Storm给我的感觉不那么好，所以就在开发层面总结一下好了，然后在搜集一下有没有相关的面试题：

首先我任务就是理解：

1. 并行度
2. 流分组策略
3. Trident
4. DRPC

后两者主要是了解

并行度把握几个参数：

Workers就是机器 w

Bolt的并行度就是executor个数 e

TaskNumbers就是多少execute方法执行 t，最终是以task为单位运行的

真正落实到机器上一般是平均分配：et/w

这个里面要注意messageId和streamId一个是为了保证可靠性，一个是为了同时发出多个流，用来区分，注意多个流不是分区或者task成面的，而是bolt层面的，看拓扑图就比较清楚。

关于流分组策略，虽然官方给了很多，但我只想讲几个常用的：

Shufﬂe grouping,Fields grouping,Direct grouping还有一个localOrShuffle

Shuffle其实也不用太在意，理解为佛系分组就好

Filed grouping也好理解，同一个filed会到相同的分区，

直接分组就是保证把消息给bolt中某一个特定的task，而且这些都是试一次就会的。

Trident就好像spark抛开rdd编程而产生的sql一样，就是在编码上偷懒，我们公司还是传统的bolt编程。

关于Trident

Trident是准实时处理，从他给的几个api就可以看出，其接收到消息并不是继续流转，而是进行累计到一个batch数，即批次，（这里要考虑一下，如果消息不足批次数难道就不发送了么）

书中这么写的：Trident有连接，聚合，分组，函数，过滤器，直接支持数据库和持久化存储以及增量处理，有一致性和exact once语义。

跨batch聚合

里面的处理单元由原先的bolt编程function