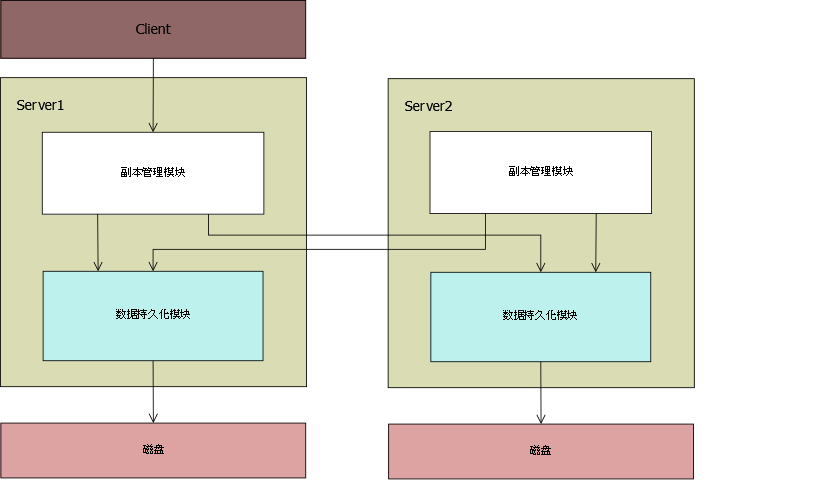
BRFS数据写入处理结构

1、**模块划分**

（1）数据持久化模块。

（2）数据副本管理模块。

如下图所示：



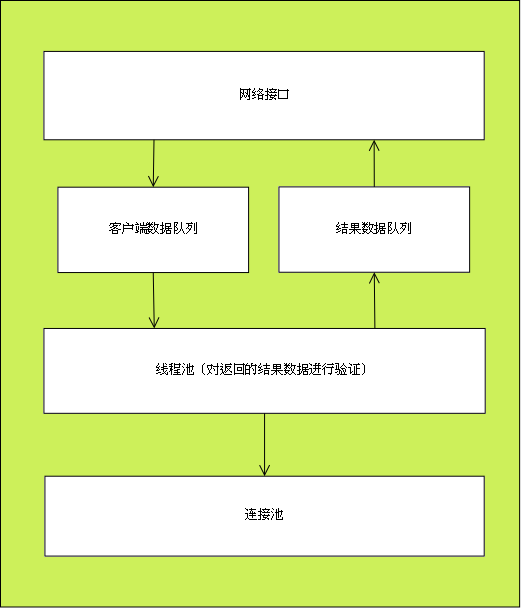
其中带箭头的线条表示数据流动方向。

Client端连接Server中的副本管理提供的网络接口，然后就可以传输客户端数据。副本管理模块会数据所属的StorageName的副本设置，连接特定数目的数据持久化模块，

连接成功后，副本管理模块会向其连接的每个数据持久化模块发送一份从客户端接受到的数据。持久化模块会向副本管理模块返回数据的存储信息，并且把数据写入到磁盘。

2、**模块内结构**

（1）副本管理模块



1、网络接口

提供客户端传输数据的入口。

2、队列

分为客户端数据队列和结果数据队列

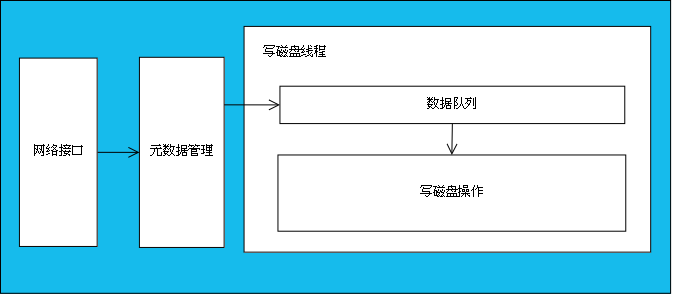
客户端数据队列保存来自客户端的数据，从网络接口接受到的数据会先保存到客户端数据队列。而客户端数据相应的存储信息保存到结果数据队列中。

3、处理线程池

线程池中的线程从客户端队列中取数据，通过网络连接发送到持久化模块（本地可以直接通过调用类接口实现），然后把得到的返回数据存储到结果队列中。对于多个副本的数据，必须每个副本返回的结果都相同时才算处理成功。（这里之所以用线程池，是因为线程中的操作是无状态的。）

4、副本管理模块对持久化模块的连接可以通过一个连接池来管理，因为多个StorageName的副本管理可以重用一个连接。

（2）数据持久化模块



1、网络接口

提供副本管理模块传输数据的入口。

2、元数据管理

维护磁盘文件大小信息，及管理磁盘文件的切换等操作，客户端数据传输到这就可以获取到其存储信息，即大文件名、偏移量，把这些信息回传到副本管理模块后，就把客户端数据放入数据队列。

3、数据队列

队列里保存元数据管理模块处理过的数据，这些数据就只是单纯的等待被写入到磁盘，

因为元数据管理模块维护这磁盘文件的大小信息，所以当一个文件无法容纳下一条数据时，它会在队列中插入一个特殊的节点，表示是时候换个文件了。

4、写磁盘

这一步就是简单的从队列中取出数据并且写入到磁盘上。

（3）一些补充

1、副本管理模块中的所有子模块可以被所有StorageName共用，而持久化模块中从元数据管理到写磁盘这个流程上的所有节点，每个StorageName都会单独的有一组。

2、不同副本所在Server之间的磁盘文件名同步问题，暂时考虑用ZK实现。

最后附上整体的详细结构图：

