**Glusterfs 客户端增加修复概要设计**

|  |  |  |  |
| --- | --- | --- | --- |
| 文 档 编 号 | NxSDS | 保 密 等 级 | 内部 |
| 作 者 | 林世跃 | 最后修改日期 | 2020-08-31 |

[文档修订记录 3](#_Toc17343)

[1. 简介 3](#_Toc29949)

[1.1. 功能背景概述 3](#_Toc14982)

[2. 总体设计 3](#_Toc23504)

[2.1. 需求规定 3](#_Toc26930)

[2.2. 运行环境 3](#_Toc22649)

[2.3. 设计思路 3](#_Toc17101)

[2.3.1配置方式 4](#_Toc8198)

[2.3.2 使用方式 7](#_Toc9648)

# 文档修订记录

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 修订内容简述 | 修订日期 | 作者 |
| V1.0 | 初稿 | 2020-08-7 | 林世跃 |
| V1.1 | 修订版 | 2020-08-26 | 张凯敏 |
| V1.2 | 修订版 | 2020-08-31 | 张凯敏 |

# 简介

## 功能背景概述

纠删码需要cpu计算能力，纠删码消耗的是计算能力更多，在一些纠删码引擎是需要一些硬件指令集加速的，如glusterfs使用到的纠删码引擎需要avx，sse等指令集支持。在arm上，cpu很弱，纠删码的编码解码速率极低，如385和A72上编码解码的速率是3m/S，远远不能保证glusterfs在arm上修复速率的保证，所以设计一套程序，在x86等高性能的机器上做glusterfs的纠删码修复

# 总体设计

## 需求规定

Glusterfs fuse挂载客户端记录读写时候错误的文件和全盘扫描，采用gfid条目形式，因为glusterfs本身自带这样的代码流程，设计实现方便也减少代码bug。mount挂载之后默认开启，服务运行在fuse进程中，定时执行检测扫描。

1. 客户端修复 【index】
2. 客户端全量修复【full】

## 运行环境

X86环境，跟随fuse客户端。

## 设计思路

1. 写入失败文件使用glus自身记录（brick/indices/xattrop），修复调用rpc访问brick上记录的错误文件条目，组织修复接口所需要的数据结构，开始修复启动.
2. 调用远程接口，定时访问远程brick上所有条目，及时发现后台误删或全盘故障等场景，开始进行修复启动
3. 原生shd修复停止【gluster volume heal {volname} disable 】

## 2.4 详细设计

1. **记录写入错误文件**

Ec\_shd\_index\_inode访问glus自身记录，修复调用rpc访问brick上记录的错误文件条目。

1. **修复损坏文件**

在原有的xlator结构中插入功能，本功能只针对与ec模式，在ec层 notify函数中调用gf\_thread\_create，获取集群brick状态和卷状态，单独启用线程处理修复服务，分别启动shdfuse\_ec\_shd\_index\_healer和shdfuse\_ec\_shd\_full\_healer,此为修复总入口，调用shdfuse\_ec\_shd\_index\_healer(根据ec\_shd\_index\_healer函数接口)和shdfuse\_ec\_shd\_full\_healer(根据ec\_shd\_full\_healer函数接口),实现间隔时间和前置条件处理，以xl\_up\_count > fragments为判断条件，index调用shdfuse\_ec\_shd\_index\_sweep，构建ec\_shd\_index\_heal所需要的结构体数据传入，最终执行修复；full调用ec\_shd\_full\_sweep来执行最终修复。