实战总结

《意向与线索分析主题》

1. 需求分析： 在数据库中一共有9张相关的表，分别是客户静态信息表、线索申诉信息表、客户线索表、客户意向表、员工信息表、班级表、校区表、学科表员工部门表，通过梳理，排除了班级表，8张表进行关联使用，并找出核心字段使用。
2. 建模实现：ODS,DWD,DWB,DM，总共分四层分析，确定了后期使用的拉链表、同步方式、存储方式、压缩方式、分割字段等细节，并且把8张表区分为2个事实表：线索申诉信息表，客户意向表，其余的为维度表。
3. 数据同步：使用sqoop采集数据 ，最后6张表使用shell脚本运行，其中user这个字段在运行时报错，在没有打开老师贴心的参考资料包时，百度找出‘user‘为shell中关键字的错误，需要加上反引号，同时需要加上反斜杠转换为无意义（大概就是这个意思）
4. 数据清洗：数据清洗上和组员有些分歧，但是讨论过后，将area, origin\_type, origin\_channel等字段进行清洗.
5. DWB+DM，这两层比较薄弱，技术上多半由组员图图实现，同时感谢他的技术支持。在B和M层中对于字段的精准抽取，没有理解，导致数据抽取过多，存在一些不必要的数据。grouping sets 里面的字段和我的目标表中的字段不一样，一直报错。
6. 其余的。。。。。还没有完全计算结束，仍需努力