#### glus纠删码修复功能迁移客户端设计文档

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **描述** | **作者** |
| 2020.08.07 | V1.0 | 《glus纠删码修复功能迁移客户端设计文档》 | 张凯敏 |
|  |  |  |  |
|  |  |  |  |

目录

[1. 背景说明 2](#_Toc29508)

[2. 方案设计 2](#_Toc834)

[2.1 修复说明 2](#_Toc14423)

[2.2 详细设计 2](#_Toc27421)

[2.2.1 设计思路 2](#_Toc6445)

[2.2.2 代码细节 2](#_Toc18054)

[2.2.3 方案细节 2](#_Toc14293)

1. 背景说明

Glus目前的应用场景主要是在arm服务器上，数据冗余的方式有纠删和副本两种选择，当数据发生故障需要修复时副本有较好的性能（25MB/s左右）可以满足场景；但是纠删需要大量的纠删码计算，由于arm的内存和cpu性能瓶颈修复数据性能较差（2MB/s-4MB/s），现在使用的磁盘有上T的容量，磁盘损坏修复太耗时间，因此需要考虑重新设计一种数据修复方案来解决此问题。

1. 方案设计

2.1 修复说明

glus集群本身修复功能是专设一个独立的shd服务来完成，分为index和full两种修复数据模式：

index是针对写入过程中磁盘故障或掉线，集群会写入表中，等到磁盘正常之后通过查表来进行修复，具有针对性，不会修复全量数据，

Full是针对已写入的正常数据因为磁盘故障导致的数据丢失，等磁盘更换后会全量修复，具有耗时久，完善度高的特点

2.2 详细设计

2.2.1 设计思路

我们需要的是修复动作由客户端来完成，计划数据检测，命令传达，集群信息校验还是借用原来的模块，再准备工作完成后具体开始修复时将其指向客户端节点，此处需要我们在每个server上提供配置文件，这样集群server任一节点均会判断此条件。

2.2.2 代码细节

在ec-heald.c中，Ec\_heal\_op操作被调用后，由给定操作来调用ec\_shd\_index\_healer和ec\_shd\_full\_healer，以index\_healer为例，syncop\_lookup -> syncop\_opendir -> syncop\_readdir(获取到xattrop下的文件)-> syncop\_getxattr，之后开始syncop\_mt\_dir\_scan ->ec\_shd\_index\_heal，每个运行有shd服务的节点均由此流程，由此接下会获取待读写文件的inodelk，此处可以由人工干预

2.2.3 方案细节

A. 新增配置文件/var/lib/glusterd/client.conf，配置好客户端节点ip；

B. 将客户端节点也添加到集群中来（会自动拉起shd服务），这样client节点也具有了恢复集群数据的前置条件

C. 在代码细节中添加筛选流程，根据配置文件选择是否由此节点继续进行修复工作。