基于Hadoop的云盘系统

卢颂贤　思科信息学院 1302 计算机科学与技术

**摘　要：**为了实现数据信息的共享，减轻工作负荷，使文件管理更科学、更规范，我们设计一个基于Hadoop的云盘系统，用于实现了数据信息的移动化，满足了用户在任何地点时间进行个人数据信息的下载、上传、删除、分享及用户文件夹的新建和修改等操作需求，实现了远程的数据共享，通过互联网进入系统可以封边快捷地享受共享资源。通过使用Strut2框架能促进服务器的信息交流，实现用户自由访问，操作方便、简单，提高网络用户的信息交互效率。

**关键词**：Hadoop云盘系统；oracle数据库；strut2框架；JavaScript；Bootstrap

**Virtual storage system base on Hadoop**

**Songxian Lu Cisco school of Informatics**

**Abstract:** In order to achieve the information sharing among database, lighten the working load, made the file management become an upward development of scientific and standard, we plan to design a virtual storage system which based on Hadoop, which become a landmark of virtual storage field. Achievement of mobile information management fulfil the need of the user to manage their characteristic storage system whatever they want. Achievement of remote data share takes a virtual role of using internet to connect with other user rapidly. Meanwhile, using the frame construction Strut2 advances the information communication of servicer, increase the Information interaction efficiency of internet user deeply, which permit user visit freely and become a simple operation system.

**Key words:** *Virtual storage system，Oracle Database，Strut2 frame construction,JavaScript；Bootstrap*

目　　录

**摘要I**

**AbstractII**

**1 绪论1**

1.1开发背景与现状................................................2

1.2开发本系统意义.................................................................................................3

1.3目标4

**2 开发平台介绍5**

2.1开发语言5

2.2数据库5

2.3 Struts2框架5

2.4文件存储平台5

2.5文件检索平台5

2.6总体技术路线5

**3、系统功能模块分析5**

3.1系统总体设计5

3.2系统模块设计5

3.2.1登录模块6

键入章标题(第 3 级)6

键入章标题(第 3 级)6

键入章标题(第 3 级)6

键入章标题(第 3 级)6

**4、数据库设计与说明5**

4.1数据库设计5

4.2数据库说明5

**5、界面设计5**

5.1登录界面5

5.2用户操作界面5

键入章标题(第 2 级)5

键入章标题(第 2 级)5

键入章标题(第 2 级)5

**6、运行环境与测试5**

6.1软件运行环境5

6.2系统测试5

**参考文献5**

**附录5**

**致谢5**

*（插入分页符）*

1. **绪论**

**1.1开发背景与现状**

网盘，又称网络硬盘，是由互联网公司推出的在线存储服务，向用户提供文件的存储、共享、备份、访问等管理文件的功能。

随着移动互联网设备的持续繁荣，个人数据将继续呈现爆炸式的增长，而多种硬件设备之间的交互需求将使网络硬盘的用户增长持续呈迅速上升的趋势。据第三方调研公司的不完全统计，全球云存储市场的份额将达到16亿美元，而金山快盘在2012年的用户数量也突破到了1000万左右。

我们最初的愿望，就是希望有个地方可以永久地存放自己的文件不怕弄丢，然现在，这个产业已经以令人乍舌的速度在发展。

最新应用的云计算储存技术，为网盘行业发展带来了新的力量，传统的网盘将逐步被云存储取代，如邮箱的文件存储，本地存储等技术。云存储是搭建在高速分布式存储的网络上的高级数据中心，它将网络中大量不同类型的存储设备通过应用软件进行加工，集合起来协同工作，建立一个较为安全的数据存储空间软件和权限访问的系统，适用于各大中小型企业与个人用户的数据资料存储、备份、回档等一系列在日常生活中的必要需求。云存储最大特点就是在于将单一的存储产品转化成为数据存储与用户服务，在这个云存储的技术支持下，网盘发展行业可以像金融行业银行一样，在单一的存储服务基础上衍生出更多增值的服务，如会员服务，权限加速，存储空间加大等，只有这种与往常不一样的改变，才能拯救网盘的现状。

网盘的原理如同分蛋糕，就是将自己公司上的服务器作为一份蛋糕，将空间容量动态分配给各个注册用户，学术上来说，就是发展网盘的网络公司将其服务器的硬盘或硬盘阵列中的一部分容量分给注册用户使用，因此网盘一般来说投资都比较大，毕竟服务器租赁的费用比较高，所以免费网盘的普通注册用户的容量比较小，一般为5G到10G不等；另外为了防止注册用户滥用网盘空间资源，往往还会附加单个文件的最大限制要求。简单来分析，免费网盘的普通用户一般只用于存储较小的文件。而收费网盘或者VIP注册用户则具有速度快、安全性能好、容量高、允许大文件存储等优点，适合与对网盘资源要求比较高的用户使用。

如今在国内的发展得较为完善的当属百度云网盘。百度云网盘是百度科技有限公司2012年正式推出的一项免费云存储服务，首次注册即可获得5GB的空间，首次上传一个文件可以获得1GB，使用手机登录百度云移动端的注册用户，即可领取1024G的永久免费云盘存储空间。 到目前为止，具备Web版、Windows客户端、Android手机客户端，Mac客户端，IOS客户端和WP客户端。在开发如此完善多元的情况下，用户可以随时随地将自己的文件上传到网盘上，普通用户单个文件最大可达3G，并可以跨越不同终端随时随地查看和分享。不仅如此，百度网盘还提供离线下载、文件智能分类浏览、视频在线播放、文件在线解压缩、免费扩容等功能。

国内网盘事业发展得非常迅速，不仅仅有百度云网盘，还有咕咕网盘、360网盘、115网盘、联想企业网盘、金山快盘、华为网盘、搜狐企业网盘、彩云网盘、天翼云等各个企业的网盘软件，包括web端、移动端、不同服务器端的应用。当然，对于用户而言，方便的网盘软件就是一个好的软件应用。

当然在国外，在较早的时候就开始发展网盘技术，如OneDrive、MediaFire、DivShare、WikiUpload、Scribd、jooce、FanBox、onShare、Mozy、5upload、Nakido、Xdrive、Mofile、G宝盘、Box、vdisk、蜂盘、FileFront等；在2012年，Google推出了Google Drive，微软的Skydrive也开始进一步整合业务。可以负责任的说，云盘系统在用户心中占的比重也越来越大，与人们日常生活，工作的联系也越来越缜密。

随着网盘市场竞争的日益激烈和存储技术的不断发展，传统的网盘技术已经显得力不从心，传输速度慢及恢复能力低、安全性差、营运成本高等瓶颈一直困扰着网盘企业。同时，因为网盘技术的发展，可能因为竞争客户群体太激烈，很多企业也开始减少在网盘方面的发展，如较为出名的360云盘，是奇虎360科技的云存储服务产品。在2016年10月20日决定停止开发，2016年11月1日起停止云盘上传服务，2017年2月1日起关闭所有云盘账号并清空数据，2016年10月20日至2017年2月1日进行会员退款。

但是，我个人认为，随着硬件和宽带成本的降低以及无线移动网络的普及，网盘的发展将会有很大发展前景，因为服务群体非常的大，届时网盘的盈利模式将不再局限与云存储，而是会成为一个更好的产品，为人们提高更方便的服务。

**1.2开发本系统意义**

开发本系统的灵感源至于我们的日常生活，在使用百度云网盘的过程中，发现云盘在各种网络状态下的使用限制，为此，我决定深入探究云盘系统的搭建与构造，并尝试完成一个具有完整网盘功能的系统。

该系统实现了文件的上传、下载、删除、分享，通过利用hadoop平台来实现数据的备份，防止了数据丢失，增强了数据的安全性。用户在使用过程中可以将自己的文件分享给其他用户，实现了数据共享。本系统还引入了Lucene全文检索技术，可以对用户的文件进行内容检索。

**1.3目标**

通过注册页面实现用户的注册账号功能，注册的同时用户可以选择邮箱验证注册和默认注册两种方式，其中默认注册方式分配给该账号的云盘空间有10G，若选择邮箱验证注册方式系统会发送验证码到用户的邮箱，以此方式成功注册后用户将获得100G云盘空间。

用户注册成功后将跳转到登陆界面，登陆成功便跳转到系统主界面，主界面将显示用户的账号、邮箱、电话等信息，用户头像为系统默认的头像，用户可自行上传更改。主界面还有新建文件夹功能、上传新文件功能、文件名搜索和文件内容搜索功能。若用户想扩大云盘容量可通过系统的签到功能进行升级。

1. **开发平台介绍**

2.1开发语言

Java是一门由Sun公司开发出来的面向对象编程语言，其既拥有C++语言的多种优点，同时也舍弃掉了C++里面很多难以理解的概念，例如多继承、指针等。得益于此Java语言有简单易用和功能强大两个特点。Java语言作为一门静态面向对象编程语言，能够快速进行面向对象编程。我们开发的系统运行于window和linux上，因此可利用Java的跨平台和可移植性进行开发，并且本人对Java语言相对熟悉。

JSP全称是Java Server Pages，即java服务器页面，由Sun公司主导多公司贡献建立成的一种动态网页技术标准。JSP可以在传统的HTML文件中插入Java script代码和jsp标签，从而形成功能更强大的JSP界面。 用JSP开发出来的Web应用能运行于多种系统上，这也符合我们要开发的系统的要求。

2.2数据库

MySQL是一种关系型数据库管理系统，其由MySQL AB 公司开发后来归属于甲骨文公司。MySQL关系数据库没有将数据保存在一个大数据库里面，而是将数据保存到了多张不同的表里面，如此一来增加了数据库的安全性与灵活性。由于MySQL拥有体积小、搭建拥有平台的成本较低，并且速度快，又是一个开源项目，因此其成了大部分中小型网站开发的不二之选。这些特点也适合我们的云盘系统。

**2.3 Struts2框架**

Struts2是基于MVC三层架构模式设计出来的一种轻量级的Web应用框架， Struts2的主要作用是建立模型与视图之间的数据交互。Struts 2的核心是WebWork，其通过在strut.xml文件中配置拦截器来控制处理用户的请求，这样配置之后业务的处理逻辑和界面的展示层完全分开了，更有利于系统的开发与维护，大大提高了灵活性。

**2.4文件存储平台**

由于开发的云盘系统是用来存储用户数据的，因此系统存储文件的安全性就显得极为重要，要保证用户数据不能因为系统原因或者机器宕机而丢失。所以数据的存储平台我们选择了hadoop平台。Hadoop是一个由Apache开发的开源的分布式系统基础架构。其实现了分布式文件系统，英文简称为HDFS。Hadoop框架有两个核心设计，分别是HDFS和MapReduce。HDFS主要负责为海量的数据提供成熟的存储机制，而MapReduce即为海量的数据集提供了高可靠的分布式计算功能。因为HDFS可以通过配置来设定每个文件的备份数，因此其有高容错性的特点，可以部署在廉价的机器上。得益于其分布式的特点，Hadoop可以通过流的形式来访问HDFS中的数据，其有吞吐量高的特点，适合于大数据集访问的应用程序。因此HDFS非常适合用来开发云盘系统，事实上百度云盘也是基于Hadoop开发的。

**2.5文件检索平台**

由于我们的云盘系统还提供了以内容搜索文件的功能，所以我们引入了Lucene全文检索机制。Lucene也是apache软件基金会的一个开源项目，其是一个全文检索架构而不是全文检索引擎。Lucene拥有完善的查询和搜索引擎，并提供部分文本分析引擎，但是其没有提供中文文本分析引擎，所以我们还要引入庖丁中文文本分词器来进行文本切分。

2.6总体技术路线

首先是开发系统的搭建，我们的数据存储平台搭建于本地虚拟机的Linux系统下。安装好linux系统后将Linux系统的IP及防火墙等配置好。在Linux系统上我们搭建了hadoop平台及MySQL数据库。

本云盘系统采用的技术路线如下：

（1）采用Struts2框架进行代码逻辑与界面的交互。

（2）虚拟机与宿主机的通信采用hadoop提供的RPC机制进行数据交互。

（3）开发系统环境：windows系统、linux系统

（4）数据库：MySQL。

（5）开发工具：IntelliJ IDEA 2016.3，Tomcat，JDK。

1. **部分功能模块分析**

**3.1系统总体设计**

**3.2部分功能模块设计**

3.2.1登录模块

用户登录系统拥有独属于自己的账号，不同等级的用户所具有的存储空间也不相同。不同的角色拥有不同的权限，所能看到的页面，操作的数据也不一样。不同的权限用户只可以对自己权限下的功能进行操作，不能越级操作。[5]

管理员账号，可以查看网盘上所有用户的存储文件信息并允许管理员进行增删查改的操作。

3.2.2注册模块

用户通过填写账号、密码、手机号码、邮箱来注册用户账号并使用云盘系统。在此处，加入了邮箱验证的功能，增强了系统的保密性与稳定性。

3.2.3搜索模块

用户可以通过选择按钮选择通过文件名或文件内容进行搜索，大大增强了系统的可用性与方便性。使用Hadoop框架，运用数据流的技术，使文件搜索速度大大增强。

3.2.4进度条模块

通过Bootstrap框架，编写了显示网盘内存空间的进度条。通过操作进行动态变化，客观可见的显示存储空间的变化，同时具有变色提醒功能，当网盘空间达到80%时，进度条会从蓝色变为具有警示意义的红色，大大提高了用户的观感体现。

3.2.5文件处理模块

通过对文件后缀的处理，给予不同文件类型的图标icon，同时可以通过点击标签对文件进行选取与分类。每个文件都具有下载、删除、分享功能，体现了网盘系统的完整性、

3.2.6任务功能模块

设有任务功能，用户可以通过完成每日任务提升网盘存储空间，从而大大增强了系统的可用性与寿命，这种人性化的设置让用户得以长久的使用系统。

1. **数据库设计与说明**

数据库E-R图

用户

文件

拥有

n

属于

1

回收站

数据字典：

User 用户信息表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 默认值 | 允许非空 | 自动递增 | 备注 |
| 用户名 | varchar(50) |  | 否 | 否 | 用户昵称 |
| 密码 | varchar(50) | NULL | 是 | 否 | 用户密码 |
| 电话号码 | varchar(50) | NULL | 是 | 否 | 用户电话号码 |
| 头像 | mediumblob | NULL | 是 | 否 | 用户头像 |
| 邮箱 | varchar(50) | NULL | 是 | 否 | 用户邮箱 |
| 签到日期 | varchar(50) | 0 | 是 | 否 | 记录用户上次签到日期 |
| 签到次数 | int(10) | 0 | 是 | 否 | 用户连续签到次数 |
| 已用空间 | double(4,2) | NULL | 是 | 否 | 用户已用云盘空间 |
| 总空间 | double(10,0) | NULL | 是 | 否 | 用户可用云盘总空间 |

file 文件信息表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 默认值 | 允许非空 | 自动递增 | 备注 |
| Id | int(50) |  | 否 | 是 | 文件唯一标记号码 |
| 文件名 | varchar(50) |  | 否 | 否 | 文件名称 |
| 数据库路径 | varchar(50) |  | 否 | 否 | 文件在数据库中的路径 |
| 拥有者 | varchar(50) |  | 否 | 否 | 文件的拥有者 |
| 标签 | varchar(50) | NULL | 是 | 否 | 文件标签 |
| 大小 | varchar(50) | NULL | 是 | 否 | 文件大小 |
| 类型 | varchar(50) | NULL | 是 | 否 | 文件类型 |
| MD5值 | varchar(50) | NULL | 是 | 否 | 文件的MD5值 |
| 云盘路径 | varchar(50) | NULL | 是 | 否 | 文件在云盘的路径 |

recycle 删除文件缓存云盘路径表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 字段名 | 数据类型 | 默认值 | 允许非空 | 自动递增 | 备注 |
| 云盘路径 | varchar(50) | NULL | 是 | 否 | 已删除文件在云盘的路径 |

1. **界面设计**

**5.1登录界面**

所有用户打开本系统都会默认进入到登录界面，输入注册后的账号与密码，点击登录后，会到数据库进行验证，如果验证成功，则会进入用户操作主界面，若验证失败则会弹出提示。如下图所示：



右上角的三色按键，用于用户的个性化设置，通过点击按键，改变登录框的背景颜色；若对于个性化设置不满时，可以点击左上角的取消按钮，背景则会恢复默认设置，同时账号，密码栏也会恢复默认设置。

默认显示账号Username与密码Password，点击输入栏时消失，，当用户输入空值时恢复Username与Passwrod，同时设有具有变色特效的登录按键与注册按键，让用户体验一个多元化的界面。

部分前端界面代码如下：

<form action="login" method="post">  
 <input type="text" id="username" name="username" value="Username" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'Username';}">  
 <div class="key">  
 <input type="password" id="password" name="password" value="Password" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'Password';}">  
 </div>  
  
 <div class="signin">  
 <input type="submit" value="Login" style="float:left">  
 <input type="submit-1" value="Register" style="float:left;" readOnly="true"  
 onclick="window.location.href='register.jsp'">  
 </div>  
</form>

**5.2注册界面**

注册界面与登录界面部分相同，为了简便用户，用户在注册时，只需填写用户的基本信息，账号、密码、手机号码、邮箱验证后，点击确认即可，若该用户名已被注册，会弹出提示。



输入邮箱，点击getCode即会向注册邮箱发送邮箱验证码。

部分前端界面代码如下：

<form action="register" method="post">  
 <**s:if** test="#session.username!=null">  
 <input type="text" id="register\_username" name="username" value="<**s:property** value="#session.username"/>" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'Username';}" style="margin-top: 10px;">  
 <input type="text\_phonenumber" id="register\_phonenumber" name="phonenum" value="<**s:property** value="#session.phonenum"/>" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'PhoneNumber';}">  
 <input type="text\_email" id="register\_email" name="email" value="<**s:property** value="#session.email"/>" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'E\_mail';}">  
 </**s:if**>  
 <**s:else**>  
 <input type="text" id="register\_username" name="username" value="Username" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'Username';}" style="margin-top: 10px;">  
 <input type="text\_phonenumber" id="register\_phonenumber" name="phonenum" value="PhoneNumber"  
 onfocus="**this**.value = '';" onblur="**if** (**this**.value == '') {**this**.value = 'PhoneNumber';}">  
 <input type="text\_email" id="register\_email" name="email" value="E\_mail" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'E\_mail';}">  
 </**s:else**>  
  
  
 <input type="password\_text" id="code" name="code" value="Code" onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'Identifying Code';}"  
 style="margin-bottom: 0px;width: 140px;padding-left: 54px;padding-right: 54px;border-left-width: 10px;margin-left: 0px;margin-right: 20px;">  
 <input type="submit" value="Get Code" style="border-radius: 10px;width: 100px;margin-top: 5px;height: 59px;">  
 <div class="signin">  
 <input type="password\_text" id="register\_password" name="password" value="Password"  
 onfocus="**this**.value = '';"  
 onblur="**if** (**this**.value == '') {**this**.value = 'Password';}" style="margin-bottom: 15px;">  
 <input type="submit" value="comfirm" readonly="true" style="…">  
 </div>  
 </form>

**5.3用户操作主界面**

注册用户操作主界面大致分为3个部分，第一部分为界面顶部，左上角为一个返回按键，右上角如前文介绍，是一个用户个性化设置按键，在下方存在一个标签页，分为用户主页面与副页面。

第二部分为用户个人信息，显示用户名、手机号码、邮箱、用户头像、云盘存储空间以及网盘签到栏，通过签到增加用户存储空间，同时用户可以点击modify这个按钮上传头像图片。

第三部分为用户文件库，在上方显示一个新建文件夹按钮、一个上传按键、一个搜索框、一个选择按钮、一个搜索按键、动态导航栏、以及显示文件栏。每一个文件最右方都会有三个按键，分别为下载、删除、分享。

**5.4用户上传文件页面**

点击上传按键后弹出上传界面，点击选择文件选择路径，并显示文件在右边，可以通过选择类型对上传文件进行分类。

**5.5用户新建文件夹界面**

点击新建文件夹，弹出新建界面，输入文件夹名，点击确定即可。

**5.6用户分享界面**

**5.7增加存储空间页面**

1. **运行环境及测试**

**6.1系统运行环境**

数据存储平台搭建于本地虚拟机的Linux系统下。安装好linux系统后将Linux系统的IP及防火墙等配置好。在Linux系统上我们搭建了hadoop平台及MySQL数据库。

**6.2系统测试**

为确保云盘系统能健康正常的运行，测试在开发过程中是必不可少的一个组成部分。测试的工作分布在系统开发的各个阶段，如界面、功能的开发，都需要反复的调整以确保系统的正确性。

本次测试主要用到一种单元测试的测试方法，就是黑盒测试，黑盒测试顾名思义，就是类似在一个黑盒子中进行测试，无法查看内部数据，只通过输入输出来查看系统的正确性。通过测试来检查功能是否正确或遗漏或者存在数据库的访问问题等等。若发生问题，则去查看该部分的代码，这就是单元测试的方便之处，不需要再全局的寻找问题所在，只需查看某部分即可。

测试的部分界面如下：

登录、注册、验证、上传、下载、分享、修改图片等功能都能成功运行，测试成功。

（正文结束）

*（空两行）*

**参考文献：**

[1]作者.书名[M].出版地:出版社，出版年份:起止页码.*（图书）*

[2]主办方.论文集[C].出版地:出版者,出版年份.*（论文集）*

[3]作者.题名[J].刊名，年，期（卷）:起止页码.*（期刊）*

[4]World Health Organization. Factors regulating the immune response: report of WHO Scientific Group[R]. Geneva: WHO, 1970.*（科技报告）*

[5] 萧副教授.供热计量改造工作管理系统的设计与实现[EB/OL].(2013-12-19)[2014-05-22]. https://wenku.baidu.com/view/3d642236852458fb760b562d.html *(网络文献)*

**附　　录**

*（“附录”两字中间在全角状态下中间空两格，居中、小三号，宋体，加粗）*

**致　　谢**

四年的大学生涯转瞬即逝，当我完成这篇毕业论文时意味着我的大学即将告一段落，四年前的我肩负着家人的希望与期盼来到了广东外语外贸大学，在这四年间我不断成长，在学习能力、社交能力上都有很大的提升。因此我要感谢我的父母，没有他们的养育、教导，就没有现在在广外学习的我，我爱的父亲母亲，谢谢你们！

我要感谢我的指导老师周咏梅老师，作为信息学院教授，她不辞辛劳地指导我们完成这项毕业设计和毕业论文。每次出现难以解决的问题时他都会耐心地指导我们，帮助我们理清思路。我要向周老师表达我深深的敬意和最衷心的感谢，感谢他一直以来的悉心教导！

感谢广外!对于您，我们有过骄傲与自豪，有过苛责与失望，有过颓废和奋进，有过汗液和热血……四年前，不同的原因进来，四年之后的今天，我们站在岔口再次选择，或工作、或出国、或读研……就要各奔前程，每个人收获的果实不一样，但母校潜移默化的影响，对母校深深的眷恋，却将同样长久地伴随我们。四年归化于广外，此生难改其印记。

最后我要感谢与我一起完成此次设计的郑可佳同学，我们一起学习、探讨知识，当遇到迷惑或者疑问时我们总是能互相帮助、互相鼓励、互相学习。谢谢郑可佳同学，与你一起完成毕业设计的时光短暂而又精彩！

在论文即将完成之际，我的心情无法平静，从开始进入课题到论文的顺利完成，有多少可敬的师长、同学、朋友给了我无言的帮助，在这里请接受我诚挚的谢意!最后我还要感谢培养我长大含辛茹苦的父母，谢谢你们!