

面对中国网民的个人云存储

1. 综述

这篇报告调查个人云存储在国内外的发展现状，评价现有国内外云存储软件的优缺点，研究一个新的面对中国网民的个人云存储软件的前景和可行性，并简单介绍此类软件的特性、系统架构草图和团队信息。

2. 市场前景

云存储服务是随着时代的发展应运而生。网络日益普及，网速提升，个人电脑普及，娱乐全面数字化，办公数字化的需求，便携数据共享和传递，文件备份和随时随地共享文件及数据的需求，家庭环境多台电脑之间需要共享文件，个人用户办公与个人电脑之间数据同步，智能手机和PC之间的数据同步，协同办公的需要，云存储和云同步软件可以完美地解决这些问题。

世界正在从信息时代向数据时代过渡，数据将意味着核心竞争力。据统计[4]，2008年全世界共传输/处理了9.57 Zettabytes数据 (1ZB = 1000PB)。云存储的市场前景巨大，据某机构预测[1]，2014年云存储服务市场将突破2亿美元。全球IT市场咨询公司思林博德市场咨询Springboard Research发布的《中国云存储服务报告，China Cloud Storage Services Report》显示，在未来的5年，中国云存储服务市场的年复合年增长率将达到103%。Springboard认为，中国云存储服务市场将由2009年的605万美元快速增长至2014年的2.1亿美元。报告同时指出，“尽管每月每GB的存储服务价格将持续下降，但是云存储市场的总采用量的增长幅度更为迅猛，这将推动云存储市场整体规模在未来5年内的快速上扬。”

3. 理想产品特性

“同步方面，我觉得dropbox做的是最好的，尤其在增量上传方面，国内没一个可以做到这点，一个几M几十M的数据库，如果稍微改动一点点，国内的这些所谓同步软件都是需要把整个文件重新上传，可dropbox不需要这样，会自动的找出需要上传的地方进行更新，而且很快就可以上更新完毕。国内的这些同步软件真的要好好学学人家，还是把最基础的做好吧，不要整天去搞增加功能、增加空间，功能在精不在多，空间够用就行，基础的没做好，其它神马都是浮云。” – 网友1

“不过个人并不建议你用国内的网盘,国内产品极易产生说关就关的情况.何不用dropbox呐,应该比国内的网盘强多了.” – 网友2

通过对用户使用需求和感受的调查分析显示[3]，云存储产品的可用性，易用性和可靠性对用户产品的选择起到了决定性的作用。这对产品质量提出了很高的要求，要求产品在任何情况下都能够做到真正的帮助用户同步用户所需要的数据，给用户友好的提示信息，对用户的操作做出有效迅速的响应。在做好产品质量的基础之上，通过对用户需求的分析，我们希望提供给用户以下功能特性：

- 备份个人数据；
- 全平台同步支持，包括各桌面平台和移动平台；
- 通过多种方式共享文件和目录；

- 误删文件恢复；
- 文件历史版本恢复；
- 通过浏览器访问云端文件；
- 增量上传下载；
- 支持超大容量，支持单个大文件；
- 可选加密传输及存储；
- 数据安全性保障；
- 网络传输优化；
- 在线音乐和视频；
- 文件断点续传；
- 集成社交软件，支持社交软件分享；
- 简单在线编辑和协同办公；
- 商业用户以及VIP用户支持；

我们计划仿照国内外各大个人云存储软件的做法，提供免费的基本帐号给普通用户，获得用户群；对于需要更大空间或者更先进功能的用户或者企业，我们可以收取一定的会员费，或者也可以考虑按容量/实际流量收费等新模式。这里需要注意的一点是软件起点应该比较高，易用，可靠，速度快，能吸引大量免费用户来逐渐产生用户黏性。数据备份是一个很敏感的服务，如果因为软件错误导致用户数据丢失，流失的用户以后再回头来用这个软件的可能性很小，并可能产生信任危机导致其他用户离开。

4. 国内外现有云存储产品分析

国外主流云存储产品有Dropbox，Sugar Sync，Box.net等等；国内主流同类产品有酷盘kanbox，盛大Everbox，华为DBank数据银行，坚果铺子，金山快盘，够快网盘等等。国外用户习惯比较稳定，云存储产品比较成熟，几种产品拥有比较稳定的市场份额，个人云存储市场变动较小。但由于跨洲网络环境的不稳定性，以及中国特有的政治因素，国外云存储很难在国内市场立足。

与此同时，国内网民最近几年慢慢开始意识到个人云存储的需要，然而市面上云存储产品多而杂，质量参差不齐，各个产品的市场份额变化也比较迅速。国内产品的开发者包括国内各大互联网厂商，比如华为、新浪、金山、盛大等，也有发展迅速的小公司，比如酷盘公司，也有大学创业型，比如坚果铺子，以及其他高校产品[2]。

国内现有的个人存储产品大抵可以分为三类：第一类是“软件下载园”或“资源下载站”的变形升级，注重陌生人分享(Unkown Share)或资源分享(Sources Share)，典型代表是115网盘；另一类是服务于个人“文档”的无缝链接，注重的是个人文档的分享(Friends Share)和商务分享(Business Share)，这类代表有金山快盘、QQ收藏、Dropbox、Google Docs；另外一类是基于社区的分享(SNS Share)，主要是介于陌生人分享和朋友圈子分享的涵盖交际圈的资源下载，比如微盘、QQ收藏等。

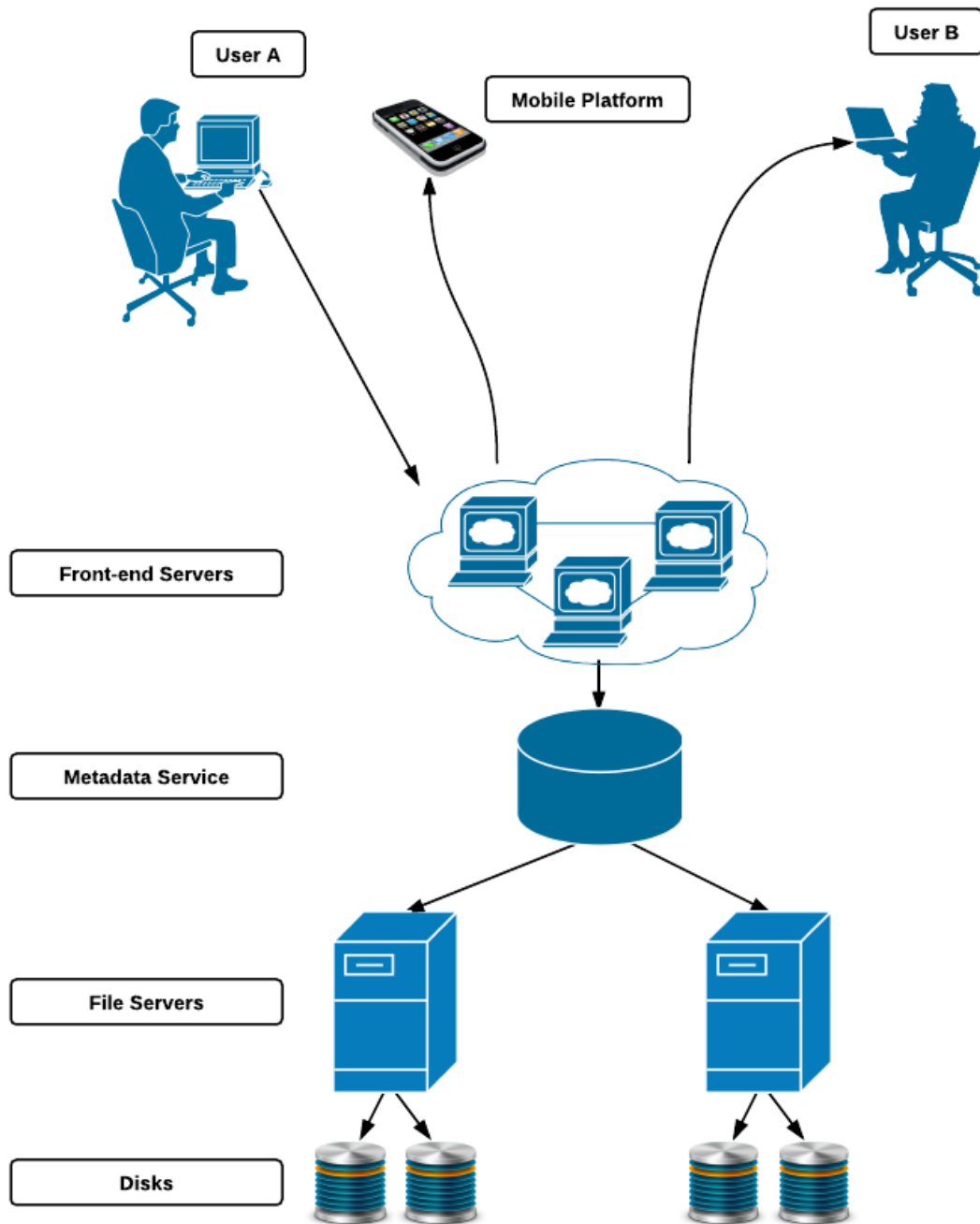
下面列出了我们对目前国内几大主流个人云存储产品的简要评测分析：

- 金山快盘
亮点：容量大，由金山软件推出，有一定的品牌优势，与QQ开放平台集成
不足：上传下载速度一般，没有增量传输
- 华为DBank
亮点：华为品牌优势，丰富的网络硬件系统经验，软件上市早，用户群较大，上传下载速度稳定
不足：仅有基本功能，平台有限
- 坚果铺子
亮点：支持三大OS；学生创业团队；用户体验与Dropbox极其相近；
不足：尚在内测阶段；速度一般(1GB文件需2小时完成上传)；限月流量机制依然对用户限制较大；据传服务器端预算有限
- 盛大Everbox
亮点：支持文件断点续传，支持很多媒体格式(音乐，视频，图片)在线浏览和播放
不足：不支持多版本恢复，平台仅限PC和Android，iOS开发中
- 酷盘kanbox
亮点：和TCL电视合作，能够在电视上直接访问酷盘内容，使用HTTPS协议
不足：不支持Mac，Linux，iPad，可能不支持增量传输

详细的功能指标列表参见文末附录的表格。

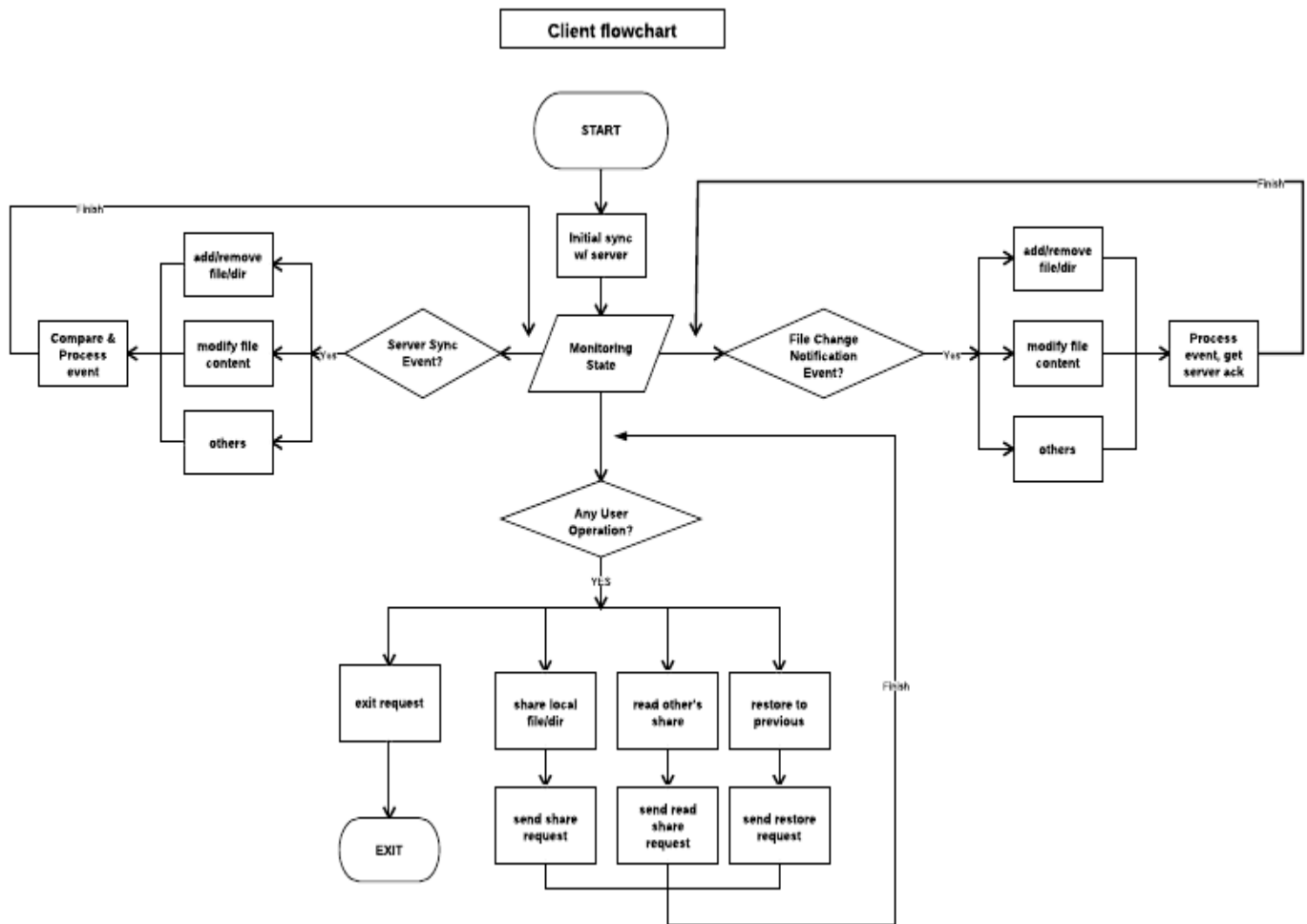
5. 系统设计

整个系统分为客户端和服务端。对于某一特定用户而言，多个平台上的多份数据需要同步，比如自己的PC电脑需要和自己的iPhone同步，同时用户可能跟其他用户分享部分数据；对于服务器而言，它提供层次型服务，包括前端服务器(Front-end Server)来处理客户端数据请求和WWW浏览等等；中间服务(Metadata Server)负责解析用户需求，查询各种元数据，将用户的数据请求发送到对应的文件服务器(File Server)；文件服务器将负责读和写相应数据。



5.1 客户端设计

客户端启动后，首先和服务端进行第一次握手和初始化操作，之后开始监视三种事件：第一种是来自服务器的同步请求，第二种是来自本地文件的更新事件，第三种是用户显式执行的操作请求，例如分享或者恢复文件等等。拟定客户端流程图如下所示：



客户端发送请求种类：

- 同步本机文件 (sync local client version to remote server)
- 更新远程服务器文件 (update server to local client)
- 恢复本机文件到某个历史版本 (restore client to an older version from server)
- 分享自己文件 (share with others)
- 读他人分享的文件 (read share from others)

实现：

- Windows截获文件操作信号(file change notification，包括file/folder add, remove, modify, rename, stat structure change): ReadDirectoryChangesW或者 FileSystemWatcher + NotifyFilters (需要.NET);
- 增量读写(incremental read/write): 需要实现一个简单高效的文件分块机制，计算每块的hash值，当一个文件（尤其是大文件）需要进行读写更新的时候，可以直接跟服务器比较每块的hash值，从而知道哪些块有变化，直接更新这些块；

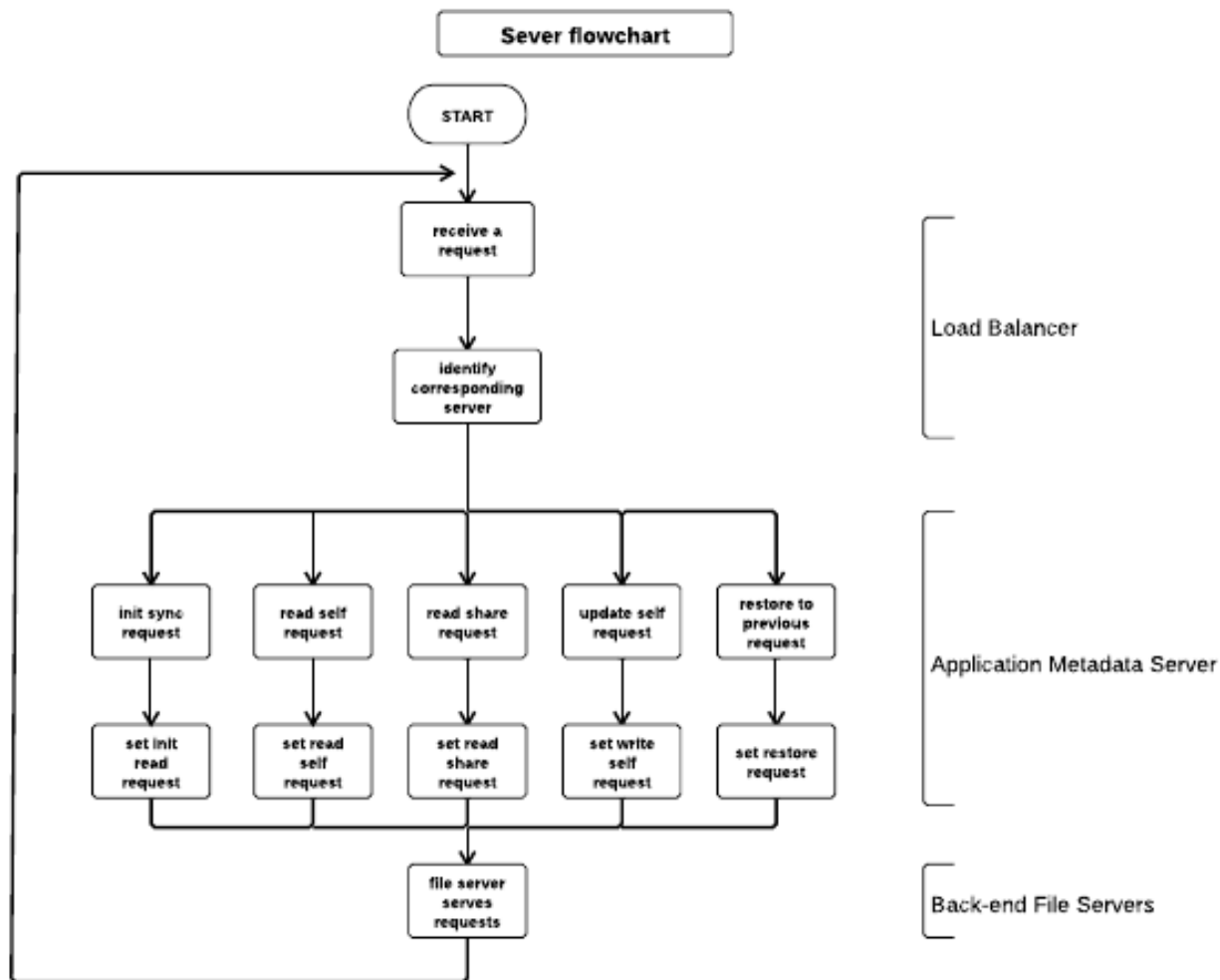
- **data deduplication:** 可以通过计算文件hash值跟服务器比较来确定这个文件是不是重复的文件；如果是，可以通过链接来“拥有”该文件。此类优化对多媒体内容比如视频、MP3、照片等会特别适用。酷盘称此技术为“秒杀”技术。
- 临时文件的探测和处理：由于云存储包含很多资料和信息，比如文档和代码。很多时候用户直接在云存储文件夹里编辑这类信息，比如编辑和编译LaTeX文档，或者MS Word类产品，甚至直接在里面编辑和编译传统源代码。Dropbox不能很好的处理这类情况，我们团队的队员都碰过这种情况。这类文件夹的特点是，文件小而多，变化频繁，中间文件多。我们也应该尽量处理这些corner case。优化的可能性之一是降低同步中间文件的优先级，优先保证同步源代码和可执行/产出，之后再尽量同步目标文件(e.g. *.obj)和其他中间产物(e.g. MS Word中间文件，迅雷下载完成之前的文件)。对于不同的应用，它们的中间产物并不一样，我们可能需要单独去寻找定义这些中间文件(case by case)。
- Mac OS里面fslogger提供一个很方便的file change notification机制，尽管苹果不官方开放相关API；Linux下面有lsyncd(Live Syncing Daemon)/inotify来管理文件变化的事件；

5.2 通信协议

- 参考NFS/CIFS的spec (Windows/Mac客户端)来决定C/S之间需要处理哪些事件以及通过哪种方式传输这些消息；
- 加密传输：HTTPS消耗大，但安全性好，可以绕敏感词监控；
- 服务器端处理的请求种类应该和客户端发送的请求种类类似；
- 国内不同ISP之间通信速度瓶颈问题和解决方案(网通，电信，联通，移动，科教，铁通，卫通，等等)；

5.3 服务器端功能要求

服务器要执行接受请求、解析请求和服务请求三层。接受请求负责接受来自客户端的各种请求；中间服务器层解析请求，并定位该用户所需要的数据的位置；底层文件服务器负责递送数据。服务器端流程如下图所示：



5.4 平台和对外API:

在设计初期，尽量考虑未来开放平台的可能性，开放平台上的活跃开发可以很强的增强用户黏性，并有利于系统的扩展性。

5.5 实验环境

- 集群服务器，SAS盘，Flash SSD（可选），Linux；
- 位于国内不同地区以及不同ISP下面的单机，用做客户端测试，Windows/Mac；

6. 团队介绍

- 丁广浩
 - 就职于Amazon.com Kindle 部门，从事网络应用以及服务开发。曾在EMC实习，从事网络文件系统namespace的相关工作。

- Northeastern University计算机硕士
- 邮件: dingguanghao@gmail.com
- 电话: (+1) 617-755-9966
- 付静
 - RPI计算机系博士四年级, 研究方向为大型机上的并行存储优化, 曾在ANL, Facebook和IBM Research实习
 - 邮件: jingfu4@gmail.com, skype: jingfu11
 - 电话: (+1) 650-804-5268
- 史亮
 - 中科院计算所博士四年级, 研究方向为信息检索
 - 邮件: nepshi@126.com
- 张力
 - 伦斯勒理工学院计算机系博士生四年级, 曾有无无线网相关开发经验, 曾在Willow Garage实习
 - 邮件: emma.lzhang@gmail.com
 - 电话: (+1) 518-269-0558

参考:

- 【1】 <http://cloud.chinabyte.com/81/11384081.shtml>
- 【2】 <http://chenguang.blog.51cto.com/350944/595980>
- 【3】 <http://www.syncoo.com/cloud-storage-product-value.htm>
- 【4】 http://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdf

附录:

国内外主流个人云存储产品特性

	Sugar Sync	Dropbox	Box	金山快盘	DBank网盘	盛大Everbox	坚果铺子	酷盘kanbox	联想网盘
数据备份									
数据可恢复	YES	YES	YES	YES	NO	YES	YES	YES	
备份任意文件夹	YES	NO	NO	NO	YES	YES	NO	NO	
实时更新上传文件	YES	YES	NO	YES	YES	YES	YES	YES	
版本恢复	YES	YES	YES	NO	NO	NO	YES	YES	
同步									
与多台电脑同步	YES	YES	有限	YES	YES	YES	YES	YES	
同步任意文件夹	YES	NO	NO	YES	NO	YES	YES	YES	
同步特定文件夹	YES	YES	NO	YES	NO	YES	YES	YES	
同步共享的文件夹	YES	YES	NO	NO	NO		YES	YES	
通过邮件上传	YES	NO	YES	NO	NO	NO	NO	NO	
与移动设备同步	YES	YES	NO	YES	YES	YES	YES	YES	

远程访问

通过浏览器访问文件	YES	YES	YES	YES	YES	YES	YES	YES
简单在线编辑	YES	NO	YES	NO	NO	NO	NO	NO

共享和协同工作

发送任意大小的文件	YES	NO	YES	NO	NO	NO	NO	NO
共享文件夹	YES	YES	YES	NO	NO	YES	YES	YES
通过权限和密码共享文件夹	YES	NO	YES	NO	NO	YES	NO	YES
发布到社交网站	YES	NO	NO	YES	WEIBO	NO	NO	NO

多媒体

自动创建相册	YES	YES	NO	NO	NO	NO	NO	NO
共享相册	YES	YES	NO	NO	NO	NO	NO	NO
同步音乐到浏览器	YES	有限	YES	NO	NO	NO	NO	NO
同步音乐到 iPhone	YES	有限	NO	NO	NO	NO	NO	NO
同步音乐到黑莓	YES	NO	NO	NO	NO	NO	NO	NO
同步音乐到Android	YES	YES	NO	NO	NO	NO	NO	NO

移动平台支持

iPad	YES	YES	YES	YES	NO	NO	YES	NO
iPhone	YES	YES	YES	YES	NO	NO	YES	YES
黑莓手机	YES	YES	NO	NO	NO	NO	NO	NO
Windows Mobile	YES	NO	NO	NO	NO	NO	NO	NO
Android	YES	YES	YES	YES	YES	YES	YES	NO
Symbian	YES	NO	NO	NO	NO	NO	NO	YES
为移动平台专门订制的网页	YES	YES	YES	NO			YES	YES

操作系统

支持Windows XP, Vista 和 7	YES	YES	NO	YES	YES	YES	YES	YES
支持Mac OSX	YES	YES	NO	NO	NO	NO	YES	NO
支持Linux	NO	YES	NO	NO	NO	NO	YES	NO
平台API	YES	YES	YES	NO	NO	NO	NO	NO
移动平台API	YES	YES	有限	NO	NO	NO	NO	NO

商业套餐

多用户商务套餐	YES	YES	YES	NO	NO	NO	NO	NO
管理员控制台	YES	YES	YES	NO	NO	NO	NO	NO
批量上传	YES	YES	YES	NO	NO	NO	NO	NO

国际化

中文	YES	NO	NO	YES	YES	YES	YES	YES
德文	YES	NO	NO	NO	NO	NO	NO	NO
日文	YES	NO	NO	NO	NO	NO	NO	NO
西班牙文	YES	NO	NO	NO	NO	NO	NO	NO
英文	YES	YES	YES	Partial	NO		YES	YES

其它

免费				YES	YES	YES	YES	YES	NO
容量限制				1TB	5 -	7-	N/A	5-	

				30GB	15GB		15GB	
单文件大小限制 断点上传		300M		< 200M				
				NO	YES	NO 邀请 码， 内测 阶 段， 限月 流量	YES	YES
备注		面向 企业	邀请 码		支持很 多在线 浏览播 放功能			商用