

学校代码： 10246
学 号： 15212010038

復旦大學

硕 士 学 位 论 文
(专业学位)

云存储中间件关键技术优化设计与实现

**Design and Implementation of Backup-Cloud
Oriented Optimized Middleware**

院 系： 软件学院

专业学位类别(领域): 软件工程

姓 名： 刘武

指 导 教 师： 韩伟力 教授

完 成 日 期： 2015 年 月 日

指导小组成员名单

韩伟力 教授

目录

摘要.....	1
Abstract.....	1
第一章 引言.....	2
1.1 背景介绍.....	2
1.2 研究内容及意义.....	4
1.3 论文组织结构.....	5
第二章 核心技术.....	6
2.1 中间件.....	6
2.2 Java 第三方开发包.....	6
2.3 WebExtension.....	6
2.4 大文件分块上传/下载.....	7
2.5 断点续传.....	7
2.6 本地缓存.....	7
第三章 云存储中间件系统优化分析.....	8
4.1 系统总体框架.....	8
4.2 优化项分析.....	8
4.2.1 交互适配管理.....	8
4.2.2 会话持久化管理.....	8
4.2.3 本地缓存管理.....	8
4.2.4 系统优化管理.....	8
第四章 云存储中间件系统优化设计.....	9
5.1 系统结构图.....	9
5.2 系统框架设计.....	9
5.3 类图设计.....	9
5.4 流程设计.....	9
5.4.1 系统总流程设计.....	9
5.4.2 交互适配流程设计.....	9
5.4.3 会话持久化管理流程设计.....	9
5.4.4 本地缓存流程设计.....	9
5.4.5 系统优化流程设计.....	9
5.4 系统接口设计.....	9
第五章 云存储中间件系统优化实现.....	10
第六章 云存储中间件系统优化结果分析.....	11
第七章 总结和展望.....	12

7.1 总结.....	12
7.2 展望.....	12
参考文献.....	12
致 谢.....	12

摘要

近年来，云存储服务因其海量的存储空间、便携的存取方式、可靠的安全特性，逐渐取代传统的数据存储模式，成为人们首选的数据备份方式。国家“核高基”科技重大专项^[1]为应对这一发展趋势，设立了云存储与云备份子课题，构建了一款自主研发的云存储服务平台。该系统以中间件^[2]作为服务器与客户端之间的交互枢纽，实现了数据传输、数据压缩、数据加密以及安全会话等基本功能。然而，持续升温的云存储行业在近几年面临着诸多的挑战，衍生出如安全性、容灾性、稳定性等问题。仅依靠基础的云存储服务已无法适应当前错综复杂的云环境，用户也对云存储平台提出了更高的要求，这刺激着各大云存储供应商对云存储的关键技术进行革新，以获取更多的市场份额。

本论文主要通过对当前云存储行业的分析，研究云存储服务中的关键技术，同时基于前期的成果，对现有云存储中间件进行优化，以适应云存储行业的发展趋势。本文拟在 Linux 平台上，通过加密的形式，保证用户信息和数据的安全性；同时，借助大文件分块传输和断点续传技术，解决大文件上传过程中遇到网络中断或服务器死机等故障而引发的数据重传问题；利用本地存储和计算能力，减少数据上传下载等 I/O 密集型服务，提高服务性能；此外，优化并完善平台框架和 API 支持机制，解决不同应用和云存储服务之间适配的问题，以支持更多的桌面 OS 组件使用云备份与云存储服务。

关键词：

Abstract

~~With the rapid development of Internet technology,~~ the network has penetrated into every corner of our lives. It needs an efficient way to store the massive resulting data. However, the traditional data storage methods has long been unable to meet the needs of users because of its limited capacity, the use of inconvenient, easy to lose data and other limitations, so cloud storage technology came into being. In just a few years, there are dozens of cloud storage products launch on the market. From abroad iCloud, Google Drive, Dropbox to the domestic Baidu cloud, Ali Cloud, Tencent Cloud, many companies who are optimistic about the emerging markets want to be in this crowd. Continued warming of the cloud storage industry in recent years is facing fierce competition, In order to occupy more market share, the challenge has also spurred the new cloud storage platforms to introduce new technologies to solve the issues derived from cloud storage like security, stability and network problems.

This paper mainly studies the key technology in cloud storage service, and optimizes a self-developed cloud storage middleware. Based on previous results, in the Linux platform, exploiting the local storage and computing power to reduce data upload and download I / O-intensive services, to improve service performance. Through the form of encryption to ensure that user information and data security. At the same time, this paper uses the technology of block upload and breakpoint resume to avoid data retransmission caused by network problem or server problem. In addition, the platform framework and API support mechanisms are optimized and refined to address the adaptation issues between different applications and multiple cloud storage services to support more desktop OS components using cloud backup and cloud storage services.

Keywords:

第一章 引言

1.1 背景介绍

随着互联网技术的飞速发展,网络数据呈现出井喷式的增长,传统的数据存储方式存在太多的局限性,已经无法满足人们的要求。在这样的背景下,云存储服务^[1],逐渐成为主流的数据存储方式。因其海量的存储空间、便携的存取方式、可靠的安全特性,越来越受大众的青睞,大型的机构如政府、医院、企业等也逐步将数据转移到云端。

为适应数据存储的新趋势,国家“核高基”科技重大专项设立了云存储与云备份子课题,拟构建支持云存储服务的底层框架,研制一款高效的云存储服务平台。至目前为止,经过充分的市场调研,设计出了一套完整的体系结构,实现了数据传输、数据压缩、数据加密以及安全会话等基本功能。系统以中间件作为服务器与客户端之间通信的桥梁,有效地分离了客户端云存储工具与基础设施之间的绑定,提高了系统的稳定性,并为系统的跨平台性提供了良好的基础^[4]。

然而,大范围普及的云存储服务在近几年迎来了巨大的挑战,除了提供云端数据存储这项基本的功能之外,人们对云存储服务平台的安全性、容灾性、稳定性投入了更多的关注。在传统的数据存储方式中,数据保存在本地,对用户来说是可管控的,而将数据存储在云端则可能产生如个人数据泄露、商业信息遭到窃取、科研成果外泄等问题,这可能对用户产生不可逆的后果。此外,云存储服务以网络作为基础设施,而网络又时常受限于硬件、软件和地域等多方面的因素,一旦出现网络波动或中断,将会影响数据在网络层的传输,尤其对较大的文件传输来说,遇到网络问题只能进行重传,这将极大地降低用户的体验度。同时,随着云端数据不断地积累,产生了大量重复冗余的数据。一方面这些冗余的数据会占用服务器大量的存储空间,另一方面用户存储在服务器的文件只是一个映射的链接,如果重复上传服务器中已存在的文件,将会占用大量的网络带宽。更进一步,介于移动互联网和智能手机的热度持续走高,人们的娱乐、办公环境逐渐趋向于平台化,针对单一应用的云存储服务已不适应当前平台多元化的格局,实现云端数据在多平台的共享,成为了云存储行业的共识。

目前,国内外成熟的云存储服务供应商为应对这一系列新的挑战,纷纷推出新的技术或是对原有技术进行革新,以最大化程度的满足用户需求。我们选取了四个国内外较为知名的云存储服务供应商进行了调研,国内的百度网盘、360云盘,国外的Google Drive、Dropbox,分别对他们的技术

批注 [J1]: 关于前期的工作

批注 [J2]: 以下瑞祺的论文作为参考文献

批注 [J3]: 安全性

批注 [J4]: 网络稳定性

批注 [J5]: 数据冗余性和本地缓存

批注 [J6]: 多平台性

特性进行了全面的分析。表 1.1 罗列出了详细的信息。

表 1.1 国内外知名云存储服务平台技术特性分析

	加密算法	数据去冗余	大文件上传	断点续传	平台数
百度网盘	128 位 SSL	全局	20G / 4M	支持	6 种
360 云盘	128 位 SSL	全局	5G / 4M	支持	5 种
Google Drive	128 位 AES SSL/TLS	无	5TB / 字节流	支持	5 种
Dropbox	256 位 SSL/TLS	单一用户	不限 / 8M	支持	7 种

数据加密是一种强有力的安全措施来保护个人信息不被窃取或篡改，根据调研结果，所有的云存储服务供应商都采用了类似的方式对数据在传输过程中及存储在云端时两种状态进行了加密。在传输过程中，百度网盘和 360 云盘使用的是 128 位的 SSL^[5]加密技术，Google Drive 和 Dropbox 分别采用了的 128 位/256 位的 AES SSL/TLS^[6] 加密技术。数据在云端时，百度网盘和 360 云盘未强制对数据进行加密，而是由用户在上传之前决定是否要将数据以加密的形式存储。反观 Google Drive 和 Dropbox，它们选择了对上传至云端数据进行强制加密，Dropbox 使用的是 256 位的 AES^[7]加密方式，而 Google Drive 采取了规模相对较小的 128 位 AES 加密技术。此外，考虑到网络的不稳定因素，各大供应商都对大文件上传技术进行了优化，以避免因网络原因而导致的大文件重传问题。百度云盘支持单个文件上传的大小上限为 20G，360 云盘为 5G，Google Drive 为 5TB，而 Dropbox 不限制单个文件的大小。在处理大文件上传的方案上，它们都将文件以固定大小的文件流形式，通过设定偏移量来对大文件进行分块传输，这样做的好处是遇到网络故障时，只需记录已上传数据流的偏移量，重传时可以根据偏移量进行文件续传而无需从头开始。针对大量数据冗余的问题，国内外的云存储服务平台选择了不一样的策略。减少数据冗余除了能为服务器节省大量的存储空间外，还能为用户提供一种新的功能，这种功能称之为“秒传技术”。所谓的“秒传技术”指的是当服务器存在当前用户待上传的文件时（通过文件的 MD5 进行校验^[7]），不会真实地上传该文件，而是会将这个文件的链接拷贝到用户的网盘中，这样能有效地解决冗余数据重复上传的问题。百度网盘和 360 云盘都支持全局的重复文件校验，为用户提供了“秒传”这项功能。Dropbox 采取的是一种相对折中的方式，它不会在全球文件系统中校验该文件是否存在，只针对单一用户进行去重处理，这样能在很大程度上地减少由去重而导致的数据安全问题，如侧信道攻击^[8]。Google Drive 对待“秒传技术”极其谨慎，尽管这样做能够大幅度地提升上传的效率，但出于安全性方面的考虑，Google Drive 未在云盘中使用该技术。同时，这些成熟的云存储供应商都会针对主流的操作系统和平

批注 [17]: 加密方式

批注 [18]: 大文件上传与断点续传

批注 [19]: 数据冗余处理

台，提供不同类型的客户端，这样使得用户可以跨平台地使用云存储服务。常见的平台有 Windows、Mac、Android、iPhone、Web，以上四个云存储供应商都提供了这些平台的客户端，其中百度云盘还支持 Windows Phone，Dropbox 是唯一支持 Linux 和 BlackBerry 智能手机的平台。

批注 [10]: 多平台支持

由此可见，各大供应商为迎合市场的需求，纷纷对云存储的关键技术做了重要的革新。而由国家“核高基”科技重大专项自主研制的云存储服务平台目前尚不具备以上技术特征，因此，亟需对原有系统中的关键技术进行优化，以保证产品的竞争力。

1.2 研究内容及意义

云存储关键技术作为产品的核心竞争力，已成为了各大供应商的研究的重点。但是，由国家“核高基”科技重大专项自主研制的云存储服务平台还停留在提供基础的云端数据存储服务上，为顺应云存储行业发展的趋势，本文主要通过对原有中间件系统进行全面分析，同时借鉴业内成熟产品的优势，从以下几个方面进行优化。

1. 持久会话信息管理：云存储中间件在客户端和服务器之间工作，多个客户端的命令通过中间件提交给服务器。用户的命令会以会话的形式保存在中间件，当中间件系统出现异常时（如网络中断、中间件死机），会在下次重启中间件时自动恢复执行未完成的会话。此外，对会话信息需要进行了物理隔离，并且以加密的形式存储在数据库中，以提高用户信息的安全性。
2. 大文件上传：利用服务器提供的高级功能，实现了大文件分块上传功能，在受到网络状况受到影响的情况下，若文件上传中断，会自动启用断点续传功能，以避免数据重传的问题。
3. 本地缓存：利用现代计算机普遍较大的存储空间以及高性能的计算能力，实现了本地缓存管理，进一步减少了系统 I/O 操作，提高了中间件性能。同时，利用服务器提供的个人存储在云端的数据信息，实现了针对单一用户的“秒传”功能。
4. 多应用适配：原有的中间件系统只支持单一的 CLI 客户端，为实现对不同客户端请求的适配，支持更多的客户端类型，对系统的跨平台性做了优化。此外，开发了一款基于 Chrome 浏览器的插件，用户可以利用插件使用云存储服务，并提供了 Java 平台的软件开发包（SDK）^[9]，Java 开发者可以利用该开发包提供的应用接口 API 集成云存储服务。

1.3 论文组织结构

本文将分七个章节对论文进行阐述。第一章介绍本文研究内容的背景以及相关的研究工作。第二章介绍本文在研究和实现过程中使用到的核心技术，包括中间件技术、加密算法、Java SDK、WebExtension、大文件上传/下载技术、断线续传技术以及本地缓存技术。第三章将分析原有云备份中间件的不足，指出本文需研究和实现的内容。第四章详细阐述云备份中间件系统的整体设计，包括优化后的系统框架、类图设计、流程设计以及接口设计。第五章介绍云备份中间件实现的细节。第六章将对云备份中间件优化后的结果进行评估，以实际的测试结果作为评估的标准。第七章对全文进行总结，并对未来进行展望。

第二章 核心技术

2.1 中间件

中间件是一种独立运行的系统应用软件，可以实现不同技术间的资源共享，被称为“软件胶水”，使得软件开发人员可以更加容易的完成输入与输出之间的通信。中间件把软件组件和企业应用程序连接起来，是一个处于应用程序和操作系统之间的软件。中间件可以通过网络连接两个独立的系统或软件，实现真正地跨平台、跨网络、跨硬件的功能。随着中间件技术的成熟，其在软件开发过程中的分工和重要性越来越明显，目前已被广泛应用于各个领域，显著降低了软件的开发成本。

2.2 Java 第三方开发包

2.3 WebExtension

附加组件(Add-ons)扩展和修改 Web 浏览器的功能。它们使用标准的 web 技术编写-JavaScript、Html、CSS 再加上一些专用的 javascript API。另一方面，附加组件可以为浏览器增加新的特性或者改变某些网站的外观。

WebExtensions（扩展）是跨浏览器的用于开发附加组件的工具。在很大程度上，与谷歌浏览器 Chrome 和欧朋浏览器 Opera 所支持的扩展 API 兼容。为这些浏览器所写的扩展在大多数情况下只需少量修改的便可在火狐浏览器 Firefox 和 Microsoft Edge 浏览器上运行。这些 API 与多线程 Firefox 完全兼容。

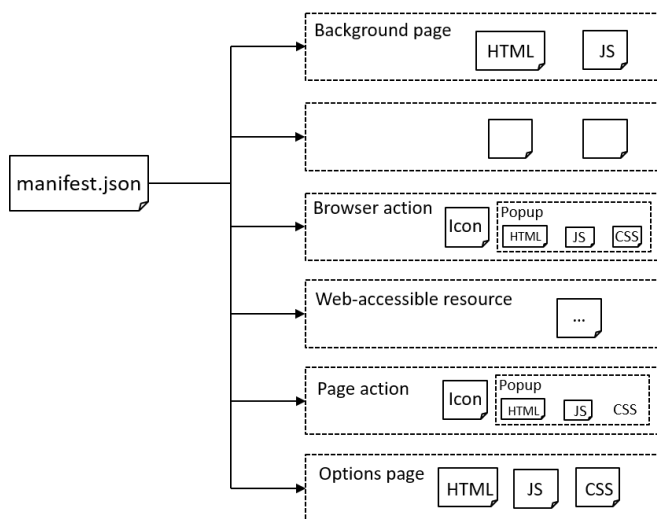


图 2-1 WebExtension 框架图

- background pages: 执行一个长时间运行的逻辑
- content scripts: 与网页进行交互
- browser action files: 在工具栏中添加按钮
- page action files: 在地址栏添加按钮
- options pages: 为用户定义一个可浏览的界面，可以改变插件设置
- web-accessible resources: 使打包好的内容可用于网页与目录脚本

2.4 大文件分块上传/下载

2.5 断点续传

客户端软件断点续传是指在下载或上传时，将下载或上传任务（一个文件或一个压缩包）人为的划分为几个部分，每一个部分采用一个线程进行上传或下载，如果碰到网络故障，可以从已经上传或下载的部分开始继续上传下载未完成的部分，而没有必要从头开始上传下载。用户可以节省时间，提高速度。

断点续传技术有以下几个特点：

- 断点续传功能，既可节约时间又可以节约金钱
- 定时下载功能，可以为将要下载的软件制定任务列表，让下载软件在规定的自动拨号上网并下载软件，下载完毕后再自动挂起 Modem，断开与 internet 的连接，甚至自动关闭计算机
- 多文件同时下载
- 致命错误发生时的关闭机制
- 预防病毒侵害的安全机制，文件下载完毕，即可自动将其发送到指定的病毒的检测软件进行病毒扫描

2.6 本地缓存

第三章 云存储中间件系统优化分析

4.1 系统总体框架

4.2 优化项分析

4.2.1 交互适配管理

4.2.3 会话持久化管理

4.2.3 本地缓存管理

4.2.4 系统优化管理

第四章 云存储中间件系统优化设计

5.1 系统结构图

5.2 系统框架设计

5.3 类图设计

5.4 流程设计

5.4.1 系统总流程设计

5.4.2 交互适配流程设计

5.4.3 会话持久化管理流程设计

5.4.4 本地缓存流程设计

5.4.5 系统优化流程设计

5.4 系统接口设计

第五章 云存储中间件系统优化实现

第六章 云存储中间件系统优化结果分析

第七章 总结和展望

7.1 总结

7.2 展望

参考文献

致 谢

复旦大学 学位论文独创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。论文中除特别标注的内容外，不包含任何其他个人或机构已经发表或撰写过的研究成果。对本研究做出重要贡献的个人和集体，均已在论文中作了明确的声明并表示了谢意。本声明的法律结果由本人承担。

作者签名： 刘武 日期： 2017.07.02

复旦大学 学位论文使用授权声明

本人完全了解复旦大学有关收藏和利用博士、硕士学位论文的规定，即：学校有权收藏、使用并向国家有关部门或机构送交论文的印刷本和电子版本；允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其它复制手段保存论文。涉密学位论文在解密后遵守此规定。

作者签名： 刘武 导师签名： 韩伟力 日期： 2017.07.02