

# JWanFS 广域网分布式存储 技术白皮书



[www.jwanfs.com](http://www.jwanfs.com) 技术网站支持

2023-06

## 目录

一、 前言 .....	3
二、 基本概念 .....	4
三、 JWanFS 广域网分布式存储系统介绍 .....	5
1. 简介 .....	5
2. 特点 .....	5
3. JWanFS 面临的挑战 .....	6
4. JWanFS 的优势 .....	6
5. JWanFS 解决的技术痛点 .....	7
四、 技术架构 .....	7
1. 核心设计目标 .....	7
2. 层级结构 .....	8
3. 数据读写流程 .....	8
4. 数据写入 .....	9
5. 数据读取 .....	11
6. 数据安全 .....	11
7. 多协议互通兼容 .....	12
8. 传输节点互联 .....	13
五、 总结和展望 .....	14

# 一、前言

数字化经济时代，算力需求呈爆发性增长，逐步成为新时代的核心生产力。随之带来的是数据量的不断增加和数据存储需求的不断变化，传统以本地存储技术的低扩展性和高扩展成本难以满足不断增长的数据存储需求，为此构建可扩展、易用、安全的跨区域空间、跨存储极限的广域网分布式新型云存储基础设施，成为算力驱动和演进的一个关键环节。

JWanFS 广域网分布式存储系统是一款跨广域网、跨多个公有云的分布式文件存储系统，它可以提供高可扩展、高可靠和高安全的文件存储服务，可以轻松扩容数据存储容量至 ZB 级，同时 JWanFS 还支持 S3、FTP、WebDAV、NFS、XRrootD 等多种数据访问接口，提供了本地挂载工具，允许用户读写 JWanFS 就像使用本地文件系统一样简单。

除了上述提到的 JWanFS 易扩展、易使用的技术优势，JWanFS 的多数据中心架构还具有更高的容灾能力。JWanFS 会将存储数据通过副本形式保存至多个数据中心，可以避免单一数据中心故障时导致服务不可用的问题，或因不可抗灾难导致的数据丢失。

本白皮书将详细介绍 JWanFS 广域网分布式存储系统的特点、挑战、优势以及技术架构。

## 二、基本概念

名词	概念
Object	对象，用户访问 JWanFS 的基本单位
Bucket	桶，用户的存储命名空间，每个桶之间的数据相互隔离，一个 Object 归属于一个 Bucket
Blob	二进制大对象，JWanFS 数据传输的基本单位，为 Object 的加密数据
Chunk	JWanFS 数据存储的基本单元，一个 Blob 由一个或多个 Chunk 组成。Chunk 默认大小为 5MB
Replica	副本，一个 Chunk 可以有多个副本
Volume	存储卷，多个 Chunk 组成一个 Volume，单个 Volume 最大为 10GB
DataNode	存储节点，多个 Volume 组成一个 DataNode，便于对 Volume 数据进行管理
DataCenter	数据中心，用于管理集群拓扑结构，分配可使用 Volume 服务
FileGateway	文件网关，提供访问协议的转换，提供用户不同协议的访问接口

JWanFS 基于 SeaweedFS 项目开发，部分逻辑架构参考 SeaweedFS，感谢 SeaweedFS 项目带来的启发。

## 三、JWanFS 广域网分布式存储系统介绍

### 1. 简介

JWanFS 广域网分布式存储系统是一款跨广域网、跨多个公有云的分布式文件存储系统，也是一款企业级私有云产品。基于先进的存储技术，它可以提供高可扩展、高可靠、和高安全的文件存储服务，并且提供用户客户端（命令行、Web 端），支持文件的上传、下载、搜索，以及其它管理功能，同时提供 SDK、API，满足用户应用接口调用。此外，JWanFS 能够协助企业建立规范的文件管理体系，支持多租户的管理，实现企业数据异地存储和备份，保障数据安全可靠。使用 JWanFS 企业可以轻松在多个公有云之上，使用公有广域网网络搭建多地分公司的数据共享文件管理系统，有效实现企业信息资源的统一高效管理，从而提高企业办公效率和运营效果。

JWanFS 能够提供跨地域等多种冗余能力的存储类型，在某些意外丢失部分数据的情况下仍能通过冗余副本数据来查找并恢复完整数据。同时，由于多份数据存放在不同区域的数据中心上，能避免一个地区的存储受到不可抗逆灾难时造成的损失。

### 2. 特点

- ◇ 实现跨地域容灾备份和保证数据一致性，实现完善的用户鉴权，身份认证，提供高安全、高可靠、高可用服务。
- ◇ 提供用户客户端（命令行、Web 端），支持文件的上传、下载、搜索，以及其它管理功能，同时提供 SDK、API，满足用户应用接口调用。
- ◇ 支持多模态类型数据的存储、分享及使用，提供用户视频、音频、文档、图片、工业数据、IOT 等多种应用业务数据的平台存储，满足用户不同应用业务统一存储需求。
- ◇ 支持多种数据访问接口协议，兼容 S3、FTP、WebDAV、NFS、XRrootD 等协议。支持多协议的数据统一视图查看，满足企业数据跨平台、跨异构系统之间的数据共享需求。
- ◇ 提供用户 Web 界面监控管理，轻松了解数据上传下载读写情况和占用空间等信息，同时也支持用户按需扩容空间容量等功能。

- ◇ 数据加密传输、加密存储等多种安全保护机制保障数据不被泄露或损坏，确保用户数据跨广域网读写、存储、传输的安全性，满足科研机构 and 保密单位对异地数据访问的高安全性要求。
- ◇ 全网存储统一调度，统一分配。当一处存储节点空间不足时会自动调用其他存储节点空间进行用户数据存储，自动适配用户各地存储容量不平衡配置的物理环境约束限制，实现平滑突破空间极限。

### 3. JWanFS 面临的挑战

- 需要存储海量数据，但部分存储节点可用存储容量不足。
- 服务节点位于全国各地，数据需要安全跨地域传输。
- 数据需要在各节点间进行流转，在其中一个节点写入数据其他异地节点可以立刻读取到数据。
- 部分节点网络安全性要求高，对公网用户数据存储、应用访问仅提供单一链路单一端口，链路条件苛刻。
- 部分节点无公网地址，但又希望作为存储节点提供其他数据中心访问。

### 4. JWanFS 的优势

- ◇ JWanFS 存储架构将各节点存储卷组合为大存储池，支持横向扩展，当一个节点存储容量不足时会自动调度其他节点的存储卷，打破存储极限，轻松存储海量数据。
- ◇ JWanFS 各节点间通过传输节点互联，数据在公网间传输只需要开放各传输节点，设置节点 IP 为白名单即可实现各节点间数据在公网上安全传输。
- ◇ JWanFS 传输数据时采用 AES 加密，进一步保障数据安全。
- ◇ JWanFS 优秀的读写架构允许跨地域读写数据实时同步，当其中一节点写入数据时，另一异地节点可以立即读取，无需等待异步复制同步。
- ◇ JWanFS 提供 AWS S3 兼容 API，支持使用第三方工具连接访问。
- ◇ JWanFS-Client 只需连接集群中的文件网关即可读写数据，且支持多协议（FTP、WebDAV、NFS、XRootD）兼容访问，统一视图管理，满足各种读写需求。

## 5. JWanFS 解决的技术痛点

JWanFS 解决了传统数据中心单一存储扩展难、无异地灾备、协议数据不互通的技术痛点。

### 1. 广域网存储扩展能力

现有单一数据中心、本地化的集群存储很难满足东数西算、人工智能、AIGC、AI 大模型、IOT 等互联网业务对跨广域网存储扩展能力的需求。JWanFS 存储架构将各节点存储卷组合为大存储池，支持横向扩展，当一个节点存储容量不足时会自动调度其他节点的存储卷，打破存储极限，轻松存储海量数据。

### 2. 数据中心异地备份

传统的数据中心往往只能在本地机房进行备份，这意味着如果该地点出现故障，数据将会丢失。而 JWanFS 多数据中心架构可以解决这个问题，使得数据备份分散在多个地点，提高了数据的可靠性。

### 3. 满足多样化的存储需求

目前单一的文件存储、对象存储、块存储服务，不能很好的满足业务多样化的存储需求。JWanFS 广域网分布式存储能够满足文件存储、对象存储、块存储混合存储，满足企业在多种业务、多协议的访问需求。

## 四、技术架构

### 1. 核心设计目标

#### 高可靠

采用多副本技术，至少保证在一台机器或一个机房不可用的情况下，数据仍不丢失。

#### 高可用

为上层 PaaS 层提供统一的存储，赋能上层在迁移时的零数据丢失以及秒级切换能力，支持跨地域部署，提供跨地域容灾能力，在单节点机器故障时可以提供无感知的故障转移能力。

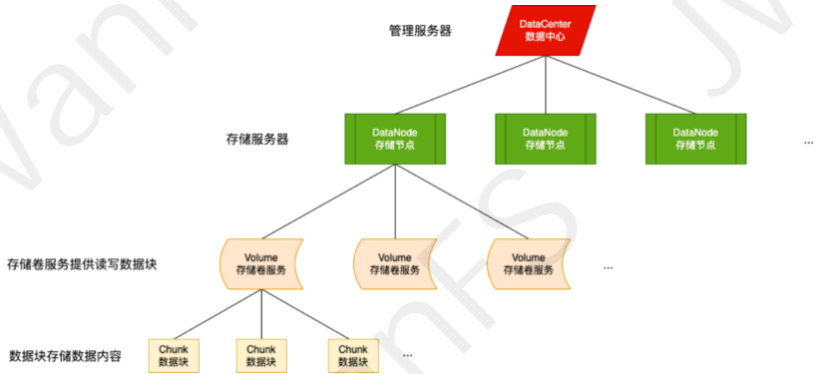
高可扩展

集群规模可扩展至数万节点，元数据节点支持水平扩展。

高性能

数据读写请求拆分、策略优化，满足高可用、高可靠的同时提高数据读写性能。

## 2. 层级结构



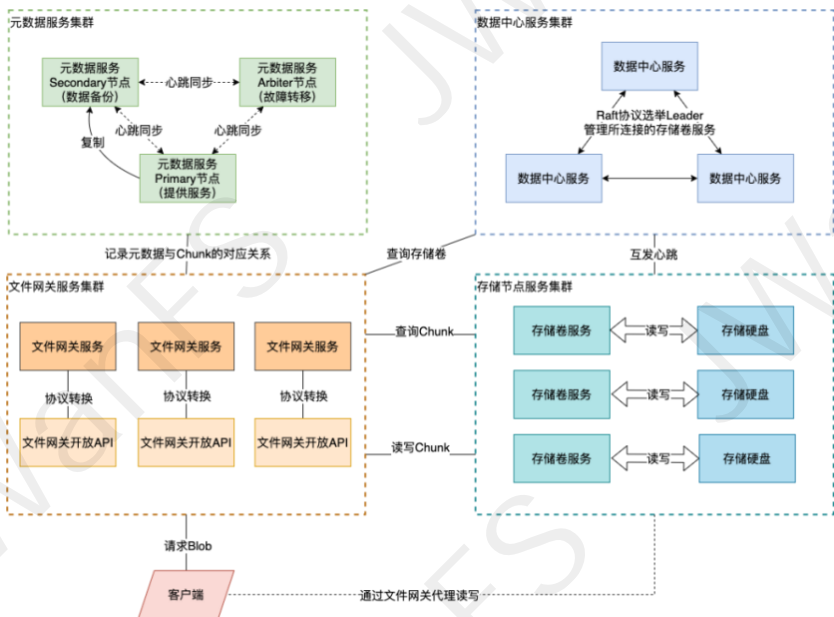
图中展示的为 JWanFS 服务的层级结构，由 DataCenter（数据中心服务）管理多个 DataNode（存储服务节点），每个 DataNode 下运行多个 Volume（存储卷服务），每个 Volume 服务保存 Chunk（数据块）。

Chunk 的 ID 由 {VolumeID}:{DataID} 构成，每份数据都有唯一的 DataID，DataID 由 DataCenter 统一分配，采用 Snowflake 算法实现分布式唯一 ID，多 DataCenter 下采用 Raft 协议选举 Leader 进行维护 ID 保证 ID 唯一性。

Volume 服务会定期将自己所持有的 Chunk 中的 VolumeID 发送给 DataCenter 服务，DataCenter 只维护 VolumeID 与 Volume 服务的拓扑信息，具体 Chunk 的信息由 Volume 服务维护。

## 3. 数据读写流程



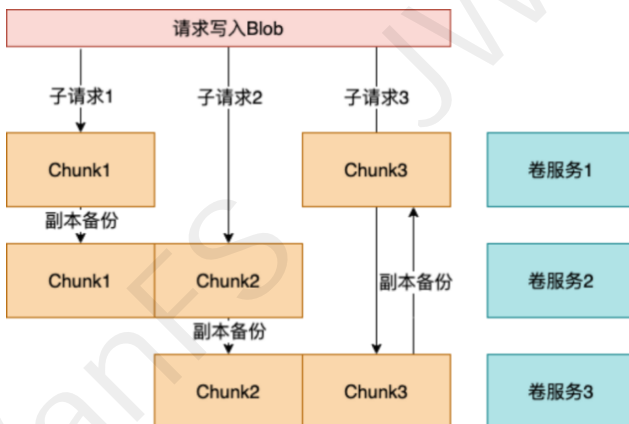


Client（客户端）读取时先访问 FileGateway（文件网关集群），FileGateway 会先从 MongoDB（元数据服务集群）获取文件的 Chunk 信息，接着请求 DataCenter 查询 Chunk 所属 Volume 服务信息，再对 Volume 服务进行读取数据。

Client 写入数据时先访问 FileGateway，由 FileGateway 向 DataCenter 请求获取可读写的 Volume，DataCenter 根据所管理的 Volume 情况，分配唯一 ChunkID 与可提供写入的 Volume，最后可以由 FileGateway 或 Client 对 Volume 进行实际数据写入，写入结束后将元数据保存至 MongoDB 中。

Client 支持命令行与 API 调用。

## 4. 数据写入



写入请求会根据 Blob 大小（如 20MB）拆分为多个 Chunk（4 个 5MB 的 Chunk）同时进行多次子请求写入，写入子请求可以请求不同 Volume 服务同时进行写入 Chunk。

Chunk 会根据写入时的备份策略进行备份，为了数据一致性，直到全部副本备份结束时才会返回客户端写入请求结束。

副本策略定义为 xyz 有如下定义：

参数	含义
x	其他数据中心的副本数量
y	相同数据中心，其他存储节点的副本数量
z	相同存储节点，不同存储服务的副本数量

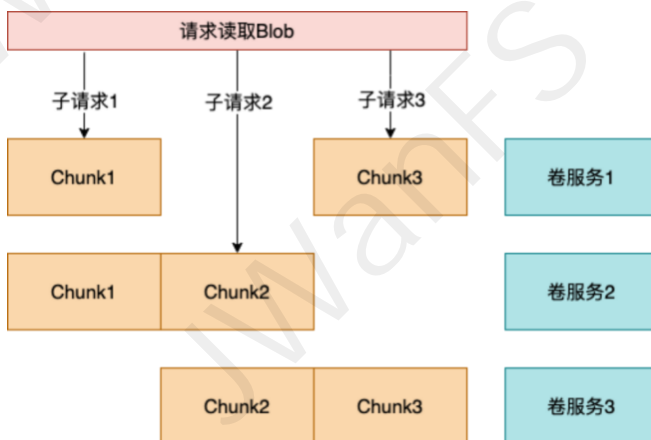
xyz 的每个值为 0-9 的数字，每种副本类型将创建  $x+y+z+1$  个副本。

可参考如下配置：

xyz 值	含义
000	只存储一份，没有备份

001	相同的数据中心，相同的存储节点复制一份，存储 2 份
010	相同的数据中心，不同的存储节点复制一份，存储 2 份
100	选择其它不同的数据中心复制一份，存储 2 份
200	选择其它不同的 2 个数据中心各复制一份，存储 3 份
110	不同的数据中心复制一份，不同的存储节点复制一份，存储 3 份

## 5. 数据读取



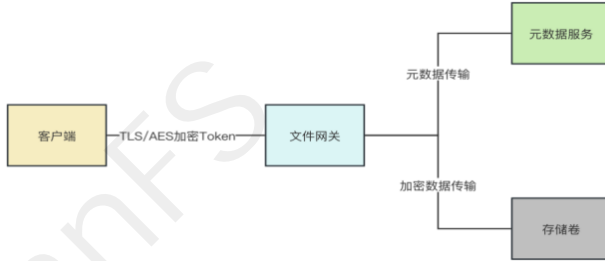
读取请求会根据请求对象进行拆分的 Chunk 进行多次子请求读取。

每次子请求对 Volume 服务都是平级的，且会根据请求中标识的数据中心存储节点信息进行读取，优先读取存储于自己数据中心内部的存储卷。

如果发现读取请求失败，则会自动重新定向至下一个副本地址进行请求。

## 6. 数据安全

数据安全主要分为两个方面，一个是面向用户的文件网关部分，一个是面向集群内部的数据存储部分。

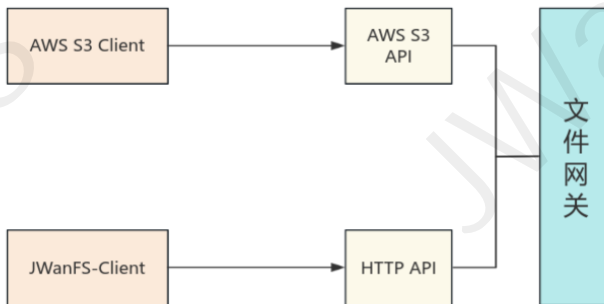


- FileGateway（文件网关）服务提供 HTTP 协议的 API，提供 Web 服务上传下载使用，为了防止接口被捕获进行非法请求，我们对请求接口参数加入了基于 AES 算法加密的 Token，Token 中包含时效性、用户操作等认证，只有合法的 Token 文件网关才会进行对应的服务处理。同时支持启动 TLS 加密传输。

- Volume 服务采用 HTTP 作为数据传输协议，传输的数据为加密压缩后的二进制数据，用户不能直接读取传输的数据内容。加密数据会直接存储至 Volume 中，用户无法直接从 Volume 获取到文件数据。

## 7. 多协议互通兼容

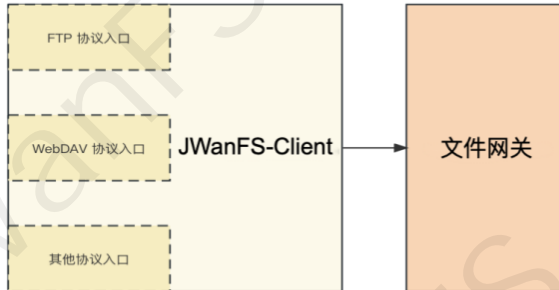
### AWS S3 应用



JWanFS 提供 AWS S3 兼容 API，兼容市面上广泛的 AWS S3 应用连接。

## JWanFS-Client

通过文件网关 HTTP API 进行连接访问。

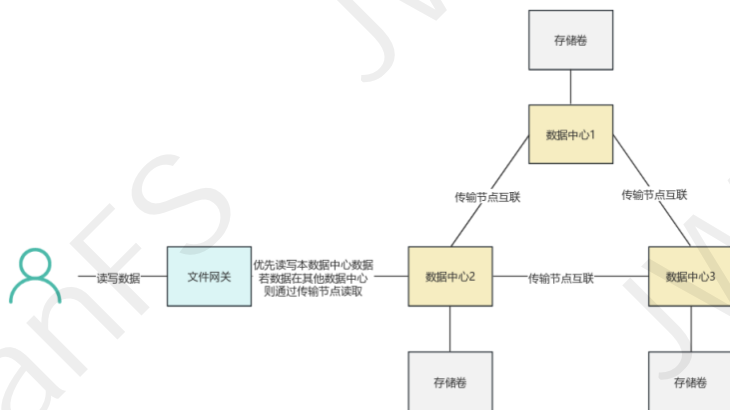


JWanFS-Client 将对文件系统的操作接口进行抽象，添加协议转换（协议解析与接口转换）层用于适配不同访问协议。用户通过简单配置即可以从 FTP、WebDAV、其他协议等读写 JWanFS 的数据，无需繁琐地重新编码对接。

JWanFS-Client 协议转换层的协议转换都是在内存中进行，且不需要依赖启动其他协议服务端，减少了一定的资源消耗，同时由于协议转换层存在，用户对数据的操作也会实时同步至其他协议。

## 8. 传输节点互联

为了适配一些高密企业数据公网传输的需求，JWanFS 允许将全部跨数据中心传输的数据通过传输节点进行传输，各数据中心只需对指向 IP 开放 1 个端口即可实现数据跨数据中心传输，大大提高了数据传输的安全性。



跨数据中心的传输调度由传输节点控制，通过集群内部域名解释实现。JWanFS 服务内部域名由以下规则构成：

【主机 ID】.【数据中心 ID】.jwanfs.server

传输时遵循同数据中心内部直连存储卷，不同数据中心通过传输节点代理访问其他节点存储卷。

## 五、总结和展望

随着未来数字经济发展，算力网络宏观战略日益落实，东数西算、大模型等新兴应用场景与算力需求形成了交替驱动的螺旋上升趋势。同时，算力需求的膨胀带动了存储容量和需求的直线上升，未来数据中心存储随着云计算的发展将占据更多的市场份额，这都给存储技术的发展提供了新的机遇。

JWanFS 广域网分布式存储系统是一个新型的分布式存储，旨在突破存储极限，以满足日益增长的存储需求。它将传统的存储功能与异地数据备份、业务高并发技术、数据安全存储等先进技术相结合，实现了存储服务的高度集成和智能化，极大地提高了存储服务的安全性和可扩展性，同时也降低了存储管理的复杂性和成本。

随着计算服务的发展，JWanFS 广域网分布式存储系统将面临以下挑战和机遇：

### 1. 网络需求高

JWanFS 广域网分布式存储系统支持用户数据跨广域网、公有云传输，网络传输速率会极大影响用户使用体验。

### 2. 技术创新

未来 JWanFS 广域网分布式存储系统需要不断创新，以满足不断变化的需求。通过引入更先进的技术，实现更智能化的数据存储和调度。

### 3. 安全与隐私

随着网络技术的发展，广域网分布式存储服务将面临更多的来自网络的安全威胁。JWanFS 广域网分布式存储系统需要加强安全防护能力，确保数据和系统的安全。

总的来说，JWanFS 广域网分布式存储系统作为新兴的存储技术，具有巨大的市场潜力和发展前景。随着数字化经济的快速发展，它将迎来更多的机遇和挑战。为了保持领先地位并满足不断变化的市场需求，JWanFS 广域网分布式存储系统将积极应对技术创新、网络需求和安全隐私等方面的挑战，不断创新和完善自身技术和服务。同时，也需要与各行业合作，共同探索更加广泛的应用场景，推动数字化经济的发展。