



LinePillar FS 并行文件系统

技术白皮书



版权所有©广东省绿算技术有限公司 2023

未经本公司许可，任何组织和个人不得以任何形式复制本文件所有、部分内容。

本公司保留所有解释权。

绿算技术有限公司

电话：010-52129203/52129206

地址：北京市海淀区西直门北大街甲43号金运大厦A座803室

CONTENTS

目录

1	Development Trend of Parallel File System 并行文件系统发展趋势	4
2	Introduction to LinePillar FS file system LinePillar FS 文件系统介绍	5
	Introduction to LinePillar FS	
	2.1 LinePillar FS 简介	5
	LinePillar FS composition	
	2.2 LinePillar FS 组成	6
	High reliability and dynamic online upgrade capability	
	2.3 高可靠以及动态在线升级能力	7
	Client Service, please	
	2.4 客户端服务	8
	Other advanced features	
	2.5 其他高级功能	9
3	LinePillar FS core advantages LinePillar FS 核心优势	10
	Support single node, dual node, clustered deployment	
	3.1 支持单节点、双节点、集群化部署	10
	Containerized storage of small files improves storage and retrieval efficiency	
	3.2 小文件容器化存储提高存储和检索效率	10
	Broad application scenario support	
	3.3 广泛的应用场景支持	11

1. 并行文件系统发展趋势

在市场需求背景下, 并行文件系统得到了快速发展, 已经被广泛视为是新一代的海量数据存储架构。其中集群技术应用于数据中心的服务器设备上已经有几年的历史了, 效果显著。而并行文件系统作为集群技术的发展和延伸, 具备了灵活的扩展性和极高的性价比。并行文件系统通过由若干个存储系统组成的集群, 能够让企业在兼顾性价比的同时轻松实现性能与容量的在线无缝扩展, 还有效解决了文件共享的效率问题。

未来体系结构基于这样一个假设: PB 级系统是动态的、大型系统不可避免地以增量的方式构建、节点故障是常态而不是例外、工作负载和特性随着时间的推移不断变化。该体系结构的主要目标是可伸缩性(达到数百 PB 甚至更高)、性能和可靠性。可伸缩性可以从多个方面考虑, 包括系统的总体存储容量和吞吐量, 以及单个客户端、目录或文件的性能。目标工作负载可能包括这样的极端情况, 即数万或数十万个主机同时读取或写入同一文件, 或在同一目录中创建文件。这种场景在超级计算集群上运行的科学应用程序中很常见, 越来越显示出未来的常规工作负载。更重要的是, 我们认识到分布式文件系统工作负载本质上是动态的, 随着活动应用程序和数据类型的变化, 数据和元数据访问会发生显著变化。

并行文件系统直接解决了可伸缩性问题, 同时通过三个基本设计特性实现了高性能、可靠性和可用性: 解耦数据和元数据、动态分布式元数据管理和可靠的智能分布式存储。

2. LinePillar FS 文件系统介绍

2.1 LinePillar FS 简介

绿算 LinePillar FS 是领先的基于 POSIX 文件系统接口的并行集群文件系统, 意味着不必重写或修改应用程序来利用 LinePillar FS。它的开发非常注重最大的性能和可扩展性, 高度的灵活性, 并且设计考虑了健壮性和易用性。

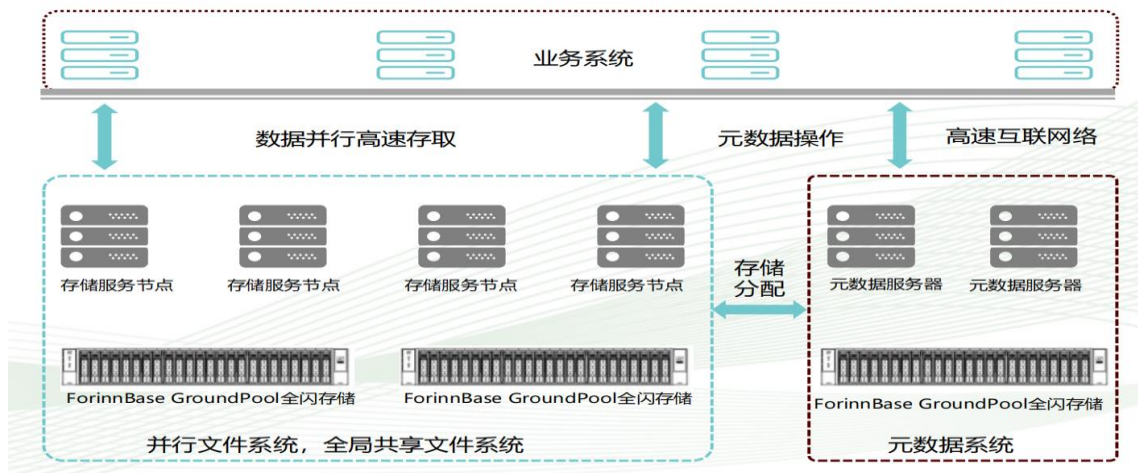
LinePillar FS 客户端通过高速网络的连接与存储服务器进行通信, 访问文件系统内的数据。LinePillar FS 通过在多个存储服务器上透明地分布用户数据, 当增加系统中服务器和硬盘的数量, 所有硬盘和所有服务器的容量和性能都聚合在一个文件系统名字空间中。从而文件系统的性能和容量就可以轻松地扩展到特定业务所需的量级, 也可以在以后系统投入生产时进行在线扩展。

LinePillar FS 文件系统的可伸缩性减少了部署许多独立文件系统的需要, 例如每个集群部署一个文件系统, 或者更糟的是每个 NFS 文件服务器部署一个文件系统。LinePillar FS 的统一名字空间将带来强大的存储管理优势, 例如避免维护在多个文件系统上暂存的数据的多个副本。

由于这个原因, 它们使用 LinePillar FS 文件系统所需的聚合存储比使用其他解决方案所需的要少得多。通过将文件系统容量与许多服务器聚合在一起, 还可以聚合 I/O 吞吐量, 并随其他服务器扩展。此外, 通过动态添加服务器, 可以在安装集群后轻松地调整吞吐量或容量。

2.2 LinePillar FS 组成

LinePillar FS 并行文件系统包含三种系统模块: 可用于访问文件系统的文件系统客户端、提供文件 I/O 服务的智能存储节点和管理文件系统中的名称和目录的元数据节点。



LinePillar FS 将用户文件块中分离出元数据: 文件块由智能存储服务器来提供, 包含用户想要存储的数据 (即用户文件内容); 而元数据是“关于数据的数据”, 例如访问权限、文件大小以及关于用户文件块如何分布在存储服务器上的信息。一旦客户端获得了特定文件或目录的元数据, 它就可以直接与智能存储服务通信来存储或检索文件块, 因此元数

据服务不再参与读或写操作。

LinePillar FS 面向所有需要大型和/或快速文件存储的应用场景。虽然 LinePillar FS 最初是为高性能计算而开发的, 但如今它几乎应用于工业和研究的所有领域, 包括但不限于: 人工智能、生命科学、石油天然气、金融、音视频媒体、影像档案。无缝可伸缩性的概念还允许快速 (但可能是不规则或不可预测的) 增长的用户很容易适应他们随着时间的推移所面临的情况。

2.3 高可靠以及动态在线升级能力

LinePillar FS 的服务器通常配备数量惊人的硬盘, 并为数十到数万个客户机提供服务。并行集群文件系统应该通过高可用性机制 (如故障转移) 完全透明地处理服务器重新启动或故障。当服务器发生故障时, 应用程序应该只感知到访问文件系统的系统调用执行的延迟。

缺少健壮故障转移机制可能会导致作业挂起或失败, 需要重新启动和群集重新启动, 这将导致严重的后果。LinePillar FS 故障转移机制提供完全透明的文件访问保障。

LinePillar FS 元数据节点配置为 Active-Active 的配对方式, 元数据在元数据节点之间实现了相互备份。假设出现某一台元数据节点宕机, 其所承担的服务则会自动切换到跟它所配对的元数据节点之上, 以确保元数据服务访问的连续性。而智能存储节点之间通常配置了多副本或者

RAID (N+M) 机制实现了跨机柜、存储节点、硬盘之间的可按需配置的冗余方式, 实现了单机最多容错四块硬盘、双机容错一个节点、集群容错四个存储节点/四个机柜等不同安全等级, 并且存储空间利用率最高可达90%以上。

2.4 客户端服务

LinePillar FS 提供了 POSIX 兼容的客户端模式, 该客户端通过 Linux 内核的虚拟文件系统接口进行本地注册, 以获得最佳性能。加载客户端时, 它将后台通过服务启动脚本启动 LinePillar FS 客户端实现文件系统挂载的方法, 它与任何其他 Linux 服务一样。

本机 LinePillar FS 客户端应该在所有应该以最大性能访问 LinePillar FS 的主机上使用。但是, 也可以通过 NFSv4 或 Samba 重新导出 LinePillar FS 挂载点, 或者使用 LinePillar FS 作为 Hadoop 的 HDFS 的替代品。LinePillar FS 还为 Windows、MacOS 提供一个原生 LinePillar FS 客户端驱动, Windows 的 LinePillar FS 客户端驱动安装并实现挂载以后, 业务可以像使用已经完成格式化的本地硬盘一样访问 LinePillar FS 的文件系统名字空间, 并且可以跟 Linux 客户端共享同一个文件系统名字空间。同时 LinePillar FS 客户端还能够以块设备的方式来给客户端提供裸设备的访问接口, 访问 LinePillar FS 服务端的块存储

的服务。

客户端运行在操作系统内核中, 以驱动的方式运行, 基本没有其他依赖包, 所以能更好的兼容各个客户端环境。另外客户端节点也可以通过 NFS、CIFS 或 Hadoop HDFS 访问 LinePillar FS。

2.5 其他高级功能

LinePillar FS 文件系统的一些附加功能如下所述。

2.5.1 访问控制列表 (ACL)

LinePillar FS 安全模型目前是 UNIX 文件系统的安全模型, 通过 POSIXACL 进行了增强。另外提供的一些额外的特性包括存储系统自身拥有的权限控制, 可以实现即使客户端使用 root 权限也无法对文件进行未授权的文件操作。

2.5.2 配额—用户和组配额

LinePillar FS 可以为目录指定一个属性, 以便该目录下的所有文件 (以及任何子目录下的递归文件) 都具有由该属性确定的容量配额。支持对目录、用户、用户组设置容量、目录数、文件数配额, 配额对所有访问接口有效, 并可在线设定、更改和取消, 配额实时性为秒级, 支持配额嵌套, 并提供配额预警功能。

2.5.3 快照克隆

LinePillar FS 提供了快照技术, 支持文件创建快照和克隆, 使用了

ROW 的方式, 尽可能减少快照的资源消耗。

快照功能可为文件或者卷创建多达无限数量的基于增量的历史时间点拷贝。当发生数据“软”故障, 比如软件程序导致的数据损坏、病毒破坏、意外删除等, 可以通过对合适的时间点标记进行“回滚”来快速恢复数据。该功能特别适用于关键性业务的连续数据保护。

3. LinePillar FS 核心优势

3.1 支持单节点、双节点、集群化部署

LinePillar FS 初始配置一台存储节点(智能存储节点)、两台存储节点即可实际上线使用, 根据需要安装指定数量的硬盘。在初期部署双台存储节点的情况下, 后续支持动态增加硬盘或者增加存储节点来扩展存储容量, 容量区间可以从数 TB 按需扩展到数百 PB, 无需停机。

LinePillar FS 系统的安全性和读写性能可以随着硬盘数量的增多而随之提升。LinePillar FS 系统更小化的初始部署能力, 方便能够让并行文件系统从业务发展之初就进入到用户的业务系统中, 并让文件系统能够持续在线扩展来支持业务的发展壮大所带来的高性能和大容量需求。

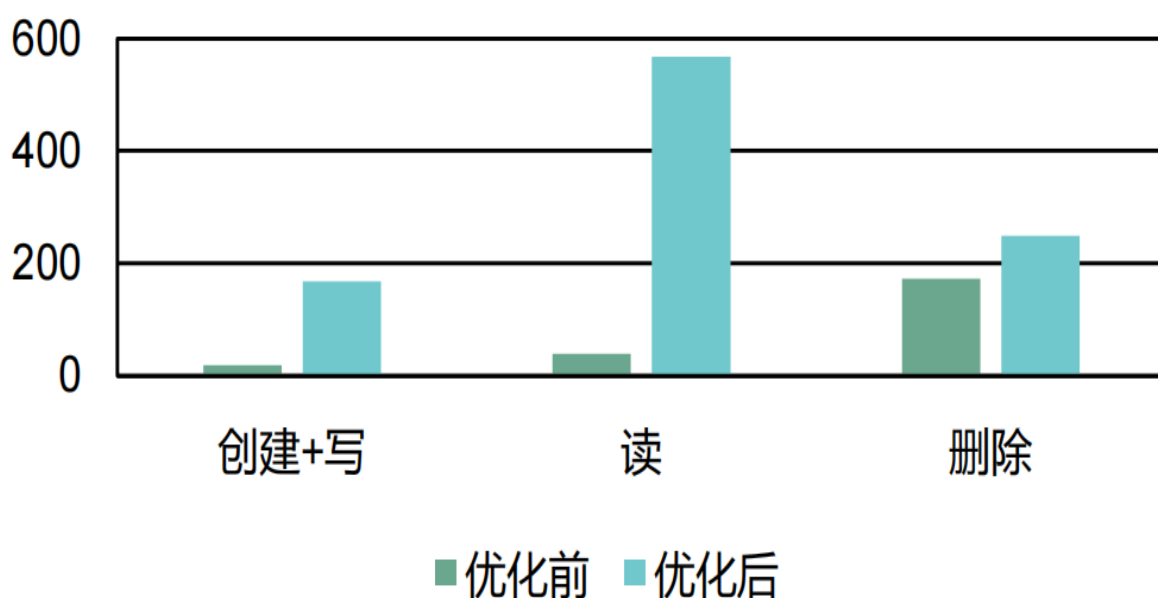
3.2 小文件容器化存储提高存储和检索效率

LinePillar FS 支持百亿级文件统一存储和高效率检索。LinePillar

FS 内部的系统采用了创新的小文件 Container 容器存储技术, 能够在前端呈现标准 POSIX 文件系统名字空间的同时, 在后端小文件落盘存储的时候实现智能的聚合成大文件的方式来实际存储。

LinePillar FS 通过采用该技术将任何大文件、小文件都会被当做大文件一样的来存储和检索访问, 完全消除了传统文件系统文件存储数量受限、文件检索效率低下等严重的数量和效率的问题, 打破了传统存储的局限, 将小文件创建、检索效率提高到传统存储的十倍以上, 可以轻松应对任何存在海量小文件的业务场景。

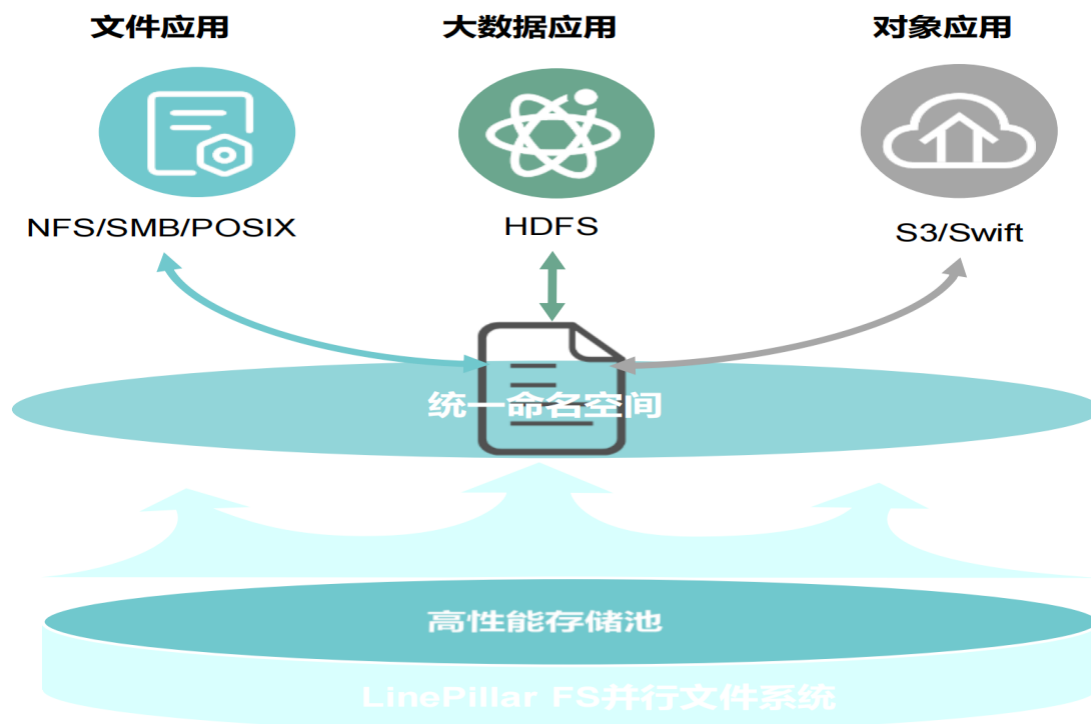
单客户端5万个4K文件, 32个并发IOPS性能



3.3 广泛的应用场景支持

LinePillar FS 能够基于创新的 RAID 冗余技术, 消除设备级别的故障影响, 并且能够同时提供文件系统、块设备、对象存储等全面的存储接

口协议, 提供更高的并发读写效率, 从而能够在同一套系统中按需分配容量、接口协议等来支持广泛的业务应用场景。





E-Mail: ls@luisuantech.com

联系电话: 010-52129203/010-52129206

运营中心: 北京市海淀区西直门北大街甲43号金运大厦A座803室

本文件内容解释权归绿算技术有限公司所有