



東南大學

# 毕业设计（论文）任务书

院（系）\_\_\_\_\_软件学院

专    业    \_\_\_\_\_软件工程

设计（论文）题目    \_\_\_\_\_基于 GFS 的分布式文件系统设计和实现

学 生 姓 名\_\_\_\_\_黄 鑫                学 号\_\_\_\_\_71112318

起 止 日 期\_\_\_\_\_2016.2 ~ 2016.6

设 计 地 点\_\_\_\_\_深圳市迅雷网络科技有限公司

指 导 教 师\_\_\_\_\_徐 造 林

顾 问 教 师\_\_\_\_\_

教研室主任\_\_\_\_\_杨 全 胜

教学院长（教学系主任）\_\_\_\_\_舒 华 忠

发任务书日期

2016 年 03 月 02 日

## 毕业设计（论文）任务的内容和要求

（包括任务内容、原始条件及数据、技术要求、工作要求等）

### 任务内容

迅雷作为中国最早的、目前最好的下载引擎之一，已经从 P2P 加速时代迈向了云加速时代；而云存储正是云加速的根基。据了解，迅雷的云存储规模已达 120PB（1PB=1024TB，1TB=1024GB），如此庞大的数据量，无疑需要一个强壮的文件系统支撑。GFS 全称 google File System，是 Google 推出的一个可扩展的分布式文件系统，用于大型的、分布式的、对大量数据进行访问的应用。它运行于廉价的普通硬件上，并提供容错功能。它可以给大量的用户提供总体性能较高的服务。迅雷的云存储文件系统正是根据实际业务情况基于 GFS 的二次开发，做了很多技术创新，在全国的分布式储存领域处于领先地位。本课题主要对 GFS（google File System）进行理论研究，再接触并研究迅雷基于 GFS 的原理开发的迅雷分布式文件系统，对迅雷分布式文件系统进行改进开发，最终为公司的云存储业务提供高效的、方便的技术支撑。

具体任务内容：

1. 课题调研，熟悉 Google File System 技术；了解工程开发技术，找出实现本系统的技术方案，完成专业英文翻译，编写开题报告。
2. 熟悉基于 GFS 的分布式文件系统，了解分布式文件系统设计开发平台及设计语言。
3. 根据设计分布式文件系统的技术方案、项目开发需求及目标，设计基于 GFS 的迅雷分布式文件系统。
4. 用程序设计语言编程实现基于 GFS 的迅雷分布式文件系统，得出相关结果，程序调试，项目验证。
5. 按毕业设计要求撰写论文。

### 原始条件及数据

Google File System、实习单位项目开发环境及条件。

### 技术要求

熟悉 Google File System，找出实现本课题的技术方案；根据项目开发需求及目标，编程实现基于 GFS 的迅雷分布式文件系统，程序调试，项目验证。

## 工作要求

---

附：普适工作要求：（如针对本课题另有特别规定的工作要求，请填于上面空白栏）

1、在深刻领会任务内容及要求的基础上，通过查阅文献资料、调查研究和方案论证，写出开题报告。然后开展实验研究、理论研究、设计、研制、开发以及数据处理、分析总结、资料整理等与任务书要求相应的工作，并撰写成毕业论文或设计说明书，独立地完成毕业设计的各项任务；

2、查找有关专业文献（10 篇以上）；

3、毕业论文或设计说明书需符合规范化要求，即：由中外文题名、目录、中外文摘要、引言（前言）、正文、结论、谢辞、参考文献和附录组成，中文摘要在 400 汉字左右，外文摘要在 250 个实词左右，中文题名字数一般不超过 20 个，设计说明书、论文或软件说明书的总字数在 1.5~2 万汉字（文、管等学科可根据具体情况，另行规定总字数，报教务处备案）。

学生应提交的软硬件的名称、内容及主要的技术指标（可按以下类型选择填写）：

☒ 计算机软件：

根据项目开发需求及目标，编程实现基于 GFS 的迅雷分布式文件系统，程序调试，项目验证。

☐ 图纸（名称、图幅、张数）：

---

---

☐ 电路板：

---

---

☐ 机电装置：

---

---

☐ 新材料、新制剂：

□结构模型:

---

□其它:

---

应提交的其它文档:

1、开题报告一份

2、与设计（论文）相关的英文资料译文一份（中文字数>5000 字，并附保留阅读痕迹的资料原文）

参考文献（至少五篇，含供学生翻译的英文资料，按规范开列）:

- [1] Ghemawat S, Gobioff H, Leung S T. The Google file system[C]//ACM SIGOPS operating systems review. ACM, 2003, 37(5): 29-43.
- [2] 郝向涛. 基于 Hadoop 的分布式文件系统技术分析及应用 [D]. 武汉理工大学, 2013.
- [3] XFS Master 主备机制: 高性能的分布式文件系统的元数据单点高可用方案. <http://djt.qq.com/article/view/322>.
- [4] 揭秘淘宝自主研发的文件系统——TFS. <http://www.infoq.com/cn/articles/tao-tfs>.
- [5] Weil S A, Brandt S A, Miller E L, et al. Ceph: a scalable, high-performance distributed file system[C]// 7th Symposium on Operating Systems Design and Implementation (OSDI '06), November 6-8, Seattle, WA, USA. 2010:307--320.
- [6] 杨德志, 黄华, 张建刚, 等. 大容量, 高性能, 高扩展能力的蓝鲸分布式文件系统 [J]. 计算机研究与发展, 2005, 42(6): 1028-1033.
- [7] 黄华, 张建刚, 许鲁. 蓝鲸分布式文件系统的分布式分层资源管理模型[J]. 计算机研究与发展, 2005, 42(6): 1034-1038.
- [8] 杨德志, 许鲁, 张建刚. 蓝鲸分布式文件系统元数据服务[J]. 计算机工程, 2008, 34(7):4-6.
- [9] 王亮. 基于 Linux 的分布式文件系统的设计与实现[D]. 华中科技大学, 2013.
- [10] 陈云云. 分布式文件系统名字空间管理[D]. 华中科技大学, 2013.
- [11] 冯幼乐. 分布式文件系统元数据管理技术研究 with 实现[D]. 中国科学技术大学, 2010.
- [12] 王敬轩. 分布式文件系统存储效率优化研究[D]. 华中科技大学, 2013.
- [13] 付松龄, 廖湘科, 黄辰林, 等. FlatLFS: 一种面向海量小文件处理优化的轻量级文件系统[J]. 国防科技大学学报, 2013, 35(2): 120-126.
- [14] Weil S A, Leung A W, Brandt S A, et al. Rados: a scalable, reliable storage service for petabyte-scale storage clusters[C]//Proceedings of the 2nd international workshop on Petascale data storage: held in conjunction with Supercomputing'07. ACM, 2007: 35-44.

## 毕业设计（论文）进度安排

起止日期	工作内容	备 注
2016.1.15~2016.3.15	背景资料学习，课题调研，确定课题技术路线，完成专业英文翻译，编写开题报告。	
2016.3.16~2016.4.1	熟悉基于 GFS 的分布式文件系统，了解分布式文件系统设计开发平台及设计语言。	
2016.4.2~2016.4.25	根据系统设计技术方案、项目开发需求及目标，设计基于 GFS 的迅雷分布式文件系统。	
2016.4.26~2016.5.25	编程实现基于 GFS 的迅雷分布式文件系统，程序调试，项目验证。	
2016.5.26~2016.6.10	撰写毕业设计论文，准备毕业设计资料，进行毕业设计论文答辩。	

注：只需按阶段作出安排，更细的安排应由学生自己在开题报告中作出。

指导教师签名：徐造林

2016 年 03 月 02 日