**个人信息**

♦ 姓名：佟永旭 性别：男 学校：沈阳工业大学 电话：15510581695

♦ 学历：本科 民族：汉 专业：土木工程 邮箱：xuriyongjiu@sina.com

**工作经验**

2018.08-至今 去哪儿网 机票事业部-数据仓库

♦ 日常工作

♦ ETL，数据清洗，报表开发，数仓底层优化，sql优化，宽表设计

♦ 支撑业务线数据应用（支付干预，航司关注度，学生用户绑定解绑……），提供画像接口

♦ 重点工作

♦ 数据开发平台 > 建表、数据清洗、报表开发的平台，优点：

1、不再需要shell，只需会写sql，降低开发成本，算法、产品人员可以在其上进行开发

2、方便任务调度，以http形式提供执执行接口，与调度系统对接，更加轻量级；sql优化，

数仓同学对用户提交的sql可以方便直观地进行优化

3、构建数据血统，可以通过开发平台上的表、任务之间的依赖关系，生成表、任务的血统关系图，

对数据仓库来说意义重大

♦ 数仓安全架构 >业务背景：其它业务、算法团队需要连接数仓元数据，不利于数据安全

实现方式：

1、收回其直连mysql元数据库的权限，转而连接hive metastore

2、数仓团队通过修改hive metastore 部分源码控制客户端只能连接修改后的hive版本，并强制开启白名单验证。同时又实现sql采集，对数仓团队本身和其它团队，即整个数据仓库上所有执行的sql进行收集，从而实现数仓完整的血统关系

♦ 实时标签系统设计与开发 >背景：业务线有相当多的实时标签需求（黑名单用户、

用户近30分钟搜索航线、学生用户绑定解绑……），以前的实现方式是web线上发布的形式，

问题是开发周期比较长，每个标签都要独立开发不能做到通用

改进方案：通过structured-streaming实现实时标签配置化，

1、从kafka解析出一张临时宽表

2、通过select临时宽表，实现业务逻辑注册成结果表

3、select 结果表将实时将数据写入hbase或redis，可实现热发布，而且可以实现实时窗口需求

优点：

1、实现了一个新的实时标签只需配置好sql即可，大节节省了开发周期

2、通用化，目前这个系统可以满足70%的实时标签需求，并减少了大量的开发工作量

3、由于数据源统一，多个标签可利用同一个临时宽表，从而节省集群资源

2017.09-2018.08 掌合天下 研发中心-数据仓库

♦ 主要工作

♦ 从零到一数据仓库建设，包括：数据建模，数据分层，宽表设计，数据ETL，

任务调度

♦ 报表开发

♦ 重点难点：业务比较复杂，为了保证扩展性，对数据建模和数据分层的要求比较高。为了兼顾

数据处理的快速，和数据存储节省空间，采用spark sql做为计算引擎，parquet做为存贮格式，

gzip和snappy做为压缩格式

2014.09-2017.09 搜房网 大数据平台

♦ 主要工作

♦ java web开发

♦ 数据产品开发：二手房画像，新房潜客推荐，数据魔方

**专业技能**

♦ 一年java web开发经验

♦ 线上故障解决能力，深入源码，定位问题，通过hive源码修改保障数据安全和构建数据血统

♦ 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

♦ 了解并使用 spark structured-streaming,flink,kafka 实时处理工具

**自我评价**

♦ 工作认真负责，勇于承担责任

♦ 做事认真踏实，生活中乐于助人，能够妥善的处理周围的人际关系

♦ 乐于接受挑战，乐于学习，有较强的上劲心

♦ 多次参加马拉松比赛