|  |
| --- |
| 技术选型与项目实现分析文档 |
| "Terra快乐分享"实现分析  创建日期：2013/02/01  创建人：梁志峰 |

修订历史：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | 修订者 | 修订日期 | 备注 | 审批日期 |
| V1.0 | 梁志峰 | 2012/03/01 | 创建原始版本 |  |
|  |  |  |  |  |
|  |  |  |  |  |

目录

引言 3

1．1技术选型 3

难点需求 4

2. 1文件上传 4

2. 1 文件下载

2. 2文件去重 与数据库设计 5

2. 3 分享与转载 与数据库设计 6

2. 4 RESTful WebService接口设计

2. 5 HDFS资源安全问题

2. 6 权限

# 功能需求

## 2．1 文件上传

1. 浏览器方面,使用Ajax(XMLHttpRequest2)来异步上传文件
2. 由于文件上传会长时间占用服务器Servlet容器的线程,如果服务器收到过多的文件上传的请求,将会使容器的线程池的线程

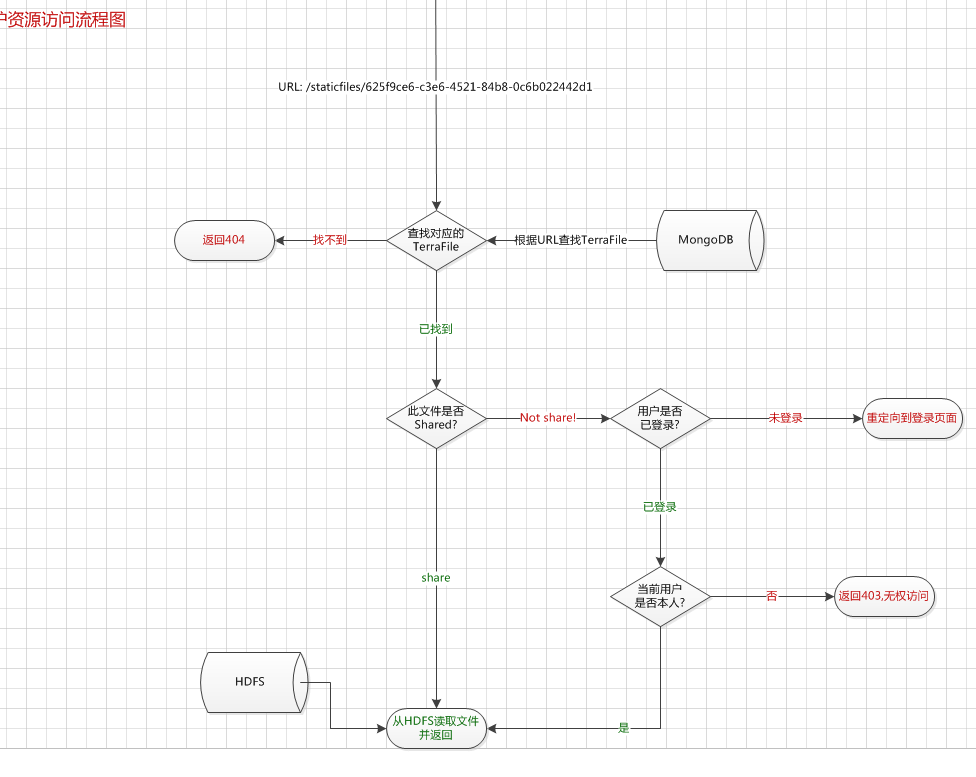
迅速耗尽,使得服务器无法处理新到来的请求,哪怕是普通页面的请求

1. 这里可以使用Servlet3.0 的AsynServlet特性: 当收到文件上传的请求后,处理完逻辑的处理后,新建一个后台线程来

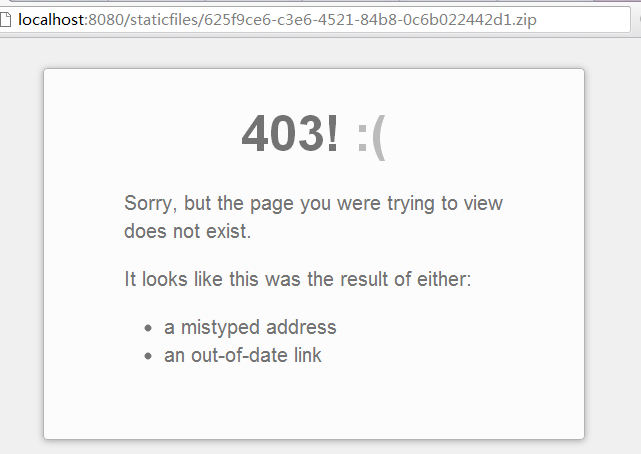
处理文件上传到HDFS的操作,保持输出流的连接,然后Servlet线程返回到容器,当该后台线程真正完成文件上传的

操作后,再将结果输出到该输出流.请求完成!

## 2．5 HDFS资源安全问题



无权限访问时返回403页面:



## HDFS文件去重

1. 文件去重是IO&CPU密集型操作,为了避免对正常业务造成影响,故选取凌晨02:00~04:00 业务较为空闲的时段进行.
2. 对新上传的文件进行MD5计算,然后将此值更新到文件的元数据表,如果发现MD5一致的文件,如果经过进一步比较

2个文件是否一致,如果一致,则将文件的元数据表的的HDFS路径指向同一份文件,并使文件引用计数+1,然后次重复

的文件删除.

2