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。

☲								
数	据集成	数据服务及应用	·····································					
4	数据传输	流程协作与通知中心						
数	据开发	7 问答中心	人: 全部 ∨		全	部 V Q 语	歸輸入监控名称、	表名称
E	实时计算	◎ 标签工厂	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
¢	数据测试中心	勤 数据服务	表: dws.dws_custloca_infor	huqi98	1/1	2021-04-02	2021-04-02	删除 试跑 引用详情
- 4	离线开发	数据治理	表负责人: yunna			19:19:14	19:19:14	
C	自助分析	🔅 模型设计中心	表: dws.dws orderscust infor	hugi98	2/2	2021-04-02	2021-04-02	
任	务运维	网 指标系统	表负责人: yunna	vunna		13:53:22	13:53:22	MULTURE 201711年1月 VULUUS
Ø	任务运维中心 (旧)	⊘ 数据管理						
Ø	任务运维中心(新)	数据质量中心	表: dwd.dwd_orders_infor 表负责人: yunna	xinzhiqiang	1/1	2021-03-31 14:58:51	2021-03-31 14:58:51	删除 试跑 引用详情
		2 数据资产中心		bugi09	5/5	2021 02 16	2021 02 16	
		② 数据地图	表负责人: yunna	yunna	yunna	16:57:27	16:57:39	删除 试跑 引用详情

产品功能详述 - 数据质量中心:创建规则监控,保证高质量数据流转 - 《有数数据开发及管理平台用户手册 V6.5.0》

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。

0	数据质量中心	< 规则模板管理							
8	质量大屏	新建规则模板							
¢	质量评估 ^								
	质量排行榜	模板来源: 🔽 系统 🔽 项目						Q、请输入模板名称	O
	表质量评分	规则模板名称 💲	引用的监控	描述		负责人	最后修改人	最后修改时间 🝦	操作
0	监控列表	自定义模板项目	0个	自定义模板			1000	2021-03-16 16:49:11	编辑删除
C	监控执行历史 ^								
	执行实例 yunna	NotNull 系统	0个 ///	这是系统预置的规	则模板	- >	unna -	2020-01-01 00:00:00	unna
	执行趋势	有效IP系统	3个	有效IP		-	-	2020-01-01 00:00:00	-
		有效纬度系统	0个	有效纬度		-	-	2020-01-01 00:00:00	
		有效经度系统	0个 Y ^{UI}	有效经度		-)	unna _	2020-01-01 00:00:00	unna
		空值系统	0个	空值		-	-	2020-01-01 00:00:00	

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:

1. 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_prod uct_sales_cookbook4_month表配置监控任务。

0	数据质量中心	< 新建监控					
6	质量大屏		1 设置监持	控对象		2 设置规则	
¢	质量评估 ^						
	质量排行榜	◆ 血控/J家 *库 yunna	dwd yunna		✓ yunna		
۲	监控列表	*表	dwd_product_sales_cookboo	ok4_month	\vee		
C	监控执行历史 ^	分区筛选	是否监控 分区字段 分区值	Internet			
	执行实例		pt_d				
	执行趋势	∨ 基础信息					
		* 监控名称	dqc_dwd-dwd_product_sales	s_cookbook4_month			
		描述,如008	请输入监控描述				

 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。

E	新建规则				Х
数据质量中心 < 新述	yunna yi		yunna 模板规则 自员	yunna E义规则	
 ● 质量 ● 质量 评估 ^ 	规则类型	 ·			
新建 质量排行榜 表质量评分 规则	yunn ^a *字段名 yu	sku_id			
◎ 监控列表	* 规则模板	非空非NULL且不等 >			
 監控执行历史 ^ 执行实例 执行趋势 	规则代码	sku_id is not null and length(sku_id):	>0 and lower(sku_id)<> " yunna	null" Y ^{unna}	
		🗌 按逆向方式执行规则 🛛 💿			
yunna	yunna 期望范围	1000 指标 112			
C.		规则命中行数 >	×	%	

 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。

⊒			新建规则					×
O	数据质量中心	< 新建		如 💮 按逆向方式执行规则	N			
6	质量大屏							
e	质量评估 ^		期望范围	指标	配置			
	质量排行榜	新建		规则命中行数	>	✓ 10		
	表质量评分	规则		规则命中行数比率	>	yunus √ 5.00	%	
٢	监控列表			2.人长仁中五小司男465				
G	监控执行历史 ^			2.13目标中主少国国口英				
	执行实例		yunna 规则控制	○ 强 规则不通过时, ○ 弱 规则不通过时, 	壬务置为失败状态, 壬务继续运行	停止运行 yunna		
	执行趋势							
			描述	cookbook示例中dwd	层表4			
				yunna yu	nna	yunna	yun 17/120	
Ξ		Ŀ-	新建即	以消				

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_co okbook4_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。

⇒	新建规则		X
◎ 数据质量中心 <	新建	yunn ⁸ yunn ⁸ yunn ⁸ yunn ⁸ yunn ⁸	
❸ 质量大屏			
● 质量评估 ^	规则类型	○ 字段级 ● 自定义SQL	
质量排行榜表质量评分	yunna *规则名称	puna 产品数量波动监控,unna punna punna punna	
 监控列表 ⑤ 监控执行历史 ^ 	第 * SQL	SOL格式限制: 1.需返回一行一列或一行多列的数据,如返回多行则只会校验算一行数据 2 可遭过解析来快速生成学程,也可手动添加 3.创建律修监控定数的是HV电要时,目定义SOL语句不能包含Kudu表 解析:每次解析后将覆盖当前已设规则,需重新配置校验条件	
执行实例	yunna	select count(*) as num,sum(ord_prod_amt) as all_prod_amt from dwd_product_sales_cookbook4_month	
		☑ 解析	
yunna	yunna 字段	列号 别名 校验类型 校验方式 期望范围,哪望率范围	yunna 操作
	±-	num 数值型 与固定值比较 > % <th< td=""><td>删除</td></th<>	删除

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。

2 数据质里中心	〈 新建监控							
)质量大屏		✓ 设置监控数	村象				2 设置规则	
质量评估	A							
质量排行榜	新建规则			模板规则(2)	目定义规则(1)			
表质量评分	规则控制/规则名称	规则类型	监控字段	校验类型	yunne	期望范围	状态	操作
监控列表			****	数值型-与固定值!	比较	>10		
监控执行历史		T 1 1 1 0 0 1	num 宣有代码	波动率型-环比上-	一周期	>5%		编辑 删除 禁用
执行实例		自定义SQL	suppa.	数值型-与固定值]	比较	>10		
执行趋势			all_prod_amt	波动率型-环比上-	—周期	>5%		
	, 	and a strate to prove						

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。

⊒											
0	数据质量中心	+ 新建监控	试跑-dqc dwo	-dwd product	sales cookbook4	1 monthyunna	yunna X	yun	ID3	yunna	
6	质量大屏	监控负责人:全部	set参数	属性 ⑦		值		✓ Q 请			0
e	质量评估 ^	监控名称 💲				无数据		间 🔷	修改时间 💲	操作	
	表质量评分	dqc_dwd-dwd ales cookbool		* 添加				4-14 39	2021-04-14 10:31:39	删除 试跑 引用详	情
0	监控列表	ID: 31	执行队列	bdms			\checkmark				
C	监控执行历史 ^ 执行实例	dqc_dws-dws_ nfor ID: 30	通知条件 yunn	✔ 试跑成功	✓ 试跑失败 yunna			4-02 14	2021-04-02 19:19:14	删除 试跑 引用详	情
	执行趋势	dqc_dws-dws_ infor ID: 28	通知方式	🖌 邮件 🔽	短信 🗌 电话	POPO Stone	e 易信	14-02 22	2021-04-02 13:53:22	删除 试题 引用详	情
		or ID: 26	yum 试题中,) 关闭弹窗不影响	yunna 试跑,可至 执行实	yunna 例 查看详情) 试路	yunna 关闭	3-31 you 51	2021-03-31 14:58:51	删除 试跑 引用详	情

 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。

0	数据质量中心	监控开始时间: 今天 自 监控负责	唐人:全部 v 表负	∩∂ 责人:全部 ∨	实例类型:全部		a 全部 \	yunna く Q 请報	yunna 俞入监控名称、表名称、实	例ID	0
8	质量大屏	监控结果:全部 重置									
¢	质量评估 ^	监控名称/监控ID/实例类型 💲	监控对象	所属任务信息	监控负责人	监控结	课 异常/计	训执行规则	开始时间-结束时间 🝦	操作	
٥	质量排行榜 yunna 表质量评分 监控列表	dqc_dwd-dwd_product_sales_c ookbook4_month ID:31	表: dwd.dwd_product _sales_cookbook 4_month 表负责人: -	实例ID: 928	Yun	 运行 	中 -/3		2021-04-14 10:34:18 2021-04-14 10:34:41	运行结算 日志	The second secon
C	监控执行历史 ^ 执行实例 unit ³ 执行趋势										

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_cookbook_month节点任务,鼠标右键,选择**配置质量监控**。

☆ ☴ ♣ 璽 凸 ⊕	CRER 盐 cookbook_flow		运行编辑		yunna 基本属性	快速还	a 原 │ 前往任务管理
	开发模式与线上模式代码隔离,但数据未需	蜀裔。因此在开发模式下立即执行时	₹ 查看版本			×	▶ 三 收起
Q、任劳或又件夹名称 V				-			任务节点
▼ 🛅 最近打开的任务	2 资源信息 □ 全选 ∨ ▶ 设置并运行	自运行结果 □保存 ⑦ 提		索节点	.∄ ⊕	Q ()	: î 数据库传输
• 🚠 cookbook_flow	yunna yunna	yunna	 里節名 复制节点 				: ᅌ 数据同步
testsql	ods_cvs_leisure_store_info	<pre>ods_cvs_opr_nation_</pre>	- 禁用节点				👯 😂 SQL
TPCH_flow			删除节点 delete	<u>}</u>			II MR MR
▶ 📩 12324		dwd_product_sales_cookb	ook4_month 				🗄 😭 Spark
DLX_TEST vunna	lwd_product_sales_cookb	dwd_product_sales_	Como	vunna			Script
		, T					控制节点
IPCH_demo							👬 🔷 选择节点
▶ 📩 cookbook示例任务							: 同 最有法
• 🗖 eti		dws_product_sales_coo	kbo				
							: 🖁 虚拟节点
xr_test yunna		yunna					

产品功能详述 - 数据质量中心:创建规则监控,保证高质量数据流转 - 《有数数据开发及管理平台用户手册 V6.5.0》

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。

Ð								-
	計画中	已編度 Acookbook	flow		开发模式	线上模式	and and a second se	熏性 快速还原 前往任务管理
Q. 任务或文件夹	名称 又	开发模式与线上	模式代码隔离,但数据未隔	离。因此在开发模式下立即执行	时,需要注意要临时修改数据	目录。		送 送 收起 任务节点
配置质量监控-dw	d_product_sales	_cookbook4_month						? 临时保存 关闭
监控名称*	dqc_dwd-dwd	_product_sales_cookbo	ok4_month •	前往数据质量中心				
执行失败的处理	● 终止本任务	○ 继续下游调度						
set 属性	属性?			值 yunna				
	十添加							

- 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
 - 分品は鹿山の 已调度 击 cookbook_flow 开发模式 线上模式 基本属性 快速还原 前往任务管理 ▶三 收起 丹发模式与线上模式代码隔离,但数据未隔离。因此在开发模式下立即执行时,需要注意要临时修改数据目录。 × Q. 任务或文件夹名称 7 任务节点 山 🔍 🔾 门 🔡 🎲 数据库传输 ⑦ 提交上线● ■ 修改记录 Q 搜索节点 ▼ 🗁 最近打开的任务 C⁰ 资源信息 □ 全选 × ▶ 设置并运行 自运行结果 □ 保存 : 🏫 数据同步 • A cookbook_flow ods_cvs_leisure_store_info ods_cvs_opr_nation_sku_di 🗄 😑 SQL testsql MR MR A TPCH_flow 质量监控 🗄 🏠 Spark 12324 dwd_product_sales_cookb. dwd_product_sales_cookb. DLX_TEST yunna 控制节点 TPCH_demo 🗄 🔷 选择节点 ▶ 📩 cookbook示例任务 :: 🗇 嵌套流 dws_product_sales_cookbo. etl 🔡 ដ 虚拟节点 test 🕨 🛅 yunqi ads_product_sales_cookbo. ▶ 🛅 默认文件夹
 - 致,可重新将任务提交上线。

产品功能详述 - 数据质量中心:创建规则监控,保证高质量数据流转 - 《有数数据开发及管理平台用户手册 V6.5.0》

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。

	已编度 蒜 cookbook_flow		开发模式	线上模式	yunna 基z	本属性 快速还原	前往任务管理
Q、任务或文件夹名称 7	① 开发模式与线上模式代码隔离,但	数据末隔离。因此在开发模式下立即执行	亍时,需要注意要临时修改数	效据目录。			×
▼ 🛅 最近打开的任务	ピ 編辑调度 ○ 取消调度 🗎 编辑	报警 ① 详细信息 2 资源信息	Q 搜索节点				± ⊛ ⊜ C3
• 🚠 cookbook_flow	yunna yuni	a yunna					
testsql	Gods_cvs_	eisure_store_info	ods_cvs_opr	_nation_sku_di			
TPCH_flow		T		T			
12324			FEIKH				
DLX_TEST yunna	ok2 Solo a dwd_produ	 t_sales_cookbook	S dwd_product	t_sales_cookbook4			
TPCH_demo				7			
▶ 📩 cookbook示例任务							
• 🛅 etl			g dws_product_sa	ales_cookbook2			
▶ 📩 xr_test			yunna	yunna			
🕨 🛅 yunqi							
▶ 🛅 默认文件夹			ads product sa	les cookbook2			
6			product_od				

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。

三 り数据中台	编辑报警					×
⁶¹ 020年日日令	添加报警 添加数据质量	添加报警		Yunga	如果配置了节点失败报警,	建议同时配置作业失败报警
 	报警对象/对象类型	报警对象	dwd_product_sales_cookbook4_month \times			
▼ □ 最近打开的任务	COOKDOOK_TIOW 多节点任务	触发规则	○ 质量异常 ● 质量检测失败	- 1	¥ -	修改删除
• 🚠 cookbook_flov	dwd_product_sales_cookbook4_	yunna 接收人*	* 🔽 负责人	yunna		
testsql	onth		□ 其他人	- 1	关 -	修改删除
TPCH_flow T2324	数据质量		报警组	- 1		
DLX_TEST yorna	yunna	yuni接收方式*	* 🔽 邮件 🔽 短信 🛛 电话 🔽 POPO	Stone		
→ 🛅 TPCH_demo		循环报警		- 1		
▶ Cookbook示例任		间隔时间	15	分钟		
r en	unna	报警次数	3	次。		
→ 🖿 yunqi	20 ann		yuu yuu	ED:24		
→ 🖿 默认文件夹	确定取消		V把LZE	-4X/F		

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。