**个人信息**

♦ 佟永旭 沈阳工业大学(2010-2014) 土木工程 15510581695

♦ [xuriyongjiu@sina.com](mailto:xuriyongjiu@sina.com) 微信: xuriyong9 Bilibili: 佟昶(chang)

**工作经验**

2020.06- 2021.10 洋钱罐(互联网金融) 数据仓库

♦ 日常工作

货后催收业务数据建模

调度系统基于 azkaban 二次开发

2019.08-2020.01 绿湾(百度系创业公司) 数据挖掘组

公司业务：为公安系统提供数据服务 离职原因：公司裁员

2018.05-2019.08 去哪儿网 机票事业部-数据仓库

♦ 日常工作

数据ETL，报表开发，数仓底层平台开发维护，sql优化，宽表设计

支持业务线数据应用（支付干预，航司关注度……），用户画像接口开发

♦ 重点工作

♦ **数据开发平台**

业务背景：其它team 会在数仓基础上洗表跑数据，存在建表不规范，任务执行效率低，使用脚本开发不便于监控

业务目标：提高开发效率，实现任务监控

实现方式：

1、提供sql接口，不再需要shell，只需会写sql，其它team同学(算法/数据分析师)可以在其上进行开发

2、建表和job 提交需要经过数仓同学审核后才能上线，避免不规范的表和sql 被提交上线

3、在任务审核时对执行效率较低的sql 进行优化

4、构建数据血缘，解析sql可以解析出表、任务之间的依赖关系，生成表、任务的血缘关系图

上线效果：

每天在该平台上跑的任务 150-200 个，cove 90% 数据需求，并获得了Q3 项目奖

♦ **数仓元数据安全架构**

业务背景：直接暴露数仓元数据给其它业务、算法团队，存在严重数据安全漏洞

业务目标：保证数仓元数据安全

实现方式：

1、收回其它团队直连mysql元数据库的权限，开启 hive metastore 服务, 用户连接Hive metastore

2、在hive 配置文件中增加白名单配置项

3、修改hive metastore 部分源码，使用户只能连接数仓团队hive版本，防止用户自行修改源码绕过白名单验证

4、加入了采集客户端sql的功能，并利用开发平台的sql 解析接口，实现了数据血缘的构建

♦ **实时标签系统**

业务背景：对于大量的业务线实时标签需求（黑名单用户、用户近30分钟搜索航线、学生用户绑定解绑、国内国际ONP……），以前的实现方式是定制化开发，带来的问题是开发周期长，每个标签都要独立开发不能做到通用

改进方案：通过 spark 实现实时标签配置化

1、接收kafka消息，利用 structured-streaming 解析出一张临时宽表

2、通过select临时宽表，使用sql实现业务逻辑，并注册成结果表

3、select 结果表并通过自定义UDF将实时将数据写入hbase或redis，实现热发布

上线效果：

1、实现了一个新的实时标签只需配置好sql即可，开发周效率提高了50%

2、通用化，目前该系统可以满足70%的实时标签需求

2016.05-2018.04 搜房网 大数据平台

♦ 数据产品开发：二手房画像，新房潜客推荐，数据魔方

♦ 数据仓库建设：数据建模，宽表设计，数据ETL

2014.07-2016.04 建筑行业

**专业技能**

♦ 有java web开发经验，java 基础功底较好，对jvm,nio,java多线程有深入了解

♦ 良好的线上故障解决能力，精读过 hadoop hive hbase 源码

♦ 熟练使用hive,spark sql,presto等数据批处理工具

♦ 了解并使用 hbase,kudu等分布式数据库