

oozie investigate

OOZIE COORDINATOR

COORDINATOR CONCEPT 概念

coordinator application: A coordinator application is a template to define the triggers or predicates to launch a workflow.

coordinator 应用程序是一个模版，用来通过定义触发器来启动workflow

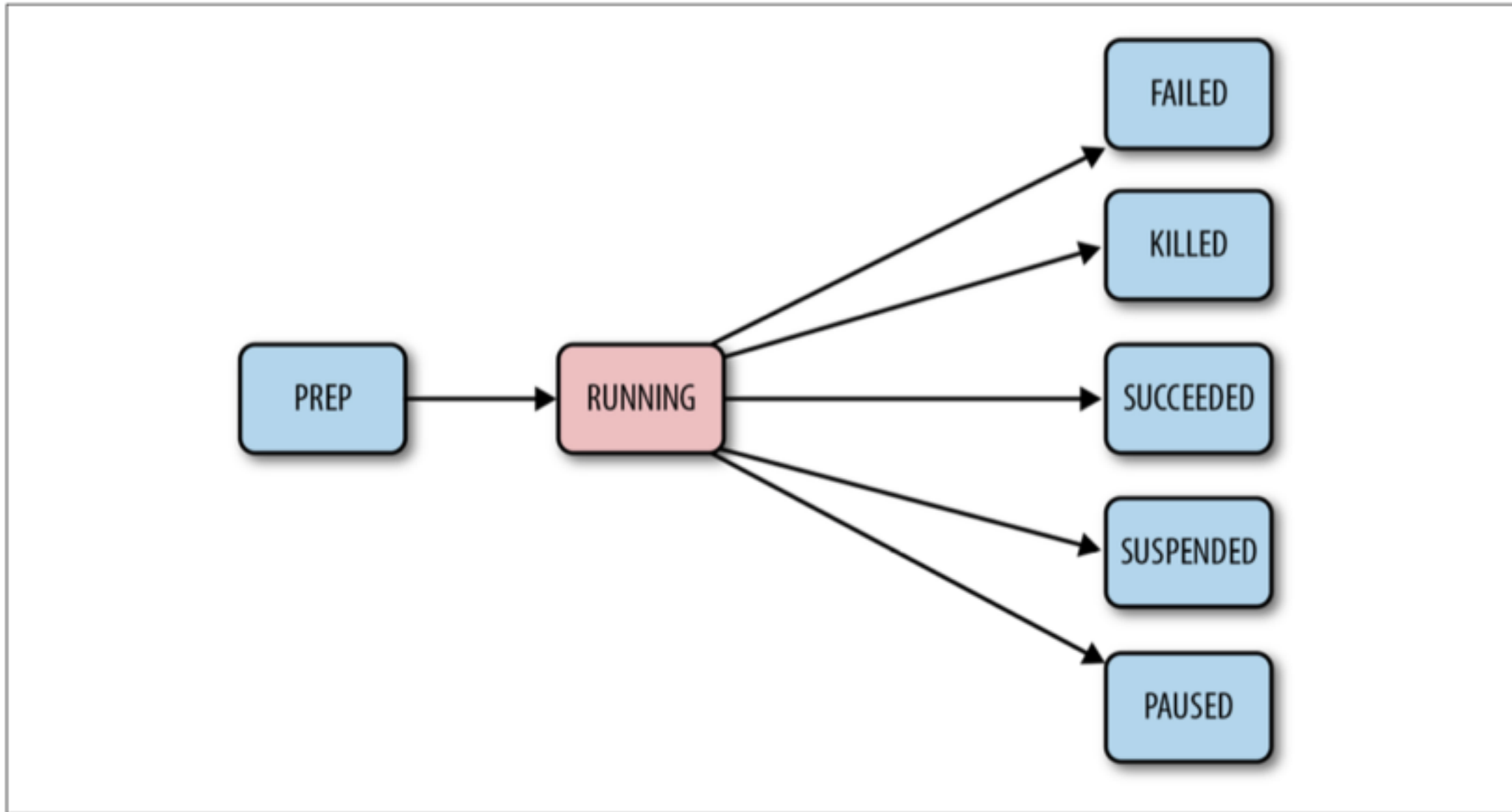
coordinator job: When a coordinator application is submitted to Oozie with all its parameters and configurations, it is called a coordinator job.

当一个coordinator 应用程序带着相应的配置文件以及参数被提交给oozie后，就会产生一个coordinator job（运行时实例）

coordinator action: A coordinator job regularly creates/materializes a new coordinator action for each time instance based on its start time and frequency.

coordinator job会依据start time 和frequency定时实例化出一个个coordinator action （运行时实例）

COORDINATOR JOB 状态图



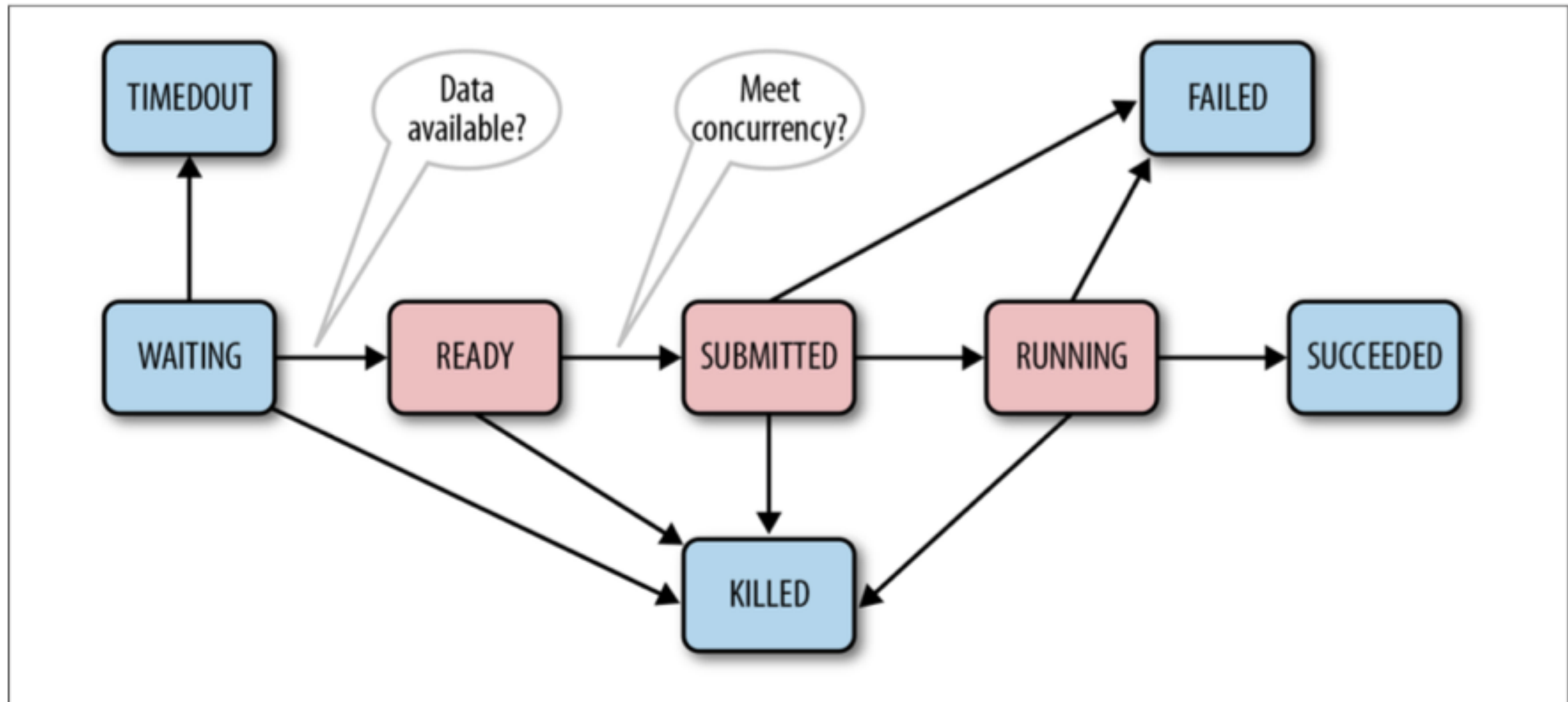
When a coordinator job is submitted to the Oozie service, Oozie parses the coordinator XML and validates the configurations.

After that, Oozie returns a coordinator job ID and puts the job in PREP state.

Oozie keeps the job in PREP state until it reaches the start time.

在达到start time之前，Oozie会保持job的状态为PREP，start time之后job的状态会变为RUNNING

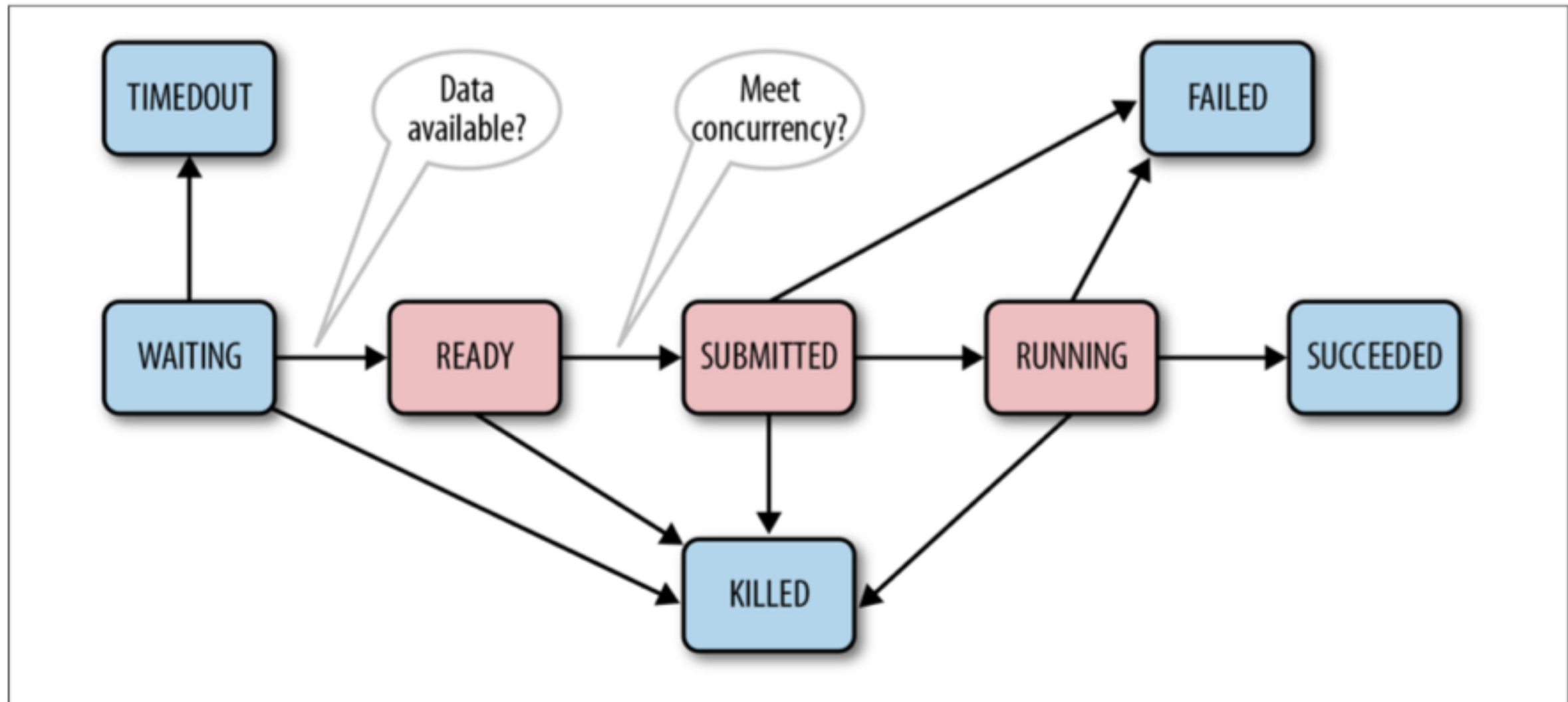
COORDINATOR ACTION 状态图



a coordinator job creates or materializes a coordinator action for a specific time instance (a.k.a. nominal time).

1. coordinator job 在每个nominal time都会创建一个action， action首先会进入WAITING状态；
2. action 如果在设定的时间内满足data trigger， 那么action状态从WAITING变为READY， 否则进入TIMEOUT
3. 处于READY状态的action如果在限定的并发数以内可以转为SUBMITTED状态， 也就表示action真正submit了一个workflow， 从而进入workflow的状态

COORDINATOR ACTION 状态图



值得注意的是READY—>SUBMITTED时，如果同时有多个coordinator处于READY状态，除了满足concurrency外，还需要考虑execution order，有三种选择：FIFO，LIFO，LAST_ONLY.

The default is FIFO (First in First Out)

The LIFO (Last In First Out) asks Oozie to execute the latest action first.

LAST_ONLY means execute *only* the last one and discard the rest.