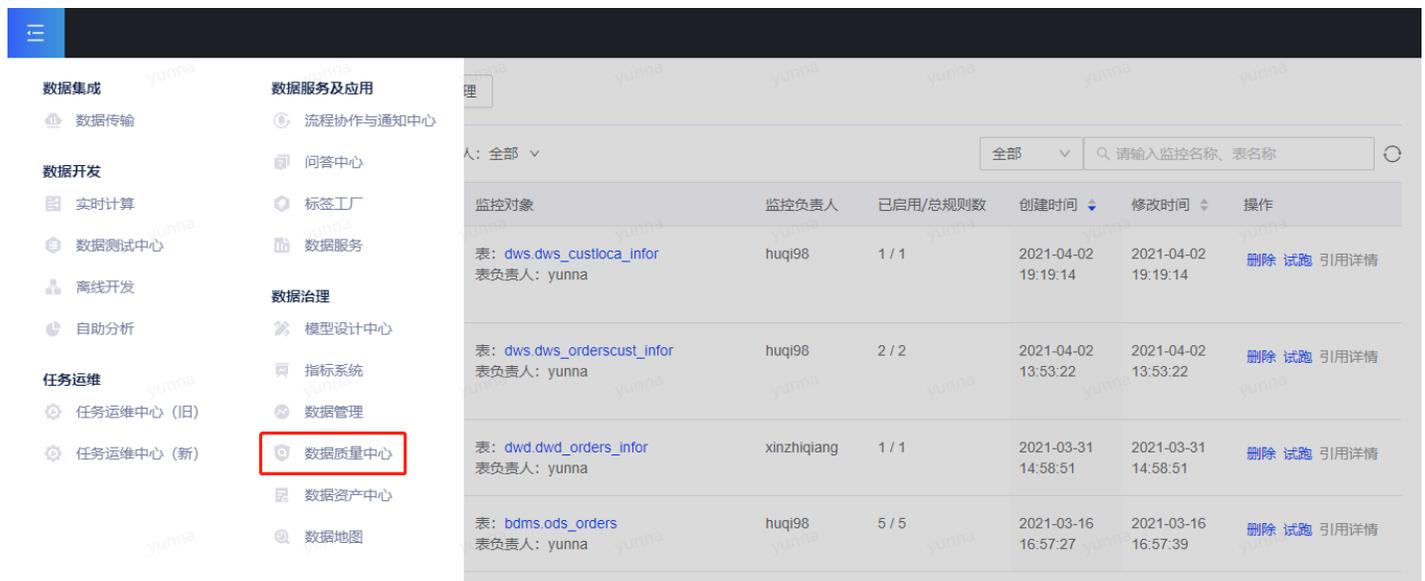


# 数据质量中心：创建规则监控，保证高质量数据流转

数据质量中心通过事前定义监控规则、事中监控数据的生成过程以及事后评估和问题追溯，依托离线开发中心的数据质量配置，提升企业数据价值。

在查看任务的运行情况之前，可以针对任务运行中的库表配置数据质量任务和监控，防止数据加工过程中，数据库和业务频繁变更或者数据订正等原因引入业务系统的脏数据，从而花费大量的时间寻找原因，耗费开发时间和精力，数据质量中心通过事前定义监控规则、事中监控数据的生成过程以及事后评估和问题追溯，依托离线开发中心的数据质量配置，提升企业的数据价值。

打开导航栏，点击**数据质量中心**，进入**数据质量中心**产品界面。



The screenshot displays the Data Quality Center interface. On the left is a navigation menu with categories: 数据集成 (Data Integration), 数据开发 (Data Development), 任务运维 (Task Operation), 数据服务及应用 (Data Services and Applications), 数据治理 (Data Governance), 数据资产中心 (Data Asset Center), and 数据地图 (Data Map). The '数据质量中心' (Data Quality Center) option is highlighted with a red box. The main content area shows a table of monitoring rules with columns: 监控对象 (Monitoring Object), 监控负责人 (Monitoring Responsible Person), 已启用/总规则数 (Enabled/Total Rules), 创建时间 (Creation Time), 修改时间 (Modification Time), and 操作 (Operations). The table contains four rows of data.

监控对象	监控负责人	已启用/总规则数	创建时间	修改时间	操作
表: dws_dws_custloca_infor 表负责人: yunna	huqi98	1 / 1	2021-04-02 19:19:14	2021-04-02 19:19:14	删除 试跑 引用详情
表: dws_dws_orderscust_infor 表负责人: yunna	huqi98	2 / 2	2021-04-02 13:53:22	2021-04-02 13:53:22	删除 试跑 引用详情
表: dwd_dwd_orders_infor 表负责人: yunna	xinzhiliang	1 / 1	2021-03-31 14:58:51	2021-03-31 14:58:51	删除 试跑 引用详情
表: bdms.ods_orders 表负责人: yunna	huqi98	5 / 5	2021-03-16 16:57:27	2021-03-16 16:57:39	删除 试跑 引用详情

进入**监控列表**，点击**规则模板管理**，系统提供通用的规则模板，比如空值、非空等，点击新建规则模板，也可以书写SQL语句，自定义规则模板。

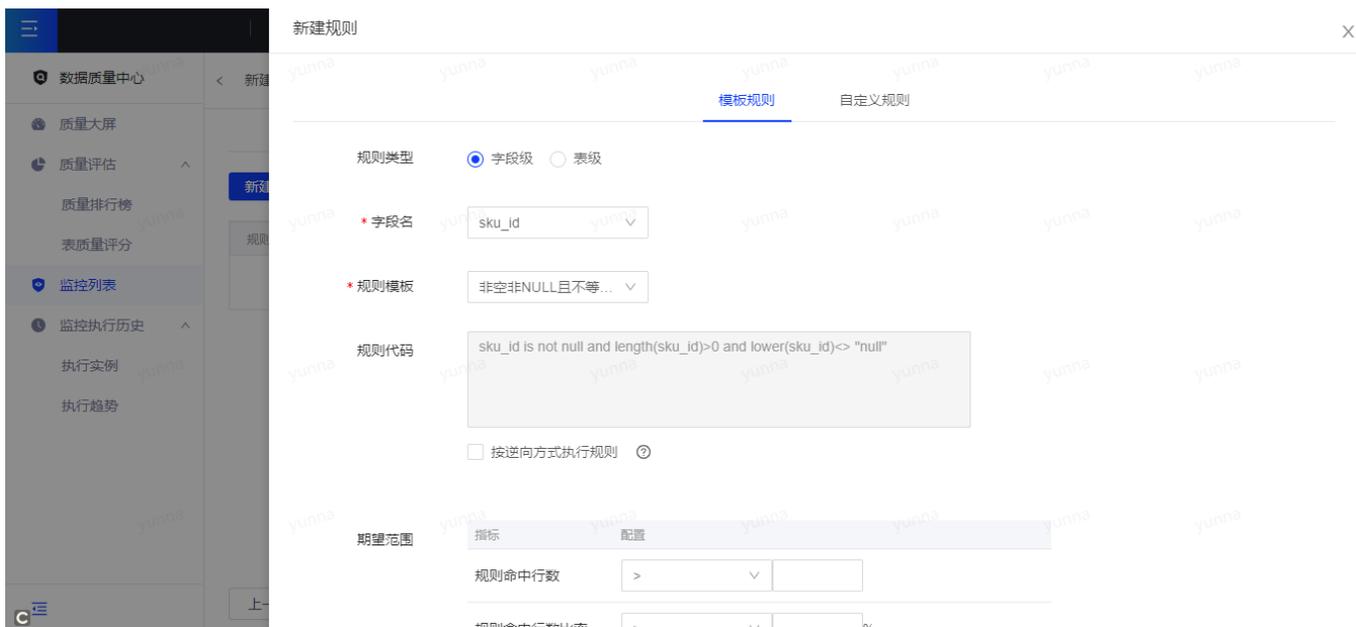


回到监控列表页面，点击**新建监控**，进入对于库表创建监控任务的流程页面：

- 首先需要选择监控对象(即平台中的Hive库表)，配置监控任务的名称内容。本示例中选择dwd库下的dwd\_product\_sales\_cookbook4\_month表配置监控任务。



- 配置**监控规则**，点击**新建规则**，支持通过规则模板或者自定义规则的方式设置监控规则，比如在模板规则中，针对字段级别，下拉选择字段sku\_id，规则模板选择非空非NULL且不等于NULL字符串，规则代码框中会填充字段名和规则模板匹配的SQL代码，用于核对监控任务是否符合要求。



3. 选择好配置的监控字段和模板后，可以设置期望范围，即符合监控规则的命中行数或者命中行数比率，2个指标中至少需要配置一项，然后针对规则的强弱进行配置，若为强规则，则当监控规则不通过时，任务会置为失败，停止运行，若为弱规则，当监控规则不通过时，任务会继续运行。点击新建，即成功创建一条监控规则。



4. 在新建规则弹窗中，若选择自定义规则，支持书写SQL语句定义监控规则，本示例中以dwd\_product\_sales\_cookbook4\_month表产品数量波动指定监控规则设置，书写完SQL语句后，可点击“解析”，产品会自动解析出SQL语句中的字段名称，也支持手动添加监控字段。同样，对于监控字段可设置期望范围或者期望值范围。



5. 点击新建，完成自定义规则的创建，监控规则创建完成之后，在规则列表中可查看针对库表创建的监控规则，可对监控规则进行重新编辑、删除或禁用操作。



6. 在监控规则与线上任务绑定之前，可以先进行监控任务的试跑，点击**完成并试跑**，监控规则使用spark引擎执行SQL语句的查询，若需要额外设置Spark内存等参数时，可通过set参数配置，同时可选择执行队列以及试跑成功或失败的通知方式。



7. 点击试跑后, 可进入执行实例查看运行情况, 支持查看运行结果和日志详情, 也可根据多维度筛选项查看其它监控实例运行情况。



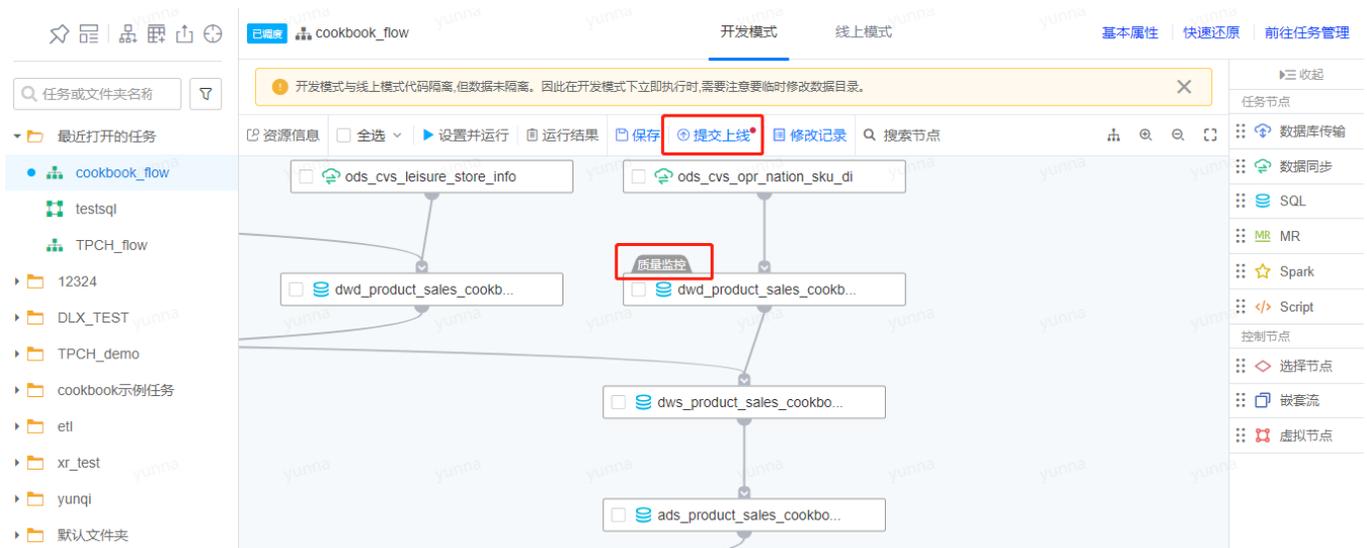
8. 监控任务运行成功之后, 需要在离线开发任务中绑定质量监控任务, 依托离线开发平台, 完成数据质量监测与任务流的关联。回到离线开发平台的任务流cookbook\_flow, 在开发模式下选择dwd\_product\_sales\_cookbook4\_month节点任务, 鼠标右键, 选择配置质量监控。



9. 在配置质量监控的弹窗中, 选择已经创建成功的监控任务dqc\_dwd\_product\_sales\_cookbook4\_month, 若监控任务执行失败, 可选择是否终止当前节点任务或者继续下游调度。



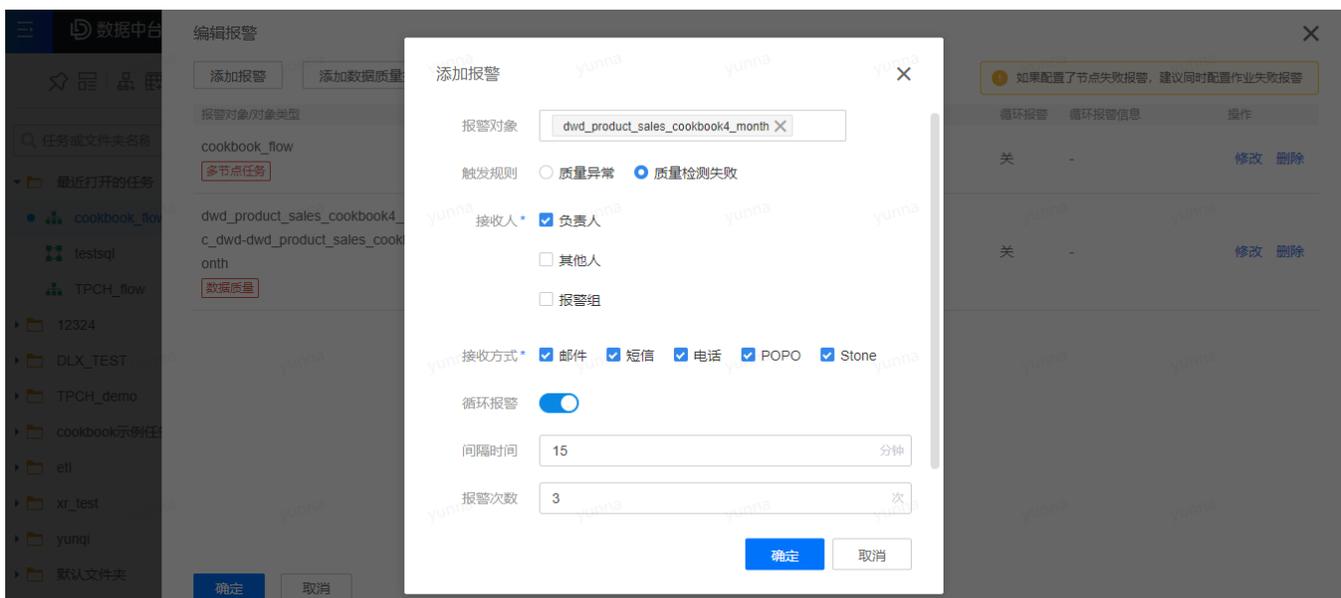
10. 节点任务配置了数据质量监控后, 节点左上方会出现明显的标志**质量监控**, 且当前开发模式和线上模式不一致, 可重新将任务提交上线。



11. 提交上线之后, 在线上模式该节点也拥有的质量监控任务, 针对任务流或任务流中的节点以及数据质量监控任务, 均可点击**编辑报警**, 对任务失败异常等内容进行报警配置。



12. 点击添加数据质量报警, 选择数据质量监控规则, 配置触发规则和接收人等内容, 也可开启是否循环报警。报警配置成功之后, 编辑报警页面会呈现针对当前任务流配置的所有报警内容, 可修改或删除操作。



以上完成了数据质量监控规则的配置, 以及在离线开发界面上与任务流的关联, 也配置了监控规则的报警, 若想要查看监控规则的执行情况, 可在数据质量中心产品中进入**监控列表**, 查看监控规则的执行情况。