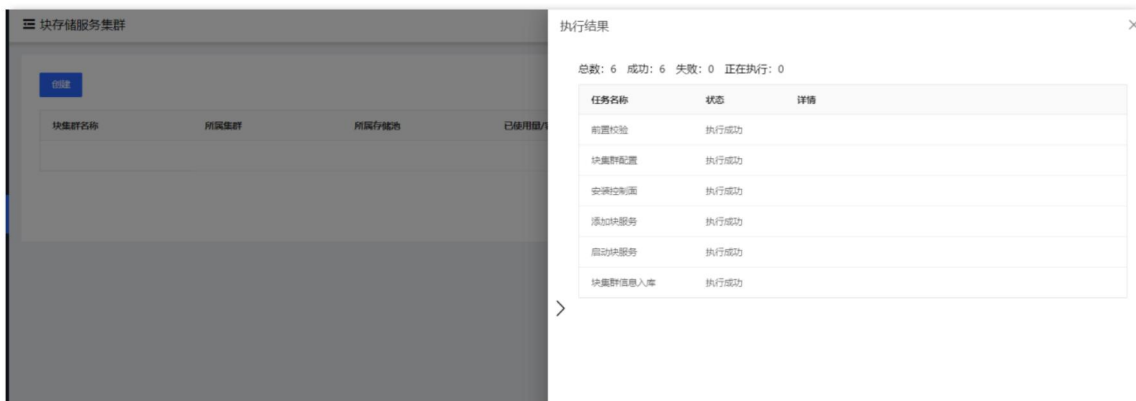
节点 IP：即至少三台；该节点主要是为了块存储服务管控以及依赖的中间件等资源部署

快照：快照是否开启，如果开启，则输入快照地址，地址目前只支持 S3 地址，格式 https://test.com/location/bucket_name/object_name

所有的存储卷快照内容都上传到该 S3 地址，需要输入 S3 对应的 bucket 与 aksk 或确保上传请求者身份的合法性；

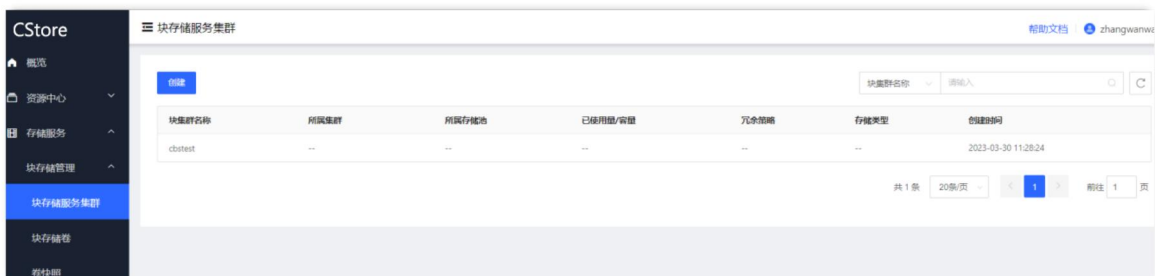


确认部署块存储服务后，会显示对应的部署列表



块存储集群列表（需要更新图）

块存储集群名称、所属集群、所属存储池名称、已使用容量/总容量、冗余策略（延承 OFS 的存储池策略）、存储类型（SSD/HDD）、创建时间



操作块存储集群

编辑更新块存储集群相关信息，内容包括：块存储平台集群名称、所属集群、节点 IP、快照是否开启、快照地址；主要针对块存储平台集群名称编辑，以及快照是否开启&快照的 S3 地址做更新

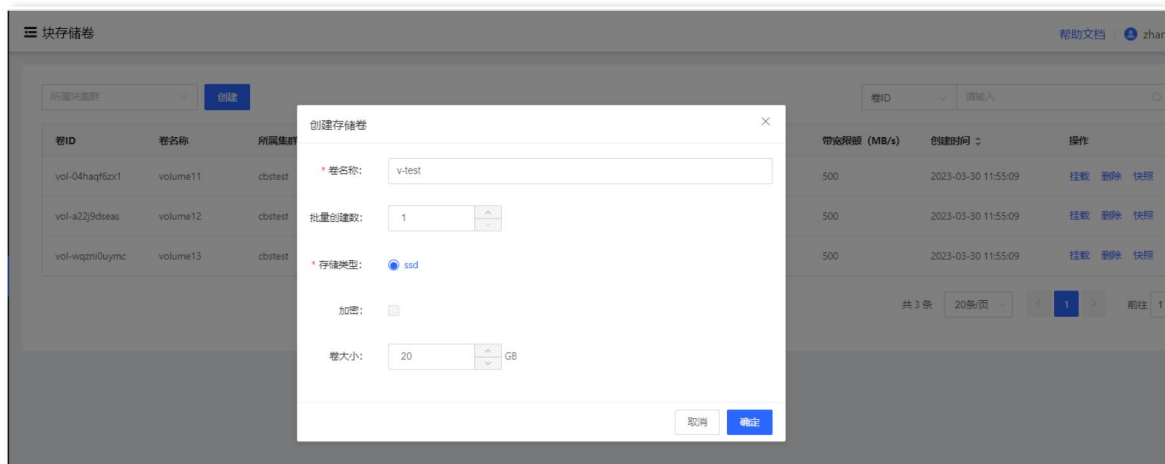
扩容块存储集群资源，内容包括：块集群名称（不可编辑，只是展示名称），所属集群名称（不可编辑，只是展示名称），节点选择（请输入或者选择节点 IP），点击确认后则弹出“执行结果”，任务执行均成功且确认后则完成

3.4.1.2. 块存储卷

块存储卷创建

块存储卷以租户维度进行操作，包括创建块存储卷，查看块存储卷列表页，对单个块存储卷进行挂载、制作快照、恢复、扩容、删除等功能。

当点击“控制台-存储服务-块存储管理-存储卷”页面上方的“创建”；



块存储卷列表

块存储实例列表页包含存储卷 ID、存储卷名称、类型、所属块存储平台集群、所属存储池、冗余策略、是否加密策略、容量、IOPS 限额、带宽限额、挂载实例、状态、创建时间的查看；

展示细节如下表所示。左上角支持对块存储平台集群维度的筛选，默认全部展示。右上角支持对存储卷名称和存储卷 ID 的搜索，搜索项目通过下拉框选择，默认为存储卷名称，搜索支持模糊搜索。

卷ID	卷名称	所属集群	存储池	容量	冗余策略	状态	IOPS限额	带宽限额 (MB/s)	创建时间	操作
vc-dh-0w	v-test1	cbstest	pool1	20GB	三副本	可用	5000	500	2023-03-30 11:55:54	挂载 删除 快照
vc-dh-x1	volume11	cbstest	pool1	20GB	三副本	可用	5000	500	2023-03-30 11:55:09	挂载 删除 快照
2-...as	volume12	cbstest	pool1	20GB	三副本	可用	5000	500	2023-03-30 11:55:09	挂载 删除 快照
vc-ni0uymc	volume13	cbstest	pool1	20GB	三副本	可用	5000	500	2023-03-30 11:55:09	挂载 删除 快照

操作：支持挂载/卸载存储卷，针对挂载目标输入主机 IP，并在卷卸载的情况下可删除卷，当存储卷为可用或已挂载状态时才能制作快照，为了保证快照能完整的捕获磁盘数据内容，建议避开业务高峰，制作快照前，停止数据读写操作。快照加密和源硬盘加密特性一致

3.4.1.3. 卷快照

快照列表页如图所示、展示字段快照 ID、快照名称、源存储卷名称、源存储卷 ID、源存储卷类型、加密、快照描述、容量(GB)、状态、创建时间、快照删除时间、操作（根据快照创建、删除）

可通过点击源硬盘名称或源硬盘 ID 跳转到对应的云硬盘详情页。

快照ID	快照名称	源卷ID	源卷名称	容量	是否加密	状态	创建时间	快照删除时间	操作
暂无数据									

3.4.2 文件存储

3.4.2.1. 文件存储集群

文件系统集群是建立在资源中心基础上，即所有的文件服务基于 ofs 集群搭建的资源池，文件集群需要关联块集群，文件系统的最终数据落盘在块集群资源上，与块集群共同使用 ofs 集群资源

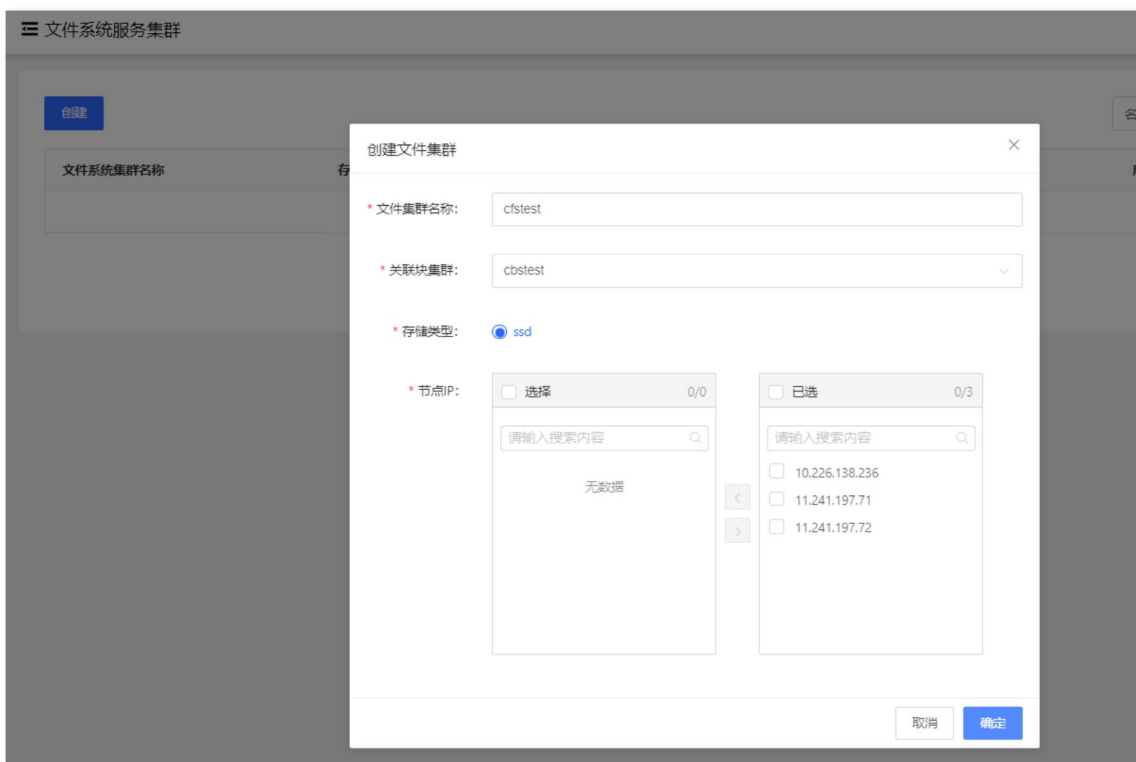
创建

有文件集群创建操作权限的管理员&超级管理员用户可以基于资源中心的资源池做文件系统集群的划分，内容包括

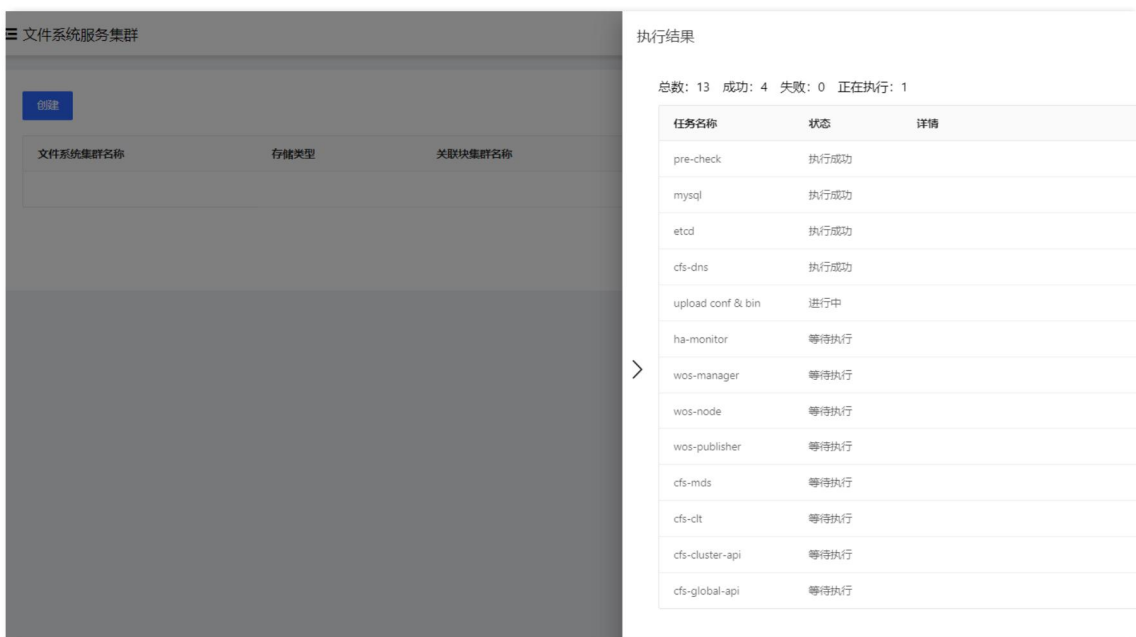
- 1、文件存储集群名称：名称不能重复；
- 2、存储类型：为 SSD 或者 HDD，当选择 HDD 时，在 hdd 下方提示：SSD 可提供高性能读写，如文件系统对性能要求较高，建议使用 SSD；默认选择 SSD
- 3、关联块存储集群：需要确认落盘在哪个块集群服务中，当存储类型选择 SSD 时，关联的块存储集群为 SSD 集群，获取的块存储集群的下拉选项应该只展示 SSD 类型的块存储集群名称；

当存储类型选择 HDD 时，关联的块存储集群为 HDD 集群，获取的块存储集群的下拉选项应该只展示 HDD 类型的块存储集群名称；

- 4、节点 IP ：文件系统部署在哪些节点上，至少选择 3 个节点但最大值目前暂不限制，同一节点不能被添加到其他文件系统集群中，故该节点 IP 应该没有被其他文件集群节点添加过；



文件系统部署进展通过列表显示：



概览信息包含：总数（即所有操作的个数总和）、成功（针对操作项状态为成功的个数总和）、失败（针对操作项状态为失败的个数总和）、正在执行（针对操作项状态为正在执行的个数总和）、等待执行（有前置条件等待部署的操作状态为等待执行）

列表展示

文件存储集群名称、存储类型（SSD/HDD；可在列表值上进行筛选，如筛选 HDD 则显示所有 HDD 相关的集群列表信息）、关联块存储集群名称、已使用量/总容量（逻辑使用量以及总容量）、冗余策略（需要调用所划分到哪个 ofs 集群的存储池数据：包含参数值为 三副本/二副本/EC：4+2）、状态（健康、告警、恢复中、更新中）、创建时间、操作

A、已使用量：指的是文件存储集群实际用户写入数据等已消耗的资源用量，即逻辑使用量；具体算法=所有文件系统用量之和，单位为 GB

B、总容量：指的是文件系统可用的总的磁盘量，是分配给文件存储集群的所有逻辑磁盘容量之和，单位为 GB

已使用量/总容量需要去掉软件应用占掉的资源量，此为客户实际线上可使用的容量

C、冗余策略：根据 OFS 资源池的冗余策略匹配，一个 OFS 资源池目前的冗余策略是唯一的

D、状态

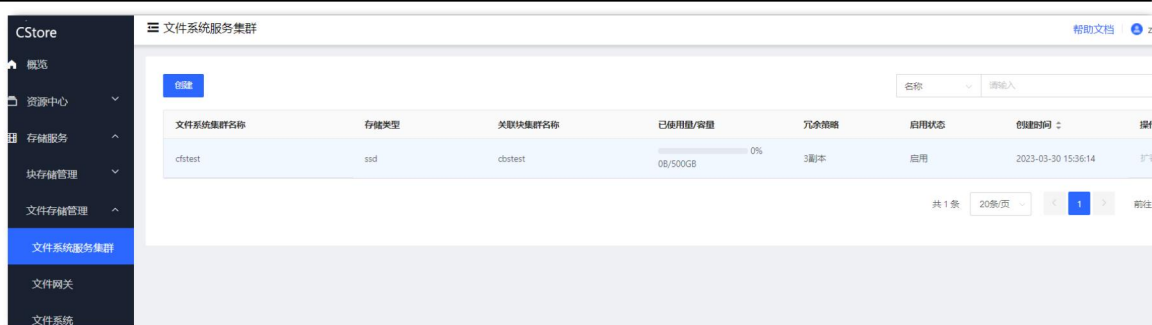
健康：根据进程监控，如所有的进程的告警级别不为危险/严重的，那集群为健康

告警：进程级别为危险/严重的，则为告警

恢复中：当告警发生后，系统已启动自修复功则为恢复中

更新中：当触发软件部署下发升级等配置更新内容时，则状态为更新中

E、创建时间:集群创建成功的时间，显示为北京时间



扩容

可针对选中的文件集群做节点资源的增加，节点扩容针对的是 IP 的资源增加；

3.4.2.2. 文件网关

文件网关主要指的是文件客户端请求文件服务时，需要通过文件网关指向后端服务，文件网关根据授权的客户端进行资源的调度和分配

创建

文件系统集群名称：选择已经创建的集群

文件网关的创建和添加：默认以文件网关组为单位，默认的网关组命名规则为文件网关组+顺序数，如文件网关组 1、文件网关组 2....

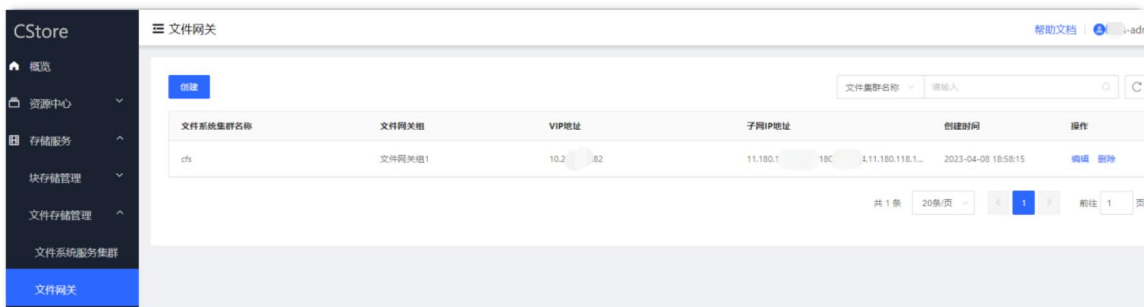
VIP 地址列表，即针对 IP 节点列表映射到的 VIP 地址，VIP 与网关 IP 地址一一对应，个数一致，VIP 输入可以是多个

网关 IP 地址列表，该部分在部署文件集群的时候可以获取，需要调取接口供用户选择，同时用户也可以模糊输入来匹配对应的 IP



列表

文件系统集群名称、文件网关组，VIP 地址（VIP 与 IP 地址在表中只展示 3 个，其余用省略号显示，当鼠标放在 IP 地址上时，显示浮框，浮框内显示对应的 IP 地址），服务节点 IP、创建时间



操作

针对 VIP 以及 IP 列表可做增加删除操作的编辑，也可以删除文件系统集群；根据文件存储集群名称搜索

3.4.2.3. 文件系统

创建

需要明确在哪个文件系统集群下创建文件系统，确保文件系统的权限以及网关选择；

文件系统名称：命名规则遵循复用对名称的所有规则

授权客户端 IP：指该文件可被读写或者访问的客户端 IP 地址

访问权限：读写、读、写

文件网关组选择



列表

文件 ID、文件系统名称、所属文件系统集群、所属存储池、已使用量/总容量、是否挂载（是/否）、冗余策略、状态（开启/停止）、创建时间、操作（启用、停止、删除、编辑）



操作

可针对文件系统做配额&限速的操作，配额可针对文件系统设置文件的个数以及文件的容量，限速，可针对文件系统的读写速度以及 IOPS 进行限速，确保文件资源的服务性能；

文件系统名称	所属文件系统	状态
test	cfs	available

文件系统名称: cfs-test

吞吐量: GB

读写速度: MB/s

读速度: MB/s

写速度: MB/s

读写IOPS:

读IOPS:

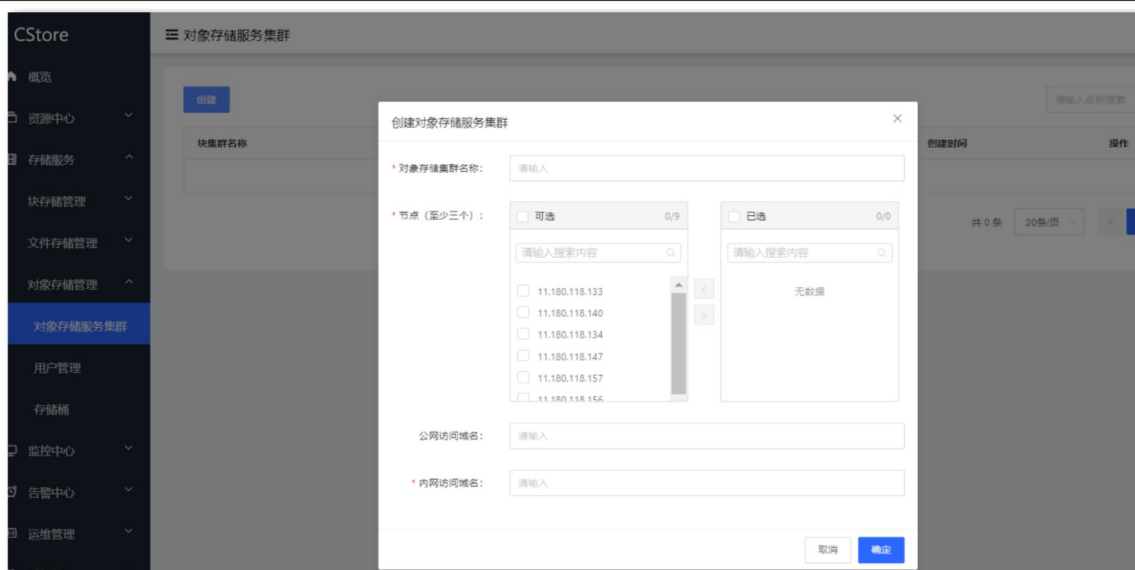
写IOPS:

取消 确定

3.4.3 对象存储

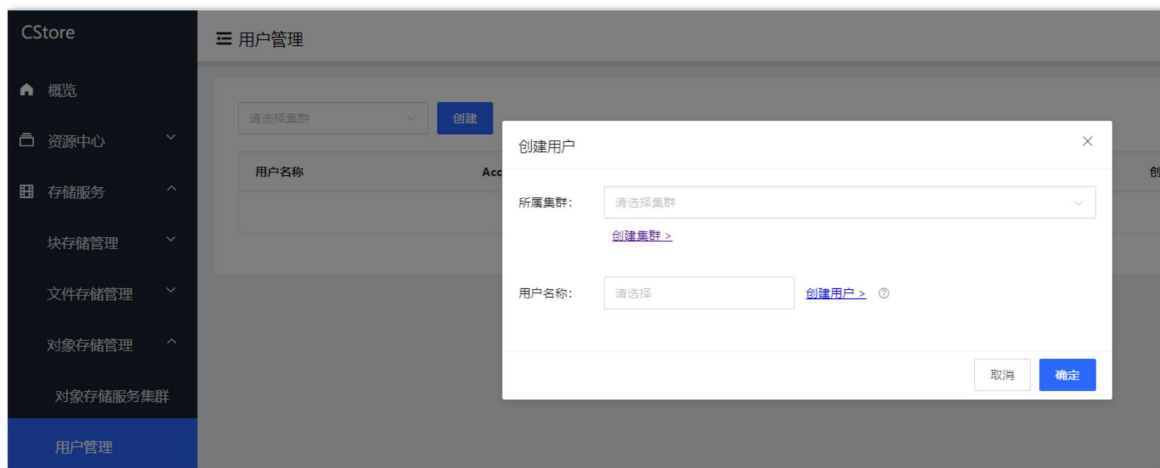
3.4.3.1. 对象存储集群

用户可基于创建的 ofs 集群，在基础上部署对象存储服务，需要有先创建对象存储集群，创建对象存储集群，需要有至少三个节点部署对象存储服务，保障服务的高可用，并针对此设置可访问的内网与公网域名；同时也支持对象存储集群的扩容



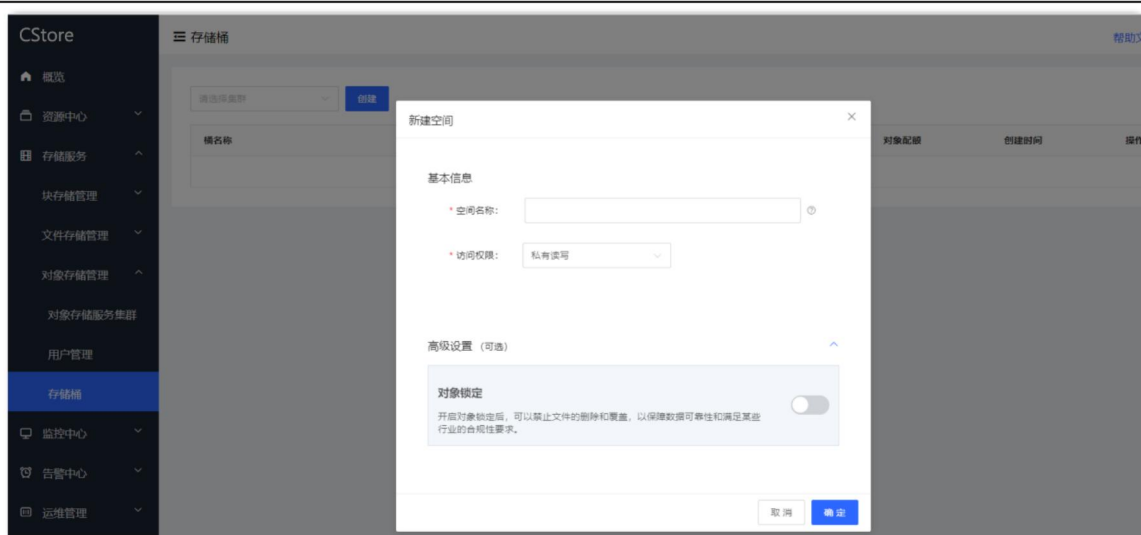
3.4.3.2. S3 用户管理

要创建 S3 用户，需要先创建可登录管控平台的用户，S3 用户创建成功后便具备可对对象存储桶以及文件操作的权限，创建用户后 Access Key ID&Access Key Secret 系统默认生成；同时可对用户进行启停的操作



3.4.3.3. 存储桶

S3 用户可针对创建的存储桶进行 object 的管理，上传下载等；

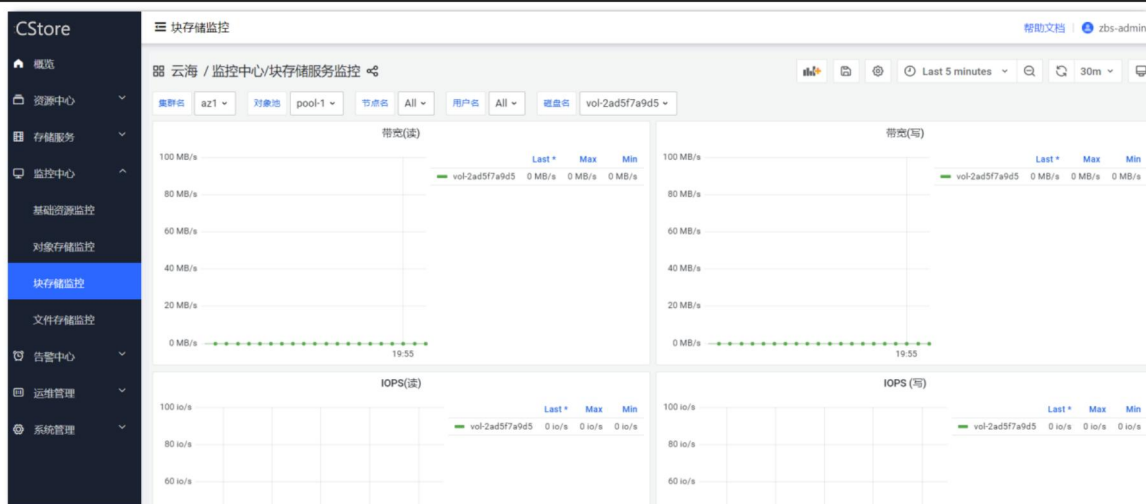


功能列表

类别	功能
空间管理	创建存储空间
	删除存储空间
权限管理	设置存储空间权限
	bucket policy 设置
Object 管理	上传文件
	下载文件
	搜索文件
	新建文件夹
	获取文件地址
	删除文件/文件夹
基础设置	生命周期管理
	访问日志
	静态网站托管
	标签
高级设置	对象锁定

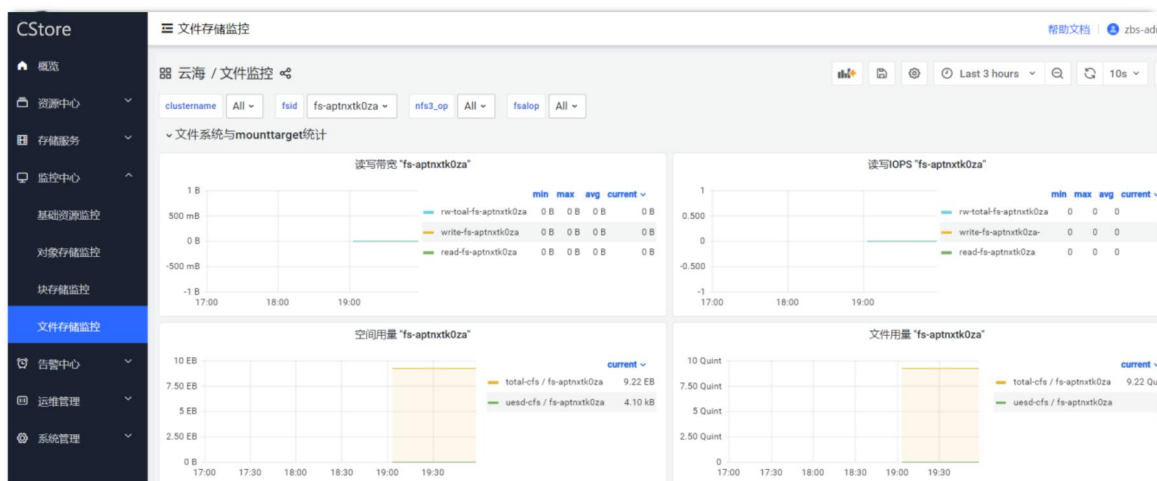
3.5 监控中心

用于监控集群、存储池、节点、文件存储、存储卷、对象存储的各项指标的状态



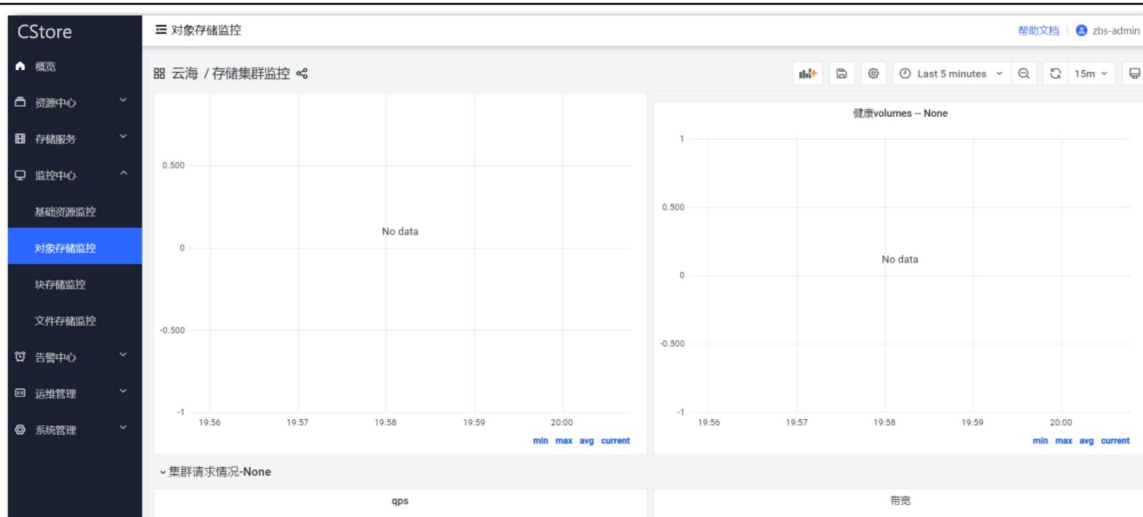
3.5.3 文件存储监控

文件监控支持从集群、存储池、节点、硬盘、文件系统等维度对读写带宽以及 IOPS、延时的监测



3.5.4 对象存储监控

对象存储可根据 S3 用户维度创建的存储桶以及集群、存储池等资源维度监测各个桶容量以及 ops、上下行带宽情况



3.6 告警中心

3.6.1 告警中心概述

告警中心分为两个模块：告警记录和告警设置模块。告警记录可以给用户一个全量的报警数据展示，数据将会落库永久保留，页面上支持各种指标维度的筛选查看和状态变更；告警设置模块作为告警记录的前置设置模块，可以对用户管理、告警策略等进行设置。其中在告警设置界面，有两个 tab 页面分别为告警策略和告警通知设置。其中告警策略放在第二项默认展示，告警通知设置放在第一项 tab。

3.6.2 告警设置

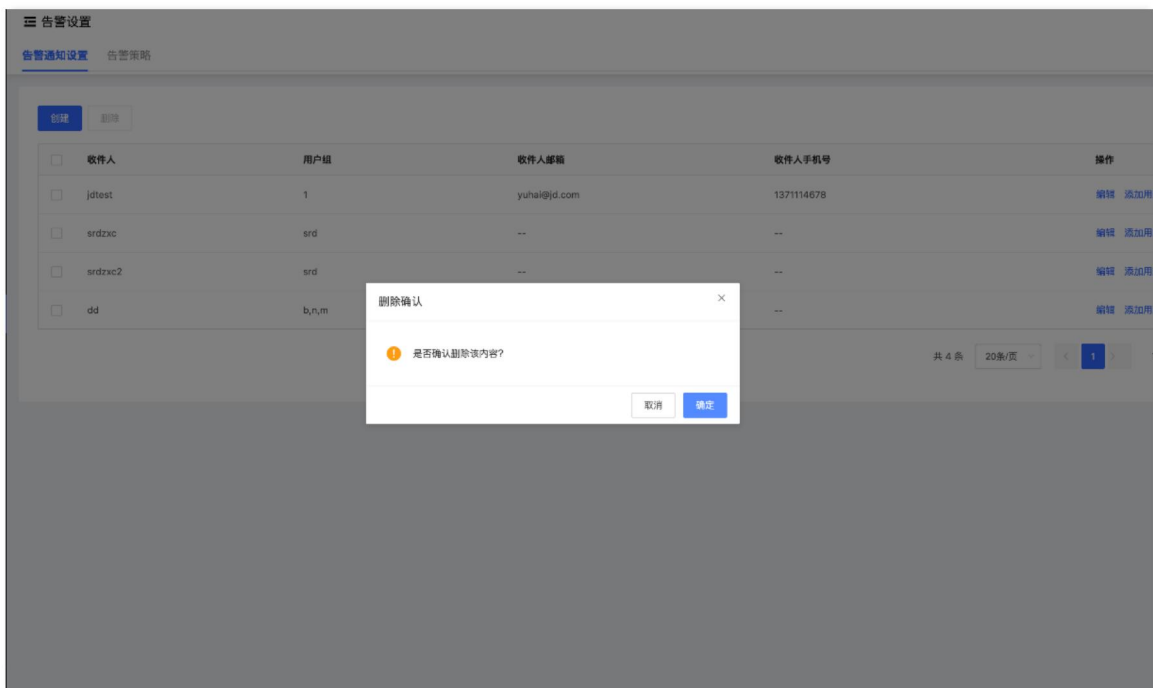
3.6.2.1. 告警通知设置

该页面功能分为收件人的创建、收件人的列表展示页和告警服务器设置三个核心的功能模块。该功能模块定义了告警通知的发送方和接收方的信息。



收件人

收件人列表页展示收件人姓名、用户组名称、邮箱、手机号字段，操作。按照创建时间降序排列，默认 20 条/页，可选 20、50、100 条/页。支持对收件人的批量选择进行删除操作。操作分为编辑和删除。



3.6.2.2. 告警服务器设置

告警服务器设置可以对告警的发送服务器进行配置设置。配置项目有下表几点：



字段名	是否必填
发送间隔	是
是否邮箱通知	通过滑块开关组件实现，当开关置开时出现以下配置项
邮箱服务器	是
端口号	是
用户名	是
密码	是
发件人邮箱	是

当第一次进行填写的时候，发送间隔默认为每 1 小时，是否邮件通知默认置是，出现邮件服务器的配置项，是否短信通知默认置否。发送间隔是一条针对所有告警策略的全局规则，描述的是向用户发送告警邮件的时间间隔，告警产生时会立即发送邮件，如告警未恢复正常，将按照发送间隔向告警接收邮箱发送告警邮件。

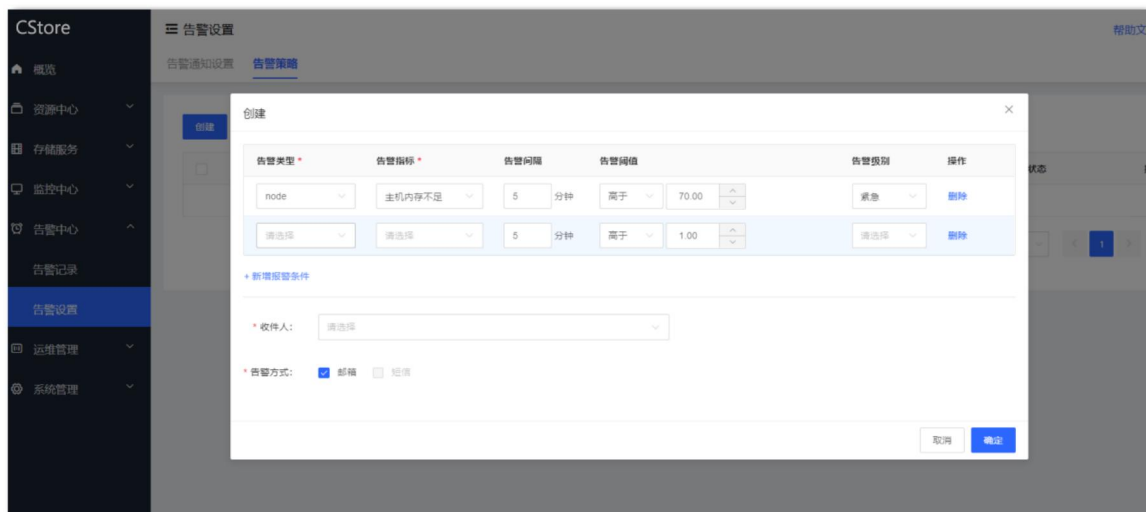
3.6.2.3. 告警策略

告警策略模块负责在告警过程中的规则模块，定义了什么样的告警信息需要通知到用户，告警信息分为哪些级别，告警的规则是什么，告警执行的方式和策略是

什么样的

告警策略

告警策略的报警条件分为定性和定量两类告警指标。定性指标是针对成功、失败之类可以直接定性的指标类型；定量的指标类型则是针对像容量等可以通过数字表征的报警指标，通常用高于，低于等作为告警的判断方式。一条报警条件由告警类型、告警指标、统计周期、告警阈值、持续周期、告警级别、操作几部分组成。操作只包含删除，删除时不需要通过弹窗二次确认，直接删除即可。其他字段如下表所示。在报警条件下可以新增报警条件，支持一次创建多条报警条件。



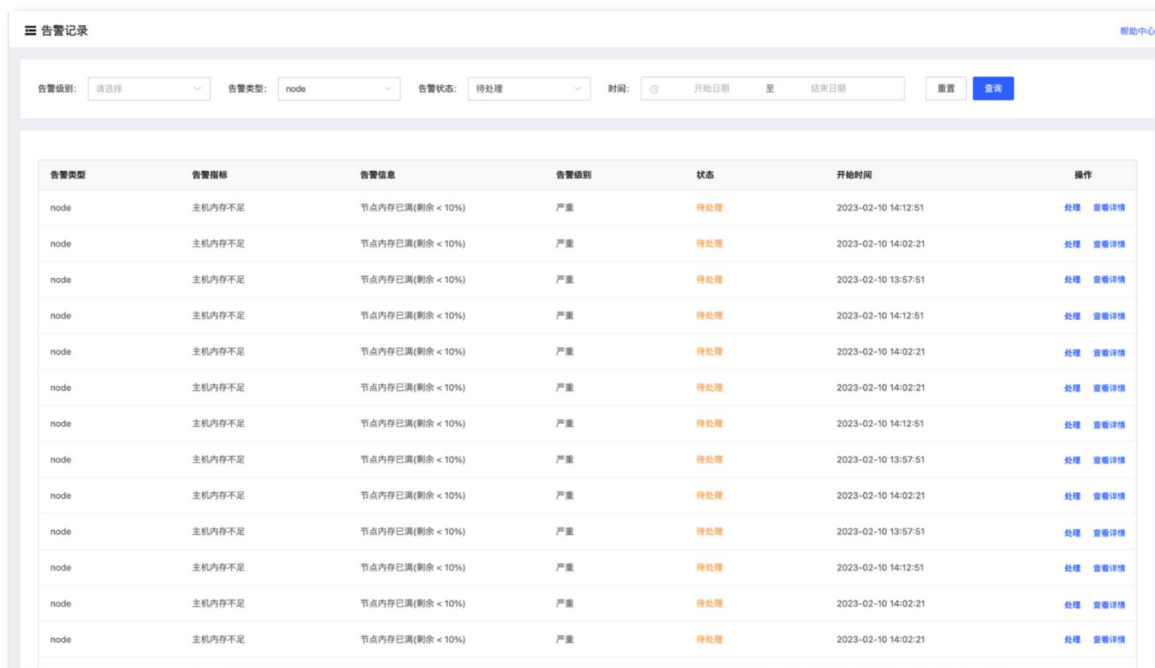
告警策略列表

告警策略列表页展示告警类型、告警指标、告警条件、告警阈值，当前阈值、告警级别、状态、操作。

告警策略							
告警通知设置 告警策略							
告警类型	告警指标	告警条件	告警阈值	当前阈值	告警级别	状态	操作
☐ node	主机内存不足	>	20%	10%	严重	开启	关闭 编辑 删除
☐ node	主机CPU未被充分利用	>	1000000%	20%	致命	开启	关闭 编辑 删除
☐ log	CFS_mdshLog.journal_space_left_warn	<	1%	0%	致命	开启	关闭 编辑 删除

3.6.3 告警记录

用户可查看系统产生的所有告警通知。可以进行多维度的筛选展示。告警记录列表页展示告警 ID、告警类型、告警指标、告警信息、告警级别、状态、告警时间，操作。



告警类型	告警指标	告警信息	告警级别	状态	开始时间	操作
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:12:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 13:57:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:12:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:12:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 13:57:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 13:57:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:12:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:12:51	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情
node	主机内存不足	节点内存已满(剩余 < 10%)	严重	待处理	2023-02-10 14:02:21	处理 查看详情

3.7 运维管理

3.7.1 日志管理

3.7.1.1. 日志列表

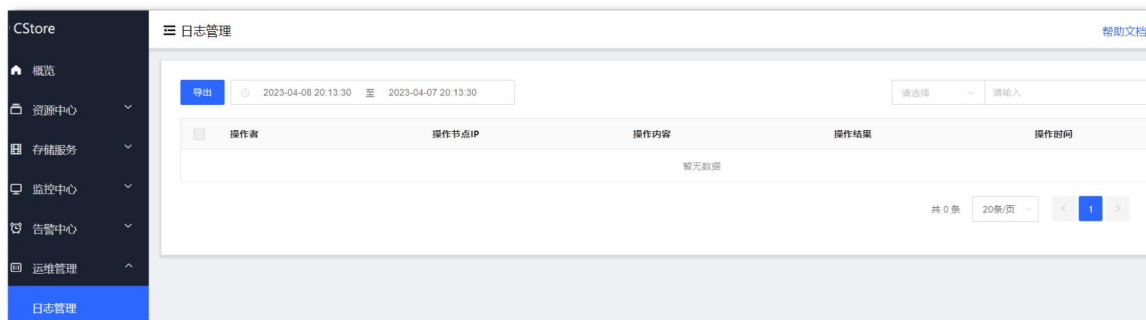
操作者：登陆者账号

操作节点 IP：针对哪个节点做的任务操作

操作内容：做的具体操作名称，如针对块存储查询资源列表，或者创建块存储集群等

操作结果：查询或者操作配置下发成本与否，参数值为 成功 or 失败

操作时间：实施具体操作内容的开始时间



3.7.1.2. 日志导出&按时间可查询

时间筛选：可以在管控平台上基于时间筛选查询的日志范围，时间具体到分钟级别

导出：根据日志列表可导出 excel 表形式

3.8 系统设置

3.8.1 权限管理

角色	权限范围
超级管理员	对存储设备有完全的控制权限，可以创建和查看各类角色的用户
管理员	拥有除权限管理&license 管理、SSL 证书管理以外的所有权限。权限包含概览、资源中心、存储服务、监控中心、告警中心、运维管理、系统管理（权限管理）
只读用户	拥有查询信息和修改用户自身密码的权限。
块服务用户	块服务用户，权限包含存储服务-块存储管理（块存储服务集群、块存储卷、卷快照）、监控中心（基础监控、对象）、告警中心（告警设置、告警记录）、运维管理（日志管理、配置管理、在线升级）
文件服务用户	文件服务用户，权限包含存储服务-文件存储管理（文件系统服务集群、文件网关、文件系统）、监控中心（基础监控、对象）、

	告警中心（告警设置、告警记录）、运维管理（日志管理、配置管理、在线升级）
对象服务用户	对象服务用户，权限包含存储服务-对象存储管理（对象存储服务集群、用户管理、存储桶）、监控中心（基础监控、对象）、告警中心（告警设置、告警记录）、运维管理（日志管理、配置管理、在线升级）
设备管理用户	拥有除存储服务以及系统设置外的所有权权限

4 产品规格表

产品核心功能及功能介绍

产品核心功能	功能介绍
概览	从集群维度展示集群的资源状态，如存储池个数、节点个数、硬盘个数，资源的利用率以及集群的健康状态，并提供性能指标如 IOPS、带宽的走势图，通经验正常值的对比，突出异常时间段，更清晰告知集群的健康状态，同时提供实时告警信息，划分危险、严重、警告、提示等四个等级的告警信息以及当前发生未处理的告警信息流，方便运维快速捕获重要信息，提升工作效率；
存储资源管理	1、支持存储集群管理，可自主自助独立部署集群的资源，横向自助化部署不同的存储应用，纵向从集群、存储池、节点、硬盘由客户根据业务体量自主部署，提供存储容量以及使用量、集群整体的 IOPS、带宽等性能指标，监测集群的状态是否正常并提供集群的告警信息；提供集群拓扑关系视图，可视化展示集群部署架构； 2、支持存储池管理，查看和管理集群中的各个存储池，同一存储池可以根据不同的节点不同硬盘创建不同的存储服务，缩小资源池管理的最小规模以及提升资源的利用率；支持存储池资源的弹性伸缩，提供副本、纠删码冗余保护策略； 3、支持存储节点，一键获取节点资源，可视化管理和查看节点信息，自动监控节点状态，预警节点异常情况并提供自助修复功能，监控节点资源的利用率以及合理性 4、支持硬盘管理，系统会自动设备插好的硬盘，包括硬盘状态、硬盘使用率、存储属性、容量、数量，提供硬盘的统一监控和管理列表

对象存储服务	1、支持 S3 用户管理，多租户特性可以使用对象存储服务，每个用户都有一对 AccessKeyId 与 AccessKeySecret，是用户访问对象存储的身份标识，并确保这些数据之间相互隔离； 2、支持桶管理，S3 用户可创建桶资源，可以设置存储空间的访问权限、防盗链规则等，每个桶的基本信息都会具备空间名称、当前的访问权限、创建时间、内网访问域名和外网访问域名；支持桶的创建和删除，Bucket 标签管理； 3、支持访问权限控制，支持创建存储空间时配置 ACL，也可以在任意时间修改 ACL，同时支持 Bucket Policy 功能，授权其他用户访问您的 OSS 资源，同时也支持用户角色的操作授权，通过不同的角色授权控制存储服务的访问权限 4、支持对象的管理，支持文件的上传、下载、删除以及分片上传 5、支持配额管理，管理员需要按需求为用户、桶为单位合理分配存储空间或对象数量 6、支持数据安全，限制访问来源，对访问来源设置黑白名单、referer 防盗链等，避免 OSS 资源被其他人盗用； 7、支持数据管理，数据同步，跨不同的 Bucket 自动、异步复制文件，它会将源 Bucket 中的对象的改动（除删除操作）同步到目标 Bucket；历史数据同步，使用历史数据同步可以将原 Bucket 的现有文件一次性复制到目标 Bucket；静态网站托管设置，存储空间配置成静态网站托管模式，并通过存储空间域名访问该静态网页 8、支持生命周期管理，可自动删除过期的 Object 及分片 9、支持 EC，副本冗余保护策略
块存储服务	1、支持块存储卷的管理，支持自助创建、删除存储卷，并批量设置卷的大小，支持挂载块存储服务；支持块存储集群的扩容 2、支持 QoS 设置，实现对存储卷的 IOPS、带宽的实时控制； 3、支持 EC，副本冗余保护策略 4、支持卷快照，自定义创建快照策略，可以为云盘周期性地创建快照备份数据，支持共享快照，将已经创建好的 数据

	盘快照 共享给其他用户使用;支持复制快照;支持快照回滚,当需要版本回退或将数据回滚到过去某一时刻,可以使用快照回滚;快照一致性级别,快照所保存数据完整性级别;
文件存储服务	1、支持文件存储集群的创建,资源的扩容、缩容,管理挂载目标 2、支持目录配额,可根据用户定义的容量标准来统计或限制存储的使用情况,包括添加配额、编辑配额和删除配额等 3、支持 POSIX、NFS v4.1 和 NFS v4.0 协议 4、服务质量,可针对文件的读写性能以及延迟等提供实时数据监控 5、数据定制,采用全用户态数据落盘机制,无内核组件的额外开销(内核文件系统开销,内核块设备驱动开销); 6、支持共享文件,一写多读架构,数据共享存储,多客户端读写一致性机制 7、高容错机制,CStore 高容错分布式协议,数据可写在任意磁盘上,只要集群中存在 3 个可用磁盘即可保障服务可用性,数据三副本强一致写入,且不会因故障降级为两副本写入 保障数据的可靠性; 无 Leader 非仲裁协议,故障时无选举时间,不会因 Leader 发生故障时,选举新的 Leader 过程会有 10-20s 左右,敏感业务会受损,甚至数据有丢失的风险 8、支持分级存储,支持通过 SSD、HDD 等不同存储介质构建分级存储能力;
监控分析	1、支持对集群、存储池、节点、硬盘维度的监控,包括资源总量/使用量,资源利用率,数据的健康状态,以及 CPU、内存的使用量,IOPS、带宽、延时等性能走势以及预警相关异常情况,盘故障:坏盘、慢盘,系统定时检测,发现故障、检测故障、隔离故障,恢复故障等提供自助修复功能 2、支持存储服务层面的监控,如对象存储服务的数据监控状态,上传下载带宽、请求数以及延时等性能指标,同时支持存储容量的使用率以及预警资源规划不合理信息;支持块

	存储服务、文件存储服务的数据监控状态，资源利用率，CPU&内存利用率、总容量以及使用量信息； 3、支持网络的监控，包括网络延时、丢包、速率、带宽、请求并发等避免网络超时、中断、震荡影响业务
告警管理	1、支持告警信息的查看和管理，包括告警类型、告警指标以及信息描述、告警级别、开始时间、持续时间、告警状态等 2、支持告警规则的设置，设置不同的告警类型，根据对应的告警指标在一定的条件下配置告警阈值以及告警级别 3、支持告警通知的设置，用户将指定类型的告警信息发送到需要通知人员的邮箱或者手机短信，以此增加告警渠道。
运维管理	1、支持日志管理和查看，支持系统触发以及操作人员行为触发产生的日志进行记录，并支持日志的导出，便于排错、审计和跟踪； 2、支持配置管理，用户可以管理系统配置信息，包括 DNS 设置、系统时间等设置；
系统管理	支持用户权限管理，根据不同的用户角色分配不同的菜单以及操作权限，通过系统管理可以管理用户的访问权限，查看和修改资源基本配置以及存储服务信息等功能。