# 星云系统简介

星云系统是一个整合用户行为数据收集，行为数据清洗，数据持久化，后台数据分析，数据报表生产发送，推荐系统物料数据接口API等一整套，全流程的数据系统。重点解决页面数据收集过程中，页面埋点接入复杂，页面埋点以侵入式方式，页面用户数据收集单一，不灵活等痛点问题。

星云数据系统收集模块采用WebSocket的数据请求交互模式，采用Xpath的定位方式进行位置定位，数据收集类型分为浏览、点击、曝光，用户设备ID区别用户，并区分各用户手机型号、浏览器型号，对用户行为进行全流程，浏览全链路监控收集。服务端采用Kakfa 消息通道，方面各业务线对原始数据的不同业务处理。实时数据通过一层MongoDb间接存储，最终T+1的方式持久化的Hadoop的Hive中，方便后续数据分析。

后台数据分析，通过编写SQL，可以实时计算出SQL结果，并通过创建邮件模版页面，对应制定SQL任务，得出数据报表，通过定时任务，定时将数据报表发送邮件。

通过用户行为数据加上相关业务数据使用Spark MLlib，对用户进行全角度，多方案业务数据推荐，并结合数据回流，ABTest 评估数据对推荐算法实时调整优化。

通过kafka原始消息通过可以做到实时数据报表、实时数据推荐等。