storm出现的背景

互联网从诞生的第一时间起，对世界的最大改变就是让信息能够实时交互，从而大大加速了各个环节的效率。正因为大家有对信息实时响应、实时交互的需求，所以软件行业除了个人操作系统之外，数据库（更精确的说是关系型数据库）应该是发展最快、收益最为丰厚的产品了。记得十年前，很多银行别说实时转账，连实时查询都做不到，但是数据库和高速网络改变了这个情况。

随着互联网的更进一步发展，从Portal信息浏览型到Search信息搜索型到SNS关系交互传递型，以及电子商务、互联网旅游生活产品等将生活中的流通环节在线化。对效率的要求进一步提升了对实时性的要求，而信息的交互和沟通正在从点对点向信息链，甚至信息网的方向发展，这样必然带来数据在各个维度的交叉关联，数据爆炸已不可避免。因此流式处理加NoSQL产品应运而生，分别解决实时框架和数据大规模存储计算的问题。早在7、8年前，诸如UC伯克利、斯坦福等大学就开始了对流式数据处理的研究，但是由于更多地关注于金融行业的业务场景或者互联网流量监控的业务场景，以及当时互联网数据场景的限制，造成了研究多是基于对传统数据库处理的流式化，对流式框架本身的研究偏少。目前这样的研究逐渐没有了声音，工业界将更多的精力转向了实时数据库。

2010年，Yahoo！对S4的开源、2011年Twitter对Storm的开源，改变了这个情况。以前互联网的开发人员在做一个实时应用时，除了要关注应用逻辑计算处理本身外，还要为了数据的实时流转、交互、分布大伤脑筋。但是现在情况却大为不同，以storm为例，开发人员可以快速地搭建一套健壮、易用的实时流处理框架，配合SQL产品、NoSQL产品或者MapReduce计算平台，就可以低成本地做出很多以前很难想象的实时产品。例如，一淘数据部的量子恒道品牌旗下的多个产品就是构建在实时流处理平台上的。