

1、 演讲主题；

动态实时标签处理平台架构

2、 演讲主题简介；

1).概要介绍：

在携程国际业务上，因为面临的市场多，产品和业务复杂多样，投放渠道多，引流费用高，因此我们需要对业务和产品做更精细化的管理和优化，满足市场投放和运营的需要，降低成本，提升运营效率，提升转化率。为此我们提出研发动态实时标签处理平台（以下简称 CDP），为 Trip 业务增长解决“Grow Revenue”和“Reduce Costs”的问题。

基于我们的业务需要，我们将业务数据标签筛选的场景分为两大类

第一类是实时触发场景，根据业务需要，配置动态规则，实时订阅业务系统的变更消息，筛选出满足动态规则条件的数据，通过消息的方式推送到下游业务方。

第二类是标签持久化场景，根据业务需要，将业务系统的实时业务变更消息按照业务需要加工成业务相关的特征数据持久化存储到存储引擎，业务根据业务需要组装查询条件查询引擎数据，主要是 OLAP（分析类）和 OLTP（在线查询）两大类查询。

为了解决以上问题，我们设计开发了一套“实时动态标签处理系统”，业务方只需要按照我们的基本算子规则配置提交任务，系统就会自动解释执行规则，按照配置要求执行数据处理操作，目前支持的基本算子有 Stream（流式数据接入目前支持 QMQ 和 Kafka）、Priority（优先级判断）、Join、Filter（过滤）、Sink（数据输出，目前支持 TiDB、Redis、QMQ）等等，这个我们在整体设计里面会详细介绍，通过规则和动态计算的方式提升数据处理和开发效率，降低开发成本。

流式数据采用类 Kappa 架构，标签持久化采用类 Lambda 架构。

2).内容大纲：

1.背景

2.系统架构设计

2.1 系统总体架构设计

2.2 动态标签处理设计

2.3 标签持久化

3.动态实时标签处理

3.1 实时标签处理流程

3.2 流式数据处理算子设计

3.3 流式数据处理 UDF 设计

3.4 流式数据处理 DAG 设计

4.标签持久化

4.1 标签持久化存储方案

4.2 标签动态更新

4.3 OLTP 查询服务支持

4.4 OLAP 查询服务支持

5.业务应用

3、 演讲适合听众范围（即用户需要什么样技术基础，才能更好地交流和学习，并从中获益）；

在工作中对流式数据处理、系统/平台架构有一定实践或思考的同学。比如：

1).对流式数据处理学习比较多，但是实践少的同学

2).对围绕流式数据处理，标签持久化存储和使用开展工作，渴望寻找技术方向的同学