实习在字节做的大数据相关的工作，但实际做的事情大部分是为了业务服务的，所以一般按照操作流程来就没什么难点。

但公司发展太快，文档不全，甚至没有，所以就会在工作的时候卡壳

**业务上**比如有次我创建的新任务要代替旧的任务，所以要把旧任务关闭，再开启新任务，但我当时不知道这个顺序，以为这样会导致数据在关闭和打开之间会有gap，所以先打开了新任务再关闭了旧任务，但这样造成了数据重复，不过我从刚入职把我做的每个需求都会写一个操作文档，这样后面别人做的时候也可以看，

我的有个文档cover了我们组负责的整个数据流里各个操作步骤

所以之前就被几个同事拿去参考了，还是挺有用处的

**技术上**比如要从es里取数，不知道怎么拼接取数的dsl，所以先去接口测试工具里看请求的dsl和返回的dsl是怎么样的。根据这个去用数据结构来构造请求并且从返回的结构体里拿到想要的字段和字段值；还有会去打日志看读取的索引和集群，有次发现配置的runtimeprofile读取不到集群，后面才发现是没有配置物理集群，没有配置物理集群是因为没有数到es里，所以再追上游发现是hive2es的离线任务没有运行成功，因为是我创建的任务所以没有权限

**项目里的难点**比如我在开发发帖网站的时候觉得用户点赞和关注后如果是和其他逻辑同步进行的，这样就会要有些等待时间，而且点赞和关注是比较频繁的操作，所以如果数据量大的话就会造成更多延迟所以我想到用消息队列来实现，在触发事件之后，方法就能继续向下执行，因为消费者会新建一个线程进行操作，因为是异步的所以处理效率高，可以提高性能

而且如果帖子是个热帖，点赞的人非常高，可以用消息队列攒着慢慢处理