采集数据网站：[okcoin.cn](http://okcoin.cn) [okcoin.com](http://okcoin.com) [huobi.com](http://huobi.com)

采集各项币种：[okcoin.cn](http://okcoin.cn) （国内比特币，国内莱特币） [okcoin.com](http://okcoin.com)（国际比特币，国际莱特币） [huobi.com](http://huobi.com)（国内比特币，国内莱特币，国际比特币） 共7项

采集币种内容：ticker（实时行情）depth（市场深度，okcoin有1，0.1，0.01 三种。huobi有0.01。）trades（交易记录） kline（K线，有多个档）

采集方法：websocket 和 rest。

数据保存方法：使用数据库，可以为MySQL。有没有可能试一试 Hbase 和Mongodb。

疑问：

1. 采集币种内容问题：okcoin的市场深度有3个档，是否都采集，还是只采集0.01。
2. 采集币种内容问题：kline的数据有多个档，是否都采集。还是只存储最小时差1min的数据，其余数据以后用1min数据计算得到。
3. 采集方法问题：个人认为使用websocket或者rest其中一种方法采集就足够了。 而使用websocket + rest的方法，需要保证采集数据一致性。当然双保险也不错。
4. 数据存储问题：使用关系数据库MySQL还是使用非关系数据库Mongodb。因为Mongodb这种使用键值对存放，非结构化，易于扩展。也许更适合。
5. 数据存储问题：市场深度数据不易保存，严老师邮件中最后没有提到市场深度，是不是就不存储市场深度了。。

解答：

1. 只采集市场深度0.01
2. 只采集K线1min
3. ticker用websocket+rest。trades、depth、kline用rest
4. 数据库用Mongodb
5. 市场深度按时间区分存储。

okcoin.cn okcoin.com huobi.com三个网站对比：

1. okcoin.cn 是国内比特币、莱特币网站，okcoin.com是国际比特币、莱特币网站，但2个网站的API接口几乎完全一致，可以直接使用之前的程序。
2. huobi.com是国内比特币、莱特币、国际比特币的网站，API接口也有rest 和 socket接口，程序需要稍微改动。目前我的进度是基本看完huobi.com的API文档，还没有编程连接测试。难度不大。

采集系统设计：

1. 主要任务为采集数据，运行环境应为本地服务器，长期采集数据。最好编写为后台服务。服务器程序，去除图形界面编写。
2. 为适应多种网站采集及提高扩展性，设计改用配置文件（数据库信息、采集网站、采集任务等） + 任务调度（使用线程池完成各项采集任务） + 数据处理（筛选数据，剔除无用数据，统一格式） + 数据库存储
3. 服务器程序，若要监控，再增加log日志系统。（暂无）

采集系统运行方式：

1. 编写配置文件。包括采集数据网站，内容、数据库信息等。
2. 主程序调用Config类对配置文件处理，初始化系统，创建Schedule类使用线程池发送各项采集任务指令。主程序应是个启动服务的后台进程，xx\_start.sh xx\_stop.sh
3. 由各项采集任务方法采集，回调DataFilter类处理数据。
4. 最后回调database类存入数据库。

预计改进系统的优点：

1. 易于修改：只需改变配置文件，可增加或减少采集任务。
2. 易于扩展：如增加新的采集网站，只需增写配置文件，采集任务方法，以及数据库建表。无需改变系统整体结构。
3. 模块切分：统一接口，如需更改数据库种类，只需修改配置文件，新增改写database类即可。