Flink SavePoint的使用

checkPoint and savePoint

说到savePoint, 一定会想到checkPoint, 先来说下两者的区别:

checkpoint: 应用定时触发,用户保存状态,会过期,内部应用失败重启的时候使用,默认情况下cancel时,会删除之前的checkpoint。

savepoint: 用户手动执行,不会过期,相当于备份,可以在cancel的时候使用。一般可以用于修改并行度,程序升级等等。

所以如果想要在程序停掉后重启,数据一致的情况下,强烈推荐使用cancel时做savepoint,当然,如果程序中没有涉及到state,可以不用。

savePoint的使用

1.在flink-conf.yaml中配置savepoint存储位置,不是必须设置,如果设置了,那么做savepoint的时候,可以不指定位置,系统会默认使用你设置的位置。如果没有设置,那么在做savepoint的时候,必须指定位置,不然就会报错。

state.savepoints.dir: hdfs://namenode:9000/flink/savepoints 或者是 hdfs:/flink/savepoints

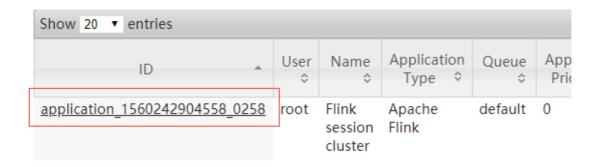
2.触发savePoint【直接触发或者是在cancel的时候触发】

flink savepoint jobId [targetDirectory] [-yid yarnAppId]【针对on yarn模式需要指定-yid参数】

flink cancel -s [targetDirectory] jobId [-yid yarnAppId]【针对on yarn模式需要指定-yid参数】

这里的jobId和yid我这里截图说明一下:

这个是yid



这个是jobld:



所以这里对应的命令就是:

- flink savepoint 2f0511c8fe430acb81a968aff0cbb5b9 -d hdfs:/user/root/flink -yid application_1560242904558_0258 或者flink savepoint 2f0511c8fe430acb81a968aff0cbb5b9 yid application_1560242904558_0258 这里-d是指定位置
- flink cancel -s hdfs:/user/root/flink 2f0511c8fe430acb81a968aff0cbb5b9 -yid application_1560242904558_0258 或者flink cancel -s 2f0511c8fe430acb81a968aff0cbb5b9 yid application_1560242904558_0258 这里-s后面可以跟路径也可以不跟路径,取决于 state.savepoints.dir有没有配置

从指定的savepoint启动job

flink run -s savepointPath [runArgs]

例子: 我上面已经做了一个savepoint, 保存的路径为hdfs:/user/root/flink 生成的文件是 savepoint-402ba9-e7e4915ab401

那我要从这个savepoint中启动程序,对应命令就是:

nohup flink run -m yarn-cluster \

- -yn 3 -yjm 1024 -ytm 1024 -p 3 \
- -s hdfs:/user/root/flink/savepoint-402ba9-e7e4915ab401 \
- -c com.sp.writetohbase.writetohbase \
 /usr/local/myjob/writetokafka/MyFlinkTTest-1.0-SNAPSHOT.jar &

多了一个-s 其他参数和正常启动时的参数一样

如果savepoint时报错,可以去jobmanager的日志上去查看具体原因。