

应用场景如下：

网盘（网络 U 盘、网络硬盘）因为其便捷性被很多人青睐（目前国内多家网盘运营商停止服务，不是技术问题，而是管理与盈利问题，与本次作业无关，本次作业只讨论技术性问题），一个便捷的网盘服务，可以跨平台、跨地域，实现文件的存储、共享等功能；不需要随身携带存储介质，只要能上网，就能访问文件。

网络同步盘是网盘的功能之一，具体形式为：机器上运行网络同步盘的客户端，在网盘上指定一个目录、本机上指定一个目录，两者绑定之后，将文件拖入本机目录，就会自动上传到网盘的指定目录下，如果用户再在另一台机器上运行网络同步盘的客户端并以同一账号登录，则网盘同步盘的文件会即时更新到另一台机器的本地目录中。

作业基本要求：

实现一个简单的网盘同步盘系统，具体要求如下：

- 1 网盘的服务器操作系统确定为 CentOS 8，数据库系统使用 CentOS 8 自带的 MariaDB 10，编程语言为 C/C++，字符集为 GBK/GB18030
- 2 为了支持网盘同步盘，首先要实现网盘客户端，要求在 Windows/**Linux**/MacOS/Android/iOS 中任选其一并编程完成（编程语言不限），支持的基本功能如下：
 - 2.1 新用户的注册，要求可以自定义用户名和密码，检查用户名是否重复
 - 2.2 已有用户的登录，通过验证用户名和密码，允许合法的用户登录
 - 2.3 用户名和密码在传输过程中不必考虑加密问题（即抓包可看到明文），但对密码的长度和强度要有一定要求，服务器上存储的密码必须是加密形式
 - 2.4 用户登录成功后，可以看到自己账户下网盘的目录结构（允许为每个新用户定义一个初始目录结构，下含若干目录）
 - 2.5 可以指定目录结构中的某个目录为同步目录，同步目录要允许更换
 - 2.5.1 本要求的“网盘目录结构”为网盘的整体要求，网盘整体是不要求与本地同步的
 - 2.5.2 将网络同步盘理解为整个网盘中指定的某个下级目录
 - 2.5.3 网盘的整体目录结构，除用于指定/更换同步目录外，其余不做要求（即不需要做非指定同步目录的上传/下载）
 - 2.6 可以将本机上的某个目录与服务端指定的同步目录绑定/解绑，允许更换本机目录
 - 2.7 **不需要**限定每个用户的可用空间大小
 - 2.8 客户端应该能看到同步过程的各种信息（日志）
- 3 单机与服务端文件的初始同步（下面所称的目录均指本机/服务器的同步目录）的规则如下：
 - 3.1 如果服务器的目录下有文件，要全部下载到本机的目录下
 - 3.2 如果本机的目录下有文件，要全部上传到服务器目录下
 - 3.3 服务器上的文件，如果内容相同（以 MD5 码/SHA 码的计算值为准），则真实文件 **只允许** 保留一份，然后产生两个链接（服务器上真实文件的命名规则可自行定义），因此相同文件向服务器的上传要做到“秒传”（例：本机的 CentOS8.iso 文件，已同步到服务器上，再复制一份为 CentOS8-dual.iso，则服务器秒传）
 - 3.4 如果服务器目录下的文件和本机目录下的文件名相同且内容相同（以 MD5 码/SHA 码的计算值为准），则 **不允许** 再次上传/下载，要做到“秒同步”
 - 3.5 如果服务器目录下的文件和本机目录下的文件名不同但内容相同（以 MD5 码/SHA 码的计算值为准），则以服务器上的文件名为准更改本机文件名
 - 3.6 如果服务器目录下的文件和本机目录下的文件名相同但内容不同（以 MD5 码/SHA 码的计算值为准），则最终应该有两个文件（命名规则可自行定义）
 - 3.7 本机目录/服务器目录均可能包含下级子目录
 - 3.8 要求初始同步完成后，本机和服务器对应目录的内容应保持一致
 - 3.8.1 假设子目录包含最多 3 层，例：本机指定的同步目录为 D:\test，则目录层次最深为 D:\test\dir1\dir2\dir3，dir3 下只有文件而不再有子目录

- 3.8.2 假设单个文件最大不超过 4GB
- 3.8.3 本机的 0 字节文件不要上传
- 3.8.4 如果本机/服务器的下级子目录不同名但内容相同（例：本机同步目录为 D:\test，假设 D:\test\dir1 的内容和服务器同步目录的 dir2 子目录下的内容完全相同同名，则最终会有 dir1/dir2 两个子目录，且服务器上的文件为秒传）
- 3.8.5 **（额外加分，可不完成）** 某些编辑时产生的临时文件不应该同步（例如，docx 文档编辑时，Word 会自动在同目录下建立~开头的临时 docx 文件），不同步文件的匹配规则要求能放入配置文件中，随时可修改（配置文件的语法可参考正则表达式）

- 4 单机与服务端在初始同步完成后的同步规则（假设客户端始终在运行）如下：
 - 4.1 如果在本机目录下删除文件，则服务器对应删除（要考虑服务器上该真实文件还有其它链接存在，无其它链接存在等情况）
 - 4.2 如果在本机目录下删除下级子目录，则服务器对应删除
 - 4.3 在客户端运行时，除非解除绑定，**不允许**删除本机目录，如果本机目录不存在/在客户端未运行或解除绑定后被删除，则需要重新建立本机目录，按初始同步规则进行
 - 4.4 如果在本机上增加文件/子目录，则服务器对应增加
 - 4.5 如果在本机上修改文件名/子目录名，则服务器对应修改
 - 4.6 **不需要**考虑直接在服务端增加文件
- 5 多机（多个客户端在不同机器上以相同用户名登录）文件的同步规则如下：
 - 5.1 按两个单机-服务端的同步规则进行处理，服务端要考虑同步顺序的控制以避免冲突
 - 5.1.1 如果A机的 test1.dat 与B机的 test2.dat 文件内容相同，同步后的命名规则应如何？
 - 5.1.2 如果 A 机和 B 机同时增加新文件 new.dat，但内容不同，应如何处理？
 - 5.2 A 机上删除文件/子目录后，B 机同步删除
 - 5.3 A 机上增加文件/子目录后，B 机同步增加
 - 5.4 A 机上文件/子目录改名后，B 机同步更改
 - 5.5 如果 A 机/B 机上有内容相同（MD5 码/SHA 码的值相同）的文件要同步到服务器上，可以选择先将某台机器的文件同步，完成后另外一台机器再“秒同步”
 - 5.6 **（额外加分，可不完成）** 基于 5.5 的情况，如果能做到多个客户端同时上传同一文件（各客户端的上传内容不重复，最终形成一个完整的文件）

5.7 几个提示

- 5.7.1 想做到多客户端即时文件级的同步，很难，要定义一个合理的同步规则，例如，必须 A 机全部同步后才能开始同步 B 机，即服务端只支持某用户单机同步，多机当做若干台单机按时序排队，...
- 5.7.2 数据保护原则，多客户端同步的情况下，如果同步规则可删可留均合理的情况下，优选保留，哪怕扔垃圾箱，也有个后悔的机会
- 5.7.3 服务端的实文件与虚文件夹问题
- 5.7.4 完备的日志很重要，客户端也可以用日志来保留上次运行结束时的状态

- 6 异常处理
 - 6.1 文件的上传/下载要支持“断点续传”，即如果本次上传/下载因为某种原因在未完成时中断，下次要在原有基础上继续传输，**不允许**重传已传输部分的内容
 - 6.2 单机同步时，在上次单机同步已完成的情况下，如果对本机目录在离线状态（**离线状态**包括客户端未运行/本机目录解除绑定/当前网络不通等情况）时做了增加/删除/修改，则再次恢复到在线状态（**在线状态**是指客户端已运行，绑定了本机目录，且网络连通）时的规则如下：

故障原因			进行的操作			故障恢复后的处理规则
客户端运行	目录绑定	网络连通	增加	删除	修改	
Y	Y	N	Y	/	/	服务器增加
			/	Y	/	服务器删除
			/	/	Y	服务器修改
Y	N	任意	任意			恢复绑定后按初始同步规则进行

N	Y	任意	Y	/	/	服务器增加
			/	Y	/	服务器删除 (问题:本机怎么知道文件被删除?)
			/	/	Y	服务器修改
N	N	任意	任意			恢复绑定后按初始同步规则进行

6.3 多机同步时，在上次多机同步已完成的情况下，要考虑如下情况

6.3.1 A 机在线但 B 机离线，A 机进行了增加/删除/修改，则 B 机在线后要同步更新 B 机

6.3.2 A 机离线但 B 机在线，A 机进行了增加/删除/修改，则 A 机在线后要同步更新 B 机

6.3.3 A/B 机均离线，A/B 机均进行了增加/删除/修改，则按连通的先后顺序依次处理

6.4 **不考虑**手工在服务端删除文件/目录的情况

6.5 注：不同操作系统对文件系统的支持细节可能不同，因此，文件名含空格、文件名大小写不同的情况该如何处理，可自行指定规则

7 同步完成后，文件内容被修改后的更新规则

7.1 在同步完成后，如果实时编辑本机目录下的某个文件（文件名不变），则用本机文件去更新服务器文件（服务器上原版本是否保留可自行制定规则）

7.2 在多机同步时，如果某台机器上的文件被修改，则要依次更新服务端及其它机器上的本地目录中的对应文件

重点和难点提示：

- 1、服务端实际文件的存储和网盘文件夹的目录结构中各文件的对应关系是怎样的？对应的数据库结构应该如何设计？
- 2、一个文件（指 MD5 码/SHA 码相同）和多个用户的链接关系应该如何表示？不同用户的不同文件名、不同存储位置如何对应同一个文件？对应的数据库结构应该如何设计？
- 3、断点续传的分配和判断规则，服务端应如何支持？

作业要求：

- 1、**6 月 15 日（第十六周二）**前提交设计报告（包括数据库的设计、存储方案的设计、用户目录结构的设计等），之后会通知设计部分完成较好的小组准备 PPT，于 **6 月 17 日（第十六周周四）**课上进行设计部分的汇报并当场讨论、取长补短（预计 3-4 组）
- 2、设计部分讨论完成后，会给出作业的**验收要求**
- 3、本次作业会安排当面验收汇报，验收通过后再将所有内容打包后网上提交
- 4、当面验收的时间暂定为 **6 月 30 日~7 月 2 日之间的某天**后，具体待定（要求 6 月 29 日前完成全部程序）

额外加分作业（以下分值均为最高分值，不是具体得分）：

- 1、设计部分完成的较好的小组被选中进行汇报的，可以得到 1 分的额外加分
- 2、如果客户端/服务器之间的用户的登录认证及文件的传输过程均采用加密方式（OpenSSL），可以得到 2 分的额外加分
- 3、完成 3.8.5 要求的，可以得到 1 分的额外加分
- 4、完成 5.6 要求的，可以得到 1 分的额外加分
- 5、在完成某个包含众多小文件的大文件夹（比如 QQ 的聊天记录文件夹，总大小 $\geq 1.5\text{GB}$ ，文件数量 ≥ 20000 ）从 A 机到 B 机的同步过程中，传输正确的所有组按速度（上传速度、写入数据库速度）排名，第 1 名可得到 3 分，与第 1 名相差时间 10%以内的可得到 2 分，相差时间 20%以内的可得到 1 分（得到 2 分/1 分的小组数量不限）
- 6、其它在当面验收汇报过程中的特色功能，可以根据实际情况得到不同加分
- 7、**本作业的必做要求实际得分在 60 分（百分制）以上，才能得到额外加分，否则不得分**

【分组及组队申报方法:】

- 1、分组规则见“20212-000001-W0001. 开篇-注意事项”
- 2、本次作业申报小组成员及选题目的截止时间为 **6 月 15 日 23:59:59（第十六周周二）**
- 3、如果不申报，默认为“20212-G00101. 小组作业 - 编程技巧类 - TCP 协议代理”的分组
- 4、申报信息发 QQ 群，不接受私聊