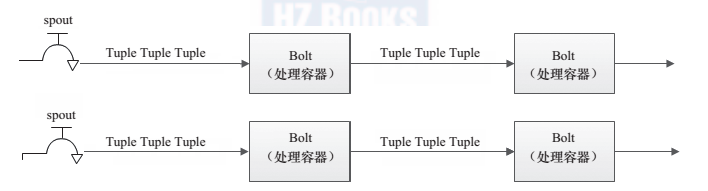
Storm的设计思想

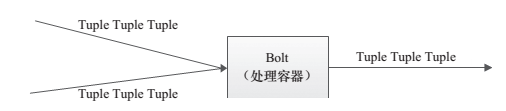
在Storm中也有对于流（Stream）的抽象，流是一个不间断的、无界的连续Tuple（Storm在建模事件流时，把流中的事件抽象为Tuple即元组）。Storm认为每个流都有一个Stream源，也就是原始元组的源头，所以它将这个源头抽象为Spout，Spout可能连接Twitter API并不断发出推文（Tweet），也可能从某个队列中不断读取队列元素并装配为Tuple发射。

有了源头即Spout也就是有了流，同样的思想，Twitter将流的中间状态转换抽象为Bolt，Bolt可以消费任意数量的输入流，只要将流方向导向该Bolt，同时它也可以发送新的流给其他Bolt使用，这样一来，只要打开特定的Spout（管口），再将Spout中流出的Tuple导向特定的Bolt，由Bolt处理导入的流后再导向其他Bolt或者目的地。

假设Spout就是一个一个的水龙头，并且每个水龙头里流出的水是不同的，想获得哪种水就拧开哪个水龙头，然后使用管道将水龙头的水导向到一个水处理器（Bolt），水处理器处理后使用管道导向另一个处理器或者存入容器中。下面为Spout、Tuple和Bolt之间的关系和流程。

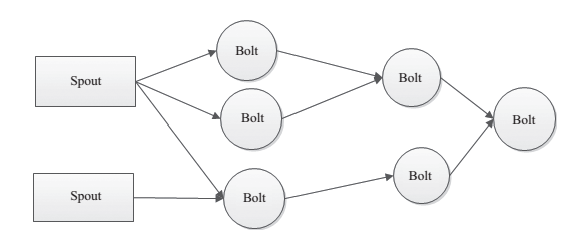


**Spout、Bolt顺序处理数据流图**



**Bolt多输入数据流图**

为了增大水处理效率，可以在同一个水源处接上多个水龙头并使用多个水处理器，如图所示



**多Spout、多Bolt处理流程图**

对应上文的介绍，可以很容易地理解上图，这是一张有向无环图。Storm将这个图抽象为Topology（即拓扑），拓扑是Storm中最高层次的一个抽象概念，提交拓扑到Storm集群执行，一个拓扑就是一个流转换图。图中的每个节点时一个Spout或者Bolt，图中的边是指Bolt订阅了哪些流。