

Flink SQL 实战：广告点击流和曝光流的场景应用

作者：[fc13240](#)

原文链接：<https://ld246.com/article/1608374973896>

来源网站：[链滴](#)

许可协议：[署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

本文主要介绍在流式场景中 join 的实战。大家都知道在使用 SQL 进行数据分析的过程中，join 是经常使用的操作。在离线场景中，join 的数据集是有边界的，可以缓存数据有边界的数据集进行查询，Nested Loop/Hash Join/Sort Merge Join 等多表 join；而在实时场景中，join 两侧的数据都是无界的数据流，所以缓存数据集对长时间 job 来说，存储和查询压力很大，另外双流的到达时间可能不致，造成 join 计算结果准确度不够；因此，FlinkSQL 提供了多种 join 方法，来帮助用户应对各种 join 场景。

本文主要介绍 regular join/interval join/temporal table join 这种 3 种 join 的实战应用，主要包含以下几个部分：

- 数据准备
- Flink SQL join 之 regular join
- Flink SQL join 之 interval join
- Flink SQL join 之 temporal table join
- 总结

01 数据准备

一般来说大部分公司的实时的数据是保存在 kafka，物料数据保存在 MySQL 等类似的关系型数据库，根据 Flink SQL 提供的 Kafka/JDBC connector，我们先注册两张 Flink Kafka Table 以及注册一张 Flink MySQL Table，明细建表语句如下所示：

- 注册 Flink Kafka Table, 作为两条需要 join 的数据流；对于 点击流，我们定义 Process time 时属性，用来做 temporal table join，同时也定义 Event Time 和 watermark，用来做双流 join；对于 光流，我们定义 Event Time 和 watermark，用来做双流 join。

```
DROP TABLE IF EXISTS flink_rtdw.demo.adsdw_dwd_max_click_mobileapp;
CREATE TABLE flink_rtdw.demo.adsdw_dwd_max_click_mobileapp (
  ... publisher_adspace_adspaceId INT COMMENT '广告位唯一ID',
  ... audience_behavior_click_creative_impressionId BIGINT COMMENT '受众用户点击的广告创
的ImpressionId',
  audience_behavior_click_timestamp BIGINT COMMENT '受众用户点击广告的时间戳(毫秒)',
  ... procTime AS PROCTIME(),
  ets AS TO_TIMESTAMP(FROM_UNIXTIME(audience_behavior_click_timestamp / 1000)),
  WATERMARK FOR ets AS ets - INTERVAL '5' MINUTE
) WITH (
  'connector' = 'kafka',
  'topic' = 'adsdw.dwd.max.click.mobileapp',
  'properties.group.id' = 'adsdw.dwd.max.click.mobileapp_group',
  'properties.bootstrap.servers' = 'broker1:9092,broker2:9092,broker3:9092',
  'properties.sasl.jaas.config' = 'org.apache.kafka.common.security.scram.ScramLoginModule r
quired username="kafka-administrator" password="kafka-administrator-password";',
  'properties.security.protocol' = 'SASL_PLAINTEXT',
  'properties.sasl.mechanism' = 'SCRAM-SHA-256',
  'avro-confluent.schema-registry.url' = 'http://schema.registry.url:8081',
  'avro-confluent.schema-registry.subject' = 'adsdw.dwd.max.click.mobileapp-value',
  'format' = 'avro-confluent'
);
```

- 注册 Flink Mysql Table, 作为维度表

```
DROP TABLE IF EXISTS flink_rtdw.demo.adsdw_dwd_max_show_mobileapp;
```

```
CREATE TABLE flink_rtdw.demo.adsdw_dwd_max_show_mobileapp (  
  ... audience_behavior_watch_creative_impressionId BIGINT COMMENT '受众用户观看到  
广告创意的ImpressionId',  
  audience_behavior_watch_timestamp BIGINT COMMENT '受众用户观看到广告的时间(毫秒)',  
  ... ets AS TO_TIMESTAMP(FROM_UNIXTIME(audience_behavior_watch_timestamp / 1000  
)  
)  
  , WATERMARK FOR ets AS ets - INTERVAL '5' MINUTE  
) WITH (  
  'connector' = 'kafka',  
  'topic' = 'adsdw.dwd.max.show.mobileapp',  
  'properties.group.id' = 'adsdw.dwd.max.show.mobileapp_group',  
  'properties.bootstrap.servers' = 'broker1:9092,broker2:9092,broker3:9092',  
  'properties.sasl.jaas.config' = 'org.apache.kafka.common.security.scram.ScramLoginModule r  
quired username="kafka-administrator" password="kafka-administrator-password";',  
  'properties.security.protocol' = 'SASL_PLAINTEXT',  
  'properties.sasl.mechanism' = 'SCRAM-SHA-256',  
  'avro-confluent.schema-registry.url' = 'http://schema.registry.url:8081',  
  'avro-confluent.schema-registry.subject' = 'adsdw.dwd.max.show.mobileapp-value',  
  'format' = 'avro-confluent'  
)  
);
```

更多内容请访问: [Flink SQL 实战: 广告点击流和曝光流的场景应用](#)