这两天遇到一个很奇怪的问题，利用receiver和direct两种方式来消费kafka中的数据，官网说direct形式的效率更高但是实际效果却不是很好没有那么快。

这是因为经过查看运行时间表发现，每个批次都等待了3秒的时间。因此联想到spark.locally.wait的默认等待时间是三秒因此可能是为了等待数据的本地化导致的时间延长，因此把时间调低来解决。本地化策略是为了减少数据的网络传输因此在batch比较大的时候本地化策略会有比较好的效果，而在数据量小的时候等待时间比处理时间还要长就得不偿失了，因此该参数的配置需要不断的调试。