

Flink SQL 如何实现数据流的 Join?

作者介绍:

林小铂，网易游戏高级开发工程师，负责游戏数据中心实时平台的开发及运维工作，目前专注于 Apache Flink 的开发及应用。探究问题本来就是一种乐趣。

本文授权转自: <http://1t.click/b5k3>

无论在 OLAP 还是 OLTP 领域，Join 都是业务常会涉及到且优化规则比较复杂的 SQL 语句。对于离线计算而言，经过数据库领域多年的积累，Join 语义以及实现已经十分成熟，然而对于近年来刚兴起的 Streaming SQL 来说 Join 却处于刚起步的状态。

其中最为关键的问题在于 Join 的实现依赖于缓存整个数据集，而 Streaming SQL Join 的对象却是无限的数据流，内存压力和计算效率在长期运行来说都是不可避免的问题。下文将结合 SQL 的发展解析 Flink SQL 是如何解决这些问题并实现两个数据流的 Join。

<https://ververica.cn/developers/how-does-flink-sql-realize-data-flow-join/>