# 新教育项目意向与分析总结

**维度: 年月日. 线上,线下. 新老学员 各地区 各学科 各校区 各来源渠道 各咨询中心**

**指标: 用户人数**

一;需求文档和建模设计文档

在写需求文档和建模设计文档时,第一次写没有考虑的很全面,很多表关系没有梳理清楚,以至于后来DWB层在joi时就卡住了,推倒重来很多次,浪费了很多时间.

二;ODS层和DWD层建设

在做ODS层时和DWD层时发现HIVE支持的类型很少,常用的也就string,int,之前这方面完全模糊,还有USER在hive里是关键字,所以建表时需要加反引号,

三; DWB和DM层建设

在做DWB时层因为需求文档和建模文档的不成熟,很多字段没有去看数据,导致在join 6张表时,很多指标和维度无法计算,缺少字段时,hive报错不明显,白白浪费很多时间

DM层时,在做grouping sets 时 ,判断的所有维度字段都需要用()括起来,最后还要用()把所有的维度组合括起来,当时漏加了最外面的(),运行居然不报错,可运行之后DM层的宽表里time\_type和group\_type字段只有一个值,排错排了我好几个小时,最后才找到原因,还有

总结:经过这个项目,更加了解数仓每层的功能职责,hive里支持什么字段类型,加深了之前学的函数印象,比如grouping grouping sest cast....学会了看jobhistory报错日志,解决了好几个报错,然后再hive建表时字段名单词一定要仔细认真,尤其在复制粘贴时,很容易报错,还有需求文档和设计文档一定要花时间去理清表关系,理解需求,画出表关系图,这对后面的工作有很大的帮助

总结人:涂茂林