

Data Kaleido 造数平台

（使用手册）

版本：0.1.1

时间：2022 年 5 月

目 录

1 造数模型管理	1
1.1 创建造数模型	1
1.2 编辑造数模型	5
1.2.1 增加字段	5
1.2.2 删除字段	6
1.2.3 移动字段位置	6
1.2.4 修改字段参数	7
1.2.5 修改数据生成参数	8
1.2.6 修改数据输出参数	8
1.2.7 数据预览	9
1.2.8 导入造数模型	9
1.3 保存造数模型	10
1.4 删除造数模型	10
1.5 导出造数模型	11
2 造数任务管理	12
2.1 生成造数任务	12
2.2 查看任务详情	13
2.3 下载数据文件	14
2.4 删除造数任务	14
3 数据类型参数	17
3.1 数据类型总览	17
3.2 参数说明	20
3.2.1 字段固定参数	20
3.2.2 数据类型的特定参数	21
Int	21
Float	23
Date	24
Time	27
Date&Time	29
Char	31

Varchar	32
Mix	34
Birthday	34
UUID	35
4 数据生成参数	36
5 数据输出参数	37
文件格式: CSV	37
文件格式: SQL	37
文件格式: JSON	38

1 造数模型管理

DK 可以对造数模型进行管理。

1.1 创建造数模型

可以通过四种不同的方式创建造数模型。

前三种在顶部导航的造数模型列表页面实现。

进入顶部导航栏造数模型列表页面后，点击右上角的创建模型按钮，将弹出创建模型页面，如图 1。

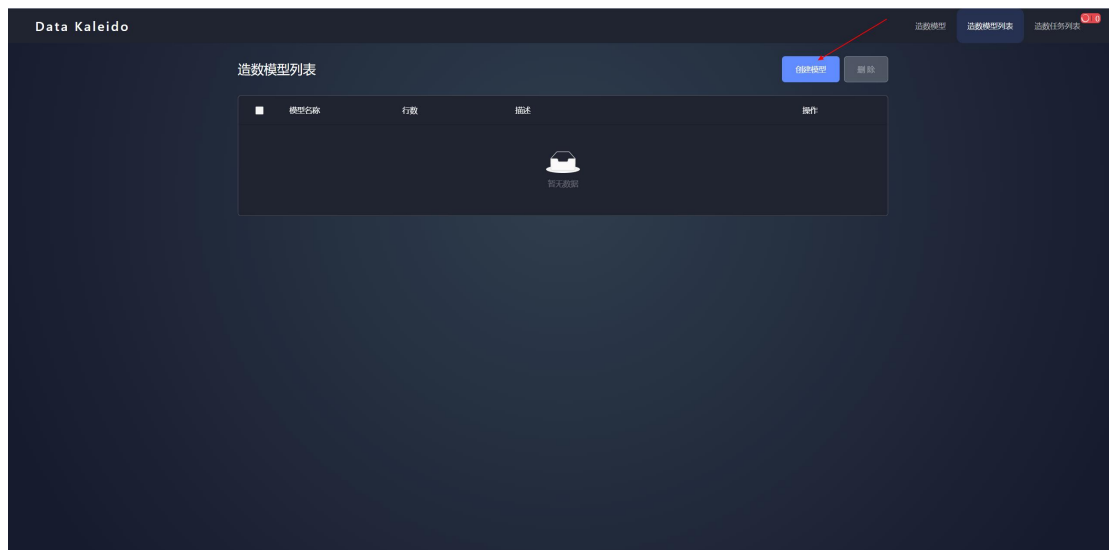


图 1

在创建模型页面中，有三种不同的方式创建新模型，如图 2。

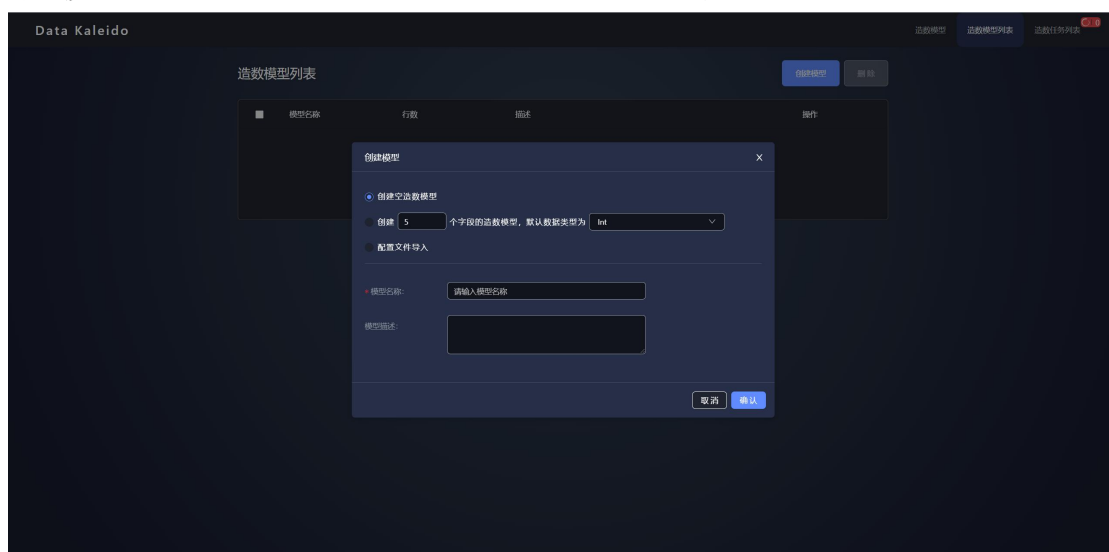


图 2

第一种：创建空模型。

点击左边单选框，选中第一种创建方式，输入模型名称(必填)、模型描述信息，如图 3。

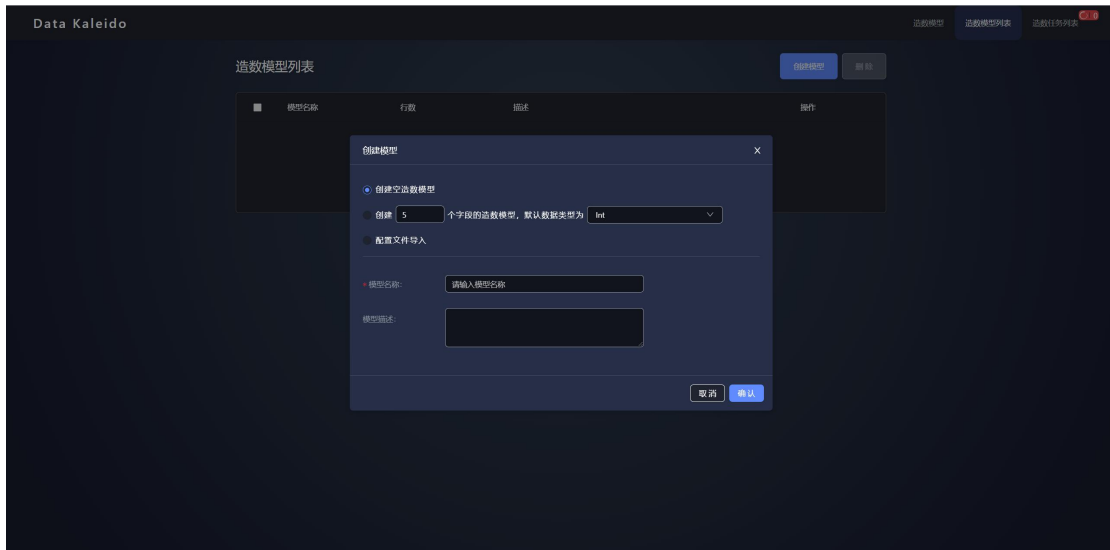


图 3

点击右下角**确认**按钮，完成模型的创建，如图 4。

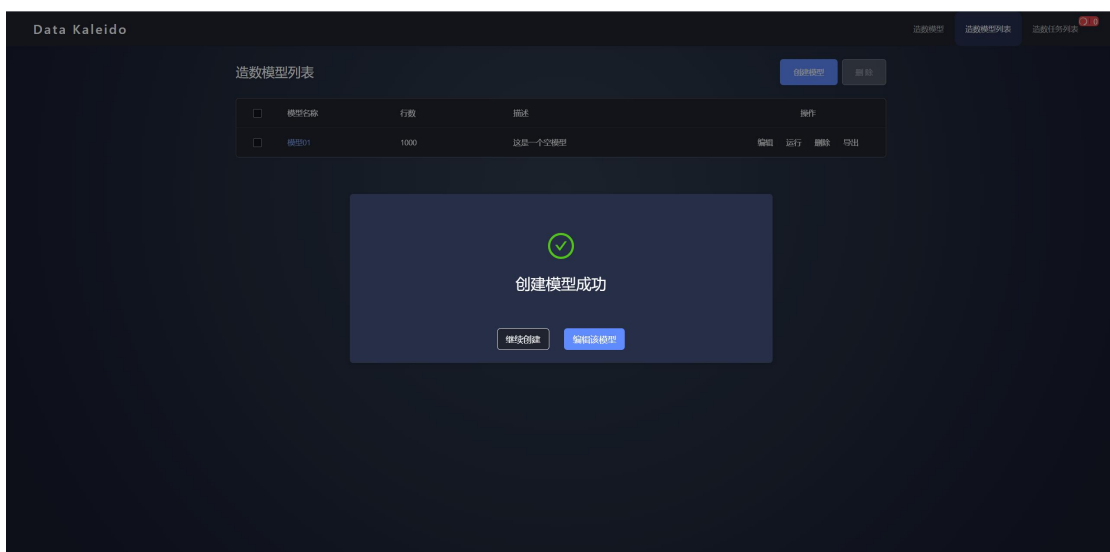


图 4

第二种：创建指定字段数量、以及默认数据类型的造数模型。

点击左边单选框，选中第二种创建方式，输入字段数量、选择默认数据类型，接着输入模型名称(必填)、模型描述信息，如图 5。

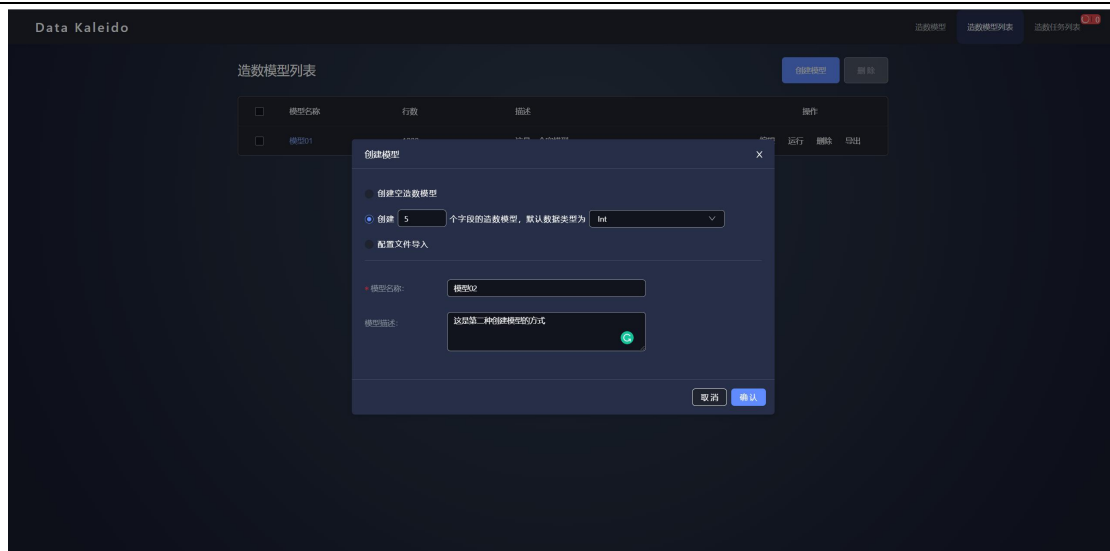


图 5

点击右下角**确认**按钮，完成模型的创建，如图 6。

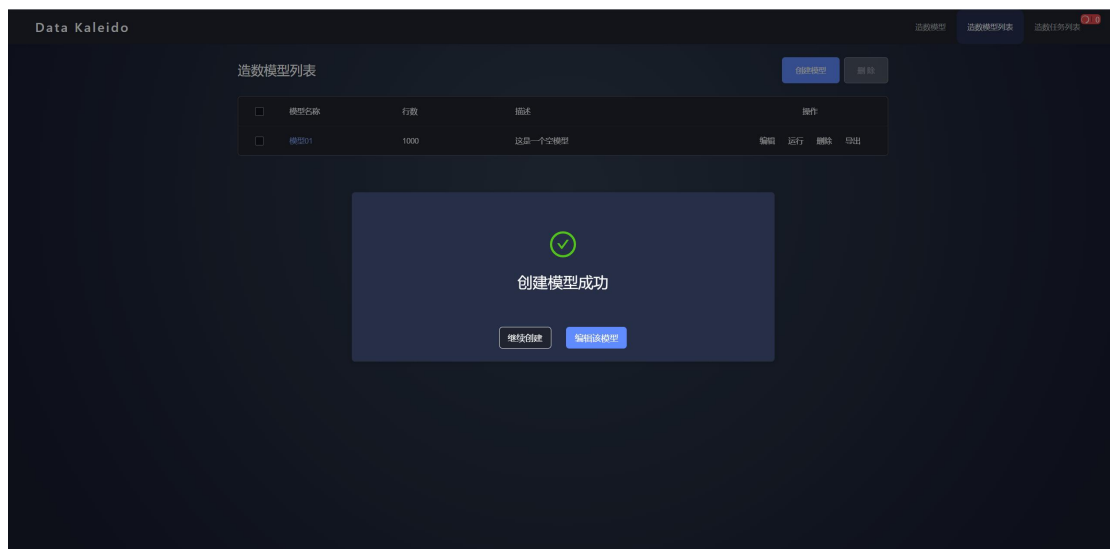


图 6

第三种：通过导入配置文件，创建新模型。

点击左边单选框，选中第三种创建方式；点击页面中的虚线框，弹出配置文件选择页面，选中需要导入的配置文件；或者通过鼠标，直接将配置文件拖入虚线框，如图 7。

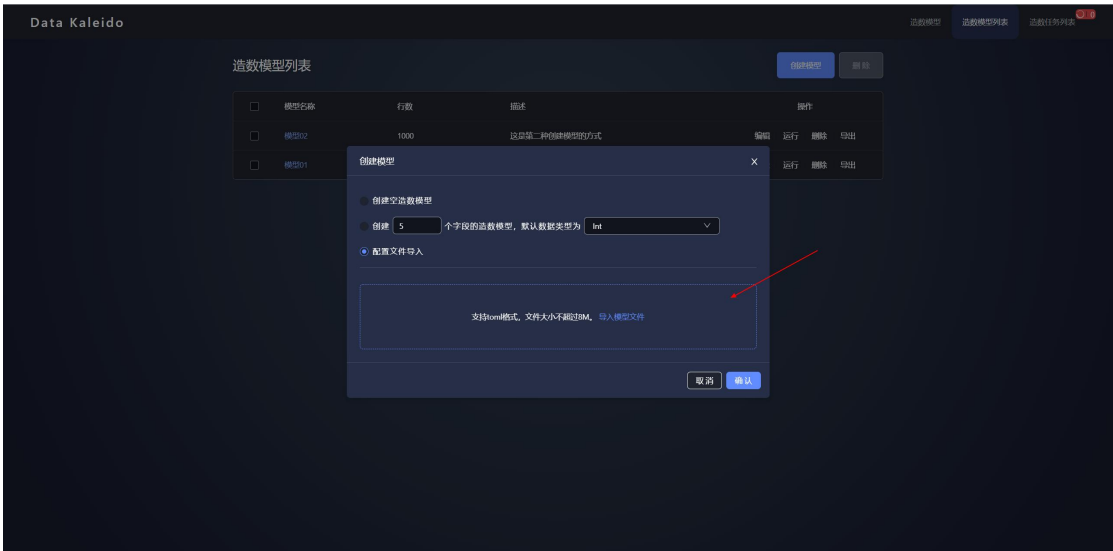


图 7

导入完成后，在虚线框下面会出现配置文件名，如图 8。

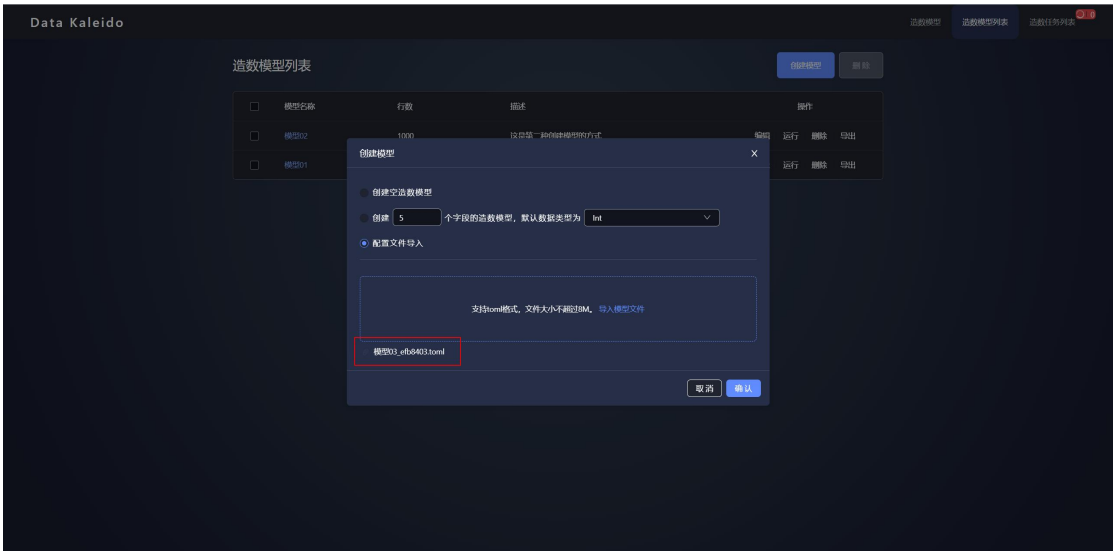


图 8

点击右下角的**确认**按钮，完成模型的创建，如图 9。

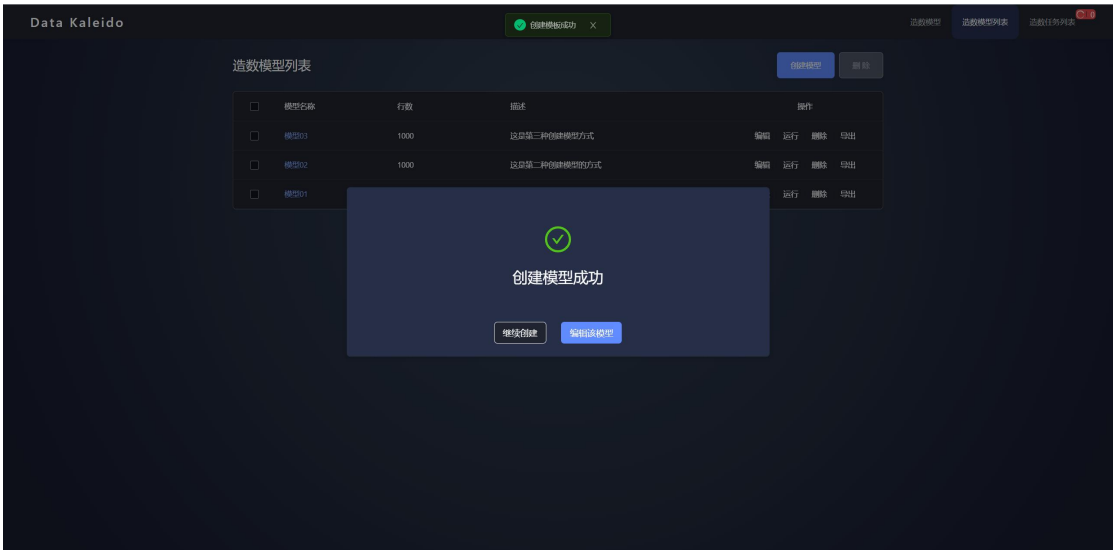


图 9

最后一种是在顶部导航的**造数模型**页面实现。

点击底部的**保存模型**按钮，弹出**保存模型**按钮，输入模型名称(必填)、模型描述信息，点击**确认**按钮，完成模型创建，如图 10。

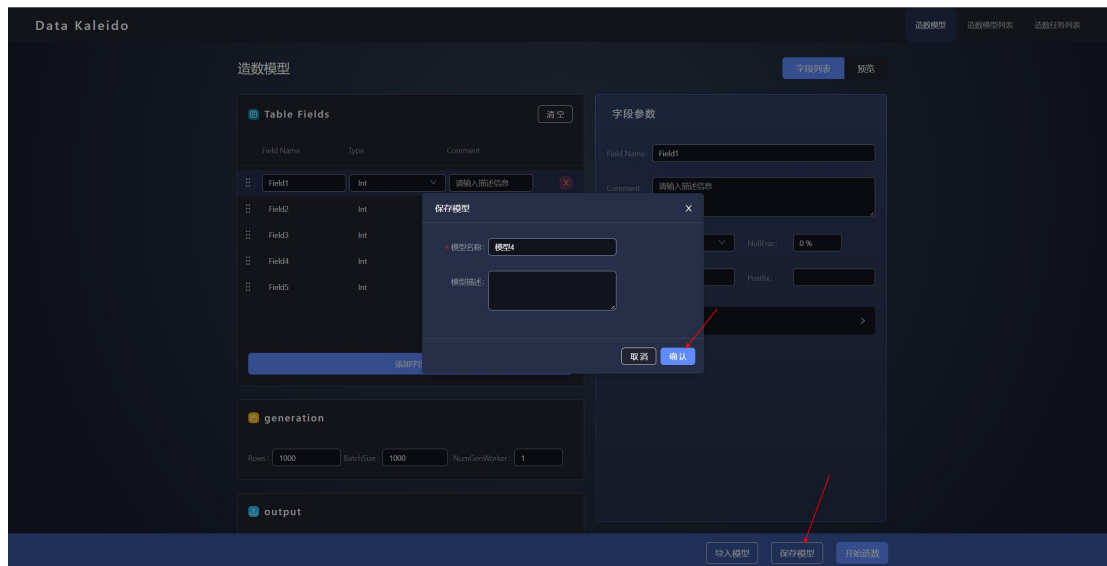


图 10

1.2 编辑造数模型

在顶部导航栏中的**造数模型**页面，可以编辑造数模型参数信息，具体操作如下。

1.2.1 增加字段

在左侧的 Table Fields 表单中，输入批量添加的字段数量，点击**添加字段**按钮，完成批量增加字段操作，如图 11。新添加字段的默认数据类型为 Int。

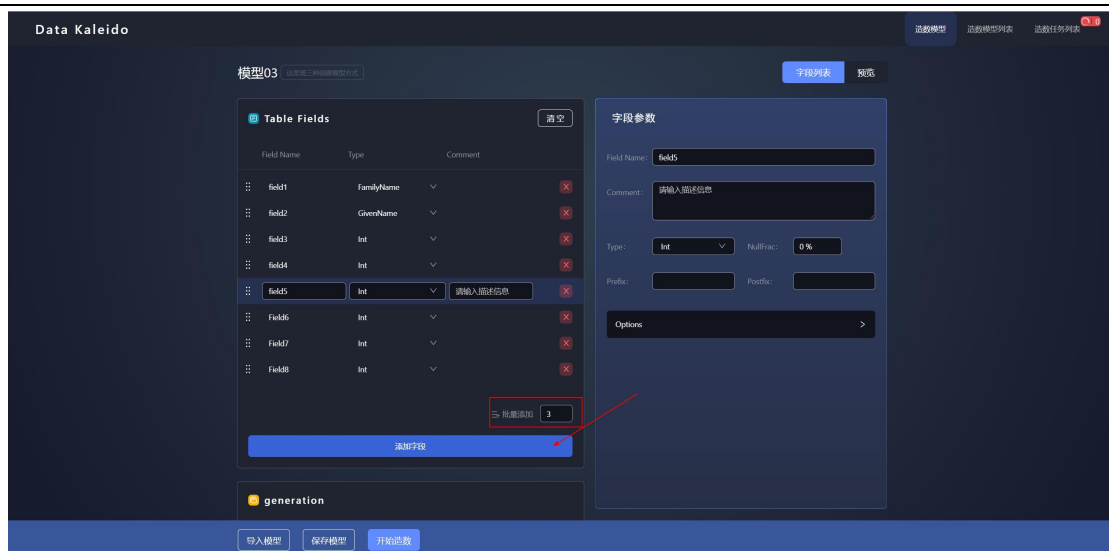


图 11

1.2.2 删除字段

在左侧的 Table Fields 表单中，选中将要删除的字段，点击右侧删除按钮，完成单个字段的删除操作，也可以点击右上角的清空按钮，将造数模型的字段全部删除，如图 12。

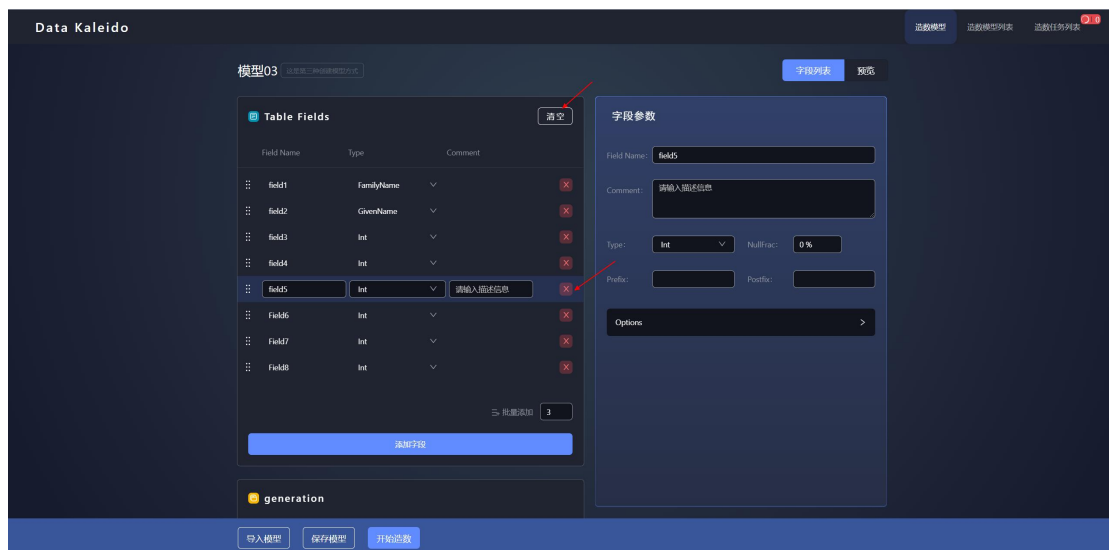



图 12

1.2.3 移动字段位置

在 Table Fields 表单中，使用鼠标左键长按选中字段左侧图标 ，通过上下拖拉，移动字段的位置，如图 13。

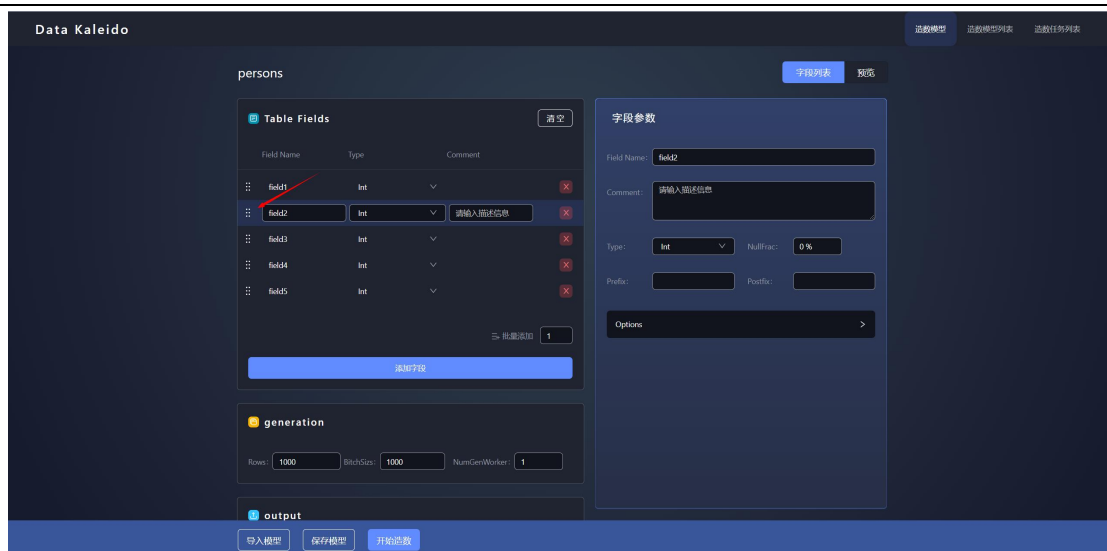


图 13

1.2.4 修改字段参数

如图 14，在左侧的 Table Fields 表单中，选中要修改的字段，右侧的字段参数表单会展示该字段的所有参数，用户可以根据自己的需求对参数进行修改。

- 每个字段均存在以下 6 个参数：
 - Field Name: 字段名称，不为空，且唯一
 - Type: 数据类型
 - Comment: 字段备注信息
 - NullFrac: 空值比例(取值范围:0~100)
 - Prefix: 前缀
 - Postfix: 后缀
- Options 是不同数据类型的可选参数，依赖于参数 Type，详细说明见第 3 节。

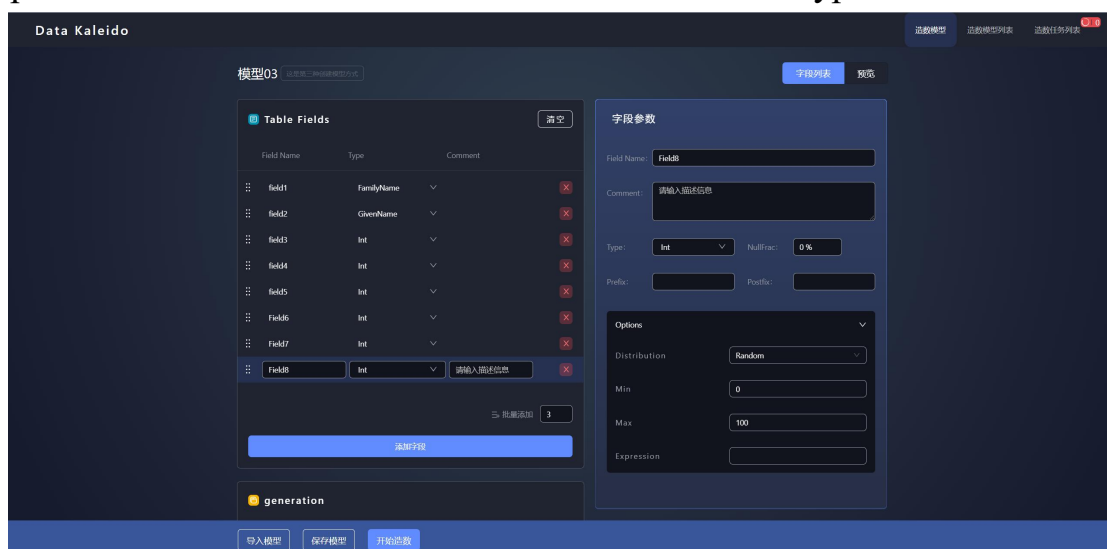


图 14

1.2.5 修改数据生成参数

如图 15，在 generation(数据生成)表单中，可以修改的参数如下：

- Rows: 构造数据的总行数
- BatchSize: 批量生成的数据行数
- NumGenWorker: 数据生成的并发数

详细说明见第 4 节。

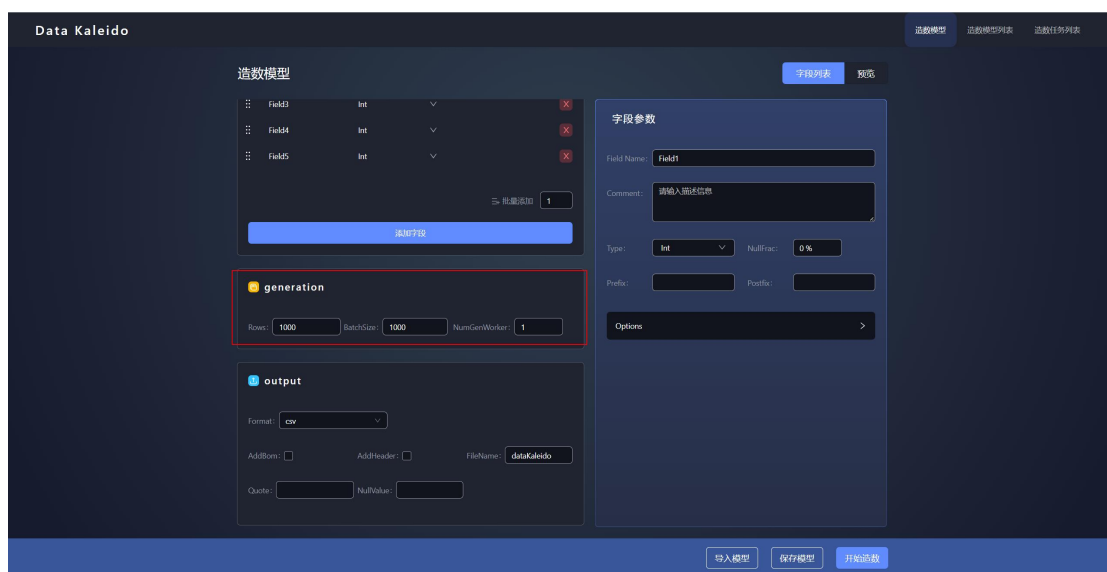


图 15

1.2.6 修改数据输出参数

如图 16，在 output(数据输出)表单中，可以修改数据的输出格式 Format。不同的输出格式，output 表单下方会出现不同可选参数，其详细说明见第 5 节。

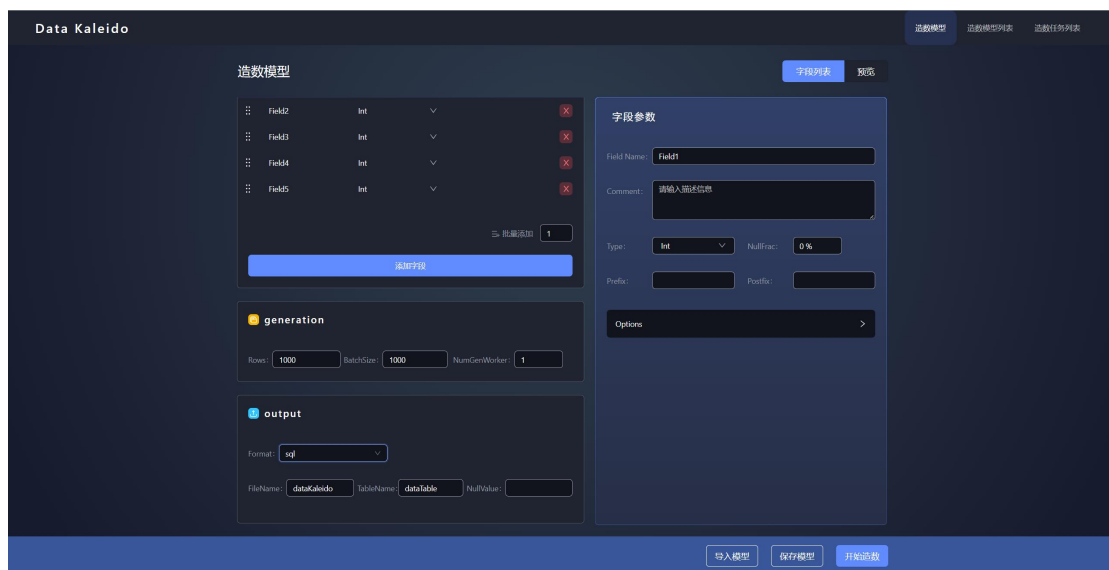


图 16

1.2.7 数据预览

点击右上角**预览**按钮，切换到预览页面，可以查看要构造的数据，如图 17。默认的预览数据量为 30 行。

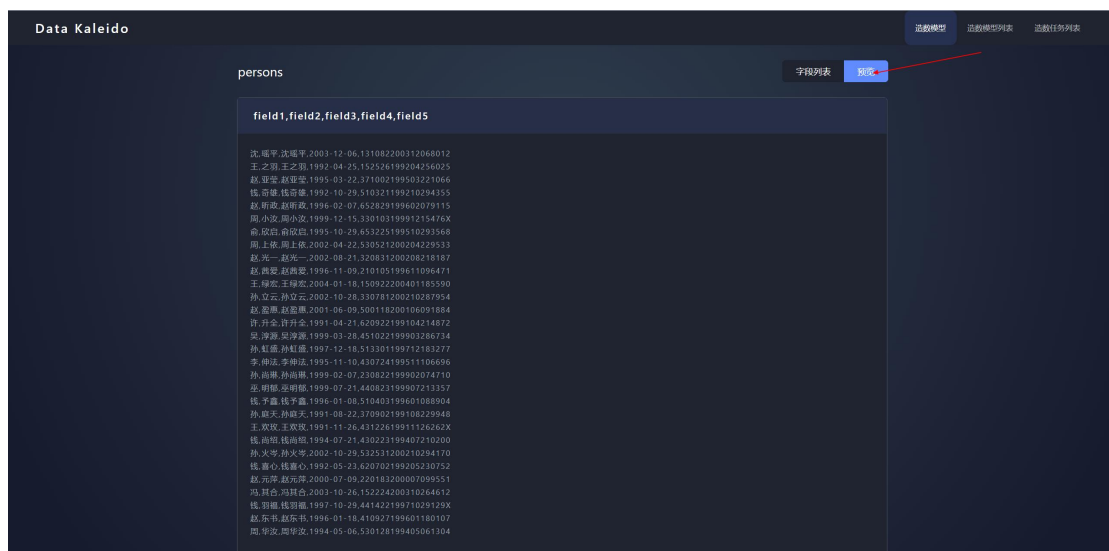


图 17

1.2.8 导入造数模型

如图 18，点击底部的**导入模型**按钮，弹出导入模型页面；点击页面中的虚线框，弹出配置文件选择页面，选中需要导入的配置文件；或者通过鼠标，直接将配置文件拖入虚线框；最后点击右下角确认按钮，完成模型导入。

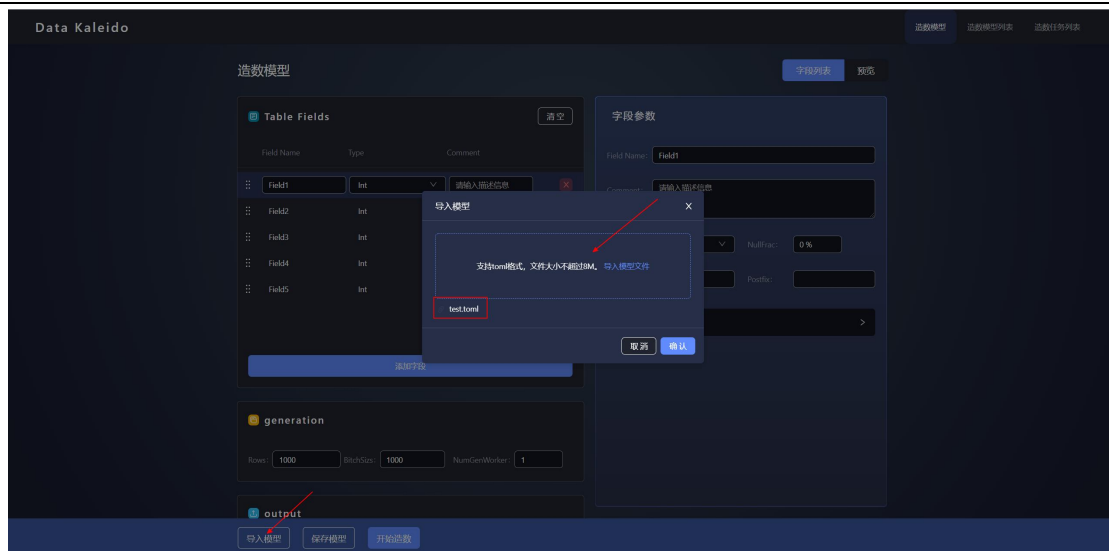


图 18

1.3 保存造数模型

在顶部导航栏中的**造数模型**页面，编辑好模型后，点击底部的**保存模型**按钮，将修改的模型保存，防止定义好的参数丢失，如图 19。

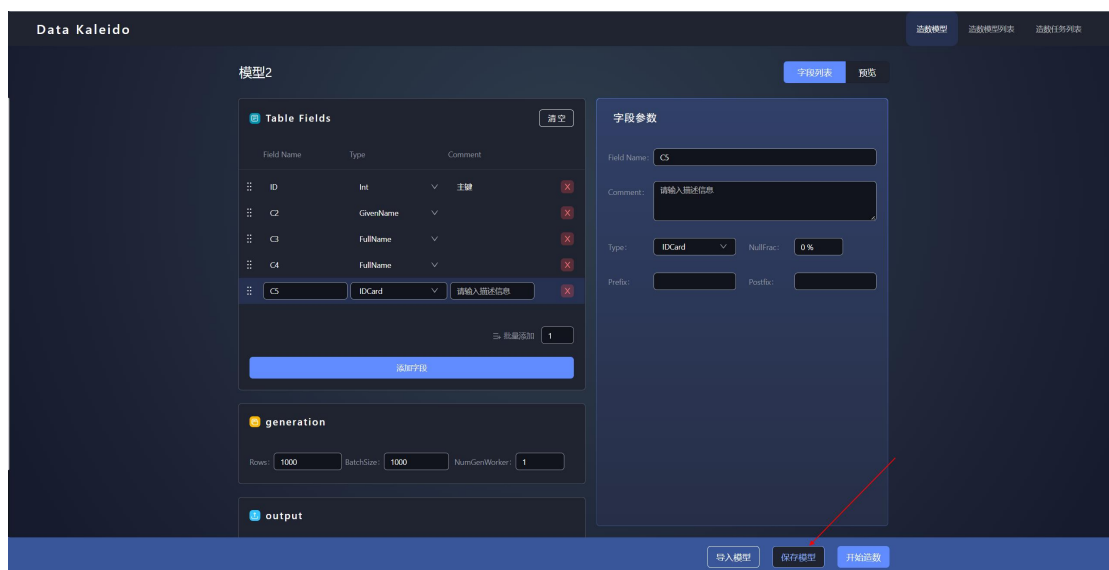


图 19

1.4 删除造数模型

进入顶部导航栏中的**造数模型列表**页面，可以通过以下两种方式删除已保存的造数模型：

第一种：删除单个造数模型。将鼠标移动至指定造数模型所在行，点击右侧操作

列中的**删除**按钮，完成该模型的删除，如图 20。

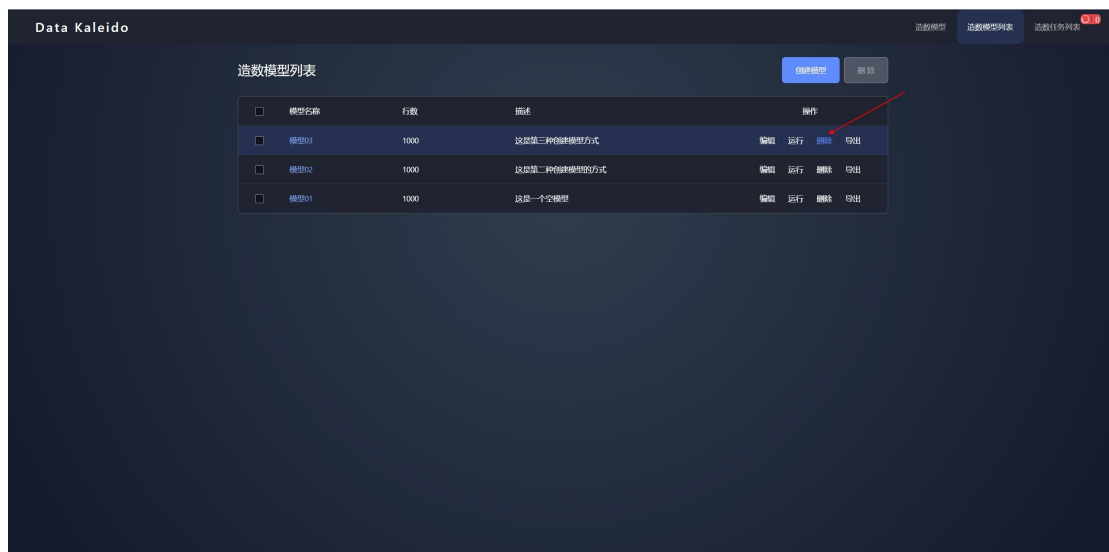


图 20

第二种：批量删除多个造数模型。在左侧多选框选，选中多个造数模型，点击右上角的**删除**按钮，完成模型的批量删除，如图 21。

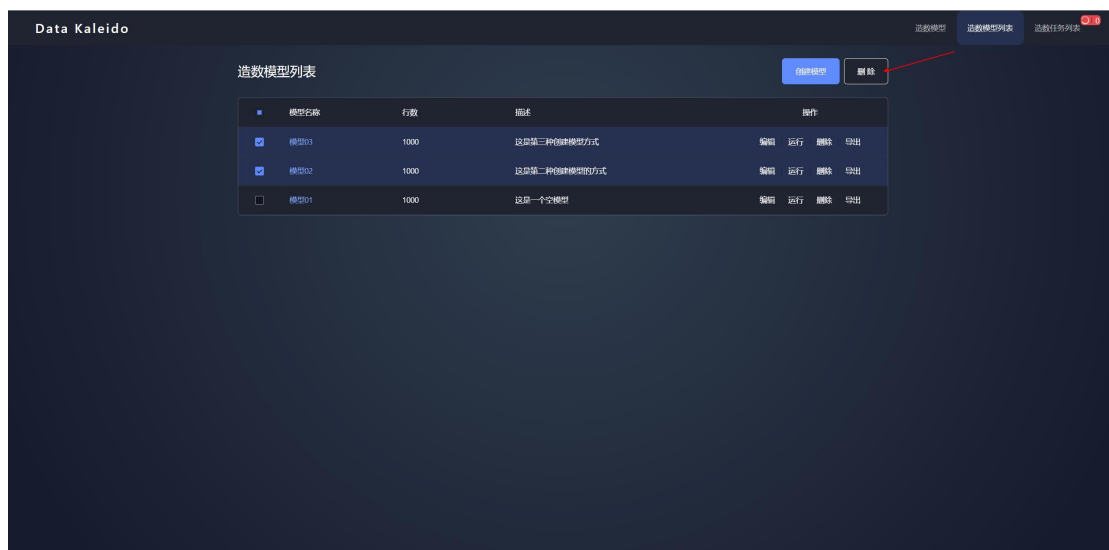


图 21

1.5 导出造数模型

进入顶部导航栏中**造数模型列表**页面，将鼠标移动至指定模型所在行，点击右侧操作列中的**导出**按钮，可以将造数模型导出至配置文件，如图 22。

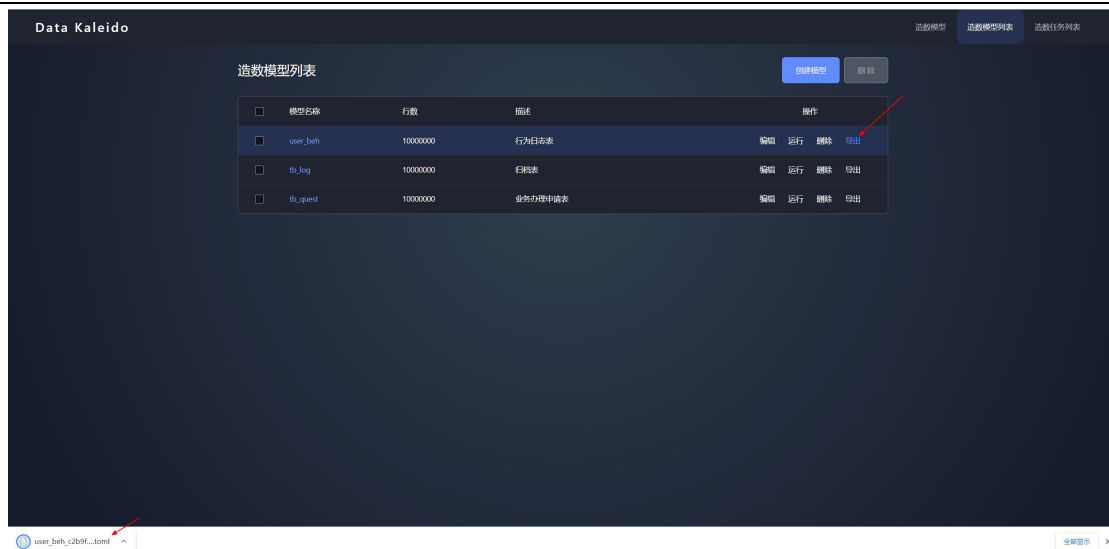


图 22

2 造数任务管理

DataKaleido 可以对造数进行管理。

造数任务是来自造数模型。每个造数任务会生成一个数据集文件，任务完成后，可以下载该数据集文件。从完成时间起，造数任务的有效期为 7 天，超过该时间，任务自动变成失效状态，构造的数据集也会被删除。

2.1 生成造数任务

可通过以下两种方式生成造数任务

第一种：进入顶部导航栏**造数模型列表**页面，使用鼠标移动到指定模型所在行，点击操作列中的**运行**按钮，可以生成造数任务，如图 23。

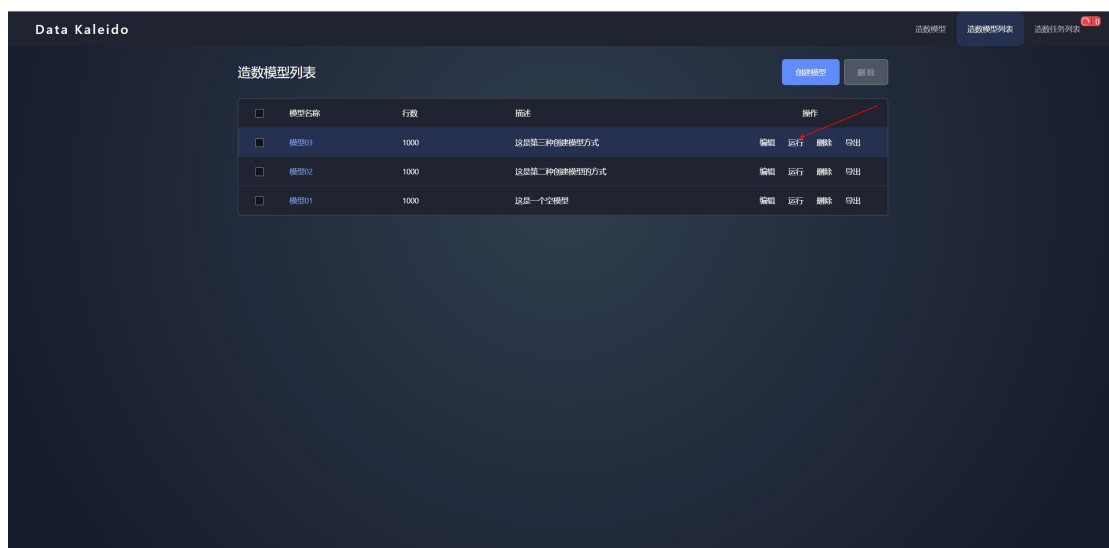


图 23

第二种：进入顶部导航栏**造数模型**页面，点击底部的**开始造数**按钮，同样可以生成造数任务，如图 24。

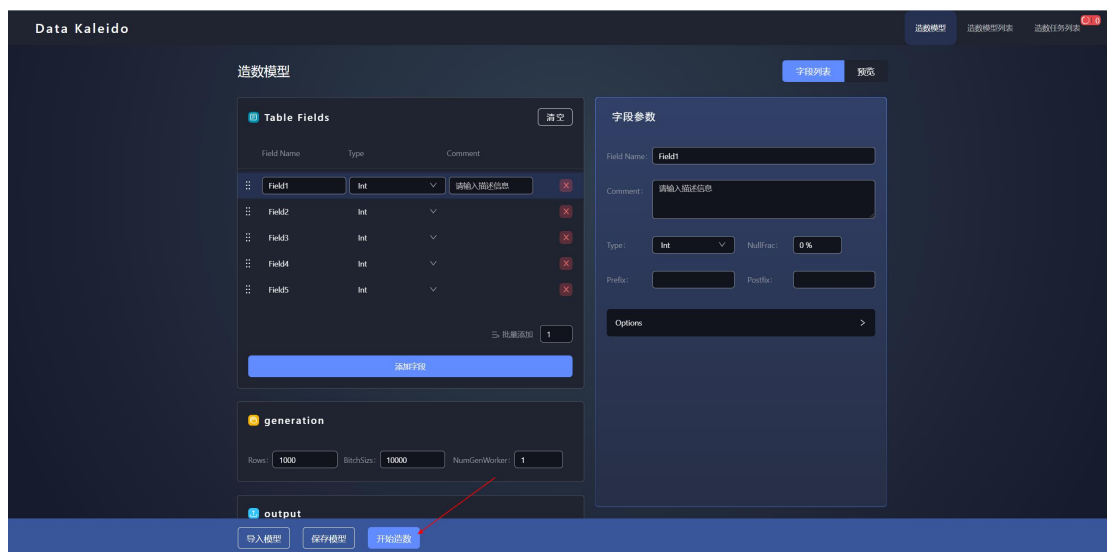


图 24

2.2 查看任务详情

在顶部导航栏中的**造数任务列表**页面，将鼠标移动到指定造数任务行，点击造数任务的 id，会弹出该任务的详细信息，如图 25。

任务详细信息包括：任务 id、创建时间、结束时间、剩余可下载时间和对应造数模型的字段列表等信息，如图 26。

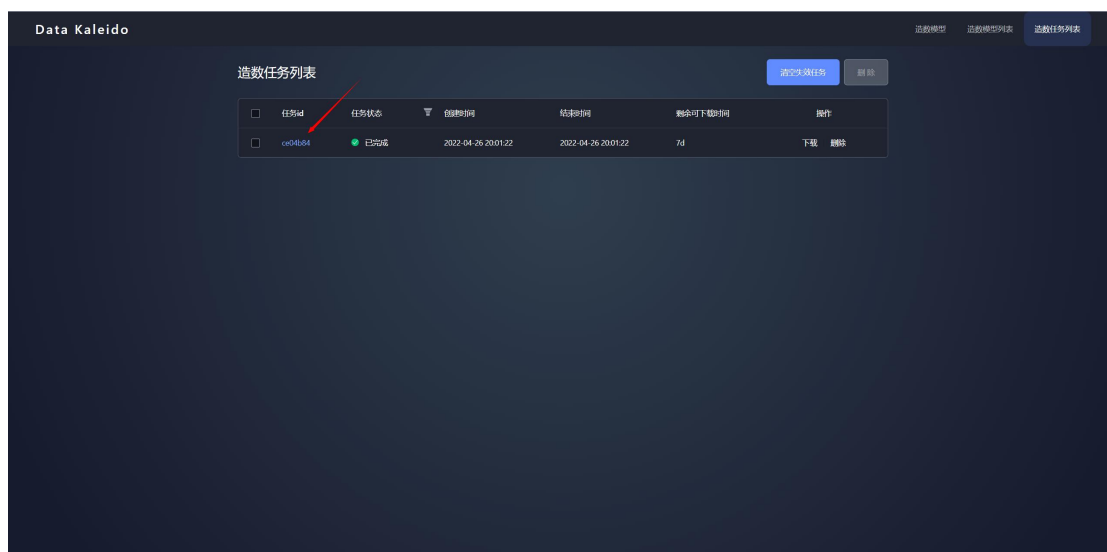


图 25

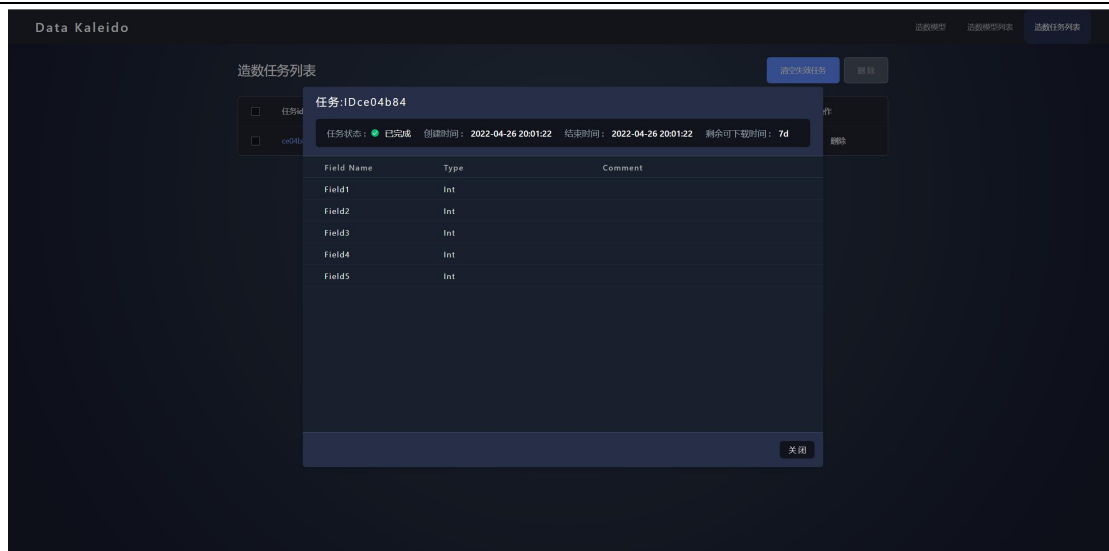


图 26

2.3 下载数据文件

在顶部导航栏中的**造数任务列表**页面，将鼠标移动到指定造数任务所在行，点击右侧操作列的**下载**按钮，可以下载构造好的数据集文件，如图 27。需要注意的是，只有成功状态的造数任务，可以进行数据下载。

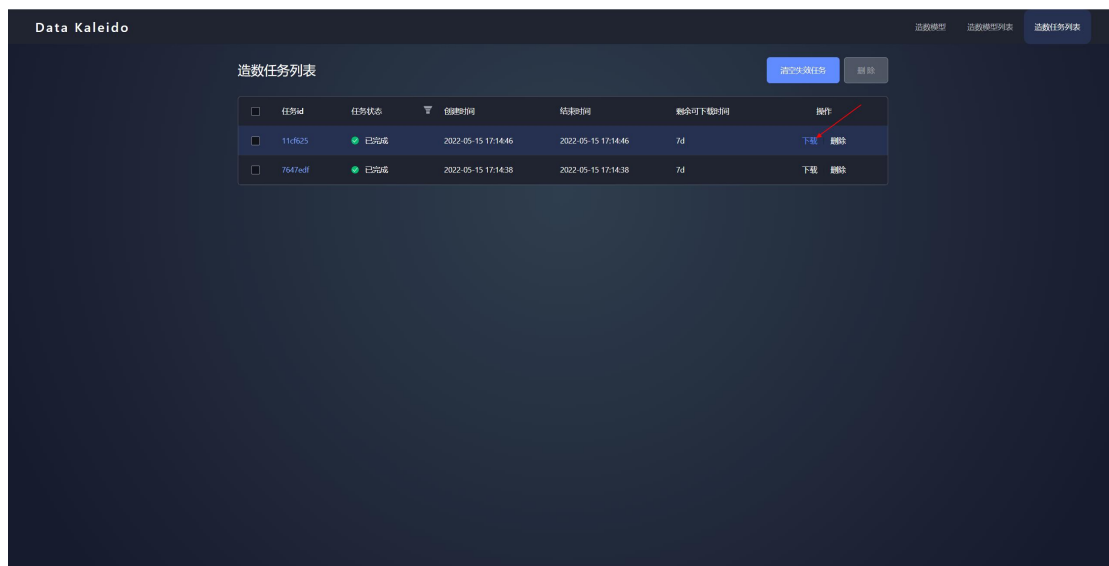


图 27

2.4 删除造数任务

在顶部导航栏中的**造数任务列表**页面，将鼠标移动到指定任务所在行，点击右侧操作列的**删除**按钮，可以删除对应造数任务，如图 28。

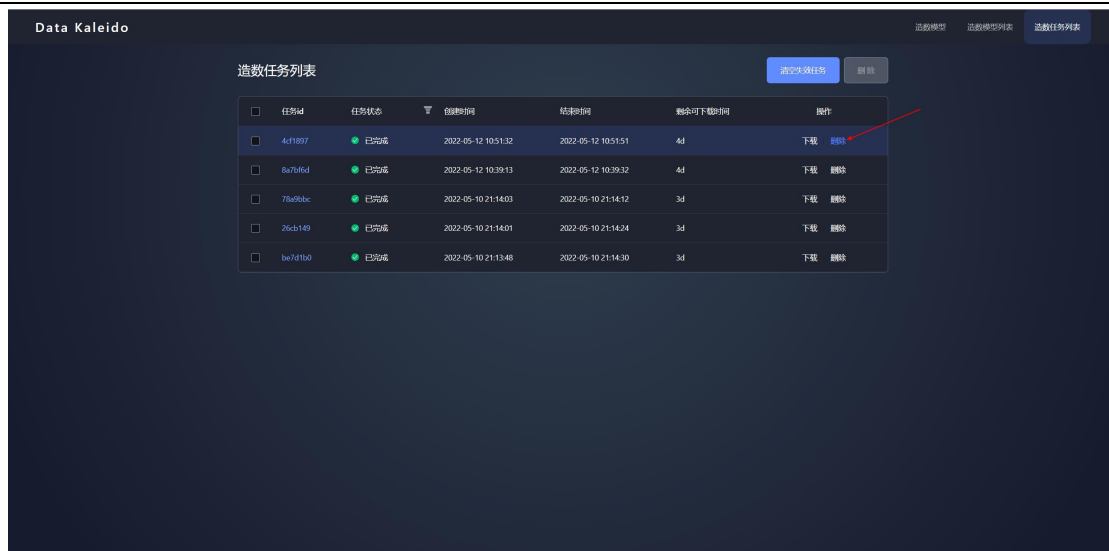


图 28

也可以选中多个任务，点击右上角**删除**按钮，进行批量删除，如图 29。

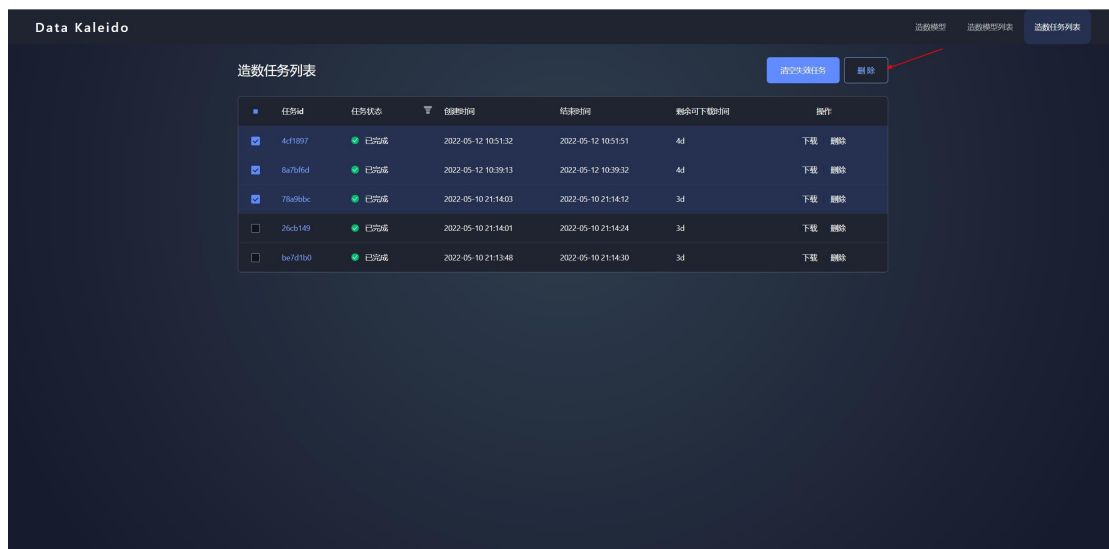


图 29

对于已经失效的造数任务，可以点击**清空失效任务**按钮，完成失效任务的一键删除，如图 30。

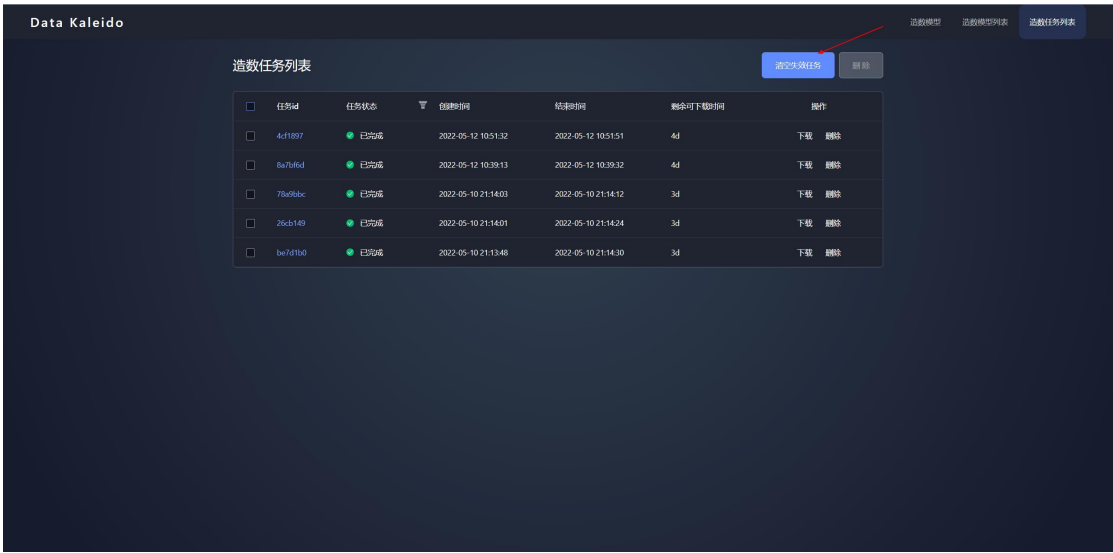


图 30

3 数据类型参数

3.1 数据类型总览

DK 共提供了以下 40 种数据类型。

大类	数据类型	造数说明	样例数据
基本(10种)	Int	按照指定的分布，或者计算表达式，生成 Int 类型数据。	1 10
	Float	按照指定的分布，或者计算表达式，生成 Float 类型数据。	0.1 10.02
	Date	按照指定分布，生成 Date 类型数据。	2022-04-07 2022/04/07
	Time	按照指定分布，生成 Time 类型数据。	08:00:00 18:50.59.9
	Date&Time	按照指定分布，生成 Date&Time 类型数据。	2022-04-07 08:00:00 2022/04/07 18:50.59
	Bool	随机生成 Bool 类型数据。	true false
	Char	按照指定的数据格式，随机生成固定字符数量的字符串。	16--WG--8J - wo 73--WO--Cz - lb
	Varchar	按照指定的数据格式，随机生成字符数量可变的字符串。	05-tt- 36-ah-FM-y
	List	按照指定分布，在自定义数据集中取值，用于生成数据。	

	Mix	<p>Mix 是一个混合类型，是由多个子字段组成。除了 Mix 类型本身，这些子字段可以是其它的任意数据类型。</p> <p>其造数的原理为：将多个子字段生成的数据，使用符号 Connector 连接在一起，组合成 Mix 类型数据。</p>	
Person 相关(13种)	FamilyName	随机生成中国人的姓氏。	李 钱
	GivenName	随机生成中国人的名字。	政和 俞启
	FullName	<p>随机生成中国人的姓名。</p> <p>备 注：与 FamilyName、GivenName 关联</p> <p>FullName = FamilyName + GivenName</p>	李政和 钱俞启
	Birthday	随机生成出生时间	1996-11-05 1999-11-27
	Age	<p>随机生成年龄。</p> <p>备注：与 Birthday 关联，由出生时间计算获得</p>	25 22
	IDCard	随机生成中国人的身份证号码。	522633199611055421 451082199911279890
	Gender	<p>随机生成中文性别。</p> <p>备注：与 IDCard 关联，由 IDCard 中顺序码计算获得</p>	男 女
	Phone	随机生成电话号码。	18212817005 15828250902
	Email	随机生成电子邮件地址。	AaliyahConnelly@Herman.com RaoulLindgren@Schulist.org

	JobTitle	随机生成职业名称。	园艺师 法官
	Education	随机生成学历名称。	高中 本科
	AcademicDegree	随机生成学位名称。	学士 硕士
	FamilyAddress	随机生成家庭住址。	陕西省汉中市南郑区产旅路 6362 号砖旁厚小区 15 单元 1736 室
Location 相关 (4 种)	Province	随机生成中国版的省市区数据， 之间存在相互映射关联。	山西省,临汾市,汾西县
	City		辽宁省,阜新市,细河区
	District		黑龙江省,牡丹江市,爱民区
	Street	随机生成中国版的街道数据。	江西省,宜春市,丰城市 正阳区街道 长甸街道
IT 相关 (9 种)	APPName	随机生成 app 名称	哔哩哔哩 QQ 音乐
	APPVersion	随机生成 app 版本号	2.16.18 1.9.9
	DomainName	随机生成域名	sendspace.com ucla.edu
	IPV4	随机生成 IPV4 地址	59.198.180.79 186.10.185.217
	IPV6	随机生成 IPV6 地址	f418:f2e5:d796:e7df:44f:446d:2a80:ef43
	MACAddress	随机生成 MAC 地址	56:3b:06:39:4a:0c
	MD5	随机生成 MD5 编码	cfdc208495d565ef66e7dff9f98764da
	URL	随机生成 URL	http://www.bloomberg.com/pay/extension/region/degree/council
	UUID	随机生成 UUID	fc44d154-6dd0-46a9-99f0-3bab2af25b97

Bank 相关(4 种)	BankCardNo	随机生成银行卡号	6216670019165820387 52894819874293833
	BankCardType	随机生成银行卡类型	信用卡 借记卡
	BankName	随机生成银行的名称	中国建设银行 中国邮政储蓄银行
	BankEnglishAbbreviation	随机生成银行卡英文简称 备注：与 BankName 关联，是对应银行的英文简称	CCB PSBC

3.2 参数说明

在字段参数设置中，均存在 3.2.1 节中的 6 个固定参数；部分数据类型还有其特定的可选参数，其详细说明见 3.2.2 节。

3.2.1 字段固定参数

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Name	字段名称	string	"Field" + n(n 为造数模型中的字段总数) eg: 在含有 5 个字段的造数模型中添加一个字段，新增字段的默认名称为："Field6"	否	任意字符串，但是在同一个造数模型下，每个字段名称唯一。
Type	数据类型	string	"Int"	否	可选项值，必须是 3.1 节中给出的 40 种数据类型之一
NullFrac	空值比例	int	0	否	$0 \leq \text{NullFrac} \leq 100$
Comment	注释	string	""	是	
Prefix	前缀	string	""	是	

Postfix	后缀	string	""	是	
---------	----	--------	----	---	--

3.2.2 数据类型的特定参数

Int

参数名称	描述	数据类型	默认值	是否可以 为空	取值范围
Min	最小值	int64	0	否	$-1 * (1 \ll 63 - 1) \leq \text{Min} < (1 \ll 63 - 1)$
Max	最大值	int64	100	否	$-1 * (1 \ll 63 - 1) \leq \text{Max} < (1 \ll 63 - 1)$, $\text{Min} \leq \text{Max}$
Distribution	数据分布类型	string	"Random"	否	可选取值: "Random" "Sequence" "Custom_FrequencyHistogram" "Custom_EqualHeightHistogram"
Step	步长	int64	1	否	$1 \leq \text{Step} \leq \text{Max} - \text{Min}$
IsPrimaryKey	是否为主键	bool	false	否	
Expression	计算表达式	string	" "	是	只能引用 Int、Float 类型的字段 eg: $c1 * 10 + 100$, 其中 c1 为其他字段的名称(Int 或者 Float 类型)

FrequencyHistogramFile	频率直 方图文 件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值以及其频率</p> <p>第一列中的取值必须是 Int 类型数据</p> <p>eg:</p> <p>value,Frequency</p> <p>10,2</p> <p>20,2</p> <p>100,6</p>
EqualHeightHistogramFile	等高直 方图文 件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值，必须是 Int 类型数据</p> <p>第二列的值必须大于或等于第一列的值</p> <p>第三列是其频率</p> <p>eg:</p> <p>LowerValue,UpperValue,Frequency</p> <p>1,10,100</p> <p>11,100,100</p> <p>100,101,10</p>

不同的数据分布对应的可选参数：

数据分布	可选参数	造数原理
Random	Min、Max、Expression	<p>提供以下方式构造 Int 类型数据</p> <p>1、在取值范围[min,max]内，随机生成 Int 类型数据。</p> <p>2、按照 Expression 中的表达式，计算出要生成的 Int 数据。</p>
Sequence	Min 、 Max 、 Step 、 IsPrimaryKey	在取值范围[min,max]内，按照步长 Step 序列化生成 Int 类型数据。
Custom_FrequencyHistogram	FrequencyHistogramFile	基于 FrequencyHistogramFile 文件中给出的 Int 类型数据的频率直方图信息，进行造数。

Custom_EqualHeightHistogram	EqualHeightHistogramFile	基于 EqualHeightHistogramFile 文件中给出的 Int 类型数据的等高直方图信息，进行造数。
-----------------------------	--------------------------	---

Float

参数名称	描述	数据类型	默认值	是否可以 为空	取值范围
Min	最小值	float64	0	否	
Max	最大值	float64	100	否	Min <= Max
Distribution	数据分布类型	string	"Random"	否	可选取值： "Random" "Sequence" "Custom_FrequencyHistogram" "Custom_EqualHeightHistogram"
Step	步长	float64	0.01	否	$0 < \text{Step} \leq \text{Max} - \text{Min}$
Decimals	精确小数位	int	2	否	$-1 \leq \text{Decimals} \leq 20$
Expression	计算表达式	string	" "	是	只能引用 Int、Float 类型的字段 eg: $c1 * 1.2 + 100.1$ ，其中 c1 为其他字段的名称 (Int 或者 Float 类型)
FrequencyHistogramFile	频率直方图文件	string	""	否	文件格式为 csv、大小限制在 100M 以下 文件的内容如下： 第一行为表头 第一列和第二列分别是取值以及其频率 第一列中的取值必须是 Float 类型数据 eg: value,Frequency 11.1,2 12.2,2 13.3,6

EqualHeightHistogramFile	等高直 方图文 件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值，且必须是 Float 类型数据</p> <p>第二列的值必须大于或等于第一列的值</p> <p>第三列是其频率</p> <p>eg:</p> <p>LowerValue,UpperValue,Frequency</p> <p>1.1,10.1,100</p> <p>11.1,100.1,100</p> <p>111.1,1001.1,10</p>
--------------------------	-----------------	--------	----	---	---

不同的数据分布对应的可选参数

数据分布	可选参数	造数原理
Random	Min、Max、Decimals、Expression	<p>提供以下方式构造 Int 类型数据</p> <p>1、在取值范围[min,max]内，随机生成 Float 类型数据。</p> <p>2、按照 Expression 中的表达式，计算出要生成的 Float 类型数据。</p>
Sequence	Min、Max、Step、Decimals、IsPrimaryKey	在取值范围[min,max]内，按照步长 Step 序列化生成 Float 类型数据。
Custom_FrequencyHistogram	FrequencyHistogramFile Decimals	基于 FrequencyHistogramFile 文件中给出的 Float 类型数据的频率直方图信息，进行造数。
Custom_EqualHeightHistogram	EqualHeightHistogramFile Decimals	基于 EqualHeightHistogramFile 文件中给出的 Float 类型数据的等高直方图信息，进行造数。

Date

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
------	----	------	-----	--------	------

Min	最小值	time.Time	time.Now() - 1Month eg: 2022-04-24	否	Min >= 1678-01-01
Max	最大值	time.Time	time.Now() eg: 2022-03-24	否	Max <= 2261-12-31 Min <= Max (时间 Min 在时间 Max 之前)
Distribution	数据分布类型	string	"Random"	否	可选取值: "Random" "Sequence" "Custom_FrequencyHistogram" "Custom_EqualHeightHistogram"
Step	步长	string	"1D"	否	Max - Min >= Step step 的时间单位有: Y、M、D、h、m、s、i、o、n(分别对应年、月、日、时、分、秒、毫秒、微秒、纳秒)。
Format	输出时间格式	string	"YYYY-MM-DD"	否	可选取值: "YYYY-MM-DD" "YYYY/MM/DD" "YYYY-MM" "YYYY/MM"

FrequencyHistogramFile	频率直方图文件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值以及其频率</p> <p>第一列中的取值必须是 Date 类型数据</p> <p>eg:</p> <p>value,Frequency</p> <p>2021-11-22,2</p> <p>2021-11-23,2</p> <p>2021-11-24,6</p>
EqualHeightHistogramFile	等高直方图文件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值，且必须是 Date 类型数据</p> <p>第二列的值必须大于或等于第一列的值</p> <p>第三列是其频率</p> <p>eg:</p> <p>LowerValue,UpperValue,Frequency</p> <p>2021-01-01,2021-02-01,2</p> <p>2021-02-01,2021-03-01,1</p>

不同的数据分布对应的可选参数

数据分布	可选参数	造数原理
Random	Min、Max、Format	在取值范围[min,max)内，随机生成指定时间格式的 Date 数据。
Sequence	Min、Max、Step、Format	在取值范围[min,max)内，按照步长 Step，生成指定时间格式的 Date 数据。

Custom_FrequencyHistogram	FrequencyHistogramFile Format	、 基于 FrequencyHistogramFile 文件中给出的 Date 类型数据的频率直方图信息，进行造数。
Custom_EqualHeightHistogram	EqualHeightHistogramFile Format	、 基于 EqualHeightHistogramFile 文件中给出的 Date 类型数据的等高直方图信息，进行造数。

Time

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Min	最小值	time.Time	00:00:00	否	
Max	最大值	time.Time	23:59:59	否	Min.UnixNano() <= Max.UnixNano() (时间 Min 在时间 Max 之前)
Distribution	数据分布类型	string	"Random"	否	可选取值： "Random" "Sequence" "Custom_FrequencyHistogram" "Custom_EqualHeightHistogram"
Step	步长	string	"1s"	否	Max - Min >= Step step 的时间单位有：h、m、s、i、o、n(分别对应时、分、秒、毫秒、微秒、纳秒)。
Format	输出时间格式	string	"hh:mm:ss"	否	可选取值： "hh:mm:ss" "hh:mm:ss.s" "hh:mm:ss.sss" "hh:mm:ss.ssssss"

FrequencyHistogramFile	频率直方图文件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值及其频率</p> <p>第一列中的取值必须是 Time 类型数据</p> <p>eg:</p> <p>value,Frequency</p> <p>01:01:01,2</p> <p>08:08:08,2</p> <p>12:12:12,6</p>
EqualHeightHistogramFile	等高直方图文件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值，且必须是 Time 类型数据</p> <p>第二列的值必须大于或等于第一列的值</p> <p>第三列是其频率</p> <p>eg:</p> <p>LowerValue,UpperValue,Frequency</p> <p>01:01:01,05:05:05,2</p> <p>08:08:08,12:12:12,2</p>

不同的数据分布对应的可选参数

数据分布	可选参数	造数原理
Random	Min、Max、Format	在取值范围[min,max)内，随机生成指定时间格式的 Time 数据。
Sequence	Min、Max、Step、Format	在取值范围[min,max)内，按照步长 Step，生成指定时间格式的 Time 数据。

Custom_FrequencyHistogram	FrequencyHistogramFile Format	、 基于 FrequencyHistogramFile 文件中给出的 Time 类型数据的频率直方图信息，进行造数。
Custom_EqualHeightHistogram	EqualHeightHistogramFile Format	、 基于 EqualHeightHistogramFile 文件中给出的 Time 类型数据的等高直方图信息，进行造数。

Date&Time

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Min	最小值	time.Time	time.Now() - 1Month eg : 2022-04-24 12:00:00	否	Min >= 1678-01-01 00:00:00
Max	最大值	time.Time	time.Now() eg : 2022-04-24 12:00:00	否	Max <= 2261-12-31 23:59:59 Min.UnixNano() <= Max.UnixNano() (时间 Min 在时间 Max 之前)
Distribution	数据分布类型	string	"Random"	否	可选取值： "Random" "Sequence" "Custom_FrequencyHistogram" "Custom_EqualHeightHistogram"
Step	步长	string	"1D"	否	Max - Min >= Step step 的时间单位有：Y、M、D、h、m、s、i、o、n(分别对应年、月、日、时、分、秒、毫秒、微秒、纳秒)。

Format	输出时间格式	string	"YYYY-MM-DD hh:mm:ss"	否	<p>可选取值：</p> <p>"YYYY-MM-DD hh:mm:ss"</p> <p>"YYYY/MM/DD hh:mm:ss"</p> <p>"YYYY-MM-DD hh:mm:ss.s"</p> <p>"YYYY/MM/DD hh:mm:ss.s"</p> <p>"YYYY-MM-DD hh:mm:ss.sss"</p> <p>"YYYY/MM/DD hh:mm:ss.sss"</p> <p>"YYYY-MM-DD hh:mm:ss.ssssss"</p> <p>"YYYY/MM/DD hh:mm:ss.ssssss"</p>
FrequencyHistogramFile	频率直方图文件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值以及其频率</p> <p>第一列中的取值必须是 Date&Time 类型数据</p> <p>eg:</p> <p>value,Frequency</p> <p>2021-11-22 00:00:00,2</p> <p>2021-11-23 00:00:00,2</p> <p>2021-11-24 00:00:00,6</p>

EqualHeightHistogramFile	等高直 方图文 件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值，且必须是 Date&Time 类型数据</p> <p>第二列的值必须大于或等于第一列的值</p> <p>第三列是其频率</p> <p>eg:</p> <p>LowerValue,UpperValue,Frequency</p> <p>2021-01-01 00:00:00,2021-02-01 00:00:00,2</p> <p>2021-02-01 00:00:00,2021-03-01 00:00:00,1</p>
--------------------------	-----------------	--------	----	---	--

不同的数据分布对应的可选参数

数据分布	可选参数	造数原理
Random	Min、Max、Format	在取值范围[min,max)内，随机生成指定时间格式的 Date&Time 数据。
Sequence	Min、Max、Step、Format	在取值范围[min,max)内，按照步长 Step，生成指定时间格式的 Date&Time 数据。
Custom_FrequencyHistogram	FrequencyHistogramFile 、Format	基于 FrequencyHistogramFile 文件中给出的 Date&Time 类型数据的频率直方图信息，进行造数。
Custom_EqualHeightHistogram	EqualHeightHistogramFile 、Format	基于 EqualHeightHistogramFile 文件中给出的 Date&Time 类型数据的等高直方图信息，进行造数。

Char

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
------	----	------	-----	--------	------

Format	生成数据的格式	string	"##-@@-^^-**"	否	<p>按照指定的数据格式，随机生成固定字符数量的字符串。</p> <p>使用“#”表示随机数字(0~9)。</p> <p>使用“@”表示随机小写字母(a~z)。</p> <p>使用“^”表示随机大写字母(A~Z)。</p> <p>使用“*”表示随机数字或字母(0~9、a~z、A~Z)。</p> <p>任何其他字符（包括中文）将被逐字包含。</p> <p>eg: "##-@@-^^-** " => "99-fp-CD-St"</p>
Length	生成数据的字符数量	int	11	否	<p>Length > 0</p> <p>备注：</p> <p>1、如果按照 Format 生成的数据的字符数量小于 Length，则自动按照"*"类型的数据补充。</p> <p>2、生成字符的数量 Length 不包括参数 Prefix 和 Postfix 字符数量。</p>

Varchar

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Format	生成数据的格式	string	"##-@@-^^-**"	否	<p>按照相应的数据格式，随机生成可变字符数量的字符串。</p> <p>使用“#”表示随机数字(0~9)。</p> <p>使用“@”表示随机小写字母(a~z)。</p> <p>使用“^”表示随机大写字母(A~Z)。</p> <p>使用“*”表示随机数字或字母(0~9、a~z、A~Z)。</p> <p>任何其他字符（包括中文）将被逐字包含。</p> <p>eg: "##-@@-^^-** " => "99-fp-CD-St"</p>
MinLength	生成数据的最小字符数量	int	2	否	MinLength >= 0

MaxLength	生成数据的最大字符数量	int	11	否	MaxLength >= MinLength 备注： 1、如果按照 Format 生成的数据的字符数量小于实际 Length，则自动按照 "*" 类型的数据补充。 2、生成字符的数量 Length 不包括参数 Prefix 和 Postfix 字符数量。
-----------	-------------	-----	----	---	--

List

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Values	用户自定义字符串集合	string	""	是	
Separator	不同字符串之间的分隔符(用于分割参数 Values 中的字符串集合)	string	","	否	
ListFile	文件名，引用 ListFile 文件中的字符串集合(每一行为一个字符串)	string	""	是	文件格式为 csv、txt，文件大小限制为 100M 以下
IsPrimaryKey	是否为主键	bool	false	否	
Distribution	数据分布类型	string	"Random"	否	可选取值： "Random" "Sequence" "Custom_FrequencyHistogram"

FrequencyHistogramFile	频率直方图文件	string	""	否	<p>文件格式为 csv、大小限制在 100M 以下</p> <p>文件的内容如下：</p> <p>第一行为表头</p> <p>第一列和第二列分别是取值以及其频率</p> <p>第一列中的取值可以是任意字符串</p> <p>eg:</p> <p>value,Frequency</p> <p>str1,2</p> <p>str2,2</p> <p>str3,6</p>
------------------------	---------	--------	----	---	--

不同的数据分布对应的可选参数

数据分布	可选参数
Random	Values、Separator、ListFile
Sequence	Values、Separator、ListFile、IsPrimaryKey
Custom_FrequencyHistogram	FrequencyHistogramFile

Mix

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Connector	连接符	string	""	是	
SubFields	子字段	[]FieldModel	默认有一条 Int 类型字段	否	除了 Mix 类型，子字段的数据类型可以其他任意类型

Birthday

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Min	最小值	time.Time	1970-01-01	否	Min >= 1937-01-01

Max	最大值	time.Time	2006-01-01	否	Max <= time.Now() Min.UnixNano() <= Max.UnixNano() (时间 Min 在时间 Max 之前)
-----	-----	-----------	------------	---	--

UUID

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
Format	UUID 的版本	string	"Version4"	否	可选取值： "Version1" "Version2" "Version3" "Version4" "Version5"

4 数据生成参数

DK 的数据生成参数如下表所示。

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
TotalRows	生成数据的总行数	Int64	1000	否	$0 < \text{TotalRows} \leq (1 \ll 63 - 1)$
BitchSize	批量生成数据的行数	int	1000	否	$100 \leq \text{BitchSize} \leq 100000$
NumGenWorker	生成数据的并发数	Int	1	否	$0 < \text{NumGenWorker} \leq 100$

5 数据输出参数

DK 支持 3 种数据文件输出格式，相关参数的详细描述信息如下。

文件格式：CSV

提供两种编码格式：默认格式为 **UTF-8**；若参数 `AddBOM = true`，则编码格式为 **UTF-8-BOM**。

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
FileName	输出文件名	string	"dataKaleido"	否	
Quote	字段引用符号。 输出的每个字段数据使用该符号引用起来。	string	"	是	
AddHeader	是否在第一行输出字段名称	bool	false	是	
AddBOM	是否添加 BOM	bool	false	是	
NullValue	空字段的值	string	"	是	

文件格式：SQL

编码格式为 UTF-8。

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
FileName	输出文件名	string	"dataKaleido"	否	
TableName	表名	string	"dataTable"	否	
NullValue	空字段的值	string	"	是	

备注：为了避免插入数据库时 insert 语句事务过大，每生成一个 BatchSize 数据，会加一行 "commit;" 命令。

文件格式：JSON

编码格式为 UTF-8

参数名称	描述	数据类型	默认值	是否可以为空	取值范围
FileName	输出文件名	string	"dataKaleido"	否	