

Big-data-homework1设计说明书

任务概述

目标

实现简单文件同步器

运行环境

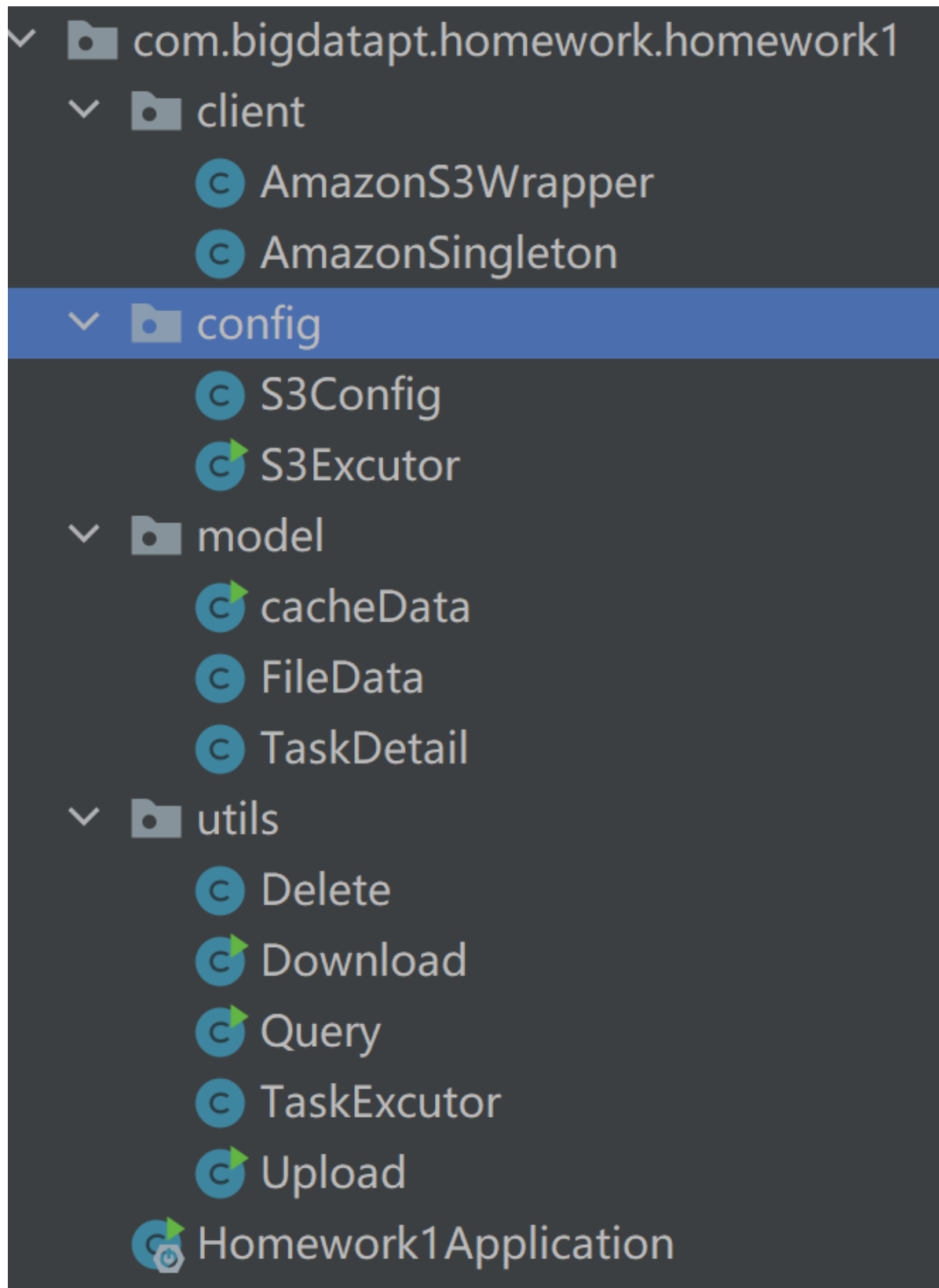
windows, jdk13

需求概述

- 制定本地某个目录与S3的某个bucket进行文件同步
- 程序启动时把bucket中的文件同步到本地，需要解决文件冲突
- 本地添加修改了文件，需要上传到S3；本地删除了文件，也要删除S3上对应的文件
- 对于超过20MB的文件，需要使用分块上传/下载，传输中断、程序重启可以继续原来的进度

总体设计

总体结构和模块外部设计



模块功能

client

- AmazonSingleton
提供amazon s3的单例实例

config

- S3Config

设定参数

```
bucketName:所需关联的桶的名称  
baseFilePath:本地关联的文件夹地址  
accessKey:S3 accessKey  
secretKey:S3 secretKey  
serviceEndpoint:S3服务器地址  
signingRegion:S3服务器注册地区  
partSize:大文件传输分块大小  
threadSize:大文件和小文件的大小分界线  
RefreshInterval_Second:定时任务刷新闻隔  
cachePath:缓存信息保存地址  
filterFile:文件传输过程中需要忽略的文件名  
Overdue_Second:缓存文件过期时间
```

- S3Excutor

定义初始化任务，定时任务，整个程序的核心执行器

model

- cacheData

结构化缓存，信息包括任务列表，上传etag列表，上传请求，S3文件快照

提供读取本地缓存文件以及保存缓存文件至本地的接口

- FileData

结构化文件信息，包括文件相对路径名，时间戳，是否为文件，文件大小

- TaskDetail

结构化任务信息，包括任务类型，数据大小类型，数据路径信息，对于大任务的分片执行位置

utils

- Delete

提供删除S3端文件的接口以及将需要删除的文件列表聚合成删除任务列表的接口

- Upload

提供上传S3端文件的接口以及将需要上传的文件列表聚合成上传任务列表的接口

- Download

提供下载S3端文件的接口以及将需要下载的文件列表聚合成下载任务列表的接口

- Query

提供查询S3端文件和查询本地文件的接口

- TaskExcutor

任务调度中心，提供的接口包括执行任务列表，单个任务或执行缓存中的任务列表，同时调度好相应的缓存更新功能

处理流程

1. 在程序启动时，检查本地cacheData.dat是否有保存之前剩余的任务信息
2. 如果cacheData.dat中保存有剩余任务信息，则先执行剩余的任务
3. 之前剩余的任务执行完成后，S3Excutor初始化，检查本地是否有保存S3服务器上文件的镜像信息，如果有则将镜像信息与现在S3服务器上保存的文件进行比较，下载镜像中没有的文件；再拿本地文件与S3上的文件进行对比，上传S3服务器中没有但是本地有的文件，如果遇到文件冲突，则以最后修改的文件为准；若没有S3服务器文件的历史快照，则将S3服务器上的文件与本地服务器上的文件进行同步下载及上传
4. S3Excutor初始化完成后将本地现有文件快照缓存，启动定时任务，每隔一定的时间执行一次
5. 每个定时任务执行时将本地文件与本地文件快照进行对比，将添加的文件和修改的文件上传，将删除的文件在云端删除

功能实现

- 单个bucket与本地路径绑定：在config中配置基本参数，AmazonSingleton通过单例模式绑定好一个S3客户端
- 启动时更新本地文件：S3Excutor设置启动任务，utils.Query查询本地文件与S3端文件，S3Excutor判定需上传的文件列表以及需下载的文件列表，utils中的工具类将文件列表聚合成任务列表，交由TaskExcutor执行
- 解决文件冲突：程序启动时本地上传，云端下载，遇到冲突的文件以最新的为准
- 执行时本地文件的变动同步到云端：定时任务每次执行后保存一份本地文件的快照，每次将快照与当前文件列表对比得出任务列表
- 大文件分块上传，中断重传：config中设置大小文件阈值以及分块的块大小，执行大文件传输任务时每传输完一个块就将任务列表中当前任务的currentpos标志位加一，同时将信息更新到缓存文件中，并且传输开始时将requestresult也缓存起来，这样传输中断时程序重启，优先读取缓存文件，获取任务列表，执行被中断的大数据传输任务时读取requestresult，得到request的id，然后读取currentpos得到大数据传输任务的分块位置，从该分块位置继续执行传输任务