文件服务器开发白皮书

1. 概述

本文通过设想的功能来说明文件服务器的开发内容和期望的目标。Wiki上有一个[旧的说明](http://develop.centit.com/wiki/pages/viewpage.action?pageId=61734955)。

1. 功能说明

系统的功能分为三个部分：client端、服务程序（war包）、文件服务使用接口（jar包）。

* 1. Client端

Client端是一个上传文件的表单，目前应用的一个开源的，应该是可以使用的，需要整理一下。Client端需要支持：

1. 单文件上传。
2. 多文件上传。
3. 断点续传。
4. 隐藏属性提交，不是所有的属性都是必须的，应用系统可以设置一些属性，随着文件一起提交记录在服务端的文件信息表中。其中有一个归宿应用系统是必须的。
   1. 服务程序
      1. 系统注册

这是一个后台维护功能，他维护这个应用服务器给那些应用系统使用，各个应用对文件的存储策略（包括位置），系统的url地址（一个系统可能有多个url地址），这个用于跨站过滤的。

Client端提交时必须带上所属系统的代码，这样服务程序知道把文件存放到哪儿。

* + 1. 文件写入
       1. 接收客户端上传文件

接收客户端上传的文件，并按照对应的系统存放在响应的地方。目前存放只考虑文件系统。但在设计时希望做成接口，这样后面可以换存储策略，比如支持写入数据库、分布式文件系统。

文件写入成功时同时写入文件的目录信息，向客户端返回文件的ID号。

客户端断点续传和文件的时序，这部分我不知道现在是怎么实现的，我的设想是这样的。

1. 客户端提交文件上传
   1. 计算文件大小，md5码。
   2. 上传文件大小、md5码、归宿系统和其他属性。
2. 服务端接收文件大小和md5码
   1. 判断系统中是否有这个文件
   2. 如果没有记录文件信息，返回上传起点为0
   3. 如果有，判断文件是否完整，
      1. 如果完整返回上传成功标记，包括文件编号。
      2. 如果没有完整，返回上传起点为上次接收到的点。
3. 客户端获取上传起点
4. 客户端上传文件片段，美容包括，md5码，起点，文件片段（这个大小一遍是固定的，除了最后一个片段）。
5. 服务端接收到文件片段
   1. 写入文件片段。
   2. 判断是否完成
      1. 如果完成返回上传成功标记，包括文件编号。
      2. 如果没有完成返回下一块文件上传起点。
6. 客户端重复3~5知道完成
7. 客户端告知应用系统文件上传完成，并提交文件编号。

以文件方式存储的文件，文件名可以为md5编码。

* + - 1. 附加操作

对文件进行加密。第一期可以搞一个最简单的加密算法，后面可以通过策略定义加密算法。

* + 1. 全文检索

全文检索采用Solr平台。

文件的文档中结构化数据。主要是权限上的划分，预计可以做的权限在以下几个维度进行划分组合：

1. 文档所属系统
2. 文档所属业务实例，比如：某个具体的业务办件。
3. 文档所属业务类别，比如：报销流程。
4. 文档所属人员。目前只能有一个归属人员。
5. 文档所属机构。目前只能有一个归属机构。
   * 1. 文件访问

文件目录信息表中记录文件的访问策略。

1. 对不保密的文件，可以直接通过md5编码或者文件id来访问。
2. 对于保密的文件，需要应用系统申请一个文件访问操作，并同时生成一个令牌。客户端通过文件操作的编号和令牌来访问。
3. 客户端程序通过接口程序来访问，这个需要通过系统中的白名单来验证。
   * 1. 文件同步

不同的文件服务器之间的文件同步。这个功能是这样设想的：

1. 系统中定义需要将新的文件同步到策略，
   1. 接收的上传文件需要同步。
   2. 通过同步过来的文件需要再次同步。
   3. 所有新增的文件都需要同步。
2. 定义同步的目标和策略。
3. 同步时
   1. 先访问目标文件服务器的结构将文件的基本信息同步过去。
   2. 目标文件服务器根据文件的基本信息请求源文件信息中的文件内容。
   3. 文件服务使用接口
4. 访问文件的基础信息。
5. 修改文件的基础信息。
6. 获取文件内容。
7. 分配文件访问授权，并获取对应的令牌。