Worker、Task、Executor三者之间的关系

Storm集群中的一个物理节点启动一个或者多个Worker进程，集群的Topology都是通过这些Worker进程运行的。

然而，Worker进程中又会运行一个或者多个Executor线程，每个Executor线程只运行一个Topology的一个组件（Spout或Bolt）的Task任务，Task又是数据处理的实体单元。

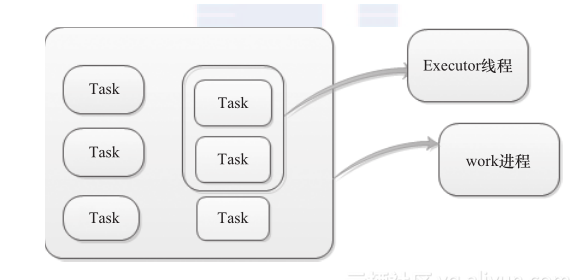
Worker是进程，Executor对应于线程，Spout或Bolt是一个个的Task;

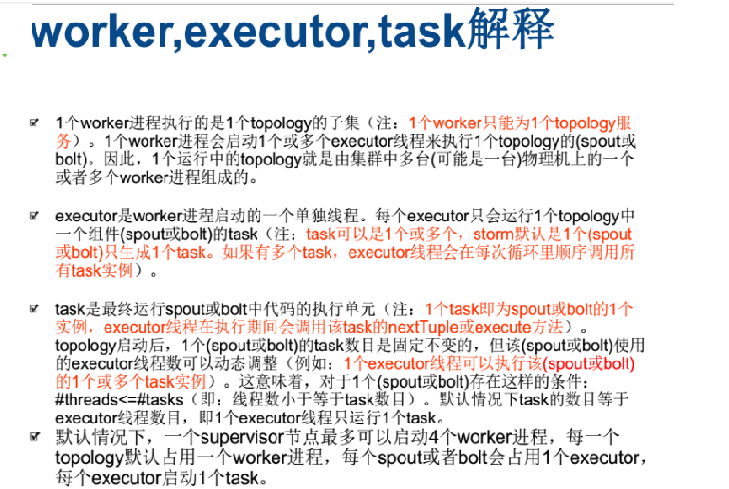
同一个Worker只执行同一个Topology相关的Task；

在同一个Executor中可以执行多个同类型的Task，即在同一个Executor中，要么全部是Bolt类的Task，要么全部是Spout类的Task；

在运行时，Spout和Bolt需要包装成一个又一个Task。

Worker、Task、Executor之间的关系如图所示。





storm程序主要由Spout和Bolt组成。Spout和Bolt在运行期间会生成task实例（new Spout或者new Bolt）。那这些task实例是需要在线程（executor）里面运行的，而线程是需要在进程（worker）里面执行的。

【补充】

Spout与Bolt的运行流程：

bolt declareOutputFields

spout declareOutputFields

bolt prepare

spout open

spout nextTuple

spout  nextTuple

bolt execute