存储方案技术建议书

曙光信息产业（北京）有限公司

2024年

目录

[1.1. 整体方案 2](#_Toc164165008)

[1.1.1. 异构融合 2](#_Toc164165009)

[1.1.2. 极致性能 2](#_Toc164165010)

[1.1.3. 持续服务 2](#_Toc164165011)

[1.2. 存储节点设计 3](#_Toc164165012)

[1.3. 存储网络设计 3](#_Toc164165013)

[1.4. 存储软件设计 7](#_Toc164165014)

[1.4.1. 提供单一命名空间，多协议并发访问 8](#_Toc164165015)

[1.4.2. 性能线性增长 8](#_Toc164165016)

[1.4.3. 稳定可靠全冗余架构 9](#_Toc164165017)

[1.4.4. 副本/EC纠删码数据冗余 9](#_Toc164165018)

[1.4.5. 故障域隔离 12](#_Toc164165019)

[1.4.6. SSD Cache 13](#_Toc164165020)

[1.4.7. 统一管理，简易运维 14](#_Toc164165021)

[1.5. ParaStor分布式存储产品规格 15](#_Toc164165022)

[2. 存储系统集成方案 18](#_Toc164165023)

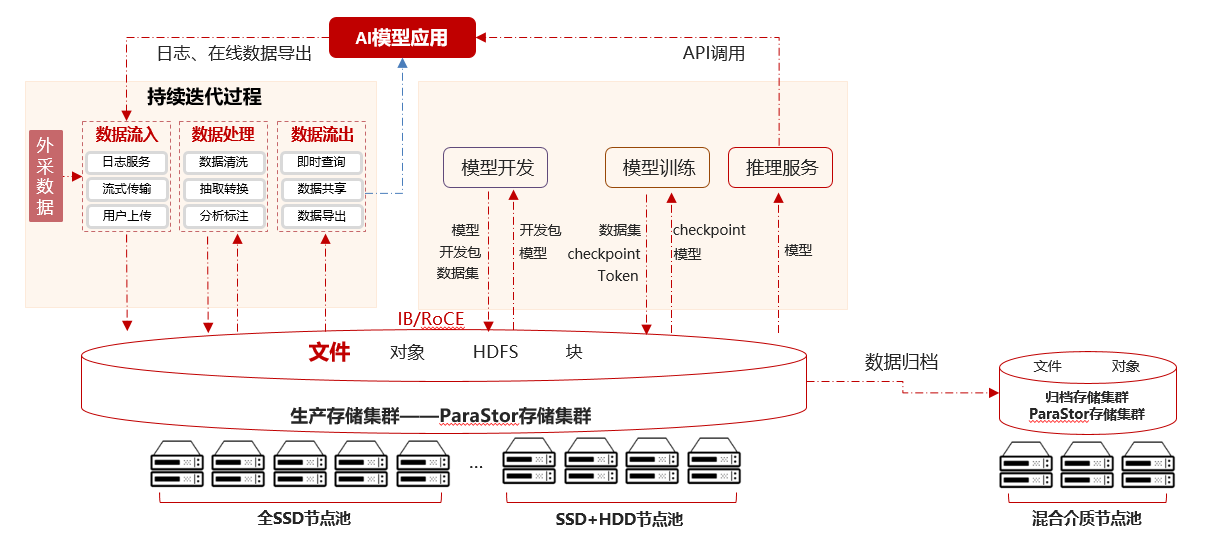
[2.1. 项目建设原则 18](#_Toc164165024)

[2.2. 项目建设目标 18](#_Toc164165025)

[2.3. 存储系统配置清单 19](#_Toc164165026)

* 1. 整体方案

曙光面向行业客户的AI应用开发场景，推出以ParaStor高性能分布式存储为核心的AI存储解决方案。



* + 1. 异构融合

ParaStor通过一个数据平台、一套存储系统完成异构数据池化，在一个集群内支持多个存储池，一个存储池内可同时支持文本、图片、音频、视频等多种类型数据存储，同一份数据又可以被前端不同业务场景以文件、对象、大数据（当前仍有部分客户的训练数据集预处理是在大数据平台等完成）以及视频的存储方式并行访问。

* + 1. 极致性能

ParaStor存储系统提供持续的低访问延迟与高数据吞吐带宽。在AI应用开发过程中，需要频繁从数据集读取Tokens。但是每个Token通常只有几个字节大小，需要存储系统的实时高并发的海量小IO能力，需要极低的延迟（如百毫秒级读写延时）来保障性能；此外，在存储模型Checkpoint时，需要高带宽来支撑数据快速写入。Parastor存储通过多级缓存、XDS数据加速等技术，提高数据Shuffle、读batch阶段的元数据（小文件）与数据访问效率，提供极高数据写带宽缩短checkpoint耗时，提升算力效率。

* + 1. 持续服务

ParaStor存储系统通过部件、节点、系统、方案等多级可靠性机制，保障存储集群在极端环境下的稳定运行，提供企业级数据保护特性。根据不同行业AI业务场景需求和客户对于核心重要数据的保护需求，ParaStor存储方案也能够通过全闪、混闪、带库、光盘四种介质，支持文件、对象、大数据、视频、块协议，兼顾性能&成本，满足大容量、多协议共享，冷数据的长期保存和归档。

* 1. 存储节点设计

ParaStor全闪存储集群采用全对称方式进行部署，每个存储节点可以同时提供集群管理、元数据和数据服务。

管理服务主要管理存储系统的磁盘、网络、存储节点以及客户端访问存储的授权等，提供存储集群管理软件，包括存储性能监控、故障告警、系统调试优化、物理/逻辑磁盘故障检测和存储使用报告等功能，任意2个管理服务节点故障不影响高性能计算机系统内节点对数据的读写。

元数据服务主要实现存储系统元数据的统一管理、存储，并承担前端计算节点数据访问所触发的元数据访问需求，每个存储节点配置1~2块≥3.84TB NVMe SSD作为元数据存储空间，元数据分布式存放，采用3副本冗余机制，元数据访问无瓶颈，任意2个存储节点故障不影响高性能计算机系统内节点对数据的读写。

数据服务主要用来管理和池化底层硬件平台的存储资源，与元数据服务一起对前端计算节点提供存储空间和存储访问带宽。数据节点对存储容量、存储密度、计算性能以及I/O通道均有较高要求。数据采用副本或N+M:b纠删码冗余机制，即每个数据条带包含M份校验数据，单个节点分区内任意b个数据节点故障不影响高性能计算机系统内节点对数据的读写。

2U24全闪数据节点基于高性能64位处理器平台开发的一款高性能双路服务器，每节点配置≥512GB Cache，2个1Gb管理网络接口，配置4个100Gb/200Gb

数据网路接口，可提供≥100TB的NVMe SSD存储容量，1+1冗余电源，内嵌高性能数据存取引擎，用于并行处理所有客户端的数据访问请求，内嵌高性能数据恢复引擎，支持节点间replication（多副本）/N+M（纠删码）数据冗余方式。

* 1. 存储网络设计

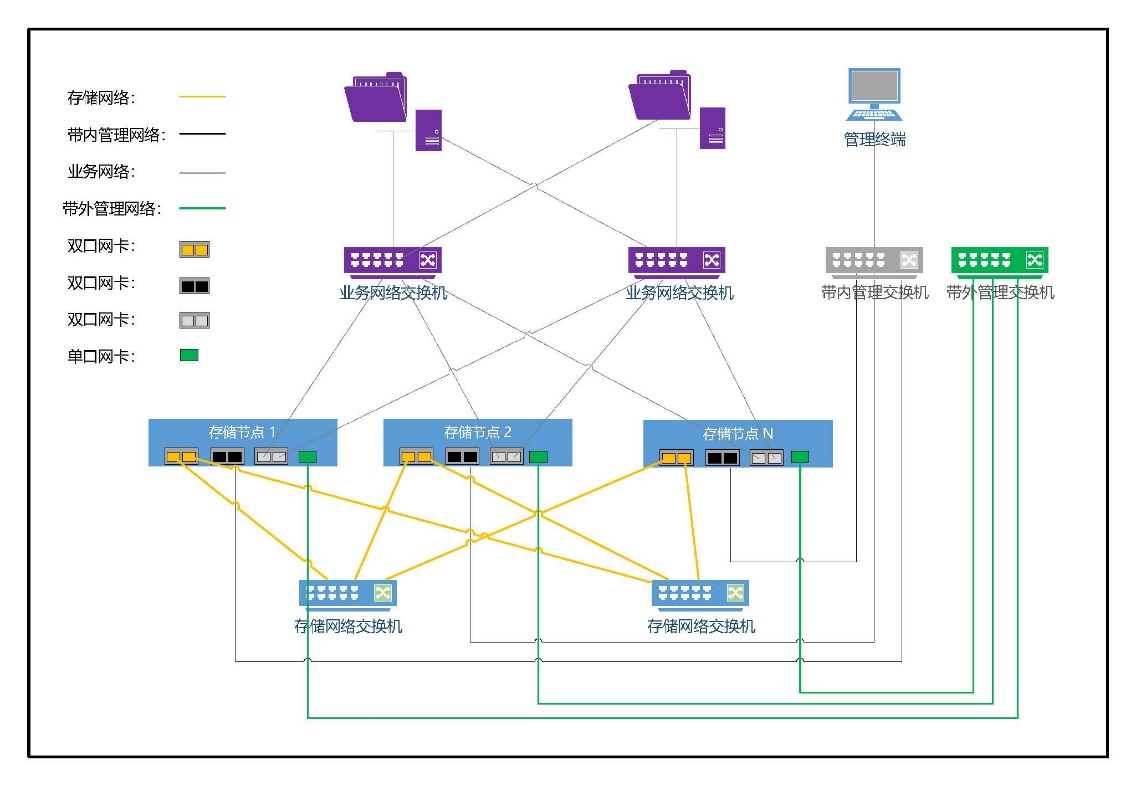
ParaStor全闪分布式统一存储的网络分为4个部分：带内管理网络、带外管理网络、存储网络及业务网络。

带内管理网络：用来同步系统中所有硬件节点的配置信息、并基于策略抓取或推送实时监控所需的硬件和进程等状态信息，数据传输量不大，通常千兆网络即可，配置千兆或者万兆接口。

带外管理网络：提供了一种不依赖于服务器的处理器、BIOS或是OS来工作的无代理管理系统，可以更加独立的进行硬件组件运作状况的监测以及紧急事件的处理，一般主板集成IPMI接口。

存储网络（又称后端网络）：是指存储节点间相互通信的网络。存储内部数据通信均由数据网络提供传输通道，也是ParaStor存储私有POSIX协议的访问链路，配置100Gb/200Gb IB或以太网接口。

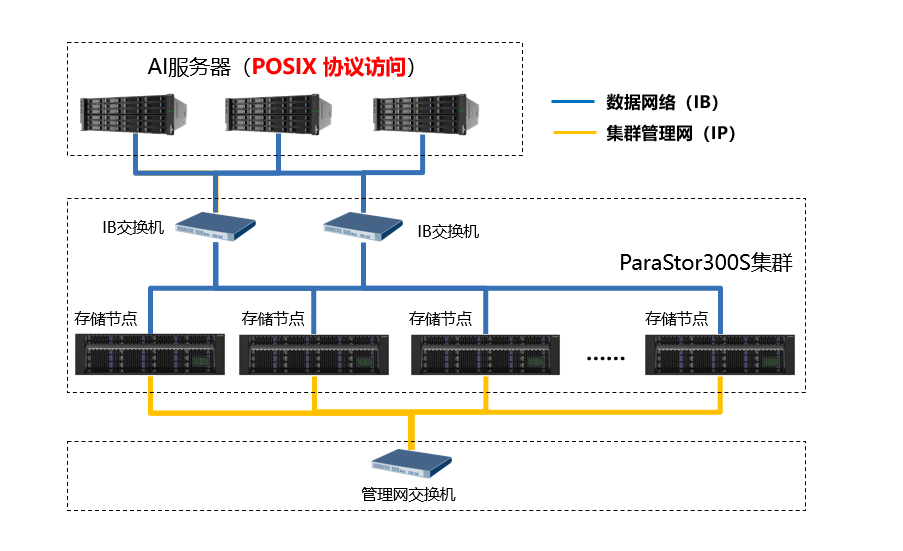
业务网络（又称前端网络）：是指存储系统与外部客户端进行协议通信的网络，一般用于计算节点通过NAS或S3协议访问存储，配置10/25/100Gb/200Gb 以太网接口。



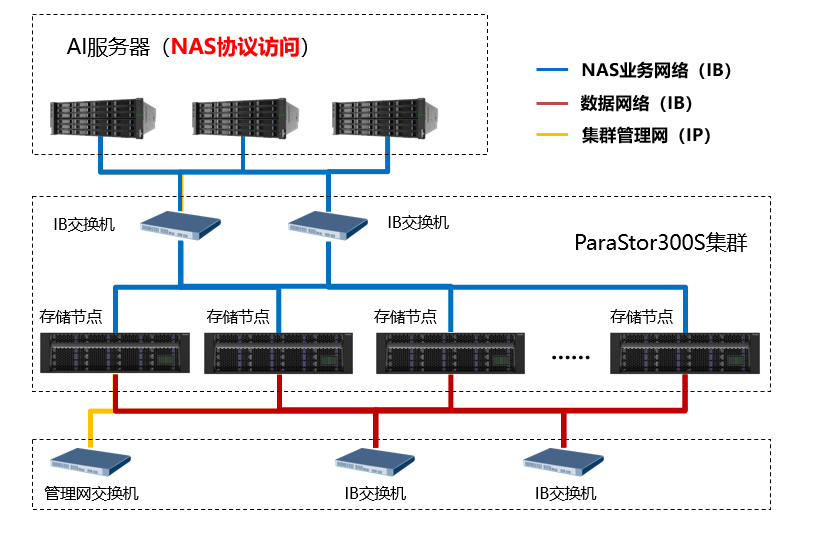
AI业务场景开发过程中，面对复杂的工作负载，要极速处理高分辨率模拟、超大型数据集和高并行算法，就需要搭建一个高性能无损网络。推荐采用POSIX协议+IB高速网络的组网方式。InfiniBand作为可完全卸载计算资源的网络计算平台，可提供巨大性能，在降低成本和复杂性的同时在高性能计算(HPC)、AI和超大规模云基础设施中带来超强性能，在AI应用开发与训练场景中广泛部署。

ParaStor300S的网络分为数据网络和业务网络，数据网络是存储集群内部通信网络。

通过POSIX协议为AI服务器提供数据访问接口时，由数据网络提供数据传输（推荐POSIX协议+分布式存储数据组网方式）



通过集群NAS协议为AI服务器提供访问接口时，由NAS业务网络为客户提供数据传输。



进行AI应用开发时，面对复杂的工作负载，要极速处理高分辨率模拟、超大型数据集和高并行算法，就需要搭建一个高性能无损网络。InfiniBand 作为可完全卸载计算资源的网络计算平台，可提供巨大性能，在降低成本和复杂性的同时在高性能计算 (HPC)、AI 和超大规模云基础设施中带来超强性能。InfiniBand是一个用于高性能计算的网络通信标准，具有极高吞吐量和极低延迟，在AI应用开发与训练场景中广泛部署。InfiniBand 具有如下核心特性：

* 简单网络管理
* 领先的高带宽性能
* 最大化提升CPU利用率，加速计算

曙光AI存储解决方案推荐存储组网使用X台IB交换机连接到存储集群，存储集群内的每个存储节点配置X个100Gb/200Gb IB网口，分别连接到X台IB交换机。对于集群存储内的每个存储节点而言，规划两条连接到存储网的IB链路，可保证网络链路冗余。

集群存储连接的X台IB交换机在大型AI计算中心组网中，担任Leaf节点交换机角色。

根据可用空间XX PB的需求，共配置XX个存储节点，需要XX台IB交换机，每台IB交换机提供XX个100Gb/200Gb接口。

根据英伟达DGX SuperPOD最佳实践，DGX SuperPOD每节点IO带宽约为40GB/s，每20个DGX SuperPOD划分为一个SU，此时需要Leaf节点交换机为4台，Spine节点交换机为2台。

备注：ParaStor300S网络支持1GbE/10GbE/56Gb IB/100Gb IB/200Gb IB，项目中根据实际需求进行配置。

* 1. 存储软件设计

ParaStor全闪分布式统一存储（以下简称ParaStor存储）是曙光存储公司依托多年的存储开发经验并根据存储产品应用现状和存储技术未来发展趋势，推出的新一代高端全闪存储产品，能够满足大中型企业用户对海量数据存储、高扩展性、高可靠性、高性能和方便管理的要求。提供了文件、块、对象、大数据多存储服务。以业界领先的性能、多种效率提升机制为支撑，为用户提供了高性能、全方位的存储解决方案，使用户投资收益比最大化，能够满足自动驾驶、自然语言处理、人脸识别、智能制造视觉检测等AI应用场景。

ParaStor存储整体逻辑架构如下图所示：



核心模块主要包括：

（1）智能本地磁盘管理：全自主研发的本地磁盘管理系统，负责管理物理介质并通过抽象简化向上层调用者提供通用的访问接口；

（2）数据管理：有效管理元数据、数据，整合存储资源，对外提供统一的存储空间；

（3）数据保护：对数据提供跨磁盘、跨节点、不同级别、不同维度的冗余保护，保障系统的高可用性；

（4）协议管理：并行处理所有客户端的数据访问请求；

（5）WebUI管理：提供统一的控制管理界面，用以管理整个存储系统。

* + 1. 提供单一命名空间，多协议并发访问

全局单一命名空间可以为用户构建EB级容量的虚拟存储池，所有客户端可以在同一个命名空间下共享访问所有的数据，可以简化海量数据管理的复杂性。

ParaStor300S架构中，存储节点构成节点池，同一节点池划分为多个存储池，将物理资源进行统一管理和池化，在存储池的基础上再次划分为不同存储卷（文件系统），实现多层级细粒度的资源管理方式。

ParaStor300S对外支持多种存储访问协议：

1. POSIX：提供符合标准POSIX语义的文件系统接口，是ParaStor300S提供的内核态的Linux私有客户端访问协议，支持基于Infiniband的RDMA协议，一般适用于高速网络的高性能计算应用领域。
2. 集群NAS：通用的网络文件系统接口，支持NFS、CIFS/SMB、FTP；业务客户端无需安装任何代理程序，按照标准协议挂载后，即可访问；使用简单便捷，适用于绝大多数的非结构化数据共享的场景。
3. S3：兼容Amazon S3访问接口，适用于对象接口应用场景。

同一集群可以同时提供POSIX、集群NAS、S3对象接口，满足前端不同的业务应用。

* + 1. 性能线性增长

ParaStor300S是一款基于对象的分布式集群系统，数据切片成固定大小的对象，条带化地分布到多个存储节点的多个磁盘上，提供并发的传输通道实现并行的I/O访问，有效地提高了存储系统的读写带宽和IOPS。同时系统聚合性能随着节点规模的增加而线性增长，消除传统的Scale-up纵向扩展存储架构的性能局限性。除此之外，ParaStor300S通过多项技术进一步提升性能：

1. 内部高速互联：集群内部网络可以采用56Gb/100Gb Infiniband高速网络，基于RDMA协议，获得更高的带宽、更低的延迟。
2. 内核态POSIX协议：在高性能应用领域中，可以采用内核态的私有客户端协议，并发访问能力更强。
3. SSD Cache：存储节点配置一定数量的SSD，作为本地节点的二级读缓存，提升数据随机访问性能。
   * 1. 稳定可靠全冗余架构

可靠性是存储系统的基本要求，尤其对于中大规模的分布式系统，稳健性更是重中之重。ParaStor300S采用冗余架构设计，从部件、网络、节点、数据冗余等各个层面保证数据的安全性及系统的稳健性。

1. 部件可靠性

ParaStor300S基于曙光通用高性能服务器构建，选用的是曙光当前阶段最为成熟、稳定且市占率很高的硬件产品。CPU、内存、电源、风扇等主要部件均是冗余配置，单个部件的故障不会影响整个节点的正常运行。同时ParaStor300S系统的WebUI管理界面集成了部件的监控功能，部件发生故障，将给出准确的告警信息，便于用户及时定位维护。

1. 网络可靠性

存储系统采用冗余网络拓扑，任一链路发生故障，均不影响系统的正常服务。

1. 节点可靠性

所有节点并行提供服务，在冗余规则之内发生节点故障，不影响存储服务。

1. 数据冗余

ParaStor300S支持多副本、EC纠删码两种数据冗余方式，可以针对节点池设置不同的冗余策略，保障数据一致性。

* + 1. 副本/EC纠删码数据冗余

用户数据切片成固定大小的对象，存储到不同节点的不同磁盘上，为了保证数据安全，根据原始数据通过一定的算法得到特定的校验（冗余）数据，一并存在磁盘上；当任一数据块故障时，可以通过算法反校验得以恢复，进而保证数据的完整性。ParaStor300S支持多副本、EC纠删码两种数据冗余方式。

1. 副本（Replication）：文件切分成对象后，由原始数据拷贝得到冗余数据，对于M副本，冗余数据为（M-1）份。由于仅是数据的拷贝，系统开销较小，但空间利用率较低，对于M副本来说，空间利用率为1/M，如双副本时，空间利用率为50%，3副本时，空间利用率为33.33%。采用M副本冗余机制时，存储系统允许任意（M-1）块数据盘故障，或者允许任意（M-1）个数据节点故障，如双副本时，存储系统允许任意故障1块数据盘或1个节点。ParaStor300S支持2~4副本，即M=2~4。
2. EC纠删码：EC（Erasure Code）是一种技术，将N份原始数据经过特定算法得到M份校验数据，并能通过N+M份中的任意N份数据，还原得到原始数据。具体算法包含Encode和Decode两个过程，将原始的N份数据变为N+M份称为Encode，由N+M份数据中的任意N份得到原始数据称为Decode，即允许任意故障M份数据。Erasure Code可以认为是RAID的通式，任何RAID都可以转换为特定的Erasure Code。在传统的RAID中，仅支持少量的磁盘分布，当系统中存在多个分发点和多节点时，RAID将无法满足需求。比如RAID5只支持一个盘失效，即使是RAID6也仅支持两个盘失效，Erasure Code算法可以支持多块盘同时失效，因此可以满足客户对更高可靠性的要求。

ParaStor300S系统的纠删码算法支持N+M:b冗余方式，其中M表示存储系统允许任意故障的磁盘数，b表示存储系统允许任意故障的数据节点数，如8+2:1表示存储系统任意故障2块磁盘或1个数据节点。

N+M:b参数说明：N表示原始数据的份数，M表示校验数据的份数，这（N+M）份数据中，最多允许同时故障M份数据。ParaStor300S的选盘机制尽量将同一段中的对象分布到不同节点的不同磁盘上，因此M表示存储系统允许故障的最大磁盘数。b表示存储系统允许故障的最大节点（OStor，下同）数。需要特别注意的是，N+M:b表示存储系统允许任意故障的最大磁盘数是M，或者允许任意故障的最大节点数是b，这两个任意性无法同时满足；同一节点上多于M块磁盘甚至所有磁盘发生故障，相当于节点故障，与任意故障的磁盘数无关联。

N+M:b配比有多种数据分布方式，满足此配比的最小节点数为[b(N+M)/M]，[]表示向上取整。

如满足4+2:1所需的最小节点数为3，见图3-2：



* + - * 1. 满足4+2:1数据分布的节点示意图

按照图3-2的数据分布，任意故障一个节点，最多有两个对象故障，数据完整性得以保障，可以正常读取；但是此时数据处于降级状态，为了避免再次故障导致数据丢失，需要尽快进行数据重建。按照最优的选盘机制，修复的2个对象尽量放置到不同的节点上，此时系统并没有多余的节点可用。在实际应用中，仅配置满足数据冗余的最小节点数是不够的，为了保证b个节点故障时可以进行最优的数据修复，至少需要配置[b(N+M)/M+b]个节点，[]表示向上取整。

如满足4+2:1可修复的最小节点数是4，如图3-3所示：



* + - * 1. 满足4+2:1数据修复的节点示意图

M与b最大支持4，常见的N+M:b配比方式如表3-1所示：

|  |  |  |  |
| --- | --- | --- | --- |
| N+M:b | 满足数据分布的最小节点数b(N+M)/M | 满足最优的数据修复的最小节点数[b(N+M)/M]+b | 空间利用率N/(N+M) |
| 4+2:1 | 3 | 4 | 66.7% |
| 6+2:1 | 4 | 5 | 75% |
| 4+2:2 | 6 | 8 | 66.7% |
| 8+2:1 | 5 | 6 | 80% |
| 8+2:2 | 10 | 12 | 80% |
| 10+2:1 | 6 | 7 | 83.3% |
| 10+2:2 | 12 | 14 | 83.3% |
| 16+1:1 | 17 | 18 | 94.1% |
| 16+4:2 | 10 | 12 | 80% |
| 16+4:4 | 20 | 24 | 80% |

N+M:b推荐配置

在实际应用中，为了保证节点故障后，可以快速修复，建议按照满足数据修复的节点数量设置冗余方式。

* + 1. 故障域隔离

N+M:b纠删码冗余中，M、b最大数值均是4，即存储系统允许任意故障的最大磁盘数是4，或者允许任意故障的最大节点数是4。在系统规模较大时，磁盘动辄成百上千块，节点多则几十台，冗余度为4的存储系统远远无法满足客户需求，为此ParaStor300S存储系统引入故障域的概念，在副本/纠删码的基础上，通过磁盘分组、节点分区进一步提高系统的冗余度。

1. 节点分区

若干节点构成不同的分区，之间的物理节点相互隔离，各个分区分别设置副本或EC纠删码冗余方式，某一分区的节点故障不会影响其他存储池的服务状态。

如图3-4所示，存储系统共有18个存储节点，若采用8+2:1冗余方式，仅允许任意故障1个节点。若采用节点分区，可以将18个节点分成3个域，每一个域包括6个节点。每个分区均可以设置为8+2:1，此时每个分区均可以任意故障1个节点，整个存储系统最多允许故障3个节点而不发生数据丢失。若某一分区同时故障1个以上的节点，则该分区无法正常提供服务，但是其余的分区不受任何影响。



* + - * 1. 节点分区

分区通过缩小节点的故障域，进一步提升系统的可靠性。

1. 磁盘分组

以特定节点分区说明为例，多个分区时情况类似。

分组指的是将所有存储节点（OStor）上的磁盘划分为若干池（Pool），磁盘池之间相互隔离，遵从分区设置的N+M:b冗余方式。某一分组的磁盘故障不影响其他分组的服务状态。

如图3-5所示，6个OStor组成一个分区，若采用8+2:1冗余方式，仅允许任意故障2块磁盘。若采用磁盘分组，每个节点的若干（图中为2块）磁盘组成一个分组，图中不同颜色表明一个磁盘分组，每一分组内包含12块磁盘。由于分区采用的是8+2:1冗余方式，每个磁盘遵从同样的冗余度，即每个分组均可以任意故障2块磁盘，整个分区最多允许故障8块磁盘而不发生数据丢失。选盘算法规定同一文件不会跨分组存放。若某一分组同时故障2个以上的磁盘，只会影响相应的文件访问，其余文件不受影响。



* + - * 1. 磁盘分组

分组通过缩小磁盘的故障域，进一步提升系统的可靠性，ParaStor300S默认设置磁盘分组。

* + 1. SSD Cache

SSD相较于机械硬盘，带宽及响应时间有着明显的优势，并且容量远大于普通内存。将SSD作为缓存资源，可以显著降低存储系统的响应时间，有效提高数据的访问频率。

ParaStor300S系统中，存储节点配置一定数量的SSD，组成介于HDD与内存之间的二级本地缓存池。

当SSD作为读缓存时，利用SSD盘对随机小文件读取速度快的特点，通过智能算法将访问频度高的随机小文件热点数据存放到SSD上，应用程序再次访问该数据时，可以直接从SSD上获取。由于SSD盘的数据读取速度远远高于机械硬盘，因此可以显著缩短热点数据的响应时间，从而提升系统的性能。

* + 1. 统一管理，简易运维

提供基于Web的图形化监控管理界面，提供系统配置管理、监控告警等功能，直观易懂的图形化界面方便用户实时监控系统的状态，简化了安装和维护过程，提高了管理效率。

同一Web界面可以管理多套ParaStor300S存储系统，便于多集群的统一运维。

* 1. ParaStor分布式存储产品规格

|  |  |  |
| --- | --- | --- |
| **产品名称** | | ParaStor300S并行分布式云存储系统 |
| **系统架构** | 系统架构 | 基于X86服务器构建的分布式集群架构，Scale-Out横向扩展 |
| 动态扩展 | 单一集群支持3~4096个节点；支持节点、硬盘在线动态添加和安全移除；数据自动迁移，保证各存储节点容量和性能的负载均衡 |
| **访问接口** | 存储协议 | 支持Linux POSIX文件系统接口，Linux(2.6.32以上内核版本的Linux) x86\_64 |
| 支持NFS协议 |
| 支持CIFS协议 |
| 支持FTP协议 |
| 兼容Amazon S3访问接口 |
| 网络接口 | 支持1GbE、10GbE |
| 支持56Gb/100Gb Infiniband，支持IPOIB、RDMA协议 |
| **系统功能** | 配额 | 支持对文件系统/目录的配额  支持对用户/组的配额  支持针对容量、文件数设置统计配额（仅监控使用情况）、限制配额（可设置软配额/硬配额） |
| 权限管理 | 支持NIS，Microsoft Active Directory，LDAP  支持ACL高级权限管理 |
| 负载均衡 | 存储节点的NAS连接支持基于连接数、轮询、带宽、CPU使用率等方式的负载均衡 |
| SSD Cache | SSD作为二级读缓存，加速本地存储节点的数据读访问请求，提升系统性能 |
| 快照 | 支持基于文件/目录的快照技术，针对同一文件或目录创建多个时间点的文件映像，实现对快照目录/文件的内容保留 |
| **可靠性** | 数据保护模式 | 纠删码：N+M:b，最大支持任意4个数据控制器或者任意4块磁盘同时失效，空间利用率最高可达94% |
| 副本：Replication（2X-4X），最大支持任意3个数据控制器或者任意3块磁盘同时失效，适用于虚拟化、海量小文件等场景 |
| 故障域隔离 | 节点分区：允许不同节点分区中同时失效N+M：b或者是Replication规则允许失效的最大节点数和硬盘数 |
| 硬盘分组：对同一个节点分区中的硬盘进行横向分组，允许不同硬盘组中同时失效N+M：b或者是Replication规则允许失效的硬盘数 |
| 磁盘巡检 | 智能感知磁盘的状态，精准预测磁盘故障，并自动纠错或主动处理 |
| 数据重构 | 硬盘或节点失效可自动触发数据重构，重构过程无需管理员干预，数据恢复优先级可动态调整 |
| **管理运维** | 统一管理 | 同一Web界面可以集中管理多套存储系统 |
| 安装部署 | 支持集中式配置与部署，在Web界面可实现存储系统的在线升级 |
| 状态监控 | 支持对系统总体状态进行监控，支持对各节点和服务状态进行监控 |
| 性能监控 | 支持对系统总体性能、资源进行监控，以及各节点的性能、资源进行监控 |
| 信息查询 | 支持对系统及各节点状态、性能、资源信息进行查询 |
| 告警 | 支持对系统软硬件故障和系统服务状态异常进行告警，除界面告警以外，支持Email、短信、SNMP等方式 |
| **系统规格** | 系统容量 | 单一命名空间支持EB级空间 |
| 文件数 | 单一文件系统支持千亿文件数量 |
| 存储节点数 | 单一集群最大支持4096个 |
| 聚合IO带宽 | 随存储节点数量增加呈线性增长，可以达到数百GB/s |
| 集群NAS协议 | 单一访问区允许同时导出的NFS共享目录为65535个 |
| 单一访问区允许同时导出的SMB共享目录为65535个 |
| 单一访问区允许同时导出的FTP共享目录为65535个 |
| 文件快照数量 | 目录路径上快照数量上限：1000个  集群内文件快照数量上限：20000个 |

1. ****存储系统集成方案****
   1. 项目建设原则

（1）系统性：本系统应保证其在功能上的完整性，易于操作，维护管理方便。所有日常维护工作要求全部实现不停机在线式操作；

（2）维护性：提供有效的系统监控手段和软硬件工具，监测网络通信系统、存储系统的运行状态，提供有效的故障排除手段。在系统出现故障时，应能够在较短的时间内恢复系统运行。存储设备中的大部分部件支持热插拔。

（3）可靠性：系统应确保7×24小时无故障工作。必须考虑存储设备的冗余备份。必须保证系统的一个模块发生故障时，不影响到其他模块的正常工作。

（4）扩展性：系统间采用松散耦合架构。用户可以根据自己业务增长的实际情况，对相应的业务子系统进行独立升级和扩容；系统支持在规模、处理能力、存储容量、网络负载带宽、工作流程等方面的升级扩充，在线平滑升级。

（5）开放性：提供的系统硬件、软件对招标人不能有私密性。当存储系统需要与国家安全测评中心业务系统协同工作时，应能按照要求提供相应的接口及测试方案，存储系统可以兼容不同厂家的产品和设备。

（6）安全性：确保系统内存储的原始数据、分析结果数据不被非法修改和非法获取，保证数据资料存储安全，设备安全、数据传输安全，内部网络安全保护，软件容错及质量控制等。

（7）标准化：为实现国家安全测评中心平台内系统之间的互连互通，以及提高本系统内各模块的可替换能力，系统必须采用标准化的设备模块和接口。

（8）经济性：保证整体系统的安全性和稳定性的前提下，提供性能价格比较高的软硬件设备。

（9）先进性：采用国际上先进的、成熟的计算机及网络技术。系统总体设计具有一定的超前性，确保系统能够在业界具有领先地位，同时在3~5年内不会落后于主流技术。

* 1. 项目建设目标

本次存储资源共享平台满足XXX项目规定的如下需求：

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX。

* 1. 存储系统配置清单

|  |  |  |
| --- | --- | --- |
| 产品型号 | 技术规格及配置 | 数量 |
| UStor 224 | 高性能64位处理器，256GB Cache，2个管理网络接口，内部网络：2个100Gb数据接口，提供84TB的存储容量（7.68TB NVME SDD×23），配置7.68TB高速共享空间（7.68TB NVME SSD×1），1+1冗余电源。  内嵌高性能数据存取引擎，用于并行处理所有客户端的数据访问请求，内嵌高性能数据恢复引擎，支持节点间replication（多副本）/N+M（纠删码）数据冗余方式。 | 8 |
| UStor 460 | 高性能64位处理器，256GB Cache，2个管理网络接口，内部网络：2个100Gb数据接口，提供的存储容量（20TB HDD×6），配置7.68TB高速共享空间（7.68TB NVME SSD×2），1+1冗余电源。  内嵌高性能数据存取引擎，用于并行处理所有客户端的数据访问请求，内嵌高性能数据恢复引擎，支持节点间replication（多副本）/N+M（纠删码）数据冗余方式。 | 20 |
| IB交换机 | 40口 200Gb IB交换机，含双电源，双风扇模块（端口侧出风） | 36 |

备注：数据安全保护机制为N+M:b，空间利用率为N/(N+M)，实际可用容量=所有数据盘裸容量之和\*空间利用率\*90%=XX TB（90%是硬盘折算率，主要是硬盘标称容量与格式化成文件系统后的容量差值）。