Flink on zeppelin从checkpoint&savepoint 自动恢复任务

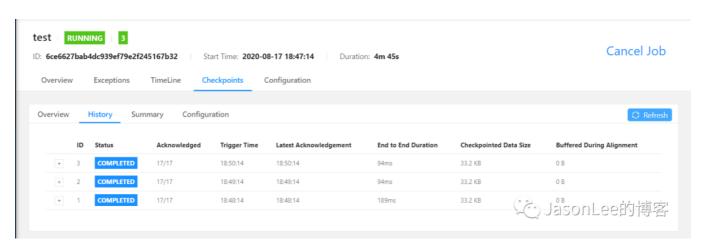
在上一篇的文章中介绍了Flink on zeppelin怎么配置checkpoint,并且从指定的checkpoint恢复任务,但是这个过程还是相对复杂,对用户来说不是很友好,在最新的版本中,zeppelin支持了自动从checkpoint&savepoint恢复任务,只需要添加几个参数即可,下面就来介绍一下如何使用.

1.checkpoint自动恢复

1.1 指定checkpoint恢复



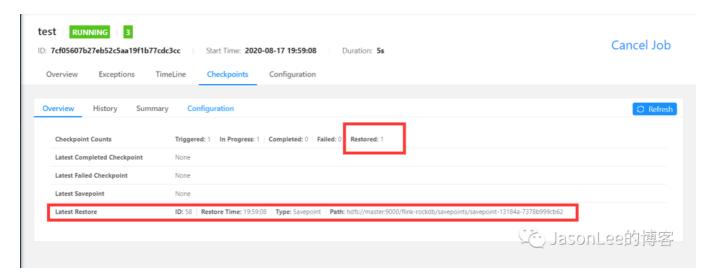
先启动一个简单的任务,点击右上角的Flink Job 看下Flink的web ui 任务的checkpoint情况.



可以看到checkpoint已经完成了3次了,现在把任务停掉,从checkpoint恢复任务.



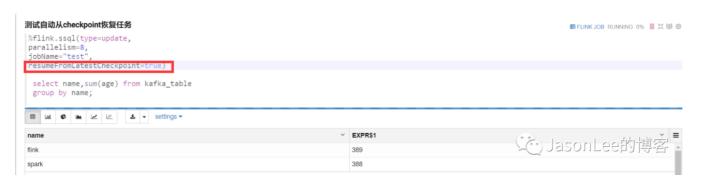
上一篇文章说的 execution.savepoint.path 这个参数也可以直接设置到 paragraph 里面, 这样不同的任务可以设置从不同的checkpoint启动.下面的 sum 结果不是从0开始的 是接着上一次的结果计算的,说明从checkpoint启动成功了.再到 UI 上面看一下



可以看到任务是从上一次的checkpoint恢复的.

1.2 自动从最近的checkpoint恢复

如果不想设置 想直接启动的时候就从最近一次成功的checkpoint恢复,设置一下 resumeFromLatestCheckpoint=true 这个参数就可以了.



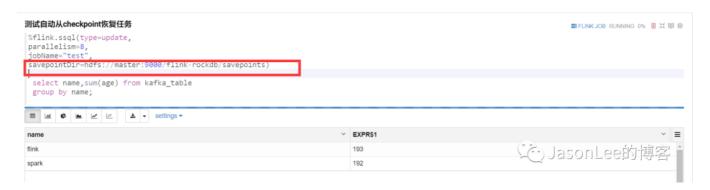
zeppelin会定期的通过Flink的restful api 去获取checkpoint的信息,并把最近一次成功的保存在 note 里面 这样在下次启动的时候就能直接从保存的 checkpoint 路径启动任务,但是如果你设置保留了多个 checkpoint,不想从最近的启动,这个时候就不能用这种自动启动了,需要用上面那个参数去指定你想要恢复的checkpoint路径即可.

2.savepoint自动恢复

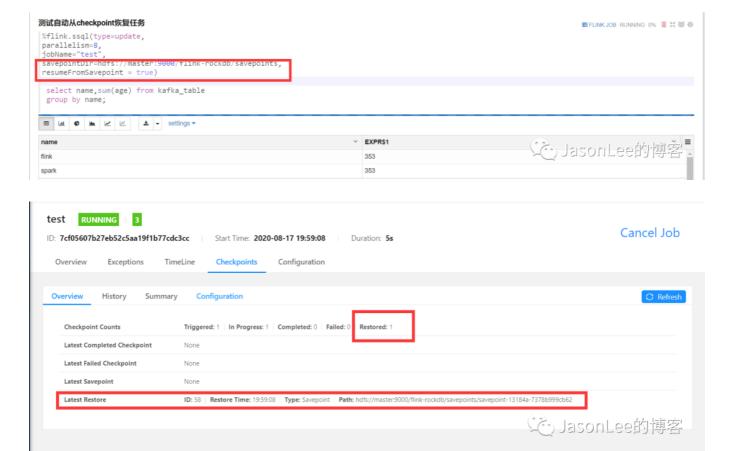
savepoint不像checkpoint那样定期的自动触发,它是需要我们手动触发的,在用户cancel job的时候才去做savepoint. 正常是需要用下面的命令触发的.然后启动的时候指定 savepoint的路径启动.

flink cancel —s jobid —yid appid

现在在zeppelin里面只需要在 Paragraph 里面配置 savepointDir 参数就可以了,取消任务的时候会自动触发savepoint,保存到指定的目录里面. 然后重启的时候需要设置 resumeFromSavepoint = true 这个参数,就可以自动从savepoint恢复任务了. 下面演示一下这个过程.



先把任务运行起来,然后等任务做了几次checkpoint后把任务停掉,然后添加 resumeFromSavepoint = true 参数直接启动任务就可以了.



任务也是可以正常启动的,不管是从checkpoint还是savepoint都变的非常简便.

总结:

自动从checkpoint&savepoint恢复任务其实是一个必要的需求,zeppelin已经帮我们做了这个功能,这样恢复任务就变的非常简单,配置几个参数就可以直接启动任务了.