# 数据挖掘-参数设置

在数据挖掘中,通过参数筛选用户可查询出需要的数据。

操作步骤增量抽取场景

目前只有部分节点支持参数设置功能,如关系数据源、数据查询、关系目标表(追加)、关系目标表(覆盖)、关系目标表(插入或更新)、SQL脚本、源库SQL脚本。

## 操作步骤

1、新建实验,拖入关系数据源节点,并选择数据源、SCHEMA和表。



2、在公共空间针对本ETL定义一个参数,具体如何配置,参考公共设置-参数定义,下图是配置好的参数示例。

Ş,			Q <b>+</b>	👚 中文 🖻 🗘 🚳 🧐 😭 管理员 🗄
<b>_</b>	Ⅲ 脊 号航   新建参数 常用设置			辺索工具 Q
	[?]参数定义 参数即报表查询筛选条件,用户可通过 改变参数查询条件值改变报表	22 数据格式 用于设置数据集字段显示的格式类型	於較規则 數据脫敏是指对某些敏感信息通过脫敏 規则进行数級的变形,实現敏	中 告警设置 实现分析资源数据告警而提供的格式方 案
	<b>转换规则</b> 查询时把数据库中各字段的内容按需要 转换为熟悉的业务术语	☆ 流程管理 对流程进行管理及新建		
•	更多设置〉			

											_
Ş					C	2 + 🕋	中文 🖸	ا 🗘	🤨 😨	管理员	
	Ⅲ 倄 导航   新建参数 🗙										
Ð	✓	☑ 显示参数	□ 允许多选	□ 手工输入	□ 校验合法性	□ 不显示标题	亙	🗹 输	出合并		
BÒ	> 🛅 DEFAULT	✔ MDX使用引号	允许为空 (存储过程数据集或均	真报属性)	参数缓存: 系统设置 🗸						
	> 🗀 计算字段	备选值设置: *	● SQL ○ 対象 ○ 静态列表 ○ 函	数						-15'	Q,
	> 🗀 过滤器		1 select 饮料,调味品,点	心,日用品,谷类/麦片,肉/家禽,物	寺制品,海鲜 from Categories						
	> Q、业务视图										
	> 🗟 业务主题										
	> 阔 用户属性										
ณ์ไ	> [1] 参数定义	默认值设置:	● SQL ○ 対象 ○ 静态列表 ○ 函	数						15	0
	> 6 函数列表		1								
•				<表达式编辑器,拖动对象	支持: 表、视图、参数、字段、函数	效、过滤器、业9	<b>将属性、用户属</b>	性>			
•						E-		下—	(星友(S)	¥i₹	1(C)
答疑	Q 7=					-	201	i ≫ (N)	PE17(3)		()

#### 3、 点击 **参数设置** 按钮。

#### 注意:在ETL中,需要先保存后,才能对参数按钮进行编辑。原因:涉要设置参数的话,首先需要知道是给哪个资源设置。

Q O		初始化	》参数	属性	帮助
┢ 数据源			A.4		
🗟 文本数据源		æ	天条数据	源*必填(?)	
□• Kafka数据源		Q	数据	原	
s≞b 关系数据源		ସ		and built and	
₿o 示例数据源		_	n	brunwing	
▶ 数据集		0	SCH	ema	
□• 数据查询		<b>~</b>	D	EFAULT	~
■ Excel文件	★系数据源				
□o 读取Excel sheet			表名		
■ 目标源			7**	品类别表	$\sim$
■ 已训练模型					
■ 数据预处理				分区设置	
■ 特征工程			选择列(	D	
🖬 统计分析				はほろい	
➡ 评分卡分析				龙马丰为	
📄 文本分析			SQL语句	(只限于where;	2后语句,
▋ 机器学习			如: nam	ie = zs)	
■ 脚本模块					
■ 自定义模块	参数设置				
■ 服务					

4、配置参数,选择参数类型、参数定义、参数值并保存。这里需要在参数定义中选择刚刚设定好的参数。

; 1月14(八町) 王参叙名/	及参数值		
参数类型参数	参数定义*	参数值 (可选) (7	> +
参数	SMS-20409-sql	饮料	
		_	The State
	参数类型       参数       参数	参数类型 参数定义* 参数 ✓ 请选择 参数 SMS-20409-sql	参数送型     参数定义*     参数値(可选)(       参数         参数     SMS-20409-sql     饮料

#### 参数设置项说明如下:

设置项		说明
参数名	参数的名称。	
参数类型	值	参数值:固定参数值
		类型: • int: 整数 • bigint: 整数(当整数超过int范围使用) • string: 字符串 • double: 双精度浮点型
	参数	参数定义:在产品中定义的参数
		参数值(可选):获取参数定义中的参数值。不选择参数值为动态参数,选择参数值后为静态参数值。

5、在关系数据源参数的SQL语句输入框中,只需要输入where之后的语句。可输入静态条件,如name=zs,或使用含参数的语句: CategoryName=\${param} (表头真名=\${参数名})

》参数	属性	帮助								
关系数据派	夏*必填⑦									
数据源	Inc.									
SM	SMS_20409_northwii 🗸									
SCHE	SCHEMA									
DE	FAULT	~								
表名										
cat	egories	~								
	分区设	E								
选择列*必	埴⑦									
	选择列									
SQL语句()	只限于where	之后语句,如:								
name = z	s)									
Categ	oryName	=\${param}								
		//								

#### 在查看输出中选择表头真名,可查看表头真名。

前显示 4 列 / 总共 4 列, 10 条 / 总共有 10 条数据		列筛选 请选择	表头真名 表头别:
# CategoryID	A CategoryName	Ab Description	A» Picture
1	饮料	软饮料、咖啡、茶、啤酒和淡啤酒	[B@e34780c
2	调味品	香甜可口的果酱、调料、酱汁和调味品	[B@aa13c16
3	点心	甜点、糖和甜面包	[B@6d972f46
4	日用品	乳酪	[B@1a3d7c86
5	谷类/麦片	面包、饼干、生面团和谷物	[B@37e493d5
6	肉/家禽	精制肉	[B@34a658b3
7	特制品	于果和豆乳	[B@609fac7e
8	海鮮	海菜和鱼	[B@11b725a7

#### 6、筛选出产品名称为"饮料"的数据,如图:

٩	查看输出			×
	当前显示 4 列 / 总共 4 列, 1 条 / 总共有 1 条数据	列筛选	请选择	💛 表头真名 🔵 表头别名
	# CategoryID	Ab CategoryName		Ab Description
	1	饮料		软饮料、咖啡、茶、啤酒和淡啤酒
			_	
	提示: 点击单元格可查看超出的内容。注意: 表头中 🔆 表示特征列, * 表示标签	51		下载预览数据

### 增量抽取场景

增量抽取可以显著减少数据传输量和处理时间,比如需要前三天至今天抽取新增的数据,但不希望全量抽取,可以使用参数来实现增量抽取。

1、首先连接好数据源和目标源节点。

"S			Q + 1	¥ 🖻	L <sup>1</sup> 💩	💬 (î	會 管理员	
	Ⅲ 倄 导航   ETL增量抽取 ×							: 🗆
<b>—</b>	O		运行成功:1秒	Ø »	参数	属性	帮助	
	<ul> <li>数据源</li> <li>过本数据源</li> <li>FTP数据源</li> </ul>		6	2 4	关系数据源	必填 ⑦		
	<ul> <li>■ Kafka数据源</li> <li>▲ 关系数据源</li> </ul>	▲ 关系数据源 ②	٥	2	数据源 north	wind	~	
	<ul> <li>■。示例数据源</li> <li>☑ 数据集</li> </ul>		( (		SCHEM.	A .ULT	~	
<u>111</u>	<sup>B</sup> ● 数据查询 <sup>B</sup> ● Excel文件 <sup>B</sup> ● 读取Excel sheet		(	<i>&gt;</i>	表名	ŧ		
4	<ul> <li>■ 目标源</li> <li>■ 数据预处理</li> </ul>	3 关系目标表(追加)			U+4	☆	。 置	
	■ 脚本模块			ì	选择列 ⑦			
				S	SQL语句(只	选择列 限于when	」 e之后语句,:	如:
そう				^ r	name = zs)			

2、在参数定义中编写sql获取当天的日期。

"S						Q + 🖀	A 🗘 💩		管理员
	Ⅲ 倄 导航 । etl_date 🗙								: 🗆
Ŧ	✓	☑ 显示参数	□ 允许多选	□ 手工输入	□ 校验合法性	□ 不显示标题	<b>Z</b>	输出合并	
BQ	> 🖫 DEFAULT	✔ MDX使用引号	□ 允许为空 (存储过程数据	集或填报属性)	参数缓存:系统设置 🗸	•			
uu	> 🗀 计算字段	备选值设置: *	● SQL ○ 对象 ○ 静态列表	○ 函数					1 Q
۲. ۲	> 🗅 过滤器		1 select Now ( )						
	> Q、业务视图								
	> 💼 业务主题								
	> 🖪 用户属性								
<u> í</u> í	> [n] 参数定义	默认值设置:	● SQL 〇 对象 〇 静态列表	○ 函数					₫ Q
	> 🛱 函数列表		1						
4									
				<表达式编辑器,拖动对象支	持: 表、视图、参数、字段、	函数、过滤器、业	务属性、用户属性>		
<b>⑦</b> 答题						上一步(P)	下一步(N)	保存(S)	关闭(C)
<b>=</b> %	Q 7								

#### 2、返回到ETL界面中,把刚刚新建的参数定义在ETL里。

"S				Q + 🖀 🖻 🗳	🕨 💩 💬 🏾 管管理员 🗄
	Ⅲ 倄 导航   ETL增量抽取 ×   成本事实			N	: 🗆
			< ▶ < ඕ [?]	初始化 🔯 🧖 属性	E
82	■ 数据源 ◎ 文木数据源			④ 别名	
5% 2	So FTP数据源	①参数设置 *注意:请依次配置参数名及参	参数值,点击 + 添加参数	×	山增量抽取
. (2)	So Kafka数据源	参数名* 参数类型	参数定义* 参数值(可选) ⑦	描述	
(fr	5 <sup>-</sup> 6 关系数据源 <sup>S</sup> o 示例数据源	Date 参数 ~	请选择 へ 请选择 ~	+	
\$	☑ 数据集		system > 请假申请参数 >		
ഷ്	Bo 数据查询		功能演示 > etl_date		
<u>1111</u>	⇒ 读取Excel sheet		产品类别ID		
4	■ 目标源		崔		
	■ 数据预处理				
				创建的	时间 2022-04-08 11:28:19
				更新	时间 2022-04-08 11:28:19
Ð			确定	取消	
答疑		L			

\$			Q	+ 🏠	ê (	0 💿	管管理员	÷
Ē		RETU#122mm ● R A R C Ø → → R [?]		初始化 🖸	ð » ≉	教 属性	帮助	
BÊ 2	■ 数据源 ■ 文本数据源	○○会教員器 *注意:連体次配置会教名及会教师 占击 + 法加会教		œ		northwind		
	⇒o F1P数据源 Bo Kafka数据源 da 关系数据源 Bo 示例数据源	参数名*         参数定义*         参数值 (可选) ⑦           参数         请选择	× +			DEFAULT 表名		
4	<ul> <li>■ 数据查询</li> <li>■ 数据查询</li> </ul>	Date 参数 etl_date	Ē			分区	。 设置	
≝ ∕₽	<ul> <li>◎● Excel又件</li> <li>◎● 读取Excel sheet</li> <li>■ 目标源</li> </ul>	不设定参数的值			选择	<b>郓列 ⑦</b> 选	郓	
	<ul><li>■ 数据预处理</li><li>■ 脚本模块</li></ul>				SQI nan	L语句(只限于wh me = zs) Orderdate>=\${	ere之后语句,如 Date}-3	4:
<b>⑦</b> 答疑		4652	主 取	۲ ۲				

3、在关系目标源中的SQL语句编辑框中填写如图的内容,即可抽取前三天至今新增的数据。

\$		Q +	🖀 🖻 🔎 💩 💬 管 管理员 🗄
	Ⅲ 🏠 导航 🗆 ETL增量抽取 🗙		: 🗆
Ē	O	■ ETU增量抽取 ④ 昆 □ 昆 Ø × ▶ × 昆 [?] 运行成功:	0秒 🔯 》 参数 属性 帮助
	<ul> <li>数据源</li> <li>文本数据源</li> <li>6. FIP数据源</li> <li>6. Kafka数据源</li> <li>6. Kafka数据源</li> <li>6. 示例数据源</li> <li>6. 数据查询</li> <li>6. Excel文件</li> </ul>	<ul> <li>              ◆系数確源</li></ul>	Q Q SCHEMA Q DEFAULT ○ 表名 ② 订单表 分区设置
4	<sup>圓</sup> • 读取Excel sheet ■ 目标源	节点类型:关系目标表(追加) 抽取的sql语气	近年が、低
	<ul><li>■ 数据预处理</li><li>■ 脚本模块</li></ul>	关系数据源       当前显示 15 列 / 总共 15 列, 0 条 / 总共有 0 条数据	く         SQL语句(只限于where之后语句,如:           表头别名         name = zs)
<b>⑦</b> 答疑		# 订单编号 A 顾客编号 # 雇员编号 @ 订单日期 @ RequiredDate @ ShippedDate # ShipVia	# F Orderdate<= DATE_FORMAT(DATE_SUB('\${ Date)",INTERVAL 3 DAY),"%Y- %m-%d %H:%i:%s")aj