BULLFROG®

分布式统一存储软件

技术白皮书

文档版本	V3.0(正式版)
发布日期	2023-4-30

上海讯真信息科技有限公司

目 录

1 BULLFROG®分布式统一存储 概述

1.1 概述

在数据爆炸时代，人们可以获取的数据成指数级的增长。云服务、移动性、大数据和分析、社交网络和专业的互联网产品等技术的发展彻底改变了商务和日常生活的几乎每个方面。当谈及存储时，数据雪崩问题正在逼近现今的数据中心，因此现在必须采取措施应对这种巨大的数据量。

数据中心和专家必须确定能够灵活响应数据增长的大规模可扩展基础设施的类型。例如，云服务供应商近几年便面临这样的挑战。此类数据密集型环境的经验表明每当千万亿字节范围内需要巨大数据量以及线性可扩展性能时，传统的存储架构就会快速达到限制。在这种情况下，必须克服以下挑战：

- RAID重建耗时、危险
- 高可用性会导致成本激增
- 难以预料的数据增长会导致配置过度或不足
- 数据迁移极其耗时
- （预计）停机的重大问题
- 每十亿字节的成本较高
- 性能问题

传统企业级存储成本高，扩展能力无法满足业务系统预料之外的高速增长，重大问题的停机时间过长，PB 级别数据存储的性能问题诸如之类的问题，传统企业级存储已经难以支持超大数据量（PB 级）存储。

因此，云服务、移动性、大数据和分析、社交网络和专业的互联网产品等数据中心平台技术正在将数据增长推向无法想象的高度，并且在未来几年内仍将继续保持这种态势。而应对这种发展的唯一方式就是通过高容量的存储基础设施实现可线性扩展的性能。

而上海讯真的 BULLFROG®存储系统使您能够轻松地自行应对数据激增问题。这一超大规模、由软件定义的系统适用于具有巨大在线数据量的所有环境，并以始终如一的高性能水平提供几乎没有限制的可扩展能力。

1.2 目标用户/适用场景

上海市上海讯真信息科技有限公司的 BULLFROG®分布式统一存储可以为如下类型的用户/使用场景提供分布式存储服务：

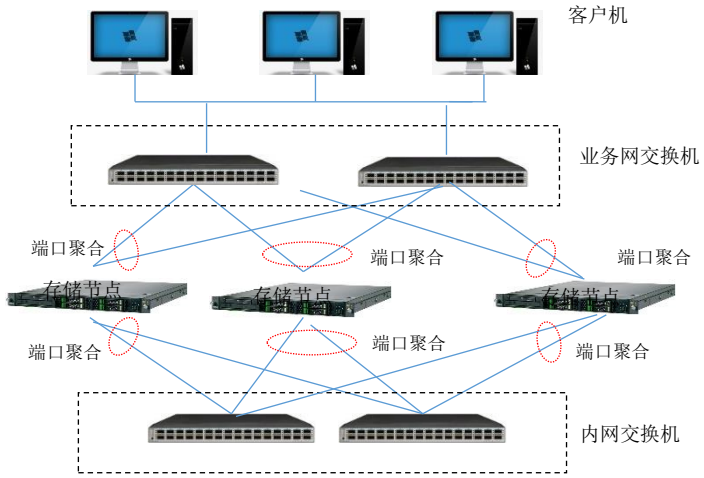
应用领域	要求	分布式存储的优势
媒体行业	可快速并可靠地使用已处理的大量数据	■ I/O性能高，适于视频和音频流 ■ 卓越的容量可扩展性 ■ 每十亿字节的低成本
医疗行业	存储和读取大量医疗影像数据	■ 零停机，服务质量高 ■ 可按需扩展，降低前期投资 ■ 无需中断即可进行维护和组件更换
教育行业	大量数据的保存、读取、归档	■ 极高的容量可扩展性 ■ 高性能始终如一 ■ 无前期成本，扩展选项几乎无限制 ■ 无需中断即可维护和组建更换
银行，保险公司	客户和企业数据具有较高的在线可用性	■ 横向扩展架构，可随时在线 ■ 无（计划）停机时间 ■ 低成本实现灾难恢复理念 ■ 超级可扩展的容量和性能
具有各种研发项目的企业	具有大量文档、照片、图形、视频的项目数据	■ 降低成本 ■ 数据访问性能高 ■ 可按需扩展，降低前期投资

1.3 产品架构

1.3.1 硬件架构

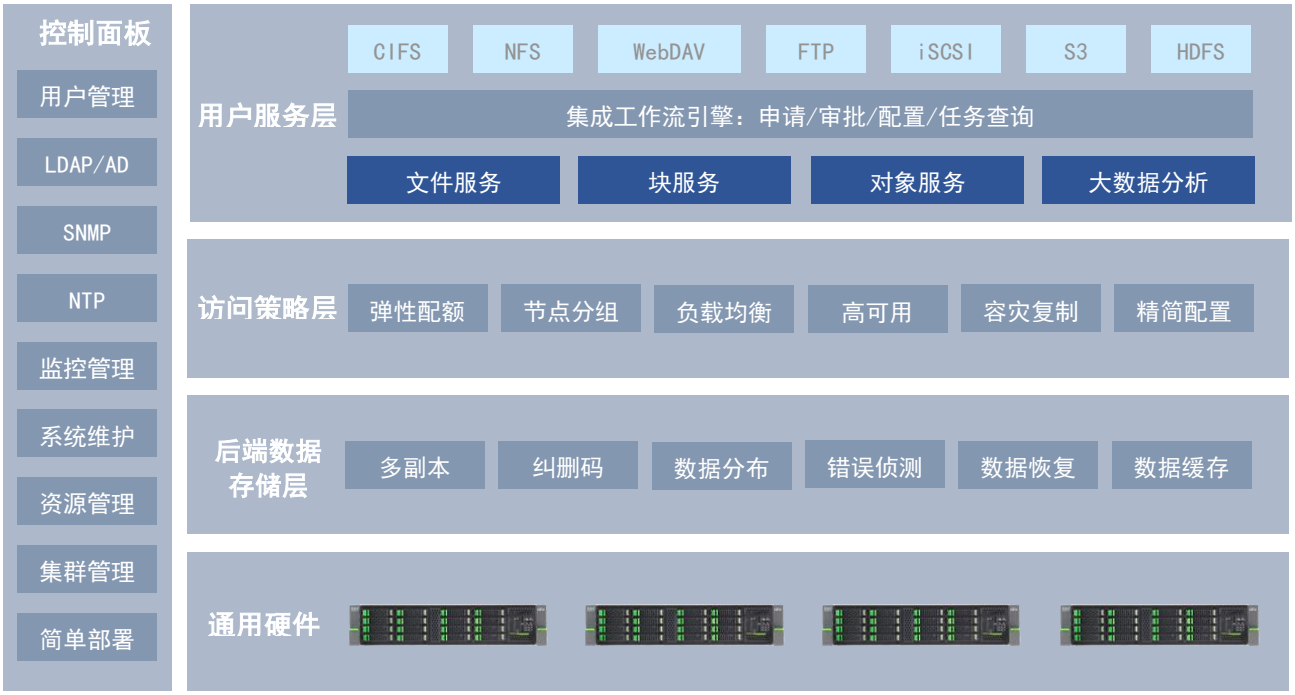
BULLFROG®以标准 x86 服务器作为存储节点，数据通过哈希分布算法均匀存放在系统的各个存储节点。BULLFROG®为分布式去中心化存储，无须部署元数据节点，每个节点都是存储节点，并可以作为管理节点。每个节点的功能一致，地位均等，数据均衡分布。

BULLFROG®存储节点理论上可无限扩展。当前支持配置最多 65536 个存储节点，实现高达 100 PB 的容量。



1.3.2 软件架构

BULLFROG®支持高性能硬件和多种存储接口，在配置上有丰富的灵活性和可靠性，提供 7*24 小时不间断的存储服务。BULLFROG®配有图形用户界面，为系统管理员提供所见即所得的监视和操作界面。



2 BULLFROG®分布式统一存储 功能服务

2.1 分布式文件系统

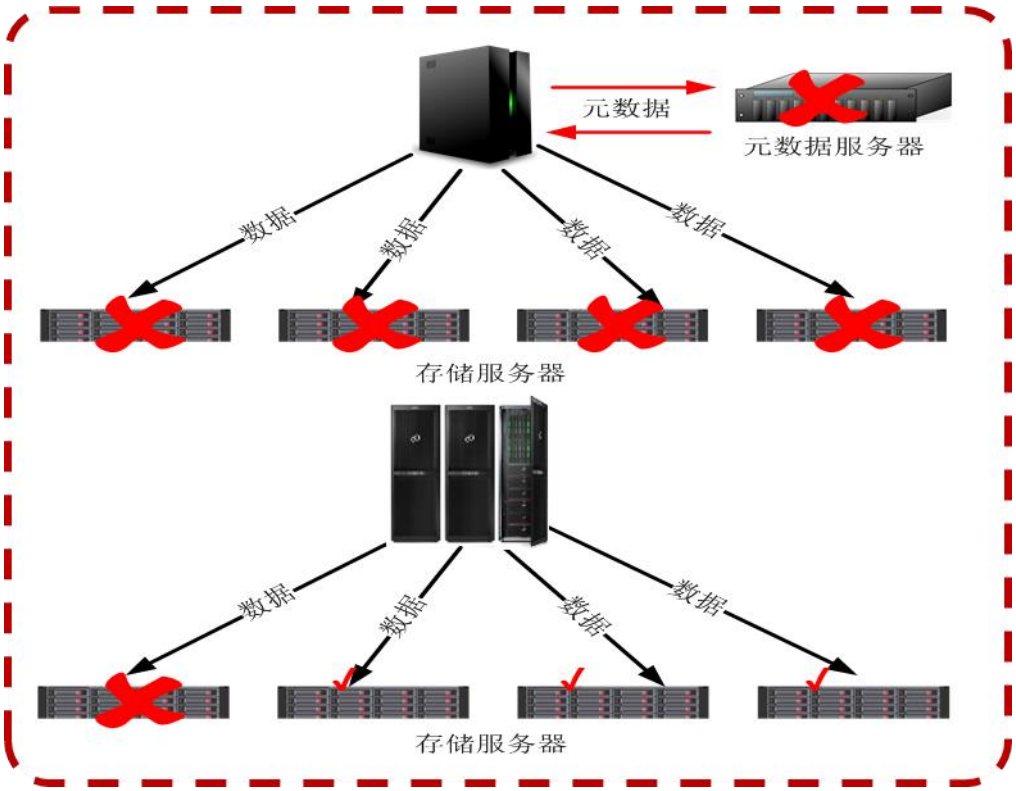
BULLFROG®采用全对称分布式架构，为开放的、软件定义的横向扩展的存储平台，可在零停机情况下实现存储容量和性能的无限模块化线性扩展，从而可以即时、经济、有效地在线访问海量数据。单系统可同时支持块存储、文件存储和对象存储。

BULLFROG®通过采用成本与性能优化的企业级存储组件以及数据自动复制和自动修复功能，在实现容量和性能的高扩展性同时保证了高可靠性。独特的数据分布算法消除了性能瓶颈和单点故障，确保了容量和性能随着每添加的节点线性扩展。此外，采用了专用连接协议和缓存技术大幅提升了存储性能。

BULLFROG®可轻松用于物理，虚拟和云环境下的非结构化数据管理，例如：媒体内容（视频、图像、音频文件等），备份镜像和近线归档，虚机镜像，以及大数据 （日志、RFID 数据及其它机器数据），可广泛应用于广电媒体、电信运营商、能源勘探、科研院所以及公共事业（如城市、机场视频监控）等行业。

2.1.1 分布式存储卷

BULLFROG®不同于一般的分布式存储系统，它采用全对称分布式架构，没有元数据节点，所有数据分散存储在多台独立的存储服务器上，构成一个虚拟的海量存储卷。



下面介绍分布式存储卷的工作原理和管理机制。

工作原理：

BULLFROG®存储系统将海量数据存储映射为一个统一的存储卷，该存储卷可以通过 CIFS，NFS 或 FTP 映射到一个访问共享目录，也可以通过 iSCSI 映射成一个本地磁盘。写数据的流程大体上如下：

1) 当收到数据写入请求时，通过标准的FUSE协议或VFS传到存储系统内部；
2) 通过内部的DHT算法动态计算数据应该保存在相应存储节点上；
3) 通过Replica（副本）或Erasure Code（纠删码）算法保证数据的安全性；
4) 通过特征模块进行数据的进一步处理；
5) 最后通过POSIX协议保存在相应的存储节点的文件系统里。

管理机制：

BULLFROG®提供 Web UI 画面进行存储系统的监控和管理，包含以下功能：

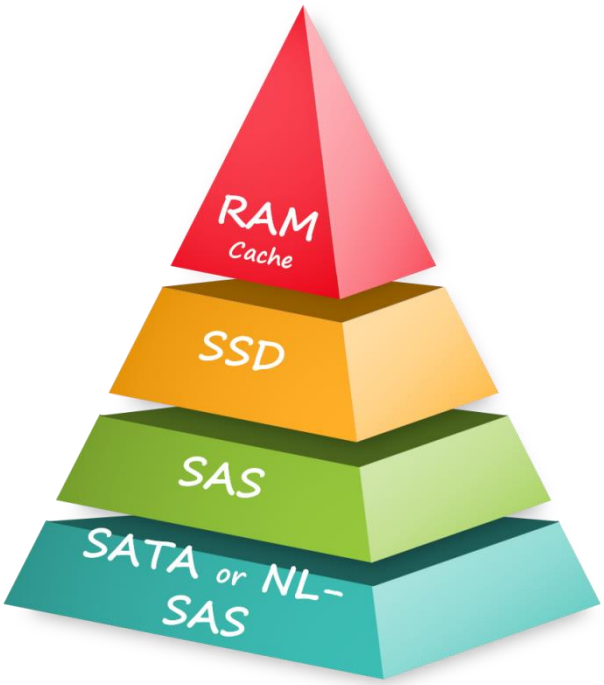
1) 提供系统的统一监视画面，包括存储卷的运行状态，使用情况等；
2) 提供创建存储卷，节点分组，高可用和负载均衡配置；
3) 提供用户和权限管理；
4) 提供文件服务，对象服务，块服务，大数据分析；
5) 提供和SNMP网管系统联携的设置；
6) 提供NTP管理；
7) 提供日志功能；
8) 提供监控和告警功能；
9) 提供文件审计功能（定制功能）。

2.1.2 单一命名空间

无论内部节点数量是多大，硬盘总容量是多大，BULLFROG®都可以对用户提供统一的命名空间。也就是说，用户能够访问一个容量最高支持 100PB 的卷，对于应用程序来说，只需要正常读写该卷即可。从逻辑上，用户可以认为自己使用的是一套超大容量和超高性能的专业存储，而无需考虑物理上的部署细节，比如，数据到底存放在哪个节点的哪个硬盘上。

基于数据访问频度，数据可在快（SSD）、慢（HDD）层间自动迁移。

本系统支持分级存储技术，对数据进行分门别类，分级的存储管理。分级存储技术对集群中的存储资源划分成不同的存储池，数据在存储的时候根据用户预先设置好的策略存放到不同的存储池中，譬如，性能要求高的数据可以存放到 SSD/SAS 组成的存储池中，性能要求不高的可以存放到 SATA 盘组成的存储池中。此外，整个系统会定期扫描元数据，将匹配策略的数据自动从一个存储池中迁移到另外一个存储池中。自动分级存储技术可以通过不同的存储设备来服务不同需求的应用业务，提高整个存储系统的可用性，并且定期将数据备份到慢速廉价设备上，便于用户归档存放。



图中各层解释如下：

RAM Cache: 每个节点都会利用 RAM Cache 进行性能加速。Cache 的大小和节点内存大小直接相关。

SSD：采用 SSD，可以带来非常高的 IOPS，尤其适合对热点数据频繁读取的应用。因为 SSD 存放的是持久化的数据，且价格日趋下降，所以它在存储系统中所占的比例会越来越大。用户可以在 SSD 上面建立一个高速的卷，直接对它进行读写操作。也可以利用 SSD 做 Cache，采用全透明的方式使用 SSD，让系统自动把频繁读写的数据缓存在 SSD 上面，进而提高 IO 性能。

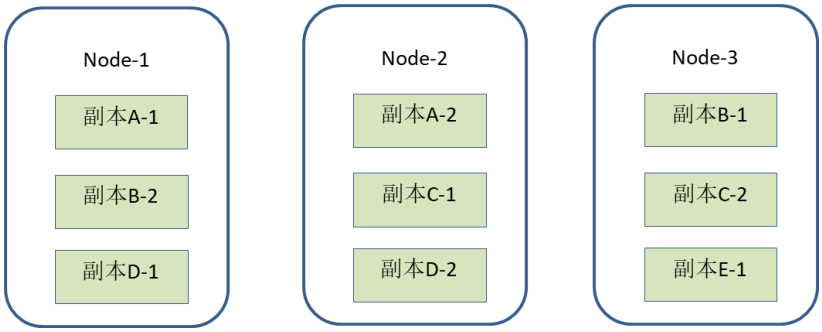
SAS：用户可以使用高速的 SAS 硬盘（比如， 15000 转）用来存放既有容量要求又有性能要求的重要数据。

SATA：用户可以使用企业级的 SATA 硬盘，用来存放有巨大容量要求的数据。

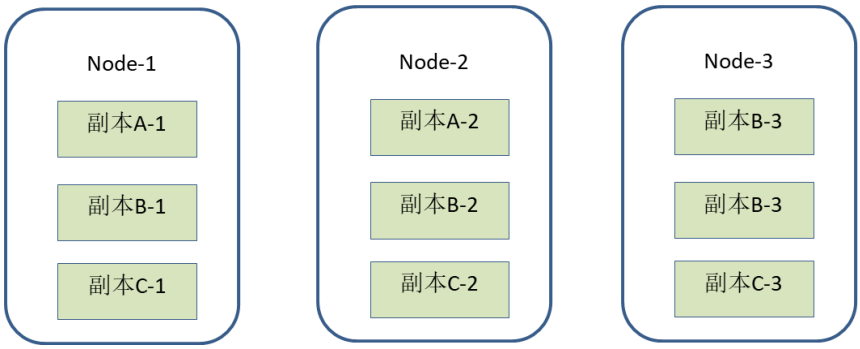
2.1.3 多副本提高数据可靠性

分布式存储系统多采用多副本的方式实现容错。每一个文件有多个存储副本（支持大于 2 个及以上副本），分别存储在不同的存储节点上。支持多副本及跨节点保护技术：全局数据灵活多级别冗余设置数据保护，最大可达 N+4 保护级别或 8 倍副本以上的镜像保护。

副本的分布策略考虑了多种因素，如网络拓扑，磁盘利用率，内网带宽等。对于每次读写，数据强一致性特性保证必须将所有副本全部写入才视为成功写入。在其后的过程中，如果相关的副本出现丢失或不可恢复的状况，在磁盘或者节点恢复后，存储系统会自动进行数据重构和复制，从而确保一定的副本个数，在数据恢复的过程中不妨碍数据的读写访问。在有多副本的情况下,只要保证至少有一个副本都不会造成数据丢失，而且随着副本数目增多，整个系统的可靠性越大。如下为 2 副本和 3 副本的存储示意图。



2 副本存储示意图



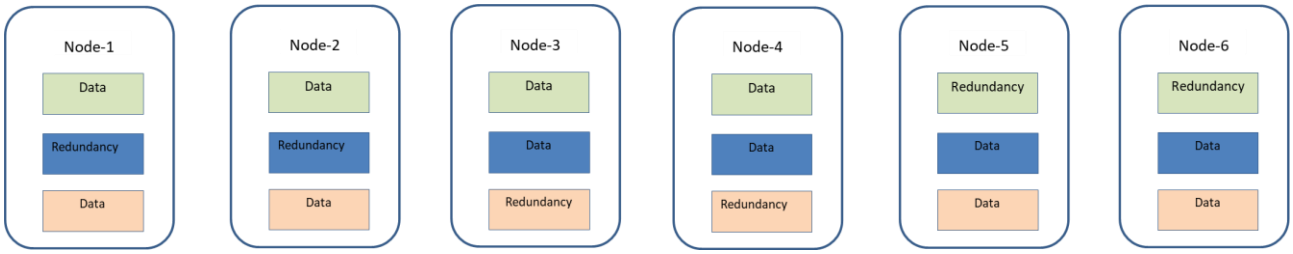
3 副本存储示意图

2.1.4 Erasure Code 提高磁盘使用率

Erasure Code 的出现给分布式存储提供了分布式 RAID 特性。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，所以支持多个盘失效的算法也就是 erasure code 是解决这一问题的办法。

Erasure Code（EC），即纠删码，是一种前向错误纠正技术，主要应用在网络传输中避免包的丢失， 存储系统利用它来提高存储可靠性。相比多副本复制而言， 纠删码能够以更小的数据冗余度获得更高数据可靠性， 但编码方式较复杂，需要大量计算 。纠删码只能容忍数据丢失，无法容忍数据篡改。Erasure Code 可以将 n 份原始数据，增加 m 份数据(用来存储 erasure 编码)，并能通过 n+m 份中的任意 n 份数据，还原为原始数据。定义中包含了 encode 和 decode 两个过程，将原始的 n 份数据变为 n+m 份是 encode，之后这 n+m 份数据可存放在不同的磁盘或节点上，如果有任意小于 m 份的数据失效，仍然能通过剩下的数据还原出来。也就是说，通常 n+m 的 erasure 编码，能容 m 块数据故障的场景，这时候的存储成本是 1+m/n，通常 m<n。因此，通过 erasure 编码，我们能够把副本数降到 1.x，提高了磁盘的使用率。

下图为 EC4+2 的示意图，每个文件都被切分成很多条带，并且每个条带又通过算法分布在不同的节点上，在 EC4+2 的情况下，一个条带被分为 4 块数据块和 2 块冗余块。当任意 1 个或 2 个数据块异常时，都可以通过剩下的数据块和冗余块将异常的数据块计算出来，保证数据不丢失，存储系统正常读写。



EC 存储示意图

如下表格为不同配置策略时	最低节点数	硬盘利用率

的磁盘利用率 纠删码策略		
2+1	3	67%
4+1	5	80%
8+1	9	88%
4+2	6	67%
6+2	10	75%
8+2	10	80%
6+3	9	67%
8+3	11	73%
8+4	12	67%

纠删码与副本对比：

三副本和纠删码是分布式存储中常见的两种数据保护机制。由于纠删码存在写放大问题，小块数据的写性能不足，通常仅适用于视频、备份、容灾等对 IO 性能要求不高的业务场景。在虚拟化、私有云、数据库等块存储场景，最常见的是三副本机制，即数据块按某种随机规则，保持在三个不同节点上的不同磁盘上。

	磁盘利用率	计算开销	网络开销	重构效率
多副本（3）	1/3	无	高	高
纠删码（n+m）	n/(n+m)	高	较高	低

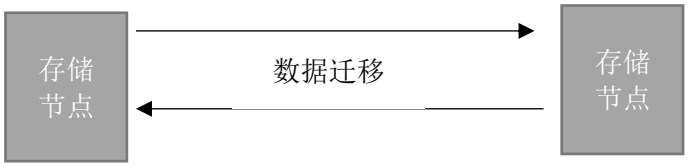
2.1.5 数据再平衡

在如今的大数据时代，数据量的增长呈指数级的态势，需要永久化的存储数据也随之增长，传统的存储系统容量已经很难应对数据量爆炸式的增长，因此分布式存储系统便应运而生。分布式存储系统即通过网络和软件定义存储系统，组合多个独立的服务器来解决存储容量以及扩展的问题。对于分布式存储系统，增加节点以及减少节点是日常运维中经常会碰到的问题。

以增加节点为例，随着数据的增长，企业需要增加存储节点来扩充分布式存储集群的容量，而增加新的存储节点后会出现集群内各个节点的数据存储量分布不均衡的情况，旧数据只存在原有存储节点上，新增存储节点只会存储后续新增的数据，长此以往，原有节点会出现负载过高等情况，而新增存储节点却没有被利用，于是，存储节点之间就产生了数据的不均衡。

现在常用的数据均衡算法是对文件计算哈希值。再根据哈希 Map 将文件存放到规划好的存储区域。但是这样的方式在后续的运用中，可能会给系统造成没有大量必要的性能浪费。比如哪怕修改了文件的一个字节，都会让文件的哈希值发生变化，从而触发系统的数据再平衡，强行进行文件的数据迁移，造成额外的 CPU 开销和硬盘的读写损耗。

在本分布式统一存储中，采用了更加智能化的数据再平衡算法，在充分确保数据均匀分布的状态下，减少没有必要的数据迁移。



2.1.6 使用 RDMA 技术降低网络延时

BULLFROG®可采用 softRoCE 的方式实现远程内存访问（RDMA：Remote Direct Memory Access），降低网络延迟提高性能。

TCP 方式的网络节点之间的数据交换

普通网卡集成了支持硬件校验和的功能，并对软件进行了改进，从而减少了发送数据的拷贝量，但无法减少接收数据的拷贝量，而这部分拷贝量要占用 CPU 的大量计算周期。

普通网卡的工作过程如下：先把收到的数据包缓存到系统上，数据包经过处理后，相应数据被分配到一个 TCP 连接；然后，接收系统再把主动提供的 TCP 数据与相应的应用程序联系起来，并将数据从系统缓冲区拷贝到目标存储地址。这样，制约网络速率的因素就出现了：应用通信强度不断增加和主机 CPU 在内核与应用存储器间处理数据的任务繁重使系统要不断追加主机 CPU 资源，配置高效的软件并增强系统负荷管理。问题的关键是要消除主机 CPU 中不必要的频繁数据传输，减少系统间的信息延迟。

RDMA 方式的网络节点之间的数据交换

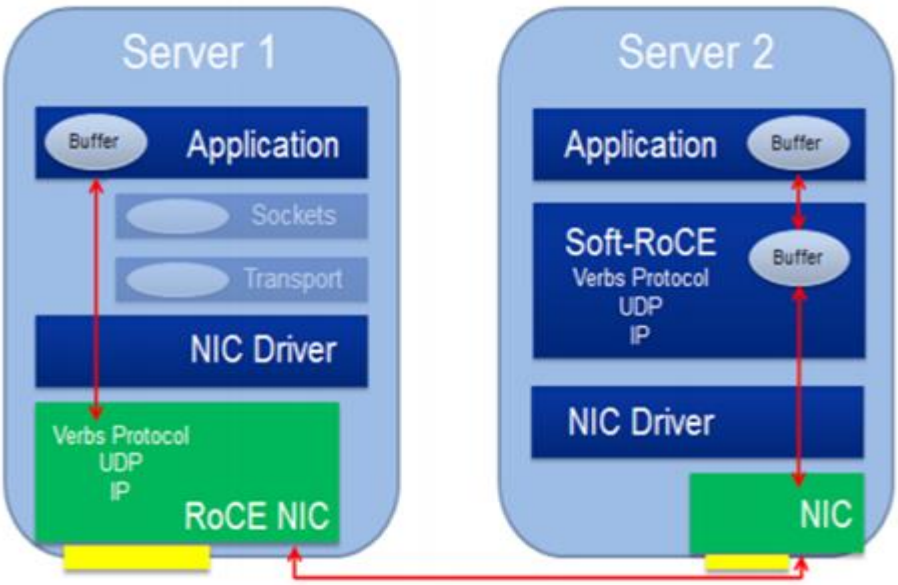
RDMA 是通过网络把资料直接传入计算机的存储区，将数据从一个系统快速移动到远程系统存储器中，而不对操作系统造成任何影响，这样就不需要用到多少计算机的处理功能。它消除了外部存储器复制和文本交换操作，因而能腾出总线空间和 CPU 周期用于改进应用系统性能。通用的做法需由系统先对传入的信息进行分析与标记，然后再存储到正确的区域。

采用 RDMA 技术后与传统的 TCP 方式对比

1）减少了数据在用户层和内核层的拷贝

2）降低了 CPU 的使用率

BULLFROG®通过 softRoCE 技术，在以太网卡上实现了低成本的 RDMA 解决方案

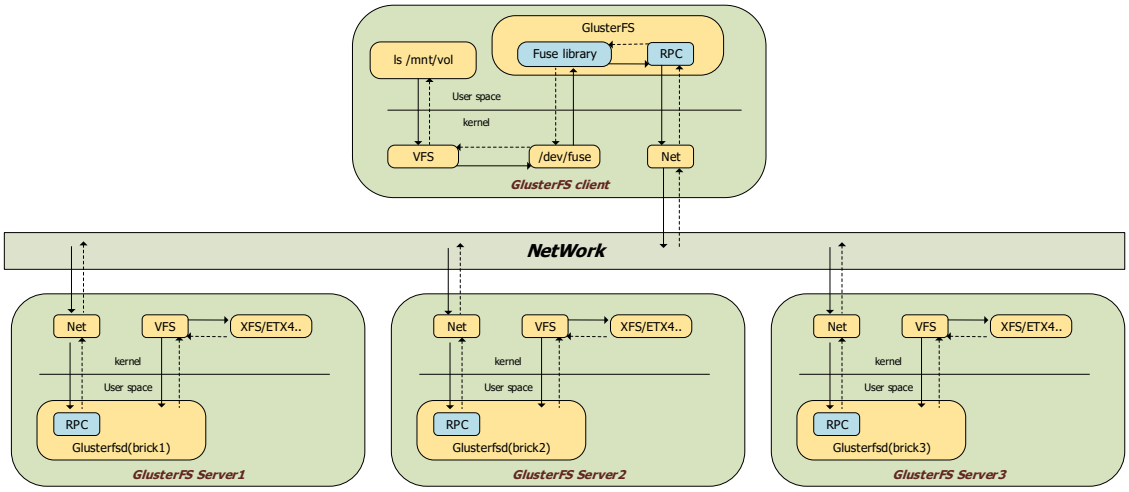


不同于 RoCE（RDMA over Converged Ethernet），softRoCE 适用于任何以太环境，无需依赖 NIC、switch、L2QoS 等支持。softRoCE 的目标是在所有支持以太网的设备上都可以部署 RDMA 传输，对上通过 librxe 与 RDMA stack（libibverbs）耦合，对下通过 rxe.ko 与 linux stack layer3 耦合，用户通过某个 eth NIC 的 UDP 隧道为虚拟的 RDMA 设备传输 RoCE 数据。

2.1.7 减少 Brick 数量以提高海量文件场景下的 ls 性能

BULLFROG®分布式系统由于采用了无元数据中心的架构，海量文件场景下 ls 效率存在如下瓶颈

- 内核/用户态之间的来回切换
- 大量网络通信而导致延迟
- 挂载brick多了以后会引发链式调用，每个brick都要响应数据请求



因此，BULLFROG®的推荐实践：采用 Raid5 减少单节点上的 brick 数量以提高海量文件场景下的 ls 效率。

2.2 文件存储服务

BULLFROG®支持通过多种协议(POSIX，NFS，CIFS，FTP，HTTP/HTTPS 等)访问内部的存储数据，且不用安装客户端，使用标准协议就可以读写数据。用现有的这些传统网络文件协议，能够兼容用户现有的应用环境，无需大力气修改应用架构，可满足用户的各种复杂需求。

对于普通的企业用户，可用通过 CIFS（Samba），NFS，HTTP(S) ，FTP 协议，以文件方式来访问存储。客户端的 OS 可以是 Windows，Linux，Unix，VMWare ESX，ESXi，XEN，Hyper-V，KVM 等。 无需安装任何其他软件，直接使用系统自带的功能模块，便可以加载对应的数据卷，且多个客户端可以同时对同一个卷进行读写。

2.3 对象存储服务

BULLFROG®提供对象存储服务，支持 Amazon S3 接口对对象存储服务进行访问，提供的服务包括：创建、修改、删除桶，上传、下载、删除对象、权限管理等，支持的 API 一览以及支持程度可参考《S3 Rest API Reference》。

BULLFROG®对象存储通过先进的算法，对对象存储的应用场景，尤其是海量小文件的应用做了优化。通过小文件聚合的方式，将大量小文件聚合成大文件存储到硬盘中，降低了在高并发高压力的情况下对磁盘 IO 的影响。在文件写入或者修改时，采用附加写的方式进行写入到磁盘文件中，将随机无序写入变为顺序写入，极大的提高了对象存储的 IO 性能。能在纯机械硬盘的服务器上，提供接近友商纯 SSD 配置服务器的性能，提供更长的耐久性和更高的性价比。

同时 BULLFROG®对象存储通过先进的数据管理算法，在存储了几千万甚至上亿对象后，能够保证性能不明显衰减。相对于友商的对象服务（例如 mino），在存储了大量数据后，会有显著的性能下降。相对于 Ceph，配置和扩容更加简单，且性能近似线性增长，在面对爆炸性的海量扩容和性能需求方面能充分满足用户需求。

2.4 块存储服务

BULLFROG®提供基于标准 iSCSI 协议的块存储服务，支持多路径(Multipalpath)高可用。尤其在海量大文件和高并发的情况下，可提供高速稳定的块访问方式供客户使用。支持 Linux/Windows/Mac 等常用平台。

当前支持的块存储的功能一览如下所示：

- Create volume
- Delete volume
- Attach volume (支持以 multipath 形式提供卷的访问)
- Detach volume
- Clone volume
- Snapshot

创建块存储服务时，推荐使用副本模式。纠删码虽然磁盘利用率较高，但是在小文件的情况下会降低 IO 性能。

2.5 大数据存储服务

数据挖掘、深度学习这些前沿的技术都依赖于大规模数据集的处理，HDFS 是 Hadoop 项目的核心子项目，是分布式计算中数据存储管理的基础，是基于流数据模式访问和处理超大文件的需求而开发的，可以运行于廉价的商用服务器上。它所具有的高容错、高可靠性、高可扩展性、高获得性、高吞吐率等特征为海量数据提供了高可用存储，为超大数据集（Large Data Set）的应用处理带来了很多便利。

BULLFROG®集成了 Hadoop 大数据分析与应用，支持用户对 BULLFROG®上存储的数据进行大数据分析。用户在首先申请文件服务，将此文件服务作为其他系统的数据存储后端，将大数据存放存储集群中。无需将存储的数据导入到其他模块，用户上传算法后，充分利用集群的计算力和高并发高带宽，迅速的对海量数据进行分析计算，分析结果会自动导入相关路径供用户使用。大数据服务能帮助企业轻松构建大规模数据集的应用，充分挖掘信息价值。

2.6 企业网盘

BULLFROG®支持企业网盘服务。BULLFROG ®企业网盘通过资料库来分别管理文件，资料库可以单独同步，也可以加密，且密码不会保存在服务器端，保证数据安全。

BULLFROG®企业网盘后端对接分布存储系统，提供容量集群管理，扩容，高可用和故障恢复等功能。

BULLFROG®企业网盘实现了与存储的代码级集成，无需额外软硬件即可实现网盘服务的高可用，确保任何节点的故障下网盘服务能够正常运行。网盘对客户端数量，容量，上传下载速度和文件大小不做限制。

BULLFROG®企业网盘对接本地 LDAP/AD 服务器，提供用户的存储访问认证，保证访问安全。

BULLFROG ®允许用户创建群组，可以在群组内开通企业网盘服务，群组内共享和同步文件，进行团队协同工作。

用户可以通过 Web UI、移动端、虚拟网盘等多种平台来访问 BULLFROG ®企业网盘，各个平台之间数据相互同步，提供版本控制，不用担心异常覆盖。

提供各种常见的文件管理操作，支持并实际配置网盘文件和文件夹上传和下载，复制，移动，重命

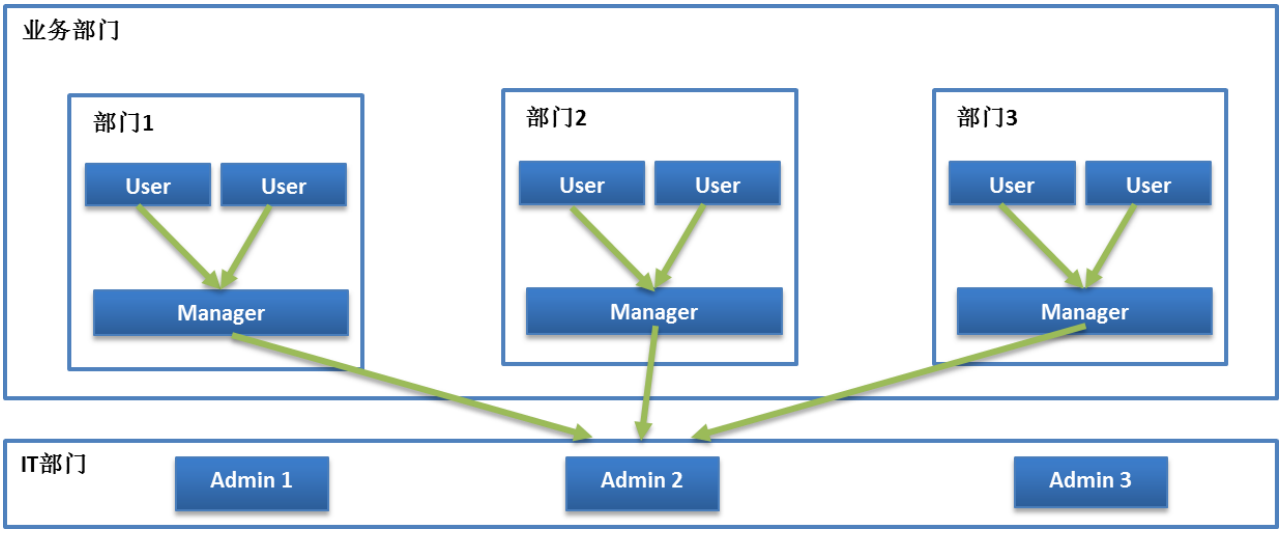
名，删除和重建文件和文件夹，支持文件在线预览和编辑，文件和文件夹收藏，消息通知，文件和文件夹共享功能，支持共享管理，上传链接等网盘功能模块

2.7 存储即服务

BULLFROG®管理软件内置流程引擎，具备云存储 Storage as a Service 功能，可配置云存储全容量许可，将存储服务开放给业务部门和 IT 支援部门使用。

BULLFROG®管理软件集成存储资源多级申请审批工作流功能，集成空间动态管理， 集成权限管理，可提供 Web 界面全图形化的根据人员级别对存储资源进行申请，审批，监 控等自助化流程服务。

流程引擎中定义 3 个角色，分别为用户，经理以及 IT 管理员，并为他们分配了不同的流程处理权限。



- 用户：

  依据业务实际需求，通过存储管理软件提供的 Web UI 申请资源。

- 经理：

  作为用户的上级，站在业务层面上对用户的申请进行判断。

  如果申请符合业务需求，审批通过用户的申请；否则打回申请。

- IT管理员：

  作为公司的 IT 技术人员，站在 IT 系统的层面上对用户的申请进行判断。

如果申请符合 IT 管理规范，审批通过用户的申请；否则打回申请。

引入流程处理引擎的优点

1） 提高效率

审批流转自动化，过程中无需借助传统的通信手段（例如Email），全程只需要通过Web UI即可完成。

2） 协同内外，快速响应

将申请人、审批人、资源三方集成起来，无论何时、何地，整个企业都有机地紧密整合在一起，协同工作。

3） 监控全面，提升执行

管理层可以全面把握资源申请情况，了解和分析任务状态，从而全面掌握企业的运营效率等。

自助服务类型

- 文件存储
- 对象存储
- 块存储
- 大数据服务

2.8 数据保护

2.8.1 远程数据复制

在大型系统中，经常会涉及到跨多个数据中心的需求。在对服务质量和灾备要求更高的场景中，会规划将机房部署在地理位置分散的多个数据中心内。在此类多数据中心部署中，通常会使用跨地域复制机制提供额外的冗余，以防某个数据中心故障、自然侵害或其他事件导致服务无法正常运作。Bullfrog 提供了远程数据拷贝功能 Geo- replication。

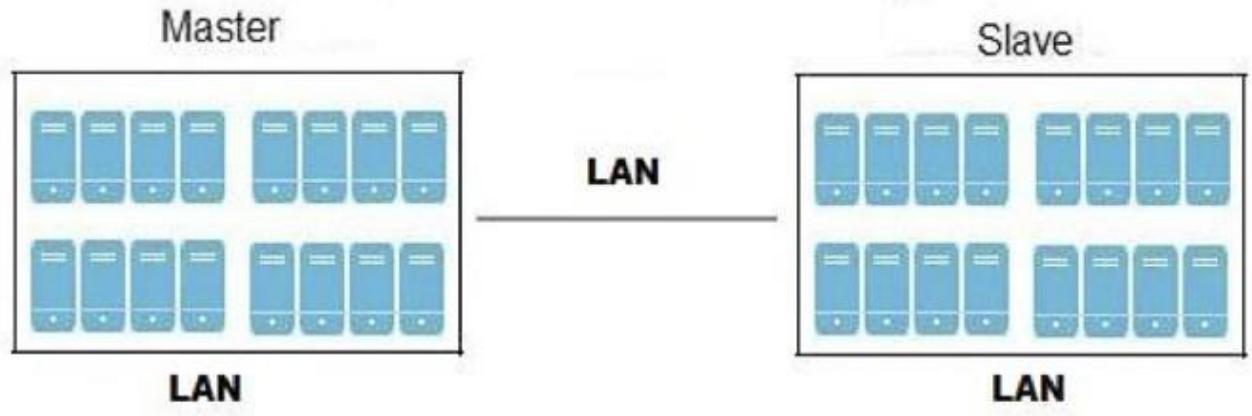
BULLFROG®提供远程数据复制功能，支持一对一、一对多、多对多复制方式，支持单向和双向复制；支持增量复制。

Geo-replication 是指跨数据中心的广域异地复制，BULLFROG®提供了一种持续，异步，增量的数据备份策略，可以通过局域网，广域网，Internet 对物理上分离的两处数据中心进行备份，增强数据的可靠性。使用 Geo-replication，我们能够在存储环境建立数据冗余，提供数据的灾难恢复功能。

Geo-replication 使用主-从模式，主存储卷和从存储卷之间使用异步复制和镜像机制，主存储卷和从存储卷通常不在一个集群中。主存储卷可以是集群中的一个卷或者一个访问目录，从存储卷可以是一个 mount 点。实际生产环境可以采用阵列或者 NAS 系统作为备份数据存储中心，该路径就是阵列或者 NAS 系统的 mount 点。

异地复制有如下使用场景：

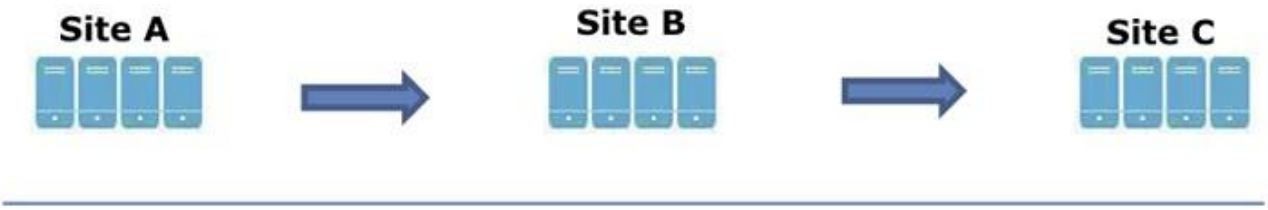
1）通过局域网进行异地复制

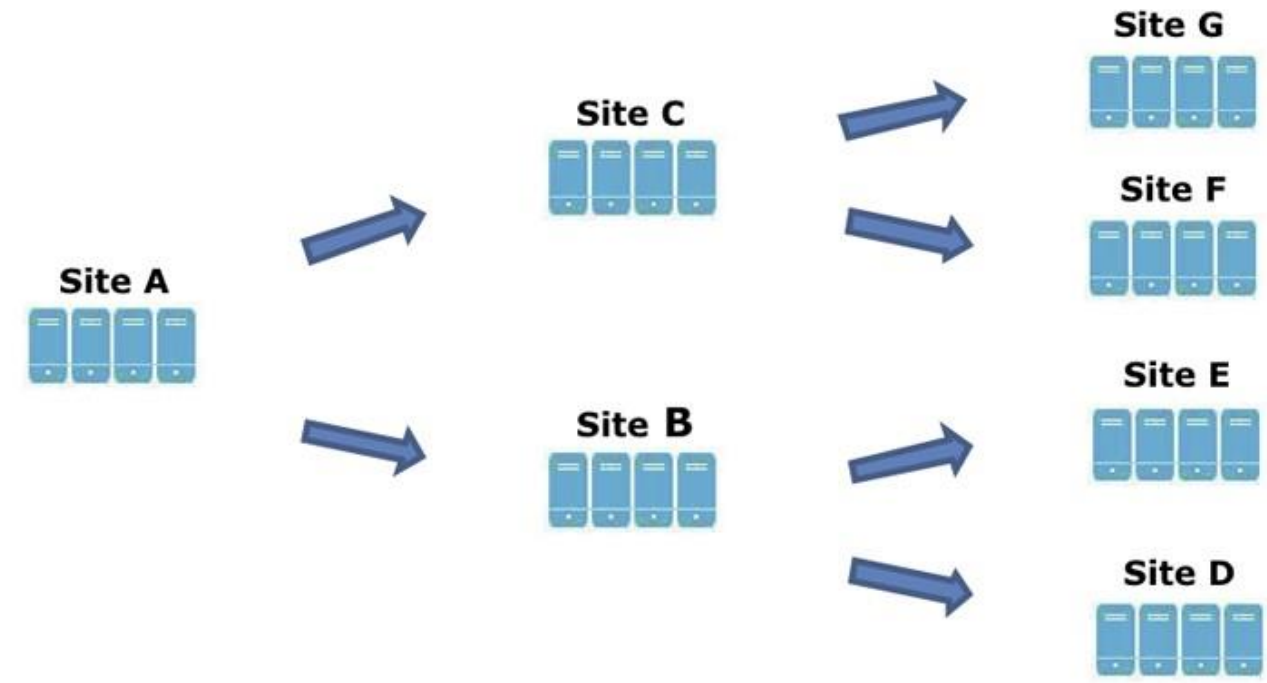


2）通过广域网进行异地复制

3）通过Internet进行异地复制

还可以配置多站点进行级联异地复制：





在主服务器上使用启动异地复制功能，异地复制开启后，主从卷之间自动开始异地复制。异地复制完成后，当主存储卷出现故障的时候，可以使用从服务器/backup 目录下的数据进行灾难恢复。

2.8.2 NDMP

网络数据管理协议（NDMP）是一种基于企业级数据管理的开放协议。NDMP 中定义了一种基于网络的协议和机制，用于控制备份、恢复、以及在主要和次要存储器之间的数据传输。基于 TCP/IP 的行业标准协议，专为 NAS 环境中的备份而设计。NDMP 协议建立在传输于 TCP/IP 链路上的 XDR 编码信息基础上。数据可以通过 NDMP 备份，不受操作系统或平台限制。由于这种灵活性，它不再需要通过应用程序服务器传输数据，从而减少了应用程序服务器上的负载，并提高了备份速度。

传统 NAS 备份方式，传统的网络备份依靠一个安装在所有待备份服务器上的备份代理程序。这些服务器同时访问 NAS 设备，数据通过网络从其他服务器传送到备份服务器上，也就是说，这些服务器先从 NAS 设备中将需要备份的数据读出，然后再通过网络将这些数据传送到备份服务器上备份。

NDMP(网络数据管理协议)是一个专门为 NAS 设备的数据备份系统设计的协议。简单来讲，它可以让 NAS 设备直接向其所连接的磁带设备或者位于网络上的备份服务器发送需要备份的数据，这个过程不需要任何备份客户端代理的参与。

相对于传统备份模式，NDMP 备份模式主要有 LAN-Free，对服务器性能无影响(Server Free)2 个优点。

2.8.3 WORM 数据归档

WORM 技术的全称是 Write Once Read Many(一次写入，多次读取)，即通过 WORM 技术存储在介质中的数据可以做到不会因各种意外而被清除或被修改，确保了数据的完整性和正确性。WORM 由国家严格定义了资料存盘的法规，并要求政府机关，医疗，金融系统必须严格遵循这种规范，BULLFROG® 作为一款通用的存储软件同样支持 WORM 技术。

WORM 技术分为两种类型，其中一种 WORM 技术是通过磁带介质本身来实现的，WORM 磁带是只可以写一次的磁带，从磁带来讲，WORM 磁带与 RW 磁带是两种不同的磁带，本身的性能、内部构造与 RW 完全不一样。用户只要在磁带机中插入 WORM 的磁带，它就变成只写一次的了，这种基于介质的 WORM 可靠性高。第二种 WORM 技术是通过软件技术实现的，本身磁带只有一种——即 RW 的磁带，当用户想要作成 WORM 文档的时候，需要事先将文件共享目录设置成 WORM，这样出来的数据就具有 WORM 属性，只能写一次，不能重复写。相对于基于介质的 WORM 技术，这种 WORM 技术的可靠性要差一些，但可以更灵活的设置，满足成本和功能的双重限制。

BULLFROG®采用软件方式进行 WORM 的支持。在开启 CIFS 共享时，可以设置 WORM 生效期间。在 WORM 生效期间内，文件共享内的文件为只读属性，不可以写、修改、删除。WORM 期间过期后，文件属性恢复可以读写模式。WORM 只对 CIFS 共享生效，对 NFS、FTP 等并不生效。

2.8.4 断电保护

BULLFROG®支持 UPS 联动保护模块，采用缓存断电保护技术，当存储突然断电 后，通过专用 Cache 保护模块，将缓存数据下刷到硬盘中进行永久保存，能够保证缓存中的 数据不丢失。断电保护需要 UPS 支持。

2.9 数据压缩

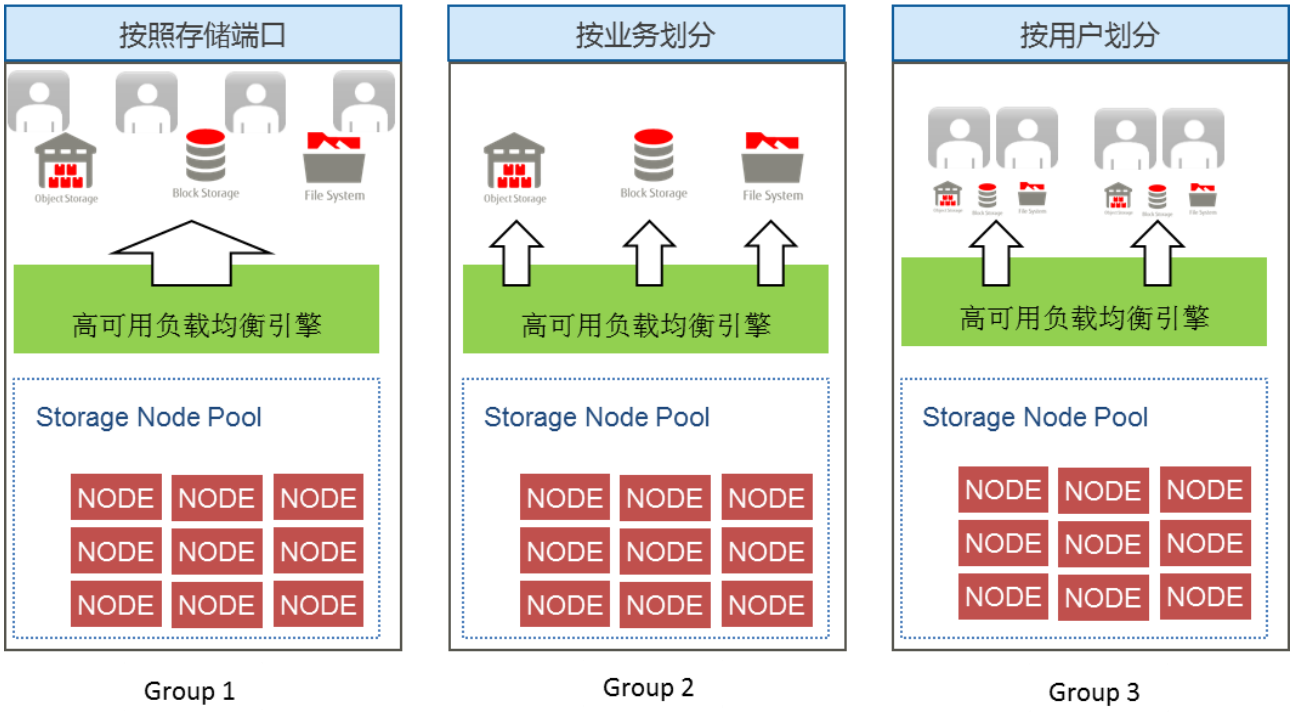
数据压缩技术，就是用最少的数码来表示信号的技术。由于数字化的多媒体信息尤其是数字视频、音频信号的数据量特别庞大；如果不对其进行有效的压缩就难以得到实际的应用。因此，数据压缩技术已成为当今数字通信、广播、存储和多媒体娱乐中的一项关键的共性技术。通过数据压缩技术，将数据在传输之前进行压缩，将压缩后的数据进行传输，到达对向侧后再实行解压。通过可逆的压缩算法，使得数据传输系统的功能丰富度和使用体验都有很大的提升。BULLFROG®支持传输过程中的数据压缩，极大的提高数据传输效率。在本设备中使用 deflate 算法对数据进行压缩和解压。写数据时客户端压缩数据，服务端解压数据；读数据，则相反。网络中传输的数据为压缩后的数据，从而节约带宽，提高数据的传输效率。数据的压缩比率根据数据类型的不同也不尽相同。最大的压缩比比率可以达到 98%。数据压缩功能适用于内部交换网络环境较差的场景，通过牺牲部分 CPU 算力和 IOPS 的代价，以达到提高吞吐量的目的。

2.10 高可用负载均衡

2.10.1 节点分组

BULLFROG®提供灵活的节点分组策略，每个组的节点可以有不同的硬件配置，提供多样性的服务，而且提供独立的负载均衡服务。如下图所示，可以按不同的应用场景把集群中的节点划分成不同的节点组，同一组的节点对外提供相同的服务。每一节点组的存储接口、硬件、服务甚至是网段都可以单

独配置，以适用不同的应用场景。BULLFROG®对这些节点组进行集中式管理，大大的提高了集群系统的易用性。



2.10.2 高可用

BULLFROG®采用多种技术支持系统的高可用性，在任意插拔一台或多台服务器（取决于节点组中节点的个数）的一根网线，或者任意切断一台或多台服务器的电源的情况下，可以实现自动故障切换，将此节点从集群中切离，虚拟 IP 自动漂移，不需要用户切换访问地址；当节点恢复后会自动加入到集群中，实现自动故障恢复。提供 7×24 小时不间断服务。

具体如下：

1. 内网采用网络绑定模式时，任意插拔单个节点的一根网线，在线业务不中断，该节点可以正常提供服务。若单个节点的内网全部断开，该节点的在线业务中断，几分钟之后会恢复。但是，中断发生时该节点写入的数据不保证完整性，需重新传输。

内网/业务网支持以下 bond 模式:

- mod=0 (平衡抡循环策略)

- mod=1(主-备份策略)
- mod=4 (IEEE 802.3ad 动态链接聚合)
- mod=6 (适配器适应性负载均衡)

mod 6 的情况下，在网卡切换的过程中，有可能造成数据的暂时中断，几分钟之后恢复。

2. 业务网通过虚拟 IP 的自动漂移，保证对外的读写业务的不中断。

在业务网断网的情况下，几秒之内能够迅速的实现虚拟 ip 的自动漂移，实现读写业务的不中断。

3. 任意切断一台服务器的电源时，读写业务不会中断，但不保证数据的完整性，需要安装 UPS 保护数据的完整性。

三个节点的情况下，可以支持一个节点断电或者断内网。四个节点及以上的情况下，可以在保证数据完整性的情况下支持两个节点断电或者断内网。对象服务的情况下，不支持相邻两个节点同时断电或者断内网。若多台节点同时断电或者断内网，可能会导致业务无法访问或者数据丢失。正在上传或下载的文件可能会因为网络中断而异常退出，需要重新发起上传或者下载请求。

基于多种客户端在 nfs/cifs 协议下，高可用如下所示：

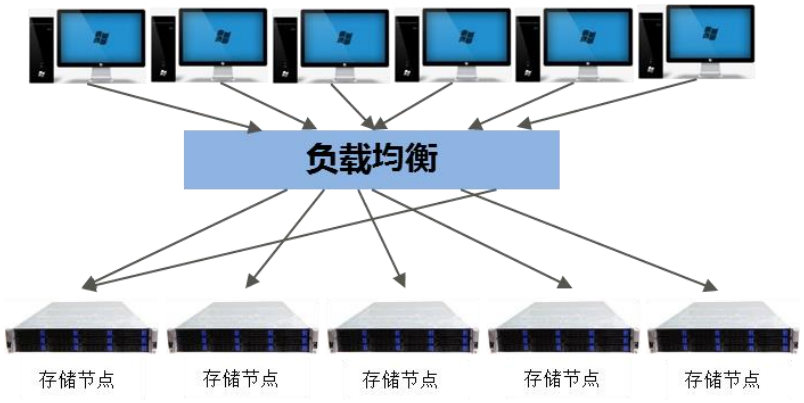
客户端	协议	任意插拔一根网线（聚合状态）	任意插拔一根业务网网线（非聚合状态）	任意切断一台电源	任意插拔一块磁盘	备注
centos	nfs	○	○	○	○	只支持 nfs v3
windows	nfs	○	○	○	○	修改 mount type 为 hard 模式 修改方法： powershell Set-NfsClientConfiguration -MountType HARD
mac	nfs	○	○	○	○	10.10 版本不支持 10.14 版本支持
centos	cifs	○	○	○	○	-
windows	cifs	○	△※	△※	○	※不支持透明故障转移，需要客户端手动重连
mac	cifs	○	-	-	○	-

HTTP 服务不支持高可用，下载或上传的过程中，发生断网或者断电，客户端需要刷新链接，并重新下载或上传。

2.10.3 负载均衡

BULLFROG®负载均衡服务是基于域名请求的负载均衡，默认采用轮询方式进行负载均衡。仅在域名请求的时候进行干预，不会参加到实机的数据流业务中，对系统的整体性能影响较小，不会成系统的性能瓶颈。负载均衡服务可以提供多个虚拟服务 IP，用户可以通过虚拟 IP 进行服务访问，也可以通过域名进行访问。

负载均衡服务属于节点组特性，每个节点组通过访问自己单独的域名，可以自动分配到节点组中的某个节点，分散节点的压力，实现负载均衡特性。



BULLFROG®内部配置了 DNS 服务器，该服务器可以与外部域名服务器进行对接。对接方法如下：

1. 域名转发

将 BULLFROG®内部域名通过配置到外部域名服务器的转发域中，当客户端访问 BULLFROG®内部域名时，由外部域名服务器转发到 BULLFROG®内部域名服务器，BULLFROG®内部域名服务器应答给外部域名服务器之后，外部域名服务器再应答给客户端。

2.子域名

将 BULLFROG®内部域名配置为外部域名服务器的子域名。

比如：

外部域名服务器域为：companyname.com

BULLFROG®内部域名为：BULLFROG.companyname.com

2.11 弹性配额

BULLFROG®在对用户提供文件访问服务时，可以提供类似于 Thin Provisioning 的配额管理。每个文件共享目录都可以设置目录级的配额，每个目录的最大配额可以设置为卷的容量，不受卷实际可用容量的限制。当整个卷的容量不足时，可以通过扩展卷实现扩容。支持并配置针对目录进行容量的随意限制和调整。用户还可以动态配置目录配额，具有高度的灵活性。

2.12 用户权限/用户管理

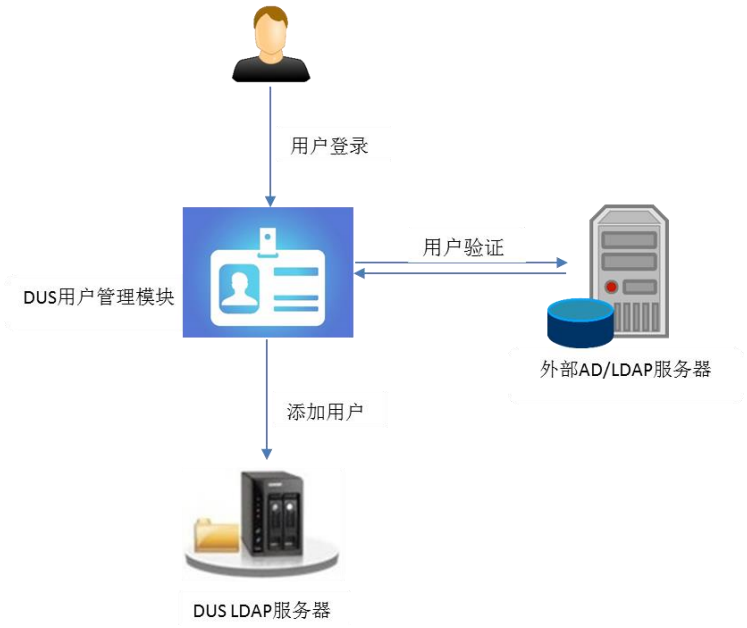
存储通过集中的方式把存储服务提供给众多客户使用，不可避免地会涉及到复杂的权限问题。BULLFROG®支持完整的权限控制，全面满足用户的实际业务需求。它支持创建足够多的用户或者组，在不同的目录针对不同的用户赋予不同的权限（可读可写、只读、禁止访问等）。它支持复杂的分级目录权限，给每一层目录设置独立的权限，严格控制内部数据的安全性和隐私性。用户的密码由用户自己管理，管理员无法查看到用户个人的密码以及网盘内容。

为了方便企业用户与 BULLFROG®的无缝对接，本产品提供了外部 LDAP/AD、本地 LDAP 两种用户管理方式。

2.12.1 外部 LDAP/AD 连接

此功能针对已经部署了 AD 或 LDAP 环境的企业用户。此功能不需要客户将用户重复导入到 BULLFROG®中，即可以实现对既存用户的管理。系统管理员将用户认证配置为外部 AD/LDAP 认证方式。用户在用 AD 或 LDAP 用户名密码登录到 BULLFROG®的登录界面时，BULLFROG®用户管理模块会到外部 AD/LDAP 服务器进行账号验证，如果验证通过，则将用户信息存储到 BULLFROG®本地的 LDAP 服务器中并建立用户。实现用户的灵活导入和管理。

如下图所示：



2.12.2 本地用户管理

此功能针对没有 LDAP/AD 服务器的企业，系统管理员将用户认证配置为本地 LDAP 认证后， 可以利用 BULLFROG®自身的 LDAP 服务器进行用户管理。

2.13 系统监控

2.13.1 状态监控

BULLFROG®存储系统提供了一个一体化管理的Web UI平台，可以对存储系统进行统一的远程监控和管理，其具体功能包括：

1) 可以显示系统总容量和存储资源池的使用情况；
2) 可以监控系统中各个节点、磁盘、网络的健康状态；
3) 可以显示系统的文件服务、对象服务、块服务数量和文件数量；
4) 可以显示网络带宽和磁盘性能统计数据，提供实时、1小时内、1天内、1周内的性能数据；
5) 可以实时监测系统中每个节点服务器的运行状态（包括各个存储节点服务器的内存、CPU、网络流量的利用情况等）；
6) 可以监控磁盘的运行状态和SSD磁盘寿命；
7) 提供告警服务和SNMP与用户网管系统对接服务；
8) 提供操作日志下载功能。

2.13.2 告警功能

BULLFROG®存储系统提供了告警功能，可以对平台监控到的警告或错误级别的信息进行告警。

1) 短信服务：存储系统不可避免的需要维护，告警功能应运而生，短信告警作为一个简单稳定而又限制小的特点而被采用。基本功能如下：

    a. 启用鉴别功能：可以由用户自由决定是否启用此功能，能有减少开支，减少信息烦扰。
    b. 基本告警发送功能：短信告警启用成功后可以针对系统内没有告警并且没有处理过的错误信息对指定正常运行的接收号码进行短信告警。
    c. 接收号码信息修改功能：可以对接收号码进行增删改，灵活修改出所需要的告警级别，号码，备注。

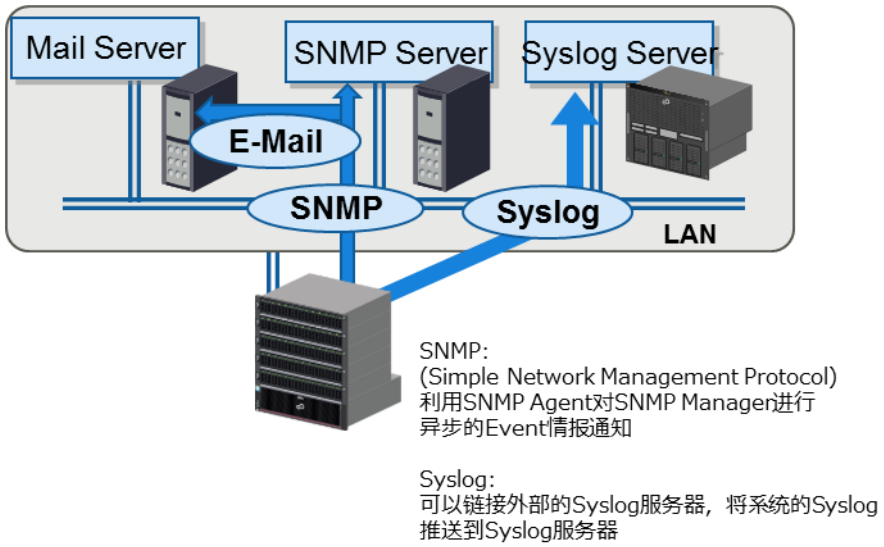
2.13.3 SNMP 和 Syslog

BULLFROG®支持 SNMP 和 Syslog 方式进行系统监控。

SNMP (Simple Network Management Protocol)协议是一个运行在网际协议（IP Protocol）之上的应用层协议。用于收集并配置网络设备信息，如服务器，打印机，交换机和路由器等。SNMP 管理环境由 SNMP Manager(运行监视装置程序的服务器)和 SNMP Agent(被管理对象，storage，switch 等)组成。SNMP 协议的处理主要有以下两种操作：SNMP Manager 主动获取 SNMP Agent 的情报或进行设定；SNMP Agent 对 SNMP Manager 进行异步通知，称为 SNMP Trap。BULLFROG®存储系统监控使用的是 SNMPTrap，通过安装在存储节点上的 SNMPAgent 对 SNMP Server 进行通知。

SNMP 协议目前定义了三个版本的网络管理协议，SNMP v1，SNMP v2，SNMP v3。SNMP v1，v2 有很多共同的特征，SNMP v3 在先前版本的基础上增加了安全和远程配置能力。

系统日志(Syslog)是一种在网络中传递消息的标准。它采用 Client/Server 架构：syslog 的发送者（Client）向接受者（Server）发送短消息（一般小于 1KB）。日志严重程度(severity levels)： 0 - Emergency (emerg) 1 - Alerts (alert) 2 - Critical (crit) 3 - Errors (err) 4 - Warnings (warn) 5 - Notification (notice) 6 - Information (info) 7 - Debug (debug) 。通常系统会将进程分组，同组进程的日志将具有相同的组标识（称为 facility），这样可以在一定程度上对日志分类。可以将各个节点上的 Syslog 发送到 Syslog Server 进行汇总，整理和分析。



2.13.4 错误侦测

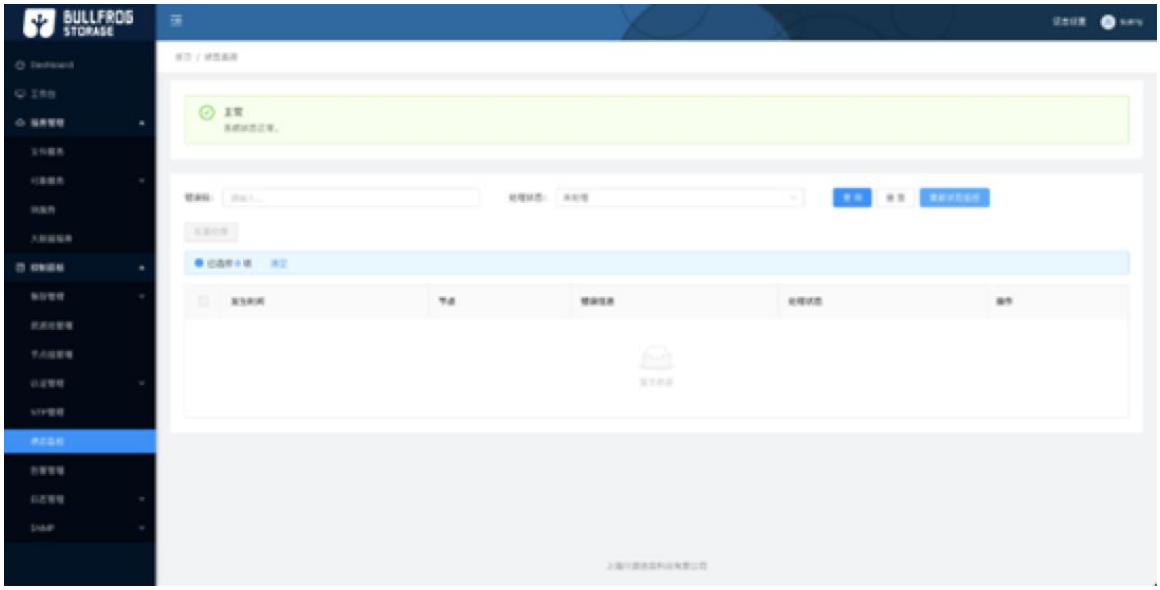
BULLFROG®支持位衰减（bit-rot）检测的错误侦测。

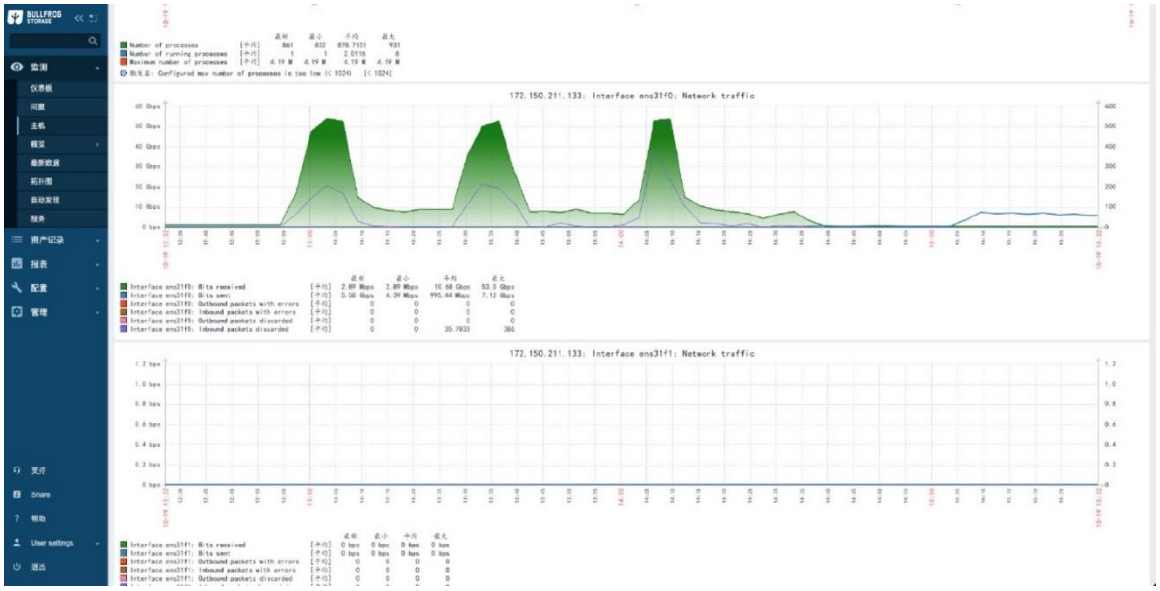
位衰减是指存储在存储介质中的数据的性能和完整性的缓慢恶化。它也被称为比特衰变、数据腐烂、数据衰变和静默数据损坏。

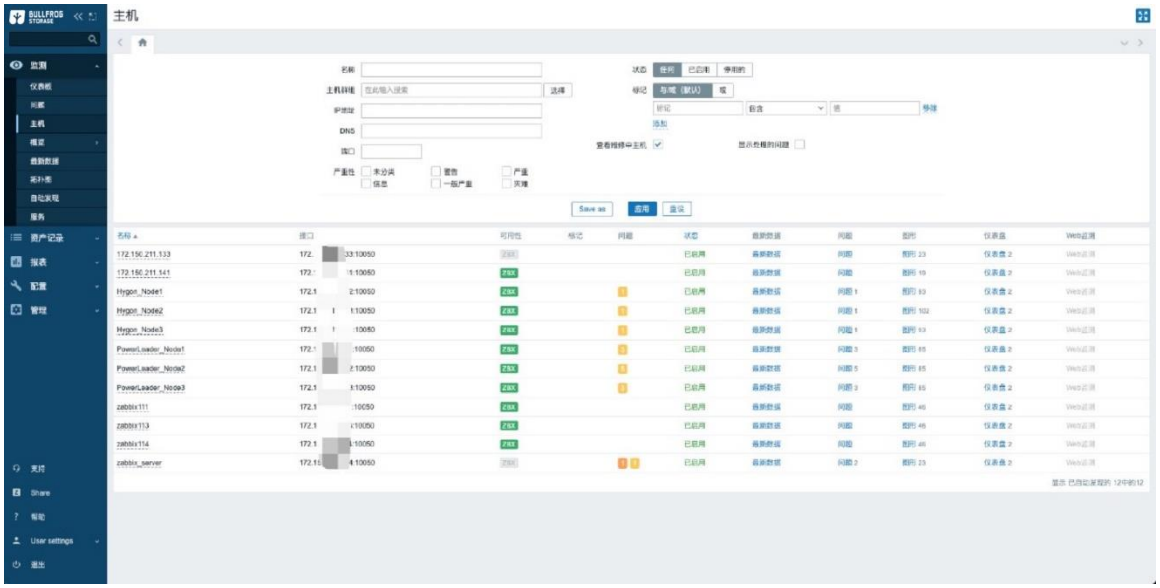
位衰减检测是一项用于定期检测磁盘内部的隐性错误的方法，相对于 RAID，BitRot 在使用 JBOD 模式时更有效。

2.13.5 数据中心统一管理

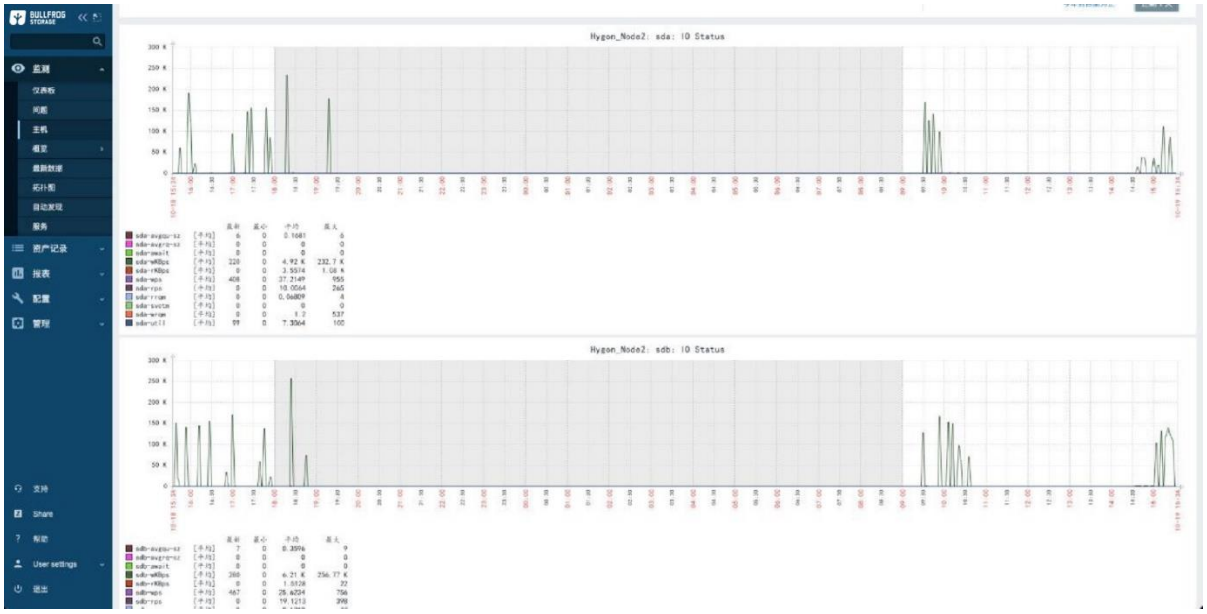
BULLFROG®分布式存储集成并实际配置数据中心监控管理模块，可纳管数据中心 内服务器，网络设备、UPS 等设备，对设备的健康状态，性能指标等实现统一监控。

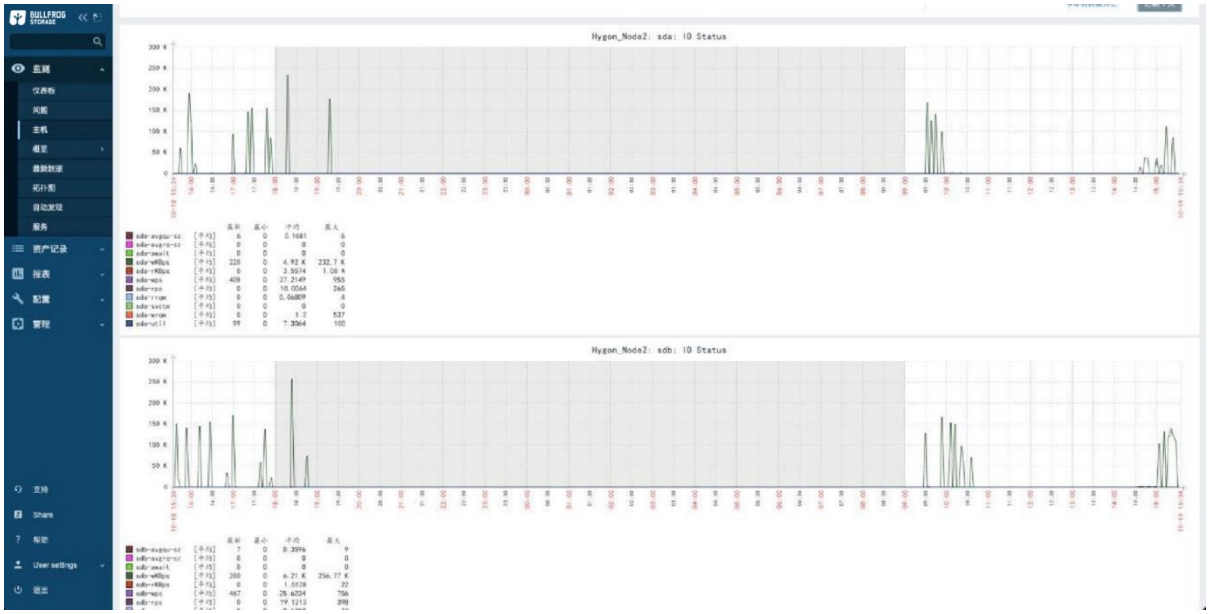






分布式存储数据中心监控管理模块具备监控数据的图形展示功能及监控指标定制化功能。





2.14 系统维护

2.14.1 故障恢复

BULLFROG®具有高可用的特性，因此某个磁盘或某个节点发生故障时，不影响整个集群的服务。BULLFROG®完善的监控机制可以实时监测到软硬件故障并提醒用户进行处理。存储系统最常出现的故障为磁盘故障。BULLFROG®的磁盘支持 RAID1/5/6/10 或者裸磁盘。当 RAID5/6/10 的磁盘发生故障时，如果不影响数据访问，可以直接采用热插拔的方式对磁盘进行替换，硬件会自动对 RAID 进行重建和数据恢复。如果裸磁盘损坏或者 RAID 无法恢复，需要手动在系统管理界面进行存储 Brick 的修复。

BULLFROG®支持副本卷和 EC 卷。在 3 副本的情况下，可以允许有两个磁盘损坏。n+m 配置的 EC 卷可以允许最多 m 个磁盘损坏而不丢失数据。BULLFROG®提供了简单易用的修复方式，用户仅仅需要按照操作指引替换新的磁盘即可以完成数据的重构。BULLFROG®还具有智能数据重构功能，在数据重构的过程中不影响系统的存储服务，在系统负荷较低的时间进行重构，不会对整个存储造成额外的压力，以致降低了系统性能。

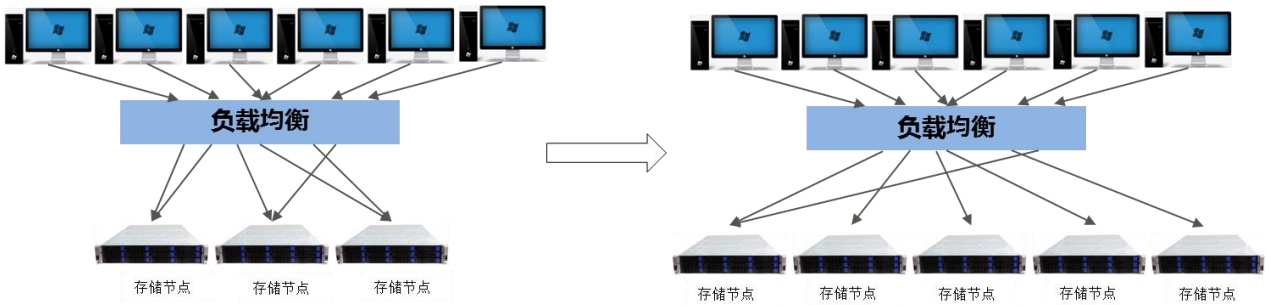
2.14.2 系统扩展

传统的存储在扩展时需要中断业务，重新配置应用程序。由于 BULLFROG®支持动态扩容，系统扩展时不需要中断系统业务，即时扩展即时使用。扩展后自动配置负载均衡。

BULLFROG®支持横向扩展和纵向扩展两种扩展方式。纵向扩展可以添加在存储节点上添加磁盘实现扩容。在不增加节点的前提下，可以利用服务器空余硬盘槽位或者附加存储硬件提供额外的硬盘空间。这是成本最低的一种扩展方式，可以充分利用原有的服务器资源，降低硬件投入。这种扩展方式可以用于对性能要求不高，但是对容量要求较高的场合。

横向扩展可以动态增加节点数量。BULLFROG®可以支持 2～288 个节点，最多可以实现 EB 级的存储。随着节点数的增加，整个集群的存储容量和吞吐量呈线性增长，提高了系统的带宽和并发数。

由于新增磁盘不会影响现有文件存储分布，新的磁盘会参与存储分布调度，不需要移动任何文件。但是负载均衡没有平滑处理，旧节点或者磁盘负载较重。BULLFROG®充分考虑到这个问题，提供了自动再平衡功能。在新建文件时会优先考虑负载最轻的节点或者磁盘，同时还可以将已经存在的文件进行再平衡，使得旧文件可以在新的磁盘或者节点上分布以实现容量负载均衡。自动再平衡功能可以降低旧节点或者磁盘的负载，提高整个系统的性能。



灵活的扩展特性可以满足用户在不同阶段对性能和容量的不同需求，按需配置，统一管理。BULLFROG®可以极大的降低用户的负担，并且在未来业务扩展时提供更高的存储空间和性能。

3 BULLFROG®分布式统一存储规格

3.1 系统规格

表3-1 BULLFROG®分布式统一存储系统规格指标

BULLFROG®分布式统一存储	
存储架构	单系统同时支持文件存储/对象存储
存储节点类型	性能型、容量型和通用型（可混用）
最小存储节点数	2（注1）
最大存储节点数	60000
前端网络类型	1GbE Ethernet或10GbE Ethernet或25GbE Ethernet或100GbE Ethernet或InfiniBand
后端网络类型	10GbE Ethernet或25GbE Ethernet或InfiniBand
数据冗余方式	多副本
数据复制方式	本地快照、远程复制
自动精简配置	支持
数据自愈功能	支持数据自动化并行快速恢复
多协议支持	NFS，CIFS，FTP，HDFS，HTTP，HTTPS，Amazon S3，SNMP，NDMP，LDAP，NIS，Microsoft Active Directory,
操作系统支持	Solaris，AIX，HP-UX，Windows，Linux，Mac OS
系统管理	Web用户界面管理控制台 支持不同级别管理用户，支持用户分权分域管理 告警支持Email，短信， Syslog 零停机扩展存储容量

	支持存储性能监控，图形化展示实时性能数据和历史性能数据		
系统维护	系统自动坏盘率检测及告警通知，坏盘可批量集中更换，无须即时更换，减少人力维护		
企业级特性	支持存储节点缓存的断电保护 支持与企业办公系统协作，提供FAAS服务		
节点类型			
存储节点类型	通用节点	容量节点	性能节点
最大容量/节点	360 TB	720 TB	43.2 TB
缓存/节点	32GB-256GB	16GB-128GB	48GB-256GB
磁盘类型	SSD/NL-SAS/SAS/SATA		
前端端口	2 × 1 GbE/2 × 10 GbE/2 × 40 GbE /2 × GbIB		
后端端口	2×10 GbE/2 × 40 GbE /2 × GbIB		

（注 1）仅文件存储服务支持 2 节点配置。对象存储服务至少需要 3 节点。

4 BULLFROG®分布式统一存储配置

4.1 配置原则

- 推荐节点个数不小于3个，可以实现完全的高可用和负载均衡
- CPU核心数不小于brick数
- 单个RAID的读写带宽接近内网带宽，消除磁盘读写瓶颈
- 硬盘推荐使用JBOD模式
- 如果对数据恢复速度高，推荐RAID5。
- 节点内存不小于32G
- 推荐配置UPS

4.2 配置介绍

存储节点类型	通用节点	容量节点	性能节点
最大容量/节点	360 TB	720 TB	43.2 TB
缓存/节点	32GB-256GB	16GB-128GB	48GB-256GB
磁盘类型	NL-SAS/SAS/SATA	SATA	SSD
前端端口	2 × 1 GbE / 2 × 10 GbE / 2 × 40 GbE / 2 × GbIB		
后端端口	2×10 GbE / 2 × 40 GbE / 2 × GbIB		

性能节点采用 PCI-e SSD 进行扩展，原始容量为 34.2 TB。

容量节点确保以高磁盘密度进行容量扩展。该存储节点的原始容量为 252.6 TB。通过在系统中混合使用基本节点、性能节点和容量节点，BULLFROG®系统可支持具有最佳配置的任何定制应用场景。

5 弹性扩展

BULLFROG®支持节点动态扩展，可配置 2-60000 个节点。

容量扩展

随着节点数的增加，存储容量线性增加。

并发能力扩展

随着节点数的增加，并发 IO 能力线性增加。

缓存扩展

采用全局缓存，随着节点数的增加，缓存也同步增加。

BULLFROG®支持业务不中断进行升级和维护。对外提供了统一的命名空间，在升级，扩容时，也能维持统一命名空间不变，扩展后能够在节点间进行自动再平衡，以保持存储系统的健壮。

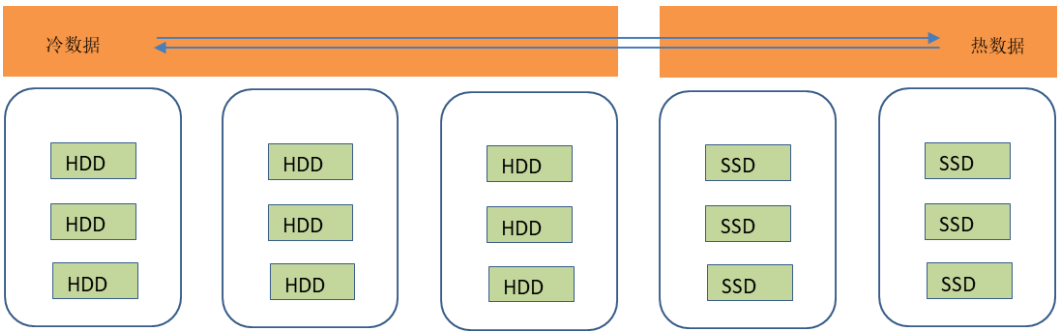
BULLFROG®采用前后端隔离的数据访问方式，并在前端网络中配置了负载均衡技术，在后端存储系统进行升级扩容时，前端网络几乎感知不到存储系统的变动。

6 高性能

相较于传统存储采用内存作为 I/O 的 cache，受制于内存的容量，缓存一般较少，BULLFROG®采用 SSD 作为缓存，以提高存储的性能。

智能分层策略基于用户的 I/O 访问频度，自动的完成冷热数据分级和数据迁移。自动分层功能能持续的监控工作负载，通过统计分析用户的活动，识别出当前的数据热点，并且把最频繁访问的数据迁移到高性能的固态磁盘上，将不常访问的数据迁移到较低性能的普通硬盘上，在迁移过程中不会产生 IO 中断或者暂停，保证用户的正常访问。通过分层，数据提升和自动再平衡可以解决性能问题，冷数据降级则可以 解决容量问题。

分级监控和识别数据的活动级别，并自动将活动和非活动数据重新平衡到最合适的存储层。在热存储层和冷存储层之间移动数据是一项计算成本很高的任务。为了解决这个问题，BULLFROG®支持在后台自动提升和降级卷内的数据，从而最大限度地减少对前端 I/O 的影响。数据的冷热取决于数据的访问速率。如果对文件的访问增加，文件将移动或保留其在热层中的位置。如果文件在一段时间内没有被访问，它将移动或保留它在冷层中的位置。



在面对对象存储海量小文件的情况下，常规的存储都会遇到极大的性能瓶颈。存储的数据量越多，性能劣化越严重，甚至造成存储停止服务。

BULLFROG®采用先进的小文件聚合方式和可扩展元数据存储服务，将海量文件聚合到单个大文件，采用 append 的方式实现写入，同时用可扩展元数据服务保证高效率多谢，使得在存储数十亿甚至百亿的文件是仍然保持高读写性能，充分利用了磁盘和网络带宽。实现低成本高性能优势。

7 系统安全

随着越来越多的存储系统采用 IP 网络进行存储内部，外部的通信，在 IP 网络上进行安全控制越来越重要。BULLFROG®在如下方面进行了安全强化。

7.1 操作系统安全

BULLFROG®采用 CentOS7.4 操作系统，会根据操作系统安全补丁和开源软件补丁定期发布安全补丁。同时针对原操作系统进行了以下基础安全配置。

- 内核裁剪和参数优化：去掉多余的内核模块。优化内核参数，更新内核补丁。
- 精简服务：最小化系统服务，降低风险。
- 强权限管理：采用严格的权限接入管理。普通用户无法登入系统，保证系统安全
- 日志审计：有详细的服务和内核日志审计功能。

7.2 网络连接安全

- BULLFROG®采用内外网隔离的测试，从物理上将内部网络和外部网络进行隔离，阻挡外部网络风险。
- 采用严格的防火墙策略，采用白名单制，只开放内网和业务网的指定端口，有效阻挡外部端口扫描和嗅探，降低网络风险。
- 业务网的接入服务采用严格的权限认证和IP验证方式，仅允许有权限和限定IP的用户进行访问。
- 对网页数据服务提供HTTPS安全访问通道，提高web服务平台的安全性。
- 防止SQL注入式攻击。

8 技术指标

本章节描述了 BULLFROG®的文件存储性能指标和对象存储性能指标和容量指标。

8.1 性能指标

8.1.1 文件存储性能指标

3 节点，万兆网络，4+2EC 30 块盘，客户端为 Linux 的情况下，顺序读写的性能指标如下：

项目	指标
CIFS IOPS（4K）读	单节点最大支持 500OP/s
CIFS IOPS（4K）写	单节点最大支持 250OP/s
CIFS（4M）稳定读	单节点最大支持 500MB/s
CIFS（4M）稳定写	单节点最大支持 600MB/s
NFS（4M）稳定读	单节点最大支持 600MB/s
NFS（4M）稳定写	单节点最大支持 650MB/s

8.1.2 对象存储性能指标

对象存储的吞吐量和 OPS 指标如下：

项目	指标
对象存储大对象(4M)读	单节点最大支持 750MB/s
对象存储大对象(4M)写	单节点最大支持 700MB/s
对象存储小对象(4K)读	单节点最大支持 6000OP/s
对象存储小对象(4K)写	单节点最大支持 2000OP/s

8.2 容量指标

容量指标如下：

项目	指标

集群最大节点数	60000
文件系统最大容量	100PB
最大文件数量（含目录）	100 亿
集群最大支持的对象存储桶数	1000 万
集群最大支持的对象数	100 亿

8.3 其它指标

其它技术指标如下：

项目	指标
1node 所需最小 CPU 核数	4
1node 所需最小内存容量	4 GB
1node 所需最小磁盘容量	20 GB(OS Disk)
1 集群最大支持的 Pool 数	72
1Pool 的最大共有数	65536
1 共有的最大容量	100 PB
1 共有同时连接用户数	每个 node 的最大连接用户数根据协议不同而区别如下。50（CIFS），50（NFS），100（FTP），500（HTTP/HTTPS）
1 集群最大用户数	1000

9 系统兼容性

协议兼容性

BULLFROG®支持 NFS，CIFS，FTP，HTTP，HTTPS，HDFS，S3/swift 等多种接口。

与云平台的兼容性

支持 Openstack Manila，Swift 等模块可以直接对接，提供 File 和 Object 存储功能。可以将私有云上的数据按照策略自动归档到公有云支持并配置。

与集中管理平台的兼容性

支持 SNMP 协议，可对接主流的网络管理平台

通用 API

BULLFROG®提供 RestAPI 接口，可对外提供特性功能的直接访问，适配。

10 术语

Brick	存储节点的最小存储单元
CIFS	Common Internet File System windows，系统在网络上进行文件共享的协议，又叫 samba
DHT	Distributed Hash Table，分布式哈希表
EC	Erasure Code，纠删码
HA	High availability，高可用性
HDD	Hard Disk Drive，硬盘驱动器，通常指机械硬盘
LB	Load Balance，负载均衡
LDAP	Lightweight Directory Access Protocol，轻量目录访问协议
NAS	Network Attached Storage，相对于 SAN，通常指文件存储，也称为网络存储器
NDMP	Network Data Management Protocol，网络数据管理协议
NFS	Network File System，即网络文件系统
NTP	Network Time Protocol，网络时间协议
RAID	Redundant Array of Independent Disks，独立冗余磁盘阵列
RDMA	Remote Direct Memory Access，远程直接数据存取
RoCE	RDMA over Converged Ethernet，是一种允许通过以太网使用远程直接内存访问(RDMA)的网络协议
SAN	Storage area network，存储区域网络
SNIA	Storage Networking Industry Association，全球网络存储工业协会
SNMP	SimpleNetwork Management Protocol，简单网络管理协议
SSD	Solid-state drive，固态硬盘
TFTP	Trivial File Transfer Protocol，简单文件传输协议