

1. 介绍

Structured Streaming是Spark2.0版本提出的新的实时流框架（2.0和2.1是实验版本，从Spark2.2开始为稳定版本），相比于Spark Streaming，优点如下：

1. 同样能支持多种数据源的输入和输出，参考如上的数据流图
2. 以结构化的方式操作流式数据，能够像使用Spark SQL处理离线的批处理一样，处理流数据，代码更简洁，写法更简单
3. 基于Event-Time，相比于Spark Streaming的Processing-Time更精确，更符合业务场景
4. 解决了Spark Streaming存在的代码升级，DAG图变化引起的任务失败，无法断点续传的问题（Spark Streaming的硬伤！！！）

原文链接：<https://blog.csdn.net/lovechendongxing/article/details/81748237>