# UML课程第一次作业

## 分组和题目确定

### 小组人员

| 姓名 | 学号 |
| --- | --- |
| 刘柏田 | 199050418 |

由于我开学后改变了上课班级，后面可能会与他人组队并有人员改动

### 课题选取

经过多次选择和考虑后，我选择的项目是 Google File System 简称 GFS。

#### 项目价值

作为谷歌的经典三部著名论文中的其中一个，该篇论文可以算是三个之中最有诚意的一篇。

该作发表于 2010 年，距今已有一定时间。

作为大数据、分布式技术的鼻祖项目，该论文至今在架构设计上仍然具有极高的学习价值，其提供了在廉价硬件设备集群上存储大文件的部署方法，对于Google的核心数据即页面搜索的存储进行了优化。

从自身业务需求出发的创新性的设计开辟了一条全新的大型存储设计参考思路：

* 将机器宕机，重启，挂掉视为常态
* 支持分块存储超大文件
* 对文件的修改尽量是追加写，而不是随机写

即便时至今日，GFS 已被现在成熟且现代化的系统、架构替代，但多少都带有 GFS 的影子。

#### 项目可行性

GFS 清晰、有条理的设计思路非常容易实现该项目，目前已经有不同语言的实现版本：

* [GFS\_Implement - Python](https://github.com/rhitayu2/GFS_Implement)
* [Google-File-System-Implementation-with-Python - Python](https://github.com/Bereket-G/Google-File-System-Implementation-with-Python)
* [goGFS - Go](https://github.com/merrymercy/goGFS)
* [mini-google-file-system - Java](https://github.com/UniqueWJL/mini-google-file-system)

且目前知名的 HDFS(Hadoop Distributed File System) 就是基于 GFS 改进而来的著名的分布式系统基础架构。

#### 规模大小

参与者

项目本身大小取决于需要存储多少的数据，最简单的一个结构是：

* Master
* ChunkServer A
* ChunkServer B

#### 业务对项目的依赖度

GFS 的设计不局限于编程语言，使用多种语言都可以实现。

目前已经没有企业级、成熟的分布式项目直接的依赖于 GFS，因其久远的设计思路已经不适合如今的硬件、网络水平，但都依赖于一些著名的分布式结构基础系统，如 HDFS，因此可以说是间接的对 GFS 有一定的依赖。

## 需求分析讨论

* 精读 Google File System 论文并分析其架构
* 建立上传、下载的用例模型，对实际案例进行模拟
* 关注 GFS 中新颖的设计理念在该设计中的影响
* 使用已经实现的项目对 GFS 进行模拟，重点理解数据的推流方式。
* 对特殊情况进行分析，如多客户的并发写入、Chunk服务器宕机、主Chunk服务器网络中断等

## 用例图

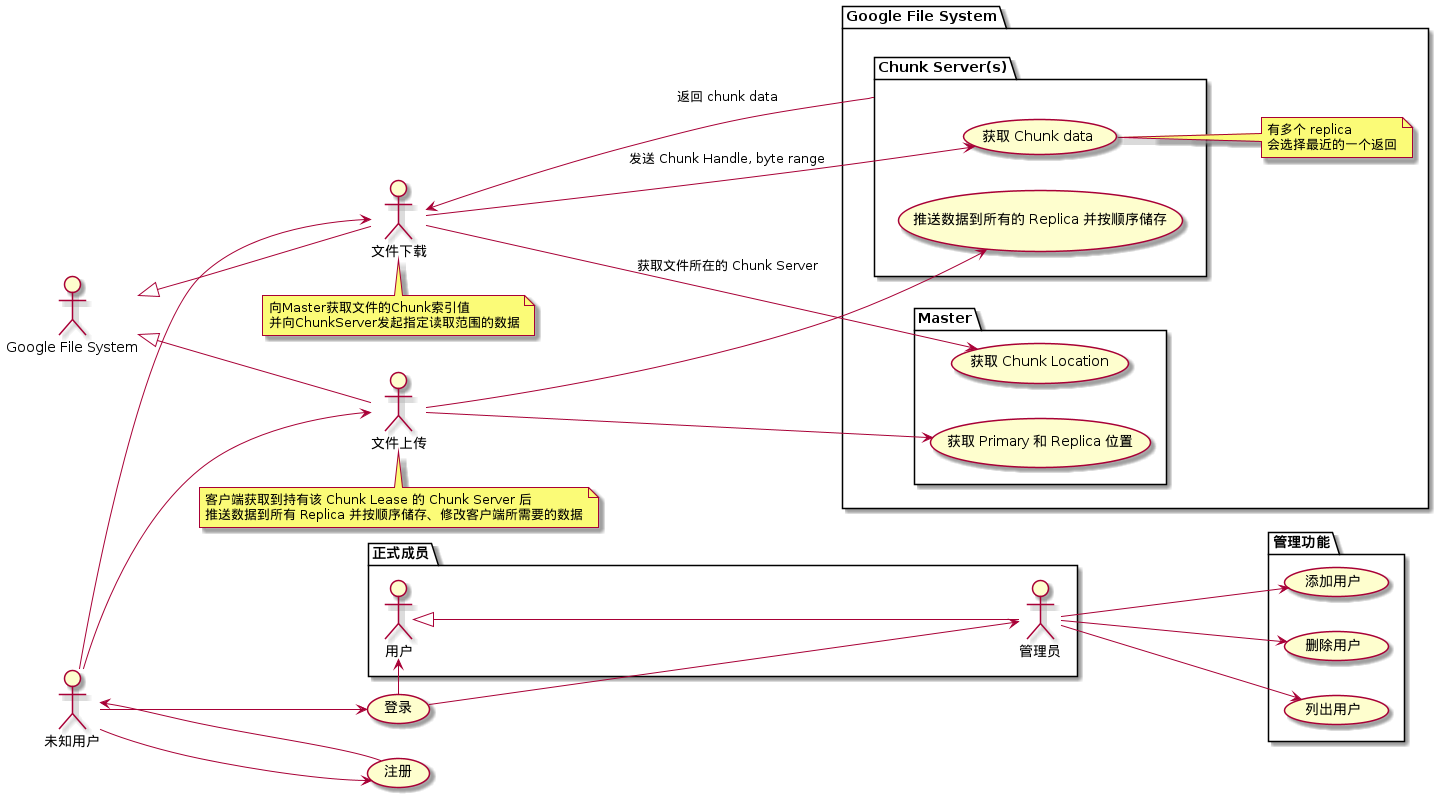


Figure 1: 用例图