Nifi采集数据流程

环境信息

CDH6开发环境nifi登录地址

<http://bigdata-cdh-004:8080/nifi/>

CDH6开发环境调度登录地址

[http://10.39.52.51:8081/zvalleycenter/login（admin/](http://10.39.52.96:12345/login（admin/123456)diaoduTEST.159）

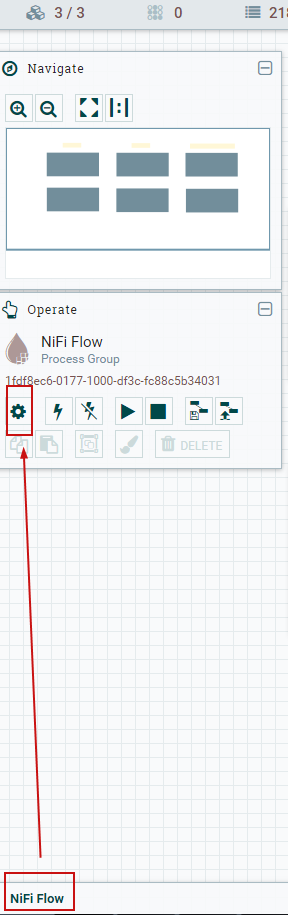
CDH6开发环境Minio登录地址

<http://10.39.52.61:8321/minio/login(admin/admin123456)>

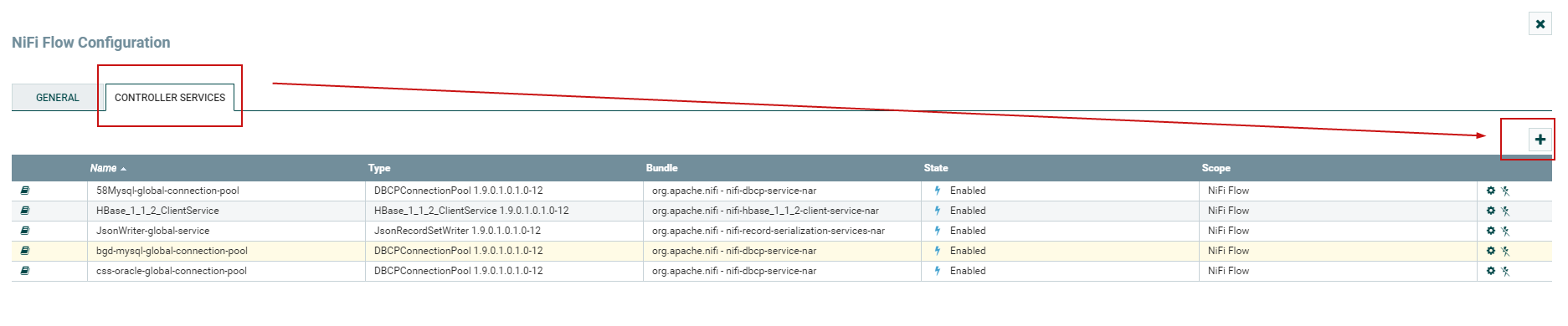
目前nifi只接入源系统为SAP（crm、ecc）和关系型数据库（mysql、oracle）的数据

nifi连接池配置（开发环境nifi可以自行配置全局连接池，生产环境连接池配置请联系大数据平台室）

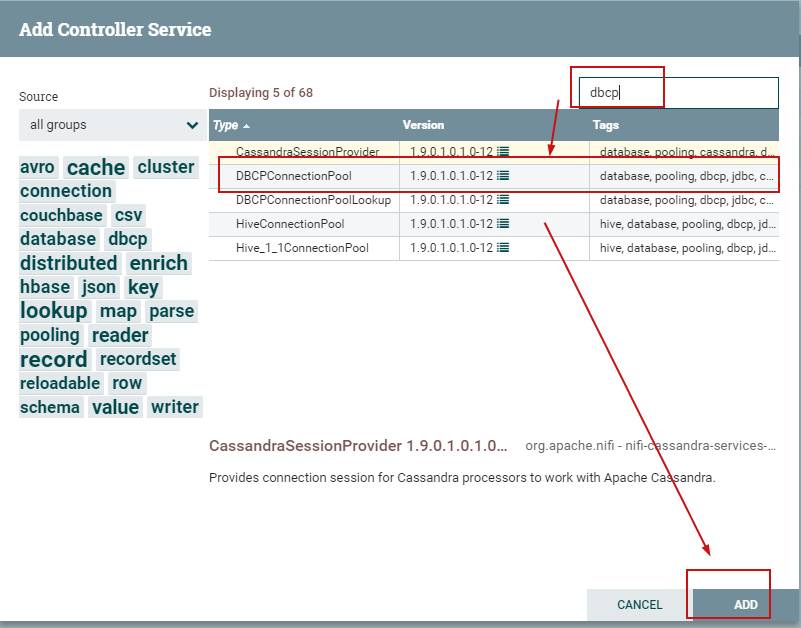
1、配置全局连接池，在nifi的根目录选择configuration



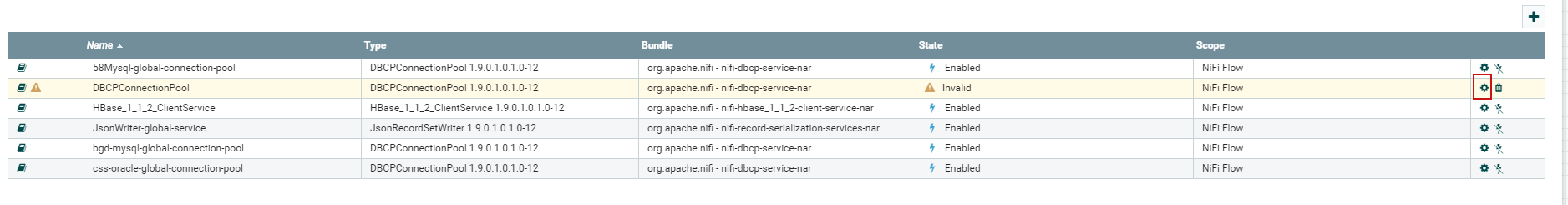
2、选择+



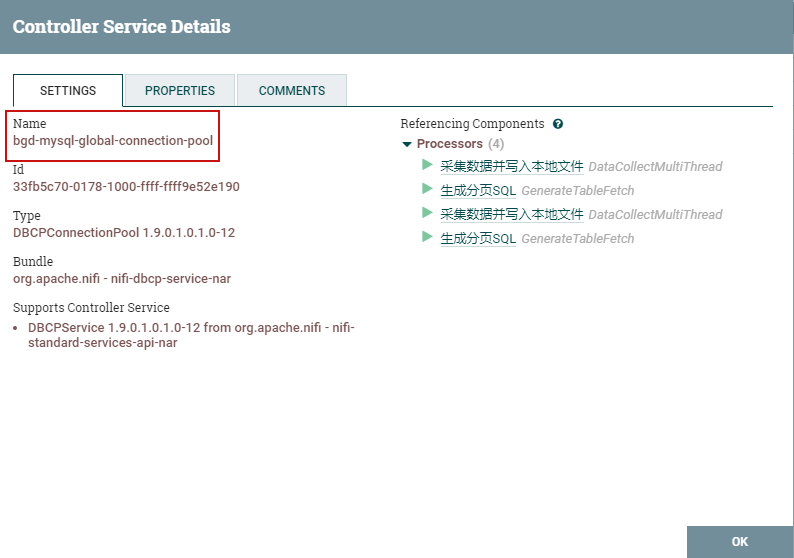
3、选择连接池组件



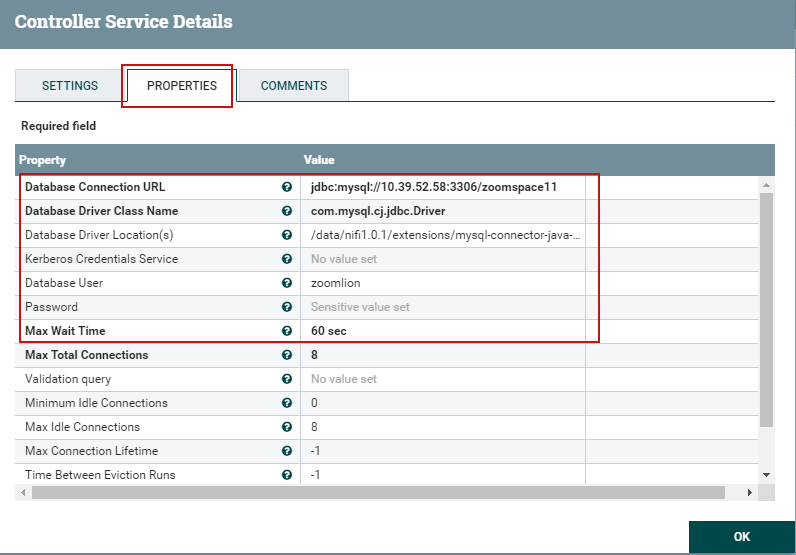
4、配置数据库信息



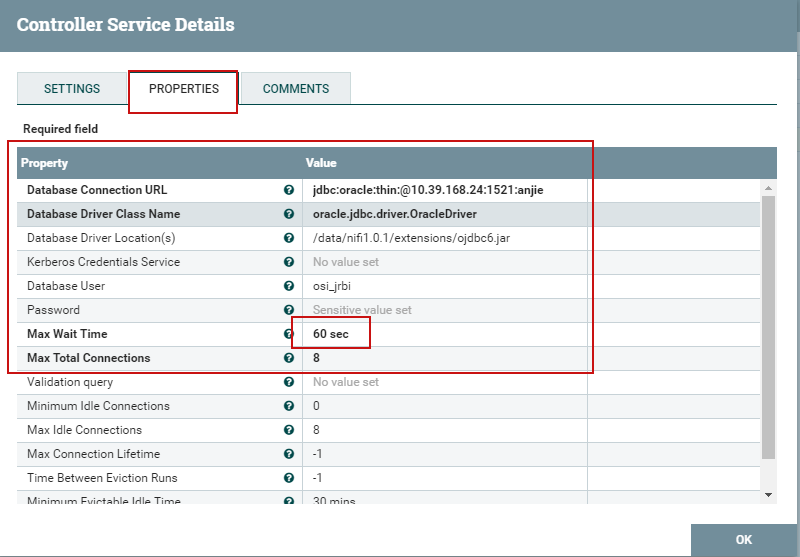
5、命名连接池名称



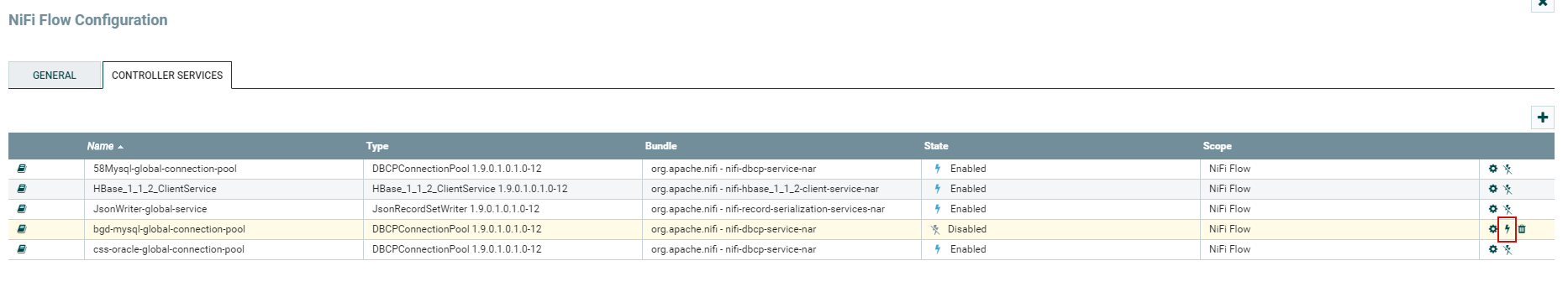
6、mysql数据库配置信息



7、Oracle数据库配置信息

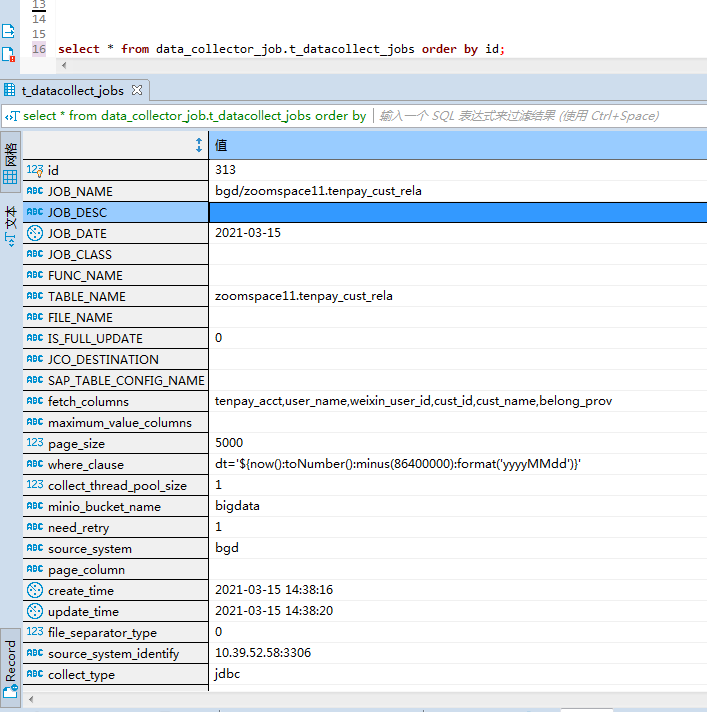


8、配置完成后启用连接池

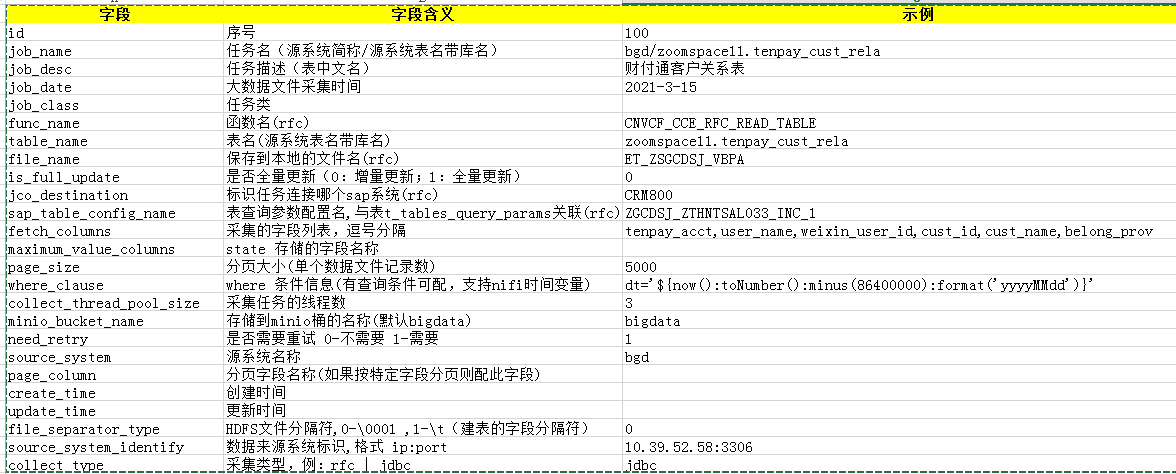


通过nifi抽取关系型数据库（jdbc）配置示例

1. 在mysql配置表data\_collector\_job.t\_datacollect\_jobs 中增加表的配置信息

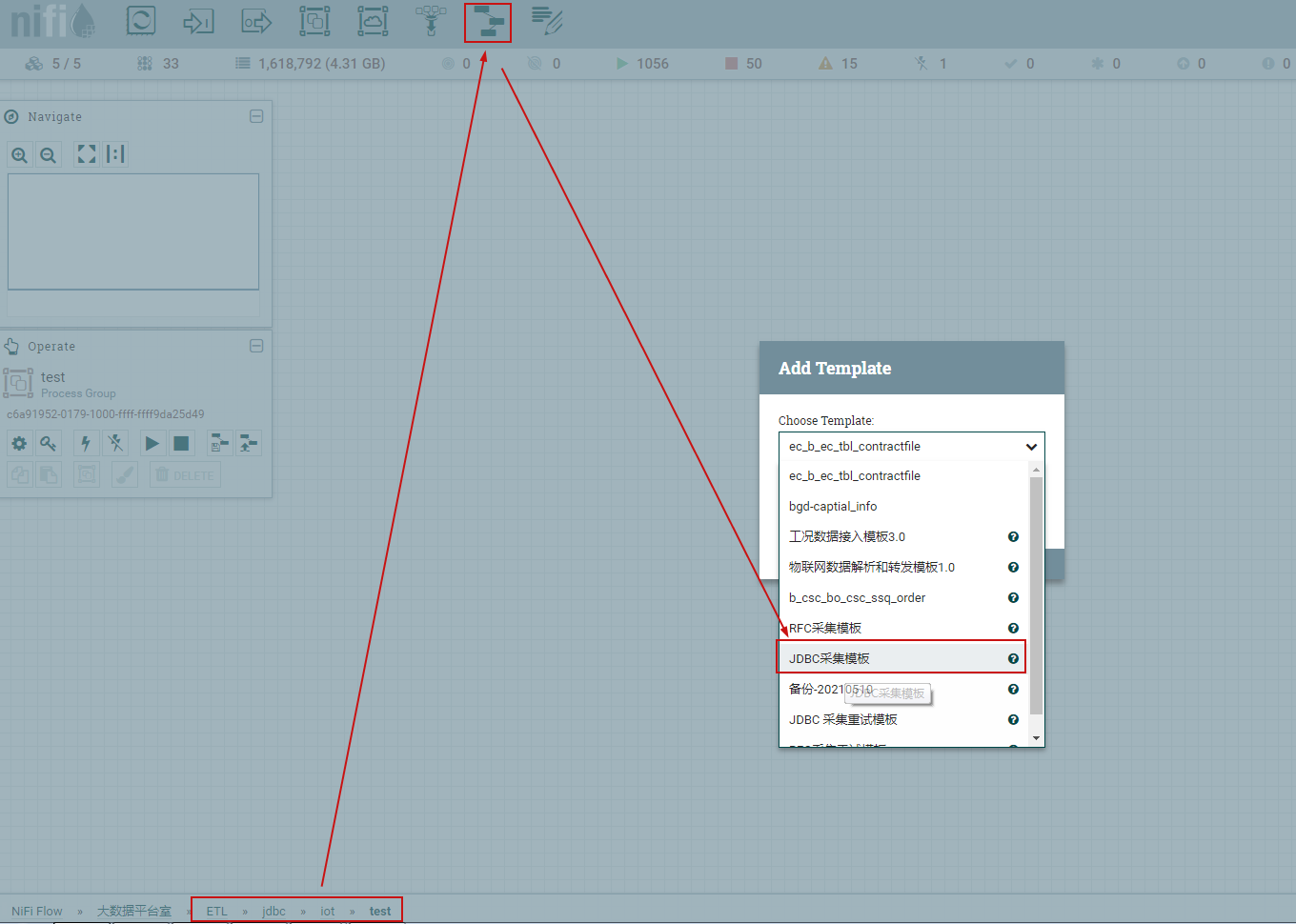


各字段说明

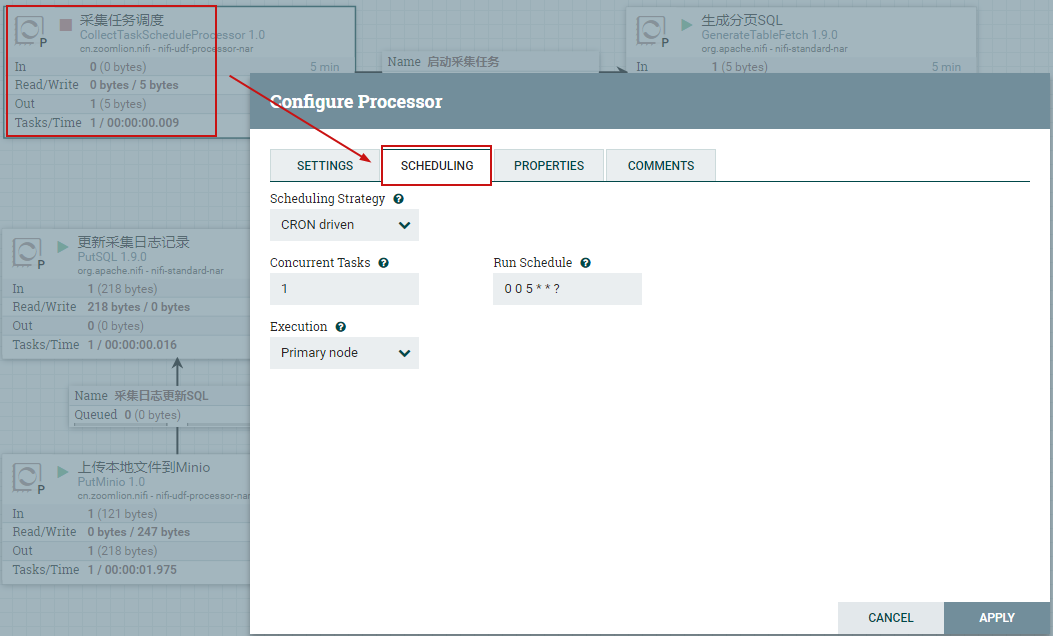


在mysql配置表中配置新增表的信息后，配置nifi抽数任务

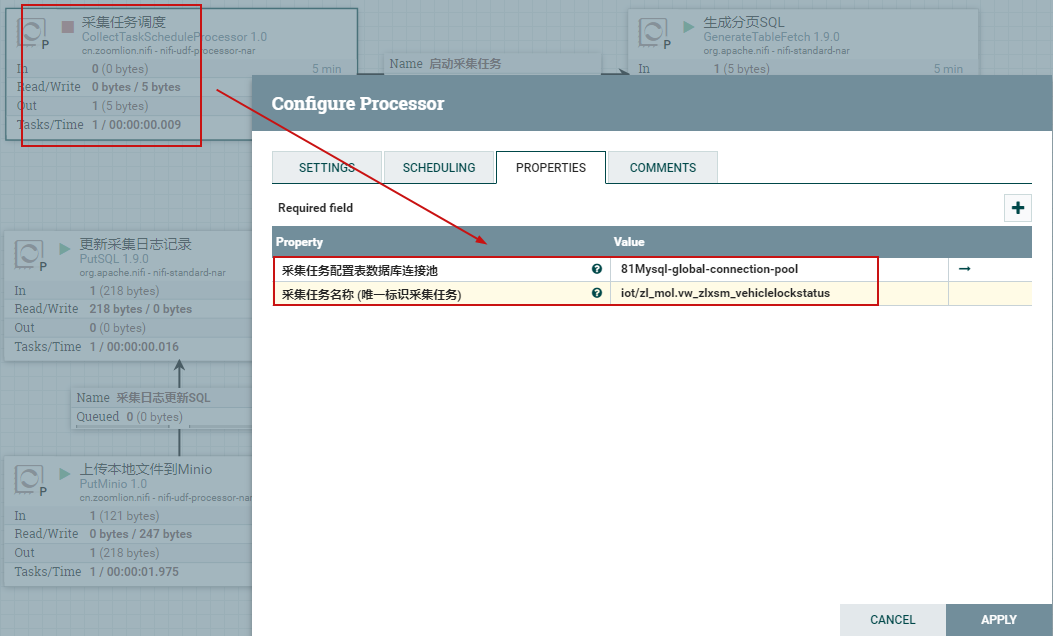
1、按目录创建表的nifi任务（以表名命名任务名），在任务里拖入jdbc数据采集模板



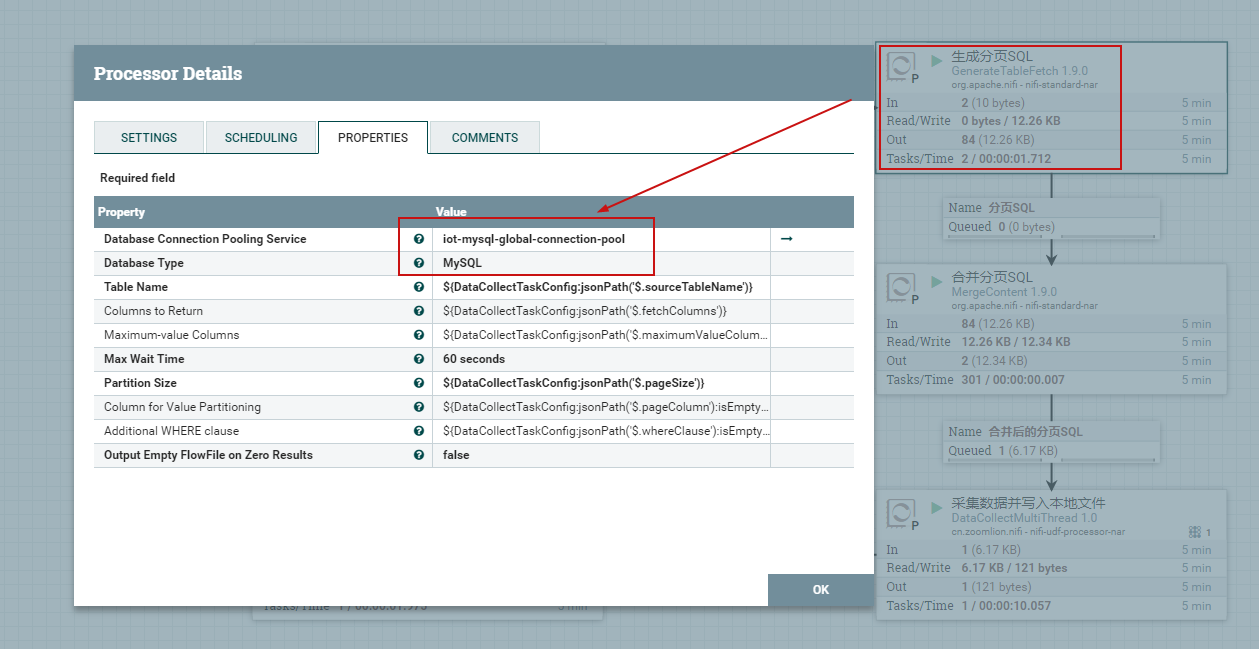
2、配置抽数任务调度信息（以实际采集频率确认）



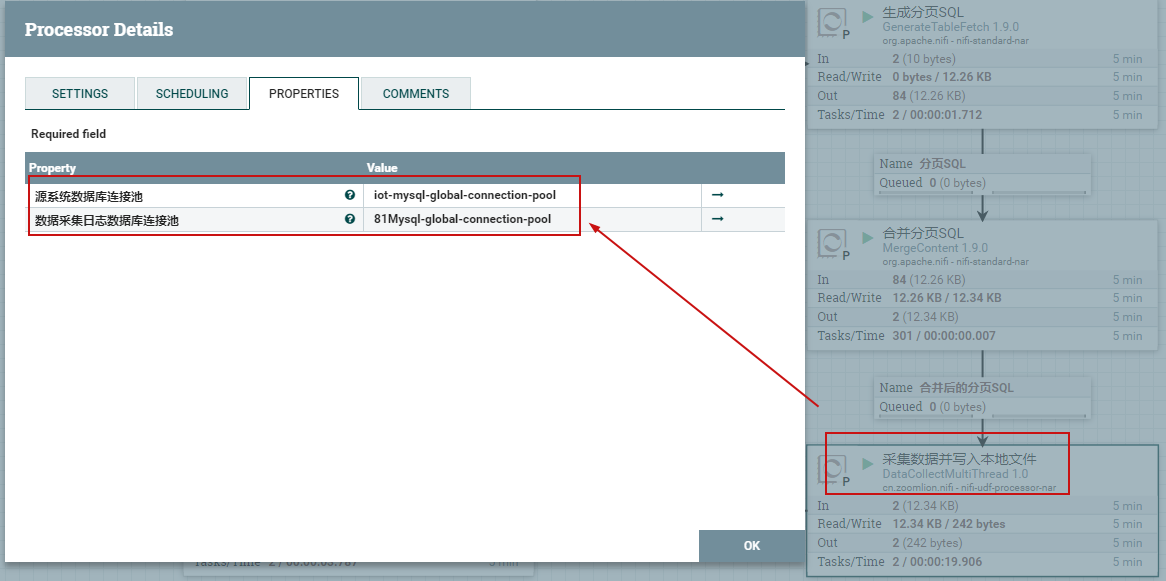
抽数任务表在mysql配置表中的连接池配置，采集任务名称为data\_collector\_job.t\_datacollect\_jobs表中的job\_name字段。



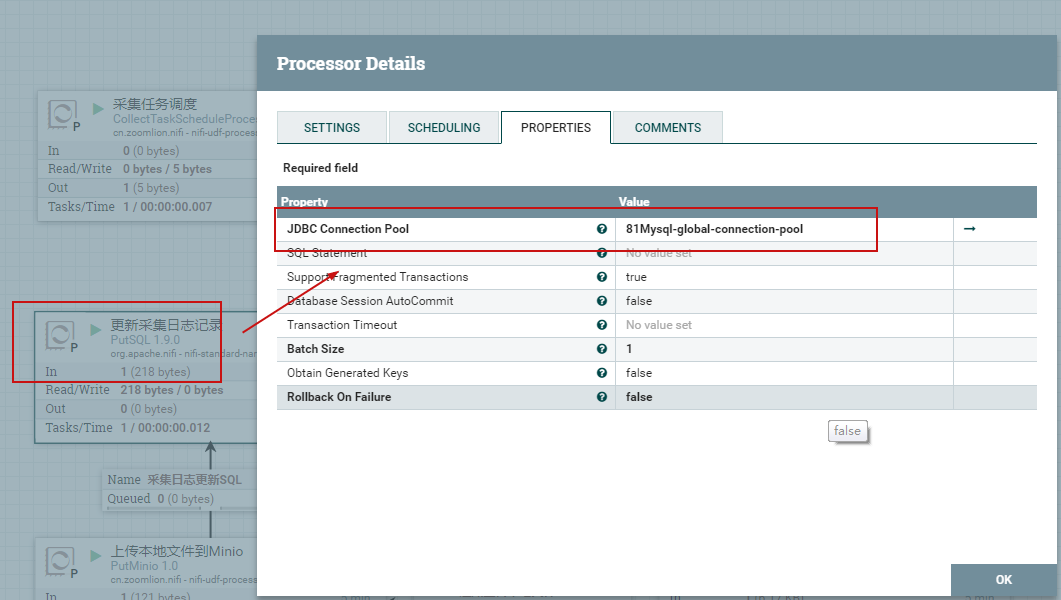
3、源系统数据库连接池配置和源系统数据库类型



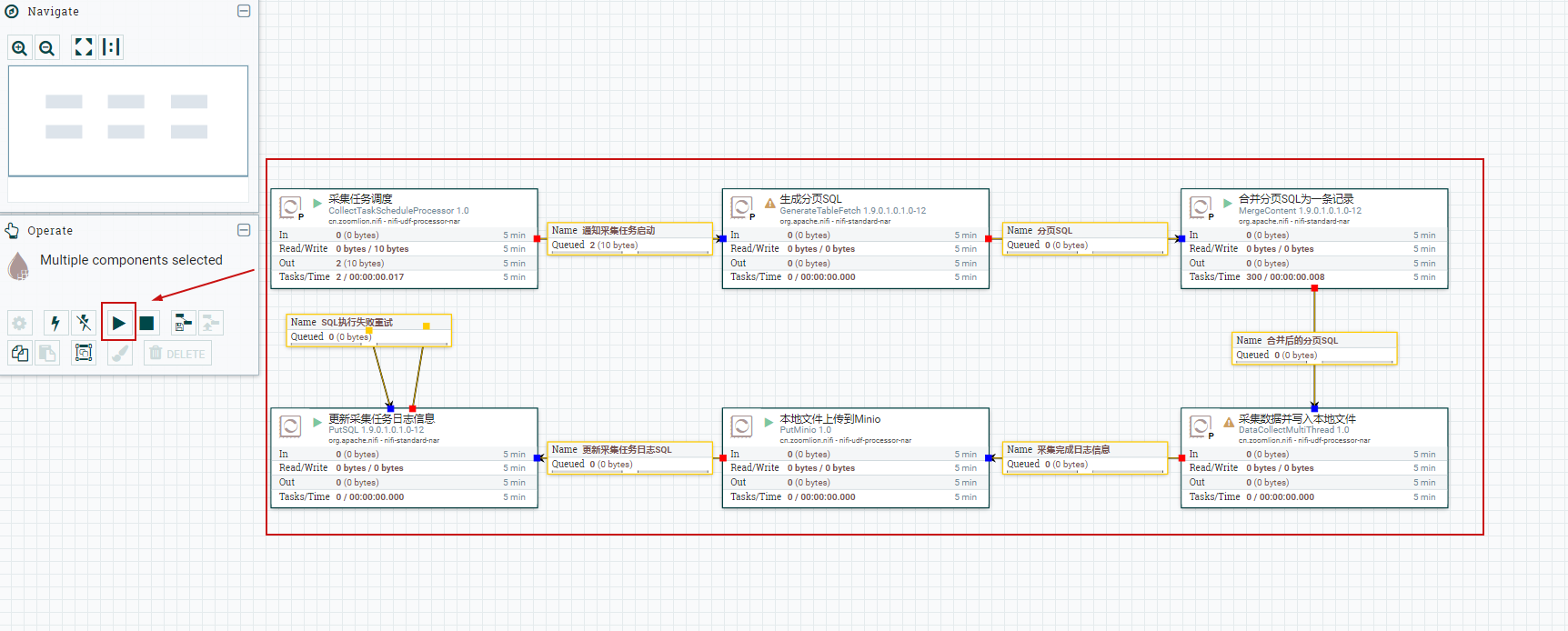
4、源系统数据库连接池和大数据data\_collector\_job.t\_datacollect\_log日志表的数据库连接池



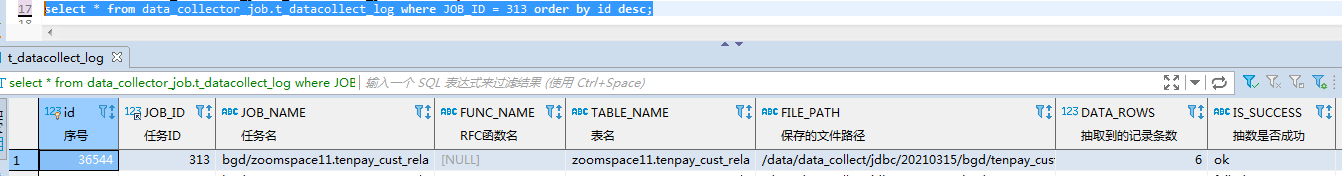
5、更新大数据data\_collector\_job.t\_datacollect\_log日志表的数据库连接池配置



抽数nifi任务配置完成后启用任务，成功执行后查看mysql日志表（以job\_id=313为例）



**select** \* **from** data\_collector\_job.t\_datacollect\_log **where** JOB\_ID = 313 **order** **by** id **desc**;



查看日志如果数据已成功上传至minio，则配置调度，把数据从minio上传至HDFS并创建分区。

sh /home/zoomspace/udp\_collect/nifi\_download.sh jdbc 20210314 bgd zoomspace11.tenpay\_cust\_rela N 20210314 20210315 bdl.b\_imp\_tenpay\_cust\_rela

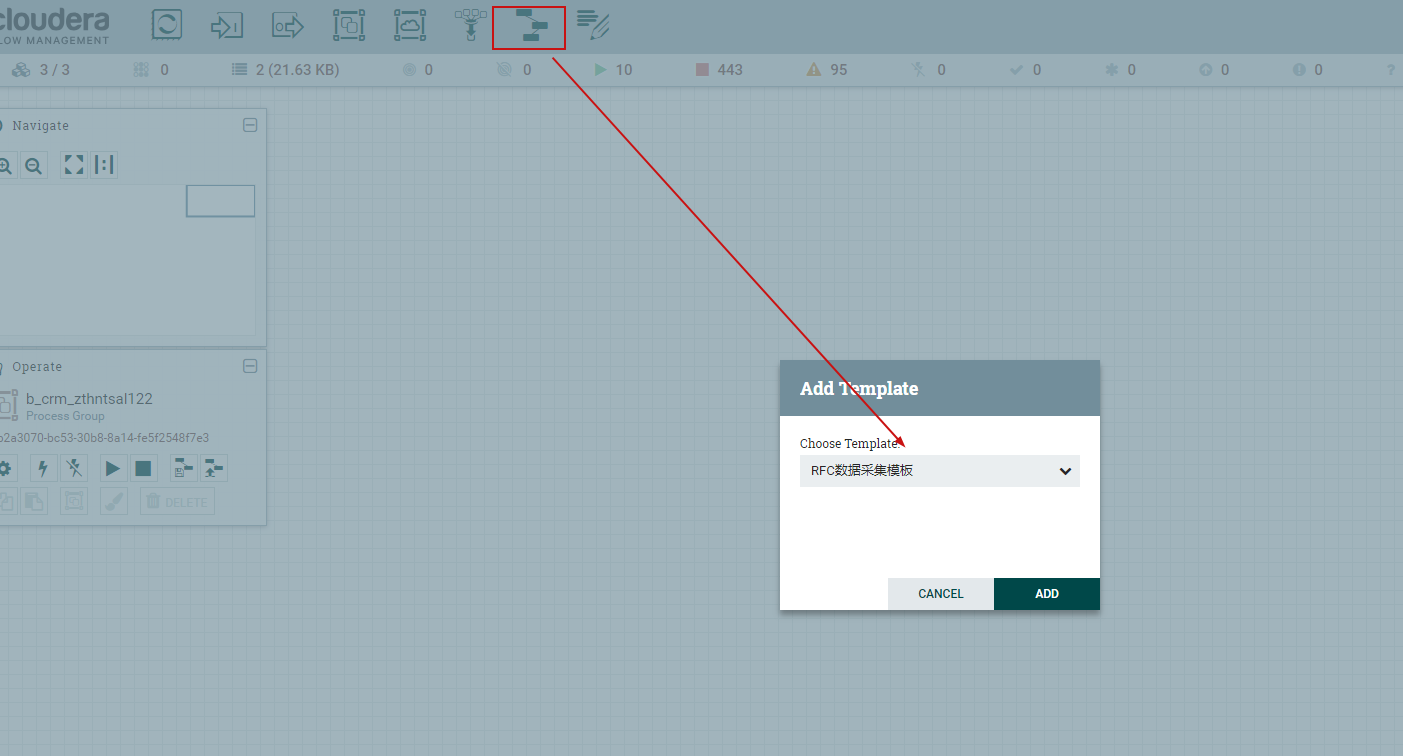
入参说明：

1. jdbc 代表是通过jdbc采集的任务，传入jdbc或sqoop均可，rfc则为任务ID
2. 20210314 代表分区日期
3. bgd 代表源系统
4. zoomspace11.tenpay\_cust\_rela 代表源系统表名，通过rfc抽取数据的表必须和data\_collector\_job.t\_datacollect\_jobs配置中job\_name字段“/”后的字符保持一致，通过jdbc抽取数据的表需和data\_collector\_job.t\_datacollect\_jobs配置中table\_name字段保持一致。
5. N 代表增全量，N增量，Y全量
6. 增量开始时间（全量此值不做要求）
7. 增量结束时间（全量此值不做要求）
8. 大数据表名（为了获取HDFS路径及创建分区时使用）

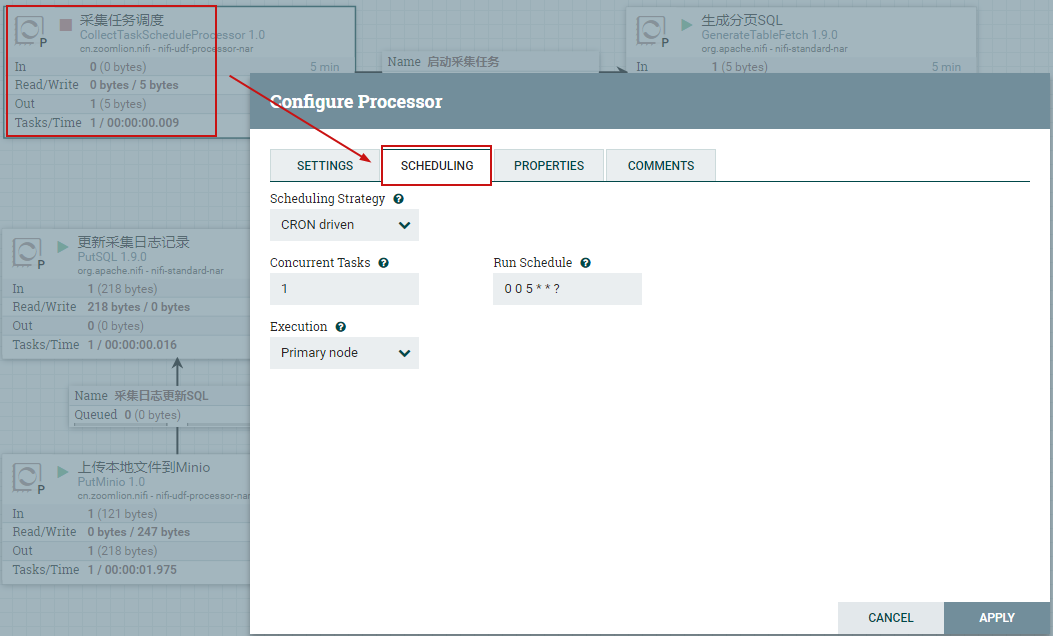
通过Nifi抽取sap(rfc)配置示例

配置Mysql的data\_collector\_job.t\_datacollect\_jobs表及把数据从minio上传至HDFS并创建分区的操作和上文相同

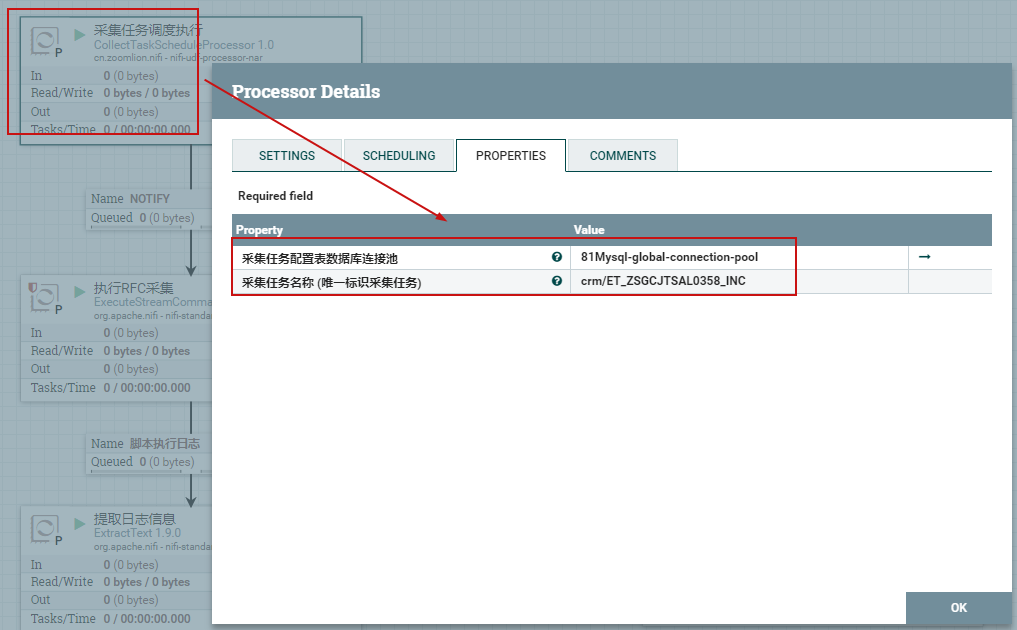
1、选择RFC数据采集模板



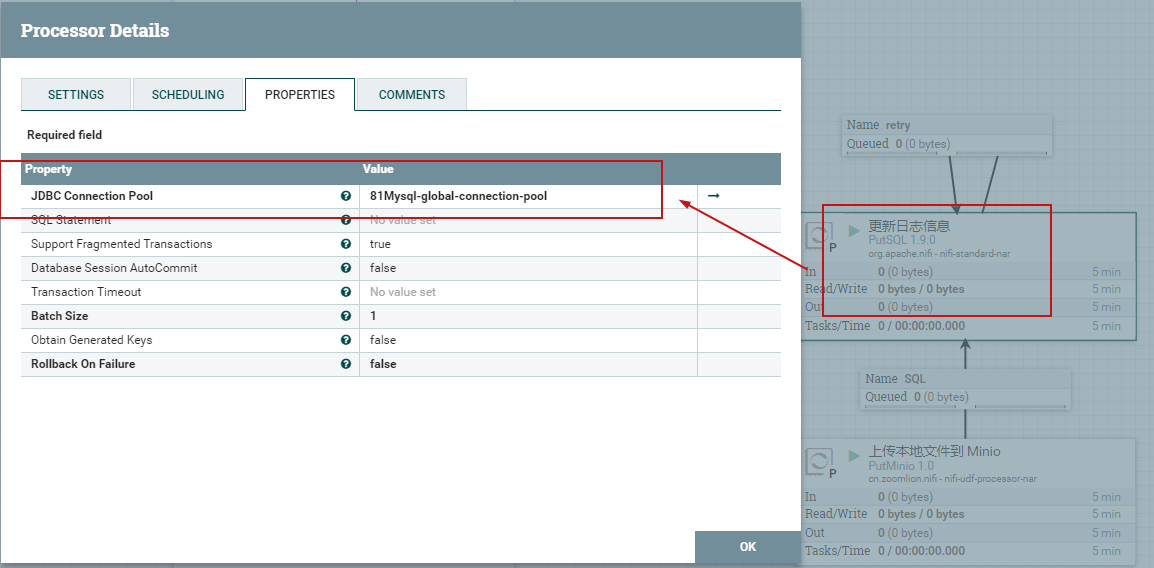
1. 配置抽数任务调度信息（以实际采集频率确认）



1. 填写连接池和采集任务标识（data\_collector\_job.t\_datacollect\_jobs表的job\_name字段）



4、选择更新日志连接池



5、查看日志信息和JDBC采集一致

6、从minio下载文件到HDFS及表添加分区

sh /home/zoomspace/udp\_collect/nifi\_download.sh 246 20210325 crm ET\_ZTHNTSAL122 Y 20210324 20210325 bdl.b\_crm\_zthntsal122