

对比点	Storm	Spark Streaming
实时计算模型	纯实时，来一条数据，处理一条数据	准实时，对一个时间段内的数据收集起来，作为一个RDD，再处理
实时计算延迟度	毫秒级	秒级
吞吐量	低	高
事务机制	支持完善	支持，但不够完善
健壮性 / 容错性	ZooKeeper, Acker, 非常强	Checkpoint, WAL, 一般
动态调整并行度	支持	不支持

Spark Streaming与Storm的优劣分析

事实上，Spark Streaming绝对谈不上比Storm优秀。这两个框架在实时计算领域中，都很优秀，只是擅长的细分场景并不相同。

Spark Streaming仅仅在吞吐量上比Storm要优秀，而吞吐量这一点，也是历来挺Spark Streaming，贬Storm的人着重强调的。但是问题是，是不是在所有的实时计算场景下，都那么注重吞吐量？不尽然。因此，通过吞吐量说Spark Streaming强于Storm，不靠谱。

事实上，Storm在实时延迟度上，比Spark Streaming就好多了，前者是纯实时，后者是准实时。而且，Storm的事务机制、健壮性 / 容错性、动态调整并行度等

特性，都要比Spark Streaming更加优秀。

Spark Streaming，有一点是Storm绝对比不上的，就是：它位于Spark生态技术栈中，因此Spark Streaming可以和Spark Core、Spark SQL无缝整合，也就意味着，我们可以对实时处理出来的中间数据，立即在程序中无缝进行延迟批处理、交互式查询等操作。这个特点大大增强了Spark Streaming的优势和功能。

来自 <http://blog.csdn.net/kwu_ganymede/article/details/50296831>