

# 文档转换中心开发说明文档

## 1、需求分析及背景

1. 将word, ppt, excel文档上传至我们服务器;
2. 服务器调用openoffice将相关文档转换成pdf文档;
3. 将pdf文档上传至七牛等云存储;
4. 整个流程提供同步和异步两种方式供使用方选择;
5. 对于步骤2/3的异步方式, 可能会有一个中间状态, 比如“文档转换中”;
6. 由于异步, 需要提供状态查询接口, 转换完成后将pdf文档链接由查询接口返回;
7. 服务能支撑一定的并发处理, 比如100个请求同时处理, 多余100的请求排队处理;
8. 文档大小做一下限制, 限制大小可配置;

## 2、需求拆解

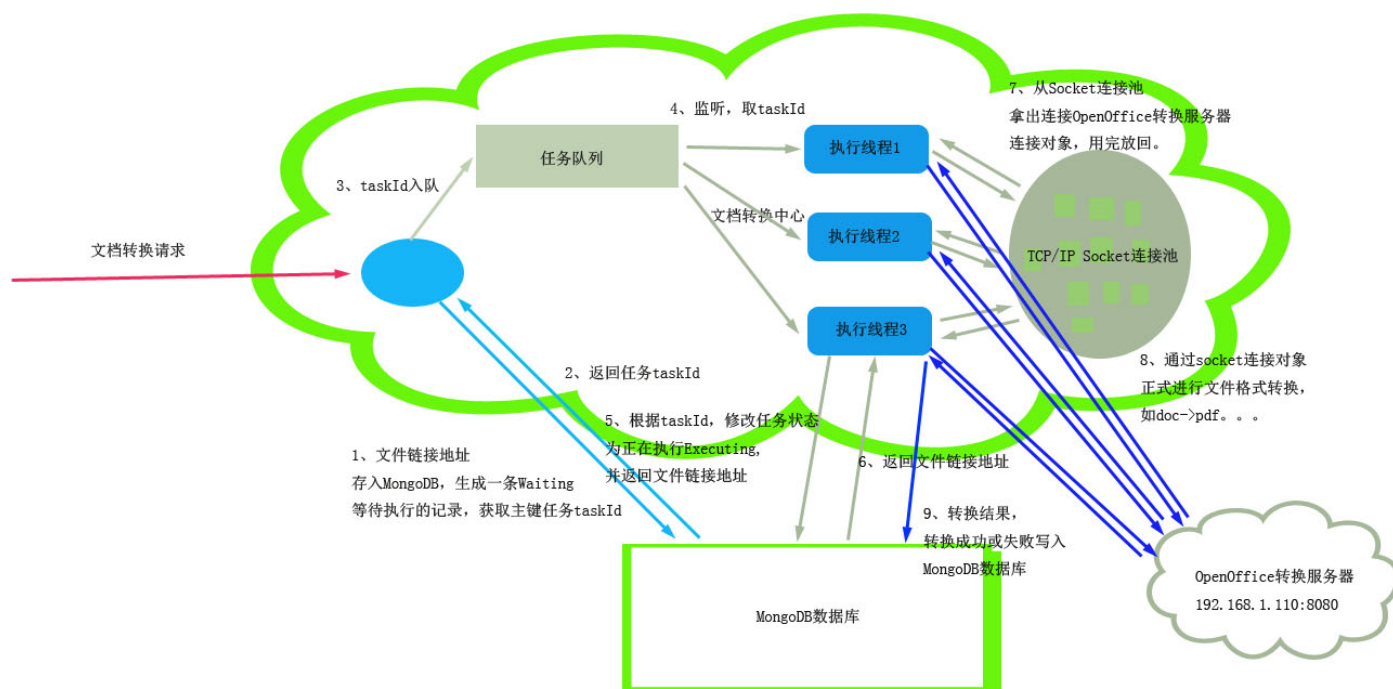
- 1.支持的待转换的文件格式doc/docx/ppt/pptx/xls/xlsx, 目标格式pdf
- 2.请求文档转换中心的参数为待转换文件的七牛链接地址  
如 (<http://okxyat5ou.bkt.clouddn.com/file/2017/4/c3fec18f86594611b635d021aa7f5c84-1.doc>)
- 3、异步请求返回任务taskId, 同步请求返回任务taskId及转换结果 (转换失败返回失败原因或者转换成功返回转换成功后的文件地址)
- 4、可以根据任务taskId查询任务转换状态 (Waiting/Executing/Failure/Success)

## 3、技术拆解

- 1、提供文档转换的服务是OpenOffice插件  
基于TCP/IP Socket请求方式。  
基于IP+端口号(PORT)监听的方式,
- 2、将连接上述OpenOffice服务器的Connection Socket连接对象池化, 类似于数据库连接池。
- 3、文档转换中心提供任务队列的实现, 有转换请求, 入任务队列, N个执行线程监听任务队列,  
任务队列有请求的某个空闲执行线程首先监听到就请求出任务队列处理,  
处理过程任务执行状态存储在mongoDB数据库中  
Waiting(待执行, 入任务队列的时候)  
Executing(正在执行, 出任务队列的时候)  
Failure/Success(执行结果, 执行线程文档转换结果反馈)
- 4、技术难点: 对象池实现细节处理及任务队列实现细节处理

## 4、实现原理图

文档转换中心实现图解



文档转换中心部署图解

