

Tableau Prep 帮助

版本 当前;上次更新:2020/8/11

Copyright © 2020 Tableau Software®。法律和隐私



目录

Tableau Prep Builder 新增功能	1
版本 2020.2.3 新增功能	1
连接到数据	1
使用 Snowflake 连接到数据时使用 OAuth 身份验证	1
版本 2020.2.2 新增功能	2
连接到数据	2
使用 Tableau Server 或 Tableau Online 连接到数据	2
使用 Alibaba MaxCompute 连接到数据	3
版本 2020.2.1 新增功能	4
连接到数据	4
连接到 Salesforce.com 数据	4
使流程数据保持最新	4
使用增量刷新仅处理最新数据	4
版本 2020.1.4 新增功能	7
连接到数据	7
使用 Impala 连接到数据	7
清理和调整数据	7
只需单击一下即可转换日期值	7
保存和共享工作	8
保持与服务器的连接	8
版本 2020.1.3 新增功能	9
安装和部署 Tableau Prep Builder	9
支持基于登录名的许可证管理	9
连接到数据	9

使用初始 SQL 将参数传递到数据源	9
清理和调整数据	9
使用详细级别 (LOD)、排名和行号计算来清理数据	9
版本 2020.1.1 新增功能	11
连接到数据	11
将流程输入映射回其原始连接	11
清理和调整数据	12
暂停数据更新以提高性能	12
与 Tableau Prep 的版本兼容性	13
检测并启用不兼容的功能	13
使流程数据保持最新	15
发布流程时为云连接器添加或嵌入凭据	15
版本 2019.4.2 新增功能	17
连接到数据	17
使用新的连接器连接到更多数据	17
清理和调整数据	17
将列转置为行变得更容易	17
其他增强功能	19
查看有关长时间运行的操作的反馈	19
支持基于登录名的许可证管理	19
版本 2019.4.1 新增功能	19
连接到数据	19
连接到云中的更多数据	20
其他增强功能	20
与 Tableau 共享使用情况数据	20
不再为 MacOS 自动安装默认数据库驱动程序	21

版本 2019.3.2 新增功能	21
构建和组织流程	21
只需单击一下，即可在所需的位置插入步骤	21
使用流程导航器轻松查看大型流程	22
检查和筛选数据	22
使用自定义数据角色对类似值进行分组	23
清理和调整数据	24
使用新列表视图对字段应用清理操作	24
通过可重用的步骤来利用您的工作和其他人的工作	25
保存和共享工作	26
从命令行运行流程时禁用 Mac OS 的钥匙串对话框	26
版本 2019.3.1 新增功能	27
连接到数据	27
使用 Google BigQuery 连接到数据	27
为流程输入连接到发布的数据源	27
检查和筛选数据	29
创建自定义数据角色	29
清理和调整数据	31
在流程中使用 R 和 Python 脚本	32
保存和共享工作	32
在构建流程时找出不兼容的功能	32
使流程数据保持最新	33
Tableau 数据管理加载项包括 Tableau Prep Conductor	33
将流程发布到 Tableau Online	33
其他增强功能	34
Tableau Prep Builder 现在提供意大利语版本	34

版本 2019.2.3 新增功能	35
检查和筛选数据	35
选择要筛选的特定值	35
通过单击复制字段	36
清理和调整数据	37
使用模糊匹配和清理建议修复无效的数据角色值	38
版本 2019.2.2 新增功能	40
连接到数据	40
使用其他数据库连接器 (ODBC) 连接到更多数据	40
在连接到数据时指定初始 SQL 命令	40
检查和筛选数据	41
在清理数据时仅保留所需的字段	41
清理和调整数据	41
重新排列流程中的步骤	41
版本 2019.2.1 新增功能	43
连接到数据	43
使用新的数据库连接器连接到更多数据	43
使流程数据保持最新	43
发布流程时查看文件安全列表位置	43
其他增强功能	44
为用户界面选择 10 种不同的语言	44
使用流程缩略图视图在 Tableau Server 上快速查找流程	45
版本 2019.1.4 新增功能	46
连接到数据	46
使用新的数据库连接器连接到更多数据	46
清理和调整数据	46

按数据角色对字段中的类似值进行分组	46
在对字段值进行分组时调整结果	48
版本 2019.1.3 新增功能	49
连接到数据	49
基于表关系快速联接数据	49
清理和调整数据	50
在工具栏中查看所有清理建议	50
其他增强功能	51
查看流程中的特定更改	51
通过“开始”页面管理最近查看的流程	52
版本 2019.1.2 新增功能	53
使流程数据保持最新	53
将流程计划为在 Tableau Server 中自动运行	53
帮助设计	53
版本 2019.1.1 新增功能	54
清理和调整数据	54
将行转置为列	54
使用通配符搜索来转置字段	54
使用说明为更改添加上下文	56
使用清理建议来快速拆分固定宽度文本文件中的数据	57
版本 2018.3.3 新增功能	58
连接到数据	58
添加新文件时自动刷新输入并集	58
清理和调整数据	58
重用清理操作	58
移除额外的空格	59

检查和筛选数据	59
快速查找字段在流程中的使用之处	59
其他增强功能	60
同时打开多个 Tableau Prep Builder 窗口	60
安装之后关闭使用情况报告	60
版本 2018.3.2 新增功能	61
清理和调整数据	61
使用清理建议快速提高数据质量	61
其他增强功能	62
通过命令行运行流程时更加灵活	62
性能改进	63
早期版本新增功能	63
安装和部署 Tableau Prep	63
更改默认存储库位置	64
连接到数据	64
清理存储在 PDF 文件中的数据	64
适用于数据库表的通配符并集	64
清理和调整数据	65
将字段值替换为 Null	66
在任何步骤中应用清理操作	66
连接到数据	68
连接到存储在 MongoDB Business Intelligence (BI) 中的数据	68
检查和筛选数据	68
为字段值指定数据角色	68
联接或合并数据	70
在联接子句中直接修复不匹配的字段	70

安装和部署 Tableau Prep	72
设置显示语言	72
连接到数据	73
连接到 Microsoft Access 中存储的数据	73
清理和调整数据	73
更改流程步骤的配色方案	73
为步骤添加说明	74
使用模糊匹配来查找并修复拼写错误	75
保存和共享工作	75
从命令行中运行流程	75
安装和部署 Tableau Prep	76
通过命令行停用 Tableau Prep	77
使用虚拟桌面支持来优化 Tableau Prep 安装	77
连接到数据	77
连接到云数据源和 Hadoop Hive	77
浏览数据	78
用于只保留所需数据的新筛选器选项	78
清理和调整数据	78
在计算字段中使用 ISO-8601 日期标准	78
在数据网格中应用清理操作	79
使用多选对“配置”窗格中的值进行分组	79
其他增强功能	80
联接或合并数据	80
使用并集建议来清理不匹配的字段	81
为所有联接类型确定不匹配的字段	82
安装和部署 Tableau Prep	83

通过命令行激活和注册 Tableau Prep (Windows)	83
连接到数据	84
连接到存储在统计文件中或 Presto 上的数据	84
在输入步骤中合并 Data Interpreter 找到的子表	84
加载表时更好的反馈	85
浏览数据	86
对“配置”窗格和“数据”网格中的字段进行重新排序	86
清理和调整数据	86
使用拖动选择功能来移除流程中的多个步骤	86
在一个操作中对多个字段组进行转置	87
合并字段时改善的字段命名	87
其他增强功能	88
Tableau Prep Builder 入门指南	89
示例文件	89
背景故事...	90
1. 连接到数据	90
检查您的工作：观看“连接到数据”的实际运行效果。	93
2. 浏览数据	94
3. 清理数据	95
清理 Orders_Central	96
查看您的更改	101
检查您的工作：观看“清理 Orders_Central”的实际运行效果。	102
清理 Orders_East	103
清理 Orders_West	105
4. 合并数据	109
合并数据	109

检查您的工作: 观看“合并数据”的实际运行效果。	112
清理产品退货数据	112
联接数据	117
清理联接结果	122
5.运行流程并生成输出	125
总结和资源	127
关于 Tableau Prep Builder	127
使用 Tableau Prep Builder	128
查看 Tableau Prep Builder 的实际运行效果	129
	130
Tableau Prep Builder 工作区一览	130
“连接”窗格	131
“流程”窗格	132
“配置”窗格	133
“数据”网格	134
Tableau Prep Builder 如何存储数据	135
Tableau Prep 视觉词典	136
连接到数据	141
启动或打开流程	141
开始新流程	142
打开现有流程	144
连接到发布的数据源(版本 2019.3.1 及更高版本)	144
连接到云数据源	150
连接到 Salesforce 数据(版本 2020.2.1 及更高版本)	151
连接到 Google BigQuery(版本 2019.3.1 及更高版本)	151
配置 SSL 以连接到 Google BigQuery(仅限 MacOS)	151

设置和管理您的 Google BigQuery 凭据	152
连接到 SAP HANA(版本 2019.2.1 及更高版本)	153
使用 ODBC 进行连接(版本 2019.2.2 及更高版本)	155
连接之前	155
使用“其他数据库 (ODBC)”进行连接	157
使用初始 SQL 查询连接(版本 2019.2.2 及更高版本)	159
运行初始 SQL	159
在初始 SQL 语句中包括参数(版本 2020.1.3 及更高版本)	160
连接到 Tableau 数据提取	161
使用 Data Interpreter 清理您的文件	161
配置数据集	162
连接到自定义 SQL 查询	163
在输入步骤中应用清理操作	164
选择要包括在流程中的字段	164
将筛选器应用于输入步骤中的字段	164
更改字段名称	165
更改数据类型	165
配置字段属性	166
配置文本文件中的文本设置	166
选择数据样本大小	167
在“输入”步骤中添加更多数据	168
在“输入”步骤中刷新数据	168
文件输入步骤类型	168
文件、数据库或 Tableau 数据提取输入步骤类型	168
在输入步骤中合并文件和数据库表	170
合并文件	171

合并数据库表(版本 2018.3.1 及更高版本)	174
在并集之后合并字段	176
在“输入”步骤中联接数据(版本 2019.1.3 及更高版本)	176
构建和组织流程	179
添加或插入步骤	179
添加步骤	179
插入步骤	181
更改流程配色方案	185
为流程步骤和清理动作添加说明	186
为流程步骤添加说明	186
为更改条目添加说明(版本 2019.1.1 及更高版本)	187
重新组织流程的布局(版本 2019.2.2 及更高版本)	190
使用流程导航工具	191
从流程中移除步骤	192
检查您的数据	194
查看分配给数据的数据类型	194
查看有关数据的大小详细信息	195
查看值或唯一值的分布	196
搜索字段和值	198
对值和字段进行排序	200
突出显示流程中的字段和值	201
在流程中跟踪字段	201
查看相关值	201
突出显示相同的值	202
筛选您的数据	202
保留或移除字段	203

可用于每种数据类型的筛选器	204
筛选器选项位于何处？	204
计算筛选器	205
“选定值”筛选器(版本 2019.2.3 及更高版本)	206
“值范围”筛选器	206
“日期范围”筛选器	207
“相对日期”筛选器	207
“通配符匹配”筛选器	208
“Null 值”筛选器	208
使用数据角色验证数据	209
为数据分配标准数据角色	209
创建自定义数据角色(版本 2019.3.1 及更高版本)	212
创建自定义数据角色	213
应用自定义数据角色	216
查看和管理自定义数据角色	218
按数据角色对类似值进行分组	219
清理和调整数据	223
关于清理操作	223
可用的清理操作	223
操作顺序	224
应用清理操作	226
选择视图	227
暂停数据更新以提高性能	229
应用清理操作	231
查看您所做的更改	232
合并字段	234

使用建议应用清理操作(版本 2018.3.2 及更高版本)	236
应用建议	237
编辑字段值	239
编辑单个值	239
编辑多个值	240
使用快速清理操作编辑多个值	240
以内联方式编辑分组或多个值	242
将一个或多个值替换为 Null	242
将多个值手动映射到标准值	243
将多个值映射到单个所选字段	243
通过选择多个值来创建组	244
添加和标识不在数据集中的值	245
使用模糊匹配将值自动映射到标准值	248
使用模糊匹配分组类似值	249
在对字段值进行分组时调整结果	250
复制步骤、操作和字段	252
复制并粘贴步骤	252
复制并粘贴清理操作	254
复制字段(版本 2019.2.3 及更高版本)	256
创建可重用的流程步骤(版本 2019.3.2 及更高版本)	258
创建可重用的步骤	258
在流程中插入可重用的步骤	259
创建详细级别和排名计算	261
计算详细级别	261
创建详细级别 (LOD) 计算	262
计算编辑器	262

可视化计算编辑器	264
计算排名或行号	265
支持的分析函数	266
创建排名或行号计算	267
计算编辑器	268
可视化计算编辑器	271
转置您的数据	273
将列转置为行	274
观看“对多个字段进行转置”的实际运行效果。	276
使用通配符搜索进行转置(版本 2019.1.1 及更高版本)	277
将行转置为列(版本 2019.1.1 及更高版本)	279
在流程中使用 R 和 Python 脚本	281
在您的流程中使用 R (Rserve) 脚本	282
先决条件	282
资源	283
针对 Tableau Server 配置 Rserve Server	283
其他 Rserve 配置(可选)	283
创建 R 脚本	284
向流程中添加脚本	286
在您的流程中使用 Python 脚本	289
先决条件	290
为 Tableau Server 配置 Tableau Python (TabPy) 服务器	290
创建 Python 脚本	291
向流程中添加脚本	293
聚合、联接或合并数据	297
对值进行聚合和分组	297

联接数据	298
检查联接的结果	300
常见联接问题	301
修复不匹配的字段以及更多操作	302
合并数据	303
检查并集的结果	304
修复不匹配的字段	306
其他合并字段选项	308
保存和共享工作	309
保存您的流程	309
在 Tableau 中查看数据样本	309
创建和发布数据提取及数据源	310
创建数据提取文件	311
以数据源形式发布	313
使用增量刷新来刷新流程数据	314
流程刷新选项	314
配置增量刷新	315
配置写入选项	317
运行流程	320
通过命令行刷新流程输出文件	321
运行流程之前	322
凭据 .json 文件要求	322
凭据 .json 文件示例	323
连接到已发布数据源	323
连接到两个数据库	324
创建凭据文件的相关提示	325

运行流程	326
在启用增量刷新的情况下运行流程(版本 2020.2.1 及更高版本)	327
命令选项	328
语法示例	330
流程连接到并发布至本地文件	331
流程连接并发布到本地文件，并为增量刷新使用短形式	331
流程连接到数据库并发布至服务器	332
流程发布至服务器，并且凭据文件存储在网络共享上	332
与 Tableau Prep 的版本兼容性	333
查找我们的版本	333
Tableau Prep Builder 不同版本之间的兼容性	336
修复 Tableau Prep Builder 的兼容性问题	336
不同版本的 Tableau Prep Builder 和 Tableau Server 之间的兼容性	337
检测不兼容的功能	337
Tableau Prep Builder(版本 2020.1.1 及更高版本)	337
Tableau Prep Builder(版本 2019.3.1 及更高版本)	339
Tableau Prep Builder(所有版本)	339
修复兼容性问题	341
确定不兼容的功能	341
从流程中移除不兼容的功能	342
不兼容的数据源	343
不兼容的功能	343
使流程数据保持最新	345
关于 Tableau Prep Conductor	346
启用 Tableau Prep Conductor	347
了解流程工作区	348

“流程概述”页面	348
“流程连接”页面	349
“流程计划任务”页面	349
“计划”页面	350
流程运行历史记录	351
流程修订历史记录	351
谁可以执行此操作	352
许可 Tableau Prep Conductor	352
在 Tableau Server 上启用和配置 Tableau Prep Conductor	352
服务器拓扑	353
后续步骤:	354
谁可以执行此操作	354
步骤 1(新安装): 安装包含 Tableau Prep Conductor 的 Tableau Server	354
安装之前	355
安装 Tableau Server 并启用 Tableau Prep Conductor	355
配置公共网关设置	355
启用 Tableau Prep Conductor	355
验证 Tableau Prep Conductor 是否已启用并正在运行	356
将节点专用于 Tableau Prep Conductor	358
多节点安装	358
后续步骤	358
谁可以执行此操作	358
步骤 1(现有安装): 启用 Tableau Prep Conductor	359
升级之前	359
准备升级:	359
配置公共网关设置	359

使用基于用户的许可证的 Tableau Server 安装	359
Tableau Server 单节点安装	360
Tableau Server 多节点安装	362
使用基于内核的许可证的 Tableau Server 安装	364
Tableau Server 单节点安装	365
Tableau Server 多节点安装	367
后续步骤	369
谁可以执行此操作	369
步骤 2: 配置 Tableau Server 的流程设置	369
发布、计划和凭据设置	369
禁用 Tableau Prep Conductor 的含义	371
为流程失败配置通知	371
启用服务器范围电子邮件通知	371
设置通知值	372
为站点配置电子邮件通知:	373
后续步骤	373
谁可以执行此操作	373
步骤 3: 为流程任务创建计划	373
创建新计划:	373
后续步骤	374
谁可以执行此操作	374
步骤 4: 安全列表输入和输出位置	375
如何将输入和输出位置放入安全列表	375
后续步骤	376
谁可以执行此操作	376
步骤 5: 可选服务器配置	377

设置流程的超时期间	377
为挂起的流程任务设置阈值	377
谁可以执行此操作	377
在 Tableau Online 站点上启用 Tableau Prep Conductor	377
步骤 1: 安装 Tableau Prep Builder	378
步骤 2: 登录到 Tableau Online	378
将流程发布到 Tableau Server 或 Tableau Online	378
发布之前	379
发布流程	383
Tableau Server	383
文件	384
数据库	386
Tableau Online	390
文件	391
数据库	391
谁可以执行此操作	396
计划流程任务	396
计划流程任务	397
谁可以执行此操作	399
管理流程	400
管理流程	400
谁可以执行此操作	401
监视流程运行状况和性能	407
在问题出现时检测问题并加以解决	407
在流程失败时收到电子邮件通知:	407
查看和解决错误	408

“流程概述”页面	408
“连接”页面	409
“计划任务”页面	409
“运行历史记录”页面	410
通知	410
Tableau Prep Conductor 进程状态	411
使用管理视图监视流程活动和性能	412
谁可以执行此操作？	412
所有用户的操作	413
特定用户的操作	413
最近用户的操作	414
后台程序任务延迟	414
非数据提取的后台任务	415
流程运行的性能	416
空间使用情况统计信息	416
谁可以执行此操作	417
开发人员资源 - REST API	418
日常生活场景	419
使用 Tableau Prep 分析医院床位使用情况	419
数据	420
初步分析	420
所需的数据结构	421
重构数据	423
床位小时矩阵	423
患者床位使用	425
Tableau Desktop 中的分析	429

总结和资源	432
使用 Tableau Prep 查找第二个日期	433
数据	434
所需的数据结构	435
重构数据	435
第一次违章日期的初始聚合	435
第二次违章日期的第二个聚合	439
为第一次和第二次违章创建完整数据集	440
创建完整数据集	443
总结	444
继续阅读 Tableau Desktop 中包含第二个日期的分析 on page 1。	445
Tableau Desktop 中包含第二个日期的分析	445
Tableau Desktop 中的分析	446
更进一步 — 转置的数据	453
转置数据的优点	461
再进一步 — 只使用计算	462
对各种方法的反思	467
Driver Infractions(司机违章)	468
Pivoted Driver Infractions(转置的司机违章)	468
LOD Driver Infractions(LOD 司机违章)	469
Tableau Prep Builder 疑难解答	471
使用命令行运行流程时的常见错误	471
错误 : “These features were found that prevent this version of the application from using this file”(发现了使此版本的应用程序无法使用此文件的这些功能)	474
发布 Tableau Prep 输出时出现“Failed to parse response from Tableau Server”(未能解析来自 Tableau Server 的响应) 错误	475

使用 Tableau Prep 登录到启用 SSL 的 Tableau Server 时出现“You are using Server version: null...”(您正使用 Server 版本: null...) 错误	476
维护 Tableau Desktop 和 Tableau Prep 的许可证	476
查看有关许可证的数据	477
跟踪 Tableau Desktop 许可证使用情况和过期数据	478
其他资源	479

Tableau Prep Builder 新增功能

浏览新功能摘要来了解当前支持的版本。

注意: 从版本 2019.1.2 开始, Tableau Prep 现在称为 Tableau Prep Builder。

版本 2020.2.3 新增功能

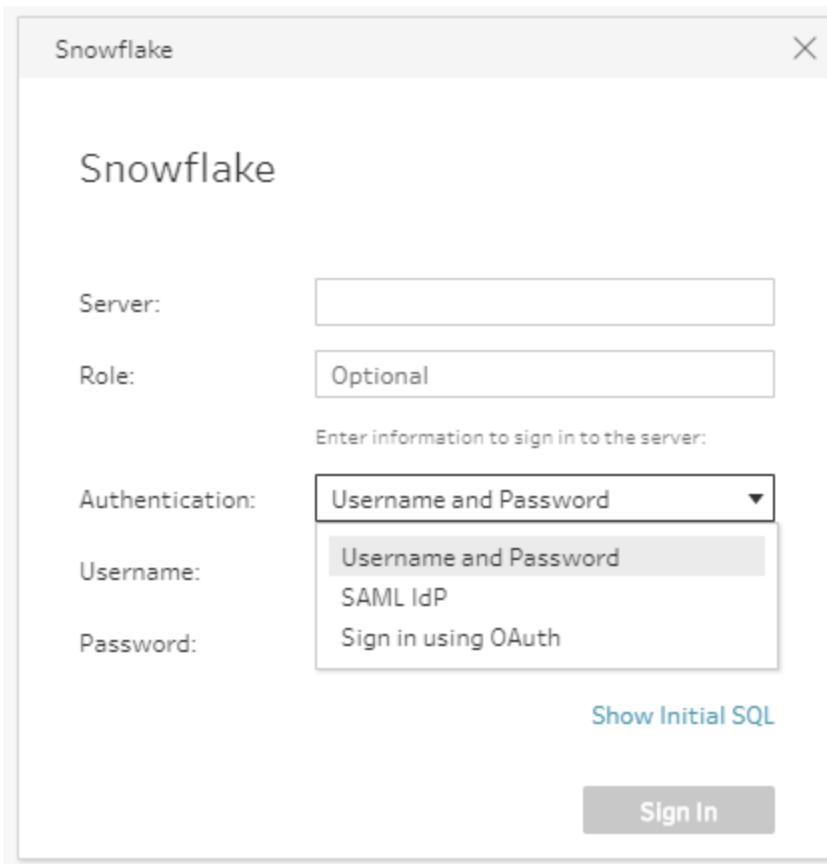
连接到数据

使用 Snowflake 连接到数据时使用 OAuth 身份验证

现在, 当您使用 Snowflake 连接到数据时, 可以使用以下任何身份验证方法, 就像 Tableau Desktop 一样:

- 用户名和密码
- SAML IdP

- OAuth



有关使用此连接器连接到数据源的详细信息，请参见 Tableau Desktop 帮助中的 [Snowflake](#)。

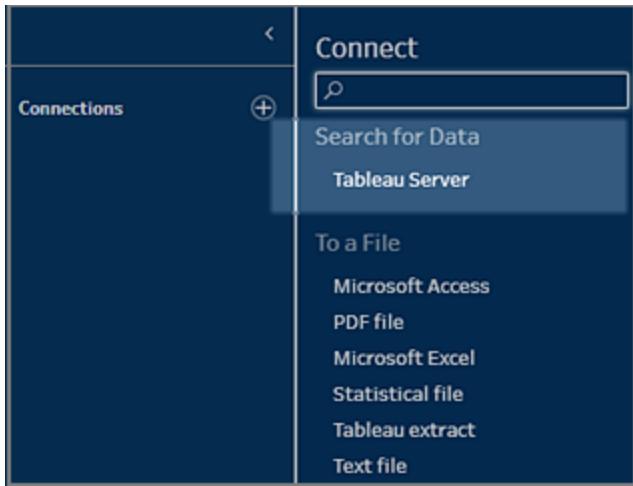
版本 2020.2.2 新增功能

连接到数据

使用 Tableau Server 或 Tableau Online 连接到数据

在以前的版本中，您可以连接到存储在 Tableau Server 或 Tableau Online 上的已发布数据源。现在，如果您有启用了 Tableau Catalog 的[数据管理加载项](#)，则还可以搜索并连接到存储在 Tableau Server 或 Tableau Online 中的数据库和表。

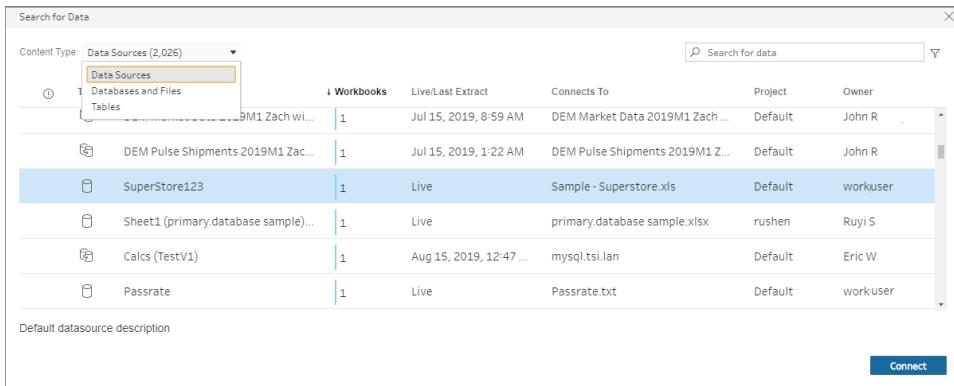
Tableau Prep 帮助



只需在“连接”窗格新的“**搜索数据**”部分选择**“Tableau Server”**并登录到服务器即可。“**搜索数据**”对话框将打开，您可以搜索并从精选的数据源中进行选择。

其他元数据(例如数据源说明、数据质量警告和认证)可帮助您找到要查找的数据源，而无需离开 Tableau Prep Builder。您还可以使用右上角的筛选器选项按连接类型、数据质量警告和认证来筛选结果。

如果没有启用了 Tableau Catalog 的数据管理加载项，则只会显示已发布的数据源，您将无法查看元数据。



有关连接到数据的详细信息，请参见[连接到发布的数据源\(版本 2019.3.1 及更高版本\)](#)在本页 144。

使用 Alibaba MaxCompute 连接到数据

如果数据存储在数据仓库中，您现在可以使用 Alibaba MaxCompute 连接到此数据源。有关使用此连接器连接到数据源的详细信息，请参见 Tableau Desktop 帮助中的 [Alibaba](#)

MaxCompute。

版本 2020.2.1 新增功能

连接到数据

连接到 Salesforce.com 数据

Tableau Prep Builder 现在可以连接到 Salesforce 数据，以便您可以清理并准备数据以便在 Tableau 中进行分析。目前不支持使用自定义 SOQL 和创建标准连接。

有关详细信息，请参见[连接到 Salesforce 数据\(版本 2020.2.1 及更高版本\)](#) 在本页 151。

使流程数据保持最新

使用增量刷新仅处理最新数据

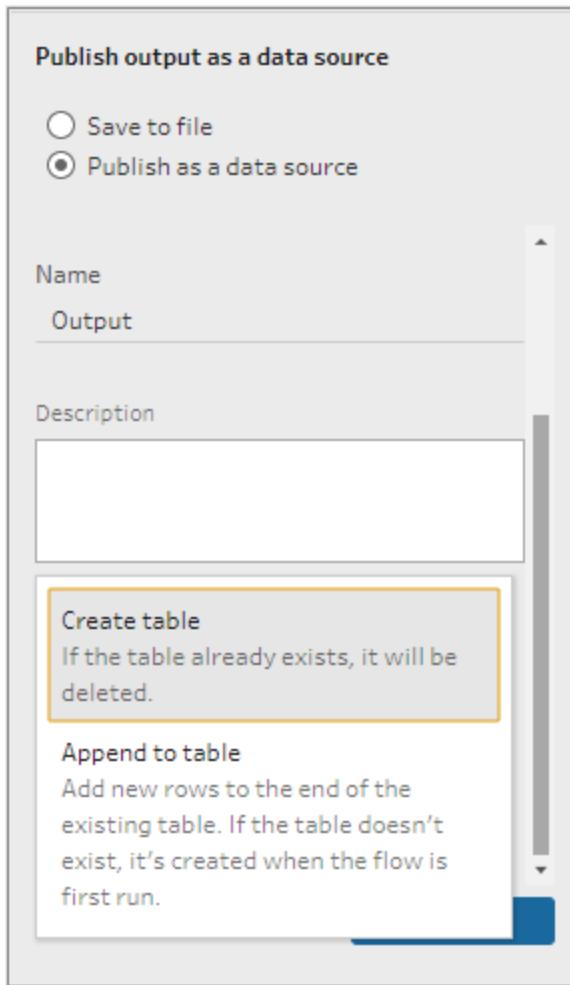
如果流程数据不断变化，现在可以通过使用增量刷新来仅检索和处理自流程上次运行以来的最新数据行，从而节省时间和资源。只需为流程配置增量刷新参数，指示 Tableau Prep Builder 如何从输入数据中检测和检索新行。您还可以配置当流程在“输出”步骤中运行时更新输出表的方式。

Tableau Prep 帮助

The screenshot shows the 'Input' settings page in Tableau Prep. At the top, there are tabs: 'Settings' (which is selected), 'Multiple Files', 'Data Sample', and 'Changes (0)'.
Connection: Text file, Orders_Central.csv (Edit), Original Table Name: Orders_Central.
Text Options: Radio buttons for 'First line contains header' (selected) and 'Generate field names automatically'.
Field Separator: Comma.
Text Qualifier: Automatic.
Character Set: A dropdown menu.
Locale: English (United States).

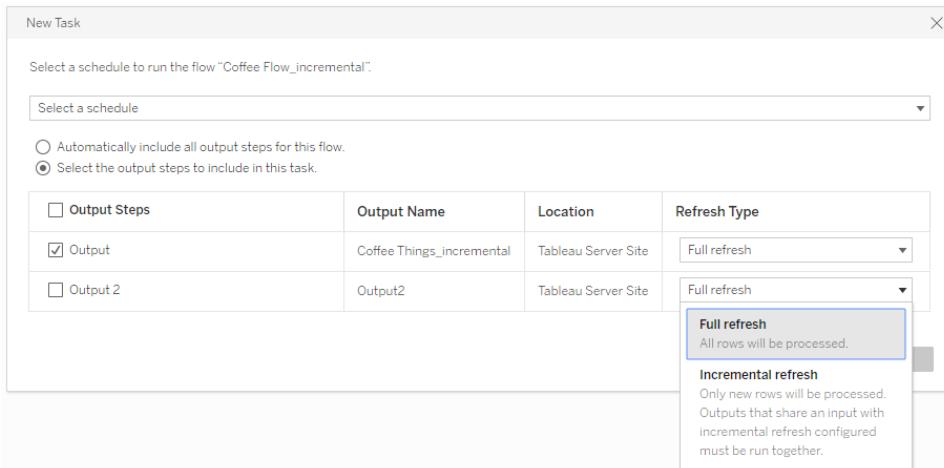
Set up Incremental Refresh: A section with a note: 'Get the latest rows for a specific field value when the values have changed since the flow was last run.' A checked checkbox labeled 'Enable'.
Identify new rows using field: A dropdown menu showing '# Row ID'.
Output: A dropdown menu showing 'Create 'Superstore Sales''.
Select the output and field with the last processed value for field "Row ID". A note explaining the purpose of the 'Row ID' field in the output.
Field name in output: A dropdown menu showing '# Row ID'.

通过“输出”步骤中的新选项，您可以选择将新数据追加到现有表或使用 Tableau Prep Builder 的现有功能，并将表数据替换为新数据。此选项可用于增量刷新和完全刷新类型。



使用 Tableau Prep Conductor 按计划设置和运行流程，或在 Tableau Prep Builder 中或者通过命令行界面以增量方式运行单个流程。

注意: 需要 Tableau Server 版本 2020.2 及更高版本(启用了 Tableau Prep Conductor)才能使用增量刷新来计划流程。



有关使用增量刷新设置和运行流程的详细信息，请参见[使用增量刷新来刷新流程数据 在本页314。](#)

版本 2020.1.4 新增功能

连接到数据

使用 Impala 连接到数据

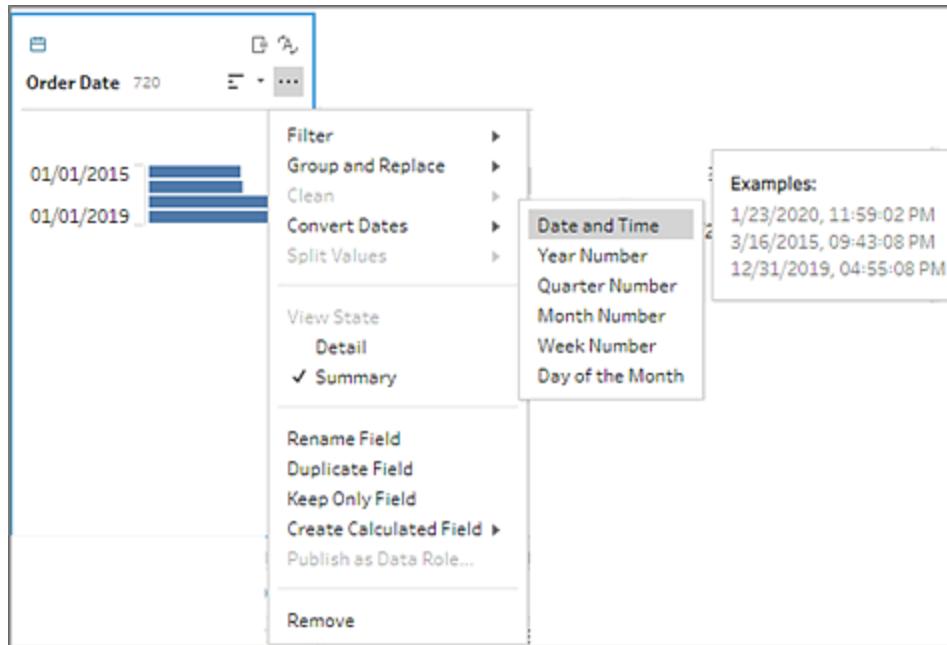
就像 Tableau Desktop 一样，您现在可以使用 Impala 轻松连接到较大的数据集。

有关使用 Impala 连接到数据源的详细信息，请参见 Tableau Desktop 帮助中的 [Impala](#)。

清理和调整数据

只需单击一下即可转换日期值

为了帮助简化数据清理，Tableau Prep Builder 添加了另一个适用于日期的快速清理菜单。如果需要转换日期值以显示 DATEPART(如年份、月份或日)，则可以从新的“[转换日期](#)”菜单中选择以下菜单选项之一，而不是编写复杂的字段计算。

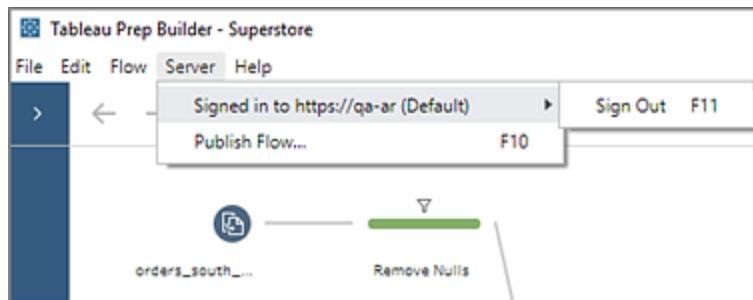


有关对日期应用清理操作的详细信息，请参见[应用清理操作 在本页231](#)。

保存和共享工作

保持与服务器的连接

每次打开 Prep Builder 时，您不再需要登录到 Tableau Server 或 Tableau Online。现在，Tableau Prep Builder 会在关闭应用程序时记住您的服务器名称和登录信息，以便下次打开应用程序时您已登录到服务器。



这个可以节省时间的选项将帮助您无缝连接以发布数据源或插入保存的流程步骤，并且，如果您需要知道某个功能是否与服务器版本兼容，兼容性警告将立即显示，而无需首先登录到服务器。

有关详细信息，请参见[以数据源形式发布 在本页313](#)。

版本 2020.1.3 新增功能

安装和部署 Tableau Prep Builder

支持基于登录名的许可证管理

基于登录名的许可证管理可帮助您为在 Tableau Server 和 Tableau Online 上具有 Creator 角色的用户管理许可。具有 Explorer 或 Viewer(查看者) 角色的用户无法使用此功能。如果将基于角色的订阅与 Tableau Server 或 Tableau Online 结合使用，则可以使用基于登录名的许可证管理来简化许可证管理，从而不必使用单独的 Tableau Desktop 和 Tableau Prep Builder 产品密钥。您只需要管理本地 Tableau Server 的一个或多个产品密钥，或者，如果是 Tableau Online，您根本无需管理任何产品密钥。

连接到数据

使用初始 SQL 将参数传递到数据源

现在，您可以使用初始 SQL 将参数传递到数据源，以提供有关 Tableau Prep 应用程序、版本和流程名称的更多详细信息。Tableau Prep Builder 支持以下参数：

- **TableauApp** - 返回用于访问数据源的应用程序 - Prep Builder 或 Prep Conductor。
- **TableauVersion** - 返回应用程序版本号。对于 Tableau Prep Builder，它将返回确切的版本号，例如 2020.1.1。对于 Tableau Prep Conductor，它将返回已启用 Tableau Prep Conductor 的主要服务器版本，例如 2020.1。
- **FlowName** - 返回 .tfl 文件的名称。

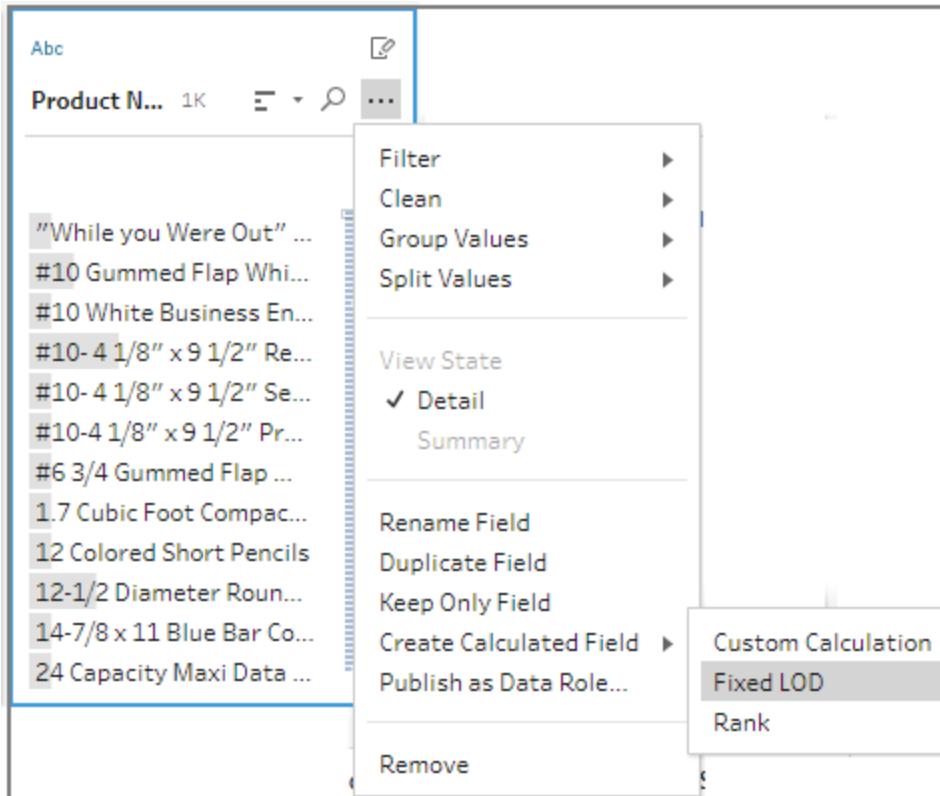
有关在使用初始 SQL 连接到数据源时使用参数的详细信息，请参见[在初始 SQL 语句中包括参数\(版本 2020.1.3 及更高版本\) 在本页 160](#)。

清理和调整数据

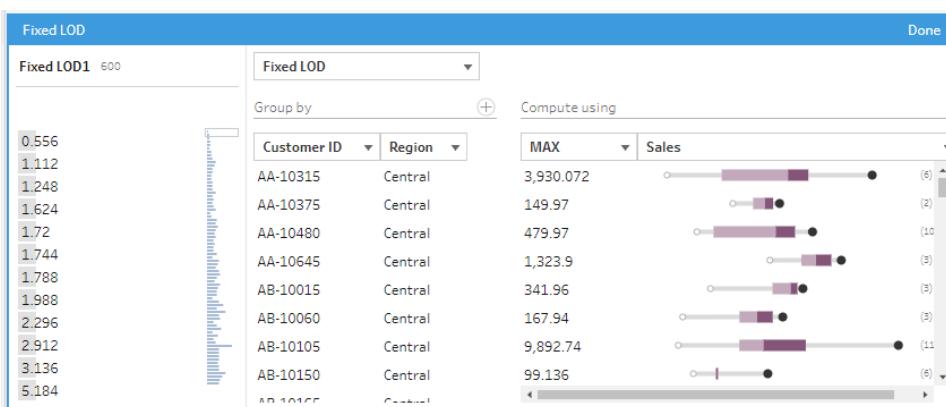
使用详细级别 (LOD)、排名和行号计算来清理数据

如果数据聚合级别与所需级别不同，或者您需要跨一组行执行计算，您现在可以使用 FIXED 详细级别 (LOD) 以及 RANK 和 ROW_NUMBER 分析函数计算来创建数据集中所需的数据。

例如，添加 FIXED LOD 计算以更改表中字段的粒度。使用新的 ROW_NUMBER() 分析函数快速查找重复行，或使用新的 RANK() 函数之一来查找具有类似数据的所选行的前 N 个或后 N 个值。



如果希望在构建这些类型的表达式时获得更多指引，您可以使用新的可视化计算编辑器。



有关使用这些计算字段清理数据的详细信息，请参见[创建详细级别和排名计算](#)在本页 261。

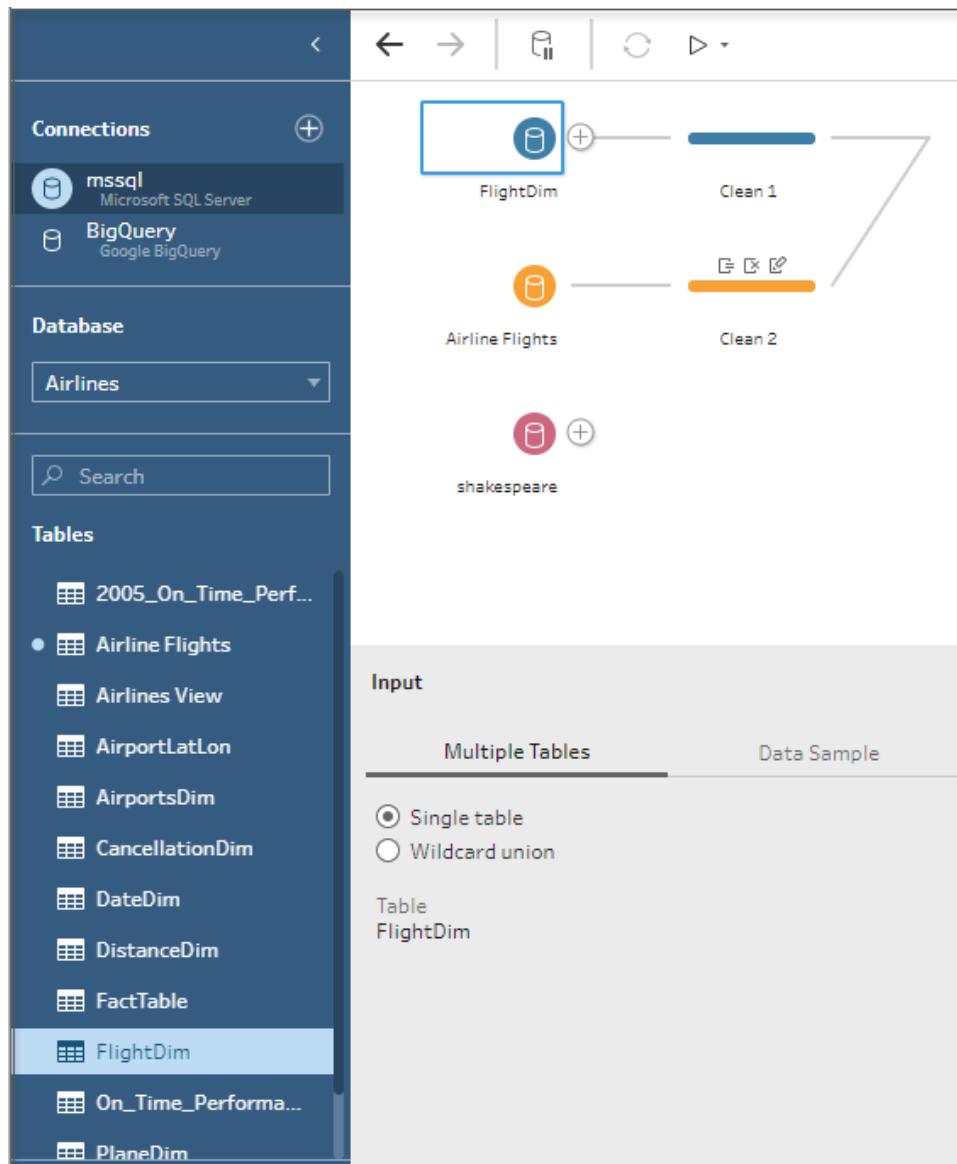
版本 2020.1.1 新增功能

连接到数据

将流程输入映射回其原始连接

如果使用多个输入来构建流程，则在需要更新或交换连接时，尝试查找用于不同输入的单个表或数据源并不总是显而易见。为了更轻松地查找流程输入和表，我们添加了一个新功能，在“连接”窗格中突出显示所选输入的连接器和表。

只需在“流程”窗格中选择一个或多个输入步骤，相关连接器和表即会在“连接”窗格中突出显示。



有关如何连接到数据以构建流程的详细信息，请参见[连接到数据](#) 在本页 141。

清理和调整数据

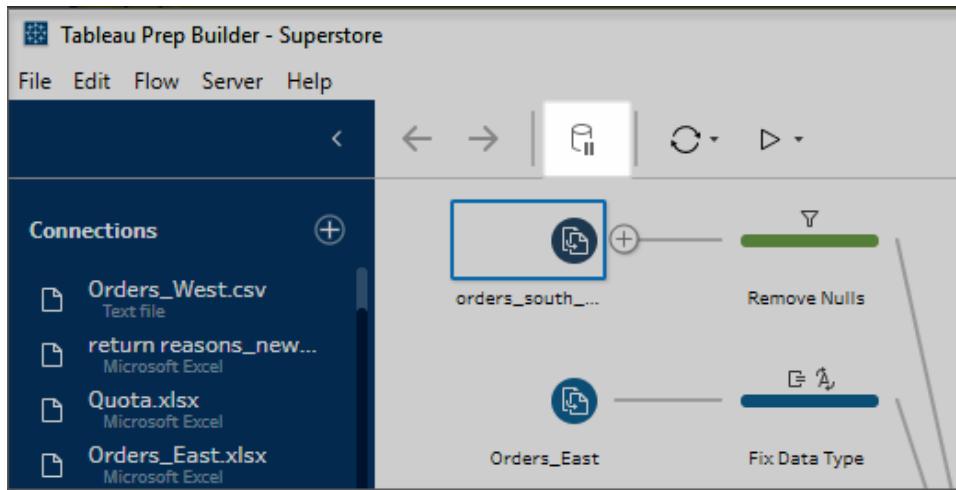
暂停数据更新以提高性能

每次对数据应用清理操作时，Tableau Prep Builder 都会刷新视图，以便您能够立即看到更改的效果。在构建新流程时，这种持续的反馈可帮助您快速了解接下来要执行的步骤，但有时应用这些更改可能需要一些时间。为了提高性能并更快地应用数据更改，您现在可以暂停这些不断进行的数据更新并批量应用更改。

Tableau Prep 帮助

只需在顶部菜单上选择“暂停数据更新”按钮，Tableau Prep Builder 即会暂停数据更新，并将“配置”窗格视图切换到“列表”视图，以便您能够快速应用更改。

系统将只显示字段元数据，而不是在每次更改时刷新视图。完成更改并希望在字段值上查看结果时，单击“恢复数据更新”按钮以切换回标准数据更新模式。



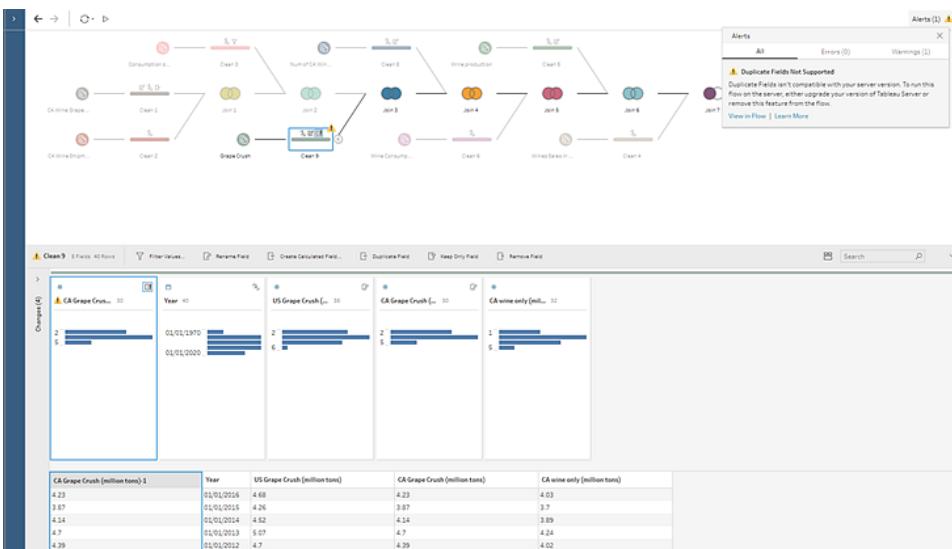
注意: 在暂停数据更新时，将禁用任何需要数据值才能执行任务的清理操作。

有关暂停数据更新以批量应用清理操作的详细信息，请参见[暂停数据更新以提高性能 在本页 229](#)。

与 Tableau Prep 的版本兼容性

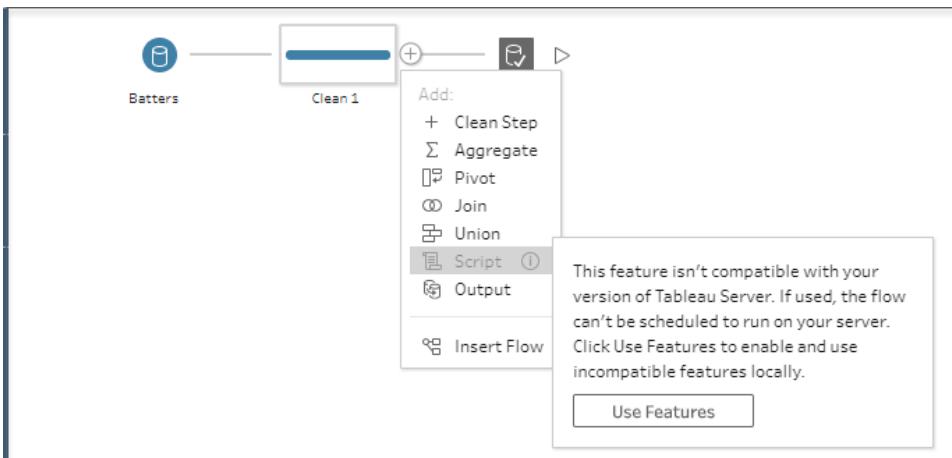
检测并启用不兼容的功能

在 Tableau Prep Builder 版本 2019.3.1 中，我们添加了警告图标和工具提示，以便您可以在构建流程更轻松地查看时哪些功能与您的 Tableau Server 版本不兼容，帮助您在计划发布流程并按计划运行流程时，做出有关流程中包含的功能的明智决策。



在此版本中，我们在处理可能与服务器版本不兼容的新功能时增加了更大的透明度。

现在，在登录到服务器后，Tableau Prep Builder 可以检测哪些功能与您的 Tableau Server 版本不兼容，并禁用这些功能。这样，如果您计划发布流程并按计划运行流程，您就不必猜测可以包括哪些功能。



如果仍要使用该功能并使用 Tableau Prep Builder 或通过命令行手动运行流程，则可以直接从菜单中启用该功能。只需单击“使用功能”按钮，该功能将应用于您的流程，并且所有不兼容的功能将启用并可供使用。

Tableau Prep Builder 版本 2019.3.1 中引入的警告选项仍将标识哪些功能与服务器不兼容，以便在想要使用 Tableau Server 中的计划运行流程的情况下，您可以轻松地查找和移除这些功能。

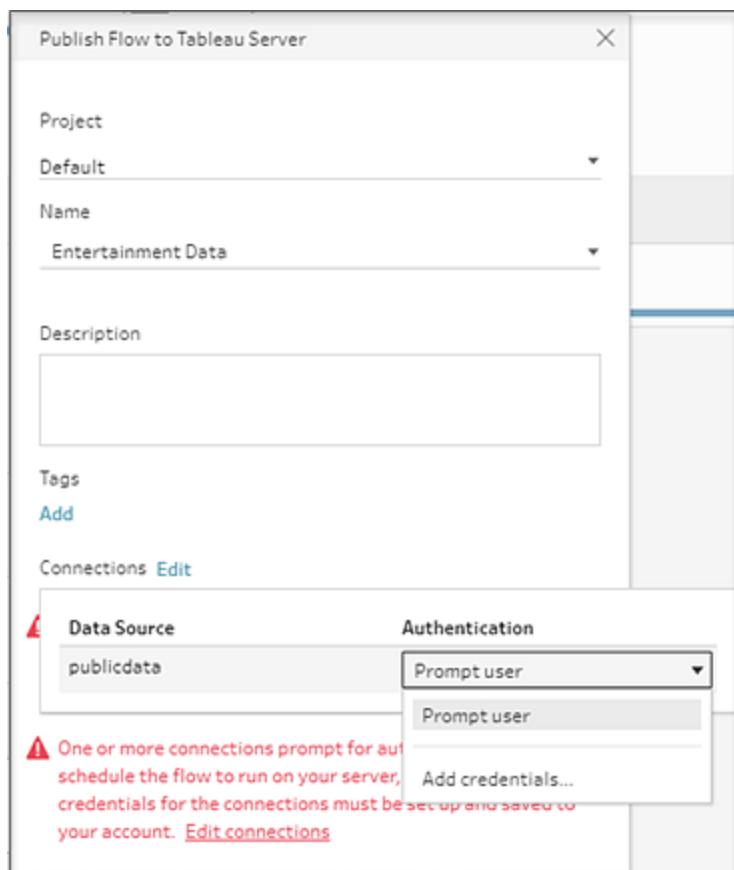
有关处理不兼容功能的详细信息，请参见与 Tableau Prep 的版本兼容性 在本页 333。

使流程数据保持最新

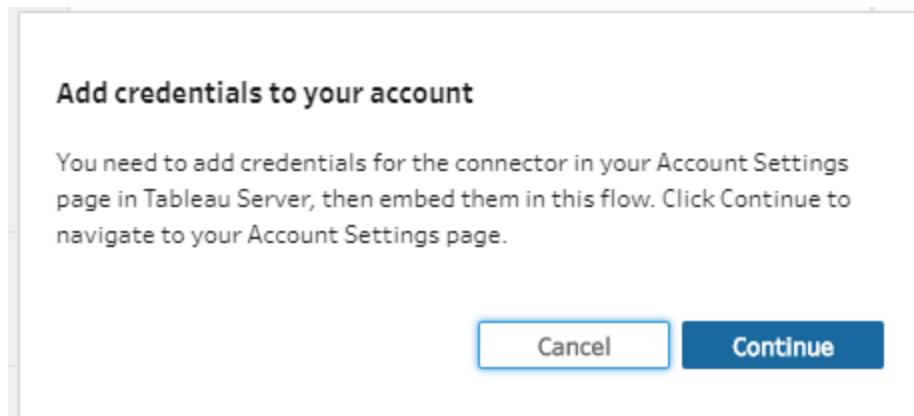
发布流程时为云连接器添加或嵌入凭据

在以前的版本中，如果要发布和计划包含云连接器的流程，则需要提前在 Tableau Server 或 Tableau Online 中配置凭据。否则，Tableau Prep Conductor 无法检测您的凭据，并且流程将无法运行。现在，只需几个简单的步骤，即可直接从 Tableau Prep Builder 的“发布”对话框中添加和嵌入云连接器的凭据。

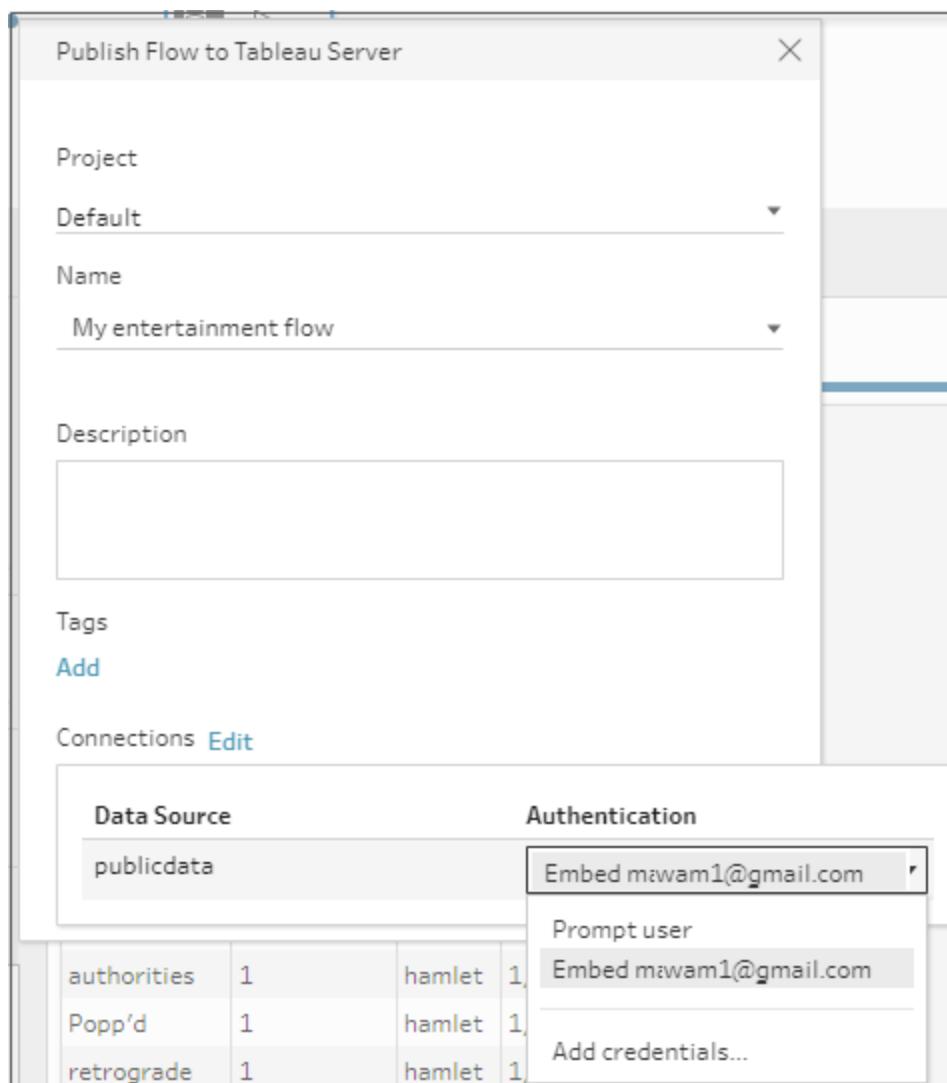
只需登录到您的服务器，并选择“服务器”>“发布流程”。在“发布流程”对话框中，将显示一条新的警告文本，指出您的输入连接需要身份验证。单击消息中的“编辑连接”链接或“连接”部分中的“编辑”链接，然后在“身份验证”下拉列表中，选择“添加凭据”。



另一条消息将确认您的选择。单击“继续”转到 Tableau Server 或 Tableau Online 中的“帐户设置”页面，然后添加凭据。



完成后，返回到 Tableau Prep Builder，单击“完成”，您的新凭据即会自动嵌入到流程中。



有关在发布期间添加云连接器凭据的详细信息，请参见[发布流程 在本页 383](#)。

版本 2019.4.2 新增功能

连接到数据

使用新的连接器连接到更多数据

在此版本中，我们新增了以下新的数据连接器：

- Alibaba AnalyticDB for MySQL
- Alibaba Data Lake Analytics
- Databricks
- Kyvos
- Qubole Presto

有关如何使用这些连接器连接到数据的详细信息，请参见 Tableau Desktop 和 Web 制作帮助的[支持的连接器](#)。有关连接到数据以构建流程的一般信息，请参见[连接到数据 在本页 141](#)。

清理和调整数据

将列转置为行变得更容易

当您需要将数据从列转置为行时，现在可以在配置窗格中手动选择字段来转置它们，Tableau Prep Builder 也可以推荐要转置的字段并为您创建转置。这两个选项都会自动生成转置步骤，以便您可以快速执行任何其他转置操作，比如重命名转置的字段或对其他字段进行转置。

若要从配置窗格中手动将列转置为行，只需选择要转置的字段，然后右键单击或按住 Ctrl 单击 (MacOS)，并从菜单中选择“**将列转置为行**”以创建新的转置步骤并转置字段。

The screenshot shows a data grid in Tableau Prep Builder. The grid has four columns labeled 'Region', '2015', '2016', and '2017'. The 'Region' column contains four rows: 'Central', 'East', 'South', and 'West'. The '2015', '2016', and '2017' columns each contain four data points. A context menu is open over the '2018' column, listing 'Merge', 'Keep Only', 'Remove', and 'Pivot Columns to Rows'.

Region	2015	2016	2017	2018
Central	100,000	105,000	120,000	145,000
East	115,000	110,000	175,000	200,000
South	125,000	150,000	100,000	100,000
West	225,000	200,000	100,000	300,000

如果 Tableau Prep Builder 检测到可转置的字段，则配置窗格工具栏中将显示一个建议选项。单击工具栏中的灯泡 图标并选择转置建议。该建议显示它将转置哪些字段。只需单击“应用”即可生成“转置”步骤并转置字段。

项。单击工具栏中的灯泡 图标并选择转置建议。该建议显示它将转置哪些字段。只需单击“应用”即可生成“转置”步骤并转置字段。

The screenshot shows a data grid in Tableau Prep Builder. The grid has four columns labeled 'Region', '2015', '2016', and '2017'. The 'Region' column contains four rows: 'Central', 'East', 'South', and 'West'. The '2015', '2016', and '2017' columns each contain four data points. A context menu is open over the '2018' column, displaying a 'Recommendations' dialog box. The dialog box contains a 'Pivot Fields' section with five items: '100,000', '145,000', '200,000', '225,000', and '300,000'. An 'Apply' button is at the bottom of the dialog.

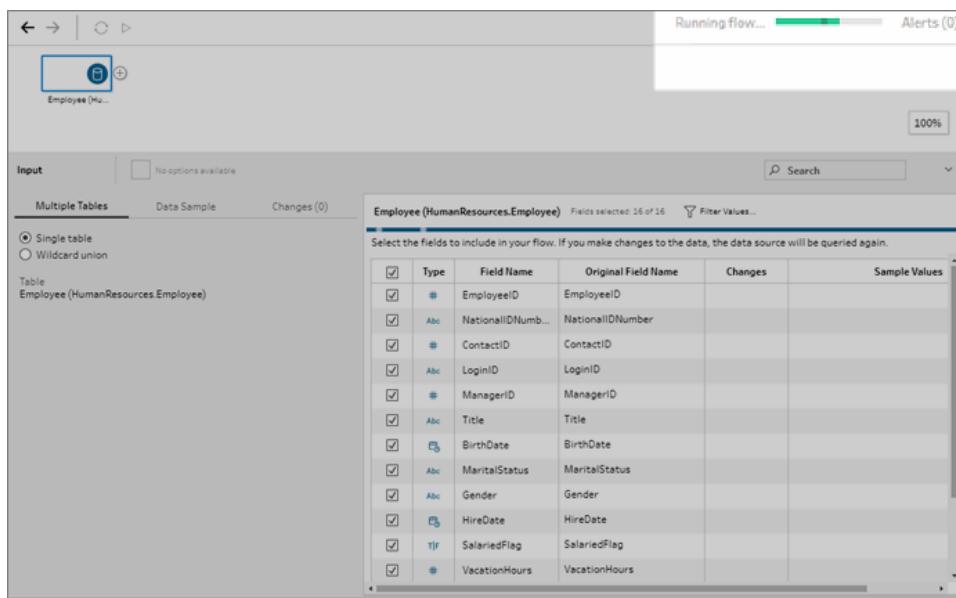
Region	2015	2016	2017	2018
Central	100,000	105,000	120,000	145,000
East	115,000	110,000	175,000	200,000
South	125,000	150,000	100,000	100,000
West	225,000	200,000	100,000	300,000

有关转置数据的详细信息，请参见[转置您的数据](#) 在本页273。有关如何使用建议来清理数据的详细信息，请参见[使用建议应用清理操作\(版本 2018.3.2 及更高版本\)](#) 在本页236。

其他增强功能

查看有关长时间运行的操作的反馈

为了在处理大型数据集或流程时为您提供更高的可见性，我们在“流程”窗格顶部新增了一个进度指示器，以在 Tableau Prep Builder 处理数据时为您提供状态更新。现在，您可以更清楚地了解 Tableau Prep Builder 正在幕后执行的操作，因此您不再需要怀疑您的数据是否仍在处理中，或者是否发生了问题。



支持基于登录名的许可证管理

基于登录名的许可证管理可帮助您为在 Tableau Server 和 Tableau Online 上具有 Creator 角色的用户管理许可。具有 Explorer 或 Viewer(查看者)角色的用户无法使用此功能。如果将基于角色的订阅与 Tableau Server 或 Tableau Online 结合使用，则可以使用基于登录名的许可证管理来简化许可证管理，从而不必使用单独的 Tableau Desktop 和 Tableau Prep Builder 产品密钥。您只需要管理本地 Tableau Server 的一个或多个产品密钥，或者，如果是 Tableau Online，您根本无需管理任何产品密钥。

版本 2019.4.1 新增功能

连接到数据

连接到云中的更多数据

在此版本中，我们新增了以下云连接器：

- Box
- DropBox
- Google Drive
- OneDrive

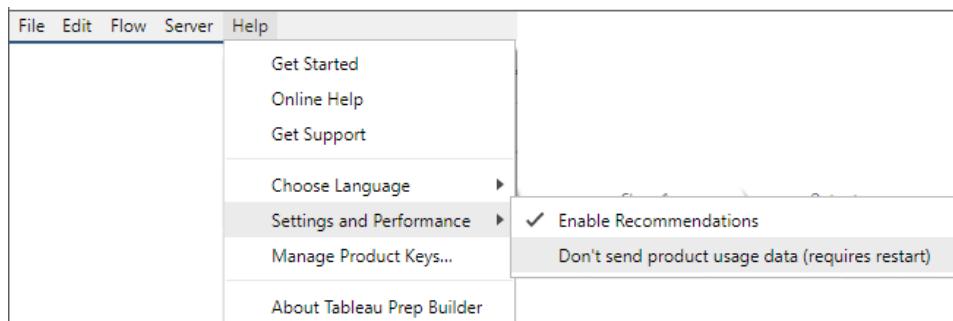
如果您计划将流程发布到 Tableau Server 或 Tableau Online，则可能需要进行其他配置来保存这些连接器的凭据。有关连接到云数据连接器的详细信息，请参见[连接到云数据源 在本页 150。](#)

其他增强功能

与 Tableau 共享使用情况数据

Tableau 会收集数据，帮助我们了解产品的使用情况，以便改进现有功能并开发新的功能。所有使用情况数据均依据 [Tableau 隐私政策](#) 收集和处理。但是，如果您不想参与，则可以选择退出。

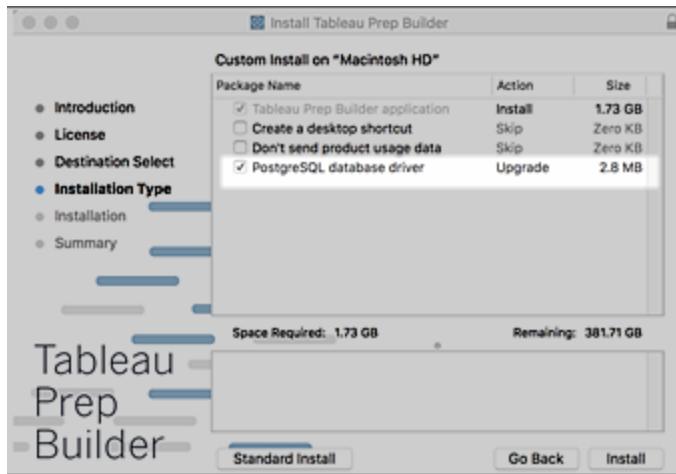
选择退出选项已更改。默认情况下已启用使用情况数据收集。如果您不想参与，请从安装程序、“自定义安装程序”(MacOS) 对话框或在安装之后从“帮助”菜单中选中**“不发送产品使用情况数据”**复选框。



有关详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[关闭使用情况报告](#)。

不再为 MacOS 自动安装默认数据库驱动程序

在 Mac 上安装 Tableau Prep Builder 时，我们只会自动安装 PostgreSQL 数据库驱动程序。如果您需要其他驱动程序来连接到数据，可以从[驱动程序下载](#)页面安装它们。



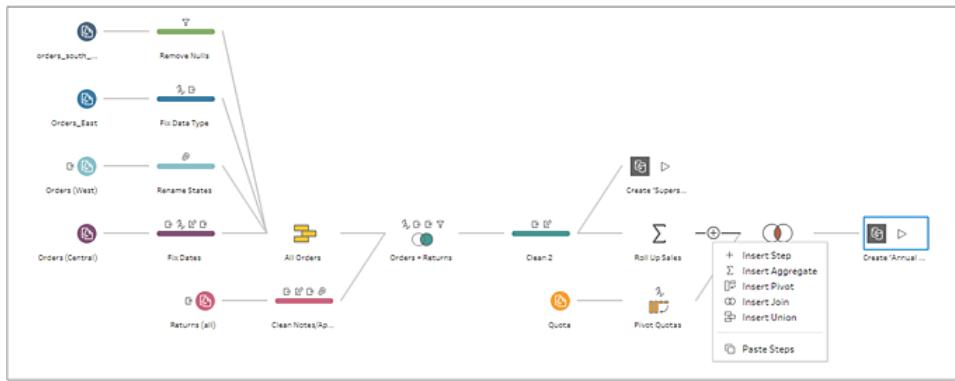
版本 2019.3.2 新增功能

构建和组织流程

只需单击一下，即可在所需的位置插入步骤

在以前的版本中，您可以在流程中的现有步骤后插入清理步骤，并且流程将向右扩展来为新步骤腾出空间。但是，在现有步骤后添加任何其他步骤类型会导致创建一个流程分支，然后您必须通过移除和添加流程线来重新连接该分支。

在此版本中，我们添加了一个新的插入菜单，使您只需单击一次便可在现有步骤之间插入任何步骤类型（甚至复制的步骤）。只需将光标悬停在要插入步骤的流程线上，并使用新的“插入”菜单选择并插入所需步骤类型，即可在所需位置插入。

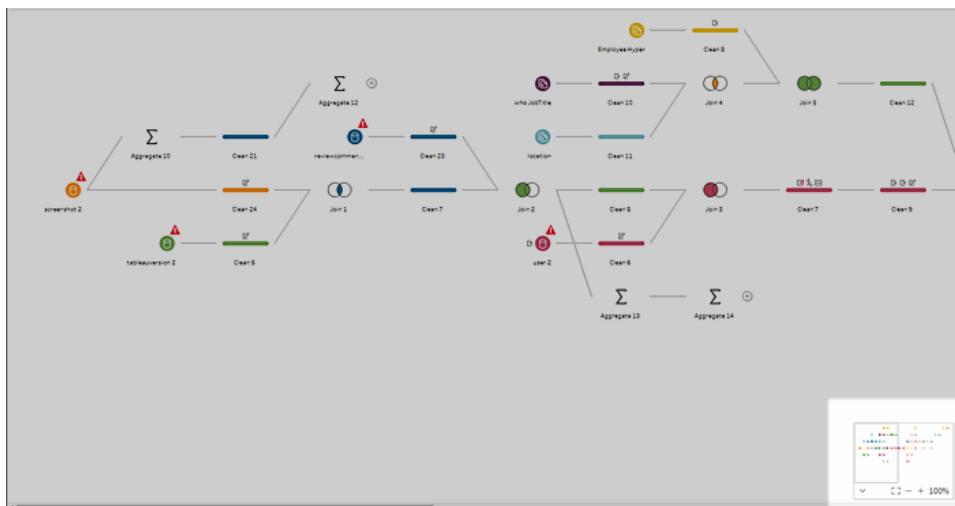


注意：“添加”菜单中的“插入步骤”菜单项已转移到这个新菜单中。此选项不可用于“输入”和“输出”步骤类型。

有关在流程中添加和插入步骤的详细信息，请参见[添加或插入步骤](#) 在本页 179。

使用流程导航器轻松查看大型流程

处理大型流程时，来回滚动以查找和修复步骤或更改可能很困难。新的流程导航工具显示在流程窗格的右下角。此工具是整个流程的微缩图像。您可以使用它放大和缩小该图像，单击图像的某些部分直接导航到流程的该部分，调整流程大小，以及执行其他操作。如果不想要显示该工具，只需单击一下便可将其最小化或还原。



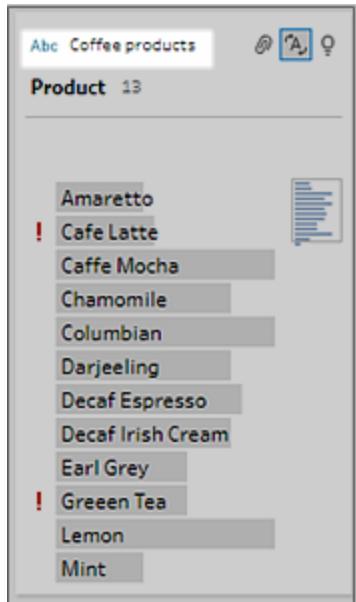
有关连接到数据并构建流程的详细信息，请参见[使用流程导航工具](#) 在本页 191。

检查和筛选数据

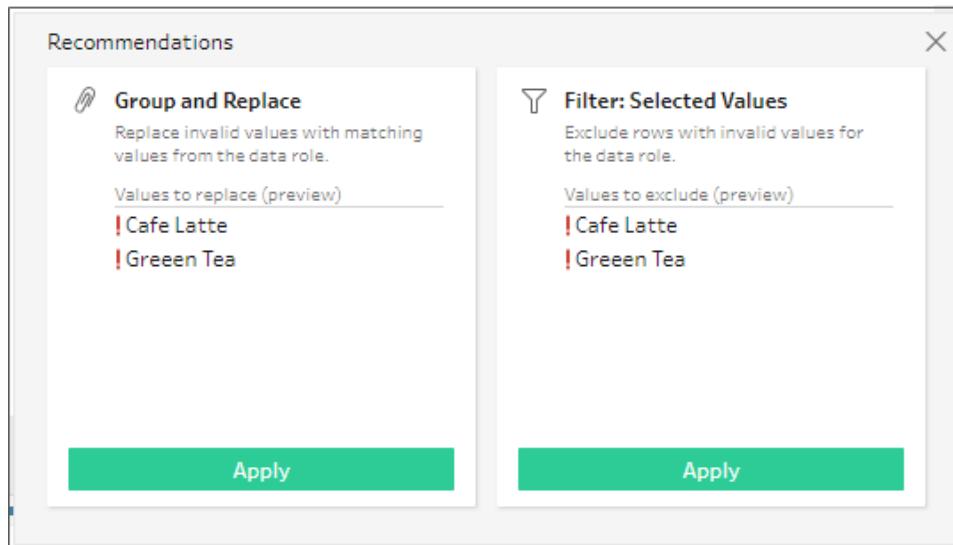
使用自定义数据角色对类似值进行分组

在 Tableau Prep Builder 版本 2019.3.1 中，我们引入了使用数据集中的字段值创建自己的自定义数据角色的功能。在此版本中，您现在可以使用这些自定义数据角色，根据拼写和发音对字段中的值进行分组，以便通过匹配无效值并将其替换为最接近的有效值来自动修复无效值。

只需将自定义数据角色应用于字段，并且任何与标准值不匹配的值都标有红色感叹号。



为了修复无效值，您可以单击字段上的“建议”图标，将建议应用于组并将无效值替换为有效值，或者从“更多选项”…菜单中选择“分组和替换”>“拼写”或“发音 + 拼写”以选择修复无效值的方式。



有关创建和使用自定义数据角色的详细信息，请参见[按数据角色对类似值进行分组](#)在本页219。

清理和调整数据

使用新列表视图对字段应用清理操作

目前，您可以在配置窗格、结果窗格或数据网格中应用清理操作。但是，一次可以看到的字段数受屏幕的大小所限。为了更轻松地扫描所有字段和执行操作，我们添加了一个新的列表视图，使您能够配置窗格或结果窗格转换为列表。选择此选项后，此视图状态将在流程的所有步骤中保持不变，但您可以随时对其进行更改。

Type	Field Name	Changes
Calculated Field	[Region] "Central"	
Calculated Field	[Order Date] STR([Order Day])+"-"+STR([Order Month])+"-"+STR([Order Year])	
Change Type	[Order Date] To Date type	
Calculated Field	[Ship Date] STR([Ship Day])+"-"+STR([Ship Month])+"-"+STR([Ship Year])	
Remove Field	[Discount] From [Discounts] to [Discount]	
Change Field	[Product Name] From [Product] to [Product Name]	
Change Type	[Ship Date] To Date type	
Remove Field	[Order Year]	
Calculated Field	[Region] "Central"	
Calculated Field	[Order Date] STR([Order Day])+"-"+STR([Order Month])+"-"+STR([Order Year])	
Change Type	[Order Date] To Date type	
Calculated Field	[Ship Date] STR([Ship Day])+"-"+STR([Ship Month])+"-"+STR([Ship Year])	
Remove Field	[Discount] From [Discounts] to [Discount]	
Change Field	[Product Name] From [Product] to [Product Name]	
Change Type	[Ship Date] To Date type	
Remove Field	[Order Year]	

在此视图中，您可以将像“移除”或“只保留”这样的操作应用于多个字段，或者快速应用像“重命名字段”或“复制字段”这样的清理操作。您还可以通过“更多选项”… 菜单应用其他

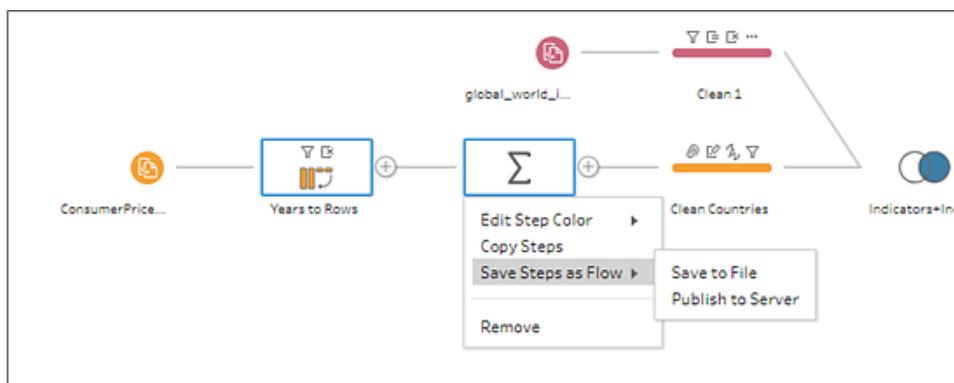
清理操作，但像“分组和替换”、“筛选”或“拆分值”这样的操作会将您返回到配置窗格或结果窗格视图来完成操作。

有关在列表视图中清理字段的详细信息，请参见[选择视图 在本页 227](#)。

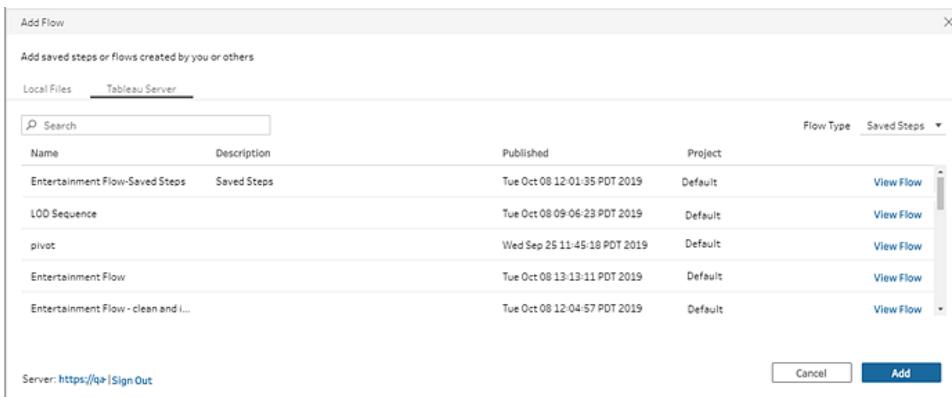
通过可重用的步骤来利用您的工作和其他人的工作

如果您通常对数据反复执行相同的操作，则可以选择一个或多个流程步骤及其关联的操作或整个流程，并将其保存以在您创建和管理的其他流程中重复使用。您可以将步骤保存在本地，或将其发布到 Tableau Server 或 Tableau Online 以与其他人共享。

若要创建可重用的步骤，只需选择要保存的步骤，右键单击并选择“将步骤另存为流程”，然后选择是将步骤保存到本地文件还是服务器。将流程步骤发布到服务器时，系统会自动添加“已保存的步骤”标记，以便以后能轻松查找流程步骤。



保存后，可以打开并插入任何流程。在流程窗格画布中右键单击或按住 Ctrl 并单击 (MacOS)，并选择“插入流程”，或从顶部菜单中选择“编辑”>“插入流程”。在“添加流程”对话框中，您可以从文件或服务器中选择并插入保存的流程步骤。该对话框会自动筛选以显示带有“已保存的步骤”标记的所有流程，但您可以更改筛选器并插入您有权访问的任何流程。



所选流程将添加到流程窗格中，然后您可以在流程中所需的步骤移动和连接步骤。有关创建和使用可重用流程步骤的详细信息，请参见[清理和调整数据 在本页 223](#)。

保存和共享工作

从命令行运行流程时禁用 Mac OS 的钥匙串对话框

在 Mac OS 通过命令行运行流程时，我们无需用户在每次流程运行时手动输入其 Mac OS 钥匙串凭据。

有关详细信息，请参见[通过命令行刷新流程输出文件 在本页 321](#)中的“禁用 SSL 验证 (MacOS)”命令选项。

版本 2019.3.1 新增功能

连接到数据

使用 Google BigQuery 连接到数据

使用 Google BigQuery 连接到云中的更多数据。与 Tableau Desktop 一样，您现在可以连接到 Google BigQuery 中存储的数据，并将流程发布到 Tableau Server 和 Tableau Online(如果已配置 Google Oauth 连接)。

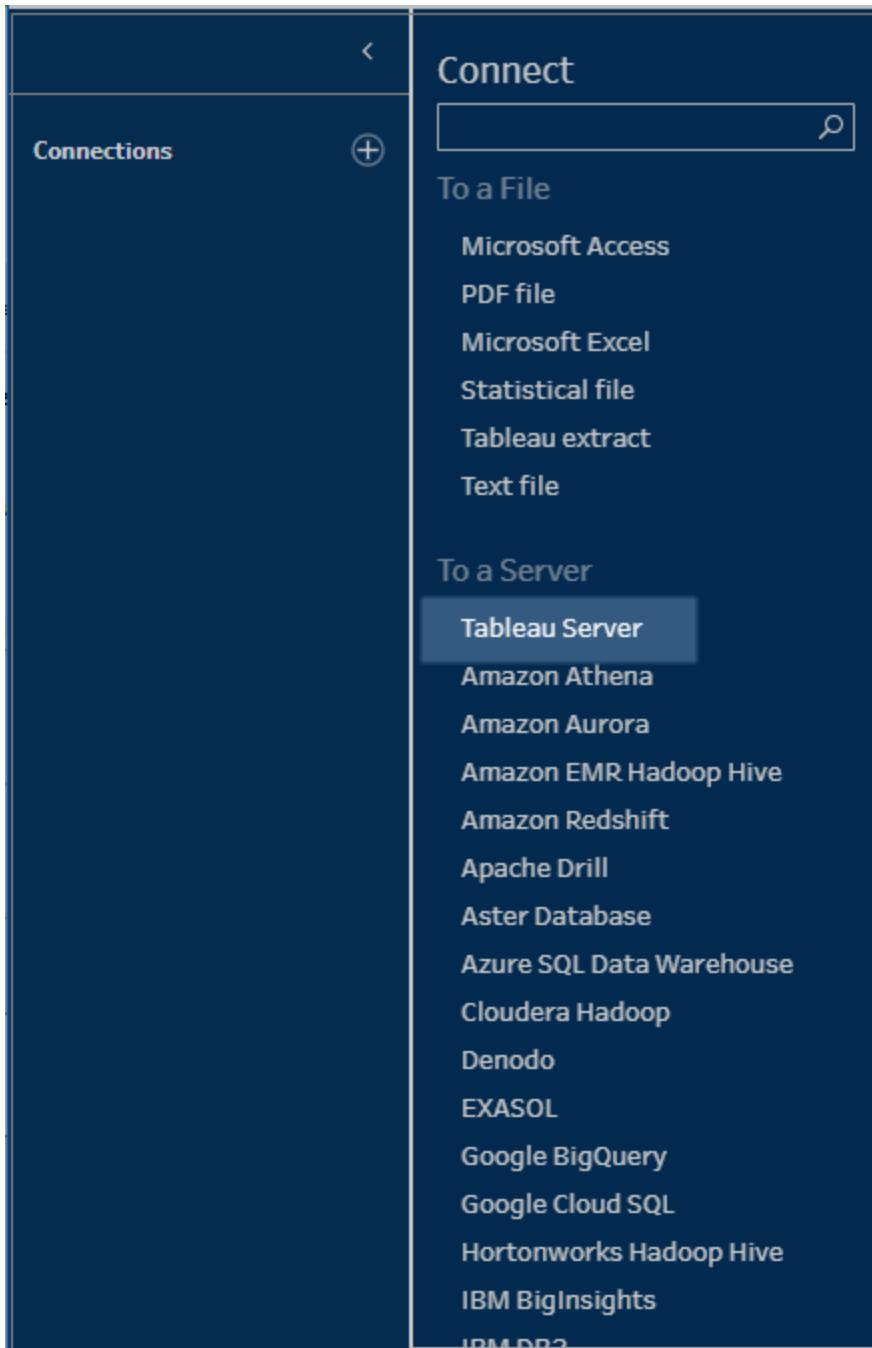
注意：Tableau Prep Builder 当前不支持使用 Google BigQuery 自定义属性。此外，包含 Google BigQuery 连接器的流程也不能通过命令行运行。作为替代，请手动运行流程，或者在 Tableau Server 或 Tableau Online 中按计划运行流程。

有关如何连接到 Google BigQuery 的详细信息，请参见 Tableau Desktop 帮助中的 [Google BigQuery](#)。有关连接到数据的详细信息，请参见[连接到数据 在本页 141](#)。

为流程输入连接到发布的数据源

无论是从 Tableau Desktop 中还是从 Tableau Prep Builder 中以输出形式生成发布的数据源并将其发布到 Tableau Server 或 Tableau Online，您现在都可以连接到发布的数据源，并使用其作为流程的输入数据源。

只需从“**连接**”窗格中选择“**Tableau Server**”，然后登录到服务器或站点。



登录后，将显示可用数据源的列表。选择数据源，或使用搜索选项查找数据源并将其拖到流程窗格以启动流程。

Tableau Prep 帮助

The screenshot shows the Tableau Prep Builder interface. On the left, there's a sidebar with 'Connections' and 'Data Sources'. A connection named 'https://near (Default)' is selected. Below it, a data source named '600topmovies3.csv (600topmovies) {Default}' is shown. The main area is titled 'Input' and displays the 'Tableau Server' tab. It shows the fields from the CSV file: 'Adjusted Gross (\$mill)', 'Budget (\$mill)', 'Day of Week', 'Director', 'Genre', 'Gross (\$mill)', 'IMDb Rating', 'Month', 'MovieLens Rating', 'Overseas (\$mill)', 'Overseas%', and 'Profit (\$mill)'. Each field has its original name and a brief description or value.

Type	Field Name	Original Field Name	Changes	Sample Values
Number	Adjusted Gross (\$mill)	Adjusted Gross (\$mill)		6,846.7, 3,901.2, 1,717.1
Number	Budget (\$mill)	Budget (\$mill)		3.9, 0.858, 75
Text	Day of Week	Day of Week		Friday, Thursday, Wednesday
Text	Director	Director		George Cukor, Victor Fleming, Sa
Text	Genre	Genre		drama, animation
Number	Gross (\$mill)	Gross (\$mill)		400.2, 267.4, 215.9
Text	IMDb Rating	IMDb Rating		8.2, 7.4, 7.2
Text	Month	Month		December, August, January
Text	MovieLens Rating	MovieLens Rating		3.83, 3.51, 3.42
Number	Overseas (\$mill)	Overseas (\$mill)		201.5, 165.2, 71
Number	Overseas%	Overseas%		0.504, 0.618, 0.329
Number	Profit (\$mill)	Profit (\$mill)		396.3, 266.5, 140.9

注意:当前不支持包括用户筛选器的已发布数据源。

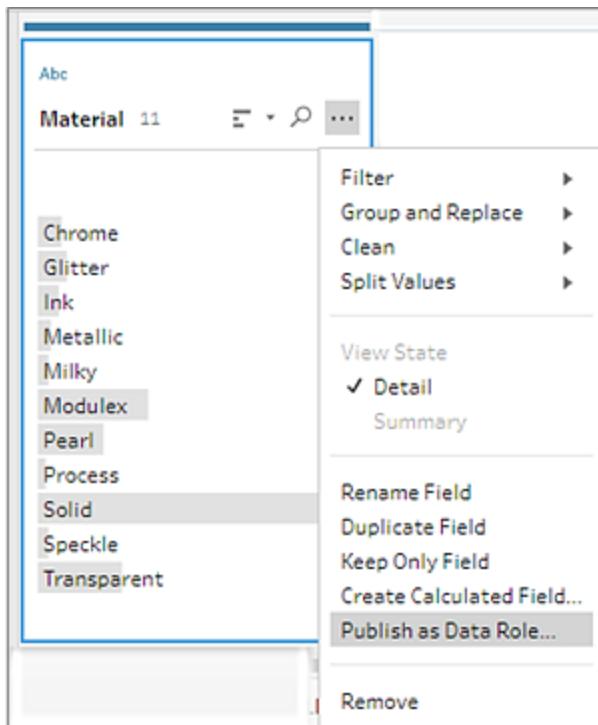
有关连接到数据和构建流程的详细信息，请参见[连接到数据 在本页 141](#)。

检查和筛选数据

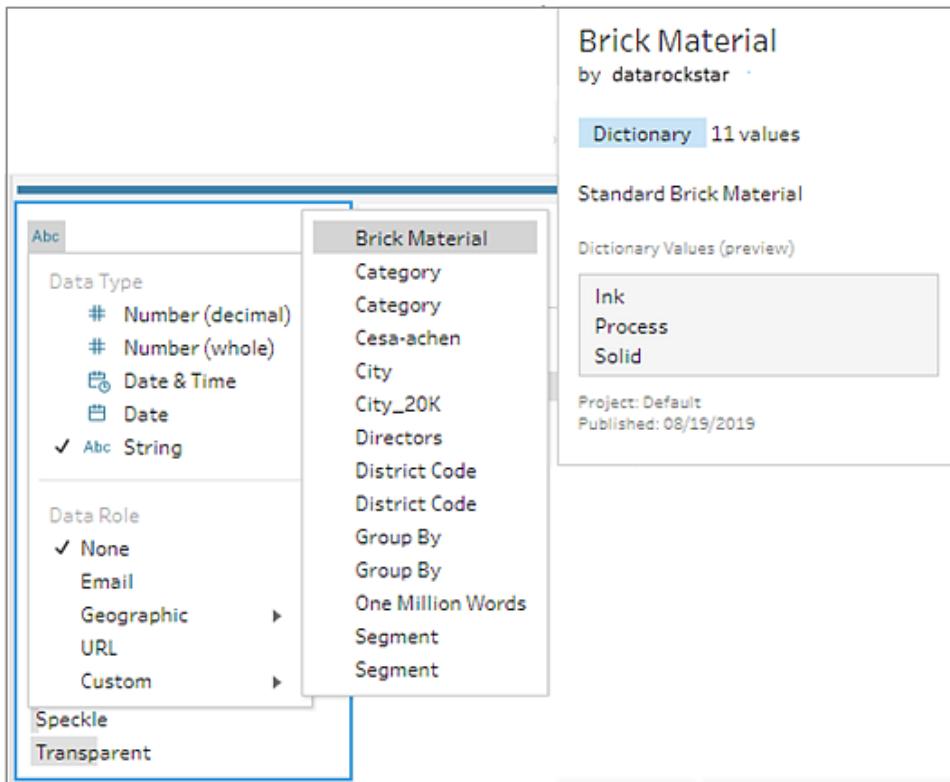
创建自定义数据角色

您可以使用数据角色来标识和验证字段值。在以前的版本中，Tableau Prep Builder 提提供了一组可应用于字段来快速标识无效值的标准数据角色。在此版本中，您现在可以使用数据集中的字段值创建自己的自定义数据角色，然后将其发布到 Tableau Server 或 Tableau Online 以便在流程中使用它们或与他人共享您的数据角色，从而比以往更轻松地标识和清理无效值。

只需选择要使用的字段，并从“更多选项” 菜单中选择“以数据角色形式发布”。



发布自定义数据角色之后，一旦登录到在其中发布了自定义数据角色的站点或服务器，您就可以从新的“**自定义**”菜单选项中选择该角色，将其应用于您的字段。



您可以在 Tableau Server 和 Tableau Online 上查看及管理发布的自定义数据角色。查看发布到站点或服务器的所有自定义数据角色、将其移动到其他项目、更改权限或将其删除。

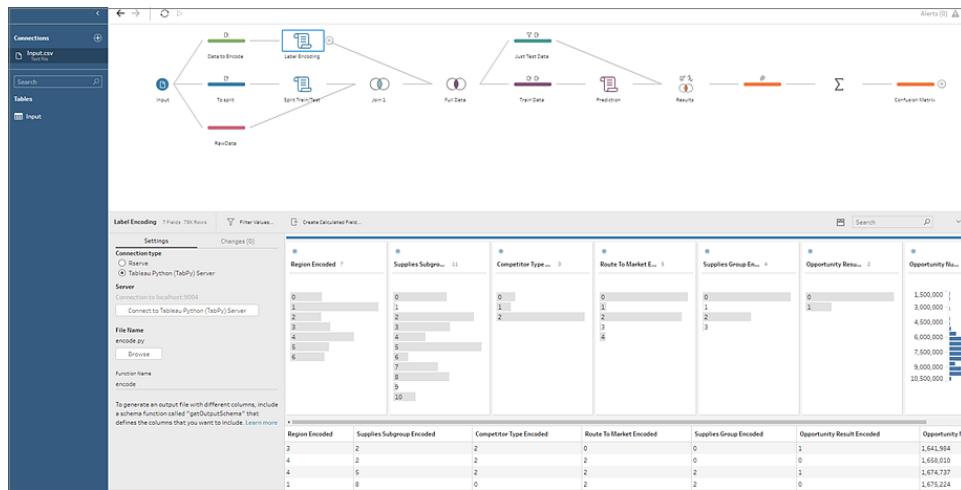
Type	Name	Project	Owner	Modified
<input type="checkbox"/>	city-datarole	Default	workuser	Jun 25, 2019, 5:37 PM
<input type="checkbox"/>	city-datarole2		workuser	Jun 25, 2019, 5:42 PM
<input type="checkbox"/>	City1		workuser	Jun 18, 2019, 11:23 AM

有关创建和使用自定义数据角色的详细信息，请参见[创建自定义数据角色\(版本 2019.3.1 及更高版本\)](#)在本页212。

清理和调整数据

在流程中使用 R 和 Python 脚本

您现在可以使用 R 和 Python 脚本来执行更复杂的清理操作，或者在流程中并入预测建模数据。数据使用 Rserve 或 TabPy 通过 R 或 Python 脚本步骤以输入形式从流程中传递，然后您可以使用 Tableau Prep Builder 的功能和函数继续清理的输出数据的形式返回。



有关如何在流程中包括脚本的详细信息，请参见[在流程中使用 R 和 Python 脚本](#)在本页281。

保存和共享工作

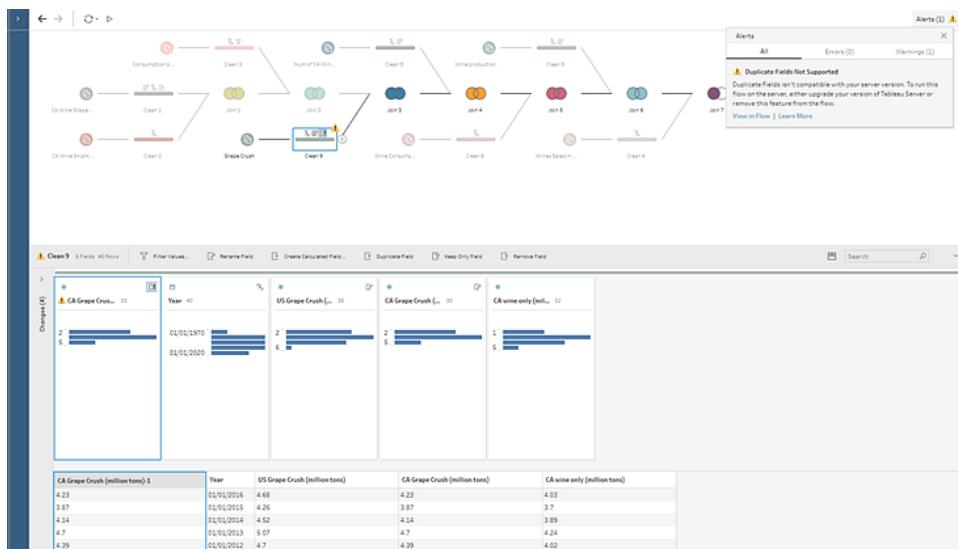
在构建流程时找出不兼容的功能

Tableau Prep Builder 每个月都会发布新版本。如果新版本中引入了新功能或连接器，并且您打算将流程发布到较旧版本的 Tableau Server 或 Tableau Online，则可能会出现兼容性问题。当前您可以在发布流程时看到是否有不兼容的功能，但在此时您已经构建了流程，因此返回删除不兼容的功能需要时间。

在此版本中，兼容性通知更加透明。您现在可以在构建流程时看到任何不兼容的功能，这样您就可以立即做出有关所包含功能的明智决策。

只需登录到 Tableau Server，在您构建流程时，如果选择的功能在您的 Tableau Server 版本中不受支持，就会显示通知。单击画布右上角的“**通知**”可查看所有通知和错误，或将光标悬停在流程窗口中步骤或注释的通知上以查看通知详细信息。

单击“[在流程中查看](#)”链接可快速导航到警告与之关联的步骤、注释、字段或更改。步骤和更改都标有通知图标并突出显示，因此您可以快速查找并解决问题，并继续处理流程。



有关解决兼容性问题的详细信息，请参见[与 Tableau Prep 的版本兼容性](#)在本页333。

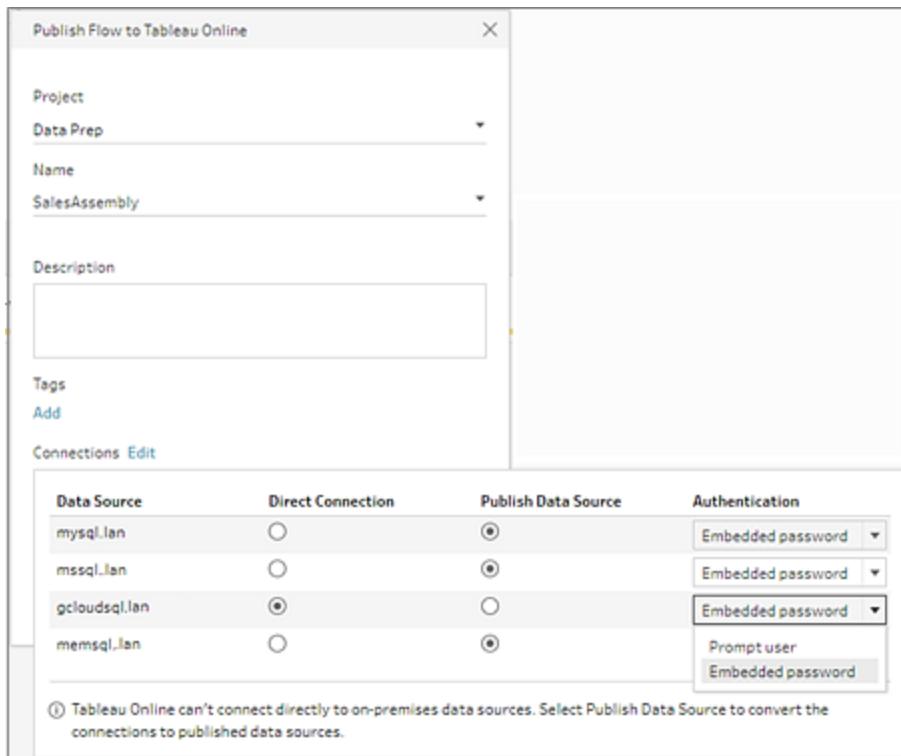
使流程数据保持最新

Tableau 数据管理加载项包括 Tableau Prep Conductor

Tableau Prep Conductor 包括在单独许可的 Tableau Server 或 Tableau Online 数据管理加载项中。有关详细信息，请参见[许可 Tableau Prep Conductor](#)在本页352。

将流程发布到 Tableau Online

您现在可以将流程发布到 Tableau Server 或 Tableau Online，并按计划运行流程以使数据保持最新。Tableau Online 不支持直接连接到本地数据源。如果将流程发布到连接到本地数据源的 Tableau Online，您可以编辑连接以自动将连接转换为发布的数据源。

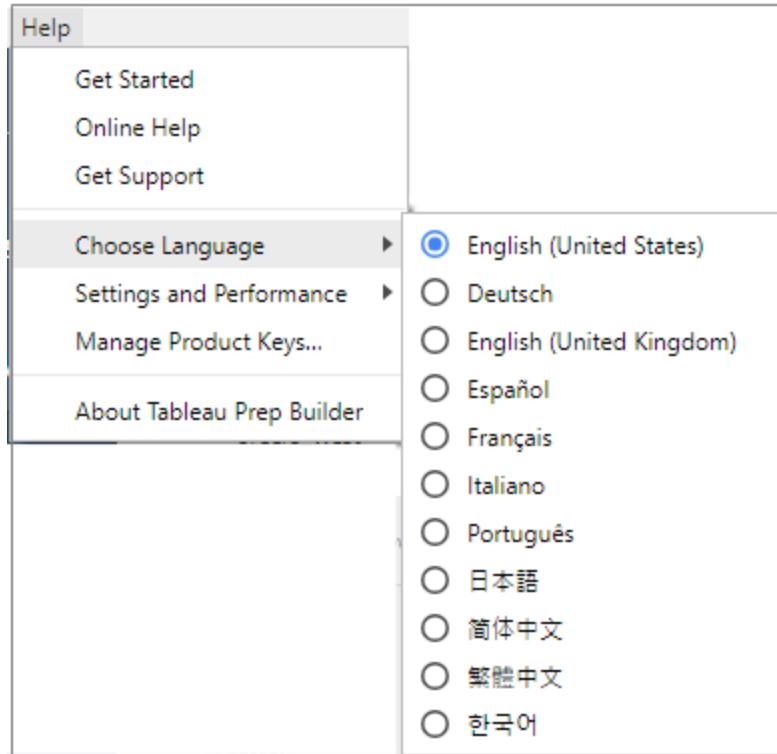


有关将流程发布到 Tableau Online 的详细信息，请参见[将流程发布到 Tableau Server 或 Tableau Online 在本页 378。](#)

其他增强功能

Tableau Prep Builder 现在提供意大利语版本

现在可以采用 11 种不同的语言之一配置用户界面(菜单、消息等)。有关配置显示语言的详细信息，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的[设置显示语言](#)

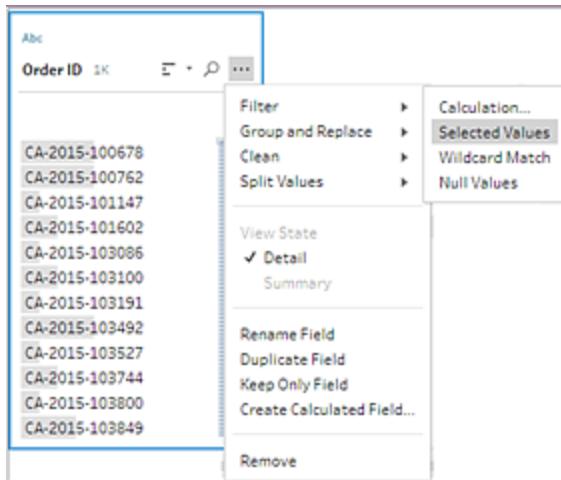


版本 2019.2.3 新增功能

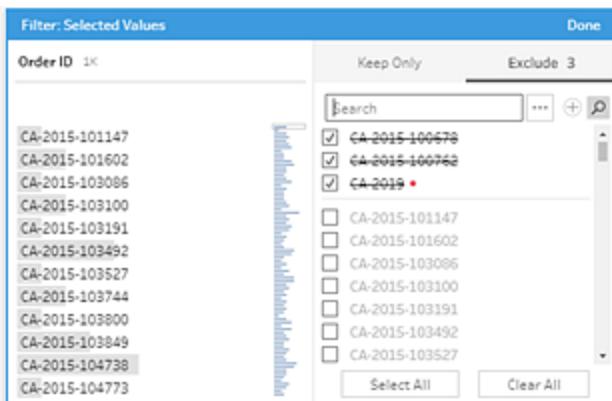
检查和筛选数据

选择要筛选的特定值

借助新的“选定值”筛选器，您现在可以选取和选择要为字段保留或排除的值，即使这些值不在样本中。在“配置”窗格、“结果”窗格或数据网格中，单击“更多选项”... 按钮，然后选择“筛选”>“选定值”。



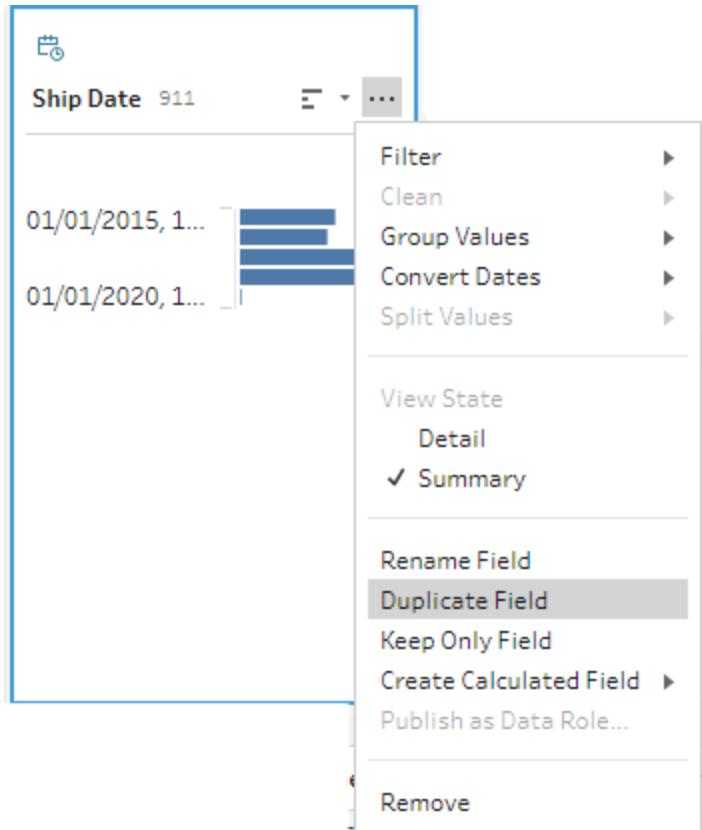
这个新的可视指示器允许您选择要保留或排除的值，并直观地显示您已选定的值。如果正在处理数据样本，并且想要包括或排除不在所处理的样本中的值，只需添加所需的值，Tableau Prep Builder 也会将保留或排除选项应用于这些新值。



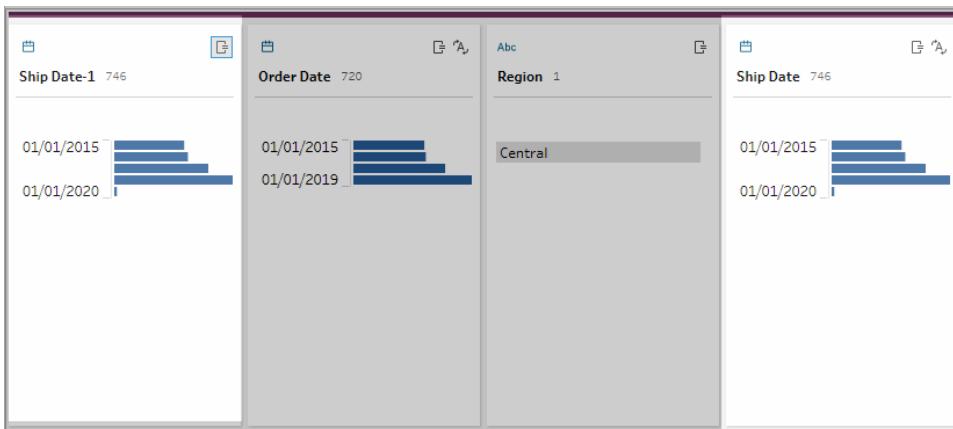
有关筛选字段值的详细信息，请参见[检查您的数据 在本页 194](#)。

通过单击复制字段

在以前的版本中，如果想要在一个字段上试验清理操作但不想更改原始数据，您将必须编写一个计算来复制该字段。现在，您只需单击便可轻松地复制字段。选择要复制的字段，然后从“更多选项”菜单中选择“复制字段”。



系统即会使用相同名称和修饰符创建一个新字段，并将其放在“配置”窗格、数据网格或“结果”窗格的最左侧。例如，“Ship Date -1”(发货日期 -1)。

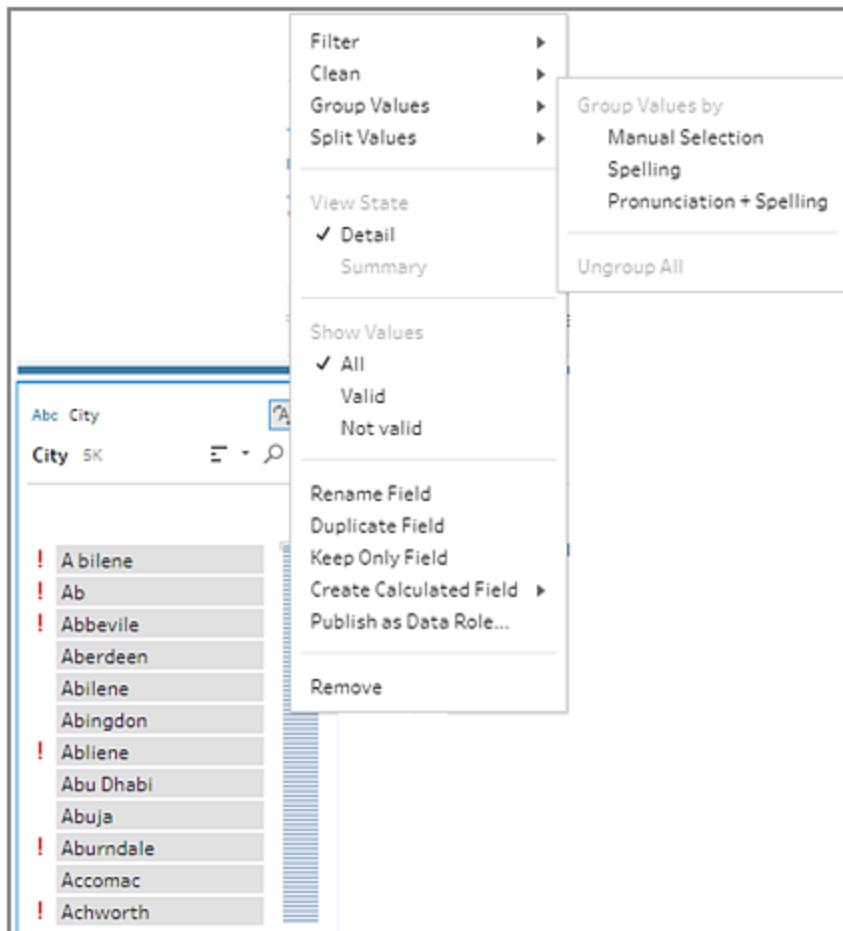


有关复制字段和其他清理选项的详细信息，请参见[清理和调整数据 在本页 223](#)[应用清理操作 在本页 226](#)

清理和调整数据

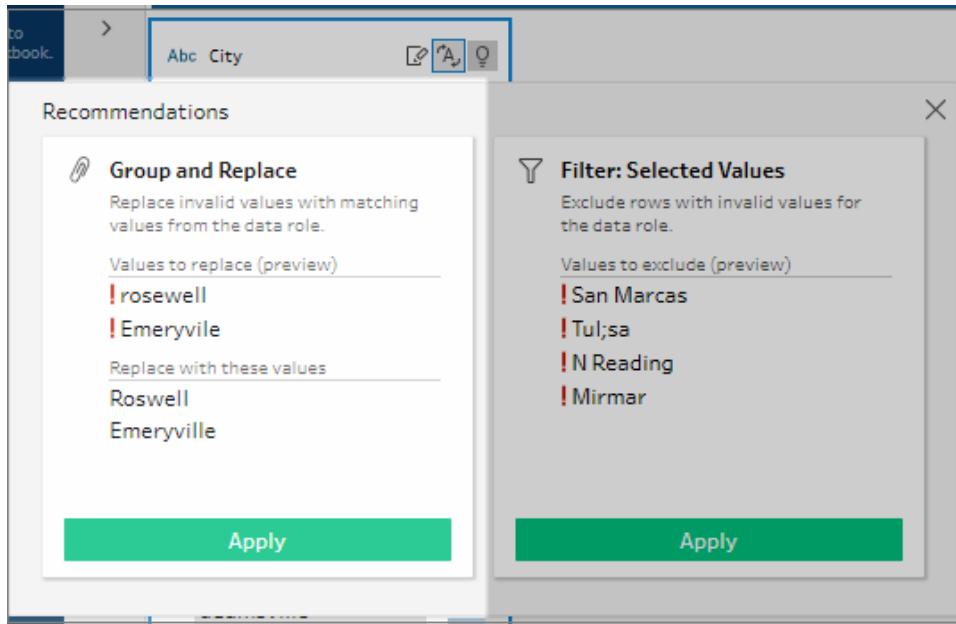
使用模糊匹配和清理建议修复无效的数据角色值

可以使用数据角色来快速确定无效的字段值。分配数据角色之后，您可以使用两个模糊匹配算法（“**拼写**”和“**拼写 + 发音**”）之一对数据角色的无效值进行快速分组，并将其与适用于该数据角色的类似标准化值匹配。



除了为您的字段建议数据角色之外，Tableau Prep Builder 现在还可使用“**拼写**”和“**拼写 + 发音**”算法建议如何为分配给数据角色的字段修复无效值。如果检测到建议的修复，则配置

卡上将出现一个灯泡 图标。单击该图标并应用建议的更改，以自动对无效值进行分组，并将其替换为适用于该数据角色的类似标准化值。



此功能仅可用于以下数据角色：

- 机场
- 城市
- 国家/地区
- 郡/县
- 州/省/市/自治区

有关将数据角色应用于字段并使用分组和替换修复数据角色的无效值的详细信息，请参见[使用数据角色验证数据 在本页 209](#)。

版本 2019.2.2 新增功能

连接到数据

使用其他数据库连接器 (ODBC) 连接到更多数据

如果需要连接到“连接”窗格中未列出的数据源，您可以使用支持 SQL 标准并实现 ODBC API 的“**其他数据库 (ODBC)**”连接器连接到任何数据源。使用“**其他数据库 (ODBC)**”连接器连接到数据的工作方式与您在 Tableau Desktop 中使用该连接器的方式类似，但有一些不同之处：

- 您只能使用 DSN(数据源名称) 选项进行连接。
- 若要在 Tableau Server 中发布并运行流程，必须使用匹配的 DSN 配置服务器。
- Windows 和 MacOS 上的连接体验是一致的。不支持 ODBC 驱动程序 (Windows) 的提示连接属性。
- Tableau Prep Builder 仅支持 64 位驱动程序。

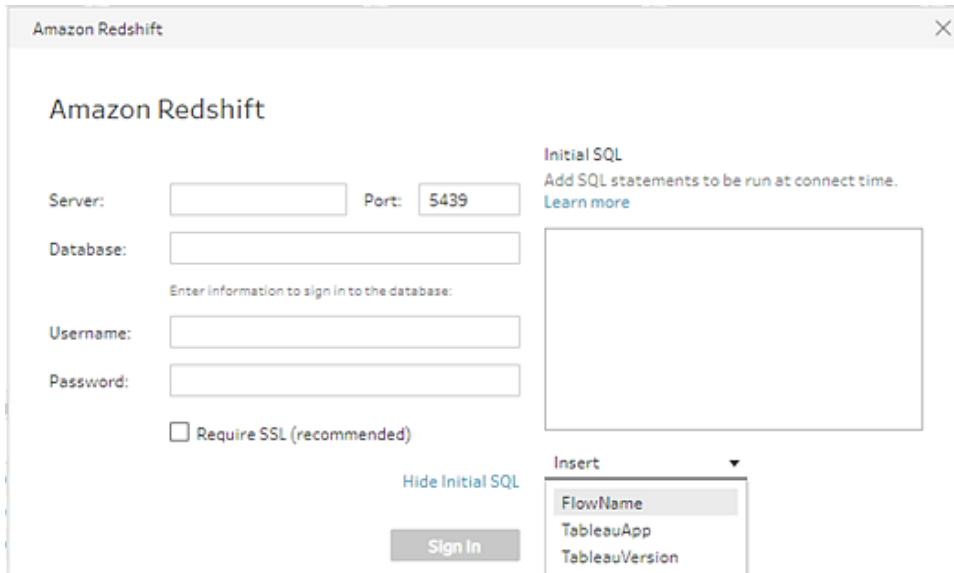
有关使用“其他数据库”ODBC 连接器连接到数据的详细信息，请参见[连接到数据 在本页 141](#)。

在连接到数据时指定初始 SQL 命令

在连接到支持初始 SQL 命令的数据库时，您现在可以指定初始 SQL 命令，该命令将在进行数据库连接时运行。例如，在连接到 Amazon Redshift 时，您可以输入 SQL 语句，以便在连接到数据库时应用筛选器，就像在“输入”步骤中添加筛选器一样。SQL 命令将在数据抽样之前应用，并加载到 Tableau Prep Builder 中。

如果更改初始 SQL 命令并刷新“输入”步骤，则会再次运行初始 SQL 命令，并更新“输入”步骤。在运行流程时，初始 SQL 命令将在处理所有数据之前运行，如果在 Tableau Server 上计划流程，则初始 SQL 将在每次数据加载时运行。

有关详细信息，请参见[连接到数据 在本页 141](#)。



注意: Tableau Prep Builder 中的初始 SQL 尚不支持参数。

检查和筛选数据

在清理数据时仅保留所需的字段

在以前的版本中，您可以通过在连接到数据后清除复选框来移除“输入”步骤中不需要的字段，在清理或操作步骤中，您可以选择一个或多个字段，并且只需右键单击便可将其移除。现在，您可以在“配置”窗格或数据网格中选择一个或多个字段，并右键单击以仅保留选定字段和移除所有其他字段。

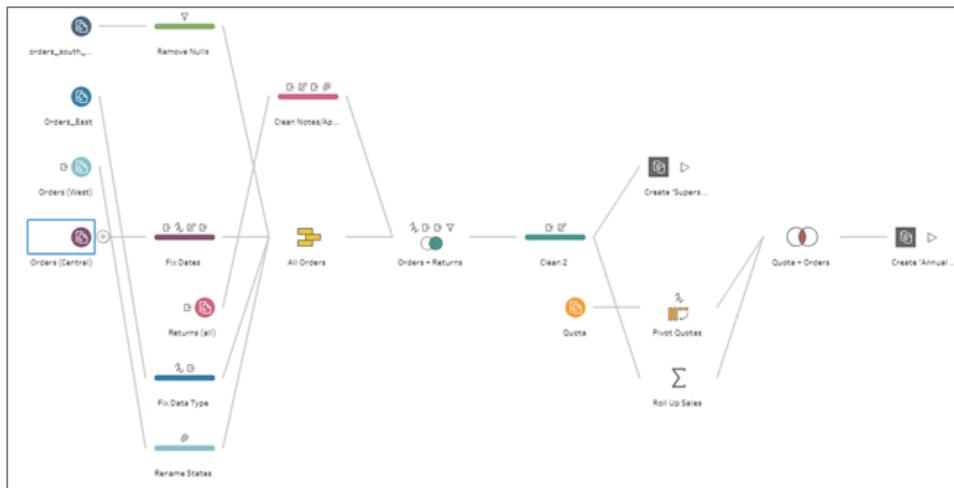
有关详细信息，请参见[保留或移除字段](#) 在本页 203。

清理和调整数据

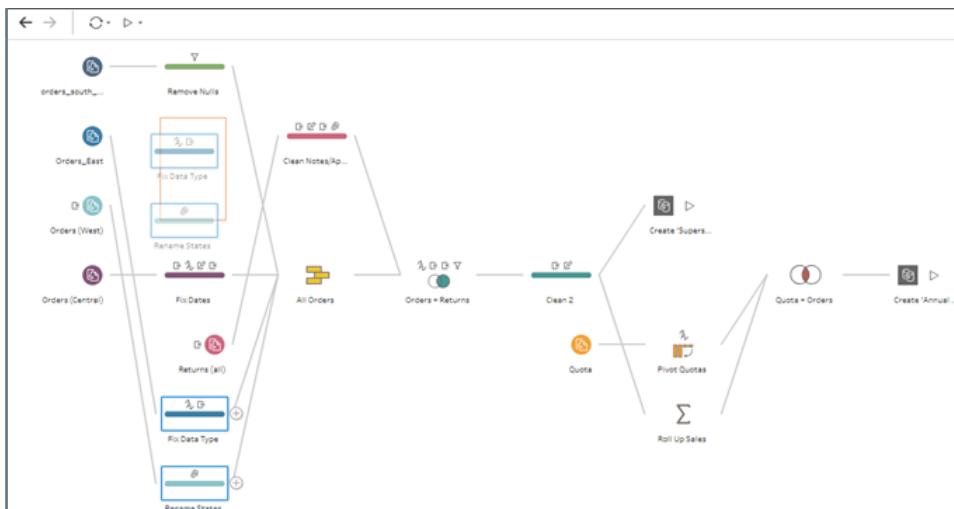
重新排列流程中的步骤

您是否曾经构建过复杂的大型流程，但 Tableau Prep Builder 的默认布局让流程很难执行？现在，您可以在画布中选择并移动步骤，以对您有意义的方式组织流程。例如，您可以修复交叉的流程线条，移动流程步骤以清理额外的空白，或者重新排列流程步骤以显示清晰的事件序列。

下面的流程混乱不清，难以执行：



使用这项新功能，您可以在流程中选择一个或多个步骤，并将其向上、向下、向左或向右拖放到所需位置。



注意: 存在一些限制，并且无法将流程步骤移到会中断从左到右处理流程的位置。将流程步骤拖到允许的位置时，将显示一个橙色的框。如果不允许该位置，则不会显示橙色框，当您尝试放置步骤时，它们将返回到原始位置。

有关详细信息，请参见[构建和组织流程](#) 在本页 179。

版本 2019.2.1 新增功能

连接到数据

使用新的数据库连接器连接到更多数据

我们增加了更多的连接器来帮助您连接到数据和清理数据。现在，您可以连接到存储在以下数据库中的数据：

- Amazon Athena
- 其他数据库 (JDBC)

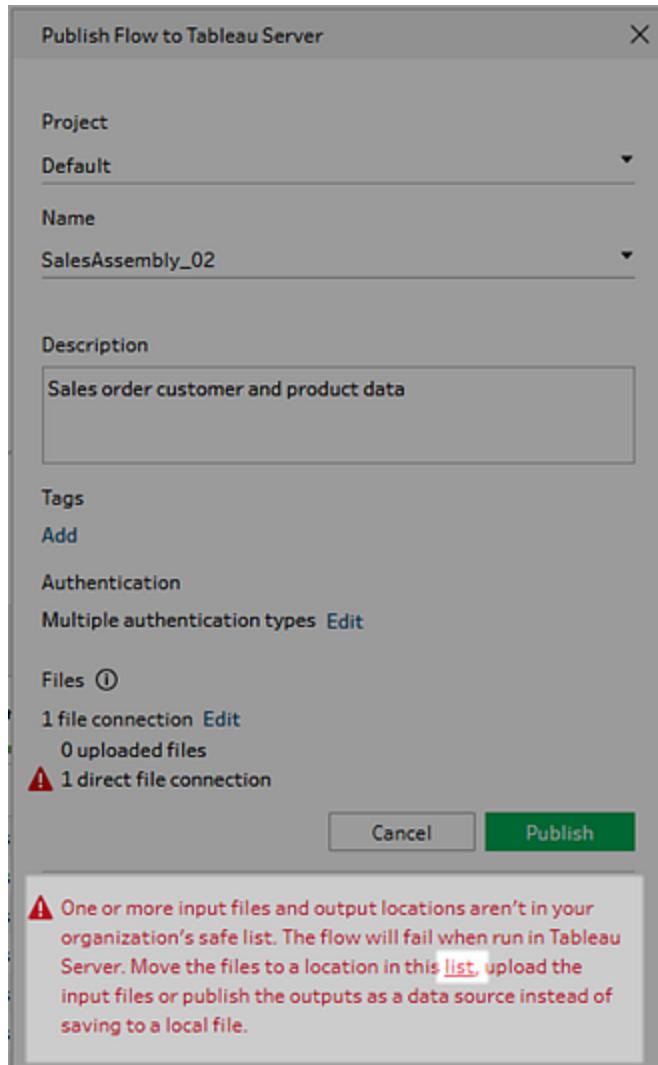
有关连接到特定连接器的详细信息，请参见 Tableau Desktop 和 Web 制作帮助的[支持的连接器](#)。有关连接到数据以构建流程的详细信息，请参见[连接到数据 在本页 141](#)。

使流程数据保持最新

发布流程时查看文件安全列表位置

您可以将流程发布到 Tableau Server 以自动按计划运行流程，并使用 Tableau Prep Conductor 刷新流程输入或输出。如果流程包括的输入或输出具有与网络共享的连接，您的系统管理员必须提前在组织的安全列表中加入文件位置，否则当您尝试在 Tableau Server 中运行流程时，流程将失败。

在 Tableau Prep Builder 版本 2019.2.1 及更高版本中，如果发布的流程包括连接到网络共享的输入或输出步骤，则在需要将文件包括在组织安全列表的情况下，您将在发布时收到通知。若要查看安全列出的位置，只需单击警告消息中的链接，然后将文件移动到显示的位置之一。



有关发布流程的详细信息，请参见[将流程发布到 Tableau Server 或 Tableau Online 在本页378](#)。有关如何将文件位置添加到组织安全列表中的详细信息，请参见[步骤 4: 安全列表输入和输出位置 在本页375](#)。

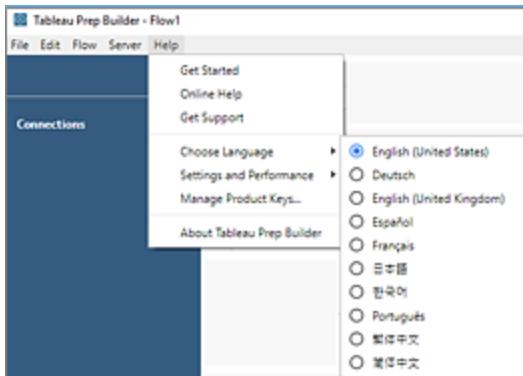
其他增强功能

为用户界面选择 10 种不同的语言

Tableau 增加了两种额外的语言，您可以从中进行选择来配置用户界面(菜单、消息等)：

- en_GB(英国英语)
- zh_TW(繁体中文)

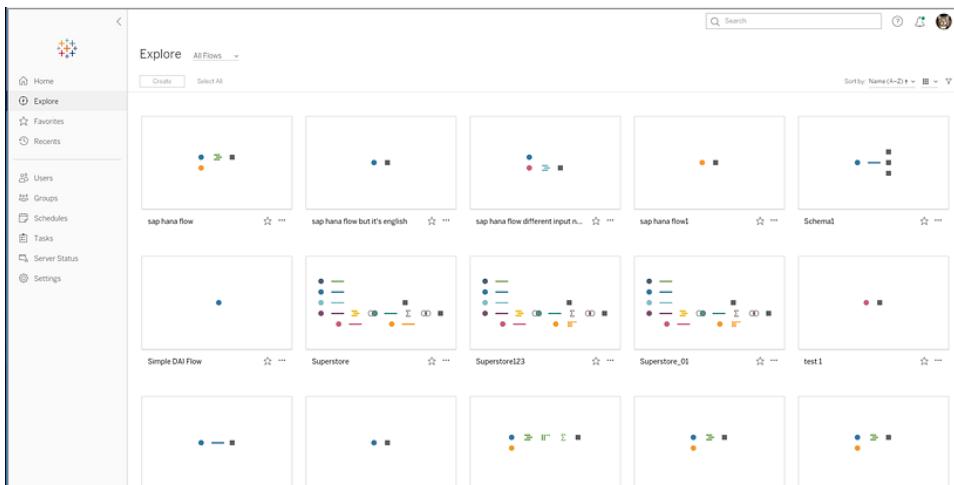
Tableau Prep 帮助



有关配置显示语言的详细信息，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的[设置显示语言](#)。

使用流程缩略图视图在 Tableau Server 上快速查找流程

将流程发布到 Tableau Server 时，您现在可以切换到缩略图视图并查看流程布局的图形表示形式，以帮助您一眼快速地找到流程。



版本 2019.1.4 新增功能

连接到数据

使用新的数据库连接器连接到更多数据

我们增加了更多的连接器来帮助您连接到数据和清理数据。现在，您可以连接到存储在以下数据库中的数据：

- SAP HANA
- Azure SQL 数据仓库
- MariaDB

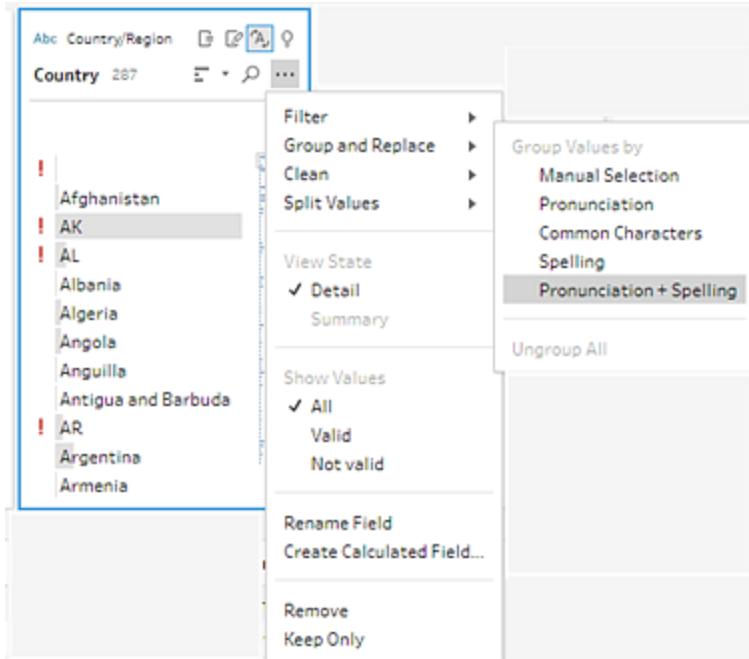
注意：使用 SAP HANA 连接到数据时，在打开流程并使用初始 SQL 查询不支持的连接时提示输入变量和参数。

有关连接到特写连接器的信息，请参见 Tableau Desktop 和 Web 制作帮助的[支持的连接器](#)部分中的连接详细信息。有关连接到数据以构建流程的详细信息，请参见[连接到数据 在本页 141](#)。

清理和调整数据

按数据角色对字段中的类似值进行分组

在 Tableau Prep Builder 版本 2018.2.3 中，我们引入了数据角色来帮助您对字段值进行标准化，并快速查找无效的值。现在，如果为字段分配数据角色，您可以使用该数据角色，通过数据角色定义的标准值来匹配值和对值进行分组。只需为字段分配数据角色，然后从“**分组和替换**”菜单中选择“**数据角色匹配**”。此新选项随后将根据拼写和发音将无效值与最相似的有效值匹配。



举例来说，如果有一个列出国家/地区(例如 Saint Martin 和 Saint Maarten)的字段，并且您为该字段分配“国家/地区”数据角色，则您可以使用新的“数据角色匹配”选项对字段进行分组，并将这些值映射到标准化数据角色值“Saint-Martin”。如果数据集中没有值的标准化拼写，则 Tableau Prep Builder 会自动为您将其添加到组中，并使用此标准值对匹配值进行分组。

The screenshot shows the 'Group Values by Pronunciation + Spelling' dialog for the 'City' field. The 'Albuquerque' group is selected, showing four members: Albuquerque, Albuerque, Albuquerue, and Albuqrque. Other cities listed include Albany, Albemarle, Albermale, Albertville, albrightsfield, Albrightserville, and so on. The dialog has a 'Done' button at the top right.

此功能适用于以下数据角色，当前仅支持美国英语。

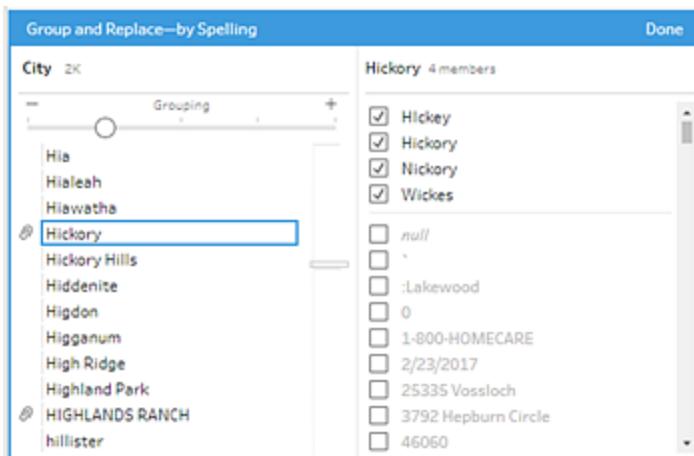
- 机场
- 城市
- 国家/地区
- 郡/县
- 州/省/市/自治区

有关将为字段分配数据角色或使用数据角色对值进行分组的详细信息，请参见[使用数据角色验证数据 在本页 209](#)。

在对字段值进行分组时调整结果

使用“**分组和替换**”清理选项按“**拼写**”或“**发音**”对类似值进行分组时，您现在可以使用字段上的简单滑块来调整分组参数的严格程度，从而使您能更大程度地控制结果。

从 5 个阈值设置中进行选择，以控制组中包括的值数量，以及创建的组数量。



有关使用模糊匹配对字段进行分组的详细信息，请参见[编辑字段值 在本页 239](#)。

版本 2019.1.3 新增功能

连接到数据

基于表关系快速联接数据

当连接到的数据库包含定义了表关系的表时(例如 Vertica 或 Microsoft SQL Server)时, Tableau Prep Builder 现在可以检测并显示此信息。

例如,现在可以看到表中的哪些字段是唯一标识符,哪些字段是字段(外键),以及这些字段的任何相关表。

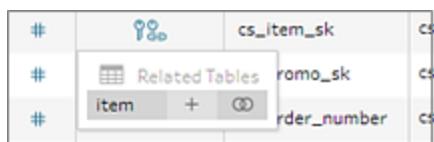
一个名为“**链接键**”的新列将显示在“输入”窗格中,唯一标识符字段、相关字段或两者的组合旁边会有图标。

The screenshot shows the Tableau Prep Builder interface with the 'catalog_sales' table selected. In the 'Input' tab, under the 'catalog_sales' table, there is a column labeled 'Linked Keys'. This column contains four rows, each with a checkmark in the first column and a small circular icon with a 'key' symbol next to it. The columns are labeled 'new name', 'original field name', 'changes', and 'sample'. The rows show:

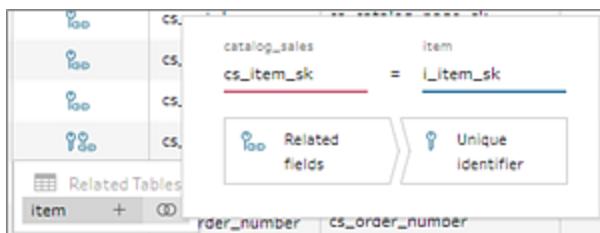
	new name	original field name	changes	sample
<input checked="" type="checkbox"/>	cs_sold_date_sk	cs_sold_date_sk	2,450,816, 2,4!	
<input checked="" type="checkbox"/>	cs_sold_time_sk	cs_sold_time_sk	33,151, 29,066	
<input checked="" type="checkbox"/>	cs_ship_date_sk	cs_ship_date_sk	2,450,886, 2,4!	
<input checked="" type="checkbox"/>	cs_bill_customer_sk	cs_bill_customer_sk	96,466, 49,302	

A tooltip above the table explains: "Linked keys identify one or more fields that link to other tables. Fields can be: A unique identifier (primary key) in this table. Related fields (foreign key) to a unique identifier in another table in the database. Both a unique identifier and related fields."

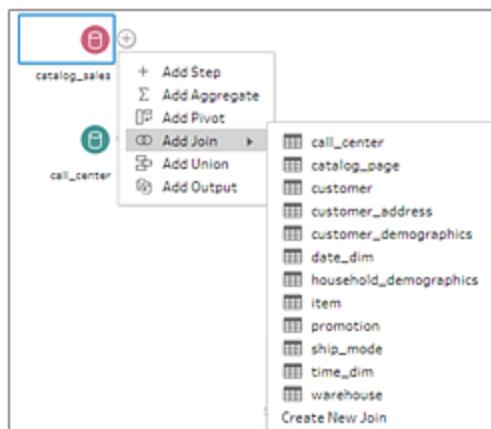
在“**链接键**”字段中,单击相关字段或组合字段以查看相关表,然后只需单击一次便可将相关表添加到流程或创建联接。



创建联接时,Tableau Prep Builder 将使用定义的字段关系来联接表,并显示定义联接子句的字段的预览。在这里,您可以单击新的“联接”步骤,以根据需要清理和编辑联接条件。



对于包含相关字段的表，您还可以查看所有相关表的列表，并通过“流程”窗格的菜单创建联接。从“添加联接”菜单选项中，从列表中选择要联接的表，Tableau Prep Builder 将基于组成两个表之间的关系的字段创建联接。

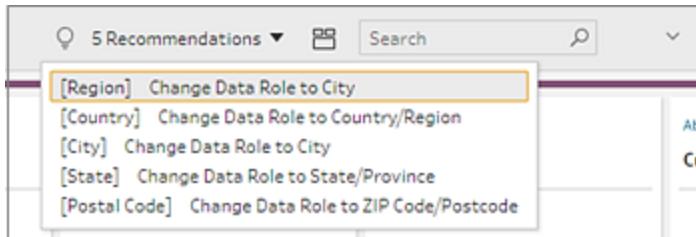


此功能可用于定义了表关系的任何受支持的数据库连接器。有关使用“**链接键**”来添加或联接表的详细信息，请参见[在“输入”步骤中联接数据\(版本 2019.1.3 及更高版本\)](#) 在本页 176。

清理和调整数据

在工具栏中查看所有清理建议

如果 Tableau Prep Builder 检测到并为您的字段提供了清理建议，则该字段的“配置”卡上会显示一个灯泡 图标。现在，您可以通过工具栏在一个方便的列表中查看所选步骤中任何字段的所有清理建议。



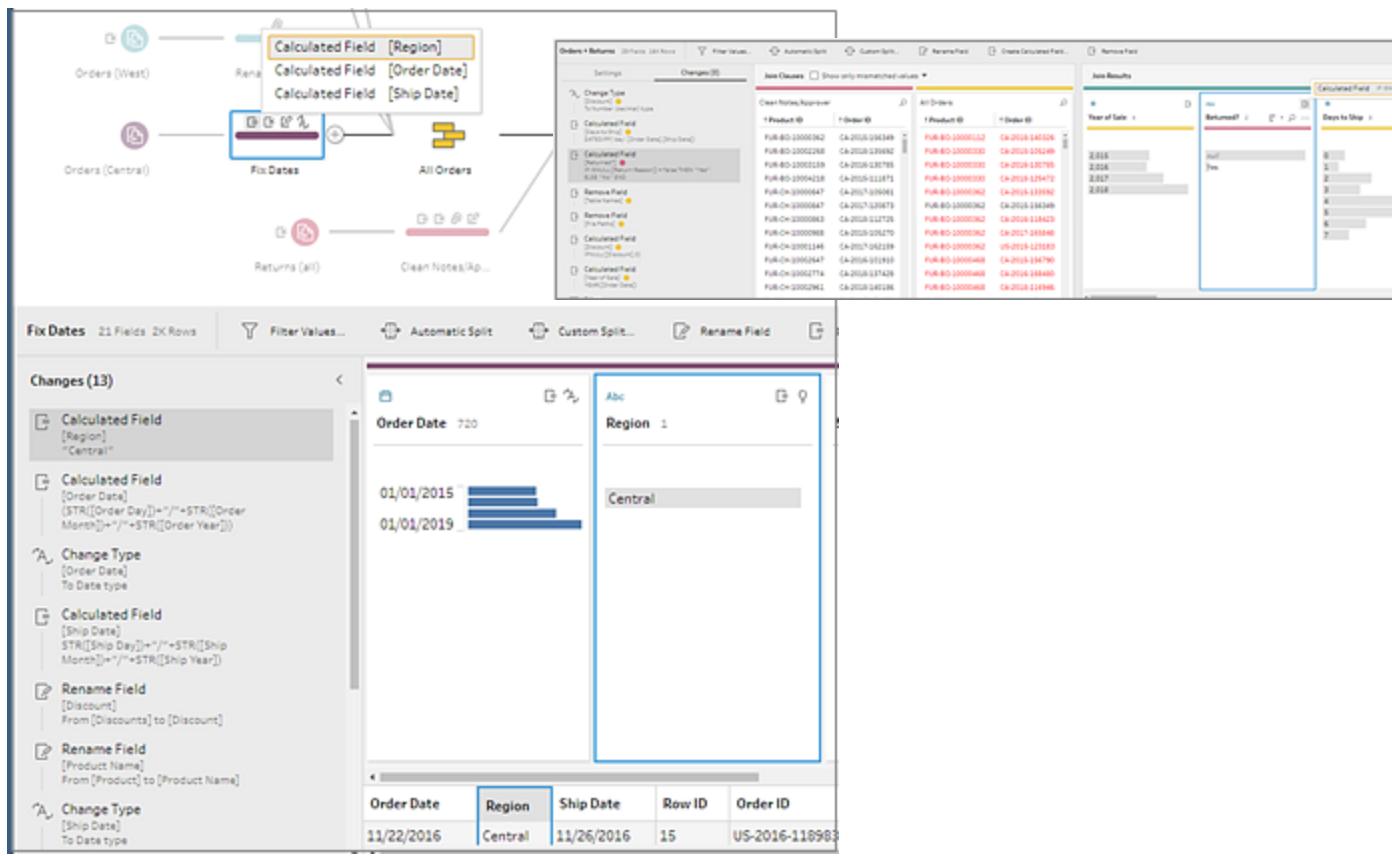
只需单击列表中的某项，Tableau Prep Builder 即可转到该字段，您可以在其中应用建议的更改。有关清理建议的详细信息，请参见[使用建议应用清理操作\(版本 2018.3.2 及更高版本\)在本页 236。](#)

注意：此选项不可用于“联接”、“输入”和“输出”步骤类型。

其他增强功能

查看流程中的特定更改

现在，通过单击“流程”窗格中某个步骤上的更改图标，或单击“配置”或“结果”窗格中的某个配置卡，您可以快速查找在流程中进行的特定更改。只需单击表示要查找的更改的图标，然后从更改列表中选择注释。更改将在“更改”窗格、“配置”或“结果”窗格以及数据网格(如果适用)中突出显示。

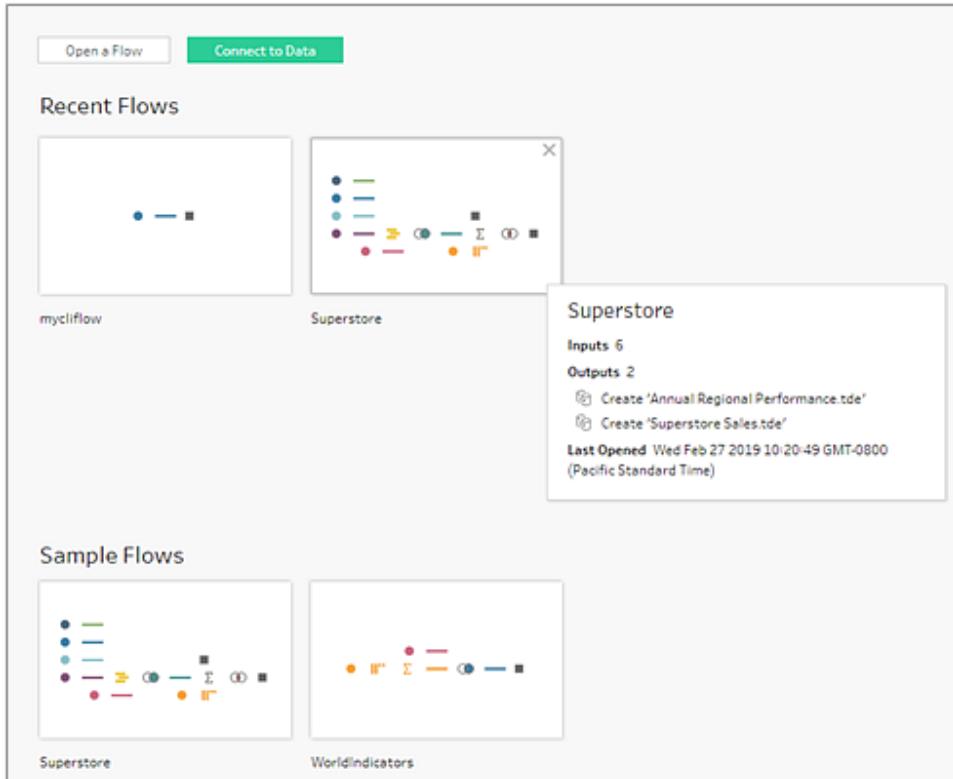


有关查看所做更改的详细信息，请参见[查看您所做的更改 在本页 232](#)。

通过“开始”页面管理最近查看的流程

当您保存流程并关闭 Tableau Prep Builder 时，该流程的缩略图将保存到“开始”页面，以便您可以快速找到最常用的流程。但作为一个数据准备摇滚巨星，您的“开始”页面可能很快就会拥挤不堪！为了帮助您快速找到所需的流程并在完成流程处理时减少杂乱程度，我们向“开始”页面添加了一些新功能。

- 流程缩略图上改进的图形可更准确地表示流程布局，从而帮助您一眼就能找到流程。
- 如果图形未提供足够的信息，请将鼠标指针悬停在一个流程缩略图上，以立即查看该流程的输入和输出详细信息，以及其上次打开时间。
- 通过从“开始”页面中移除不再需要的流程来清理杂乱情况。此选项仅从“**最近的流程**”视图中移除流程，而不会删除流程。



版本 2019.1.2 新增功能

使流程数据保持最新

将流程计划为在 Tableau Server 中自动运行

将新的数据管理加载项添加到 Tableau Server 2019.1 和更高版本，并且您可以使用 Tableau Prep Conductor 来计划和管理流程运行。将流程发布到 Tableau Server，然后安排任务运行流程，并自动更新流程输入和输出。使用错误和通知选项以及管理视图来跟踪和管理流程处理运行状况。

有关详细信息，请参见[使流程数据保持最新 在本页 345](#)。

帮助设计

进行更新了来帮助您导航内容，包括一个新的右侧菜单，用于快速扫描和选择页面上的不同主题。

版本 2019.1.1 新增功能

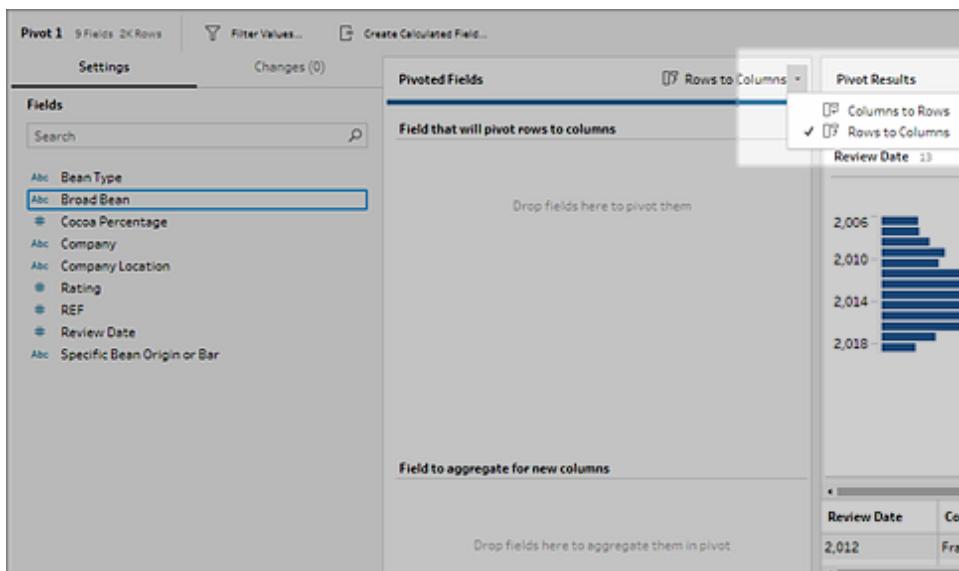
清理和调整数据

将行转置为列

如果您使用交叉表或电子表格中的数据，则其格式可能不是可在 Tableau 中轻松进行分析所需的格式。有时，您的数据可能有多个需要转换为行的数据列，或者您的列较少，但数据值却表示多个事物。

为了比以往更轻松地调整数据，您现在可以选择适合自己的转置选项。在转置步骤中，您现在可以选择以下选项之一：

- 列转置为行：**将一个或多个字段拖到“转置的字段”窗格以将列转置为行。
- 行转置为列：**将字段拖到“转置的字段”窗格以使用字段值创建新列，然后拖入字段以用于填充这些列，并选择所需的列聚合方式。



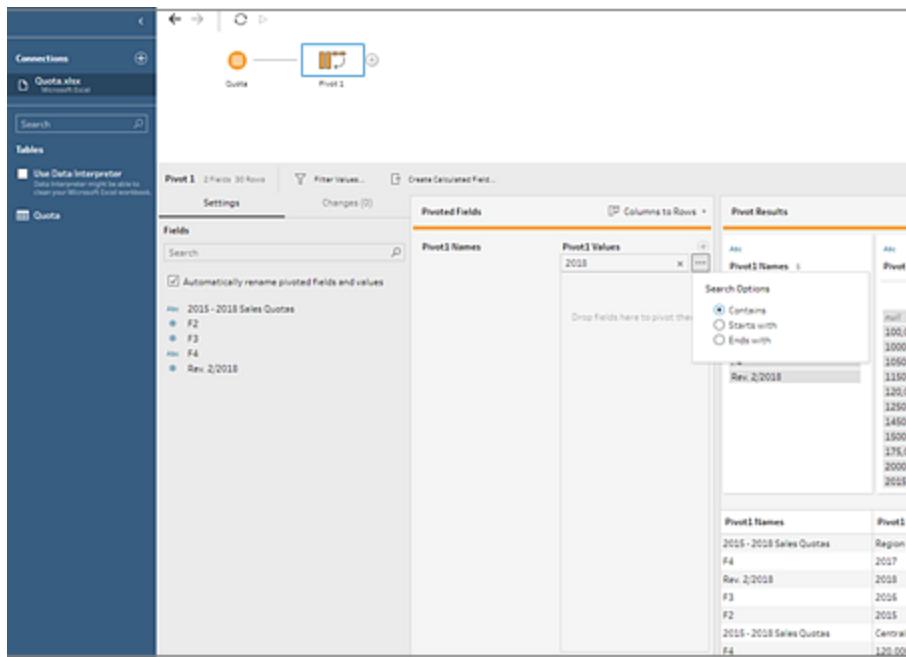
有关详细信息，请参见[转置您的数据](#) 在本页 273

使用通配符搜索来转置字段

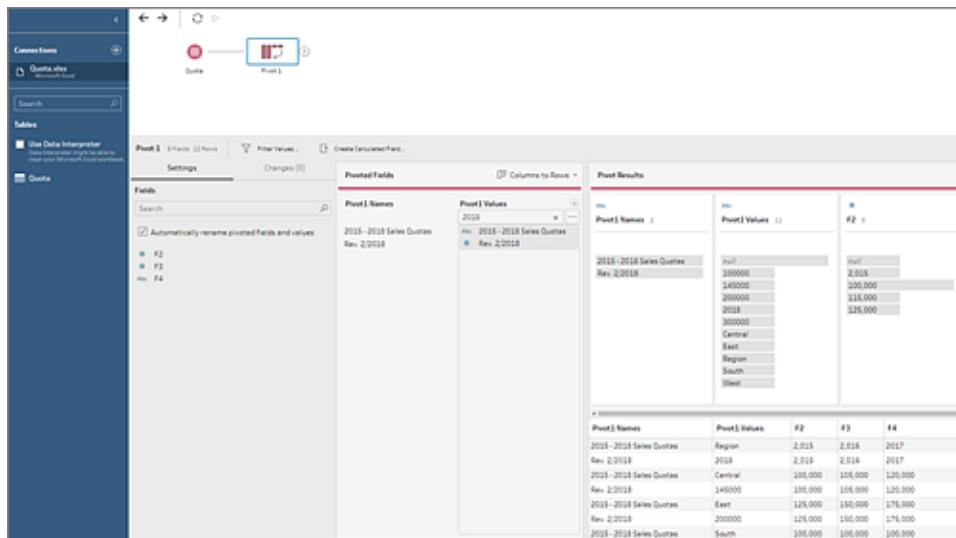
如果有包含大量字段的数据集，或者在一段时间内频繁添加或移除字段，您可以在将列转置为行时使用通配符搜索，以便根据通配符模式匹配来即时转置数据。如果添加或移除了与模式匹配的新字段，Tableau Prep Builder 会在运行流程时检测到架构更改，并且转置结果会自动更新。

Tableau Prep 帮助

若要使用通配符搜索来转置数据，请添加一个转置步骤，然后单击“**单击此处创建通配符转置**”链接打开通配符搜索模式以查找名称与所输入的值匹配的字段。单击“**搜索选项...**”按钮选择是否要包括其中包含所输入模式、以该模式开头或以该模式结尾的字段。



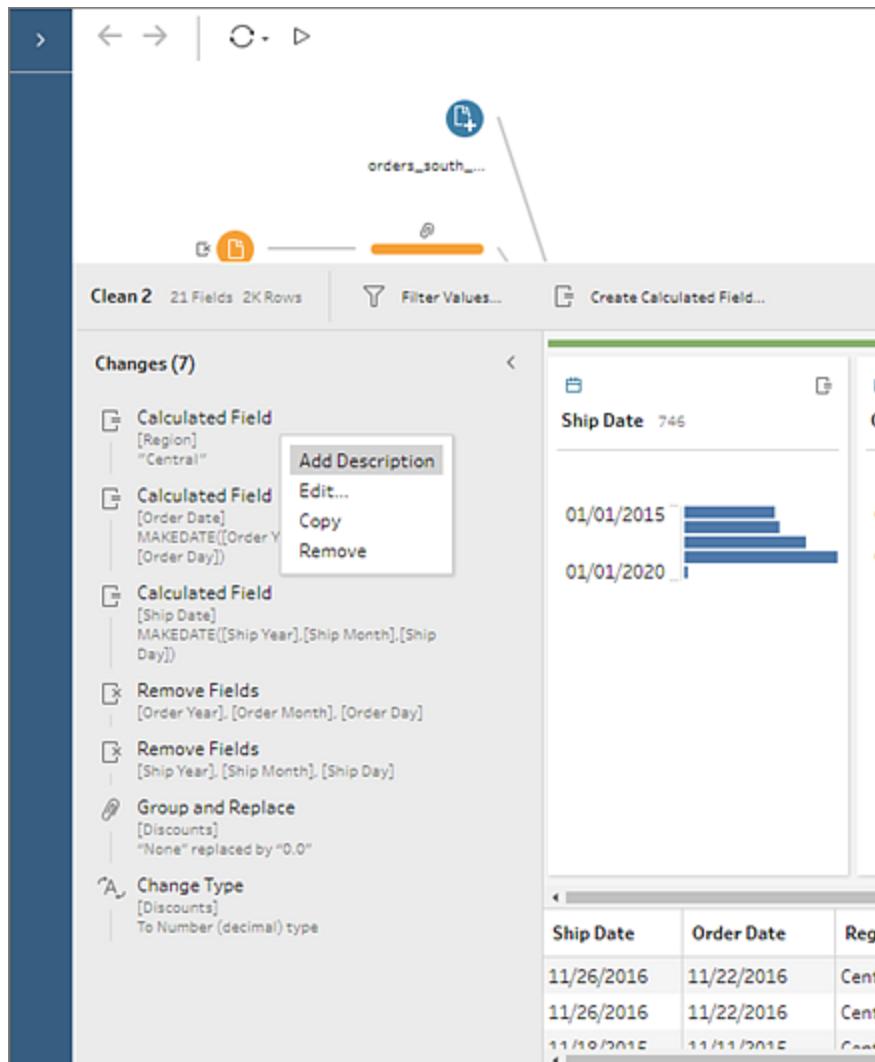
按 Enter 应用模式，Tableau Prep Builder 将自动转置字段。



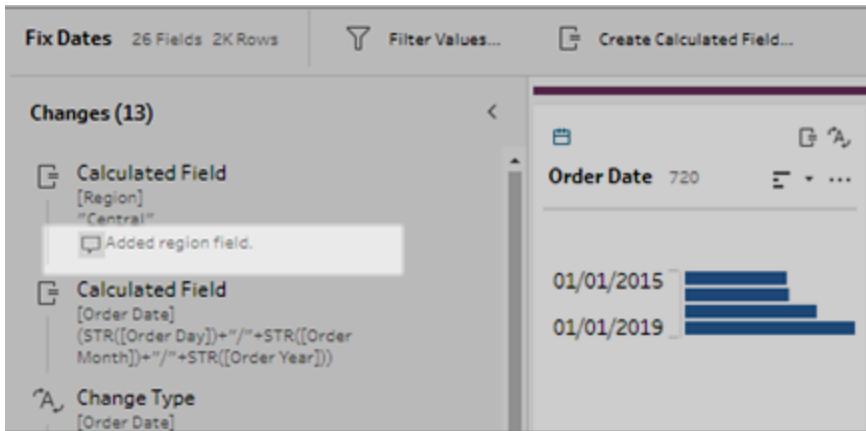
有关详细信息，请参见[转置您的数据](#) 在本页 273

使用说明为更改添加上下文

在“**更改**”窗格中为任何更改添加说明，以便为您的清理操作添加上下文。只需打开“**更改**”窗格，右键单击或者按住 Ctrl 并单击 (Mac) 列表中的某项更改，并选择“**添加说明**”。然后在出现的文本字段中键入说明。



说明将出现在为带有注释 图标的更改生成的文本的下方。

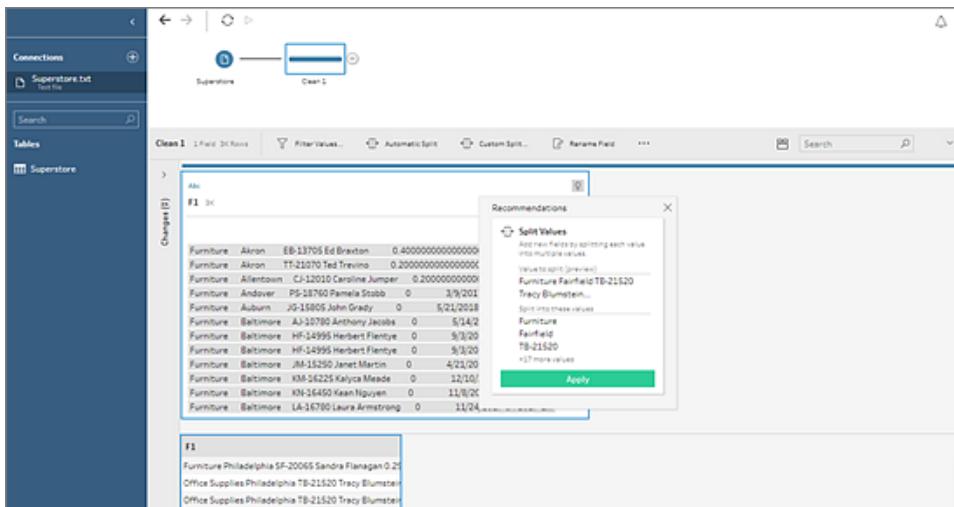


有关详细信息，请参见[为流程步骤和清理动作添加说明](#) 在本页 186。

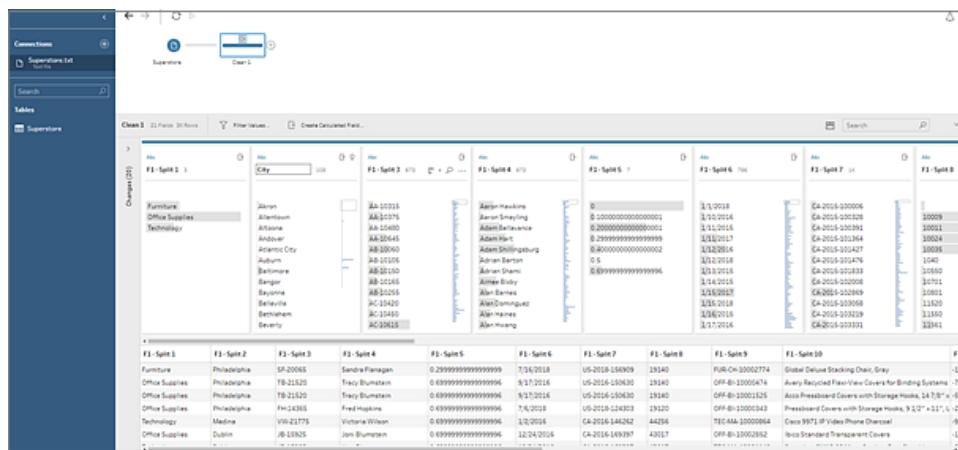
使用清理建议来快速拆分固定宽度文本文件中的数据

在以前的版本中，如果您使用存储在固定宽度文本文件(即包含的数据值字符串受固定字符限制所限的文件)的数据，则您必须对文件执行一些预清理操作，或者使用复杂的计算将值分隔出来，放在单独的列中。

现在，当您连接到这种类型的数据源时，清理建议功能将检测到这些类型的列，并为您提供拆分选项建议。只需连接到数据源，然后添加清理步骤即可。在“配置”窗格中，单击配置卡上的建议图标。



单击“应用”以拆分字段，然后检查结果并根据需要清理或重命名每个新字段。



有关详细信息，请参见[使用建议应用清理操作\(版本 2018.3.2 及更高版本\)](#) 在本页 236。

版本 2018.3.3 新增功能

连接到数据

添加新文件时自动刷新输入并集

如果在输入步骤中使用通配符模式来查找和合并 Microsoft Excel 或文本 (.csv) 文件，则在下次您打开流程或通过命令行运行流程时，添加到同一文件夹中、与模式匹配的新文件会自动包括在并集中。

注意：打包流程文件 (.tflx) 不会自动获取新文件，因为文件已与流程打包在一起。若要包括新文件，请打开流程文件 (.tfl) 以获取新文件，然后重新打包流程以包括新文件数据。

有关在输入步骤中合并文件的详细信息，请参见[在输入步骤中合并文件和数据库表](#) 在本页 170。

清理和调整数据

重用清理操作

在清理数据时，您通常会在整个流程中反复执行相同的清理操作或动作。为了使数据的清理和调整更加高效，我们添加了在整个流程中复制和粘贴这些操作或动作的功能，这样您就能执行清理操作或动作一次，然后在需要的地方重用该操作或动作。

- 复制一个或多个步骤以在流程的另一个区域中使用它们。此选项不可用于包括通配符并集的输入步骤。
- 从一个步骤的“**更改**”窗格中复制操作，然后将其粘贴在另一个步骤的“**更改**”窗格中，从而在该步骤中应用该相同操作。
- 将一个操作从“**更改**”窗格拖放到“**配置**”窗格中的一个字段，以将该操作应用于其他字段。您也可以在“**更改**”窗格中复制操作，并将其拖放到步骤中的其他字段，以将该操作应用于多个字段。此选项不可用于影响多个字段(例如计算字段)的操作。

有关详细信息，请参见[清理和调整数据 在本页223](#)。

移除额外的空格

我们添加了两个新的快速清理操作，帮助您移除字段值中额外的空格：

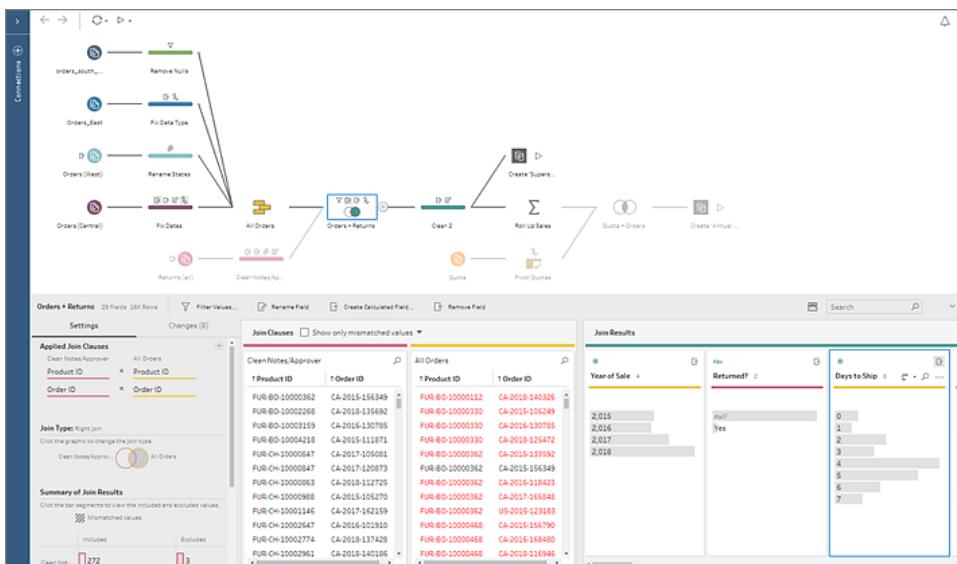
- **移除额外的空格**: 将此选项与文本字段值结合使用以移除前导和尾随空格，以及将字符之间的额外空格替换为单一空格。
- **移除所有空格**: 使用此选项从文本值中移除所有空格，包括前导和尾随空格以及字符之间的任何空格。

有关详细信息，请参见[编辑字段值 在本页239](#)。

检查和筛选数据

快速查找字段在流程中的使用之处

在构建流程时，如果未看到所有期望的字段值，您现在可以轻松地跟踪字段的每处使用位置 - 甚至能追溯到字段的来源。只需在一个步骤的“**配置**”窗格或“**结果**”窗格中选择一个字段，流程窗格中的步骤将突出显示。这将向您展示字段在流程中所采用的路径，从而能够比以往更为轻松地排查流程的故障。



有关详细信息，请参见[在流程中跟踪字段](#) 在本页 201。

其他增强功能

同时打开多个 Tableau Prep Builder 窗口

在 Windows 中或 Mac 上，单击“文件”>“新建”打开一个新流程窗口，而不是替换您正在其中工作中的当前窗口。打开流程文件(.tfl 或 .tflix)时也将打开一个新窗口，这样您就能轻松地同时处理或比较多个流程。

安装之后关闭使用情况报告

为了帮助我们改进 Tableau Prep Builder，您可以与我们共享使用情况模式数据。我们决不会收集任何具体用户数据，只会收集任何匿名使用情况数据。例如，数据源连接类型(例如.csv、sql 或 Amazon Redshift)、大数据集与小数据集、流程中的步骤数或使用的操作类型(但不会收集数据细节，例如步骤名称或使用的公式)，诸如此类。

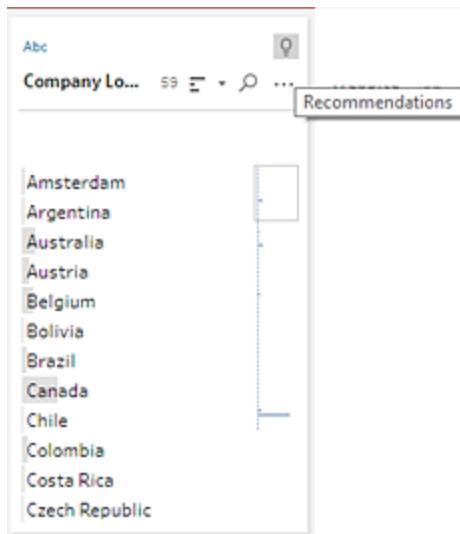
您的数据是机密的，我们决不会与任何人共享此信息。但如果要在安装之后关闭此选项，您可以这样做。转到“帮助”>“设置和性能”，并单击“**共享使用情况数据以改进 Tableau Prep Builder**”以清除复选标记。有关详细信息，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的[使用情况报告](#)。

版本 2018.3.2 新增功能

清理和调整数据

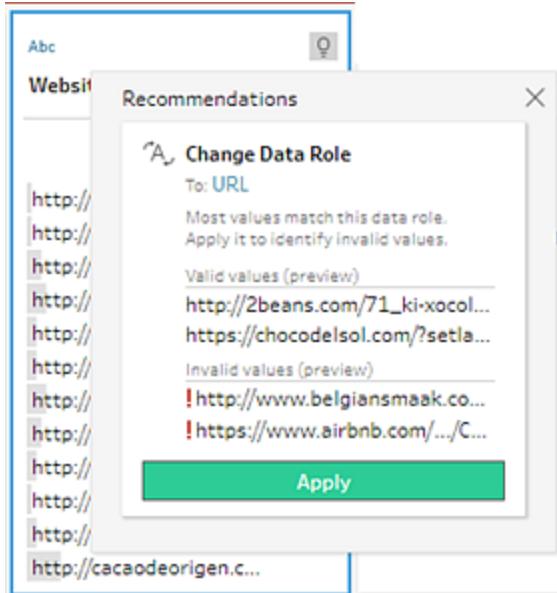
使用清理建议快速提高数据质量

为了帮助您快速找出并修复数据问题，Tableau Prep 现在可以检查数据字段并提出更改建议，您可以进行这些更改来修复问题。Tableau Prep 会检查您的数据字段，如果确定了建议的更改，字段的“配置”卡上会出现一个灯泡图标。



单击灯泡图标可查看建议的更改，然后将鼠标指针悬停在建议卡上并单击“应用”以将更改自动应用到字段。更改记录在“更改”窗格中，并且可随时移除。

在下面的示例中，Tableau Prep 将字段值确定为 URL，并建议为字段应用数据角色来帮助确定无效的值。



除“输入”、“输出”和“联接”步骤类型外的所有步骤类型中提供了此功能。此版本中的建议类型包括：

- 数据角色
- 筛选
- 将无效值替换为 Null 值
- 移除字段
- 剪裁空格

有关详细信息，请参见[使用建议应用清理操作\(版本 2018.3.2 及更高版本\)](#) 在本页 236。

其他增强功能

通过命令行运行流程时更加灵活

通过命令行运行流程时，如果您连接到数据库或将输出文件发布到服务器，则需要一个包括所有必需凭据的凭据.json 文件。在过去的版本中，如果只连接到数据库或将流程输出发布到服务器，您仍然需要在凭据.json 文件中包括空数组，否则流程将会失败。

现在，您不再需要在凭据.json 文件中包括空数组。举例来说，如果只连接到数据库，您不再需要在.json 文件中包括 outputConnection 数组。

有关如何通过命令行运行流程的详细信息，请参见[通过命令行刷新流程输出文件](#) 在本页 321。

性能改进

为了帮助您更快更高效地使用流程，我们增强了以下功能：

- 关闭产品时会立即归还 Tableau Prep 使用的内存空间。
- 如果遇到意外错误，您现在可以重试操作或过程以使您保持在流程中。
- 加载数据或在流程中查看数据时性能提高，可缩短处理时间。
- 如果连接到以下连接器，您现在可以使用搜索字段来快速查找架构：
 - Amazon EMR Hadoop Hive
 - Cloudera Hadoop
 - Hortonworks Hadoop Hive
 - MapR Hadoop Hive
 - Presto
 - Spark SQL

早期版本新增功能

版本 2018.3.1 新增功能

安装和部署 Tableau Prep

- [更改默认存储库位置](#) 下一页

连接到数据

- [清理存储在 PDF 文件中的数据](#) 下一页
- [适用于数据库表的通配符并集](#) 下一页

清理和调整数据

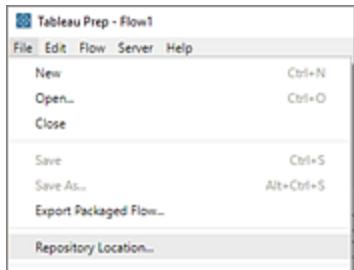
- [将字段值替换为 Null](#) 在本页 66
- [在任何步骤中应用清理操作](#) 在本页 66

安装和部署 Tableau Prep

更改默认存储库位置

默认情况下，当您安装 Tableau Prep 时，“我的 Tableau Prep 存储库”自动安装在计算机上的您的“文档”文件夹中。但如果您可存储流程和数据源的位置存在相关限制，您现在可以在安装 Tableau Prep 之后将“我的 Tableau 存储库”转移到其他目录。

若要移动该目录，请打开 Tableau Prep，然后从顶部菜单中选择“文件”>“存储库位置”。选择要将存储库转移到其中的文件夹。



有关详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[更改“我的 Tableau Prep 存储库”位置](#)。

连接到数据

清理存储在 PDF 文件中的数据

分析存储在 PDF 文件中的数据可能很难处理。现在您可以直接连接到 PDF 文件并使用 Tableau Prep 清理数据。

有关如何连接到 PDF 文件中的数据的详细信息，请参见 Tableau Desktop 帮助中的[PDF 文档](#)。

注意：数据连接器不向后兼容。包括这些连接器的流程可在以前版本的 Tableau Prep 中打开，但除非移除这些数据连接器，否则将会出错或无法运行。

适用于数据库表的通配符并集

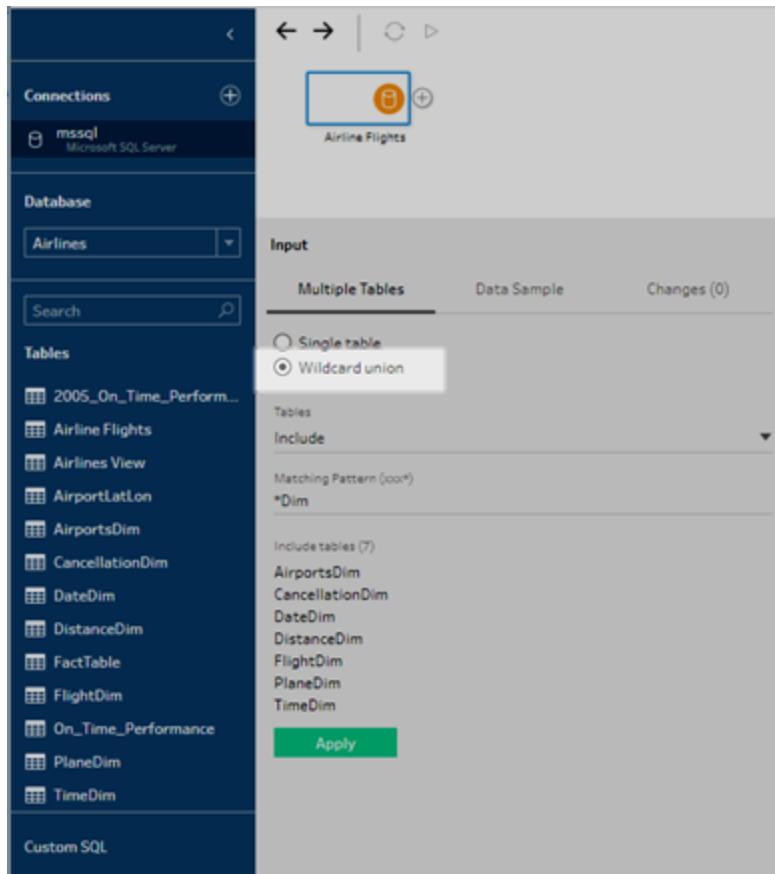
如果要处理存储在单个数据源中的数据库表，您现在可以使用通配符搜索在输入步骤中合并表。随后，您可以使用与在“输入”步骤中合并多个文件相同的步骤，在“输入”步骤中包括选定表中的所有数据。

Tableau Prep 帮助

若要使用这种类型的并集，各个表必须位于同一数据源中，并且您必须连接到支持通配符并集的数据源。

以下数据源支持这种类型的并集：

- Amazon Redshift
- Microsoft SQL Server
- MySQL
- Oracle
- PostgreSQL

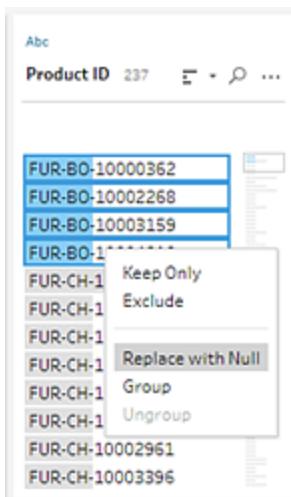


有关详细信息，请参见[在输入步骤中合并文件和数据库表](#)在本页170。

清理和调整数据

将字段值替换为 Null

如果有要包括在分析中的数据行，但想要排除特定字段值，您可以将这些值更改为 Null 值。只需选择要替换的字段值，然后右键单击(在 Mac 上按住 Ctrl 单击)所选的值，并从菜单中选择“替换为 Null”。



有关替换字段值的详细信息，请参见[编辑字段值](#) 在本页 239。

在任何步骤中应用清理操作

修复流程中的问题更加高效。现在，如果您在联接、合并、转置或聚合时发现数据中的问题，您可以直接在该步骤中修复这些问题，而不必添加额外的清理步骤。

由于操作顺序很重要，因此系统会将清理操作放在联接、合并、转置或聚合步骤之前或之后，并在该步骤的“更改”窗格中跟踪此信息。

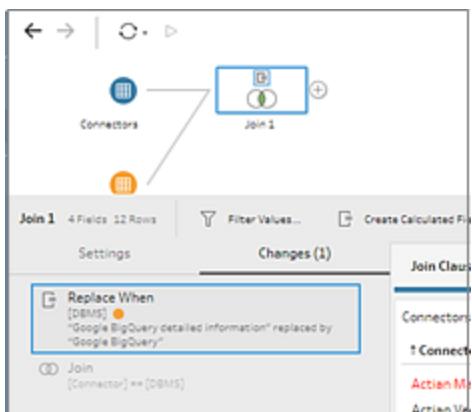
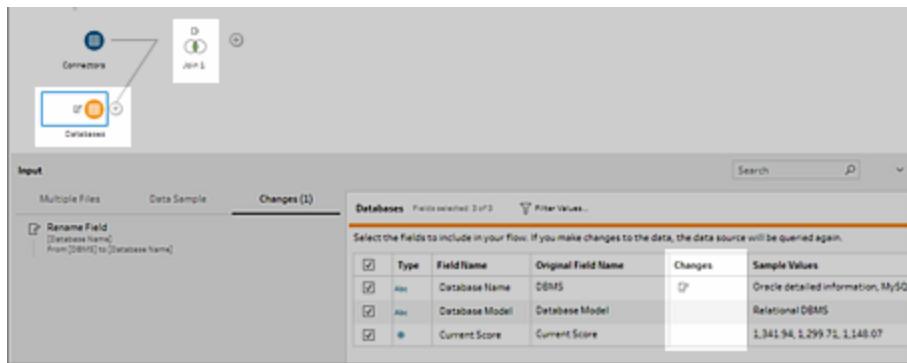


Tableau Prep 帮助

还会将更改的注释添加到“流程”窗格中的步骤。对于“输入”步骤，注释显示在步骤的左侧，并会添加到“输入”字段列表，以便您能够轻松地查看哪个字段应用了更改。对于所有其他步骤类型，注释显示在“流程”窗格中步骤的上方。



某些清理操作仍然必须在清理步骤中执行，只有部分清理操作可用于“输入”和“输出”步骤。

下表列出了可用于每种步骤类型的清理操作：

	输入	清理	聚合	转置	联接	合并	输出
筛选	X	X	X	X	X	X	
分组和替换		X	X	X	X	X	
清理		X	X	X	X	X	
拆分值		X	X	X	X	X	
重命名字段	X	X	X	X	X	X	
编辑值		X	X	X	X	X	
更改数据类型	X	X	X	X	X	X	
创建计算字段		X	X	X	X	X	
移除字段	X	X	X	X	X	X	

有关详细信息，请参见[关于清理操作 在本页 223](#)。

版本 2018.2.3 新增功能

连接到数据

- 连接到存储在 **MongoDB Business Intelligence (BI)** 中的数据 向下

检查和筛选数据

- 为字段值指定数据角色 向下

联接或合并数据

- 在联接子句中直接修复不匹配的字段 在本页 70

连接到数据

连接到存储在 **MongoDB Business Intelligence (BI)** 中的数据

如果将数据存储在 **Mongo DB Business Intelligence** 中，您现在可以连接到数据并使用 **Tableau Prep** 对其进行清理。

有关如何使用 **Mongo DB Business Intelligence** 连接到数据的详细信息，请参见 **Tableau Desktop** 帮助中的 **MongoDB BI 连接器**。

注意: 数据连接器不向后兼容。包括这些连接器的流程可在以前版本的 **Tableau Prep** 中打开，但除非移除这些数据连接器，否则将会出错或无法运行。

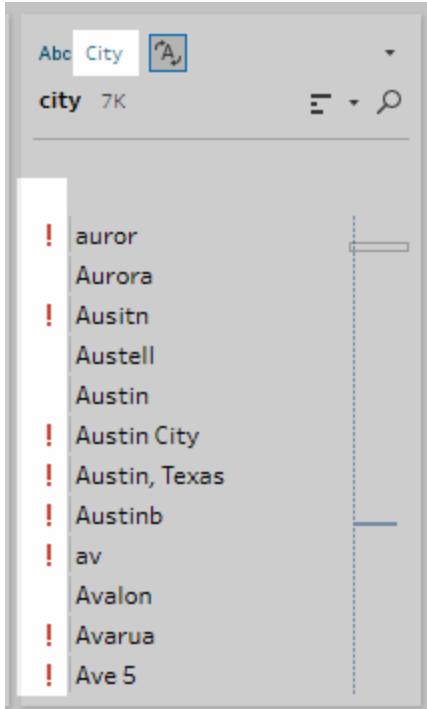
检查和筛选数据

为字段值指定数据角色

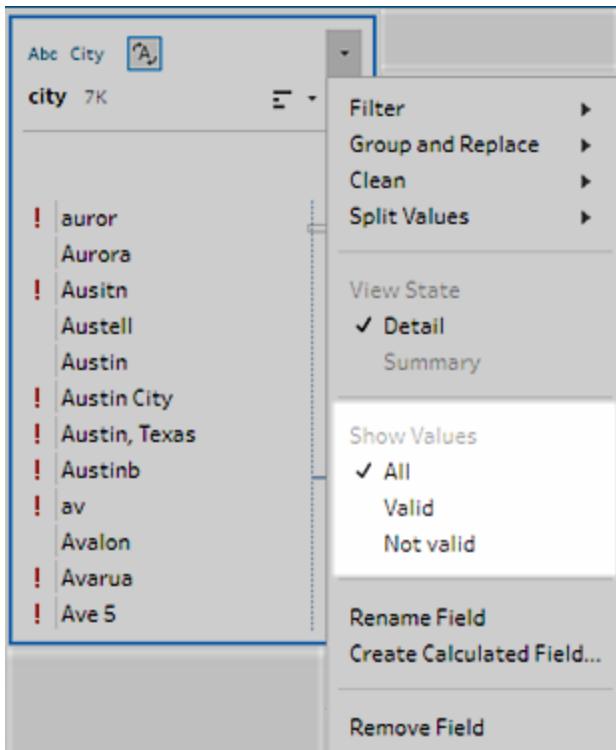
您现在可以为字段值分配数据角色，并使用 **Tableau Prep** 帮助您查找数据集中的不准确之处或离群值。

数据角色告知 **Tableau Prep** 字段值的含义或代表的内容，例如电子邮件地址或者诸如城市或邮政编码等地理角色。为字段分配数据角色之后，**Tableau Prep** 可检查字段值并标记不匹配的值，以便您能够进行更仔细的检查。

Tableau Prep 帮助



若要仅查看有效或无效的值，请使用下拉菜单上的新筛选器选项：



在此版本中，我们支持以下数据角色：

- 电子邮件
- URL
- 地理角色(基于当前地理数据，并且与 Tableau Desktop 使用的数据相同)
 - 机场
 - 区号(美国)
 - CBSA/MSA
 - 城市
 - 国会选区(美国)
 - 国家/地区
 - 郡/县
 - NUTS 欧洲
 - 州/省/市/自治区
 - 邮政编码

有关详细信息，请参见[使用数据角色验证数据](#) 在本页209。

联接或合并数据

在联接子句中直接修复不匹配的字段

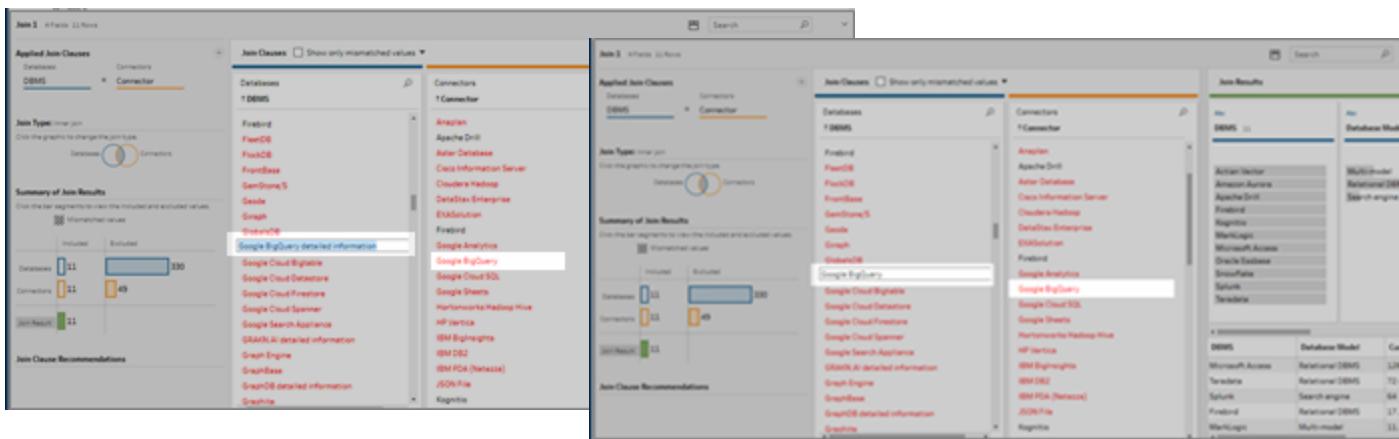
在联接两个数据表时，通常会有一些相同但由于数据录入错误而不匹配的字段值。Tableau Prep 可将不匹配的字段值标为红色，从而帮助确定联接子句中不匹配的字段。但如果能够在联接子句中修复这些字段值，岂不是更好？现在就能这样做了。

现在，当您联接两个数据集时，您可以直接在联接子句中对要联接的字段值进行编辑来修复它们。只需在字段值中双击即可对其进行编辑。

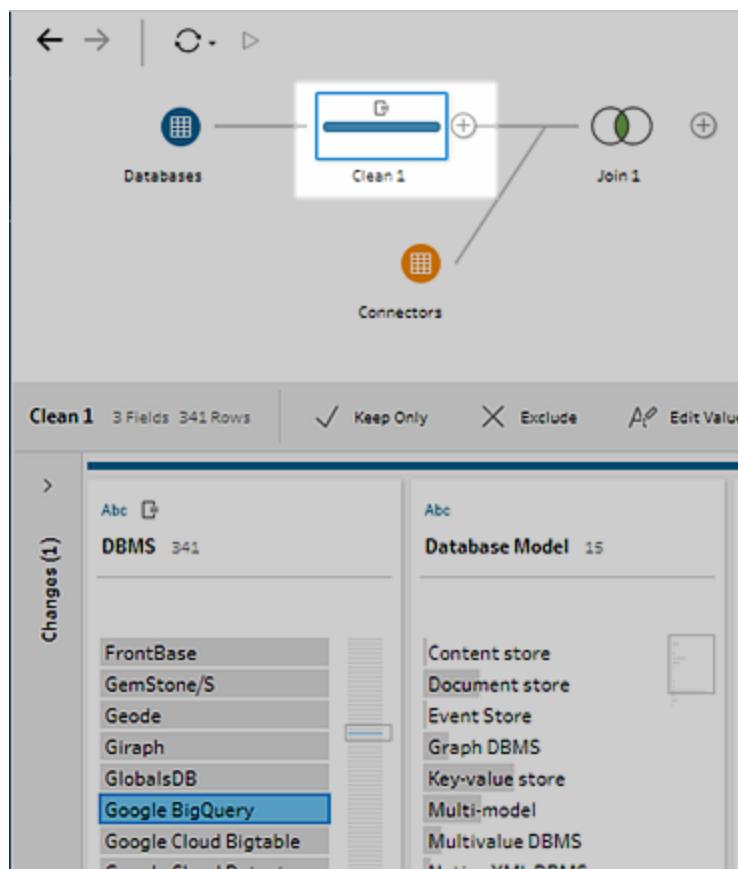
选择要更改的字段

内联编辑

Tableau Prep 帮助



系统会为相应的联接数据集将清理操作自动推回到流程中的上一个清理步骤。



联接步骤之前没有清理步骤？没问题。Tableau Prep 会自动为您添加一个清理步骤，以便从联接中捕获清理操作。

有关如何直接在联接子句中清理字段值的信息，请参见[聚合、联接或合并数据 在本页 297。](#)

版本 2018.2.2 新增功能

安装和部署 Tableau Prep

- [设置显示语言 向下](#)

连接到数据

- [连接到 Microsoft Access 中存储的数据 对页](#)

清理和调整数据

- [更改流程步骤的配色方案 对页](#)
- [为步骤添加说明 在本页 74](#)
- [使用模糊匹配来查找并修复拼写错误 在本页 75](#)

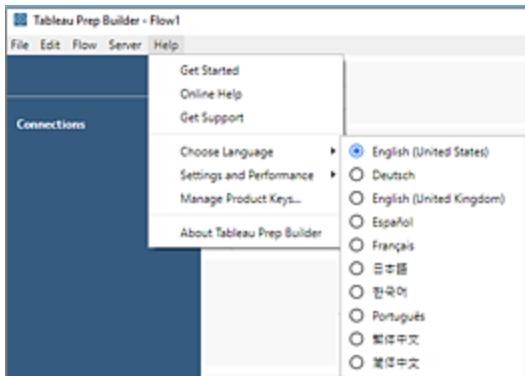
保存和共享工作

- [从命令行中运行流程 在本页 75](#)

安装和部署 Tableau Prep

设置显示语言

启动 Tableau Prep 时，它现在将检测计算机上设置的区域设置，并以相应的语言显示用户界面、日期和数字格式。如果要更改用户界面的显示语言，您可以从顶部菜单中的“帮助”>“选择语言”下选择支持的语言。



有关详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[设置显示语言](#)。

连接到数据

连接到 Microsoft Access 中存储的数据

如果使用 Microsoft Access 进行数据录入和存储数据表，您现在可以通过文件浏览器连接到（版本 2007 或更高版本中的）Access 文件，并使用 Tableau Prep 清理数据。

此连接器需要 64 位驱动程序。

有关如何将 Tableau Prep 连接到数据的详细信息，请参见 Tableau Desktop 帮助中的 [Access](#)。

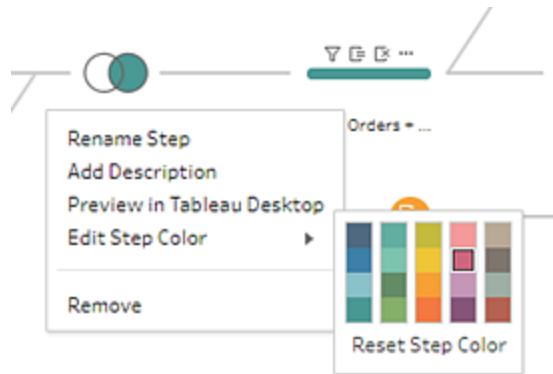
注意：数据连接器不向后兼容。包括这些连接器的流程可在以前版本的 Tableau Prep 中打开，但除非移除这些数据连接器，否则将会出错或无法运行。

清理和调整数据

更改流程步骤的配色方案

默认情况下，Tableau Prep 会为流程中的每个步骤分配一种颜色，帮助您在构建流程时更轻松地跟踪对数据所做的更改。但您可以选择配色方案。

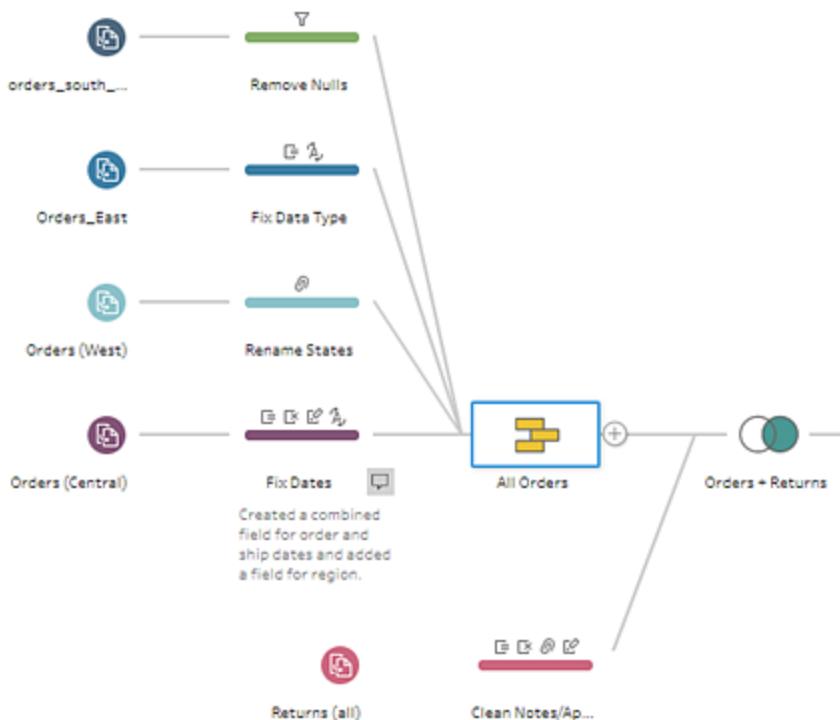
您现在可以从调色板中进行选取来更改一个或多个步骤的配色方案。只需在“流程”窗格中选择要更改的步骤，右键单击所选步骤，并从上下文菜单中选择“[编辑步骤颜色](#)”。



不喜欢它的外观？从顶部菜单中单击“撤消”，或从调色板菜单中选择“重置步骤颜色”。有关详细信息，请参见[更改流程配色方案](#)在本页 185。

为步骤添加说明

如果与其他人共享流程，传达您所做的更改以及相关原因可能很麻烦。为了简化这一点，您现在可以为流程中任何单独的步骤添加简短说明，说明将直接显示在流程窗格中。



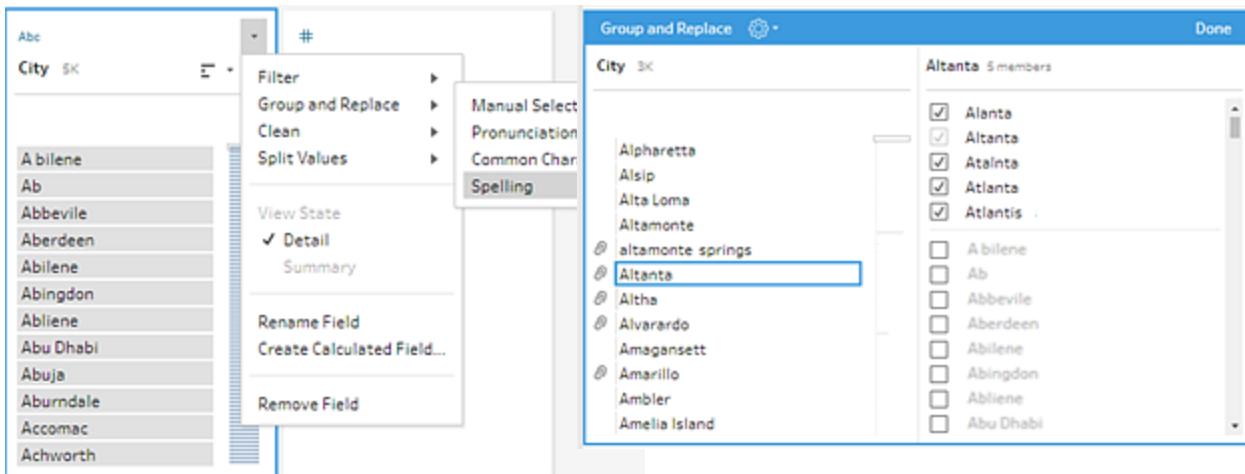
有关为流程步骤添加说明的详细信息，请参见[为流程步骤和清理动作添加说明](#)在本页 186。

使用模糊匹配来查找并修复拼写错误

为了帮助您快速确定并清理由于拼写错误而导致的相同值的多个变体，我们为模糊匹配清理功能添加了另一个选项。使用新的“**拼写**”选项来查找由于多出或缺少字母而不同的文本值，并将其分组。此选项适合于任何支持的语言。

“**拼写**”选项使用 Levenshtein 距离算法来计算两个文本值之间的编辑距离，然后在编辑距离小于默认阈值时将它们分组在一起。

有关详细信息，请参见[编辑字段值](#) 在本页 239。



保存和共享工作

从命令行中运行流程

为了使数据在 Tableau Prep 中保持最新，您会运行流程。但是，每次要运行流程时都打开 Tableau Prep 可能很费时间。为了帮助简化此流程，您现在可以从命令行中运行流程，而不必将其在 Tableau Prep 中打开。

若要从命令行中运行流程，您将需要：

- 要运行的流程 (.tfl) 文件的路径。
- 一个 .json 文件，其中包含流程为其输入步骤连接到的任何数据库的数据凭据，以及在其中发布输出的服务器的凭据。

如果流程为输入或输出连接到或发布至本地文件或存储在网络共享上的文件，则不需要此文件。

注意: 不支持连接到或发布存储在网络共享上的受密码保护的文件。

- 运行流程的计算机上的管理员权限。

此选项在 Windows 计算机(支持任务计划程序) 和 Mac 计算机上均可用。若要使用此过程，您需要 Tableau Prep 的激活版本，并且必须在安装了 Prep 的同一台计算机上运行该过程。

有关如何从命令行中运行流程的信息，请参见[通过命令行刷新流程输出文件 在本页 321](#)。

版本 2018.2.1 新增功能

安装和部署 Tableau Prep

- [通过命令行停用 Tableau Prep 对页](#)
- [使用虚拟桌面支持来优化 Tableau Prep 安装 对页](#)

连接到数据

- [连接到云数据源和 Hadoop Hive 对页](#)

浏览数据

- [用于只保留所需数据的新筛选器选项 在本页 78](#)

清理和调整数据

- [在计算字段中使用 ISO-8601 日期标准 在本页 78](#)
- [在数据网格中应用清理操作 在本页 79](#)
- [使用多选对“配置”窗格中的值进行分组 在本页 79](#)
- [其他增强功能 在本页 80](#)

联接或合并数据

- [使用并集建议来清理不匹配的字段 在本页 81](#)
- [为所有联接类型确定不匹配的字段 在本页 82](#)

安装和部署 Tableau Prep

通过命令行停用 Tableau Prep

像 Tableau Desktop 一样，如果您的计算机上不再需要 Tableau Prep，您现在可以使用 `-return` 选项通过命令行将其停用。

有关详细信息，请参见[停用产品密钥](#)。有关命令行中可用于 Tableau Prep 的其他安装程序属性的详细信息，请参见[部署 Tableau Desktop](#) 中的“安装程序选项”及相关部分。

使用虚拟桌面支持来优化 Tableau Prep 安装

就像 Tableau Desktop 一样，您现在可以配置虚拟桌面支持，针对非持久性虚拟桌面或定期重新映像的计算机优化 Tableau Prep 的安装。借助虚拟桌面支持，系统会使用 Tableau 托管的“授权运行”(ATR) 服务在预先确定的一段时间后自动停用 Tableau Prep 许可证，从而无需手动停用产品密钥。

有关如何配置此选项的详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[配置虚拟桌面支持](#)。

连接到数据

连接到云数据源和 Hadoop Hive

您们增加了对以下连接器的支持，使您能连接到云数据和 Hadoop 中存储的数据。

- Snowflake
- Amazon EMR Hadoop Hive
- Cloudera Hadoop(Hive 和 Impala)
- Hortonworks Hadoop Hive
- MapR Hadoop Hive
- Apache Drill
- SparkSQL

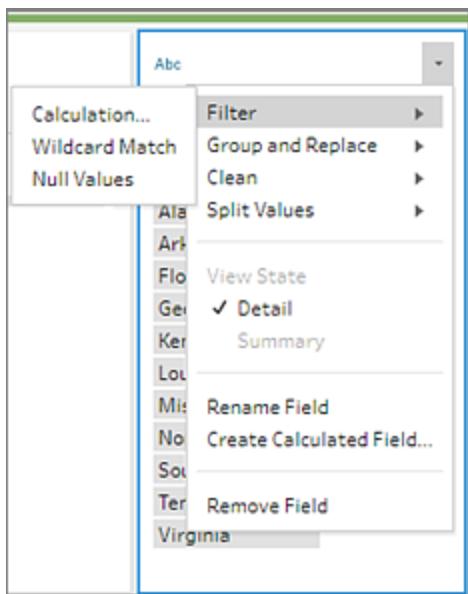
有关如何将 Tableau Prep 连接到数据的详细信息，请参见 Tableau Desktop 帮助中[支持的连接器](#)下适用于您的连接器的主题。

这些数据连接器不向后兼容。包括这些连接器的流程可在以前版本的 Tableau Prep 中打开，但除非移除这些数据连接器，否则将会出错或无法运行。

浏览数据

用于只保留所需数据的新筛选器选项

不再需要编写复杂的计算即可保留或排除 Null 值，或对文本值设置通配符匹配。作为替代，选择字段上的新筛选器选项之一即可立即看到所做更改的效果，而不必先编写计算，并随后在所做更改未产生所需结果的情况下恢复更改。



有关详细信息，请参见[检查您的数据](#) 在本页194。

清理和调整数据

在计算字段中使用 ISO-8601 日期标准

创建计算字段来支持欧洲日历更加轻松。Tableau Prep 现在为以下日期部分支持 ISO-8601 国际日期标准：

- “iso-year”
- “iso-quarter”
- “iso-week”
- “iso-weekday”

Tableau Prep 帮助

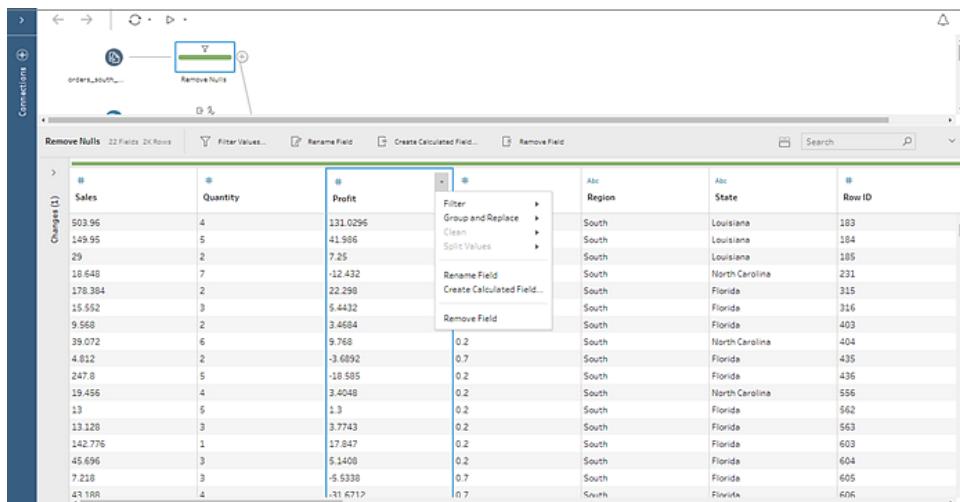
在 DATEPART、DATETRUNC、DATENAME、DATEDIFF 和 DATEADD 函数中使用这些日期部分。

例如，Week Number = STR(DATEPART('iso-year', [Week Date])) + "—" + STR(DATEPART('iso-week', [Week Date]))

有关如何在计算字段中使用日期函数的详细信息，请参见 Tableau Desktop 帮助中的[日期函数](#)。

在数据网格中应用清理操作

在以前的版本中，数据网格会向您显示数据的预览，但可用的清理选项很少。现在您可以在任何地方对数据进行操作。如果在在数据网格中处理详细值，请折叠“配置”窗格，并在数据网格中执行与“配置”窗格中的可用清理操作相同的清理操作。

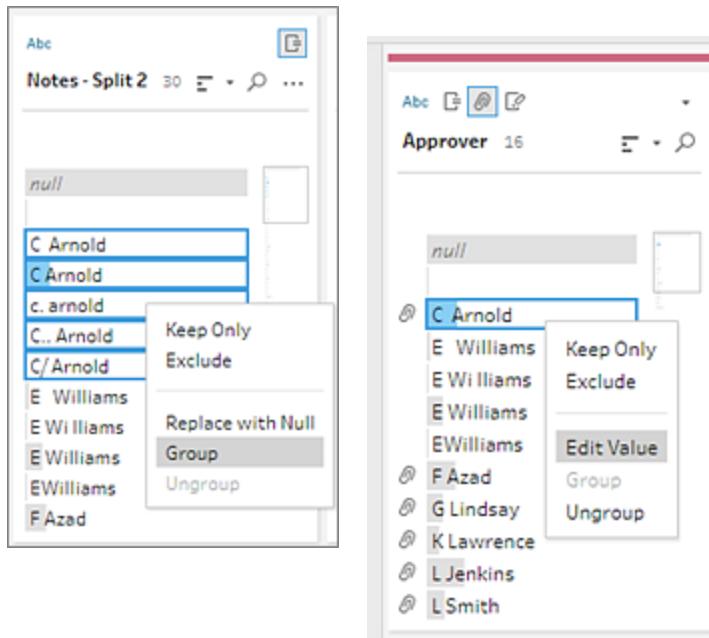


有关详细信息，请参见[清理和调整数据 在本页 223](#)。

使用多选对“配置”窗格中的值进行分组

若要快速对一组字段值进行分组，您现在可以在“配置”卡中选择多个值，然后右键单击打开菜单并选择“分组”。值将分组在您右键单击打开菜单时选择的字段值下。分组的值旁边将显示一个回形针图标。

右键单击分组的值再次打开菜单，以取消分组或编辑值。



有关对值分组的详细信息，请参见[编辑字段值](#) 在本页239。

其他增强功能

我们还提供了以下增强功能来改善可用性：

- 向流程中添加步骤时，菜单上将显示新的图标，用于提供视觉提示并帮助您了解 Tableau Prep 的视觉语言。



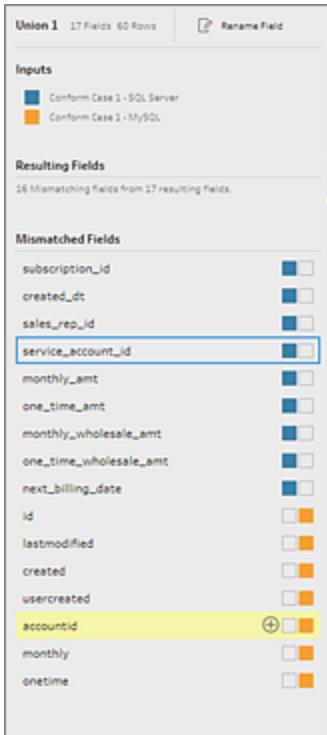
- “配置”窗格中新的动画可帮助吸引您的注意力，以便更好地了解所做更改的效果。

联接或合并数据

使用并集建议来清理不匹配的字段

在合并两个或更多文件后修复不匹配的字段更加轻松。“**并集摘要**”窗格中新的“**不匹配字段**”部分显示不匹配的字段的列表，以及这些字段的来源数据源。使用以下选项之一，直接在此部分中合并不匹配的字段：

- 在列表中选择一个字段。如果 Tableau Prep 基于类似的数据类型和字段名确定具有类似特征的字段，它将以黄色突出显示该字段，同时推荐一个匹配项。单击将光标悬停在推荐的匹配字段上时出现的加号 按钮来合并两个字段。



- 在列表中选择两个或更多字段，右键单击这些字段之一，并选择“**合并字段**”。

有关合并数据和解决不匹配字段的详细信息，请参见[合并数据](#)在本页303。

为所有联接类型确定不匹配的字段

不管您如何联接表，都能轻松确定不匹配的字段。在以前的版本中，“**联接子句**”表只会在字段彼此相等时显示排除的字段值。但现在，您可以在匹配联接子句时使用任何运算符查看此数据，例如“**End Date >= Modified Date**”。

SalesOrderID	ModifiedDate	ProductKey
43,500	01/01/2001, 12:00:00	200
45,500	01/01/2004, 12:00:00	280
47,500	01/01/2004, 12:00:00	360
49,500	01/01/2004, 12:00:00	440
51,500	01/01/2004, 12:00:00	

有关详细信息，请参见[联接数据](#)在本页298。

版本 2018.1.2 新增功能

安装和部署 Tableau Prep

- 通过命令行激活和注册 [Tableau Prep \(Windows\)](#) 向下

连接到数据

- 连接到存储在统计文件中或 [Presto](#) 上的数据 下一页
- 在输入步骤中合并 [Data Interpreter](#) 找到的子表 下一页
- 加载表时更好的反馈 在本页 85

浏览数据

- 对“配置”窗格和“数据”网格中的字段进行重新排序 在本页 86

清理和调整数据

- 使用拖动选择功能来移除流程中的多个步骤 在本页 86
- 在一个操作中对多个字段组进行转置 在本页 87
- 合并字段时改善的字段命名 在本页 87
- 其他增强功能 在本页 88

安装和部署 Tableau Prep

通过命令行激活和注册 Tableau Prep (Windows)

像 Tableau Desktop 一样，您现在可以通过包括含以下属性的命令行，从命令行中激活和注册 Tableau Prep：

- ACTIVATE_KEY=""。此安装程序运行 -activate 来应用产品密钥。
- REGISTER="1"。在安装过程中，安装程序将运行 -register 进程，并添加注册信息。

有关这些安装程序属性以及如何通过命令行激活和注册 Tableau Prep 的详细信息，请参见 [部署 Tableau Desktop](#) 中的“安装程序选项”及相关部分。

连接到数据

连接到存储在统计文件中或 **Presto** 上的数据

我们增加了两个新的连接器来帮助您连接到来自更多位置的数据。

- 统计文件。连接到 SAS (*.sas7bdat)、SPSS (*.sav) 和 R (*.rdata) 数据文件。
- **Presto**。有关如何配置 Presto 连接的详细信息，请参见 Tableau Desktop 帮助中的 **Presto**。

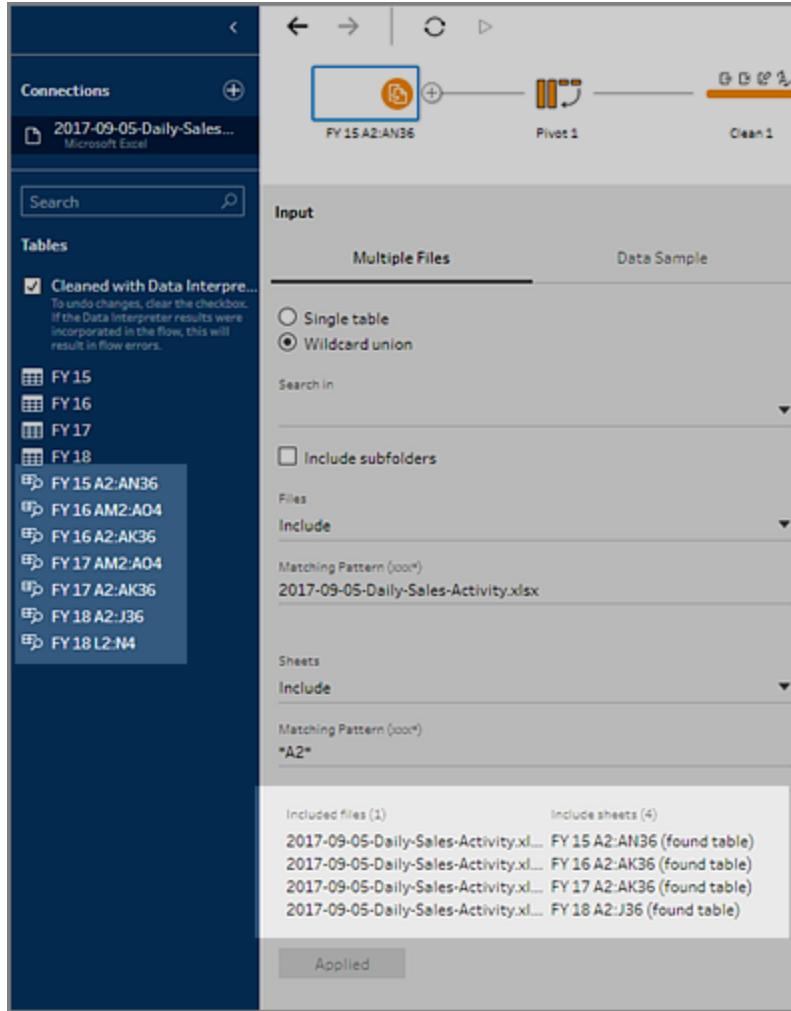
这些数据连接器不向后兼容。包括这些连接器的流程可在以前版本的 Tableau Prep 中打开，但除非移除这些数据连接器，否则将会出错或无法运行。

在输入步骤中合并 **Data Interpreter** 找到的子表

使用 **Data Interpreter** 来清理 Microsoft Excel 数据，并且现在想要合并生成的子表？您现在可以使用通配符并集来合并输出步骤中所有找到的子表。

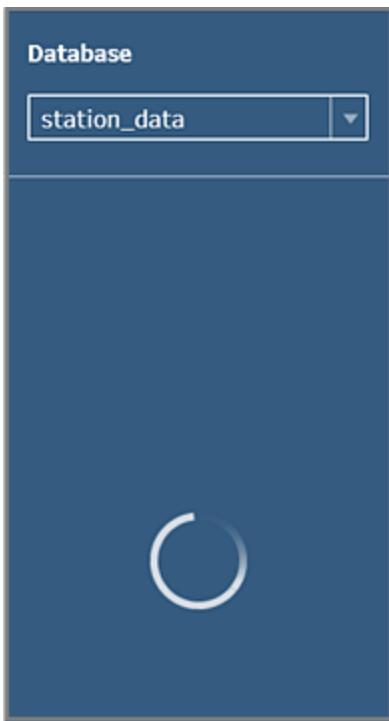
只需将其中一个子表拖到“流程”窗格，然后使用通配符搜索条件来选择要合并数据的其余子表，并将所有子表数据包括在输入步骤中。

有关详细信息，请参见[连接到数据 在本页 141](#)。



加载表时更好的反馈

首次连接到数据库时，有时可能看起来未发生任何事情。现在，新的指示符将指出数据仍在加载。



浏览数据

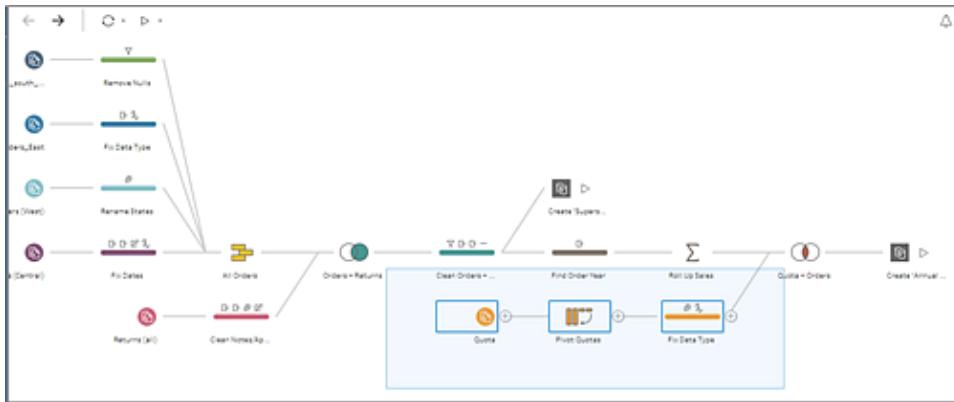
对“配置”窗格和“数据”网格中的字段进行重新排序

您现在可以通过拖放对“配置”窗格和“数据”网格中的字段重新排序，并且两个窗格将保持同步。即使您重命名字段，我们也会保持字段顺序。字段不再会自动重新排序。

清理和调整数据

使用拖动选择功能来移除流程中的多个步骤

移除流程的整个部分更加容易。您现在可在“流程”窗格中单击，并使用鼠标拖动并选择要移除的流程部分。然后右键单击即可一次性移除所有选择的步骤。有关详细信息，请参见[清理和调整数据 在本页 223](#)。



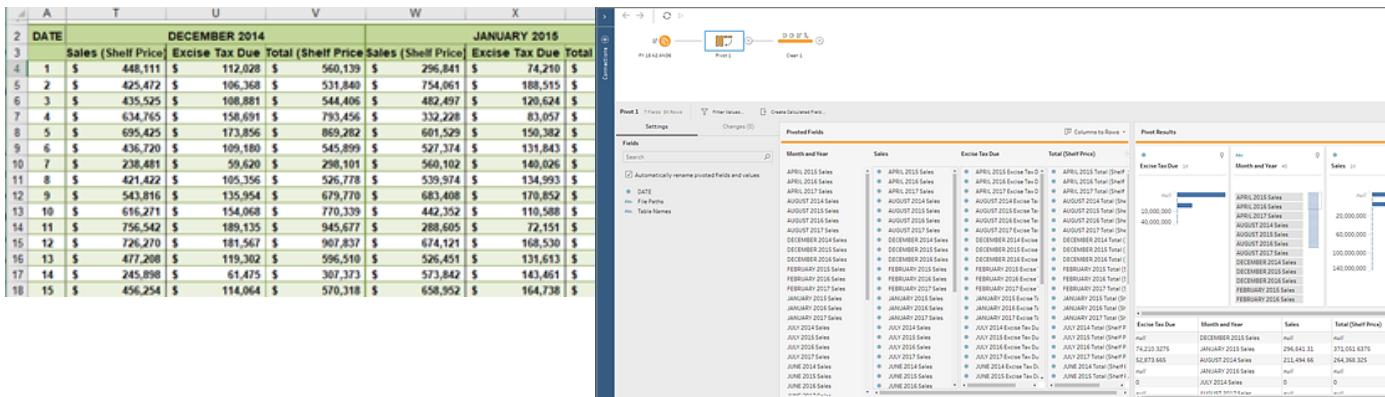
在一个操作中对多个字段组进行转置

Tableau Prep 中的转置选项已经扩展，可更好地支持更复杂的电子表格或文本文件。您现在可以执行单一转置，或选择字段组进行转置。

例如，若要对电子表格进行转置来按月和按年查看销售额、税款和总计，您可以对每组字段进行转置来获取所需的结果。有关详细信息，请参见[清理和调整数据 在本页 223](#)。

原始电子表格

对多组字段进行转置之后

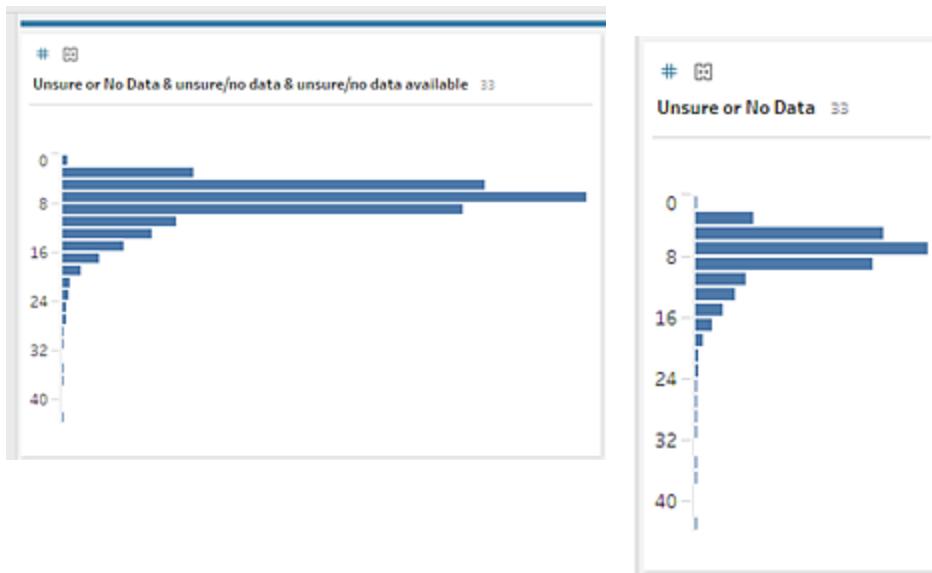


合并字段时改善的字段命名

当您合并字段时，字段名称不再是级联方式。相反，目标字段的字段名会保留。有关详细信息，请参见[合并字段 在本页 234](#)。

级联式字段命名

简化的字段命名



其他增强功能

我们还提供了以下增强功能来改善可用性和性能：

- 将光标悬停在“更改”窗格中的条目上时，将会显示一个新的编辑图标，可帮助您快速确定哪些项目是可编辑的。



- 为步骤重命名将不再会自动运行流程，使您的创作体验保持流畅。
- “连接”窗格将在连接到 Microsoft Excel 文件时自动打开，其中包含多个工作表，可帮助您快速获取数据。

Tableau Prep Builder 入门指南

注意: Tableau Prep 版本 2019.1.2 已更名为 Tableau Prep Builder。如果使用的是产品的早期版本，您仍将在产品中看到“Tableau Prep”。

本教程介绍 Tableau Prep Builder 中的常见操作。您将使用 Tableau Prep Builder 附带的样本数据集来演练为示例 Superstore 创建流程。本教程使用最新版本的 Tableau Prep Builder。如果您使用的是以前的版本，结果可能会不同。

请在进行演练的过程中观看提示，深入了解 Tableau Prep Builder 如何帮助您清理和调整数据以进行分析。

若要在继续本教程之前安装 Tableau Prep Builder，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的[通过用户界面安装 Tableau Desktop 或 Tableau Prep Builder](#)。否则您可以下载[免费试用版](#)。

示例文件

为了完成本教程中的任务，您需要安装 Tableau Prep Builder，并且需要将以下示例 Superstore 数据文件下载到单一位置：

- [Orders_South_2015](#)
- [Orders_South_2016](#)
- [Orders_South_2017](#)
- [Orders_South_2018](#)
- [Orders_Central](#)
- [Orders_East](#)
- [Orders_West](#)
- [returns_reasons_new](#)

在安装 Tableau Prep Builder 后，您还可以在计算机上的以下位置找到这些文件：

(Windows) C:\Program Files\Tableau\Tableau Prep Builder
<version>\help\Samples\en_US\Superstore Files
(Mac) /Applications/Tableau Prep Builder
<version>.app/Contents/help/Samples/en_US/Superstore Files

背景故事...

您在一家大型零售连锁店的总部工作。您的老板想要分析过去四年公司的产品销售额和利润。您推荐他使用 Tableau Desktop 来完成该操作。您的老板认为这个主意很棒，并希望您立即着手完成该工作。

在开始收集需要的所有数据时，您注意到，有人已经通过其他方式为每个区域收集和跟踪过数据。您还注意到不同文件中有大量创新数据条目，而且一个区域甚至每年都有单独的文件！

您将不得不先执行大量的数据清理工作，然后才能开始在 Tableau 中分析数据，而这项工作将需要很长时间。

在您翻找餐厅的菜单来订晚餐时，您想起来 Tableau 刚刚推出了一款名为 Tableau Prep Builder 的新产品，该产品可帮助您完成艰巨的数据清理任务。

您注册了[免费试用版](#)，决定试用一下。

1. 连接到数据

打开 Tableau Prep Builder 时，您首先看到的是一个带有“**连接**”窗格的开始页面，就像 Tableau Desktop 一样。

若要开始操作，第一步是连接到数据并创建输入步骤。您将从该处开始创建工作流(在 Tableau Prep Builder 中称之为“流程”)，并随着进展添加更多步骤来对数据执行操作。

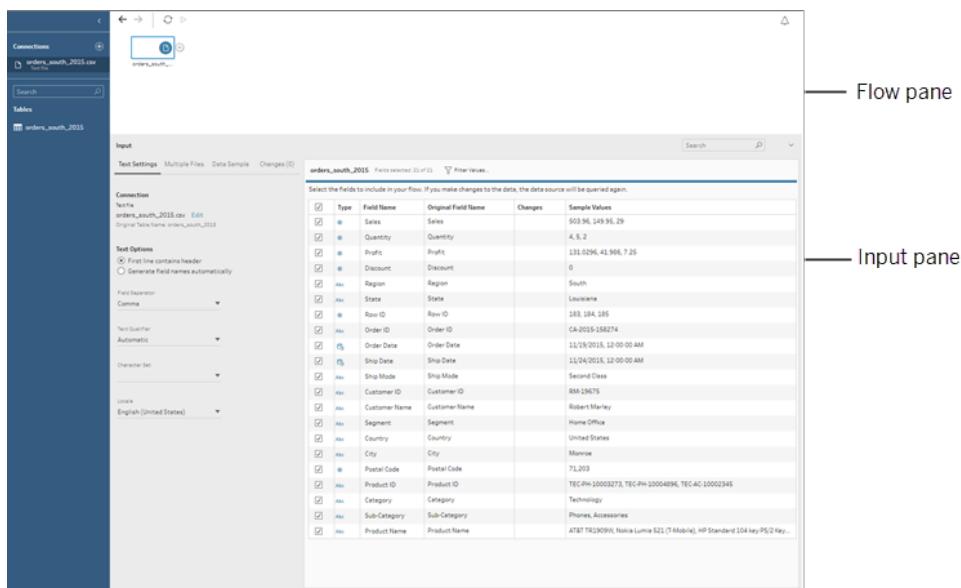
提示：输入步骤是数据的摄入点和流程的开始点。您可能有多个输入步骤，并且某些步骤可能包括多个数据文件。有关连接到数据的详细信息，请参见[连接到数据 在本页 141](#)。

不同区域的销售额数据文件以不同的格式存储，并且“南部”区域的订单实际上是多个文件。您检查了“**连接**”窗格，并发现您有很多选择来连接到数据。太好了！

由于其他区域全部四年的数据是一个文件，因此您决定先着手处理“南部”区域的文件。

1. 在“**连接**”窗格上，单击“**添加连接**” 按钮。
2. 文件是 .csv 文件，因此请在连接列表中选择“**文本文件**”。
3. 导航到文件的目录，选择第一个文件 **orders_south_2015.csv**，并单击“**打开**”以将其添加到流程。(有关文件位置，请参见[总结和资源 在本页 127](#)。)

连接到第一个文件后，Tableau Prep Builder 工作区将打开，您会看到它分为两个主要部分。顶部的“流程”窗格和底部的“输入”窗格。



在 Tableau Prep Builder 中，“流程”窗格是您的工作区，与 Tableau Desktop 很像，您可以在其中直观地与数据交互并构建流程。“输入”窗格包含有关如何摄入数据的配置选项。它还显示数据集中的字段、数据类型以及值的示例。

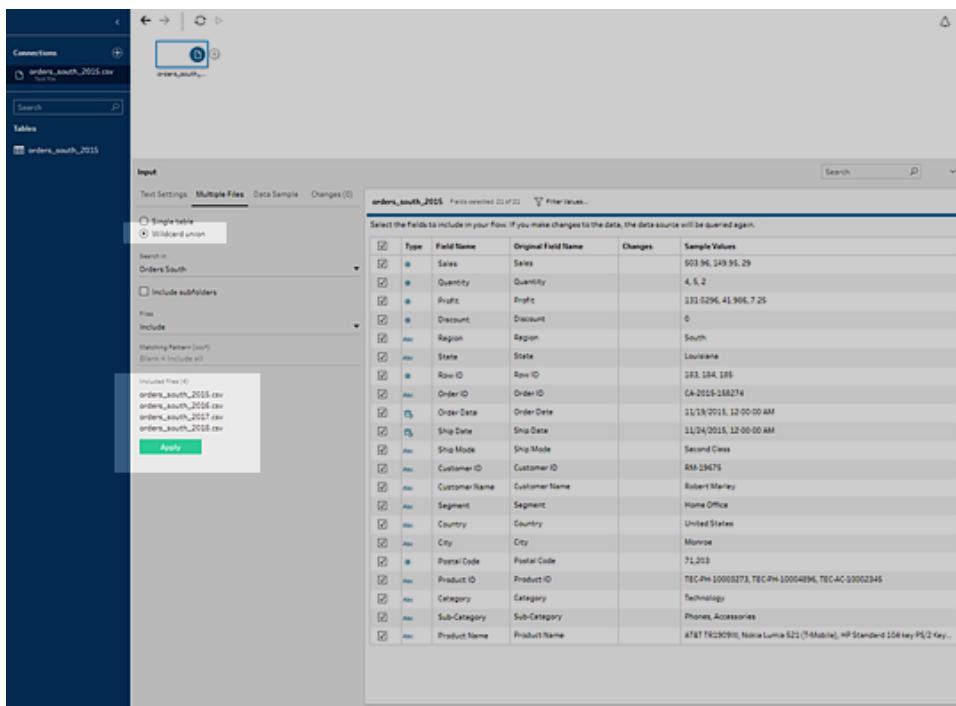
我们将在下一部分中探讨您如何才能与此数据交互。

提示: 对于单个表，Tableau Prep Builder 会在您将数据添加到流程时自动为您在“流程”窗格中创建输入步骤。否则，您可以使用拖放功能将表添加到“流程”窗格。

4. “南部”地区的订单有三个其他文件。您可以单独添加每个文件，但您想要将所有文件合并到一个输入步骤中，因此您在“输入”窗格中单击“多个文件”选项卡。
5. 您会看到一个“通配符并集”选项。选择该选项。

您注意到，您在其中选择文件的目录已填充，并且您需要的其他文件列在“输入”窗格的“已包括文件”部分中。

提示: 使用通配符并集可以非常方便的连接到单一数据源中具有类似名称和结构的多个文件并将其合并。若要使用此选项，文件必须位于同一父目录或子目录中。如果没有立即看到所需的文件，请更改搜索条件。有关详细信息，请参见[在输入步骤中合并文件和数据库表 在本页 170](#)。



6. 单击“应用”以将这些文件中的数据添加到“**orders_south_2015**”输入步骤。

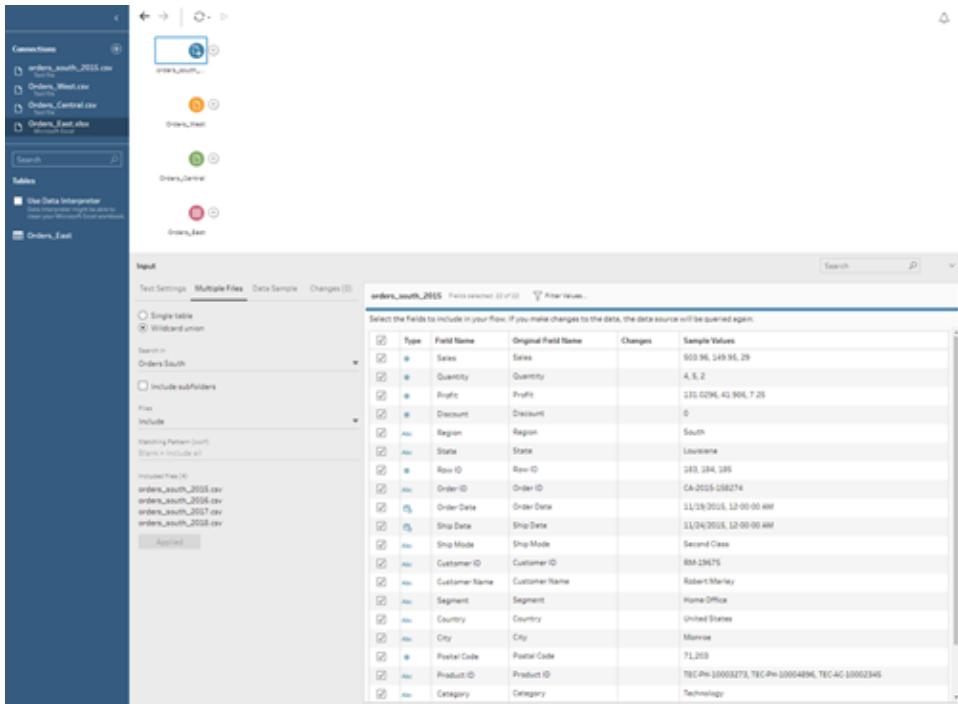
其他区域的文件全部都是单一表文件，因此您可以一次性选择所有文件并将它们添加到流程。

7. 打开文件资源管理器或 Finder，并导航到文件的目录。按住 Ctrl 单击或按住 Cmd 单击 (MacOS)，选择以下文件并将它们拖放到“流程”窗格上，将这些文件添加到流程。
(有关文件位置，请参见[总结和资源 在本页 127](#)。)

- Orders_Central.csv
- Orders_East.xlsx
- Orders_West.csv

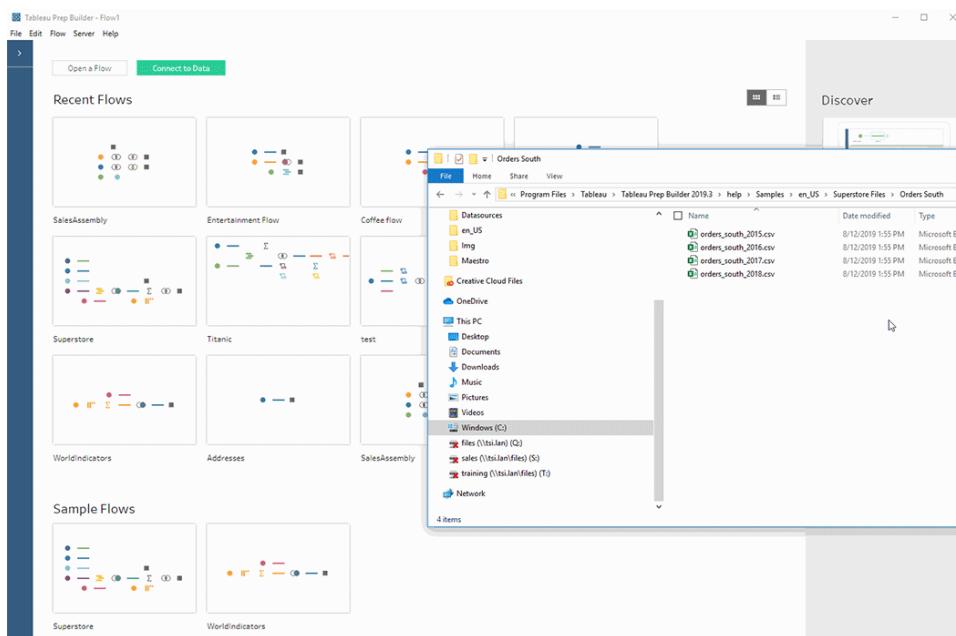
注意: 这些是不同的文件类型。如果您看不到所有这些文件，请确保您的文件资源管理器或 Finder 设置为查看所有文件类型。

Tableau Prep 帮助



检查您的工作：观看“连接到数据”的实际运行效果。

单击图像可重播



2. 浏览数据

既然您已将数据文件加载到 Tableau Prep Builder 中，您很确定自己想要将所有文件合并在一起。但在执行该操作之前，不妨先检查一下这些文件，看看是否能找出任何问题。

在“流程”窗格中选择输入步骤时，您可能会看到用于引入数据的设置、所包括的字段以及一些样本值。

在这里可以很方便地决定要在流程中包括多少数据，以及移除或筛选不需要的字段。您还可以更改未正确分配的任何数据类型。

提示：如果在处理大型数据集，Tableau Prep Builder 会自动引入数据样本，从而最大限度提升性能。如果看不到预期数据，您可能需要调整样本。可在“**数据样本**”选项卡上执行此操作。有关配置数据选项和样本大小的详细信息，请参见[选择数据样本大小 在本页 167](#)。

在“流程”窗格中，当您选择每个步骤并查看每个数据集时，您注意到需要稍后修复的一些问题，以及一个可立即在输入步骤中修复的问题。

- 在“流程”窗格中，单击“**Orders_Central**”输入步骤将其选定。在“输入”窗格中，您注意到以下问题：
 - 订单日期和发货日期在字段中已针对月、日和年分开。
 - 某些字段的数据类型与其他文件中相同字段的数据类型不同。
 - “**Region**”(区域)没有字段。

您将需要对这些字段执行一些清理操作，然后才能将此文件与其他文件合并。但您无法在输入步骤中的此处修复该问题，因此记下来以便稍后进行修复。

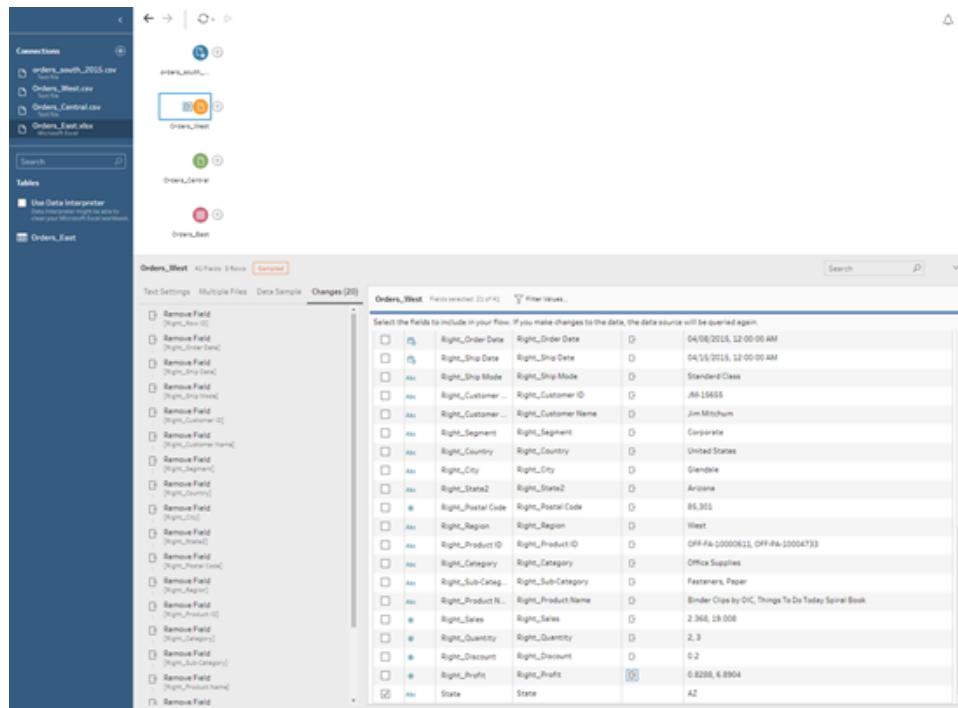
- 选择“**Orders_East**”输入步骤。

此文件中的字段看起来与其他文件完全一致。但“**Sales**”(销售额)值似乎全部都包括了货币代码。您稍后也需要修复该问题。

- 选择“**Orders_West**”输入步骤。此文件中也有一些问题。

- “**State**”(州)字段使用州名称的缩写。其他文件则使用完整拼写，因此您将需要稍后修复该问题。
- 有许多以“**Right_**”开头的字段。这些字段似乎与其他字段重复。您不想在流程中包括这些重复字段。此问题可在输入步骤中的此处修复：

若要立即修复此问题，请清除以“**Right_**”开头的所有字段的复选框。这会指示 Tableau Prep Builder 忽略这些字段，不将它们包括在流程中。



提示:当您在某个步骤中执行清理操作(比如移除字段)时, Tableau Prep Builder 会在“**更改**”窗格中跟踪您所做的更改,并在“流程”窗格中(以小图标的形式)添加注释来帮助您跟踪对数据所做的操作。对于“输入”步骤,也会向每个字段中添加注释。

在 Tableau Prep Builder 版本 2018.3.1 及更高版本中,可以在流程的大部分步骤中执行清理操作。在以前的版本中,大多数清理操作必须在清理步骤中执行。

既然您已经确定了数据集中的一些问题,那么下一步是更仔细地检查您的数据并清理找到的任何问题,以便能够合并和调整数据,并生成可用于分析的输出文件。

3. 清理数据

在 Tableau Prep Builder 中,检查和清理数据是一个反复的过程。确定了要处理的数据集之后,下一步是检查该数据,并通过对数据应用各种清理、调整和合并操作来对其进行操作。您通过向流程中添加步骤来应用这些操作。有关 Tableau Prep Builder 中的清理选项的详细信息,请参见[清理和调整数据 在本页 223](#)。

根据您尝试进行的操作,步骤有多种风格。例如,在想要对字段应用像筛选、合并、拆分、重命名这样的清理操作时随时添加清理步骤。添加聚合步骤来对字段进行分组和聚合,并更改

数据的详细级别。有关不同步骤类型及其用法的详细信息，请参见[构建和组织流程 在本页 179](#)。

提示：向流程中添加步骤时，会自动添加一条流程线将步骤相互连接在一起。您可以将这些流程线四处移动，并根据需要移除或添加流程线。

运行流程时，必须要有这些连接点，以便 Tableau Prep Builder 知道哪些步骤已连接以及步骤在流程中的应用顺序。如果缺少流程线，则流程会中断并出现错误。

清理 Orders_Central

为了解决之前发现的问题并确定是否还有任何其他问题，您首先向“**Orders_Central**”输入步骤中添加一个清理步骤。

1. 在“流程”窗格中，选择“**Orders_Central**”，单击加号 图标，并添加清理步骤。根据您的版本，此菜单选项为“添加步骤”、“添加清理步骤”或“清理步骤”。

向流程中添加清理步骤时，工作区将发生变化，您将看到数据的详细信息。

The screenshot shows the Tableau Prep Builder interface with the following components labeled:

- A.** “流程”窗格 (Workflow pane) on the left, showing the flow from 'Orders_South...' to 'Orders_Central' and then to 'Clean1'.
- B.** 工具栏 (Toolbar) at the top of the configuration pane.
- C.** “配置”窗格 (Configure pane) showing the 'Clean1' step configuration, including 'Outputs' (Row ID, Order ID, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Product ID, Category), 'Inputs' (Row ID, Order ID, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Product ID, Category), and 'Actions' (First Class, Same Day, Second Class, Standard Class).
- D.** 数据网格 (Data grid) at the bottom, displaying a list of rows with columns: Row ID, Order ID, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Product ID, and Category. The data includes various order entries and customer details.

A. “流程”窗格、B. 工具栏、C. “配置”窗格、D. 数据网格

工作区现在分为三个部分：“流程”窗格、带工具栏的“配置”窗格和“数据”网格。

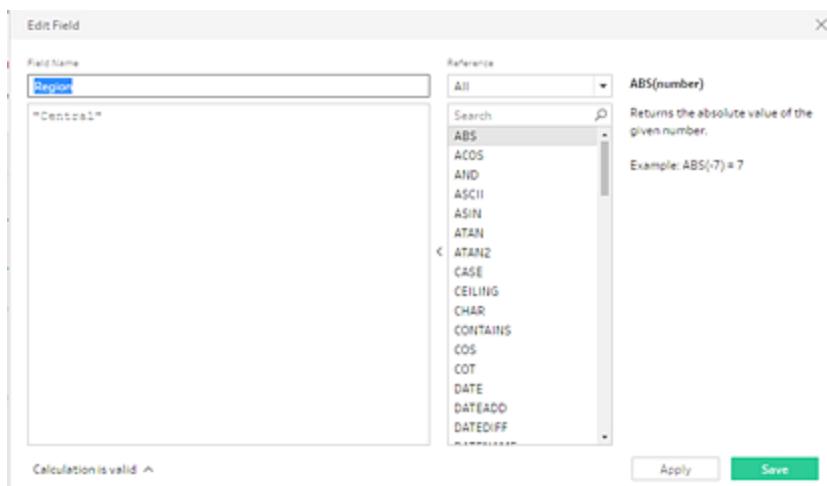
“配置”窗格显示数据的结构，同时将字段值汇总到数据桶中，以便您能够快速查看相关值，并找出离群值和 null 值。“数据”网格显示字段的行级别详细信息。

提示: “配置”窗格中的每个字段都显示在配置卡上。使用每张卡上的“更多选项” \cdots 菜单(在以前的版本中为下拉箭头)来查看和选择可用于该字段类型的不同清理选项。您也可以对字段值进行排序、更改数据类型、为字段分配数据角色，或在“数据”网格中拖放配置卡和列以重新排列它们。

使用计算字段清理数据

此数据集缺少“Region”(区域)字段。由于其他数据集有此字段，因此您将需要添加此字段，以便能够稍后合并数据。您将需要使用计算字段来执行此操作。

2. 在工具栏中，单击“创建计算字段”。
3. 在计算编辑器中，将计算字段命名为“Region”(区域)。然后输入 "Central"(包括引号)，并单击“保存”。



您能够使用计算字段来调整数据，这种灵活性让您非常喜欢。您非常高兴看到 Tableau Prep Builder 使用与 Tableau Desktop 相同的计算编辑器语言。

提示: 当您对字段和值进行更改时，Tableau Prep Builder 会在“更改”窗格跟踪所做的更改。同时，会向流程中的清理步骤和“配置”窗格中的字段添加一个表示更改的图标(注释)。我们将在进行更多更改之后查看“更改”窗格。

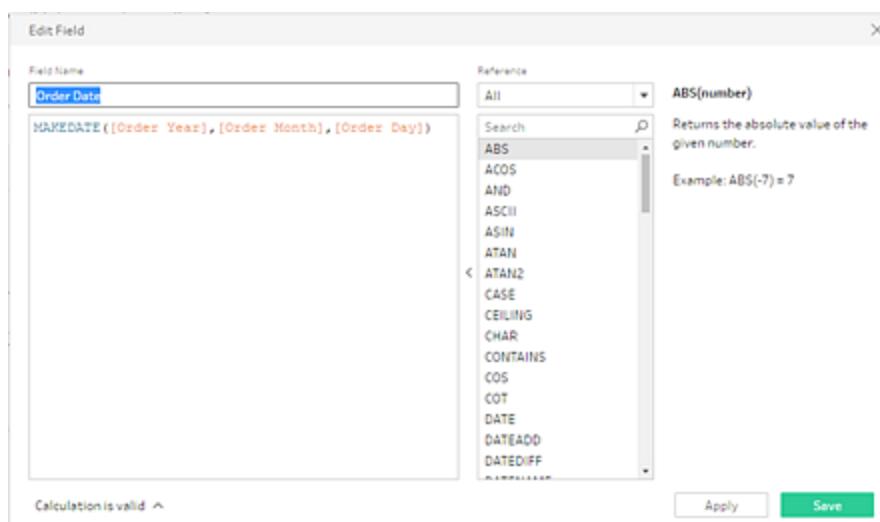
接下来，您需要解决单独订单日期和发货日期字段的问题。您需要将它们合并为两个单一字段，一个用于“Order Date”(订单日期)，一个用于“Ship Date”(发货日期)，使

它们与其他数据集中的相同字段保持一致。确保表具有相同的字段将使您能够稍后使用并集来合并表。

您可以再次使用计算字段，通过一个轻松的步骤执行此操作。

4. 在工具栏中，单击“创建计算字段”，将“**Order Year**”(订单年份)、“**Order Month**”(订单月份) 和“**Order Day**”(订单日) 字段合并为一个格式为“MM/DD/YYYY”的字段。
5. 在计算编辑器中，将计算字段命名为“**Order Date**”(订单日期)。然后输入以下计算，并单击“保存”：

```
MAKEDATE([Order Year], [Order Month], [Order Day])
```



既然有了新的订单日期字段，那么您需要移除现有字段，因为您不再需要它们。

“配置”窗格中有很多字段。您注意到工具栏右上角的“搜索”框。您不知道是否能使用该框来快速查找想要移除的字段。您决定试一下。

6. 在“配置”窗格的搜索框中，键入 **Order**。

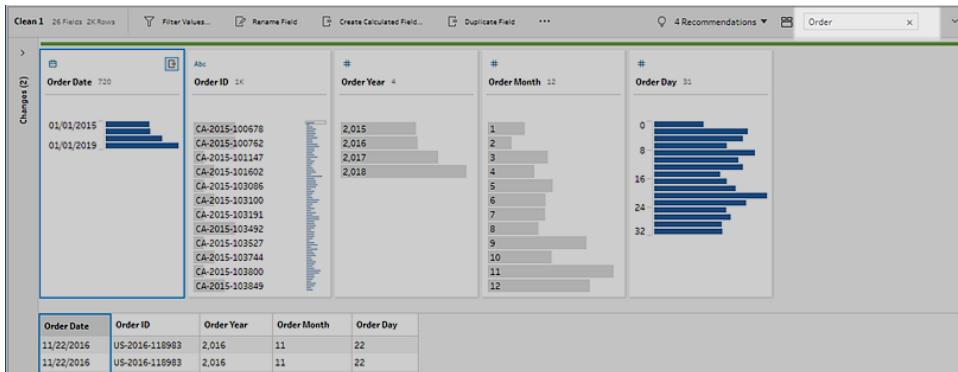


Tableau Prep Builder 将在视图中快速滚动名称中包含“**Order**”的所有字段。太酷了！

- 按住 Ctrl 单击或按住 Cmd 单击 (MacOS), 选择“**Order Year**”(订单年份)、“**Order Month**”(订单月份) 和“**Order Day**”(订单日) 字段。然后右键单击所选字段，并从菜单中选择“**移除字段**”(在以前的版本中为“**移除字段**”) 来移除它们。



- 现在重复上面的步骤 4 至 7, 为“**Ship Date**”(发货日期) 创建单一字段。自己尝试一下, 或使用下面的步骤来帮助您。

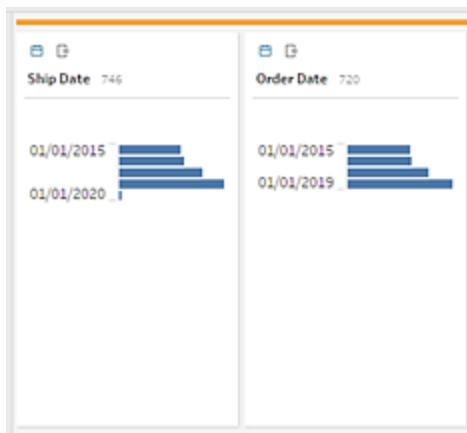
- 在工具栏中, 单击“**创建计算字段**”, 将“**Ship Year**”(发货年份)、“**Ship Month**”(发货月份) 和“**Ship Day**”(发货日) 字段合并为一个格式为“MM/DD/YYYY”的字段。
- 将该计算字段命名为“**Ship Date**”(发货日期), 并输入计算 MAKEDATE ([Ship Year], [Ship Month], [Ship Day])。然后单击“**保存**”。
- 移除“**Ship Year**”(发货年份)、“**Ship Month**”(发货月份) 和“**Ship Day**”(发货

日) 字段。搜索并选择字段, 然后从菜单中选择“移除”(在以前的版本中为“移除字段”)来移除这些字段。

提示: Tableau Prep Builder 会将“配置”窗格中的数据汇总到数据桶中, 帮助您快速查看数据的形状、查找离群值、找出字段之间的关系, 等等。

在此方案中, 订单日期和发货日期现在可按年份汇总。每个数据桶都表示从起始年份一月份到下一年一月份的一年, 并带有相应标签。由于有处于 2018 年和 2019 年下半年的销售日期及发货日期, 因此该数据的数据桶相应标有 2019 年和 2020 年年末。

若要将此视图更改为实际日期, 请单击“配置”卡中的“更多选项”... 菜单(在以前的版本中为下拉箭头), 并选择“详细信息”。



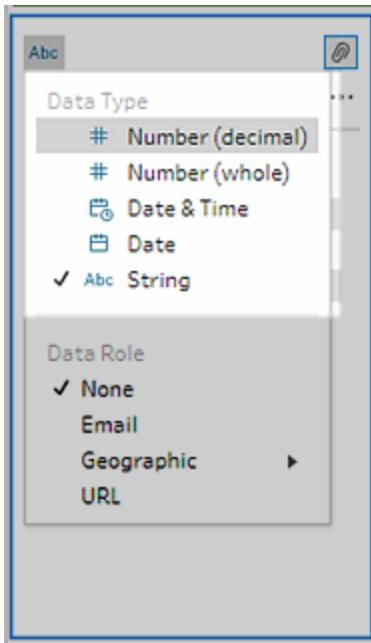
与字段直接交互来清理数据

您的数据一开始看起来很好。但在完成为订单日期和发货日期移除额外字段的操作后, 您注意到“**Discounts**”(折扣) 字段有一些问题。

- 系统为其分配了“字符串”数据类型, 而不是“数字(小数)”数据类型。
- 对于无折扣的情况, 字段值是“**None**”(无), 而不是数字值。

这将在合并文件时导致问题, 因此最好也修复该问题。

9. 清除搜索, 并在搜索框中输入 **disc** 来查找字段。
10. 选择“**Discounts**”(折扣) 字段, 双击字段值“**None**”(无), 并将其更改为数字值“**0**”。
11. 若要将“**Discount**”(折扣) 字段的数据类型从“字符串”更改为“数字(小数)”, 单击“**Abc**”, 并从下拉菜单中选择“数字(十进制)”。



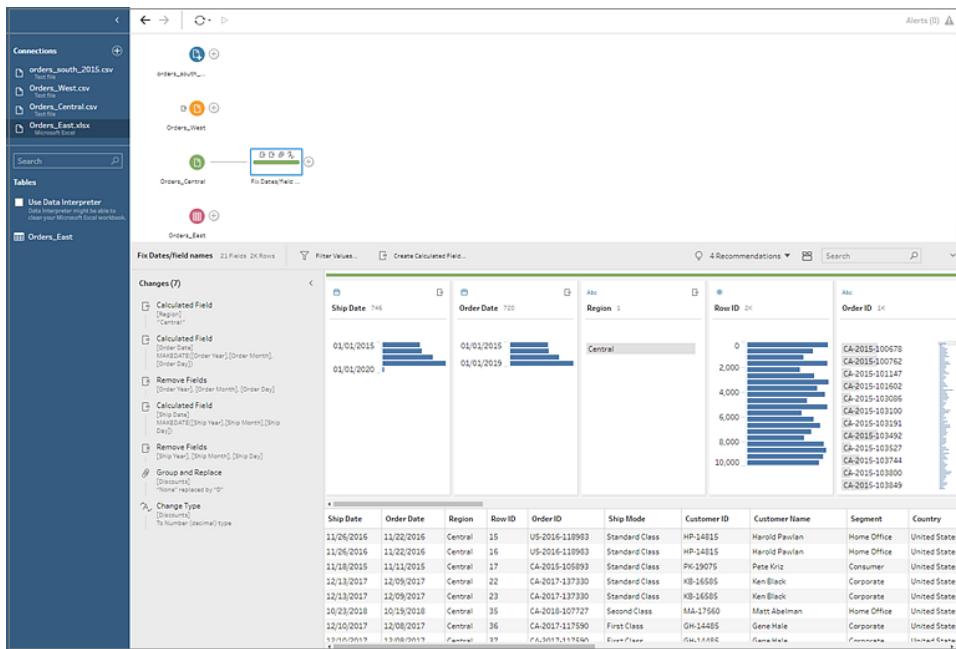
12. 最后为步骤命名，以帮助跟踪您在此步骤中进行的操作。在“流程”窗格中，双击步骤名称**“Clean 1”(清理 1)**，并键入**“Fix dates/field names”(修复日期/字段名称)**。

查看您的更改

您对此数据集进行了许多更改，并开始担心自己会不记得所进行的操作。在检查工作时，您看到**“配置”**窗格的左侧有一个名为**“更改”**的列。

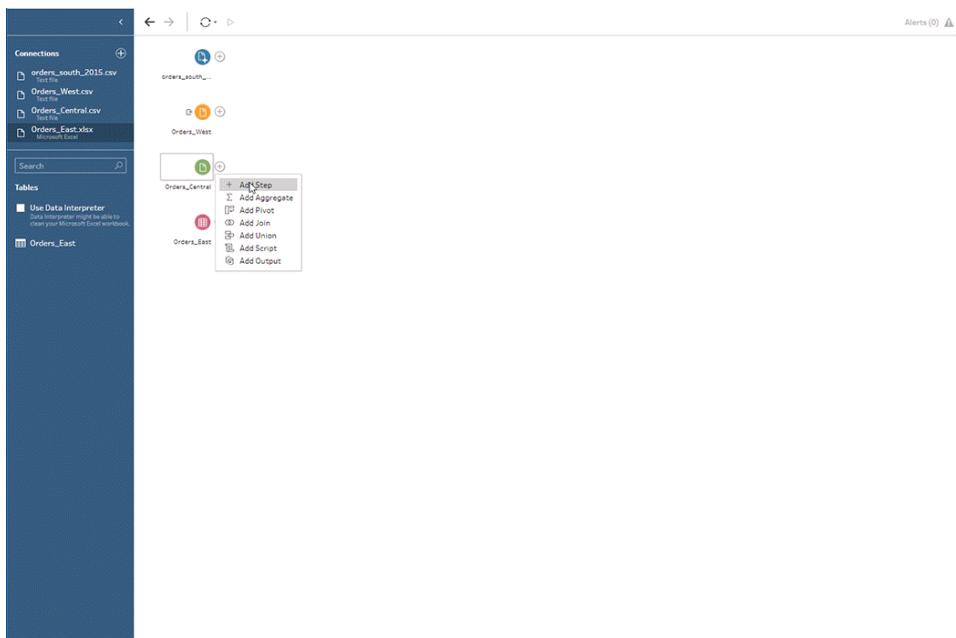
您单击箭头将其打开，看到其中列出了您刚刚进行的每项更改，这让您很高兴。在您滚动浏览列表中的更改时，您注意到，您可以删除或编辑更改，甚至能将更改四处移动来改变所做更改的顺序。

您可以轻松地查找构建流程时在任何步骤中所做的更改，并对这些更改的顺序进行试验来最大程度地利用您的数据，这让您非常喜欢。



检查您的工作：观看“清理 Orders_Central”的实际运行效果。

单击图像可重播



既然已清理了一个文件，那么您将查看其他文件来确定需要修复的其他问题。

接下来您决定查看“Orders_East”的 Excel 文件。

清理 Orders_East

在查看“**Orders_East**”文件的字段时，除了“**Sales**”(销售额)之外，大多数字段看起来与其他文件一致。为了进一步查看并确定是否还有任何其他要解决的问题，您向“**Orders_East**”输入步骤中添加一个清理步骤。

1. 在“流程”窗格中，选择“**Orders_East**”，单击加号  图标，并添加清理步骤。根据您的版本，此菜单选项为“添加步骤”、“添加清理步骤”或“清理步骤”。

通过查看“**Sales**”(销售额) 字段，您很快发现随销售额数字一起包括了 **USD** 货币代码，并且 Tableau Prep Builder 已将这些字段值解读为字符串。

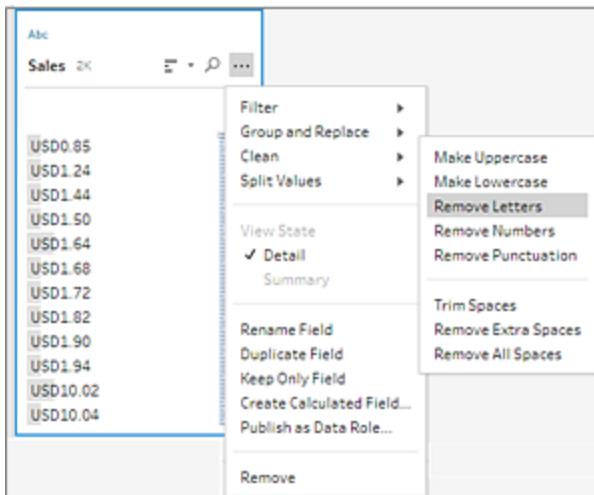
如果要获得准确的销售额数据，您将需要从此字段中移除货币代码，并更改数据类型。

修复数据类型很容易，您已经知道如何做了。但唯一的销售额数据行超过 2000 行，而修复每个单独的行来移除货币代码看起来很麻烦。

但这是 Tableau Prep Builder，因此您决定检查下拉菜单，看看是否有用于修复此问题的选项。

当您单击“**Sales**”(销售额) 字段的“更多选项”… (以前版本中的下拉箭头) 时，您看到一个名为“清理”的菜单选项，以及该选项下一个用于移除字母的选项。您决定尝试一下，看看该选项有什么用。

2. 选择“**Sales**”(销售额) 字段。单击“更多选项”… 菜单(在以前的版本中为下拉箭头)，并选择“清理”>“移除字母”。



哇！该清理选项立即从每个字段中移除了货币代码。现在您只需要将数据类型从“字符串”更改为“数字(小数)”，此文件看起来就正常了。

- 单击“**Sales**”(销售额) 字段的数据类型，并从下拉列表中选择“**数字(小数)**”来更改数据类型。

- 文件的其余部分看起来很好。为清理步骤命名以跟踪您的工作。例如，命名为“**Change data type**”(更改数据类型)。

接下来您将查看 **Orders_West** 的最后一个文件，看看是否有任何需要修复的问题。

清理 Orders_West

在您查看“**Orders_West**”文件的字段时，大多数字段看起来都与其他文件一致，但您记得看到过“**States**”(州) 字段为值使用了缩写，而不是州名称的完整拼写。为了将此文件与其他文件合并，您将需要修复此问题。因此您向“**Orders_West**”输入步骤中添加一个清理步骤。

1. 在“流程”窗格中，选择“**Orders_West**”，单击加号  图标，并添加清理步骤。

滚动或使用“搜索”来查找“**State**”(州) 字段。您将看到所有州名称值都使用简短缩写。此字段只有 11 个唯一的值。您可以手动更改每个值，但 Tableau Prep Builder 是否可能有另一种方法来完成此操作？

您单击字段的“更多选项”… 菜单(以前版本中的下拉箭头)，并看到一个名为“对值进行分组”(在以前的版本中为“分组和替换”)的选项。选择该选项时，您将看到若干选项：

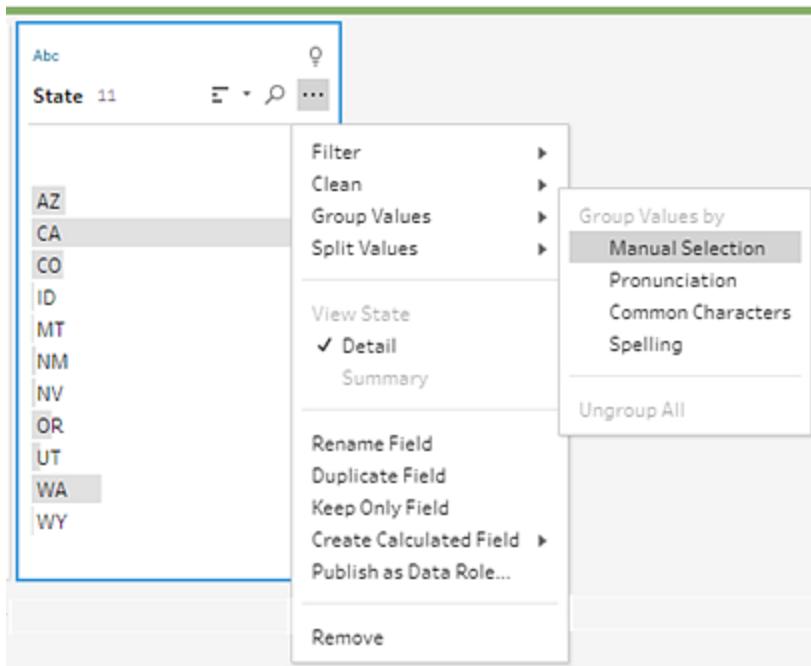
- 手动选择
- 发音
- 常用字符
- 拼写

州名称的发音不相似，它们的拼写不正确，也没有共用相同的字符，因此您决定尝试“手动选择”选项。

提示: 您可以双击字段名称或字段值来编辑单个值。若要编辑多个值，您可以选择所有值，并使用右键单击菜单选项“**编辑值**”。但如果要将一个或多个值映射到特定值，请使用下拉菜单中的“对值进行分组”选项。

有关对值进行编辑和分组的详细信息，请参见[编辑字段值 在本页 239](#)。

2. 选择“**State**”(州) 字段。单击下拉箭头，并选择“对值进行分组”(以前版本中的“分组和替换”) > “手动选择”。



一个两列的卡将打开。这是“对值进行分组”编辑器。左侧的列显示当前字段值，右侧的列显示可映射到左侧字段的字段。

您想要将州缩写映射到州名称的完整拼写版本，但在“**Orders_West**”数据集中没有这些值。您不知道是否能直接编辑名称和在该处添加名称，因此您尝试了一下。

3. 在左侧窗格的“对值进行分组”编辑器中，双击“**AZ**”以突出显示值，并键入**Arizona**。然后按 **Enter** 以添加更改。

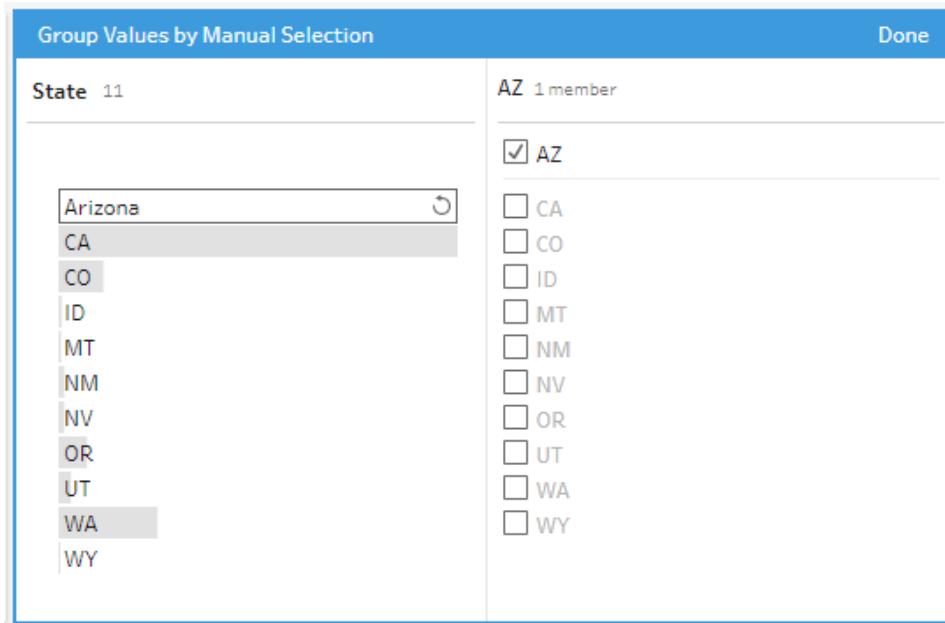


Tableau Prep Builder 为新值“**Arizona**”创建了一个映射值，并自动将旧值“**AZ**”映射到该值。如果您像这样输入了更多此区域的数据，那么为这些值建立映射关系将节省您的时间。

提示：您可以添加数据样本中没有的字段值，建立映射关系来组织数据。如果刷新数据源并添加了新数据，您可以将新数据添加到映射，而不是手动修复每个值。

当您手动添加数据样本中没有的值时，值将带有红点标记，可帮助您轻松识别该值。

4. 重复这些步骤，将每个州都映射到其名称的完整拼写版本。

缩写 州名称

AZ Arizona

CA California

CO Colorado

ID Idaho

MT Montana

NM New Mexico

NV Nevada

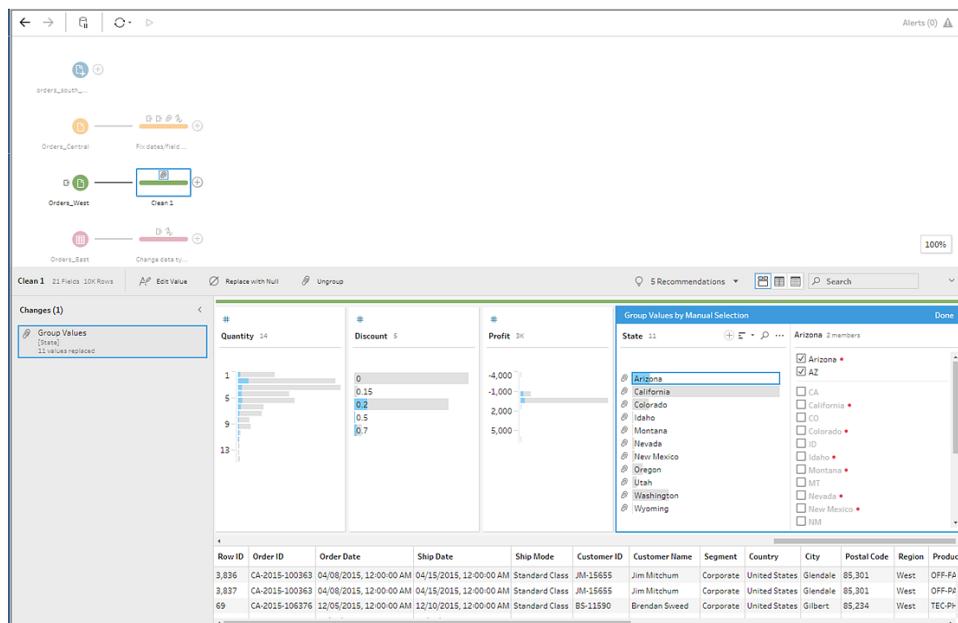
OR Oregon

UT Utah

WA Washington

WY Wyoming

然后，单击“完成”关闭“对值进行分组”编辑器。



映射了所有州之后，您查看“更改”窗格，看到其中只有一个条目，而不是 11 个条目。

看起来 Tableau Prep Builder 已将字段的类似操作分组在一起。您喜欢该功能，因为利用它可以更轻松地查找随后对数据集所做的更改。

修复“**State**”(州) 字段值是您需要在此处进行的唯一更改。

5. 为清理步骤命名以跟踪您的工作。例如“**Rename states**”(重命名州)。

您在文件中进行了大量的清理，不敢相信操作能如此快捷轻松地完成。您竟然还可以回家吃晚饭！为了确保不会丢失到目前为止所做的工作，请保存您的流程。

单击“文件”>“保存”或“文件”>“另存为”。将文件另存为流程文件 (.tfl) 并为其命名。例如“**My Superstore**”(我的超市)。

提示: 在保存流程文件时，您可以将它们另存为流程文件 (.tfl) 或打包文件 (.tflx)，并将本地数据文件与这些文件打包在一起，以便将流程和文件与其他人共享。有关保存和共享流程的详细信息，请参见[保存和共享工作 在本页 309](#)。

4. 合并数据

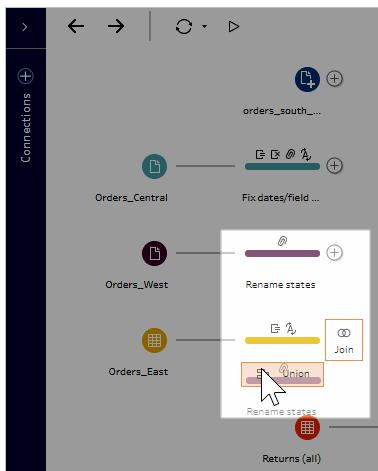
既然清理了所有文件，那么您就已最终准备好将它们合并在一起。

由于清理后所有文件都有类似的字段，因此若要将所有行合并到单个表中，您需要将这些表合并。

您记得有一个名为“**并集**”的步骤选项，但不知道是否能简单地通过拖放步骤来合并文件。您决定试一下看看。

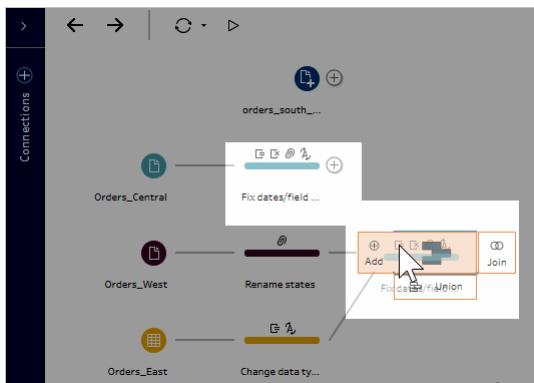
合并数据

1. 在“**流程**”窗格中，将倒数第二个清理步骤拖到流程中的最后一个清理步骤上方，并将其放在“**并集**”选项上。例如，将清理步骤**“Rename states”**(重命名州) 拖到**“Changed data type”**(更改的数据类型) 步骤，并将其放在“**并集**”选项上。



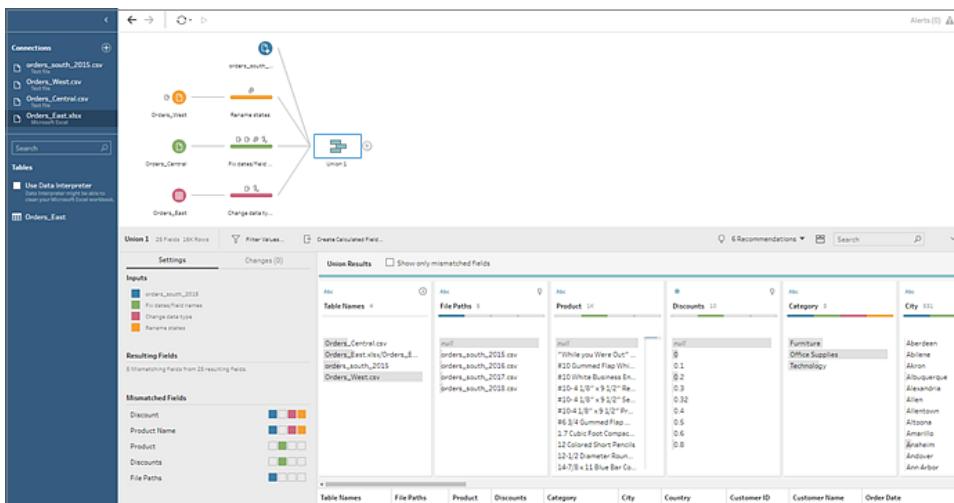
您看到 Tableau Prep Builder 向流程中添加了一个新的“**Union**”(合并) 步骤。太好了！现在，您想要向此并集中添加其他文件。

2. 将流程中的第二个清理步骤(新的“**Fix dates/field names**”(修复日期/字段名) 步骤) 拖到新的“**并集**”步骤上，然后将其放在“**添加**”上以添加到现有并集。



3. 将“orders_south_2015”步骤拖到新的“Union”(合并)步骤。将其放在“添加”上以添加到现有并集。

现在，所有文件已合并到一个表中。在“流程”窗格中，选择新的“Union”(合并)步骤以查看结果。



您注意到，Tableau Prep Builder 自动匹配了具有相同名称和类型的字段。

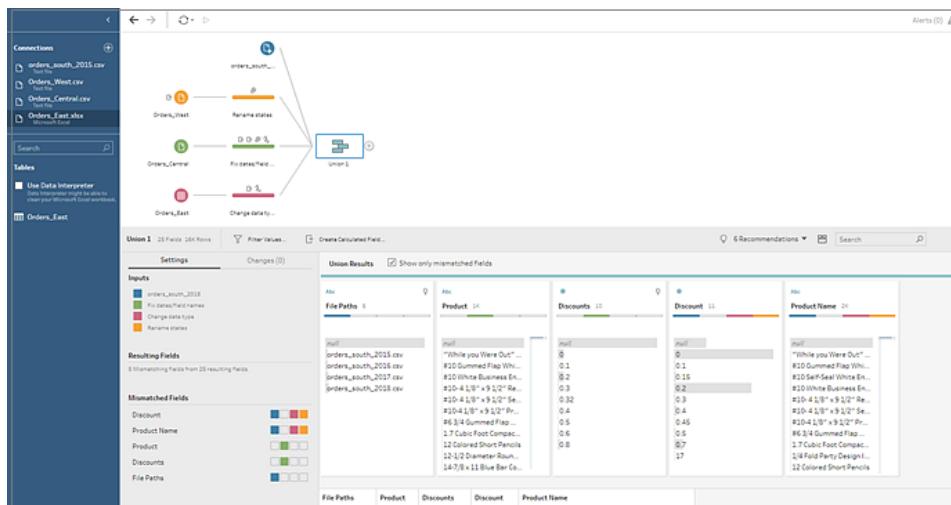
您还看到，分配给流程中步骤的**颜色**用在并集配置中指明字段来自于何处，并同时出现在每个字段顶部的**色带**中，显示该字段是否存在于该表中。

您注意到新增了一个名为“**Table Names**”(表名称)的新字段，该字段列出了并集中所有行的来源表。

不匹配字段的列表也显示在摘要窗格中，并且您可以立即看到字段“**Product**”(产品)和“**Discounts**”(折扣)只出现在 **Orders_Central** 文件中。

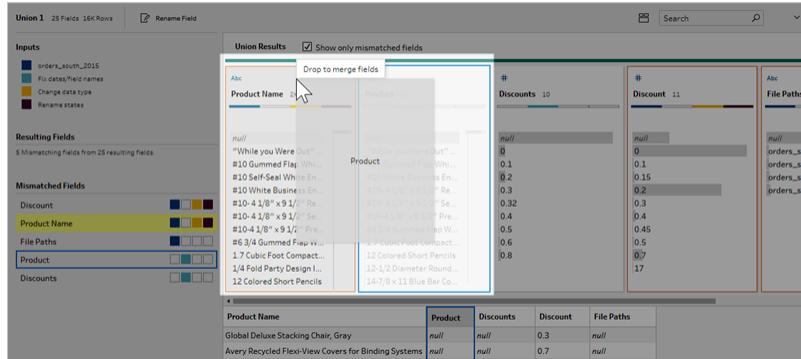
4. 若要更仔细地查看这些字段，请在“并集结果”窗格中选中“仅显示不匹配字段”复选

框。



通过查看字段数据，您很快发现数据相同，但字段名不同。您可以简单地重命名字段，但不知道是否能拖放这些字段来合并它们。您决定试一下看看。

- 选择“**Product**”(产品)字段，并将其拖放到“**Product Name**”(产品名称)字段上来合并字段。字段合并后将不再出现在窗格中。



- 重复此步骤，将“**Discounts**”(多项折扣)字段与“**Discount**”(折扣)字段合并。

现在，唯一没有匹配项的字段是“**File Paths**”(文件路径)字段。此字段显示您为“南部”区域销售订单执行的通配符并集的文件路径。您决定将此字段保留在该处，因为它有很好的信息。

提示：在合并之后修复不匹配的字段时，您有若干选择。

根据所使用的 Tableau Prep Builder 版本，您可以在“**不匹配的字段**”列表中选择一个

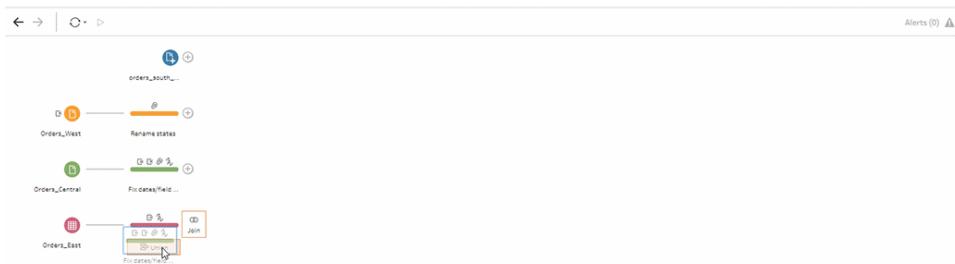
字段(版本 2018.2.1 及更高版本),如果 Tableau Prep Builder 检测到可能的匹配项,它将以黄色突出显示该匹配项。若要合并字段,请将光标悬停在突出显示的字段上,并单击出现的加号 按钮。

有关在并集中合并字段的更多方式,请参见[修复不匹配的字段 在本页 306](#)。

7. 清除“仅显示不匹配字段”复选框以显示并集中包括的所有字段。
8. 为合并步骤命名以表示此并集所包括的内容。例如,“**All orders**”(所有订单)。

检查您的工作:观看“合并数据”的实际运行效果。

单击图像可重播



您是清理天才!正当您对结果赞不绝口时,您的老板打来了电话。他忘了说他还想要您在分析中包括任何产品退货。他希望这项工作不会太麻烦。利用工具包中的 Tableau Prep Builder,这完全不成问题!

清理产品退货数据

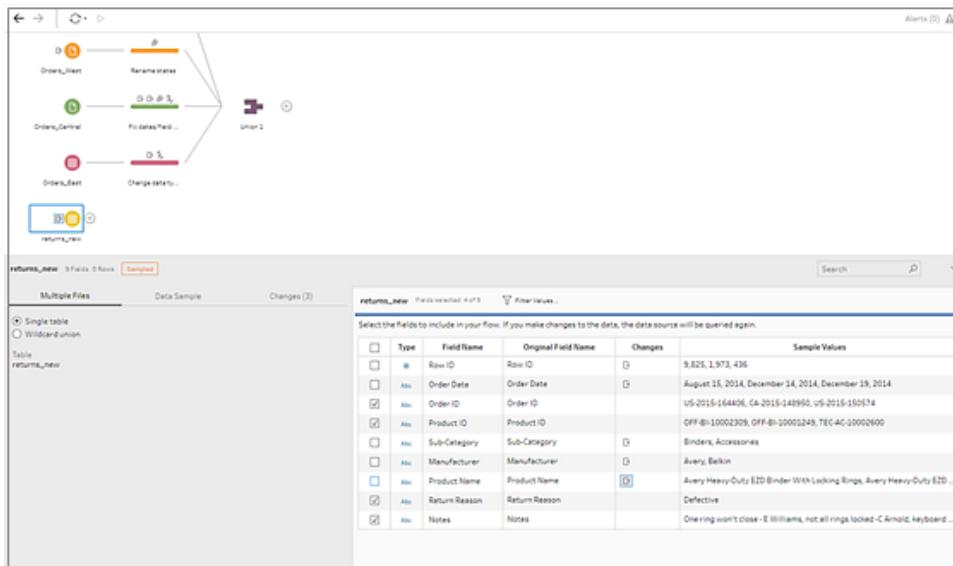
您检查老板发给您的产品退货相关 Excel 文件,发现文件有一点杂乱。您将新文件 **return_reasons_new** 添加到您的流程来仔细检查。

1. 在“连接”窗格中，单击“添加连接”。选择“Microsoft Excel”并导航到示例 Superstore 数据文件(请参见[示例文件 在本页 89 以下载该文件](#))。

2. 选择 **return reasons_new.xlsx**，然后单击“打开”将该文件添加到流程窗格。

此文件中只有 4 个您想要包括在流程中的字段：“Order ID”(订单 ID)、“Product ID”(产品 ID)、“Return Reason”(退货原因)和“Notes”(注释)。

3. 在“returns_new”的“输入”窗格中，清除字段网格顶部的复选框以清除所有复选框。然后，选中“Order ID”(订单 ID)、“Product ID”(产品 ID)、“Return Reason”(退货原因)和“Notes”(注释)字段的复选框。



4. 重命名输入步骤以更好地反映此输入中包括的数据。在“流程”窗格中，双击输入步骤名称“Returns_new”，并键入“Returns (all)”(退货(全部))。

通过查看示例字段值，您注意到“Notes”(注释)字段似乎有很多不同的数据合并在一起。

您要在此文件中执行一些清理，然后才能对数据进行进一步处理，因此，您添加了清理步骤来进行检查。

5. 在“流程”窗格中，选择输入步骤“Returns (all)”(退货(全部))，单击加号 \oplus 图标，然后添加一个清理步骤。

在“配置”窗格中，调整“Notes”(流程)字段的大小以便更好地能看到各个条目。为此，请单击并将字段的右侧外缘向右拖。

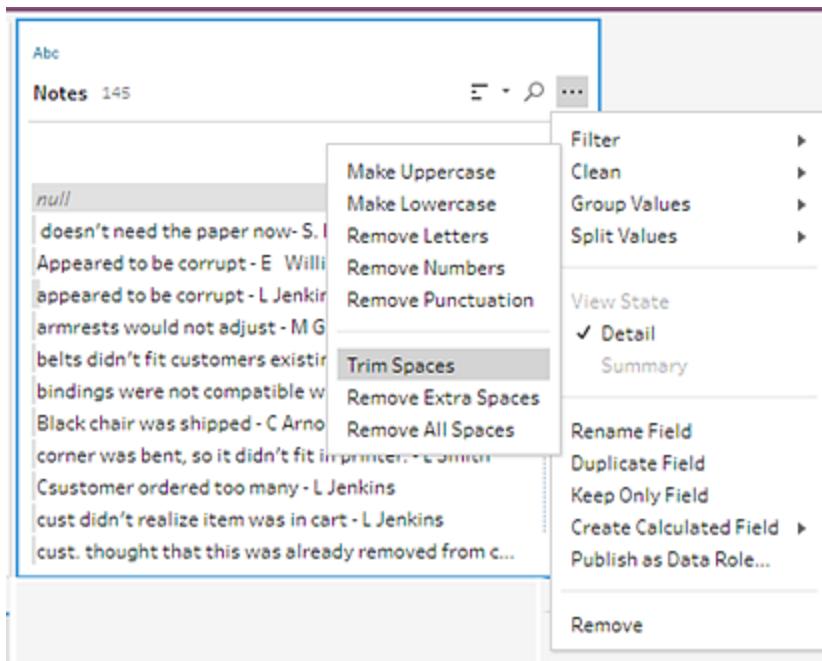
6. 在“Notes”(注释)字段中，使用字段值右侧可视的滚动条来浏览各个值。

您注意到一些内容有问题：

- 某些条目的内容中有额外的空格。这可能会导致字段被读取为 null 值。
- 审批者的名称似乎包括在退货注释条目中。为了更好地使用此数据，您需要将该信息放在一个单独的字段中。

为了处理额外的空格，您记得有一个用于移除尾随空格的清理选项，因此决定尝试一下，看看该选项是否能修复问题。

7. 选择“Notes”(注释)字段。单击“更多选项”… 菜单(在以前的版本中为下拉箭头)，并选择“清理”>“剪裁空格”。

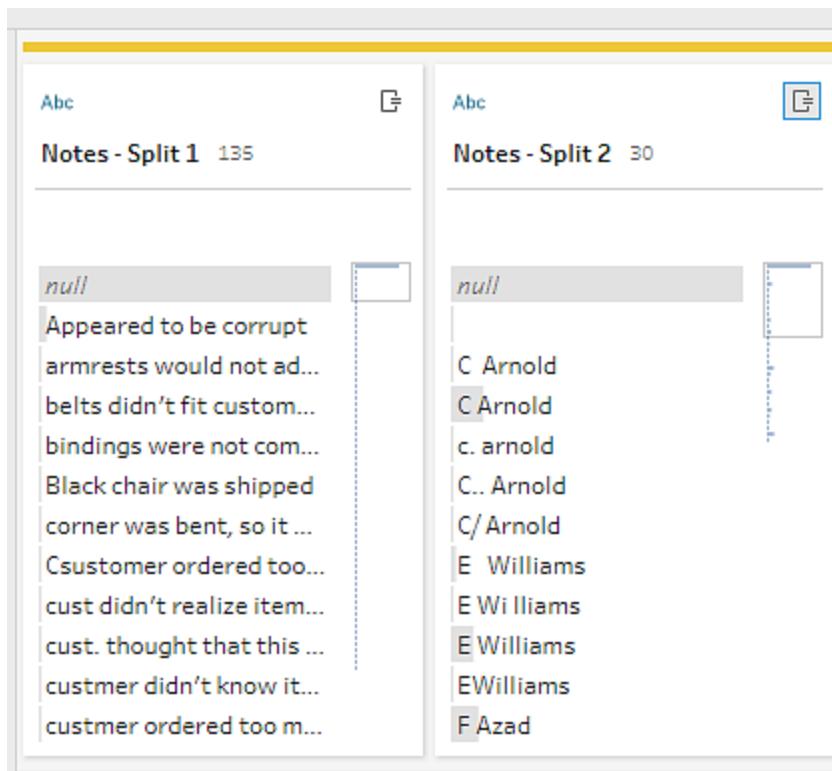


太好了！它准确地完成了所需的操作。额外的空格不见了。

接下来，您想要为审批者名称创建一个单独的字段。您看到菜单中的“拆分值”选项，因此决定尝试该选项。

8. 选择“Notes”(注释)字段。单击“更多选项”… 菜单(在以前的版本中为下拉箭头)，并选择“拆分值”>“自动拆分”。

此选项准确地完成了您希望其进行的操作。它自动将退货注释和审批者姓名拆分到单独的字段中。



就像 Tableau Desktop 一样，Tableau Prep Builder 也为这些字段自动分配了名称。因此您将需要将新字段重命名为有意义的名称。

9. 选择字段“**Notes-Split 1**”(注释-拆分 1)。在字段名中双击，并键入“**Return Notes**”(退货注释)。
10. 为第二个字段重复此步骤，并将其重命名为“**Approver**”(审批者)。
11. 最后，在您不再需要原始“**Notes**”(注释)字段时将其移除。选择“**Notes**”(注释)字段，单击“更多选项”… 菜单(在以前的版本中为下拉箭头)，并从菜单中选择“**移除**”(在以前的版本中为“**移除字段**”)。

The screenshot shows five data sources in Tableau Prep:

- Return Notes**: 135 rows, containing various notes such as "Appeared to be corrupt", "belts didn't fit custom...", etc.
- Approver**: 30 rows, showing names like null, C Arnold, c. arnold, C. Arnold, C/ Arnold, E Williams, E WIlliams, EWilliams, and FAzad.
- Order ID**: 263 rows, with IDs like CA-2015-100762, CA-2015-102652, etc.
- Product ID**: 237 rows, with IDs like FUR-BO-1000362, FUR-BO-10002268, etc.
- Return Reason**: 6 rows, with reasons like Defective, General, Incorrect Product Ordered, etc.

通过查看新的“**Approver**”(审批者)字段，您注意到字段值列出了相同名称，但这些名称的输入方式不同。您想要将这些字段分组，以消除同一值的多个变体。

也许“对值进行分组”(以前版本中的“分组和替换”)选项可帮助完成该操作？

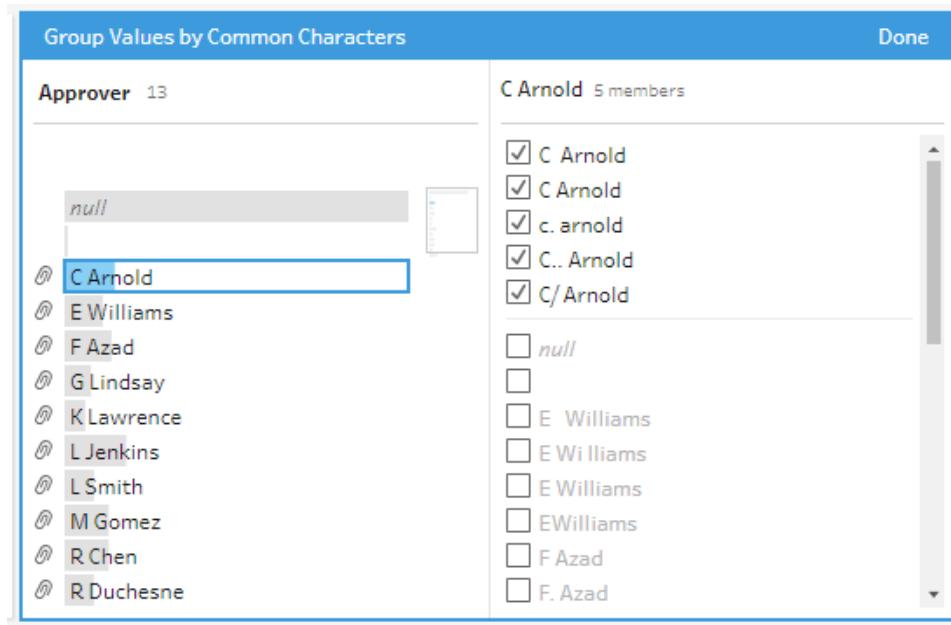
您记得有一个用于“共同字符”的选项。由于这些值共用相同字母，因此您决定试一下。

12. 选择“**Approver**”(审批者)字段。单击“更多选项”… 菜单(在以前的版本中为下拉箭头)，并选择“对值进行分组”(以前版本中的“分组和替换”) >“共同字符”。

The screenshot shows the context menu for the Approver field in Tableau Prep. The "Group Values by" submenu is open, with "Common Characters" selected. Other options in the submenu include Manual Selection, Pronunciation, and Spelling. The main menu also includes Filter, Clean, and Split Values.

此选项将每个名称的所有变体分组在一起。这完全就是您想要达到的目的。

检查了其他名称来确保其正确分组后，您单击“完成”关闭“对值进行分组”编辑器。



此文件看起来非常好。

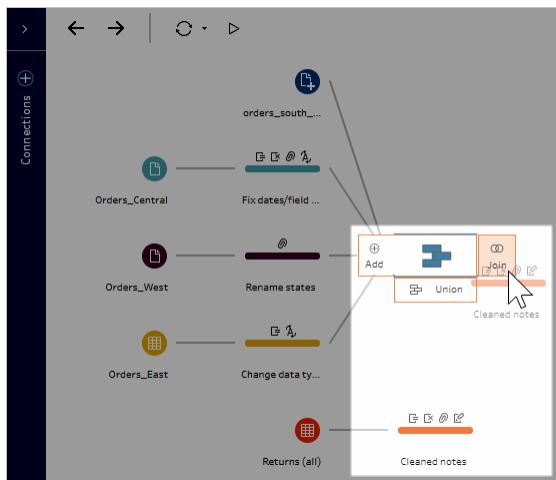
13. 为清理步骤命名以跟踪您的工作。例如，“**Cleaned notes**”(清理的注释)。

既然产品退货数据全部已清理，那么您想要将退货数据添加到合并的文件中的订单数据。但其中许多字段在合并的文件中不存在。若要将这些字段(数据列)添加到合并的数据集，您需要使用联接。

联接数据

在联接数据时，这些文件至少必须有一个共同的字段。您的文件共用“**Order ID**”(订单 ID) 和“**Product Id**”(产品 Id) 字段，因此您可以在这些字段上联接，以查看共有这些字段的所有行。在您使用拖放创建并集时，您记得有一个用于创建联接的选项，因此您试了一下。

1. 在“流程”窗格中，将“**Cleaned notes**”(清理的注释) 步骤拖到“**All orders**”(所有订单) 合并步骤上，并将其放在“联接”上。



在联接文件时，Tableau Prep Builder 会在“**联接配置**”中显示联接的结果。

处理联接可能会很棘手。您通常想要有一个具有联接中所包括因素的清晰视图，例如用于联接文件的字段、结果中的包括的行数，以及未包括或为 **null** 值的任何字段。

在 Tableau Prep Builder 中查看联接的结果时，您很高兴地看到有如此多的信息，并且可以随时交互。

Table Names	Return Notes	Approver	File Paths	Order ID	Product ID	Return Notes
orders_Central.csv		null	orders_south_2015.csv	CA-2016-11948	FUR-AP-10002311	Gen
orders_East.csv	Customer meant to order something else	E Williams	orders_south_2015.csv	CA-2018-129707	FUR-AR-10002426	Inco
orders_East.csv	Customer found comparable product at local store	R Chen	orders_south_2015.csv	CA-2018-159327	FUR-AP-10001492	No
orders_East.csv	Customer ordered too many	K Lawrence	orders_south_2015.csv	CA-2017-109691	FUR-BI-100003291	Def
orders_East.csv	customer got item cheaper at online discount store	M Gomez	orders_south_2015.csv	CA-2017-145583	FUR-FU-10001706	No
orders_East.csv		R Duchesne	orders_south_2015.csv	CA-2018-159158	FUR-FU-10001706	No

提示: 在联接配置最左侧的窗格中，您可以浏览联接并与之交互。您也可以在“**联接子句**”窗格中直接编辑值，并在“**联接结果**”窗格中执行清理操作。

在“**联接类型**”图表中单击以尝试不同的联接配置，并在“**联接结果汇总**”部分中查看联接中为每个表包括或排除的行数。

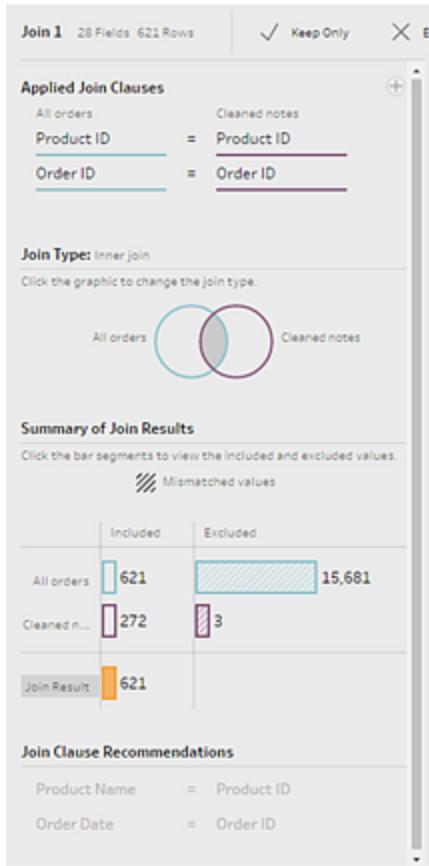
在“已应用联接子句”部分中选择要进行联接的字段，或通过“**联接子句建议**”部分添加建议的联接子句。

有关使用联接的详细信息，请参见[聚合、联接或合并数据 在本页 297](#)。

您看到有超过 13,000 行被从“**All Orders**”(所有订单)文件中排除。创建联接之后，Tableau Prep Builder 自动对“**Product ID**”(产品 ID)字段进行了联接，但您还想对“**Order ID**”(订单 ID)字段进行联接。

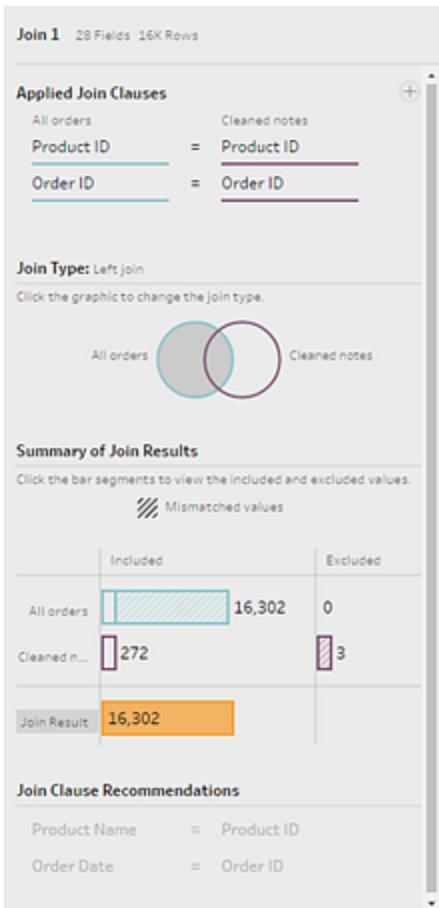
在浏览联接配置的左侧窗格时，您看到建议的联接子句列表中有“**Order ID**”(订单 Id)，因此您从该处快速添加了该字段。

2. 在联接配置左侧窗格中的“**联接子句建议**”部分，选择“**Order ID = Order Id**”，并单击加号  按钮添加联接子句。



由于“**联接类型**”设置为内部联接(Tableau Prep Builder 的默认设置)，因此联接仅包括两个文件中均存在的值。但您需要**“Orders”(订单)**文件中的所有数据，以及这些文件的退货数据。因此您将需要更改联接类型。

- 在“**联接类型**”部分，单击图表的一侧以包括所有订单。在下面的示例中，单击图表的左侧以将联接类型更改为“左”联接，并包括**“All orders”(所有订单)**合并步骤中的所有数据，以及**“Cleaned notes”(清理的注释)**步骤中的任何匹配数据。



现在，您有了销售订单文件中的所有数据，以及适用于这些订单的任何退货数据。您查看“**联接子句**”窗格，并看到一些其他文件中不存在的不同值。

例如，有很多没有对应退货数据的订单行(显示为红色)。您能够浏览此详细级别的联接相关信息，这让您非常喜欢。

您急于想在 Tableau Desktop 中开始分析此数据，但在开始分析之前，您注意到联接要有一些结果需要清理。幸好您知道怎么做！

提示：不知道数据是否足够干净？您可以从流程的任何步骤中通过 Tableau Desktop 预览数据，对其进行检查。

只需在“**流程**”窗格中右键单击该步骤，并从菜单中选择“**在 Tableau Desktop 中预览**”。

您可能会遇到您的数据和在 Tableau Desktop 中进行的更改未写回 Tableau Prep

Builder 中的数据源的情况。有关详细信息，请参见在 [Tableau 中查看数据样本 在本页 309](#)。

4. 在开始清理联接结果之前，将联接步骤命名为“**Orders+Returns**”(订单+退货)，并保存流程。

清理联接结果

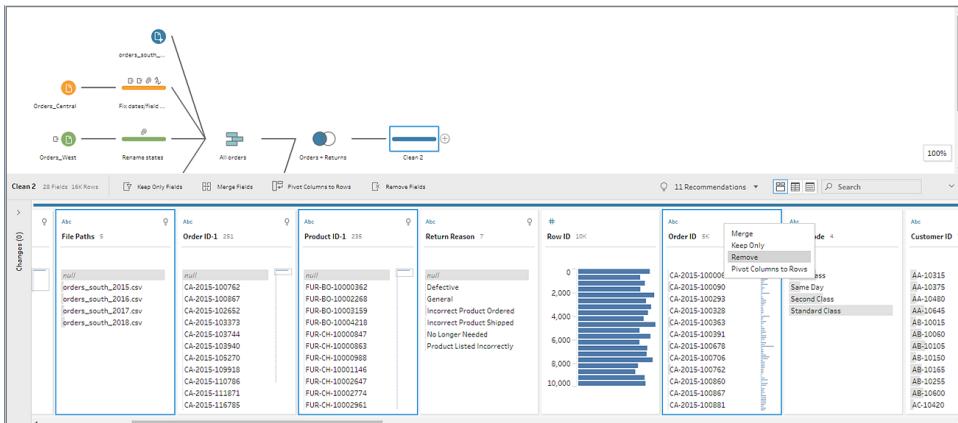
注意: 若要清理联接中的字段，您可以直接在“联接”步骤中执行清理操作。对于本教程，我们将添加一个清理步骤，以便您能够清晰地看到您的清理操作。如果要尝试直接在联接步骤中执行这些步骤，请跳过下面的步骤 1 和 3。

联接了两个步骤之后，系统为两个表都添加了“Order ID”(订单 ID) 和“Product ID”(产品 ID) 共同字段。

您想要保留所有订单中的“Product ID”(产品 ID) 字段和退货文件中的“Order ID”(订单 ID) 字段，并移除来自这些字段的重复字段。您也不需要输出文件中的“**File Paths**”(文件路径) 和“**Table Names**”(表名称) 字段，因此也需要移除这些字段。

提示: 当您使用两个文件中均存在的字段联接表时，Tableau Prep Builder 将会引入这些字段，并通过向字段名中添加“-1”或“-2”来重命名第二个文件中的重复字段。例如，“**Order ID**”(订单 ID) 和“**Order ID-1**”(订单 ID-1)。

1. 在“流程”窗格中，选择“**Orders+Returns**”(订单+退货)，单击加号  图标，并添加清理步骤。
2. 在“配置”窗格中，选择并移除以下字段：
 - **Table Names**(表名称)
 - **Order ID**(订单 ID)
 - **File Paths**(文件路径)
 - **Product ID-1**(产品 ID-1)



3. 将字段“Order ID-1”(订单 Id-1)重命名为“Order ID”(订单 ID)。

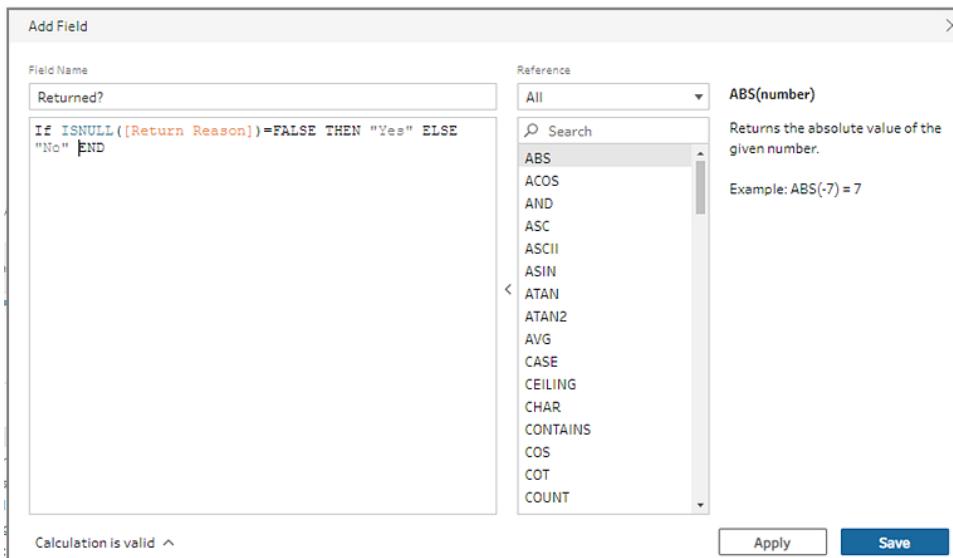
有很多产品已退货的 null 值，但没有指明退货注释或审批者。为了使此数据更易于分析，您想要添加一个值为“**Yes**”和“**No**”的字段来指明产品是否已退货。

您没有此字段，因此可以通过创建计算字段来添加它。

4. 在工具栏中，单击“创建计算字段”。

5. 将字段命名为“**Returned?**”(已退货？)，然后输入以下计算并单击“保存”。

```
If ISNULL([Return Reason])=FALSE THEN "Yes" ELSE "No" END
```

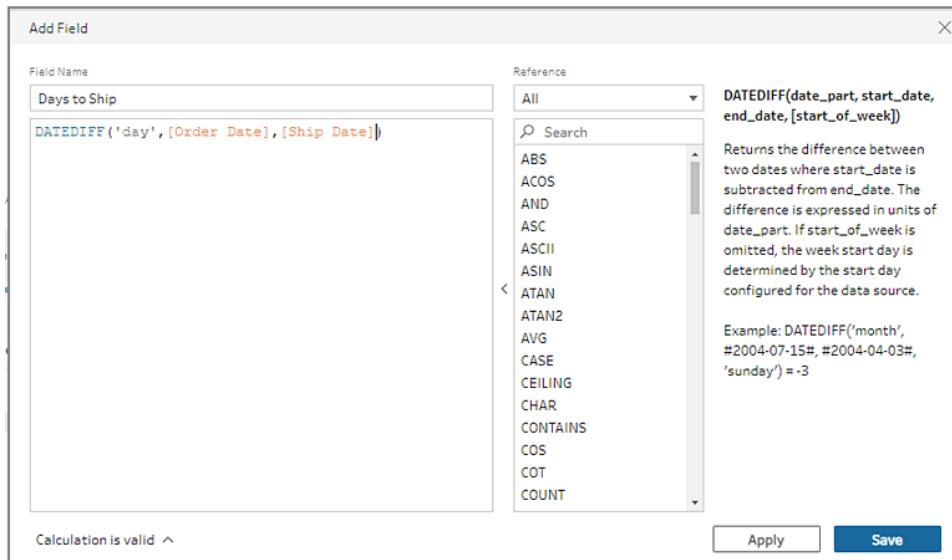


为了进行分析，您还想要知道订单发货天数，但您也没有该字段。

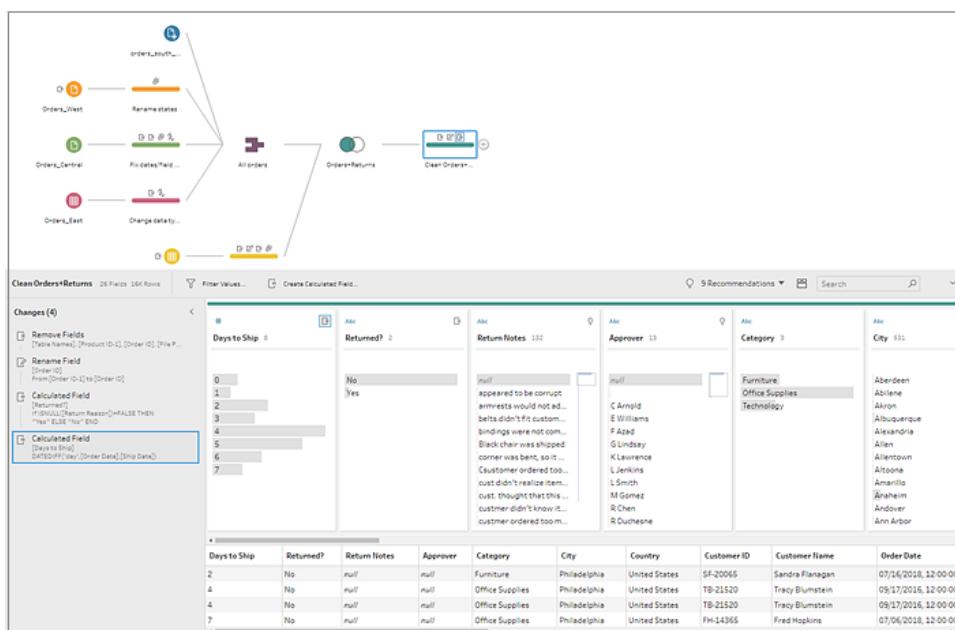
不过您有创建该字段所需的全部信息，因此添加了另一个计算字段来创建它。

6. 在工具栏中，单击“创建计算字段”。
7. 将字段命名为“Days to Ship”(发货天数)，然后输入以下计算并单击“保存”。

```
DATEDIFF('day',[Order Date],[Ship Date])
```



8. 将步骤命名为“Clean Orders +Returns”(清理订单+退货)。
9. 保存您的流程。



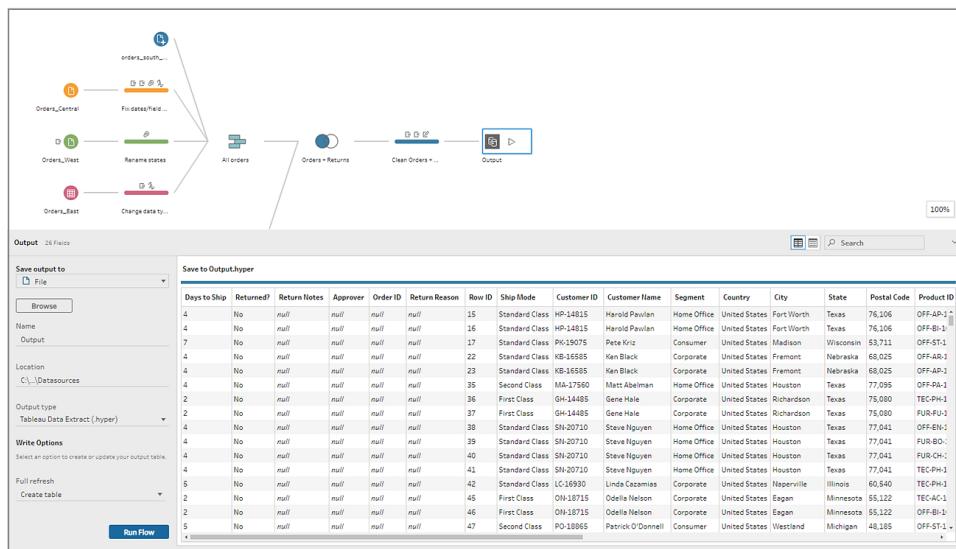
5. 运行流程并生成输出

您的数据看起来很好，并且您已准备好生成输出文件以便开始在 Tableau Desktop 中对其进行分析。您只需运行流程并生成数据提取文件即可。为此，您需要添加一个**输出**步骤。

1. 在“流程”窗格中，选择**“Clean Orders+Returns”**(清理订单+退货)，单击加号 \oplus 图标，并选择“输出”(以前版本中的“添加输出”)。

添加输出步骤时，“输出”窗格将打开，并显示数据的快照。在这里您可以选择要生成的输出的类型，并指定名称和要保存文件的位置。

默认位置为数据源文件夹中的**“我的 Tableau Prep Builder”**存储库。



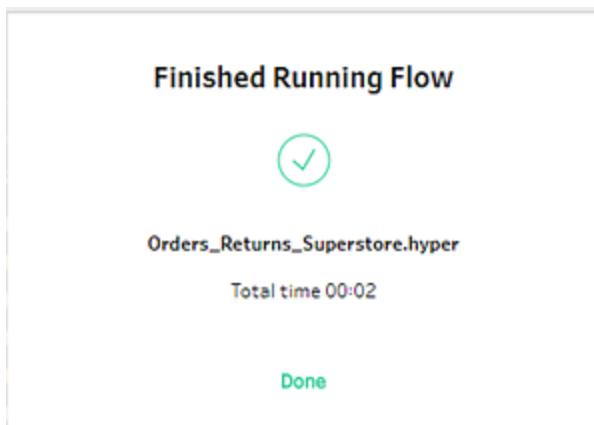
2. 在“将输出保存到”下拉列表左侧的窗格中，选择“文件”(以前版本中的“保存到文件”)。
3. 单击“浏览”按钮，然后在“将数据提取另存为”对话框中输入文件的名称，例如**“Orders_Returns_Superstore”**，并单击“接受”。
4. 在“输出类型”字段中，选择以下输出类型之一。根据您使用的 Tableau Desktop 的版本，您可以从以下选项中进行选择：
 - Tableau 数据提取 (.hyper)，适用于 Tableau Desktop 版本 10.5 及更高版本。
 - Tableau 数据提取 (.tde)，适用于 Tableau Desktop 版本 10.0 至 10.4。
 - 逗号分隔值 (.csv)(如果想要与第三方共享数据提取)。

提示: 在通过流程生成输出时,您有一些选择。您可以生成数据提取文件,或者可以将您的数据以数据源的形式发布到 Tableau Server 或 Tableau Online。有关生成输出文件的详细信息,请参见[创建和发布数据提取及数据源 在本页 310](#)。

5. 在“写入选项”部分,查看将新数据写入文件的选项。您希望使用默认选项(“**创建表**”)并将表替换为流程输出,因而此处没有要更改的内容。

提示: 从版本 2020.2.1 开始,您可以选择要如何将流程数据写回表。您可以从两个选项中进行选择:“**创建表**”或“**追加表**”。默认情况下,Tableau Prep Builder 使用“**创建表**”选项,并在运行流程时用新数据覆盖表数据。如果选择“**追加表**”,Tableau Prep Builder 会将流程数据添加到现有表,以便可以在每次流程运行时同时跟踪新数据和历史数据。有关详细信息,请参见[配置写入选项 在本页 317](#)。

6. 在“输出”窗格中单击“运行流程”,或在流程窗格中单击“运行流程”▷ 按钮以生成输出。
7. 流程完成运行时,一个状态对话框将显示流程运行成功,以及运行所花费的时间。单击“完成”关闭该对话框。



若要使数据保持最新,您可以在 Tableau Prep Builder 中或使用命令行来手动运行流程。如果您有数据管理加载项并启用了 Tableau Prep Conductor,还可以将流程发布到 Tableau Server 或 Tableau Online,并按计划运行流程。

从版本 2020.2.1 开始,您还可以选择在每次运行流程时刷新所有数据,或使用增量刷新运行流程,并且每次只处理新数据。

有关使数据保持最新的详细信息,请参见以下主题:

- [通过命令行刷新流程输出文件 在本页 321](#)
- [将流程发布到 Tableau Server 或 Tableau Online 在本页 378](#)

- 使用增量刷新来刷新流程数据 在本页 314

总结和资源

您是数据准备巨星了！您轻松地找出了不良数据并对其进行了改变！您很快清理并准备好了来自多个数据集的数据，并将其转换为可立即在 Tableau Desktop 中处理来进行分析的井然有序、干净的数据集。

想要进行更多练习？尝试使用下列文件夹中的数据文件重复执行 Superstore 示例流程的其余步骤：

- Orders_South_2015
- Orders_South_2016
- Orders_South_2017
- Orders_South_2018
- Orders_Central
- Orders_East
- Orders_West
- returns_reasons_new
- Quota

在安装 Tableau Prep Builder 后，您还可以在计算机上的以下位置找到这些文件：

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US\Superstore Files
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US/Superstore Files

想要获得更多培训？请观看 Tableau Prep Builder 的[培训视频](#)，查看这些出色的[资源](#)，或参加[现场培训](#)课程。

想要了解我们所讨论主题的详细信息？请查看 Tableau Prep Builder 联机帮助中的其他主题。

关于 Tableau Prep Builder

Tableau Prep Builder 是 Tableau 产品套件中的一个新工具，旨在让您的数据准备工作更加轻松和直观。使用 Tableau Prep Builder 来合并、调整和清理数据，以便在 Tableau 中进行分析。

注意: Tableau Prep 版本 2019.1.2 已更名为 Tableau Prep Builder。如果使用的是产品的早期版本，您仍将在产品中看到“Tableau Prep”。

使用 Tableau Prep Builder

首先，通过各种文件、服务器或 Tableau 数据提取连接到数据。连接到来自多个数据源的数据并合并数据。拖放或双击以将表放入流程窗格，然后添加流程步骤，您可以在其中使用熟悉的操作(例如筛选、拆分、重命名、转置、联接和合并)来清理和调整数据。

流程中的每个步骤都直观地呈现在您创建和控制的流程图中。Tableau Prep Builder 会跟踪每个操作，以便您能够在流程中的任何位置检查工作和进行更改。

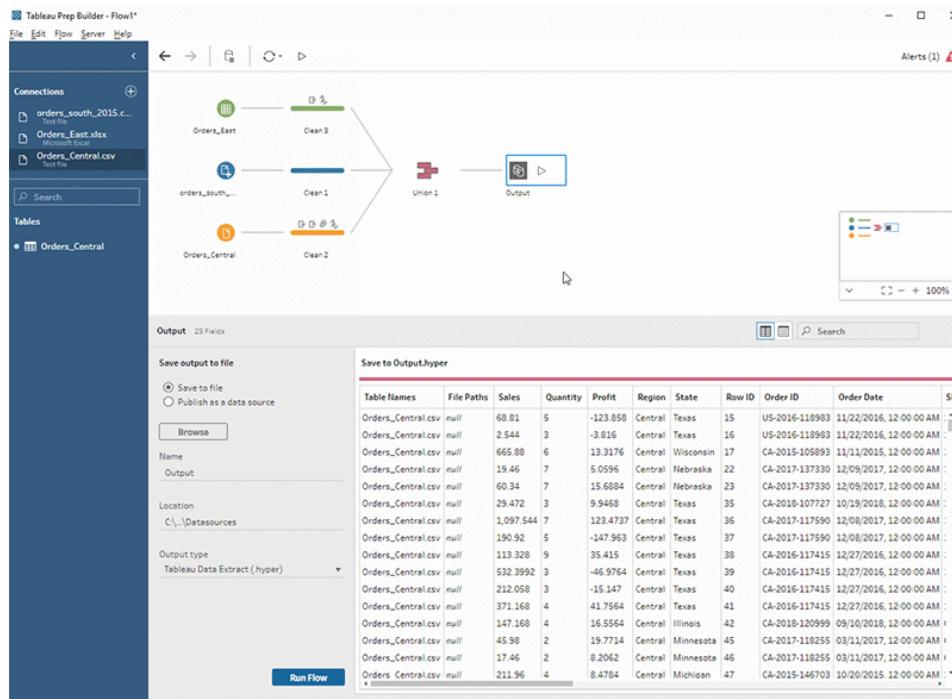
完成流程之后，请运行流程以将操作应用于整个数据集。

Tableau Prep Builder 可与其他 Tableau 产品无缝配合工作。您可以在流程中的任何位置创建数据提取、将数据源发布到 Tableau Server 或 Tableau Online、将流程发布到 Tableau Server 或 Tableau Online 以使用计划刷新数据，或者直接从 Tableau Prep Builder 内打开 Tableau Desktop 来预览数据。

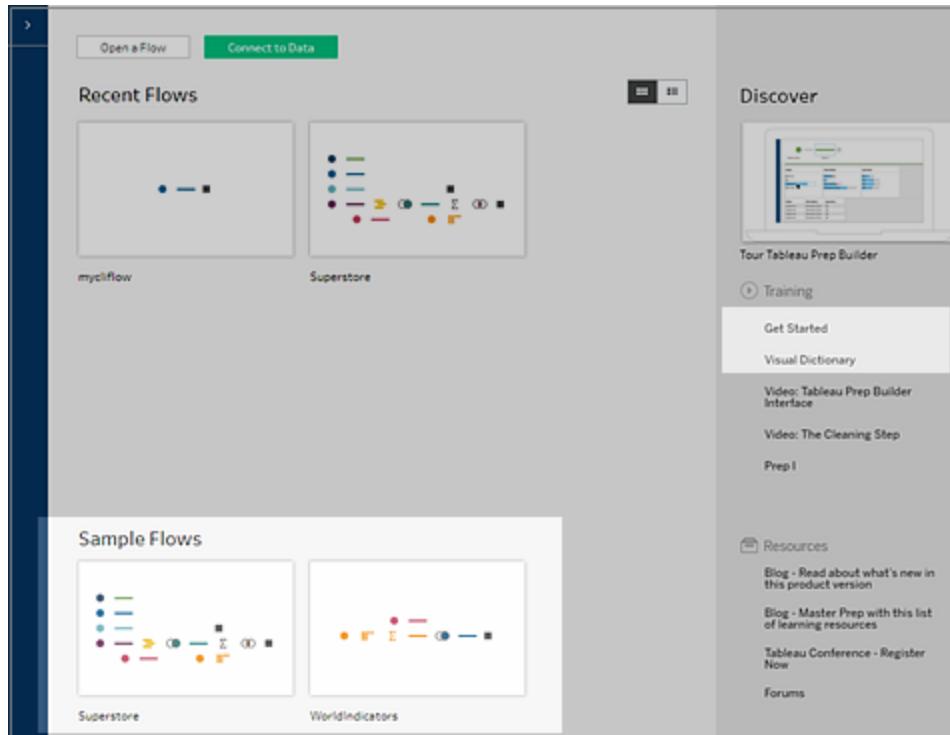
有关安装 Tableau Prep Builder 的信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[安装 Tableau Desktop 或 Tableau Prep Builder](#)。

查看 Tableau Prep Builder 的实际运行效果

单击图像可重播。



准备好试用了吗？从“开始”页面中，单击示例流程之一以浏览和体验各个步骤，尝试 **Tableau Prep Builder 入门指南** 在本页 89 实践教程来了解如何创建流程，或者尝试使用 **Tableau Prep Builder 逐步执行日常生活场景** 在本页 419。



注意: 您可以在以下位置找到流程中使用的示例数据文件:

- **(Windows)** C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en_US
- **(Mac)** /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en_US

若要详细了解 Tableau Prep Builder 如何针对性能优化您的数据, 请参见 [Tableau Prep under the hood](#)(Tableau Prep 后台揭秘)。若要详细了解 Tableau Prep Builder 及其提供的不同特性和功能, 请查看本指南中的各个主题。

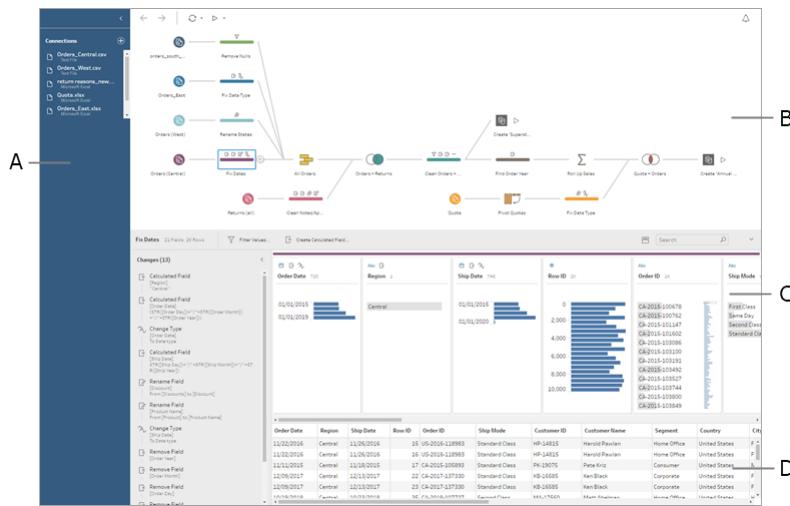
Tableau Prep Builder 工作区一览

Tableau Prep Builder 工作区由可在其中连接到数据源的“**连接**”窗格(A)和三个协调区域组成, 这些区域可帮助您与数据交互以及浏览数据:

- “**流程**”窗格 (B): 在您准备数据时直观地呈现操作步骤。这是添加步骤来构建流程的位置。
- “**配置**”窗格 (C): 汇总了数据样本中的每个字段。查看数据的形状, 并快速查找离群

值和 null 值。

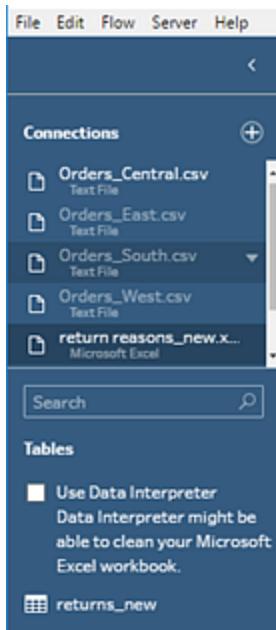
- “数据”网格 (D): 数据的行级别详细信息。



连接到数据并开始构建流程之后，您将在“流程”窗格中添加步骤。这些步骤可深入展示数据的结构，并汇总了应用于数据的操作。每个步骤都表示流程中您定义的一个不同操作类别。

“连接”窗格

工作区的左侧是“连接”窗格，其中显示您连接到的数据库和文件。添加与一个或多个数据源的连接，然后将要使用的表拖到“流程”窗格中。有关详细信息，请参见[连接到数据 在本页 141。](#)

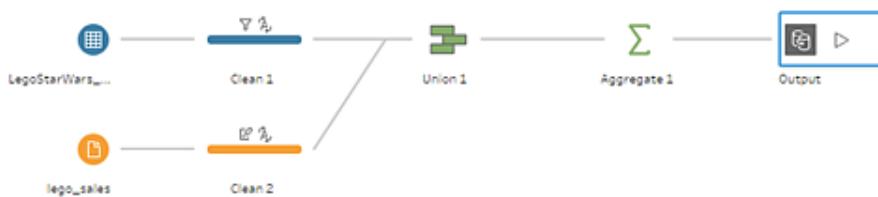


如果需要工作区中有更多空间，您可以最小化“连接”窗格。

“流程”窗格

工作区的顶部是“流程”窗格。这是构建流程的位置。在您连接、清理、调整和合并数据时，步骤会出现在“流程”窗格中，并沿顶部从左到右对齐。这些步骤指出正在应用何种操作、应用操作的顺序以及会对数据产生什么影响。例如，“联接”步骤显示您已应用的联接类型、联接子句、建议使用的联接子句，以及联接中包括的表的字段。

通过将表拖到“流程”窗格中来开始流程。在此处，您可以添加其他数据集、对数据进行转置、合并或联接数据、创建聚合，以及用 .tde 和 .tds 文件、Hyper 数据提取 (.hyper) 文件或可在 Tableau 中使用的已发布数据源格式生成输出文件。有关生成输出文件的详细信息，请参见[保存和共享工作](#) 在本页 309。



注意: 如果在 Tableau Desktop 中对数据进行更改，例如重命名字段、更改数据类型等，这些更改将不会写回到 Tableau Prep Builder。

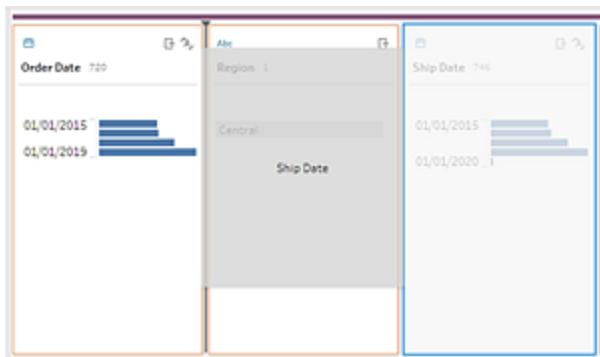
“配置”窗格

工作区的中心是“**配置**”窗格。“**配置**”窗格显示流程中任何位置的数据的结构。可以用多种方式表示数据的结构，具体情况取决于要对数据执行的操作，或您在“**流程**”窗格中选择的步骤。

“**配置**”窗格顶部是一个工具栏，其中显示您可为流程中的每个步骤执行的清理操作。“**配置**”窗格中的每个卡上还将显示一个选项菜单，您可在其中选择可对数据执行的不同操作。

例如：

- 搜索字段、对字段排序以及拆分字段
- 筛选值、包括值或排除值
- 查看并修复 **null** 值
- 重命名字段
- 使用对值进行分组或快速清理操作清理数据条目错误
- 使用自动数据解析来更改数据类型
- 通过将字段列拖放到所需的位置，重新排列字段列的顺序



在“**配置**”卡上选择一个或多个字段值，并右键单击或按住 **Ctrl** 单击 (MacOS) 以查看用于保留或排除值、将选定值分组或将值替换为 **Null** 的其他选项。

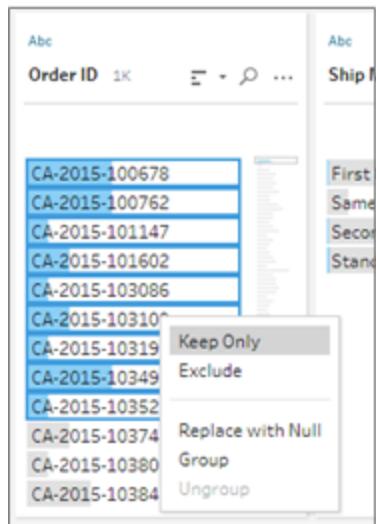
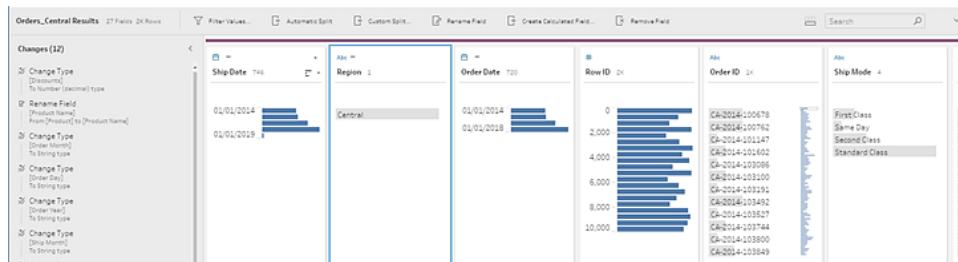


Tableau Prep Builder 会按您进行更改的顺序来跟踪您所做的更改，以便您始终能够在需要时返回并查看或编辑这些更改。使用拖放对操作进行重新排序，按不同的顺序试验和应用更改。



单击窗格右上方的箭头展开和折叠“更改”窗格，腾出更多空间来在“配置”窗格中处理数据。

有关对数据应用清理操作的详细信息，请参见[清理和调整数据 在本页 223](#)。

“数据”网格

工作区的底部是“**数据**”网格，其中显示数据中的行级别详细信息。“**数据**”网格中显示的值反映“**配置**”窗格中定义的操作。如果更愿意在更详细的级别工作，您可以像在“**配置**”窗格中一样在此处执行相同的清理操作。

单击工具栏上的“**折叠配置**”图标折叠(和展开)“**配置**”窗格，以查看您的选项。

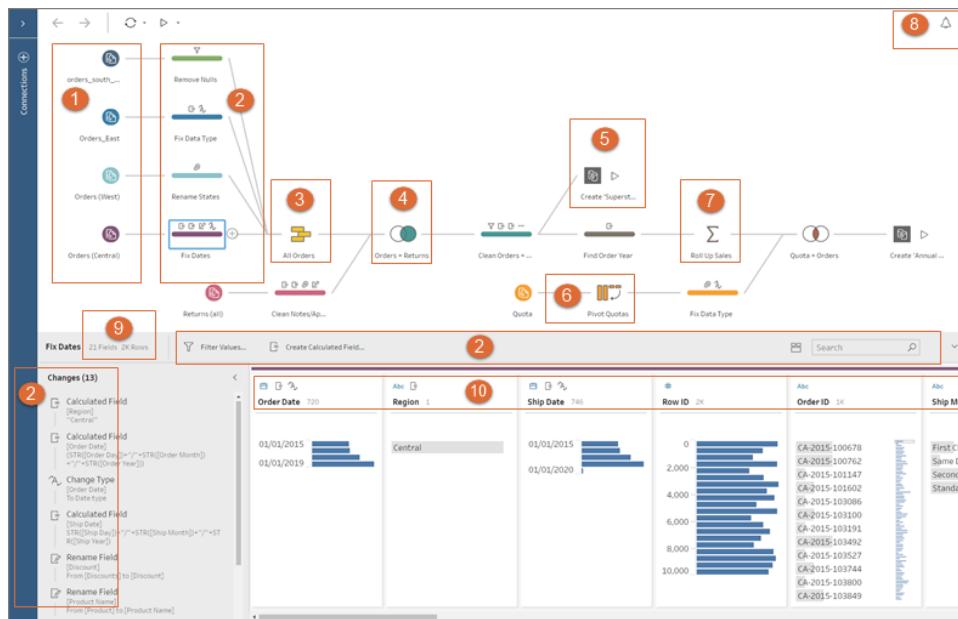
Tableau Prep Builder 如何存储数据

当您将 Tableau Prep Builder 连接到数据并创建流程时，它会将经常使用的数据存储在一个 .hyper 文件中。对于大型数据集，这可能是一个数据样本。任何存储的数据均保存在一个安全的临时文件目录下名为 Prep BuilderXXXXX 的文件中，其中 XXXXX 表示一个全局唯一标识符 (UUID)。保存流程之后，将会删除该文件。有关 Tableau Prep Builder 如何进行数据抽样的详细信息，请参见[选择数据样本大小 在本页 167](#)。

Tableau Prep Builder 还会将数据保存在 Tableau 流程 (.tfl) 文件中，以便支持以下操作(可捕获输入的数据值)：

- 输入步骤中使用的自定义 SQL
- 筛选(针对数据条目)
- 对值进行分组(针对数据条目)
- 计算

Tableau Prep 视觉词典



1

Input Step

Start your flow by dragging data to the Flow pane to create an Input step. The icon shows you the type of data source.

	Data Source
	Data Source with Wildcard Union
	Excel
	Excel with Wildcard Union
	CSV
	CSV with Wildcard Union
	Tableau Extract

Tableau Prep 帮助

	<p>2</p> <p>Clean Step, Changes Pane, and Toolbar</p> <p>Easily keep track of the changes you make to your data. Find these icons on the Clean steps in the Flow pane, in the Changes pane, and on the toolbar.</p> <table border="0"><tr><td> Calculated Field</td><td> Hide Profile Pane</td></tr><tr><td> Change Data Type</td><td> Show Profile Pane</td></tr><tr><td> Edit Value</td><td> Merge Fields</td></tr><tr><td> Exclude Values</td><td> Remove Field</td></tr><tr><td> Filter Values</td><td> Rename Field</td></tr><tr><td> Group Values</td><td> Search</td></tr><tr><td> Keep Only</td><td> Split Fields</td></tr></table>	 Calculated Field	 Hide Profile Pane	 Change Data Type	 Show Profile Pane	 Edit Value	 Merge Fields	 Exclude Values	 Remove Field	 Filter Values	 Rename Field	 Group Values	 Search	 Keep Only	 Split Fields
 Calculated Field	 Hide Profile Pane														
 Change Data Type	 Show Profile Pane														
 Edit Value	 Merge Fields														
 Exclude Values	 Remove Field														
 Filter Values	 Rename Field														
 Group Values	 Search														
 Keep Only	 Split Fields														
	<p>3</p> <p>Union Step</p> <p>Combine up to ten sources of data with similar fields in a single Union step.</p> <p> Union Data</p>														
	<p>4</p> <p>Join Step</p> <p>Connect two steps to join your data on one or more common fields. Select one of these options to choose the join type.</p> <table border="0"><tr><td> Full Anti Join</td></tr><tr><td> Inner Join</td></tr><tr><td> Left Inner Join</td></tr><tr><td> Left Outer Join</td></tr><tr><td> Full Outer Join</td></tr><tr><td> Right Inner Join</td></tr><tr><td> Right Outer Join</td></tr></table>	 Full Anti Join	 Inner Join	 Left Inner Join	 Left Outer Join	 Full Outer Join	 Right Inner Join	 Right Outer Join							
 Full Anti Join															
 Inner Join															
 Left Inner Join															
 Left Outer Join															
 Full Outer Join															
 Right Inner Join															
 Right Outer Join															

5	<p>Output Step</p> <p>Add an Output step and run your flow to apply the changes to the complete data set and generate output files.</p> <ul style="list-style-type: none"> CSV Published Data Source Tableau Data Extract▷ Run Flow
6	<p>Pivot Step</p> <p>Add a Pivot step to change columns to rows.</p> <ul style="list-style-type: none"> Pivot Data
7	<p>Aggregate Step</p> <p>Add an Aggregate step to group and aggregate your data, which changes the level of detail of your data.</p> <ul style="list-style-type: none"> Aggregate Data

Tableau Prep 帮助

8	<p>Notification</p> <p>If there's a problem with your flow or something you need to know, check notifications. Errors include a Go to Error link to help you quickly find the problem.</p> <p> No Notifications Notification Alert Error in the Step</p> <p>...</p>																		
9	<p>Profile Pane</p> <p>See the exact row count of your data and know when your data is sampled.</p> <p> Only shows when data is sampled</p> <p> Hover over row summary to see exact row count</p>																		
10	<p>Profile Card</p> <p>Identify the data type and see the options available to apply to your data when you select a field in the Profile pane.</p> <table border="0"><tbody><tr><td> Calculated Field</td><td> Rename Field</td></tr><tr><td> Change Data Type</td><td> Search</td></tr><tr><td> Edit Value</td><td> Split Fields</td></tr><tr><td> Exclude Values</td><td> Boolean Data Type</td></tr><tr><td> Filter Values</td><td> Date Data Type</td></tr><tr><td> Group Values</td><td> Date Time Data Type</td></tr><tr><td> Keep Only</td><td> Numeric Data Type</td></tr><tr><td> Merge Fields</td><td> Text Data Type</td></tr><tr><td> Remove Field</td><td></td></tr></tbody></table>	Calculated Field	Rename Field	Change Data Type	Search	Edit Value	Split Fields	Exclude Values	Boolean Data Type	Filter Values	Date Data Type	Group Values	Date Time Data Type	Keep Only	Numeric Data Type	Merge Fields	Text Data Type	Remove Field	
Calculated Field	Rename Field																		
Change Data Type	Search																		
Edit Value	Split Fields																		
Exclude Values	Boolean Data Type																		
Filter Values	Date Data Type																		
Group Values	Date Time Data Type																		
Keep Only	Numeric Data Type																		
Merge Fields	Text Data Type																		
Remove Field																			

连接到数据

若要使用 Tableau Prep Builder 来清理和准备数据，就像在 Tableau Desktop 中一样，请通过连接到数据来启动一个新流程。您也可以打开一个现有流程，并选择在何处停止。

您可以直接在开始页面上查看和访问最近的流程，因此可以轻松地找到正在进行的工作。连接到数据之后，在输入步骤中使用不同的选项来确定要在流程中处理的数据。然后添加清理步骤或其他步骤类型来检查、清理和调整数据。

启动或打开流程

Tableau Prep Builder 支持连接到常用类型的数据，以及 Tableau 数据提取 (.tde 或 .hyper)。在版本 2019.3.1 及更高版本中，您还可以连接到已发布数据源。

在版本 2020.2.2 及更高版本中，除了已发布数据源外，如果有启用了 Tableau Catalog 的 **数据管理加载项**，则可以搜索并连接到存储在 Tableau Server 或 Tableau Online 上的数据库、文件和表。有关 Tableau Catalog 的详细信息，请参见 [Tableau Server](#) 或 [Tableau Online](#) 帮助中的“关于 Tableau Catalog”。

Tableau Desktop 中支持的连接器可能在您的 Tableau Prep Builder 版本中尚不受支持。但每个新版本都会增加新连接器，因此请检查“**连接**”窗格来确定您的连接器是否可用。

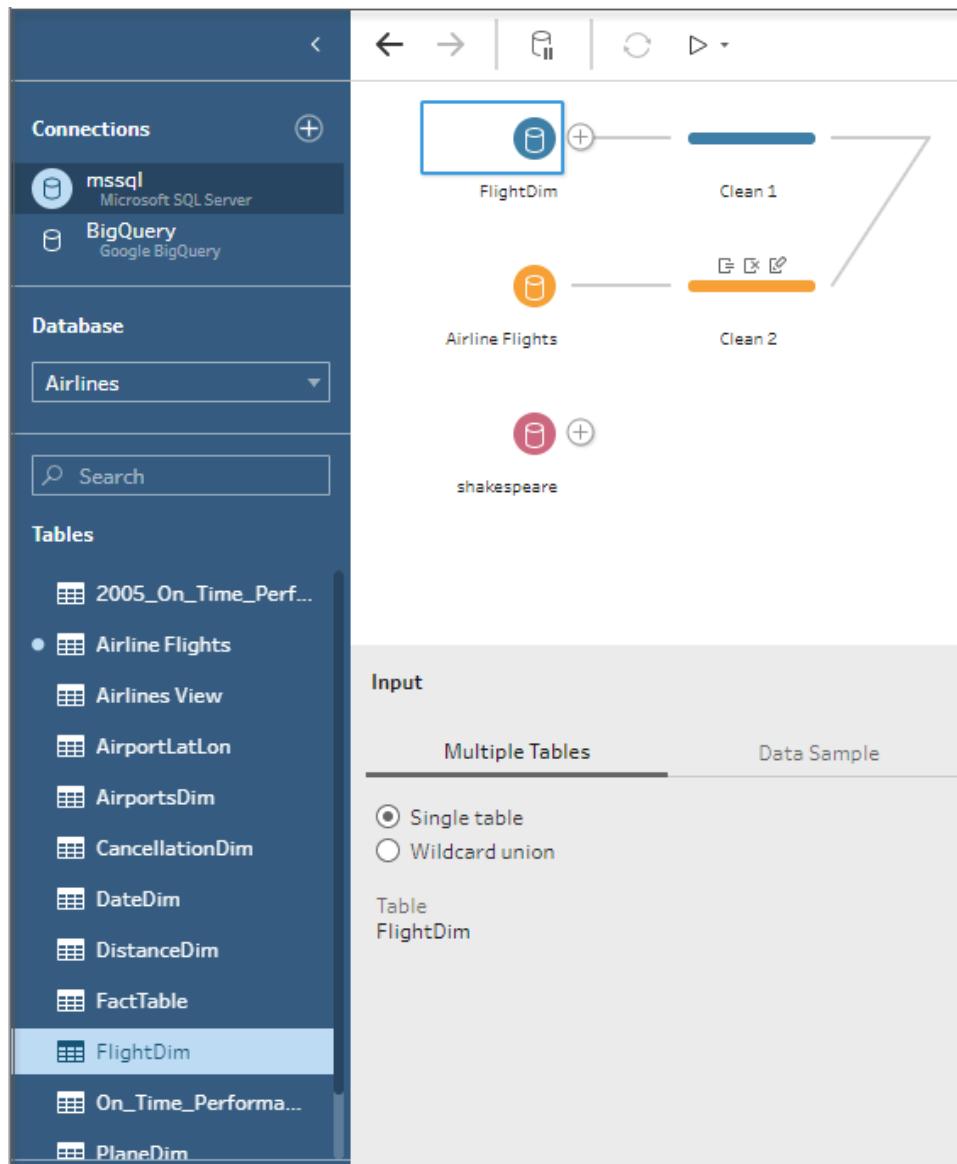
注意：如果在连接器不受支持的版本中打开流程，流程可能会打开，但是，除非移除了数据连接，否则流程可能出现错误或无法运行。

您也可以使用自定义 SQL 查询来连接到数据，就像可在 Tableau Desktop 中所做的一样。有关详细信息，请参见 Tableau Desktop 和 Web 制作帮助中的 [连接到自定义 SQL 查询](#)。

若要检查您是否能连接到数据，请打开 Tableau Prep Builder，并单击“**添加连接**” 按钮，确定您的数据是否列在左侧窗格中的“**连接**”下。

某些连接器可能会要求您下载并安装驱动程序，然后才能连接到数据。请参见 Tableau 网站上的 [驱动程序下载](#) 页面获取驱动程序下载链接以及安装说明。

从版本 2020.1.1 开始，您还可以轻松查看哪些连接器和表与您的输入步骤相关联。单击“**输入**”步骤时，关联的连接器和数据表会在“**连接**”窗格中突出显示。



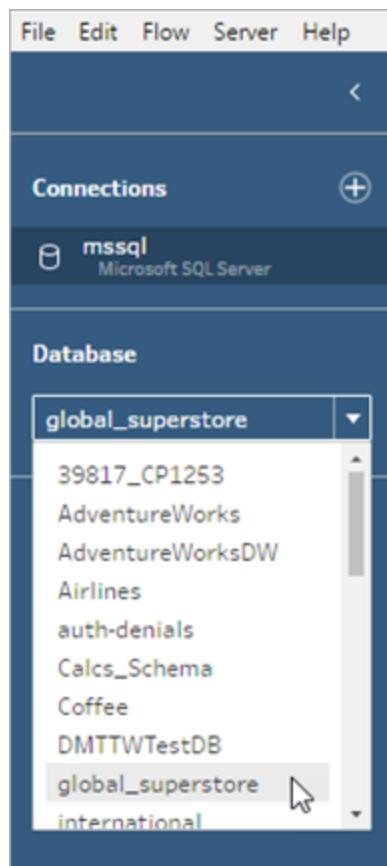
开始新流程

1. 打开 Tableau Prep Builder 并单击“添加连接” 按钮。
2. 从连接器列表中，选择文件类型或托管您的数据的服务器。如果出现提示，请输入所

需的信息以登录并访问数据。

3. 从“连接”窗格中，执行以下操作之一：

- 如果连接到文件，请双击表或将表拖到“流程”窗格以开始流程。对于单个表，Tableau Prep Builder 会在您将数据添加到流程时自动为您在“流程”窗格中创建输入步骤。
- 如果连接到数据库，请选择数据库或架构，然后双击表或将表拖到“流程”窗格以开始流程。



注意：如果连接到单一数据源中的多个文件或数据库表，您可以使用通配符搜索，并在输入步骤中合并文件或表。有关详细信息，请参见[在输入步骤中合并文件和数据库表 在本页 170](#)。

打开现有流程

若要打开现有流程，请在“开始”页面上执行以下操作之一：

- 在“最近的流程”下，选择一个流程。
- 单击“打开流程”，导航到流程文件并将其打开。

单击“文件”>“新建”或“文件”>“打开”打开新工作区。如果要同时处理多个流程，您可以使用这些选项打开多个 Tableau Prep Builder 工作区。在 Tableau Prep Builder 版本 2019.3.1 及更低版本中，如果选择“文件”>“打开”，Tableau Prep Builder 会将当前打开的流程替换为您选择的新流程。

连接到发布的数据源(版本 2019.3.1 及更高版本)

如果从 Tableau Desktop 中或从流程中以输出形式将数据源发布到 Tableau Server 或 Tableau Online，您可以连接到这些发布的数据源，并将它们用作流程的输入数据源。

注意：在发布包括发布的数据源作为输入的流程时，系统会将发布者指派为默认流程所有者。当流程运行时，它将为“**运行身份**”帐户使用流程所有者。有关“运行身份”帐户的详细信息，请参见[运行身份服务帐户](#)。只有站点或服务器管理员才能在 Tableau Server 或 Tableau Online 中更改流程所有者。

使用此功能时，以下要求适用：

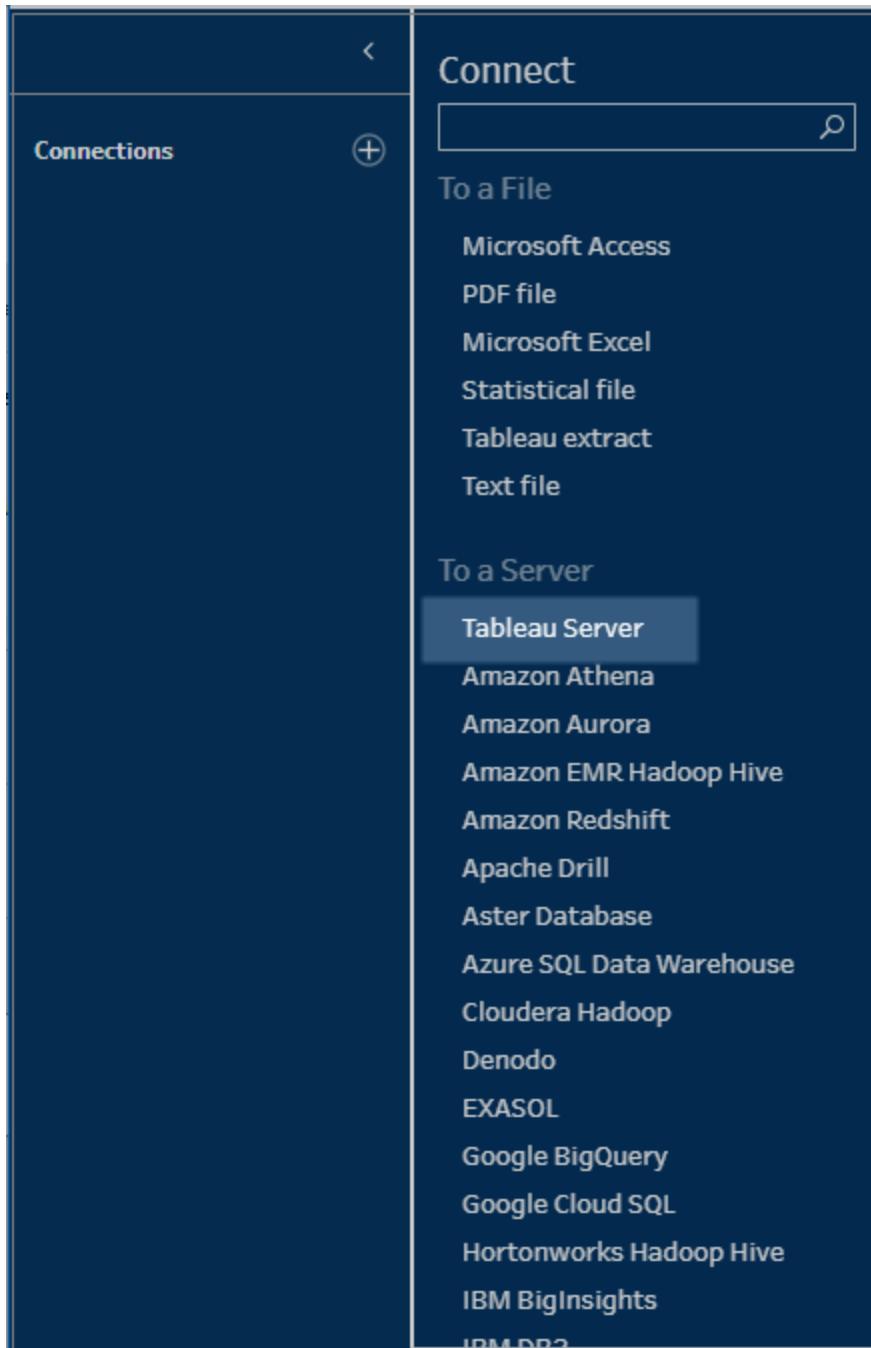
- 完全支持连接到 Tableau Server 版本 2019.3 及更高版本上可用的已发布数据源。如果连接到 Tableau Server 的早期版本上发布的数据源，则所有功能可能不可用。
- 您必须在您登录到的 Tableau Server 站点中分配有 Explorer 或更高角色，才能连接到发布的数据源。有关站点角色的详细信息，请参见 Tableau Server 帮助中的[设置用户的站点角色](#)。
- 必须嵌入凭据才能连接到已发布的数据源。如果没有为数据源嵌入凭据，请更新数据源以包括嵌入的凭据。
- 包含多维(多维数据集)数据或包含用户筛选器的已发布数据源当前不受支持。
- 不支持多服务器连接。必须使用相同的服务器或站点连接来执行以下操作：
 - 连接到发布的数据源。
 - 将流程输出发布到 Tableau Server 或 Tableau Online。

- 计划流程以在 Tableau Server 或 Tableau Online 上运行。

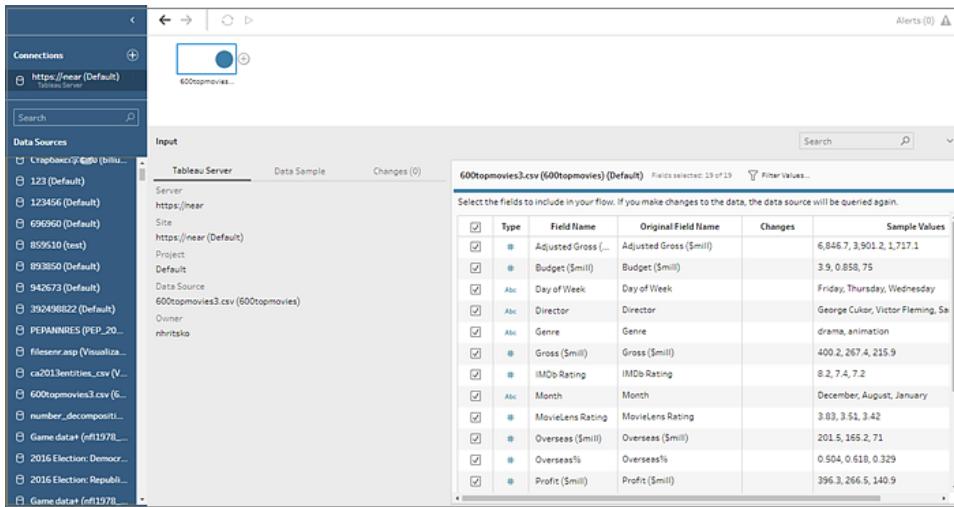
若要连接到已发布数据源，请按照 Tableau Prep Builder 版本的说明进行操作：

版本 2020.2.1 及更低版本

1. 打开 Tableau Prep Builder 并单击“添加连接” 按钮。
2. 从连接器列表中，选择“Tableau Server”。



3. 登录以连接到服务器或站点。
4. 选择数据源或使用搜索选项查找数据源，并将其拖到流程窗格以启动流程。“输入”窗格中的“**Tableau Server**”选项卡显示有关已发布数据源的详细信息。

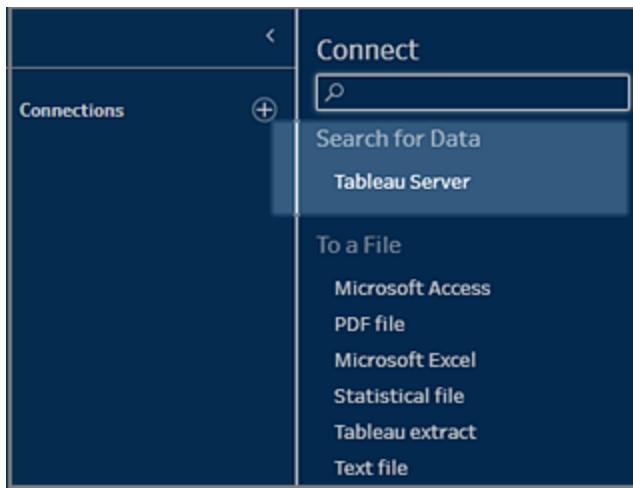


版本 2020.2.2 及更高版本

您可以直接从“连接”窗格连接到存储在 Tableau Server 或 Tableau Online 上的已发布数据源及更多数据源。如果有启用了 Tableau Catalog 的[数据管理加载项](#)，则还可以搜索和连接到数据库和表，以及查看或按有关数据源的元数据(例如说明、数据质量警告和认证)进行筛选。

有关 Tableau Catalog 的详细信息，请参见 [Tableau Server](#) 或 [Tableau Online](#) 帮助中的“关于 Tableau Catalog”。

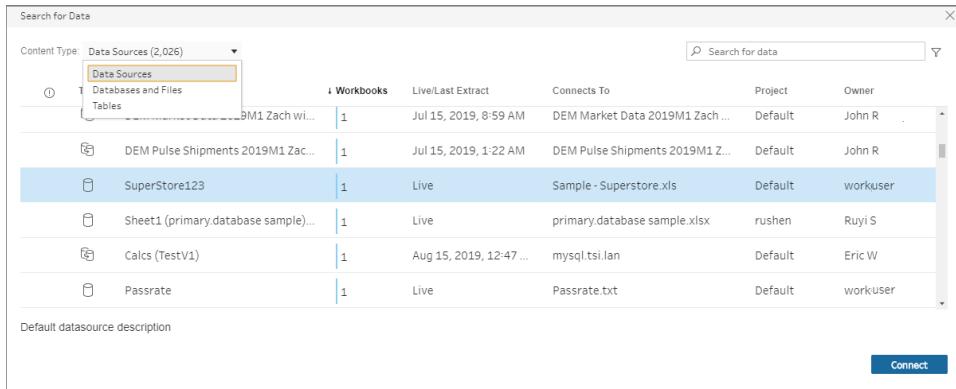
1. 打开 Tableau Prep Builder 并单击“添加连接” 按钮。
2. 在“连接”窗格上的“搜索数据”下，选择“Tableau Server”。



3. 登录以连接到服务器或站点。
4. 在“**搜索数据**”对话框中，从可用已发布数据源的列表中进行搜索。使用筛选器选项按连接类型和认证数据源进行筛选。
5. 选择要使用的数据源，然后单击“**连接**”。

如果您没有连接到数据源的权限，则行和“**连接**”按钮显示为灰色。

注意: 如果没有启用了 Tableau Catalog 的数据管理加载项，则不会显示“**内容类型**”下拉列表。列表中只会显示已发布数据源。



6. 数据源已添加到“流程”窗格中。在“**连接**”窗格中，您可以选择其他数据源或使用搜索选项查找数据源，并将其拖到流程窗格以构建流程。“输入”窗格中的“**Tableau Server**”选项卡显示有关已发布数据源的详细信息。

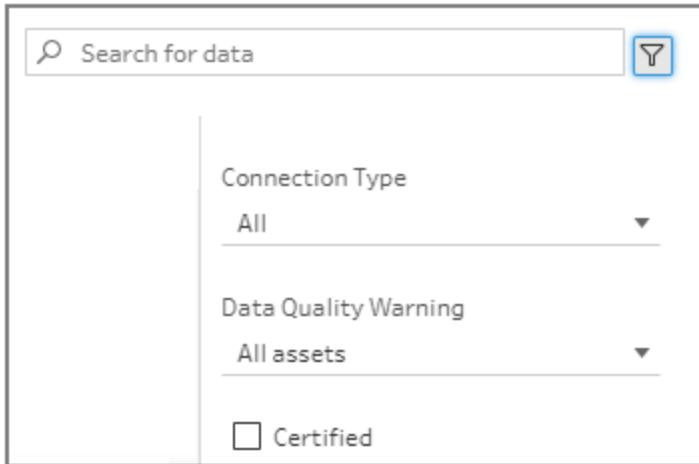
Tableau Prep 帮助

The screenshot shows the Tableau Prep interface. On the left, the 'Connections' and 'Data Sources' panes are visible, listing various connections and sources. In the center, the 'Input' pane shows a selected connection to 'Tableau Server' at 'https://near (Default)'. A preview of the data from '600topmovies3.csv' is displayed on the right, showing columns like 'Type', 'Field Name', 'Original Field Name', 'Changes', and 'Sample Values'. The data includes fields such as Adjusted Gross (\$mill), Budget (\$mill), Day of Week, Director, Genre, Gross (\$mill), IMDb Rating, Month, MovieLens Rating, Overseas (\$mill), Overseas%, and Profit (\$mill).

7. (可选)如果有启用了 Tableau Catalog 的数据管理加载项，请使用“内容类型”下拉列表来搜索数据库和表。

The screenshot shows the Tableau Catalog search interface. At the top, it says 'Search for Data'. Below that, the 'Content Type' dropdown is set to 'Data Sources (2,026)' and is highlighted with a yellow box. The interface shows three categories: 'Data Sources' (which is selected), 'Databases and Files', and 'Tables'. There is also a small help icon and a search bar.

您可以使用右上角的筛选器选项按连接类型、数据质量警告和认证来筛选结果。



连接到云数据源

您可以在 Tableau Prep Builder 中像 Tableau Desktop 那样连接到云数据源，但如果计划发布连接到云数据源的流程并在服务器中按计划运行这些流程，则需要在 Tableau Server 或 Tableau Online 中配置凭据。

您可以在“我的帐户设置”页面的“设置”选项卡中设置凭据，然后使用这些相同的凭据连接到云连接器输入。发布流程时，在“发布”对话框上单击“编辑”以编辑连接，然后在“身份验证”下拉列表中，选择“嵌入”<您的凭据>。

从版本 2020.1.1 开始，可以在发布流时直接通过发布对话框添加凭据，然后在发布时将它们自动嵌入到流程中。有关详细信息，请参见[发布流程 在本页 383](#)。

如果尚未设置已保存的凭据并在“身份验证”下拉列表中选择“提示用户”，则在发布流程后，您必须编辑连接并在 Tableau Server 或 Tableau Online 中的“连接”选项卡中输入凭据，否则流程将在运行时失败。

Tableau Prep Builder 版本 2019.4.1 中新增了以下云连接器：

- Box
- DropBox
- Google Drive
- OneDrive

有关如何使用这些连接器连接到数据的详细信息，请参见 Tableau Desktop 帮助中[特定于连接器](#)的帮助主题。

连接到 Salesforce 数据(版本 2020.2.1 及更高版本)

就像 Tableau Desktop 一样, Tableau Prep Builder 支持使用 Salesforce 连接器来连接到数据, 但有几点不同。

- Tableau Prep Builder 支持您要执行的任何联接类型。
- 当前不支持自定义 SOQL。
- 当前不支持使用标准连接创建自己的自定义连接。
- 不能将默认数据源名称更改为唯一名称或自定义名称。
- 如果计划将流程发布到 Tableau Server 并希望使用已保存的凭据, 则服务器管理员将需要使用 OAuth 客户端 ID 和连接器上的密文配置 Tableau Server。有关详细信息, 请参见 Tableau Server 帮助中的[将 Salesforce.com OAuth 更改为使用已保存凭据](#)。

Tableau Prep Builder 通过创建数据提取来导入数据。Salesforce 当前仅支持数据提取。初始数据提取可能需要一些时间, 具体取决于包含的数据量。在加载数据时, 您将在“输入”步骤中看到计时器。

有关使用 Salesforce 连接器的详细信息, 请参见 Tableau Desktop 和 Web 制作帮助的 [Salesforce](#)。

连接到 Google BigQuery(版本 2019.3.1 及更高版本)

就像 Tableau Desktop 一样, Tableau Prep Builder 支持使用 Google BigQuery 连接到数据。您必须配置 OAuth 凭据, 使 Tableau Prep Builder 能够与 Google BigQuery 通信。如果打算将流程发布到 Tableau Server 或 Tableau Online, 还必须为这些应用程序配置 OAuth 连接。

注意: Tableau Prep Builder 当前不支持使用 Google BigQuery 自定义属性。

- [为 Google 设置 OAuth](#) - 为 Tableau Server 配置 OAuth 连接。
- [OAuth 连接](#) - 为 Tableau Online 配置 OAuth 连接。

如果您在 Mac 上使用 Tableau Prep Builder, 并且正在使用代理连接到 Big Query, 则可能需要修改 SSL 配置以连接到 Google BigQuery

配置 SSL 以连接到 Google BigQuery(仅限 MacOS)

注意: Windows 用户不需要执行额外的步骤。

若要为 Google BigQuery OAuth 连接配置 SSL, 请完成以下步骤:

1. 将代理的 SSL 证书导出到文件，例如 proxy.cer。您可以在 Applications > Utilities > Keychain Access > System > Certificates (under Category) 中找到您的证书。
2. 找到用于运行 Tableau Prep Builder 的 java 版本。例如: /Applications/Tableau Prep Builder 2020.1.app/Plugins/jre/lib/security/cacerts
3. 打开“终端”命令提示符，并针对您的 Tableau Prep Builder 版本运行以下命令：

注意: keytool 命令必须从包含用于运行 Tableau Prep Builder 的 java 版本的目录中运行。在运行此命令之前，您可能需要更改目录。例如，cd /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder 2020.1.1/Plugins/jre/bin。然后运行 keytool 命令。

```
keytool -import -trustcacerts -file /Users/tableau_
user/Desktop/SSL.cer -keystore Tableau Prep Builder
<version>/Plugins/jre/lib/security/cacerts -storepass
changeit
```

```
示例: keytool -import -trustcacerts -file /Users/tableau_
user/Desktop/SSL.cer -keystore Tableau Prep Builder
2020.1.1/Plugins/jre/lib/security/cacerts -storepass
changeit
```

如果在运行 keytool 命令时收到 **FileNotFoundException**(拒绝访问)，请尝试使用提升的权限运行该命令。例如: sudo keytool -import -trustcacerts -file /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder 2020.1.1/Plugins/jre/lib/security/cacerts -storepass changeit。

设置和管理您的 Google BigQuery 凭据

在“输入”步骤中用于连接到 Google BigQuery 的凭据必须与您在 Tableau Server 或 Tableau Online 中适用于 Google BigQuery 的“我的帐户设置”页面的“设置”选项卡中设置的凭据匹配。

如果发布流程时在身份验证设置中选择不同的凭据或不选择凭据，则在 Tableau Server 或 Tableau Online 中编辑流程连接之前，该流程将失败。

若要编辑凭据，请执行以下操作：

1. 在 Tableau Server 或 Tableau Online 中的“**连接**”选项卡上，针对 Google BigQuery 连接单击“**更多操作**”***。
2. 选择“**编辑连接**”。
3. 选择在“**我的帐户设置**”页面的“**设置**”选项卡中设置的已保存凭据。

有关设置和管理凭据的详细信息，请参见以下主题：

Tableau Desktop 和 Web 制作帮助中的[管理您的帐户设置](#)。

[发布流程](#) 在本页 383，了解有关发布流程时设置身份验证选项的信息。

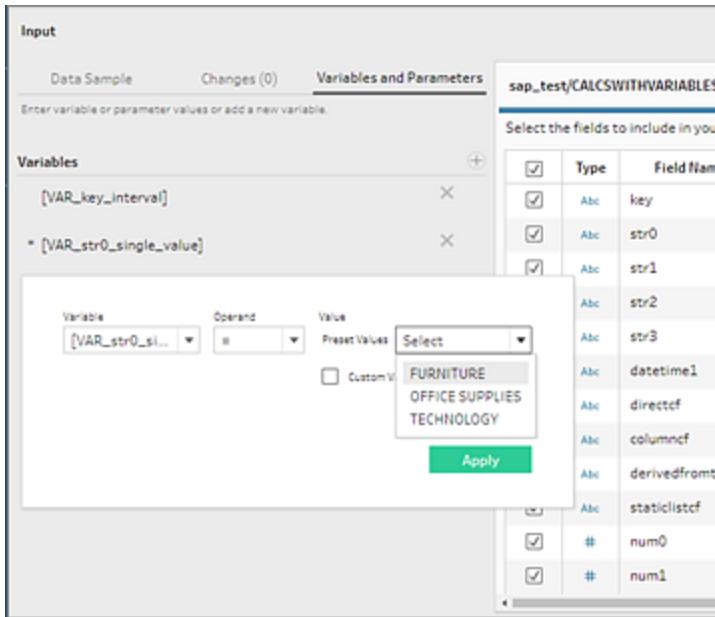
[查看和解决错误](#) 在本页 408，了解有关在 Tableau Server 或 Tableau Online 中解决连接器错误的信息。

连接到 SAP HANA(版本 2019.2.1 及更高版本)

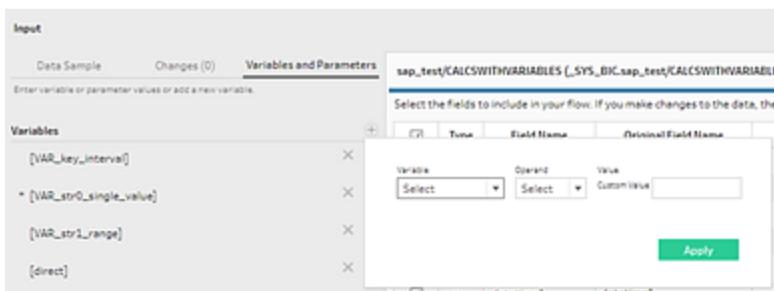
就像 Tableau Desktop 一样，Tableau Prep Builder 支持使用 SAP HANA 连接到数据，但有几点不同。使用在 Tableau Desktop 中使用的相同过程连接到数据库。有关详细信息，请参见[SAP HANA](#)。连接并搜索表之后，将表拖到画布上以开始构建流程。

不支持在 Tableau Prep Builder 中打开流程时的变量和参数提示。相反，在“输入”窗格中，单击“**变量和参数**”选项卡并选择要使用的变量和操作数，然后从预设值列表中选择，或者输入自定义值来查询数据库并返回所需的值。

注意: 从 Tableau Prep Builder 版本 2019.2.2 开始，您可以使用初始 SQL 来查询连接，并且，如果某个变量有多个值，您可以从下拉列表中选择所需的值。



也可以添加其他变量。单击“**变量**”部分中的加号按钮 并选择一个变量或操作数，然后输入自定义值。



注意: Tableau Prep Conductor 版本 2019.1 中不支持此连接器。如果创建使用此连接器的流程，则您可以发布流程，但无法使用 Tableau Server 中的计划功能运行流程。作为替代，您可以使用命令行界面运行流程以使数据保持最新。有关如何通过命令行运行流程的详细信息，请参见[通过命令行刷新流程输出文件](#) 在本页321。有关版本兼容性的详细信息，请参见[与 Tableau Prep 的版本兼容性](#) 在本页333。

使用 ODBC 进行连接(版本 2019.2.2 及更高版本)

如果需要连接到“连接”窗格中未列出的数据源，您可以使用支持 SQL 标准并实现 ODBC API 的“其他数据库 (ODBC)”连接器连接到任何数据源。使用“其他数据库 (ODBC)”连接器连接到数据的工作方式与您在 Tableau Desktop 中使用该连接器的方式类似，但有一些不同之处：

- 您只能使用 DSN(数据源名称) 选项进行连接。
- 若要在 Tableau Server 中发布并运行流程，必须使用匹配的 DSN 配置服务器。

注意: 目前不支持从包含“其他数据库 (ODBC)”连接器的命令行 (Tableau Prep Builder) 中运行流程。

- Windows 和 MacOS 上的连接体验是一致的。不支持 ODBC 驱动程序 (Windows) 的提示连接属性。
- Tableau Prep Builder 仅支持 64 位驱动程序。

连接之前

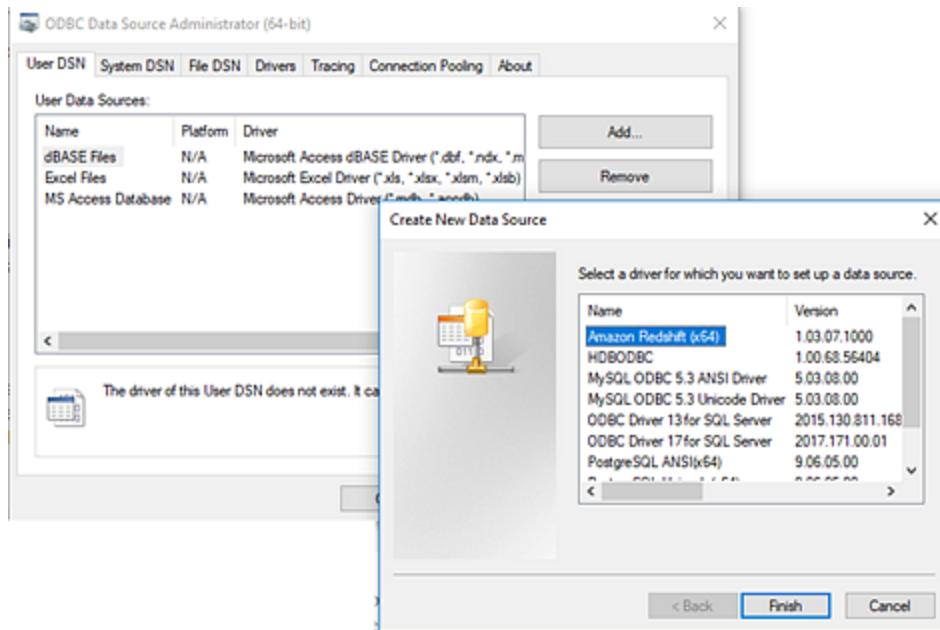
若要使用“其他数据库 (ODBC)”连接器连接到数据，您必须安装数据库驱动程序，并设置和配置 DSN(数据源名称)。若要将流程发布到 Tableau Server 并运行流程，也必须使用匹配的 DSN 配置服务器。

重要信息: Tableau Prep Builder 仅支持 64 位驱动程序。如果已设置并配置了 32 位驱动程序，则您可能需要将其卸载，然后安装 64 位版本(如果驱动程序不允许同时安装两个版本)。

1. 使用 **ODBC 数据源管理器(64 位)** (Windows) 或 ODBC 管理器实用工具 (MacOS) 创建 DSN。

如果 Mac 上未安装该实用工具，您可以下载一个(例如从 www.odbcmanager.net 下载)，或者可以手动编辑 odbc.ini 文件。

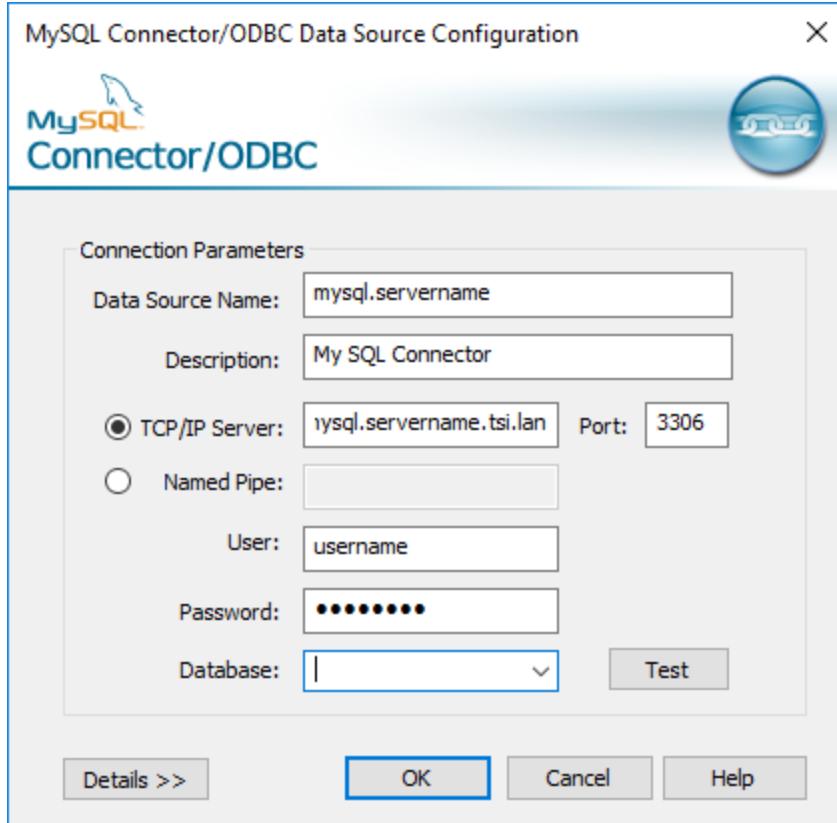
2. 在 **ODBC 数据源管理器(64 位)** (Windows) 或 ODBC 管理器实用工具 (MacOS) 中，添加一个新数据源，然后选择适用于数据源的驱动程序，并单击“完成”。



- 在“**ODBC 驱动程序设置**”对话框中，输入诸如服务器名称、端口、用户名和密码之类的配置信息。单击“**测试**”(如果对话框有该选项)验证您的连接是否设置正确，然后保存您的配置。

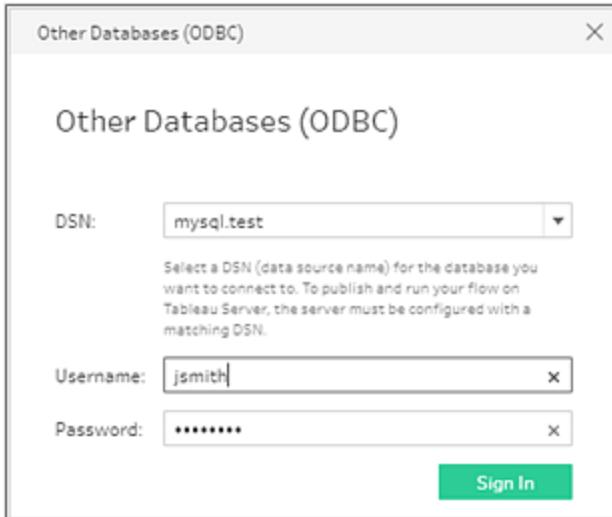
注意：Tableau Prep Builder 不支持提示连接属性，因此您必须在配置 DNS 时设置此信息。

此示例显示 MySQL 连接器的配置对话框。

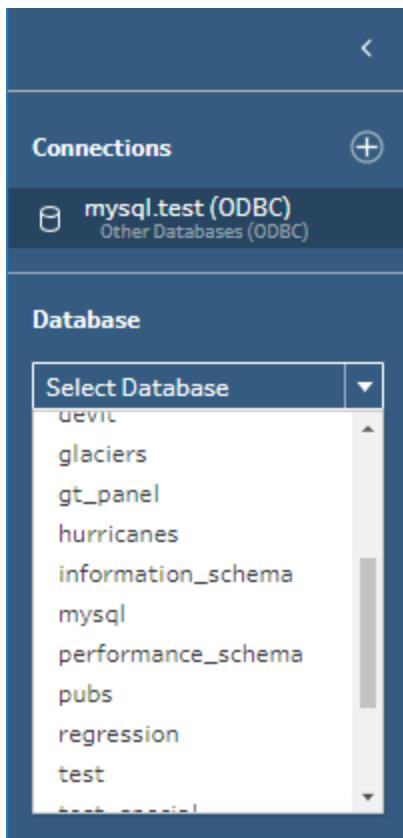


使用“其他数据库 (ODBC)”进行连接

1. 打开 Tableau Prep Builder 并单击“添加连接” 按钮。
2. 从连接器列表中，选择“其他数据库 (ODBC)”。
3. 在“其他数据库 (ODBC)”对话框中，从下拉列表中选择一个 DSN，并输入用户名和密码。然后单击“登录”。



4. 从“连接”窗格中，从下拉列表中选择您的数据库。



使用初始 SQL 查询连接(版本 2019.2.2 及更高版本)

您可以指定初始 SQL 命令，该命令将在连接到支持该命令的数据库时运行。例如，在连接到 Amazon Redshift 时，您可以输入 SQL 语句，以便在连接到数据库时应用筛选器，就像在“输入”步骤中添加筛选器一样。SQL 命令将在数据抽样之前应用，并加载到 Tableau Prep Builder 中。

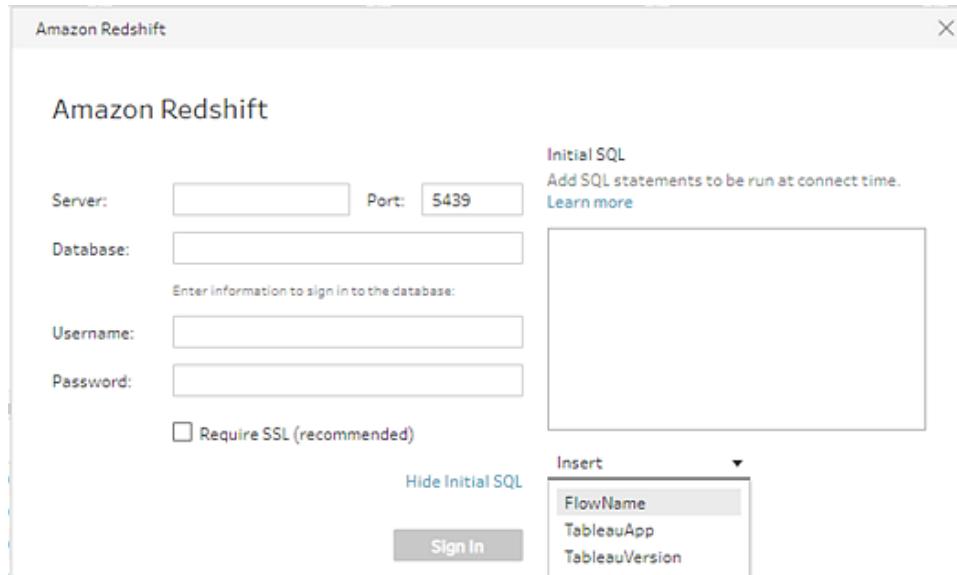
从版本 2020.1.3 开始，还可以包括参数来传递应用程序名称、版本和流程名称数据，以在查询数据源时包括跟踪数据。

运行初始 SQL

若要刷新数据并运行初始 SQL 命令，请执行以下操作之一：

- 更改初始 SQL 命令并通过重新建立连接刷新“输入”步骤。
- 运行流程。初始 SQL 命令在处理所有数据之前运行。
- 计划流程以在 Tableau Server 或 Tableau Online 上运行。初始 SQL 在每次流程运行时加载数据的过程中运行。

注意: 需要[数据管理加载项](#)以在 Tableau Server 或 Tableau Online 上运行流程。有关数据管理加载项的详细信息，请参见[使用数据管理加载项](#)。



1. 在“连接”窗格的列表中选择一个支持初始 SQL 的连接器。
2. 单击“显示初始 SQL”链接展开对话框，并输入您的 SQL 语句。

在初始 SQL 语句中包括参数(版本 2020.1.3 及更高版本)

重要信息: Tableau Prep Builder 版本 2020.2.1 及更低版本中的初始 SQL 不支持参数。如果像在 Tableau Desktop 中一样尝试并输入参数，您将收到错误。

您可以将以下参数传递给数据源，以添加有关 Tableau Prep 应用程序、版本和流程名称的其他详细信息。当前不支持 **TableauServerUser** 和 **TableauServerUserFull** 参数。

参数	说明	返回的值
TableauApp	用于访问数据源的应用程序。	Prep Builder Prep Conductor
TableauVersion	应用程序版本号。	Tableau Prep Builder: 返回确切的版本。例如 2020.1.3 Tableau Prep Conductor: 返回已启用 Tableau Prep Conductor 的主要服务器版本。例如 2020.1
FlowName	Tableau Prep Builder 中 .tfl 文件的名称	示例: Entertainment Data_Cleaned

连接到 Tableau 数据提取

连接到 Tableau 数据提取时, Tableau Prep Builder 将解包数据提取, 并且 hyper 将扩展, 从而会在将流程操作应用于生成的原始数据时使用很多临时空间。

这意味着您需要更多的 RAM 和磁盘空间来容纳该大小的文件。例如, 包含 18 列和 1200 万行、大小为 360MB(解压缩后为 8.5 GB) 的数据提取文件可能最多需要 32GB RAM、16 核和 500GB 可用磁盘空间才能在文件解压缩后支持该文件。

使用 Data Interpreter 清理您的文件

处理 Microsoft Excel 文件时, 您可以使用 Data Interpreter 来检测数据中的子表, 以及移除无关信息来帮助准备数据进行分析。打开 Data Interpreter 时, 它将检测这些子表, 并将其列为“连接”窗格的“表”部分中的新表。

然后, 您可以将它们拖到“流程”窗格中。如果使用 Tableau Prep Builder 版本 2018.1.2 或更高版本, 您可以在“多个文件”选项卡中选择“通配符并集”选项, 以在流程中包括所有找到的子表。有关在输入步骤中使用通配符并集的详细信息, 请参见[在输入步骤中合并文件和数据库表 在本页 170](#)。

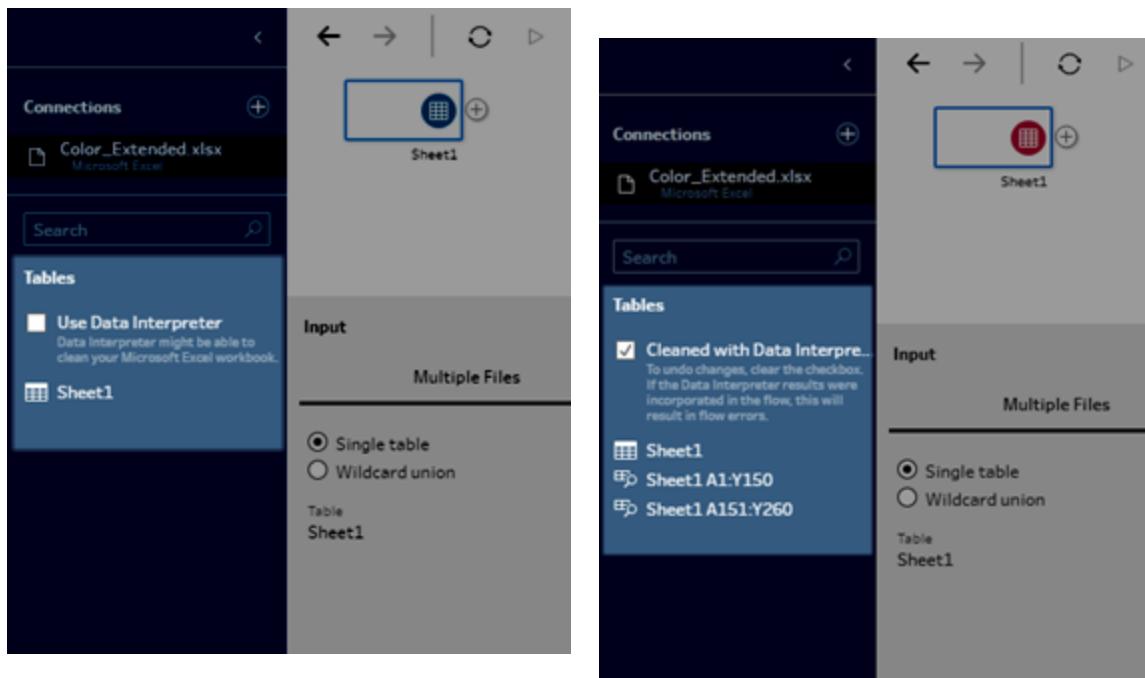
如果关闭 Data Interpreter, 则会从“连接”窗格中移除这些表。如果已在流程中使用了这些表, 则会由于缺少数据而导致流程错误。

注意: 当前 Data Interpreter 只会检测 Excel 电子表格中的子表, 不支持为文本文件和电子表格指定起始行。

下面的示例显示在“连接”窗格中对 Excel 电子表格使用 Data Interpreter 的结果。Data Interpreter 检测到两个额外的子表。

使用 Data Interpreter 之前

使用 Data Interpreter 之后



若要使用 Data Interpreter，请完成以下步骤：

1. 选择“连接到数据”，然后选择“Microsoft Excel”。
2. 选择您的文件，并单击“打开”。
3. 选中“使用 Data Interpreter”复选框。
4. 将新表拖到“流程”窗格以将其包括在流程中。若要移除旧表，请右键单击旧表的输入步骤，并选择“移除”。

配置数据集

若要确定在流程中包括多少数据集，您可以配置数据集。当您连接到数据或将表拖到“流程”窗格中时，会自动向流程中添加一个输入步骤。这始终是流程中的第一步。您可以右键单击输入步骤以对其进行重命名或将其移除。如果连接到 Excel 或文本文件，您还可以从输入步骤中刷新数据。有关如何通过输入步骤刷新数据的详细信息，请参见[在“输入”步骤中添加更多数据 在本页 168](#)。

在输入步骤中，您可以看到有关数据集的详细信息。在这里，您可以搜索字段、查看示例值，以及执行操作来缩小数据集的大小，例如选择要包括的字段、选择要处理的数据样本，或将筛选器应用于所选的字段或行。您还可以通过更改字段的数据类型或字段名来配置字段属性，或者为文本文件配置文本设置。

注意: 包括方括号的字段值会自动转换为圆括号。

	Type	Field Name	Original Field Name	Changes	Sample Values
<input checked="" type="checkbox"/>	#	Row ID	RowID		3,036, 3,837
<input checked="" type="checkbox"/>	ABC	Order ID	Order ID		CA-2015-100363
<input checked="" type="checkbox"/>	ABC	Order Date	Order Date		04/08/2015, 12:00:00 AM
<input checked="" type="checkbox"/>	ABC	Ship Date	Ship Date		04/15/2015, 12:00:00 AM
<input checked="" type="checkbox"/>	Data Type		Ship Mode		Standard Class
<input checked="" type="checkbox"/>	# Number (decimal)		Customer ID		JM-15655
<input checked="" type="checkbox"/>	# Number (whole)		Customer Name		Jim Mitchell
<input checked="" type="checkbox"/>	✓ Date & Time		Segment		Corporate
<input checked="" type="checkbox"/>	Date		Country		United States
<input checked="" type="checkbox"/>	ABC String		City		Glendale
<input checked="" type="checkbox"/>	ABC	Postal Code	Postal Code		85301
<input checked="" type="checkbox"/>	ABC	Region	Region		West

连接到自定义 SQL 查询

如果数据库支持使用自定义 SQL，您将会看到“**自定义 SQL**”显示在“**连接**”窗格靠近底部的位置。双击“**自定义 SQL**”打开“**自定义 SQL**”选项卡，您可以在其中输入查询来预选择数据和使用特定于来源的操作。查询检索数据集之后，您可以选择要包括的字段、应用筛选器，或在将数据添加到流程之前更改数据类型。

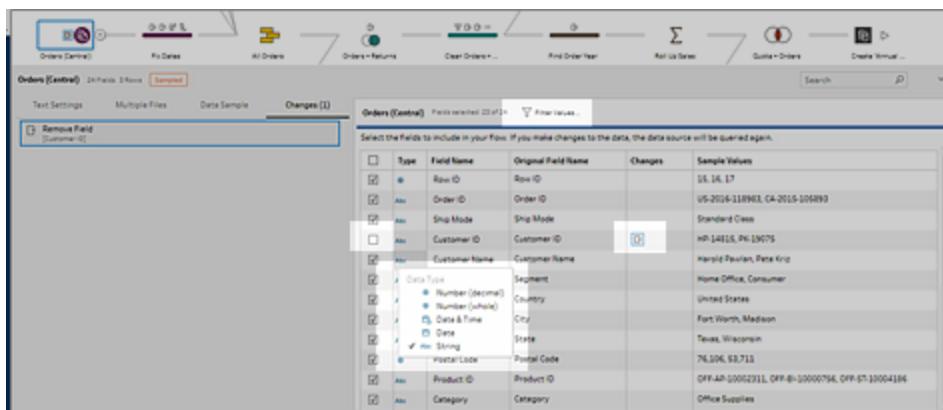
	Field Name	Original Field Name	Filters	Sample Values
<input checked="" type="checkbox"/>				

有关使用自定义 SQL 的详细信息，请参见 Tableau Desktop 和 Web 制作帮助中的[连接到自定义 SQL 查询](#)。

在输入步骤中应用清理操作

只能在“输入”步骤中执行部分清理操作。您可以在“输入”字段列表中进行以下任何更改。系统会在“**更改**”窗格中跟踪您所做的更改，并会在“**流程**”窗格中“输入”步骤的左侧以及“输入”字段列表中添加注释。

- 筛选：**在工具栏中单击“**筛选值**”，然后在计算编辑器中输入筛选条件。
- 重命名字段：**在“**字段名**”字段中，双击或按住 Ctrl 并单击 (MacOS) 字段名，并输入一个新字段名。
- 更改数据类型：**单击字段的数据类型，并从菜单中选择一个新数据类型。
- 移除字段：**清除不想包括在流程中的字段旁边的复选框。



选择要包括在流程中的字段

“**输入**”窗格显示数据集中的字段列表。您可以使用“**搜索**”字段在列表中查找字段，然后使用复选框来选择要包括或排除的字段。若要在流程中包括或排除所有字段，请切换网格顶部的复选框。

将筛选器应用于输入步骤中的字段

若要筛选字段，请执行以下操作：

1. 在工具栏中，单击“筛选值”。

The screenshot shows the 'Input' step configuration for a 'Text' connection to 'Orders_West.csv'. The 'Text Options' section includes 'First line contains header' and 'Field Separator: Comma'. The main area displays a preview of the 'Orders [West]' data with 21 rows. A 'Changes' column lists modifications made to the original field names. A 'Filter values...' button is visible at the top right of the preview area.

	Type	Field Name	Original Field Name	Changes	Sample Values
<input type="checkbox"/>	Row	Row ID	Row ID		3,826, 3,837
<input checked="" type="checkbox"/>	Attr	Order ID	Order ID		CA-2015-100363
<input checked="" type="checkbox"/>	Attr	Order Date	Order Date		04/08/2015, 12:00:00 AM
<input checked="" type="checkbox"/>	Attr	Ship Date	Ship Date		04/15/2015, 12:00:00 AM
<input checked="" type="checkbox"/>	Attr	Ship Mode	Ship Mode		Standard Class
<input checked="" type="checkbox"/>	Attr	Customer ID	Customer ID		JM15655
<input checked="" type="checkbox"/>	Attr	Customer Name	Customer Name		JIM Mitchell
<input checked="" type="checkbox"/>	Attr	Segment	Segment		Corporate
<input checked="" type="checkbox"/>	Attr	Country	Country		United States
<input checked="" type="checkbox"/>	Attr	City	City		Glendale
<input checked="" type="checkbox"/>	Attr	Postal Code	Postal Code		85301
<input checked="" type="checkbox"/>	Attr	Region	Region		West
<input checked="" type="checkbox"/>	Attr	Product ID	Product ID		000042-10000011, 000042-10004753

2. 在计算编辑器中输入筛选条件。

计算筛选器类型是“输入”步骤中提供的唯一筛选器选项。其他步骤类型中提供了其他筛选器选项。有关详细信息，请参见[筛选您的数据 在本页 202](#)。

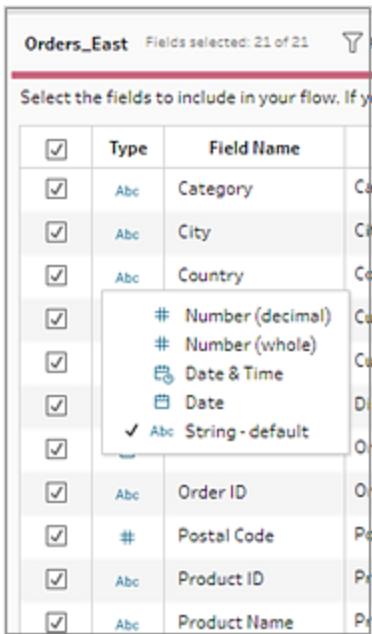
更改字段名称

若要更改字段的名称，请在“字段名称”列中选择名称，然后在字段中键入新名称。将会在字段网格以及“输入”步骤左侧的流程窗格中添加注释。系统也会在“更改”窗格中跟踪您所做的更改。

更改数据类型

若要更改字段的数据类型，请执行以下操作：

1. 单击字段的数据类型。
2. 从菜单中选择新数据类型。



您还可以更改流程中其他步骤类型中的字段的数据类型，或分配数据角色来帮助验证字段值。有关更改数据类型或使用数据角色的详细信息，请参见[查看分配给数据的数据类型](#)在本页 194 和[使用数据角色验证数据](#)在本页 209。

配置字段属性

处理文本文件时，您会看到“**文本设置**”选项卡，在该选项卡中，您可以编辑连接和配置文本属性，例如文本文件的字段分隔符。您也可以在“连接”窗格中编辑文件连接。

处理文本或 Excel 文件时，您可以纠正正在开始流程之前错误推断的数据类型。在开始流程之后，始终可以在后续步骤中通过“**配置**”窗格更改数据类型。

配置文本文件中的文本设置

若要更改用于解析文本文件的设置，请从以下选项中选择：

- 第一行包含标题(默认值)**: 选择此选项以使用第一行作为字段标签。
- 自动生成字段名称**: 如果希望 Tableau Prep Builder 自动生成字段标题，请选择此选项。字段命名约定采用与 Tableau Desktop 相同的模式。例如，F1、F2 等。
- 字段分隔符**: 从列表中选择一个字符用于分隔各个列。选择“**其他**”以输入自定义字符。
- 文本限定符**: 选择用于在文件中将值引起来的字符。

- **字符集:** 选择用于描述文本文件编码的字符集。
- **区域设置:** 选择要用于解析文件的区域设置。此设置指明要使用哪个小数分隔符和千位分隔符。

选择数据样本大小

默认情况下，为了维持最佳性能，Tableau Prep Builder 会将流程中包括的数据限制为一个数据集的代表性样本。通过根据数据集中的字段总数和这些字段的数据类型来计算最佳行数，即可确定数据样本。Tableau Prep Builder 随后即会尽快针对计算量检索最多的行数。

生成的数据样本可能包括您需要的所有行，也可能不包括，具体情况取决于样本的计算和返回方式。如果看不到预期数据，您可以更改数据样本设置以再次运行查询。

注意: 如果已对数据抽样，则“配置”窗格中会显示一个“已抽样”  指示符，并且该指示符会为您添加的每个步骤保留。所做的任何更改将应用于您在流程中处理的样本。在运行流程时，所有更改都将应用于整个数据集。

若要更改数据样本设置，请选择一个输入步骤，然后在“**数据样本**”选项卡上从以下选项中进行选择：

- **默认样本量(默认值):** Tableau Prep Builder 会计算要返回的总行数。
- **使用所有数据:** 检索数据集中的所有行，而不管大小如何。这可能会影响性能，或导致 Tableau Prep Builder 超时。

注意: 为了维持性能，即使您选择此设置，也会向“聚合”和“合并”步骤类型应用 100 万行的数据样本限制，并向“联接”和“转置”步骤类型应用 300 万行的数据样本限制。

- **固定行数:** 选择要从数据集返回的行数。推荐的行数为 100 万行或更少。将行数设置为超过 100 万行可能会影响性能。
- **快速选择(默认值):** 数据库会尽快返回请求的行数。这可能是前 N 行，或是上一次查询中数据库缓存在内存中的行数。
- **随机抽样:** 数据库返回请求的行数，但会查看数据集中的每一行并从所有行中返回一个代表性样本。在首次检索数据时，此选项可能会影响性能。

在“输入”步骤中添加更多数据

连接到数据源并开始构建流程后，您可能需要在新数据进入时刷新数据连接。您还可以在输入步骤中联接或合并数据集，以便更有效地处理更大的数据源。

在“输入”步骤中刷新数据

如果开始处理流程后输入文件或表中的数据发生变化，您可以刷新“输入”步骤为引入新数据。

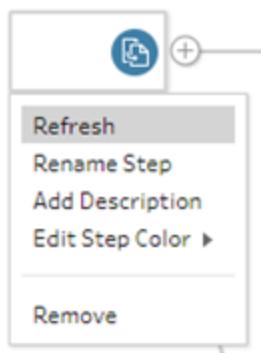
文件输入步骤类型

若要刷新文件输入步骤，请执行下列操作之一：

- 在流程窗格中的顶部菜单上，单击“刷新”按钮以刷新所有“输入”步骤。若要刷新单个“输入”步骤，请单击刷新按钮旁边的下拉箭头，并从列表中选择该“输入”步骤。



- 在流程窗格中，右键单击要刷新的“输入”步骤，并从菜单中选择“刷新”。



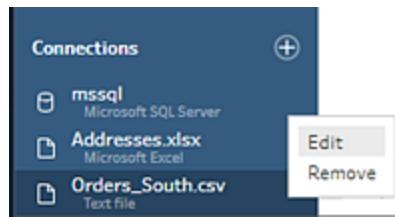
文件、数据库或 Tableau 数据提取输入步骤类型

若要刷新数据库或 Tableau 数据提取输入步骤，请执行下列操作之一：

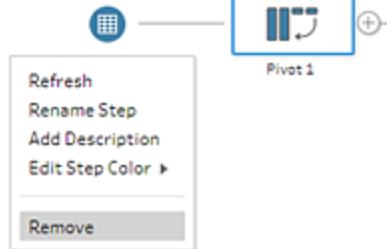
- 尝试编辑连接。

注意:为了保持性能, Tableau Prep Builder 会对大型数据集取样。如果数据已取样, 则在配置窗格中, 您可能会看到也可能不会看到新数据。您可以在“输入”步骤中的“数据样本”选项卡中更改数据的取样方式, 但可能会影响性能。有关设置数据样本大小的详细信息, 请参见[选择数据样本大小 在本页 167](#)。

- 在“连接”窗格中, 右键单击或按住 Ctrl 单击 (MacOS) 数据源, 并选择“编辑”。



- 通过登录数据库或者重新选择文件或 Tableau 数据提取来重建连接。
 - 移除“输入”步骤并将其重新添加到流程。
- 在流程窗格中, 右键单击要刷新的“输入”步骤, 并从菜单中选择“移除”。



这会使您的流程暂时处于错误状态。



- 再次连接到更新的文件。
- 将表拖到流程窗格中要向其中添加“输入”步骤的流程中的第二个步骤上。将其

放在“添加”选项上，以将其重新连接到流程。



在输入步骤中合并文件和数据库表

使用来自单一数据源的多个文件或数据库表时，您可以使用通配符搜索来搜索文件或表，然后合并数据以将所有文件或表数据包括在“输入”步骤中。若要合并文件，文件必须位于同一父目录或子目录中。

在下次您打开流程或通过命令行运行流程时，添加到同一文件夹中、与模式匹配的新文件会自动包括在并集中。打包流程文件 (.tflx) 不会自动获取新文件，因为文件已与流程打包在一起。若要为打包流程包括新文件，请打开流程文件 (.tfl) 以获取新文件，然后重新打包流程以包括新文件数据。

若要合并数据库表，表必须位于同一数据库中，并且数据库连接必须支持通配符并集搜索。以下数据库支持这种类型的并集：

- Amazon Redshift
- Microsoft SQL Server
- MySQL
- Oracle
- PostgreSQL

如果在创建并集之后添加或移除文件或表，您可以刷新输入步骤，用新数据或更改的数据来更新流程。

注意：此功能当前仅适用于 Excel 和 .csv(文本) 文件，以及存储在上面所列特定数据库中的数据表。此选项对于 Tableau 数据提取不可用。

Tableau Prep Builder 版本 2018.1.2 及更高版本中提供了适用于文件的通配符并集。

Tableau Prep Builder 版本 2018.3.1 及更高版本中提供了适用于数据库表的通配符并集。在以前的版本中编辑具有此并集类型的流程连接可能会导致错误。

如果需要合并来自不同数据源的数据，您可以使用合并步骤执行该操作。有关创建合并步骤的详细信息，请参见[合并数据 在本页 303](#)。

合并文件

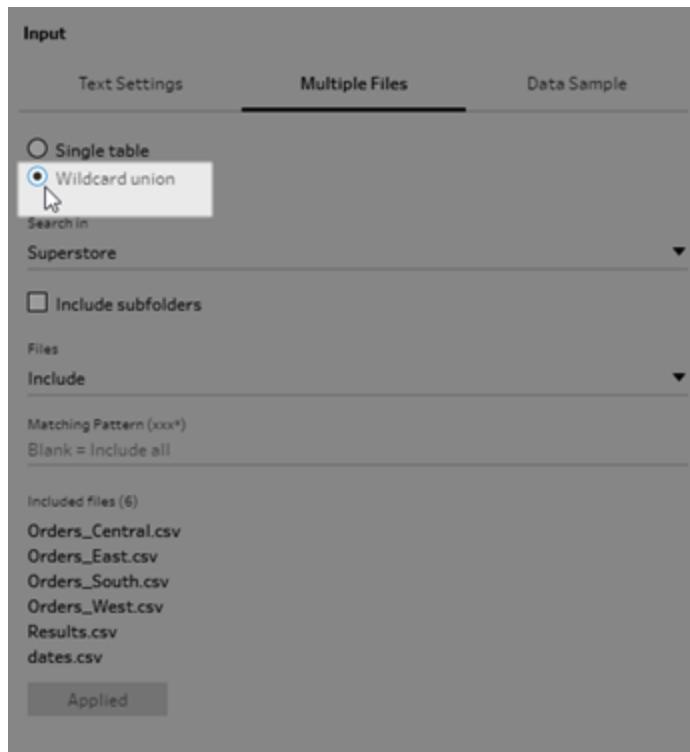
默认情况下，Tableau Prep Builder 会合并您连接到的 .csv 文件所在同一目录中的所有 .csv 文件，或者合并您连接到的 Excel 文件中的所有工作表。如果使用 Data Interpreter 清理 Excel 文件，并且使用的是 Tableau Prep Builder 版本 2018.1.2 或更高版本，您可以使用通配符搜索来合并和添加 Data Interpreter 找到的任何子表。

如果要更改默认并集，请使用以下条件来查找要包括在并集中的文件或工作表：

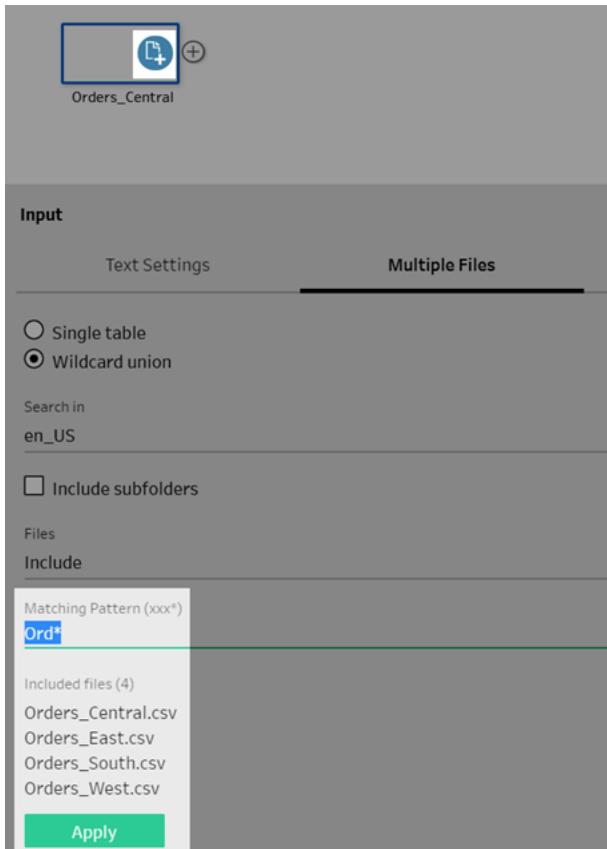
- **搜索范围**: 选择要用于搜索文件的目录。选中“**包括子文件夹**”复选框以包括父文件夹的子目录中的文件。
- **文件**: 选择是否包括或排除与通配符搜索条件匹配的文件。
- **匹配模式(XXX*)**: 输入通配符搜索模式以查找文件名中包含这些字符的文件。举例来说，如果输入 ord*，则会返回包括该文件名的所有文件。将此字段留空以包括指定目录中的所有文件。

若要在输入步骤中合并文件，请执行以下操作：

1. 单击“**添加连接**” 按钮，并在“**连接**”下单击“**文本文件**”(适用于 .csv 文件)或“**Microsoft Excel**”(适用于 Excel 文件)，然后选择要打开的文件。
2. 在“**输入**”窗格中，选择“**多个文件**”选项卡，然后选择“**通配符并集**”。

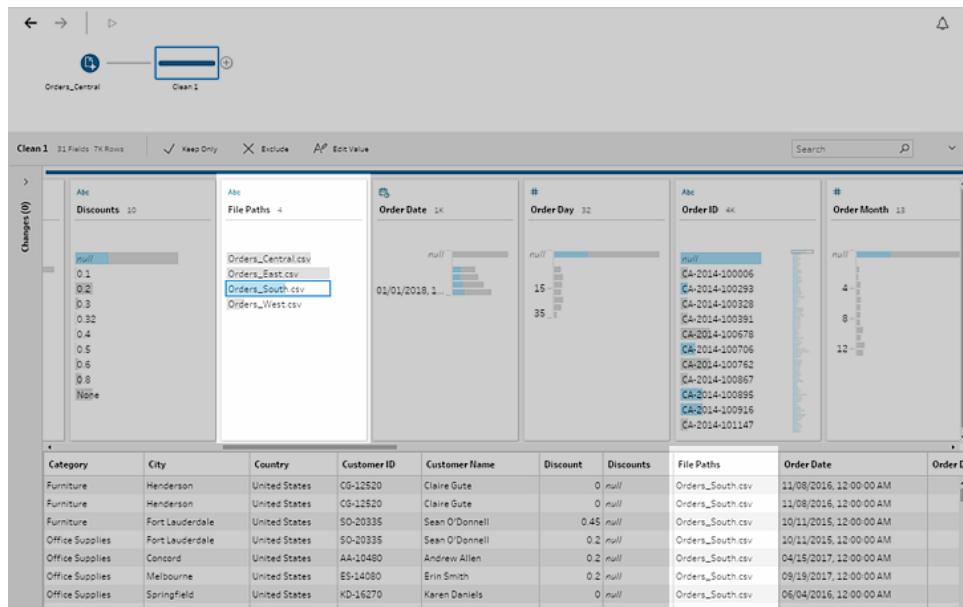


下面的示例显示一个使用匹配模式的通配符并集。“流程”窗格中“Orders_Central”输入步骤上文件图标上的加号表明该步骤包括一个通配符并集。并集中的文件列在“已包括文件”下。



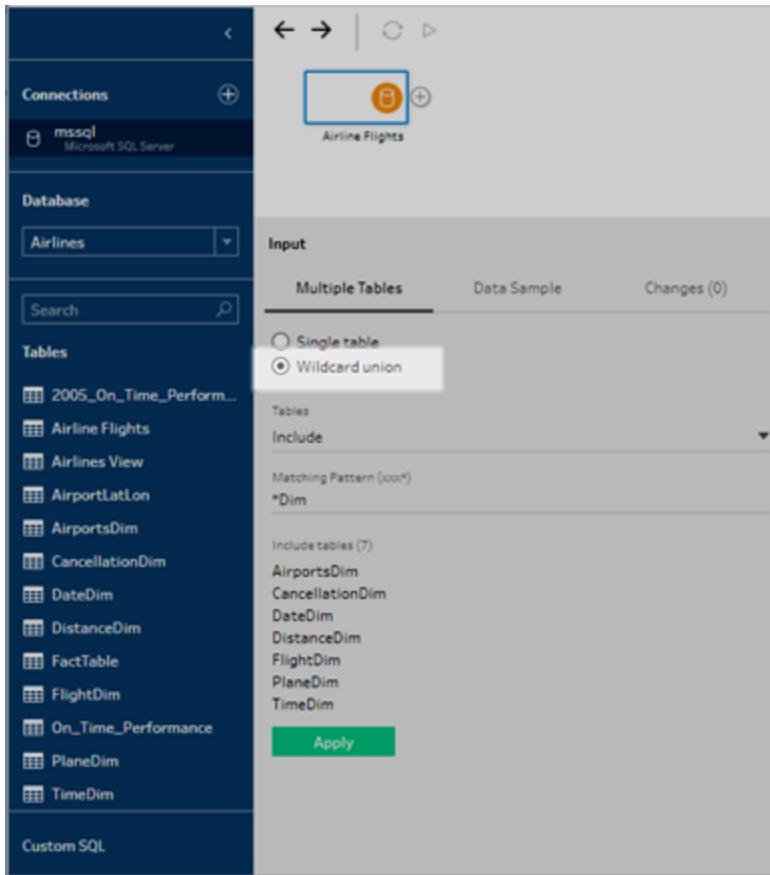
3. 使用搜索、文件和匹配模式选项来查找要合并的文件。
4. 单击“应用”以合并文件。

向流程中添加新步骤时，您可以在“配置”窗格的“文件路径”字段中看到添加到数据集的所有文件。此字段是自动添加的。



合并数据库表(版本 2018.3.1 及更高版本)

1. 单击“添加连接”按钮，并在“连接”下连接到支持通配符并集的数据库。
2. 将表拖到流程窗格。
3. 在“输入”窗格中，选择“多个表”选项卡，然后选择“通配符并集”。

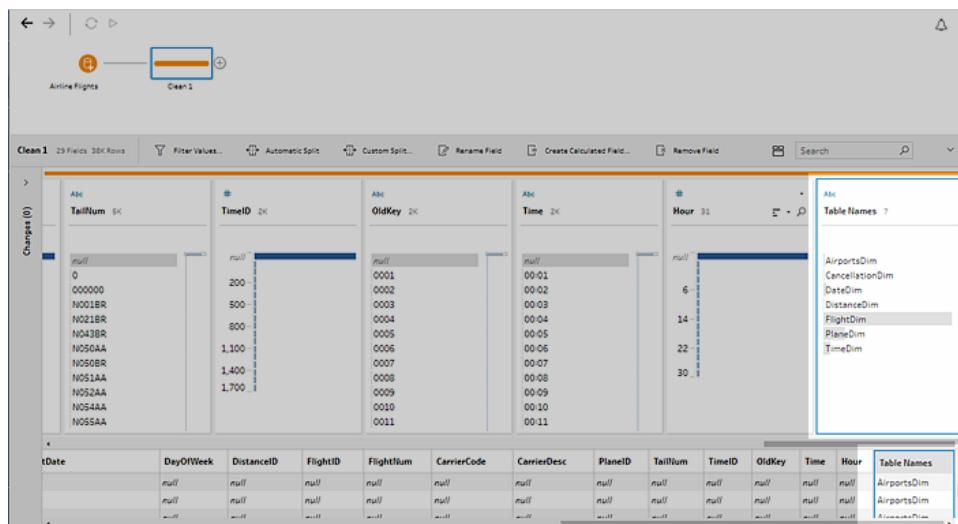


4. 使用搜索、“表”和“匹配模式”选项来查找要合并的表。

只有显示在“连接”窗格的“表”部分中的表才能包括在并集中。通配符搜索不会跨架构或跨数据库连接进行搜索来查找表。

5. 单击“应用”以合并表数据。

向流程中添加新步骤时，您可以在“配置”窗格的“表名称”字段中看到添加到数据集的所有表。此字段是自动添加的。



在并集之后合并字段

在输入步骤中创建并集后，您可能需要合并字段。您可以在除“输入”或“输出”步骤之外的任何后续步骤中执行此操作。有关详细信息，请参见[其他合并字段选项 在本页 308](#)。

在“输入”步骤中联接数据(版本 2019.1.3 及更高版本)

当连接到的数据库包含具有关系数据的表时，Tableau Prep Builder 可以检测和显示表中的哪些字段被标识为唯一标识符，哪些字段被标识为相关字段，并为这些字段显示相关的表名称。

一个名为“**链接键**”的新列将显示在“输入”窗格中，并显示以下关系(如果存在)：

- **唯一标识符**。此字段唯一标识表中的每一行。表中可以有多个唯一标识符。字段中的值必须唯一，不能为空或 null。
- **相关字段**。此字段将表与数据库中的另一个表关联。表中可以有多个相关字段。
- **唯一标识符和相关字段**。字段是此表中的唯一标识符，并将表与数据库中的另一个表关联。

Tableau Prep 帮助

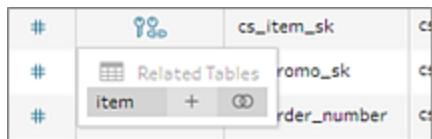
catalog_sales Fields selected: 34

Select the fields to include in your workflow.

Type	Linked Keys	new name	original field name	changes	sample
•	↳	cs_sold_date_sk	cs_sold_date_sk	2,450,816, 2,41	
•	↳	cs_sold_time_sk	cs_sold_time_sk	33,151, 29,066	
•	↳	cs_ship_date_sk	cs_ship_date_sk	2,450,886, 2,41	
•	↳	cs_bill_customer_sk	cs_bill_customer_sk	96,466, 49,302	

您可以利用这些关系快速查找并向流程中添加相关表，或者通过“输入”步骤创建联接。此功能可用于定义了表关系的任何受支持的数据库连接器。

1. 连接到包含字段关系数据(例如唯一标识符或相关字段)的数据库(例如 Microsoft SQL Server)。
2. 在“输入”窗格中，单击标记为 相关字段或 同时标记为唯一标识符和相关字段的字段。
此时将打开一个对话框，其中显示相关表的列表。
3. 将鼠标指针悬停在要添加或联接的表上，然后单击加号按钮将该表添加到流程中，或者单击联接按钮创建与所选表的联接。



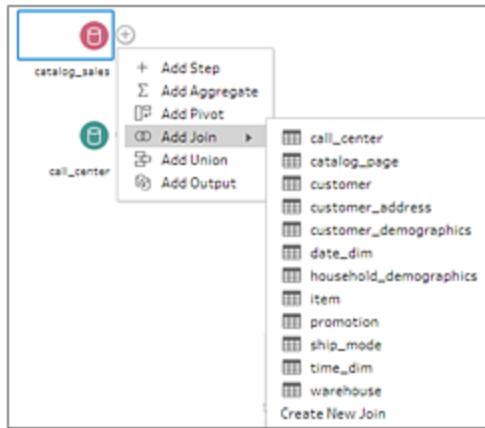
如果创建联接，Tableau Prep Builder 将使用定义的字段关系来联接表，并显示将用于创建联接的联接子句的预览。



4. 或者，您可以通过“流程”窗格的菜单来联接相关表。将鼠标指针悬停在一个步骤上，直到出现加号 图标，然后选择“添加联接”查看相关表的列表。Tableau Prep Builder

基于组成两个表之间的关系的字段创建联接。

注意: 如果表没有定义表关系，则此选项不可用。



有关使用联接的详细信息，请参见[联接数据 在本页298](#)。

构建和组织流程

连接到要包括在流程中的数据后，您可以首先通过向流程中添加新步骤或在现有步骤之间插入步骤来清理和调整数据。

若要组织流程，您可以更改默认步骤颜色、添加说明以提供步骤或清理操作的上下文，或者重新组织流程布局以便能更轻松地跟进复杂流程。

添加或插入步骤

构建流程时，您可以添加不同的步骤类型来执行所需的操作。例如，使用“**添加步骤**”选项添加用于拆分字段的清理步骤、应用筛选器或执行各种其他操作来清理不良数据。使用“**连接**”或“**合并**”步骤类型来合并数据表，或者添加“**脚本**”步骤类型，以将 R 或 Python 脚本合并到流程中。

当流程开始成形时，您可能需要回到流程中较早的步骤，并插入不同的步骤类型来执行各种操作，比如添加其他清理步骤，或者聚合数据以使用与后续步骤相同的详细级别。

您看到的菜单选项将因 Tableau Prep Builder 版本以及是添加步骤在流程中构建新步骤还是在现有步骤中插入步骤而异。如果使用的是 Tableau Prep Builder 版本 2019.3.1 或更低版本，请参考该部分来查看菜单选项。

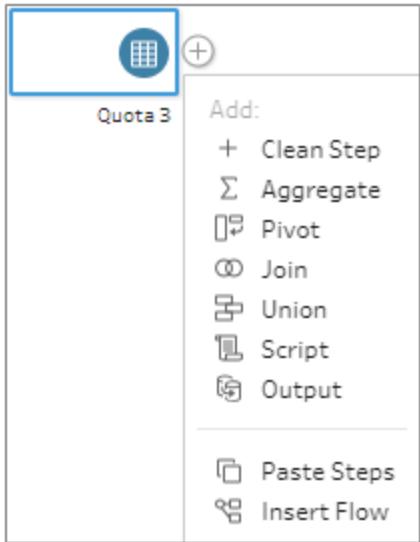
您无法使用这些菜单添加输入步骤。作为替代，您需要将表从“**连接**”窗格拖到“**流程**”窗格。有关详细信息，请参见[连接到数据 在本页 141](#)。

注意：可包括在流程中的某些功能或动作可能与您登录的 Tableau Server 版本不兼容。不能将包含不兼容功能的流程安排为使用 Tableau Prep Conductor 运行。有关管理版本不兼容性的详细信息，请参见[与 Tableau Prep 的版本兼容性 在本页 333](#)。

添加步骤

将光标悬停在步骤上，直至出现加号  图标，然后单击该图标并选择一种步骤类型。当您添加步骤以在流程中构建下一个步骤时，此菜单将出现。

注意：您的选项可能看起来不同，具体取决于您的产品版本。例如，Tableau Prep Builder 版本 2019.4.2 的此菜单中增加了“**插入流程**”。



从以下选项中进行选择：

- **添加分支:** 将流程拆分为不同的分支。从两个现有步骤之间打开“添加”菜单时，将显示此选项。

注意: 在 Tableau Prep Builder 2019.4.2 中，在两个现有步骤之间打开“添加”菜单时，此选项已替换为“添加清理步骤”。

- “**添加步骤**”或“**添加清理步骤**”: 添加清理步骤以执行各种清理操作。有关可使用的各种清理操作的详细信息，请参见[清理和调整数据 在本页 223](#)。
- **添加聚合:** 创建一个“**聚合**”步骤，您可以在其中选择要聚合或分组的字段。有关详细信息，请参见[对值进行聚合和分组 在本页 297](#)。
- **添加转置:** 创建一个“**转置**”步骤，您可以在其中执行各种转置选项，将列数据转换为行，或将数据行转换为列。有关详细信息，请参见[转置您的数据 在本页 273](#)。
- **添加联接:** 创建一个“**联接**”步骤，您可以在其中手动将其他输入添加到联接，并添加联接子句。作为替代方案，您可以通过拖放步骤来联接文件。以下示例中显示了将“Orders_Central”输入步骤拖放到“**联接**”上的操作：



有关创建联接的详细信息，请参见[联接数据 在本页 298](#)。

在 Tableau Prep Builder 版本 2019.1.3 及更高版本中，如果连接到的数据库包含具有关系数据的表，您也可以通过“流程”窗格中的菜单创建联接。有关使用此方法联接表的详细信息，请参见在[“输入”步骤中联接数据\(版本 2019.1.3 及更高版本\) 在本页 176](#)。

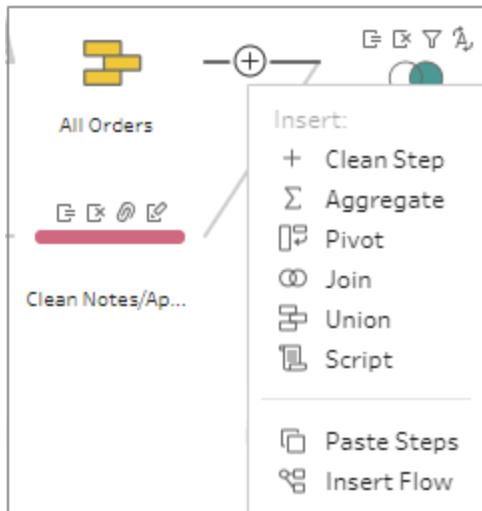
- **添加并集:** 创建一个“并集”步骤。通过将表拖动到步骤并将其放在显示的“添加”选项上，将表添加到并集中。作为替代方案，您可以将步骤拖放到另一个步骤上来合并文件。有关创建并集的详细信息，请参见[合并数据 在本页 303](#)。
- **添加脚本(版本 2019.3.1 及更高版本):** 创建一个脚本步骤，以便在流程中包含 R 和 Python 脚本。有关详细信息，请参见[在流程中使用 R 和 Python 脚本 在本页 281](#)。
- **添加输出:** 创建一个“输出”步骤，以将输出保存到数据提取文件(.tde 或 .hyper)、.csv 文件，或者将输出作为数据源发布到服务器。
- **粘贴步骤:** 添加从同一流程中复制的步骤。有关在同一流程中复制和粘贴步骤的详细信息，请参见[清理和调整数据 在本页 223](#)。
- **插入流程(版本 2019.3.2 及更高版本):** 将从其他流程保存的流程步骤添加到当前流程中。您可以将它们添加到步骤的末尾，或将其插入到现有步骤之间。在 Tableau Prep Builder 2019.4.2 中，此选项已添加到此菜单中。在以前的版本中，您可以通过流程窗格的空白区域中右键单击或按住 Ctrl (MacOS) 单击来添加流程步骤。有关在流程步骤中使用已保存流程步骤的详细信息，请参见[创建可重用的流程步骤\(版本 2019.3.2 及更高版本\) 在本页 258](#)。

插入步骤

在现有步骤之间插入一个步骤。“输入”和“输出”步骤类型在此菜单中不可用。选项因产品版本而异。如果使用 Tableau Prep Builder 的早期版本，请参考下面的“**版本 2019.3.1 及更低版本**”部分。

1. 将光标悬停在要在其中插入流程的流程线的中间，直至出现加号 \oplus 图标。然后单击该图标并选择一种步骤类型。

注意：您的选项可能看起来不同，具体取决于您的产品版本。例如，Tableau Prep Builder 版本 2019.4.2 的此菜单中增加了“插入流程”。



2. 从以下选项中进行选择：

- **插入步骤或插入清理步骤：**在现有步骤之间插入清理步骤以执行各种清理操作。有关可使用的各种清理操作的详细信息，请参见[清理和调整数据 在本页 223](#)。
- **插入聚合：**在现有步骤之间插入一个“聚合”步骤。您可以选择要聚合或分组的字段。有关详细信息，请参见[对值进行聚合和分组 在本页 297](#)。
- **插入转置：**在现有步骤之间插入一个“转置”步骤。可以执行各种转置选项，将列数据转换为行，或将行数据转换为列。有关详细信息，请参见[转置您的数据 在本页 273](#)。
- **插入联接：**在现有步骤之间插入一个“联接”步骤。您可以手动将其他输入添加到联接，并添加联接子句。作为替代方案，您可以通过拖放步骤来联接文件。以下示例中显示了将“Orders_Central”输入步骤拖放到“联接”上的操作：



有关创建联接的详细信息，请参见[联接数据 在本页 298](#)。

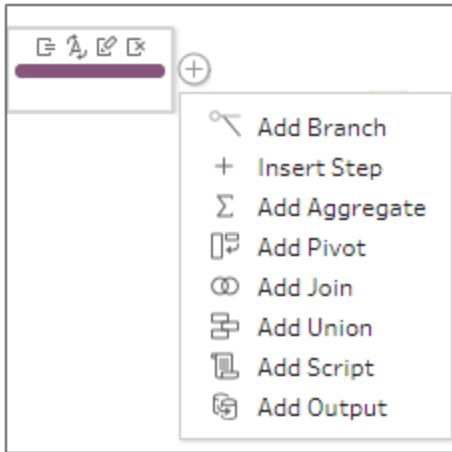
在 Tableau Prep Builder 版本 2019.1.3 及更高版本中，如果连接到的数据库包含具有关系数据的表，您也可以通过“流程”窗格中的菜单创建联接。有关使用此方法联接表的详细信息，请参见[在“输入”步骤中联接数据\(版本 2019.1.3 及更高版本\) 在本页 176](#)。

- **插入并集:** 插入一个“并集”步骤。通过将表拖动到步骤并将其放在显示的“添加”选项上，将表添加到并集中。作为替代方案，您可以将步骤拖放到另一个步骤上来合并文件。有关创建并集的详细信息，请参见[合并数据 在本页 303](#)。
- **粘贴步骤:** 在现有步骤之间插入从同一流程中复制的步骤。有关在同一流程中复制和粘贴步骤的详细信息，请参见[清理和调整数据 在本页 223](#)。
- **插入流程(版本 2019.3.2 及更高版本):** 将从其他流程保存的流程步骤插入到当前流程中。您可以将它们添加到步骤的末尾，或将其插入到现有步骤之间。在 Tableau Prep Builder 2019.4.2 中，此选项已添加到此菜单中。在以前的版本中，您可以通过流程窗格的空白区域中右键单击或按住 Ctrl (MacOS) 单击来添加流程步骤。有关在流程步骤中使用已保存流程步骤的详细信息，请参见[创建可重用的流程步骤\(版本 2019.3.2 及更高版本\) 在本页 258](#)。

版本 2019.3.1 及更低版本

1. 将光标悬停在步骤上，直至出现加号 图标，然后单击该图标并选择一种步骤类型。

插入步骤: 在步骤之间插入一个清理步骤。所有其他选项将从流程中创建一个分支。



2. 从以下选项中进行选择：

- **添加分支:** 将流程拆分为不同的分支。
- **插入步骤:** 在现有步骤之间插入清理步骤以执行各种清理操作。有关可使用的各种清理操作的详细信息，请参见[清理和调整数据 在本页 223](#)。
- **添加聚合:** 创建一个“聚合”步骤，您可以在其中选择要聚合或分组的字段。有关详细信息，请参见[对值进行聚合和分组 在本页 297](#)。
- **添加转置:** 创建一个“转置”步骤，您可以在其中执行各种转置选项，将列数据转换为行，或将数据行转换为列。有关详细信息，请参见[转置您的数据 在本页 273](#)。
- **添加联接:** 创建一个“联接”步骤，您可以在其中手动将其他输入添加到联接，并添加联接子句。作为替代方案，您可以通过拖放步骤来联接文件。以下示例中显示了将“Orders_Central”输入步骤拖放到“联接”上的操作：



有关创建联接的详细信息，请参见[联接数据 在本页 298](#)。

在 Tableau Prep Builder 版本 2019.1.3 及更高版本中，如果连接到的数据库包含具有关系数据的表，您也可以通过“流程”窗格中的菜单创建联接。有关使用

此方法联接表的详细信息，请参见在“输入”步骤中联接数据(版本 2019.1.3 及更高版本) 在本页 176。

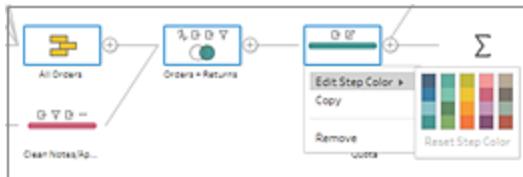
- **添加并集:** 创建一个“并集”步骤。通过将表拖动到步骤并将其放在显示的“添加”选项上，将表添加到并集中。作为替代方案，您可以将步骤拖放到另一个步骤上来合并文件。有关创建并集的详细信息，请参见**合并数据 在本页 303**。
- **添加脚本(版本 2019.3.1 及更高版本):** 创建一个“脚本”步骤以在流程中包含 R 和 Python 脚本。有关详细信息，请参见**在流程中使用 R 和 Python 脚本 在本页 281**。
- **添加输出:** 选择此选项以将输出保存到数据提取文件(.tde 或 .hyper)、.csv 文件，或者将输出作为数据源发布到服务器。

更改流程配色方案

默认情况下，Tableau Prep Builder 会为流程中的每个步骤分配一种颜色。此配色方案将在整个流程中应用，帮助您在应用清理步骤、联接、合并或聚合数据时在整个流程中跟踪数据，让您知道哪些文件受到操作的影响。

若要为步骤选择其他配色方案，请执行以下操作：

1. 选择一个或多个步骤。
2. 右键单击所选的步骤，并选择“编辑步骤颜色”。



3. 单击调色板中的一种颜色来应用该颜色。

若要将步骤颜色重置回默认颜色，请执行以下操作之一：

- 从顶部菜单中单击“撤消”。
- 按 Cntrl+Z 或 Command-Shift-Z (Mac)。
- 选择更改的步骤，右键单击所选的步骤，并选择“编辑步骤颜色”，然后从调色板的底部选择“重置步骤颜色”。

为流程步骤和清理动作添加说明

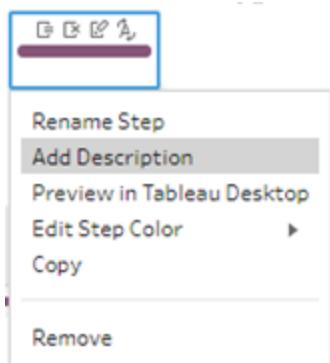
在构建流程并执行各种清理操作时，您可能需要添加说明，帮助稍后可能会查看或使用您的流程的其他人更好地理解您的步骤。您可以直接在“流程”窗格中为流程中的任何单独步骤添加说明，或为“更改”窗格中的任何清理动作添加说明来为更改提供其他上下文。说明最多可为 200 个字符长。

有关在“更改”窗格中查看更改的详细信息，请参见[查看您所做的更改 在本页 232](#)。

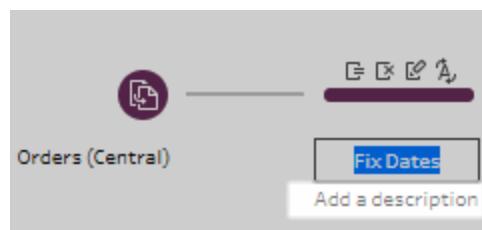
为流程步骤添加说明

当您添加说明时，会在步骤下方添加一个消息  图标。单击图标以在“流程”窗格中显示或隐藏说明文本。

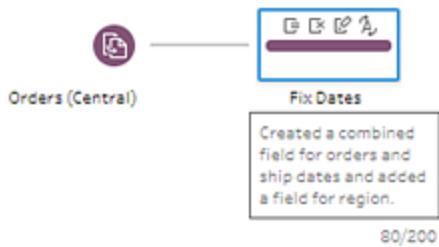
1. 在“流程”窗格中，选择一个步骤。
2. 执行以下操作之一：
 - 右键单击该步骤，并从菜单中选择“添加说明”。



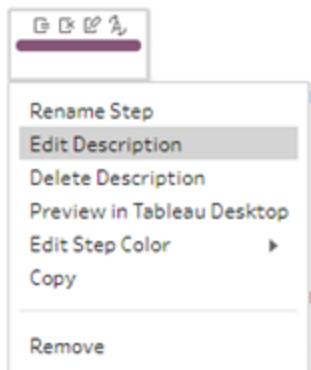
- 在步骤的名称字段中双击，然后单击“添加说明”。



3. 在文本框中键入说明。



4. 在文本框外部单击，或按 Enter 应用您所做的更改。默认情况下，说明显示在步骤的下方。若要隐藏说明，请单击消息 图标。
5. 若要编辑或删除说明，请右键单击步骤或说明，然后从菜单中选择“编辑说明”或“删除说明”。



为更改条目添加说明(版本 2019.1.1 及更高版本)

1. 在流程窗格中选择一个步骤。
2. 打开“更改”窗格或“更改”选项卡。
3. 在“更改”窗格中右键单击或按住 Ctrl 并单击 (MacOS) 一个条目，并选择“添加说明”。

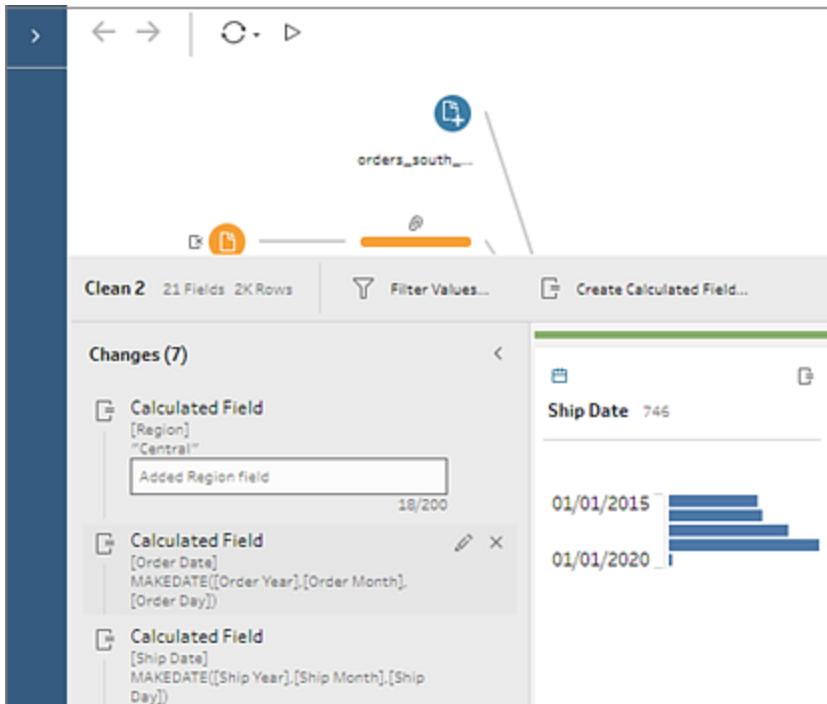
The screenshot shows the Tableau Prep interface. At the top, there are navigation icons: a left arrow, a right arrow, a circular arrow, and a double arrow. Below them is a file icon and a trash bin icon. A blue status bar indicates "Clean 2 21 Fields 2K Rows". To the right are buttons for "Filter Values..." and "Create Calculated Field...".

The main area displays a list of changes under "Changes (7)". One item in the list is a "Calculated Field" named "[Region] 'Central'". A context menu is open over this item, showing options: "Add Description" (which is highlighted), "Edit...", "Copy", and "Remove".

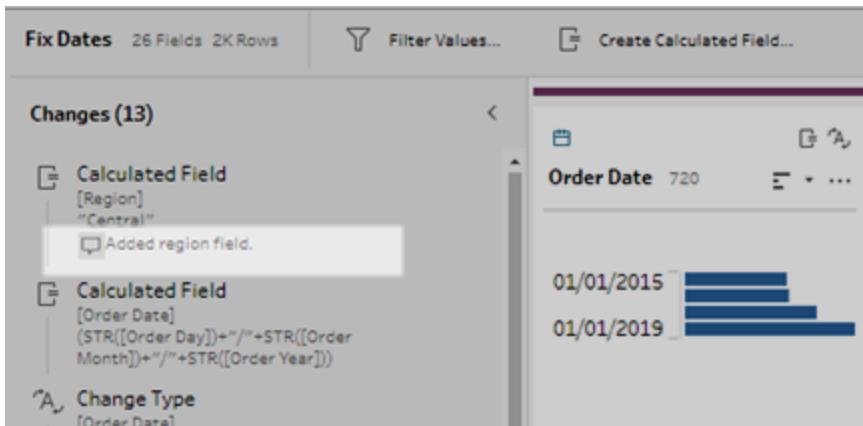
On the right side of the screen, there is a preview pane titled "Ship Date 746". It shows a histogram with two bars: one for "01/01/2015" and another for "01/01/2020". Below the histogram is a table with three columns: "Ship Date", "Order Date", and "Region". The data in the table is as follows:

Ship Date	Order Date	Region
11/26/2016	11/22/2016	Cent
11/26/2016	11/22/2016	Cent
11/10/2015	11/11/2015	Cent

4. 为更改动作输入说明。



说明将出现在为带有注释 图标的更改生成的文本的下方。



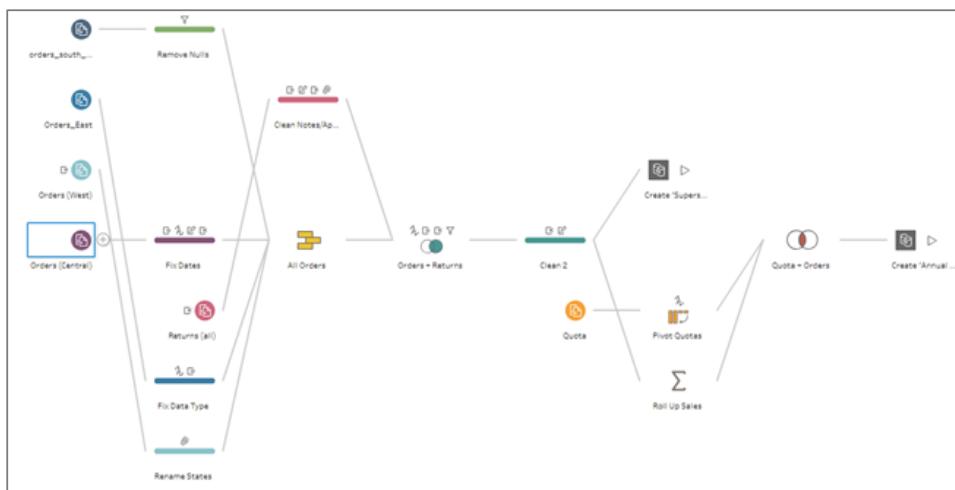
- 若要编辑或删除说明，请右键单击或按住 Ctrl 并单击 (MacOS) 更改项，然后选择“编辑说明”或“删除说明”。

重新组织流程的布局(版本 2019.2.2 及更高版本)

当您构建流程时, Tableau Prep Builder 使用默认布局。每个流程按从左到右的顺序布局和处理, “输入”步骤位于画布最左侧的开始位置, “输出”步骤位于画布右侧的结束位置。但是, 如果构建复杂的大型流程, 则流程会很快变得难以执行。

若要清理流程的布局, 您可以选择并移动步骤, 使流程布局按对您有意义的方式进行组织。例如, 您可以修复交叉的流程线条, 移动流程步骤以清理额外的空白, 或者重新排列流程步骤以显示清晰的事件序列。

例如, 下面的流程混乱不清, 难以执行:



若要清理此流程, 您可以选择并将步骤向上、向下、向左或向右拖放到画布中的新位置。无法将流程步骤移到会中断从左到右处理流程的位置。例如, 您不能将位于联接步骤之前的合并步骤拖到流程中联接步骤之后的位置。

将流程步骤拖到允许的位置时, 将显示一个橙色的框。如果不允许该位置, 则不会显示橙色框, 当您尝试放置步骤时, 它们将返回到原始位置。

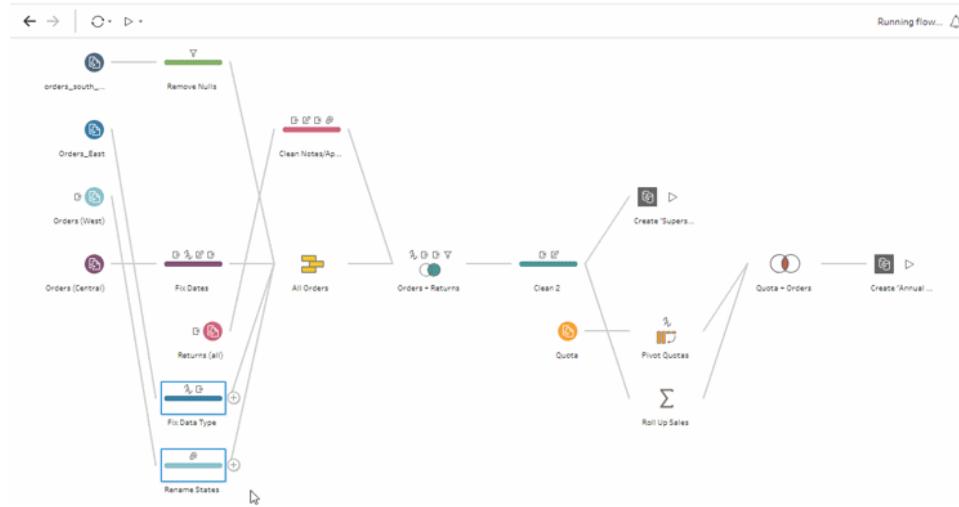
若要在流程中移除步骤, 请执行以下操作:

1. 在“流程”窗格中, 选择要移除的步骤。您可以单击特定步骤, 拖动以选择多个步骤, 或按住 Ctrl 单击或按住 Cmd 单击 (MacOS) 以选择不相邻的步骤。
2. 将步骤拖放到新位置。

注意: 如果您不喜欢所进行的重新组织移动操作, 您可以在顶部菜单中单击“**撤消**”来撤消这些操作。但是, 如果在移动步骤之间执行清理操作, 您也可以撤消这些操作。“**撤消**”选项会按您执行它们时的步骤撤消您的操作。

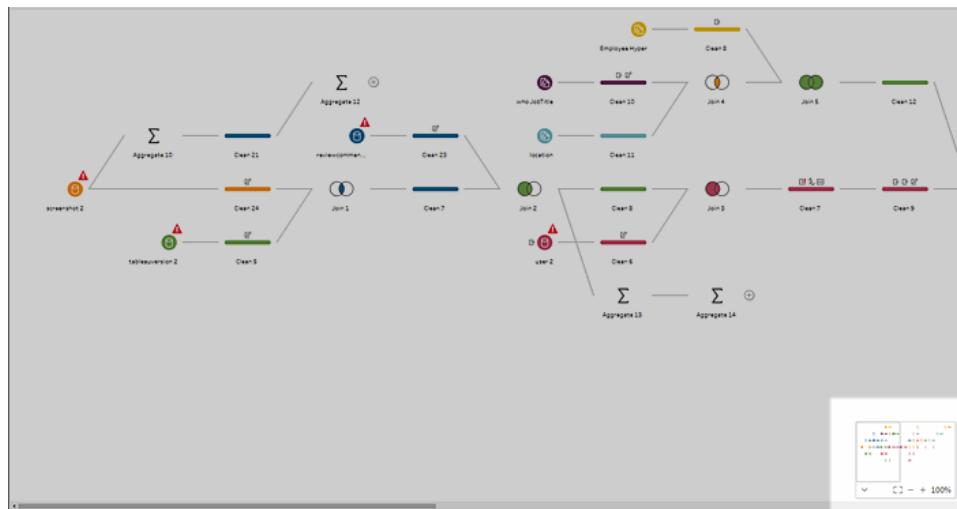
观看“重新组织流程步骤”的实际运行效果

下面的示例演示使用拖放来重新排列流程的操作。



使用流程导航工具

在处理大型流程时, 来回滚动以搜索要返回到的特定流程区域可能很困难。利用流程导航工具可以更轻松地完成此操作。流程导航器是流程的微缩版本, 显示在画布的右下角。



单击图形的任何区域可跳转到流程的该区域，或使用以下工具栏选项进行导航：



工具栏选项	说明
	折叠流程导航器图形。在折叠状态下，您可能只看到百分比指示器。只需将光标悬停在此项上来展开工具栏，然后单击向上箭头即可再次展开图形。
	展开流程导航器图形。
	更改流程的大小以适合屏幕。
- + 100%	放大和缩小您的流程。您可以单击百分比指示器将视图还原到 100% 大小。

从流程中移除步骤

在流程中的任何位置，您可以移除步骤或步骤之间的流程线。

- 若要移除步骤或流程线，请选择要移除的步骤或流程线，右键单击该元素，然后选择“**移除**”。
- 若要移除多个步骤或流程线，请执行以下操作之一：
 - 使用鼠标拖动并选择一整段流程。然后右键单击某个选定的步骤，并选择“**移除**”
 - 按 **Ctrl+A** 或 **Cmd+A (MacOS)** 选择流程中的所有元素，或按住 **Ctrl** 并单击或者按住 **Cmd** 并单击 (**MacOS**) 以选择特定元素，然后按 **Delete** 键。

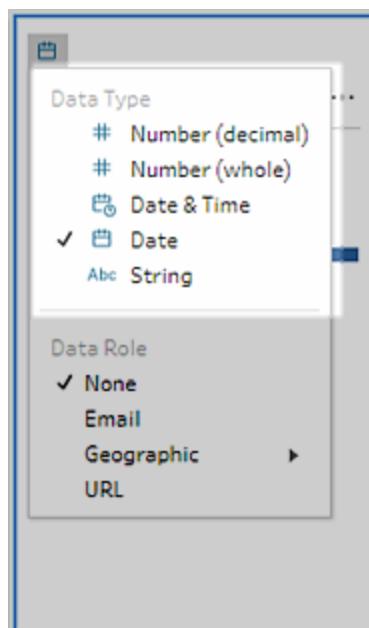
检查您的数据

使用本主题中的选项来透彻理解数据的组成，从而更好地了解所需进行的更改，以及流程中包括的操作的效果。

查看分配给数据的数据类型

像 Tableau Desktop 一样，Tableau Prep Builder 会在您将连接拖到“流程”窗格中时对字段中的数据进行解读，并自动为其分配一个数据类型。由于不同的数据库可能会以不同的方式处理数据，因此 Tableau Prep Builder 的解读可能不会始终正确。

若要更改数据类型，请单击数据类型图标，并从上下文菜单中选择正确的数据类型。您可以将“字符串”或“整数”数据类型更改为“日期”或者“日期和时间”，Tableau Prep Builder 将触发“Auto DateParse”来更改这些数据类型。像 Tableau Desktop 一样，如果更改未成功，您将会在字段中看到 Null 值，并且您可以创建计算来进行更改。

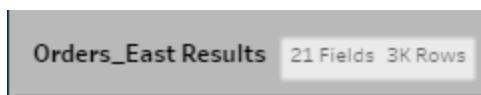


有关使用 DateParse 的详细信息，请参见 Tableau Desktop 和 Web 制作帮助中的[将字段转换为日期字段](#)。

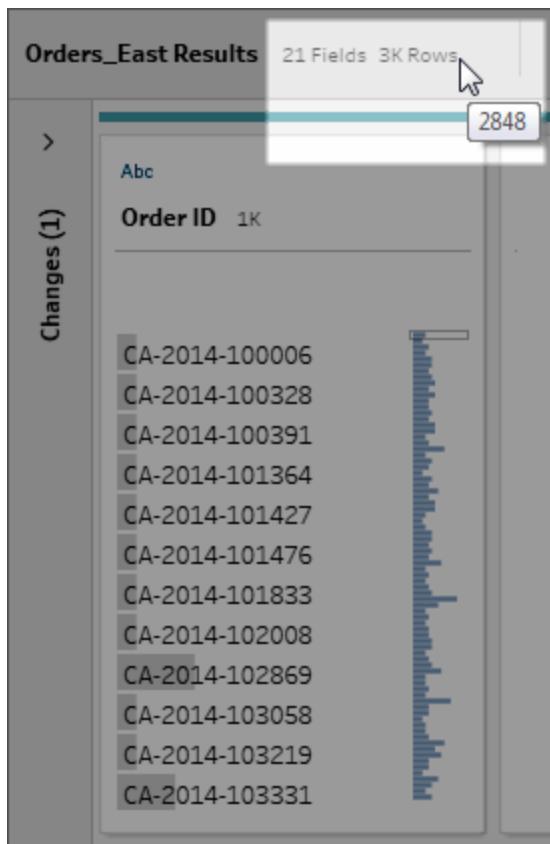
查看有关数据的大小详细信息

连接到数据后，向流程中添加一个表，然后添加一个步骤。您可以使用“配置”窗格来查看数据的当前状态和结构，并找出 null 值和离群值。

- **字段数和行数：**在“配置”窗格的左上角，您可以找到在流程的某个特定点数据中的字段数和行数的汇总。Tableau Prep Builder 会舍入到最接近的千位。在下面的示例中，数据集中有 21 个字段和 3000 行。



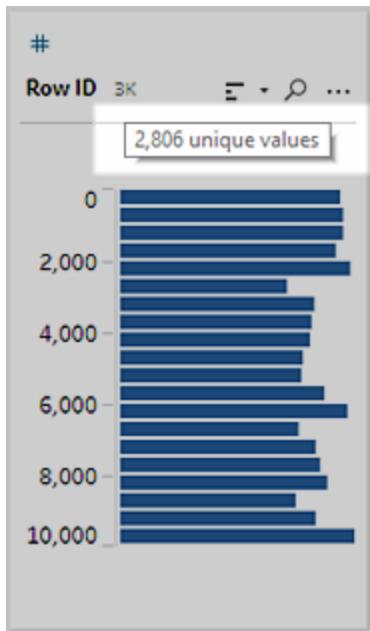
将光标悬停在字段数和行数上时，您可以看到确切的行数(本例中为 2848)：



- **数据集大小：**通过在“输入”窗格的“数据样本”选项卡中指定要包括的行数来使用数据子集。
- **已抽样：**为了让您能够与数据直接交互，Tableau Prep Builder 会使用原始数据的子

集。行数由数据类型和所呈现的字段数确定。与整数相比，字符串字段会占用更多的存储空间，因此，如果数据集中有 10 个字符串字段，与有 10 个整数字段的情况相比，您得到的行数可能会更少。“已抽样” 指示符显示在“配置”窗格中大小详细信息的旁边，用于指明这是数据集的子集。有关数据样本的详细信息，请参见[配置数据集 在本页 162](#)。

- **唯一值的数量：**每个字段标题旁边的数字表示该字段内包含的不同值。Tableau Prep Builder 会舍入到最接近的千位。在下面的示例中，“说明”字段中呈现的不同值为 3,000 个，但如果将光标悬停在数字上，您可以看到唯一值的确切数量。

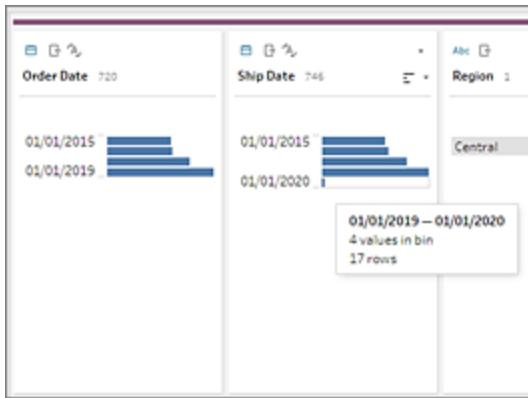


查看值或唯一值的分布

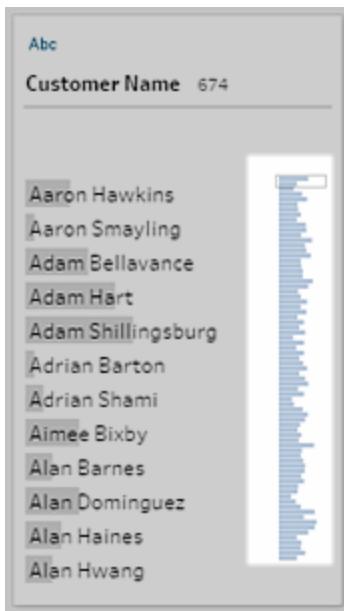
默认情况下，Tableau Prep Builder 会将字段中的数字、日期和日期时间值分组为存储段。这些存储段也称为数据桶。数据桶确保您能查看值的整体分布状况，并快速确定离群值和 null 值。数据桶大小是基于字段中的最小值和最大值计算得出的，并且 null 值始终显示在分布的顶部。

例如，将按年份对订单和发货日期进行汇总或“分桶”。每个数据桶都表示从起始年份一月份到下一年一月份的一年，并带有相应标签。由于有处于 2018 年和 2019 年下半年的销售日期及发货日期，因此将针对这些值为以下年度创建数据桶。

Tableau Prep 帮助



如果离散(或分类)数据字段包含许多行,或者其具有的分布太大,如果不滚动就无法显示在字段中,您可能会在字段的右侧看到汇总分布。您可以在分布中单击和滚动来定位特定值。

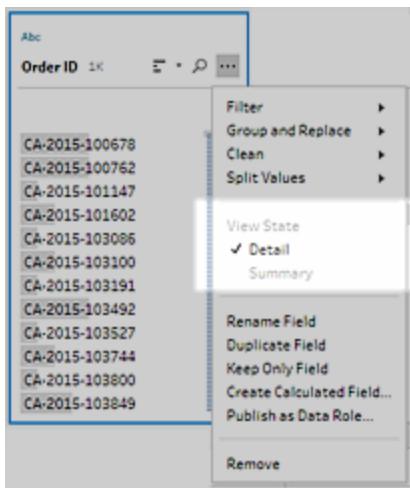


如果数据包含数字或日期字段,您可以切换显示值的详细(离散)版本或值的汇总(连续)版本。汇总视图显示字段中的值范围,以及某些值的出现频率。

这种切换可帮助您隔离唯一值(比如字段中的“3”记录数)或值的分布(比如字段中所有“3”记录的总和)

切换视图:

1. 在“配置”窗格、“结果”窗格或数据网格中,单击数值或日期字段的“更多选项”...菜单。



- 在上下文菜单中，选择“**详细信息**”以查看值的详细版本，或选择“**摘要**”以查看值的分布式版本。

搜索字段和值

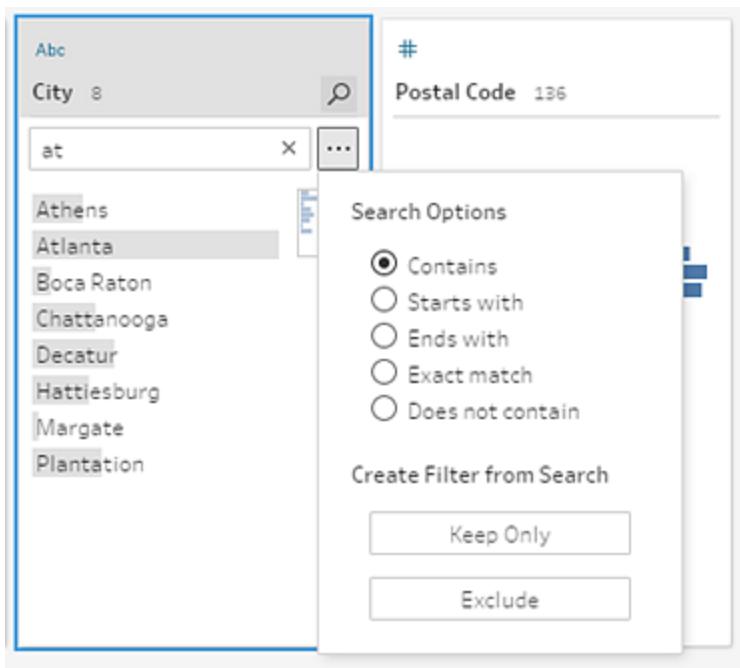
在“配置”窗格或“结果”窗格中，您可以搜索您特别感兴趣的字段或值，并使用搜索结果来筛选数据。

若要搜索字段，请在工具栏上的搜索框中输入全部或部分搜索词。

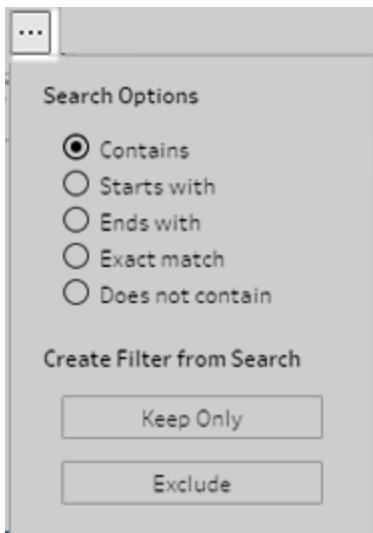
Order Date	Order ID
11/22/2016	US-2016-118983
11/22/2016	US-2016-118983
11/11/2015	CA-2015-105893
12/09/2017	CA-2017-137330

在字段中搜索值：

1. 单击字段的搜索图标 ，并输入一个值。



2. 若要使用高级搜索选项，请单击“搜索选项...”按钮。

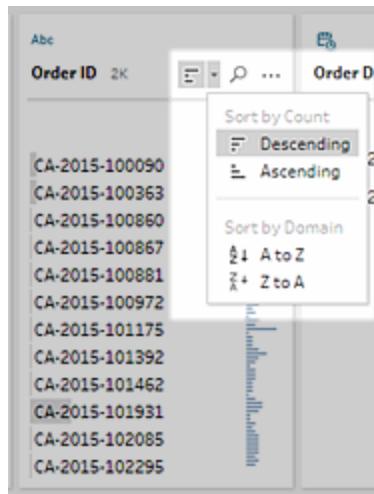


3. 若要使用搜索结果来筛选数据，请选择“只保留”或“排除”。

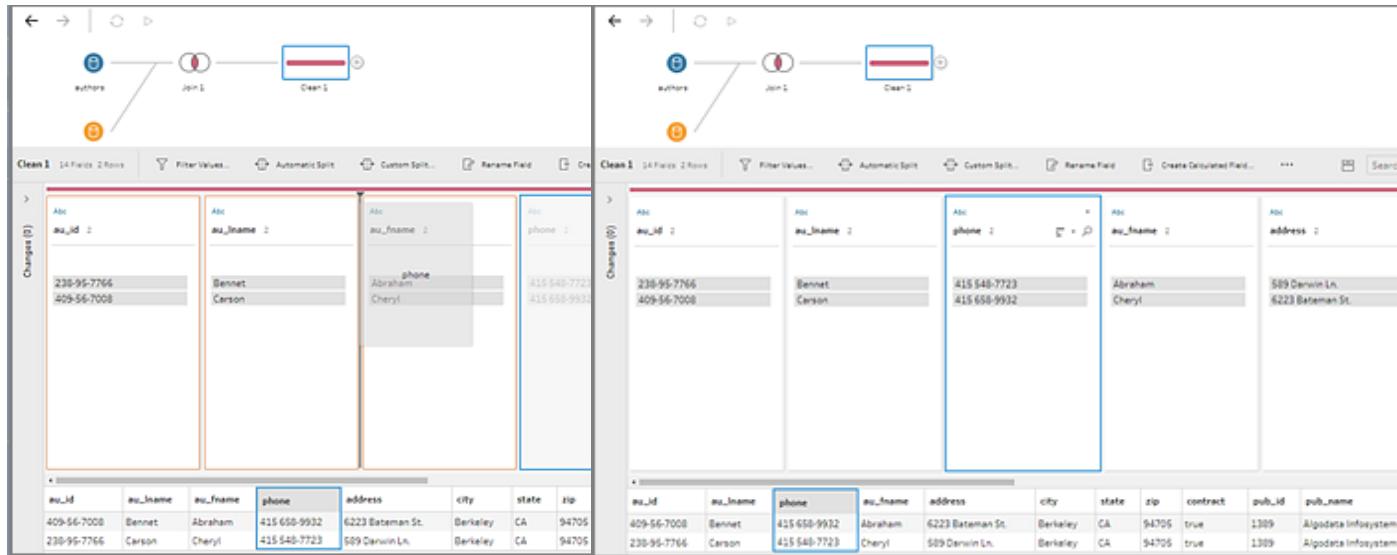
在“流程”窗格中，受影响的步骤上方将出现一个筛选器图标。

对值和字段进行排序

“配置”卡上的排序选项使您能按升序或降序对数据桶(由分布条形表示的值计数)进行排序，或按字母顺序对单独的字段值进行排序。



如果要重新排列字段的顺序，只需在“配置”窗格、“结果”窗格或“数据”网格中选择并拖动配置卡或数据网格中的字段，直至出现黑色定位线为止。然后将其放置到位。“配置”窗格和数据网格是同步的，因此字段将以相同的顺序出现在两个位置中。



突出显示流程中的字段和值

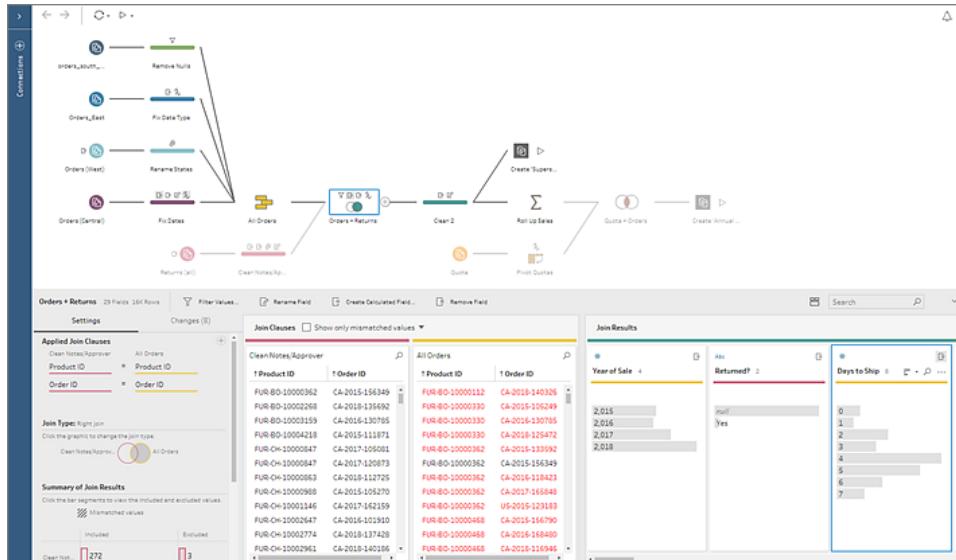
利用 Tableau Prep Builder 可轻松地查找流程数据中的字段和值。在流程窗格中跟踪字段的来源以及在流程中的何处使用了字段，或者在配置卡或数据网格中单击单独的值以突出显示相关值或相同的值。

在流程中跟踪字段

在 Tableau Prep Builder 版本 2018.3.3 及更高版本中，您可以突出显示字段在流程中的每处使用位置，甚至可以跟踪字段的来源，从而可帮助您在未看到预期结果时找出缺少的值或排查流程故障。

在清理步骤的“配置”窗格中或任何其他步骤类型的“结果”窗格中单击某个字段，流程窗格将突出显示其中使用了该字段的路径。

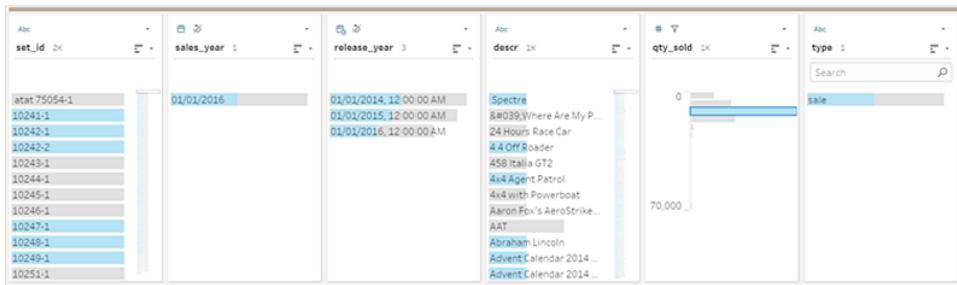
注意：此选项不可用于“输入”或“输出”步骤类型。



查看相关值

您可以使用突出显示来查找字段之间的相关值。当您在“配置”窗格或“结果”窗格的“配置”卡中单击某个值时，其他字段中的所有相关值会以蓝色突出显示。蓝色显示您选择的值与其他字段中的值之间的关系分布。

例如，若要突出显示相关值，请在“配置”窗格中单击字段中的一个值。其他字段中的相关值将变为蓝色，以蓝色突出显示的条形比例表示关联程度。



突出显示相同的值

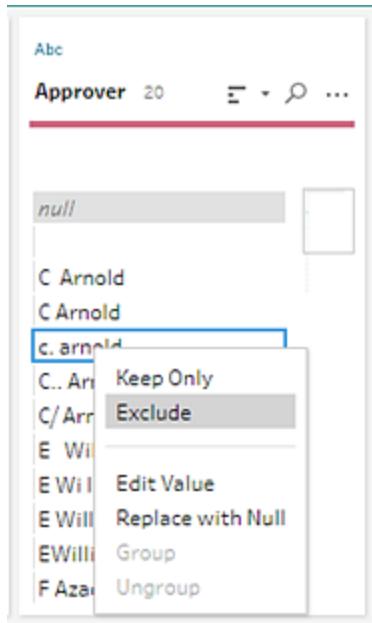
在数据网格中选择值时，所有相同的值也会突出显示。这些突出显示可帮助您确定数据中的模式或不规则情况。

Type	Customer	Purchases	Date
Cash	Wei	5	08/18/2016
Cash	Jim	7	07/15/2016
Credit	Arnold	5	06/29/2016
Credit	Lee	1	08/07/2016
Cash	Maria	2	08/30/2016
Cash	Wendy	1	07/21/2016
Credit	Max	2	07/02/2016
Credit	Juan	1	05/10/2016
Cash	Isaac	4	06/28/2016
Credit	Philip	1	08/09/2016
Credit	Lane	5	05/04/2016

筛选您的数据

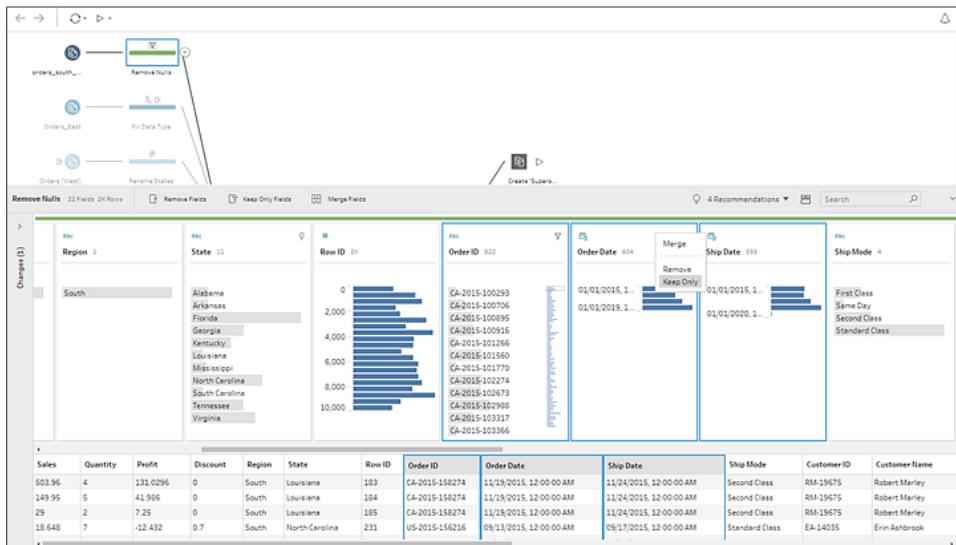
Tableau Prep Builder 提供了可用于筛选数据的各种选项。例如，使用“只保留”或“排除”以对配置卡、数据网格或结果卡中的字段的特定值进行单击筛选，或者针对更复杂的筛选需求从各种筛选器选项中进行选择。也可以保留或移除整个字段。

在流程的任何步骤中对数据进行筛选。如果要只更改特定值，您可以选择“**编辑值**”以内联方式编辑值，或将值替换为 Null。有关编辑字段值的详细信息，请参见[编辑字段值 在本页 239。](#)



保留或移除字段

在处理流程中的数据时，您可能需要移除不需要的字段。在任何清理或操作步骤的“配置”窗格或数据网格中，选择一个或多个字段，并右键单击或按住 Ctrl 单击 (MacOS)，然后选择“**移除**”以移除所选字段，或选择“**只保留**”(版本 2019.2.2 及更高版本) 以仅保留所选字段，并移除所有未选择的字段。

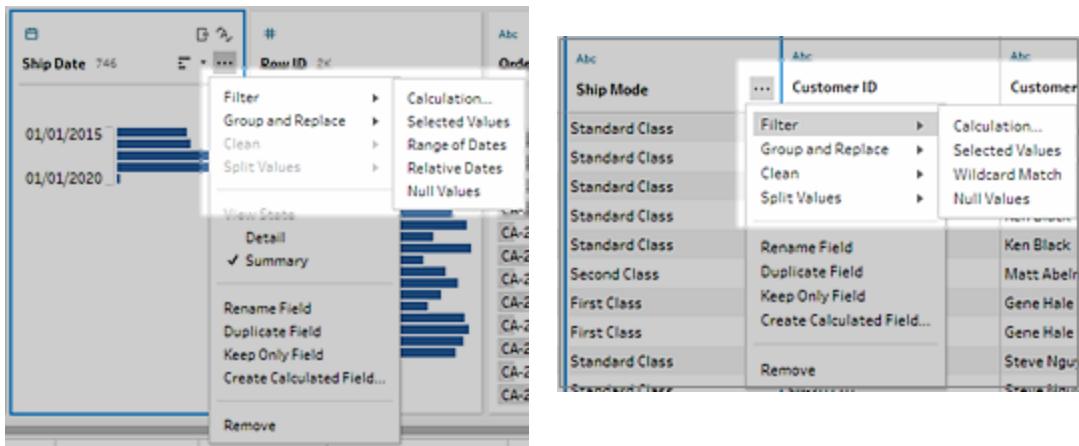


可用于每种数据类型的筛选器

数据类型	可用筛选器
字符串	计算、通配符匹配、Null 值、选定值
数字	计算、值范围、Null 值、选定值
日期及日期和时间	计算、值范围、相对日期、Null 值、选定值

筛选器选项位于何处？

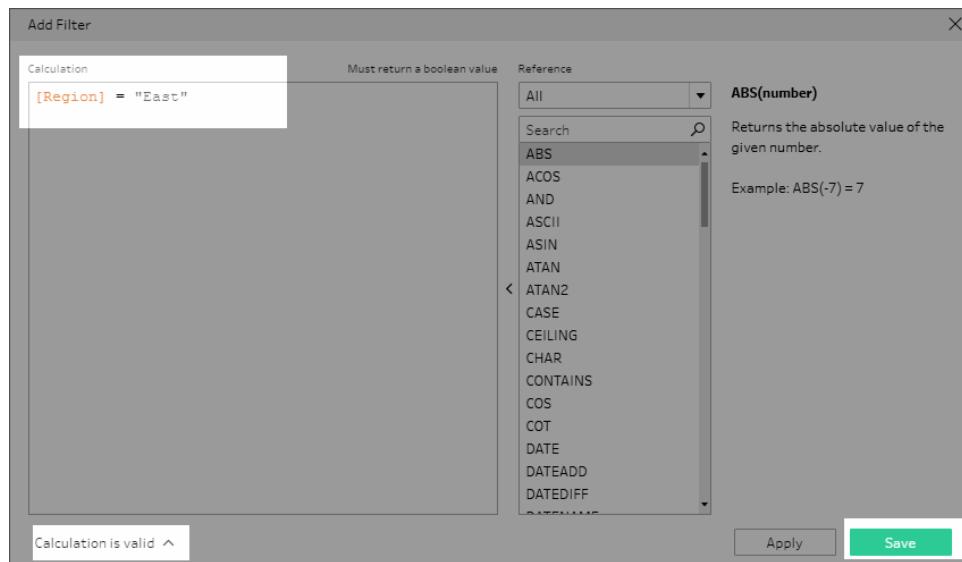
若要查看可用于字段的不同筛选器选项，请在配置卡、数据网格或结果窗格中单击“更多选项”... 菜单。若要查看数据网格上的菜单，您必须先单击“隐藏配置窗格” 按钮，然后单击“更多选项”...。



计算筛选器

当您选择“计算”时，“添加筛选器”对话框将打开。输入计算，验证其是否有效，并单击“保存”。

注意：在“输入”步骤中，这是唯一可用的筛选器类型。所有其他筛选器类型都可以在配置卡、数据网格或结果窗格中使用。

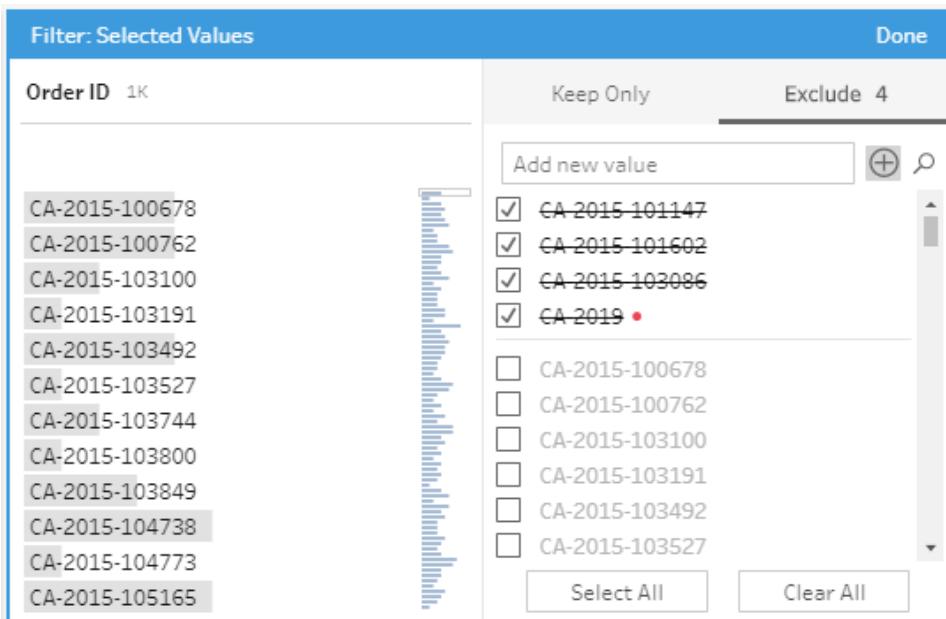


“选定值”筛选器(版本 2019.2.3 及更高版本)

借助新的“选定值”筛选器，您可以选取和选择要为字段保留或排除的值，即使这些值不在样本中。在右侧窗格中，单击“只保留”或“排除”选项卡选择操作，然后输入搜索词以搜索

值，或单击“添加值”以添加位于数据集中但未包括在样本中的值。单击“完成”应用筛选器。

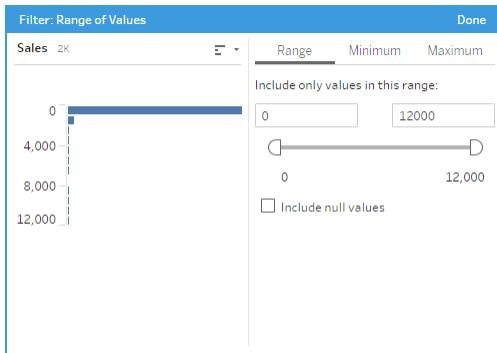
注意：此筛选器选项对于“聚合”或“转置”步骤类型不可用。



“值范围”筛选器

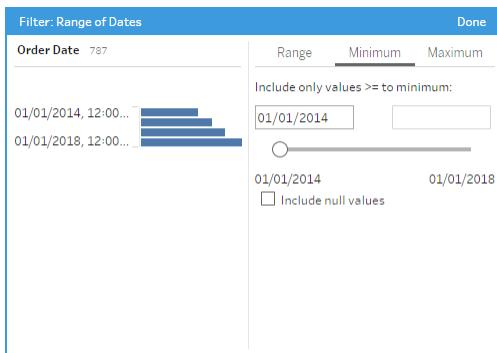
筛选出位于特定范围内的值。当您选择“值范围”时，您可以指定范围，或者设置最小值或最大值。

Tableau Prep 帮助



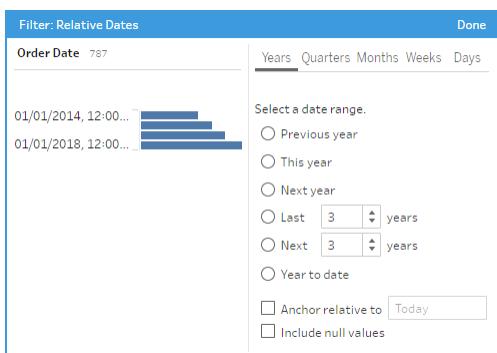
“日期范围”筛选器

筛选出位于特定日期范围内的值。当您选择“日期范围”时，您可以指定日期范围，或者设置最早日期或最晚日期。



“相对日期”筛选器

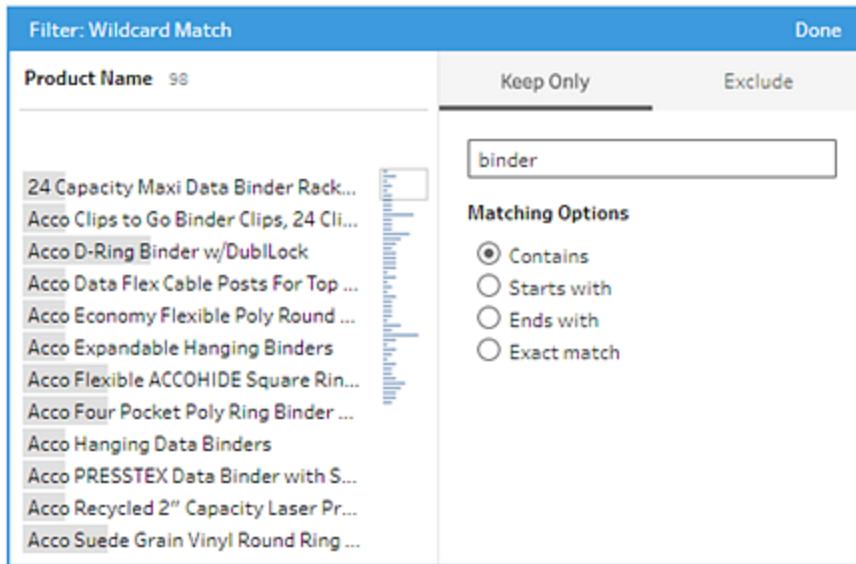
使用“相对日期”筛选器来指定要在数据中查看的年、季度、月、周或天的确切范围。您也可以配置相对于特定日期锚点，并包括 **null** 值。



“通配符匹配”筛选器

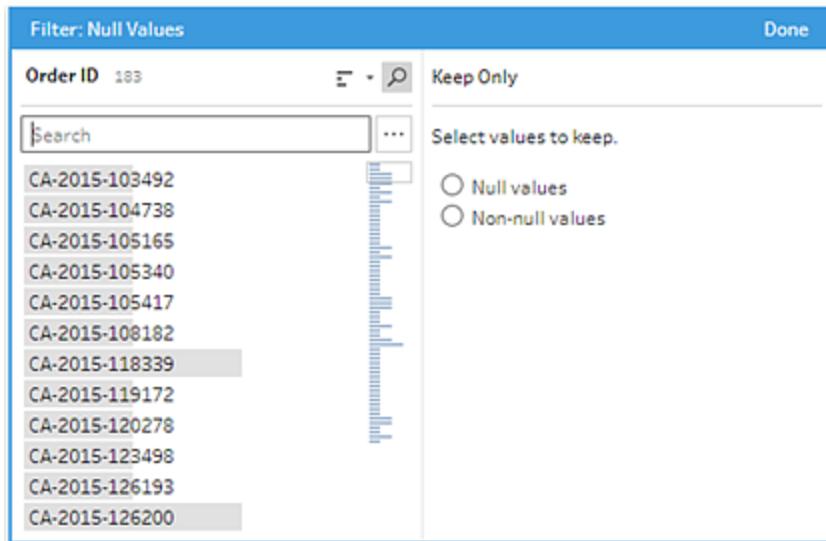
如果选择“通配符匹配”，您可以筛选字段值以保留或仅排除符合某个模式的值。在筛选编辑器中，选择“只保留”或“排除”选项卡，输入要匹配的值，然后设置“匹配选项”条件以返回所寻找的值。

筛选的结果显示在筛选编辑器的左侧窗格中，以便您能够查看和试验结果。有了所需的结果之后，单击“完成”应用更改。



“Null 值”筛选器

如果选择“Null 值”，您可以筛选所选字段中的值，以仅显示 null 值或排除所有 null 值。



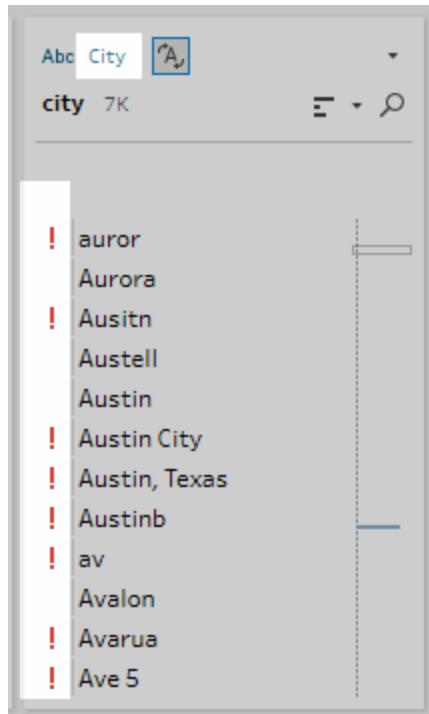
使用数据角色验证数据

使用数据角色来快速确定字段中的值是否有效。Tableau Prep Builder 提供了一组您可从中选择的标准数据角色，或者，您可以使用数据集中的唯一字段值来创建自己的数据角色。

分配数据角色时，Tableau Prep Builder 会将为数据角色定义的标准值与您的字段中的值进行比较。任何不匹配的值都用红色感叹号标记。您可以筛选字段以仅查看有效或无效的值，并采取相应的措施来修复这些值。将数据角色分配给字段后，您可以使用“**分组和替换**”选项来进行分组，并基于拼写和发音将无效值与有效值匹配。

为数据分配标准数据角色

采用与分配数据类型相同的方式将 Tableau Prep Builder 提供的数据角色分配给字段。数据角色确定数据值所代表的意义，使 Tableau Prep Builder 能够自动验证值并突出显示对于该角色无效的值。



举例来说，如果您有地理数据字段值，您可以分配数据角色“城市”，Tableau Prep Builder 即会将这些字段中的值与一组已知的域值或模式进行比较来确定不匹配的值。

注意：系统会对每个字段进行单独分析，因此国家/地区“美国”中的州“华盛顿”中的城市“波特兰”可能不是有效的城市和州组合，但并不会将其标识为无效，因为它是有效的城市名称。

Tableau Prep Builder 提供以下数据角色：

- 电子邮件
- URL
- 地理角色(基于当前地理数据，并且与 Tableau Desktop 使用的数据相同)
 - 机场
 - 区号(美国)
 - CBSA/MSA
 - 城市
 - 国会选区(美国)
 - 国家/地区
 - 郡/县
 - NUTS 欧洲

- 州/省/市/自治区
- 邮政编码

提示:在 Tableau Prep Builder 版本 2019.1.4 及更高版本中,如果为字段分配地理角色,则还可以使用该数据角色,通过数据角色定义的标准值来匹配值和对值进行分组。有关使用数据角色对值进行分组的详细信息,请参见[清理和调整数据 在本页 223](#)。

若要为字段分配数据角色,请执行以下操作:

1. 在“配置”窗格、“结果”窗格或数据网格中,单击字段的数据类型。
2. 为字段选择数据角色。

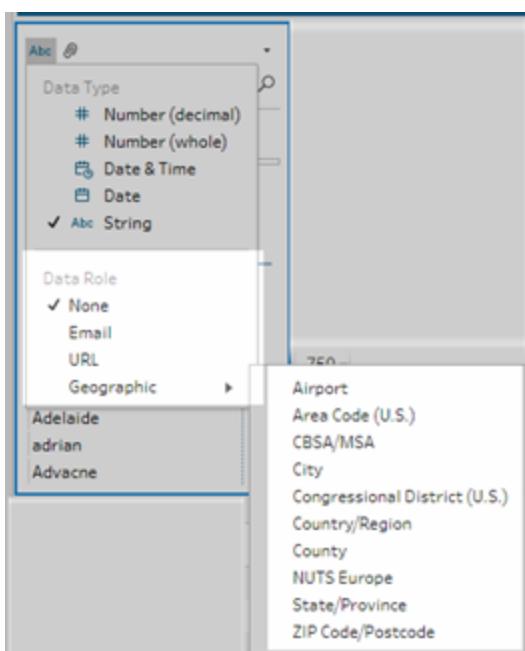
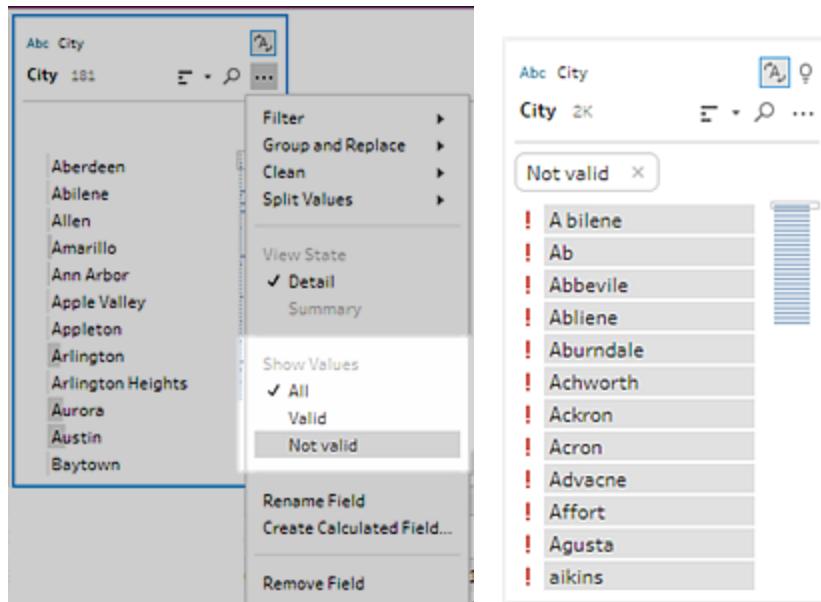


Tableau Prep Builder 会将字段的数据值与所选数据角色的已知域值或模式(对于电子邮件或 URL)进行比较,并用红色感叹号标记任何不匹配的值。

3. 单击字段的下拉箭头,并从“显示值”部分选择一个选项,以显示所有值或仅显示对于数据角色有效或无效的值。



4. 使用字段的“更多选项”… 菜单中的清理选项来更正无效的任何值。有关如何清理字段值的详细信息，请参见[关于清理操作 在本页 223](#)。

创建自定义数据角色(版本 2019.3.1 及更高版本)

您可以使用数据集中的字段值创建自己的自定义数据角色，来为您或他人可用于在清理数据时验证字段值的一组标准值。选择要使用的字段，将任何清理操作应用于该字段(如果需要)，然后将该字段发布到 Tableau Server 或 Tableau Online 以在流程中使用它或与他人共享您的数据角色。

在创建自定义数据角色之前，请查看以下内容：

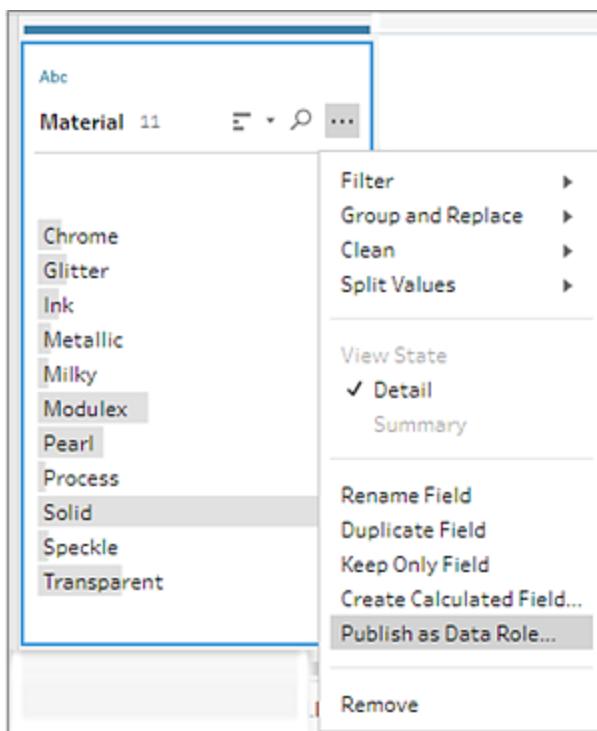
- 您可以依据数据集中的单个字段创建自定义数据角色。不支持依据字段组合创建自定义数据角色。
- 只能为分配有“字符串”和“数字(整数)”数据类型的字段创建自定义数据角色。
- 当您创建自定义数据角色时，Tableau Prep Builder 会在您的流程中创建一个专用于发布数据角色的输出步骤。
- 不支持将自定义数据角色发布到同一流程中的多个站点。如果发布流程，您必须将自定义数据角色发布到在其中发布流程的同一站点或服务器。
- 自定义数据角色特定于您在其中发布这些角色的站点、服务器和项目。具有该位置权限的所有用户都可以使用自定义数据角色，但必须登录到站点或服务器才能选择或

应用该角色。将为自定义数据角色分配“所有用户”组的默认权限，分配给新项目的“所有用户”权限，而不是“无”。

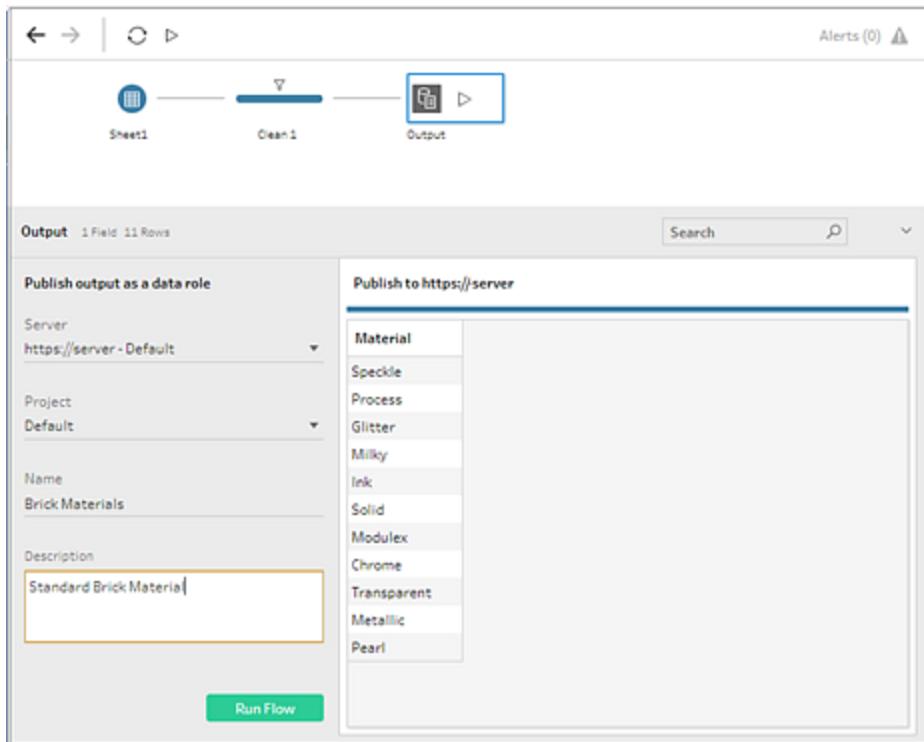
- 自定义数据角色不是特定于版本的。应用自定义数据角色时，将应用最新版本。
- 发布到 Tableau Server 或 Tableau Online 后，具有站点、服务器和项目访问权限的用户可以查看该位置中的所有数据角色。
 - 具有适当权限的用户可以移动、删除或编辑数据角色的权限。
 - 您可对自定义数据角色设置的权限以及执行的操作与可对流程设置的权限和执行的操作类似。有关详细信息，请参见[管理流程 在本页 400](#)。有关设置权限的详细信息，请参见 Tableau Server 帮助中的[权限能力](#)。
- 若要编辑数据角色，您必须在 Tableau Prep Builder 中进行更改，然后使用相同的名称重新发布数据角色以覆盖它，就像编辑已发布的数据源一样。

创建自定义数据角色

1. 在“配置”窗格、数据网格或“结果”窗格中，选择要用于创建自定义数据角色的字段。
2. 为该字段单击“更多选项”…，并选择“以数据角色形式发布”。



3. 选择要在其中发布数据角色的服务器和项目。



4. 单击“运行流程”以创建数据角色。发布过程成功完成后，您可以在 Tableau Server 或 Tableau Online 中查看数据角色。根据 Tableau Server 或 Tableau Online 站点上的负载，处理数据角色可能需要一段时间。如果您的数据角色无法立即使用，请等待几分钟，然后再次尝试选择。

Explore / Default / Brick Materials



Brick Materials ☆ ...

Owner datarockstar · Modified Aug 9, 2019, 3:39 PM

Definition

About

Role Type	Data Type
Dictionary	String

Description
Standard Brick Material

Values (11)

Name
Chrome
Glitter
Ink
Metallic
Milky
Modulex
Pearl
Process
Solid
Speckle
Transparent

应用自定义数据角色

1. 在“配置”窗格、“结果”窗格或数据网格中，单击要在其中应用自定义数据角色的字段的数据类型。
2. 选择“自定义”，然后选择要应用于字段的数据角色。

重要信息：请确保登录到在其中发布了数据角色的站点或服务器，否则您将看不到此选项。

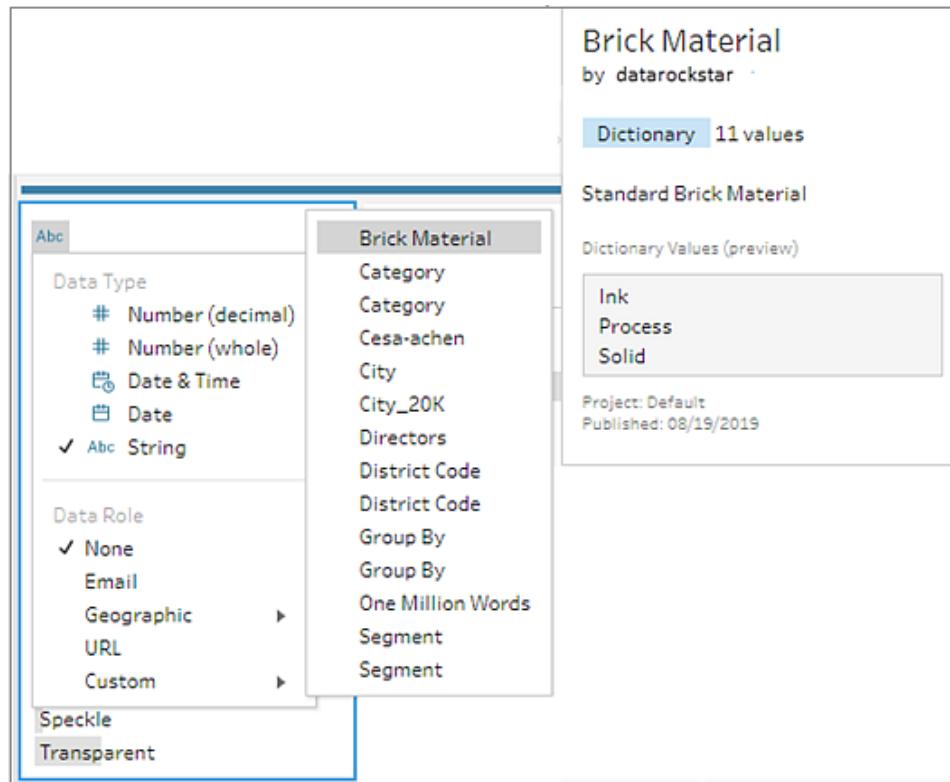
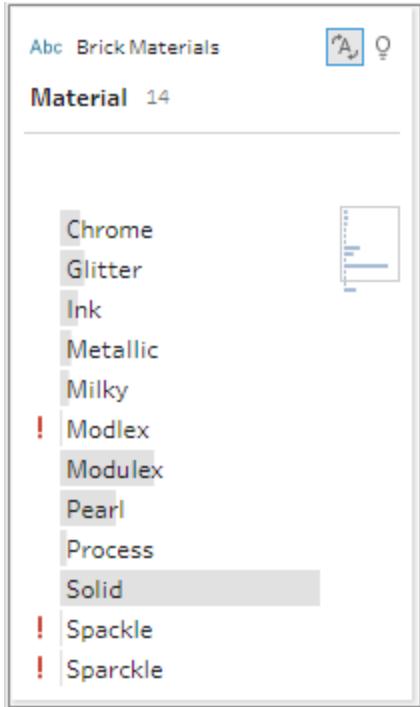
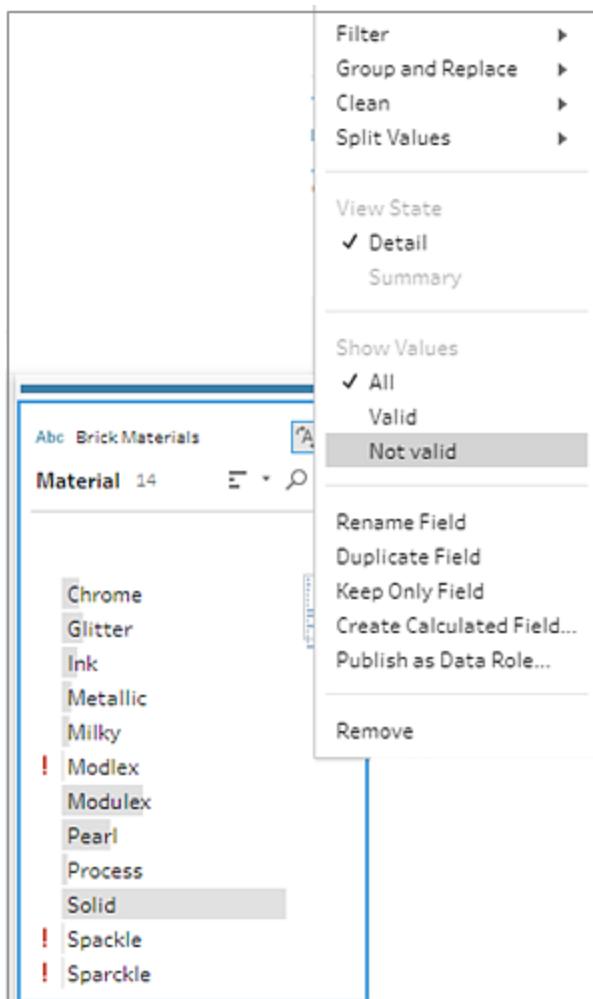


Tableau Prep Builder 会将字段的数据值与所选数据角色的已知域值进行比较，并用红色感叹号标记任何不匹配的值。



3. 单击字段的下拉箭头，并从“显示值”部分选择一个选项，以显示所有值或仅显示对于数据角色有效或无效的值。



4. 使用字段的“更多选项”… 菜单中的清理选项来更正无效的任何值。有关如何清理字段值的详细信息，请参见[关于清理操作 在本页 223](#)。

查看和管理自定义数据角色

您可以在 Tableau Server 和 Tableau Online 上查看及管理发布的自定义数据角色。您可以查看发布到您的站点或服务器的所有自定义数据角色。针对所选数据角色单击“更多操作”…，以将其移到其他项目、更改权限或将其删除。

Type	Name	Project	Owner	Modified
<input type="checkbox"/>	city-datarole	Default	workuser	Jun 25, 2019, 5:37 PM
<input type="checkbox"/>	city-datarole2		workuser	Jun 25, 2019, 5:42 PM
<input type="checkbox"/>	City1		workuser	Jun 18, 2019, 11:23 AM

按数据角色对类似值进行分组

注意: 在 Tableau Prep Builder 版本 2019.1.4 和 2019.2.1 中，此选项标为“**数据角色匹配**”。

如果为字段分配地理数据角色，您可以使用数据角色中的值，基于拼写和发音对数据字段中的值进行分组和匹配，从而使值标准化。在 Tableau Prep Builder 版本 2019.2.3 中，您可以使“**拼写**”或“**拼写 + 发音**”对无效值进行分组，并将其与有效值匹配。

这些选项使用数据角色定义的标准值。如果数据集示例中没有标准值，Tableau Prep Builder 会自动添加该值，并将值标记为不在原始数据集中。有关为字段分配数据角色的详细信息，请参见[为数据分配标准数据角色](#)在本页 209。

若要使用数据角色对值进行分组，请完成以下步骤。

1. 在“配置”窗格、“结果”窗格或数据网格中，单击字段的数据类型。
2. 为字段选择以下数据角色之一：
 - 机场
 - 城市
 - 国家/地区
 - 郡/县
 - 州/省/市/自治区

在 Tableau Prep Builder 版本 2019.3.2 中，您还可以从自定义数据角色中进行选择

**标准数据角色(版本
2019.1.4 及更高版本)**

自定义数据角色(版本 2019.3.2 及更高版本)

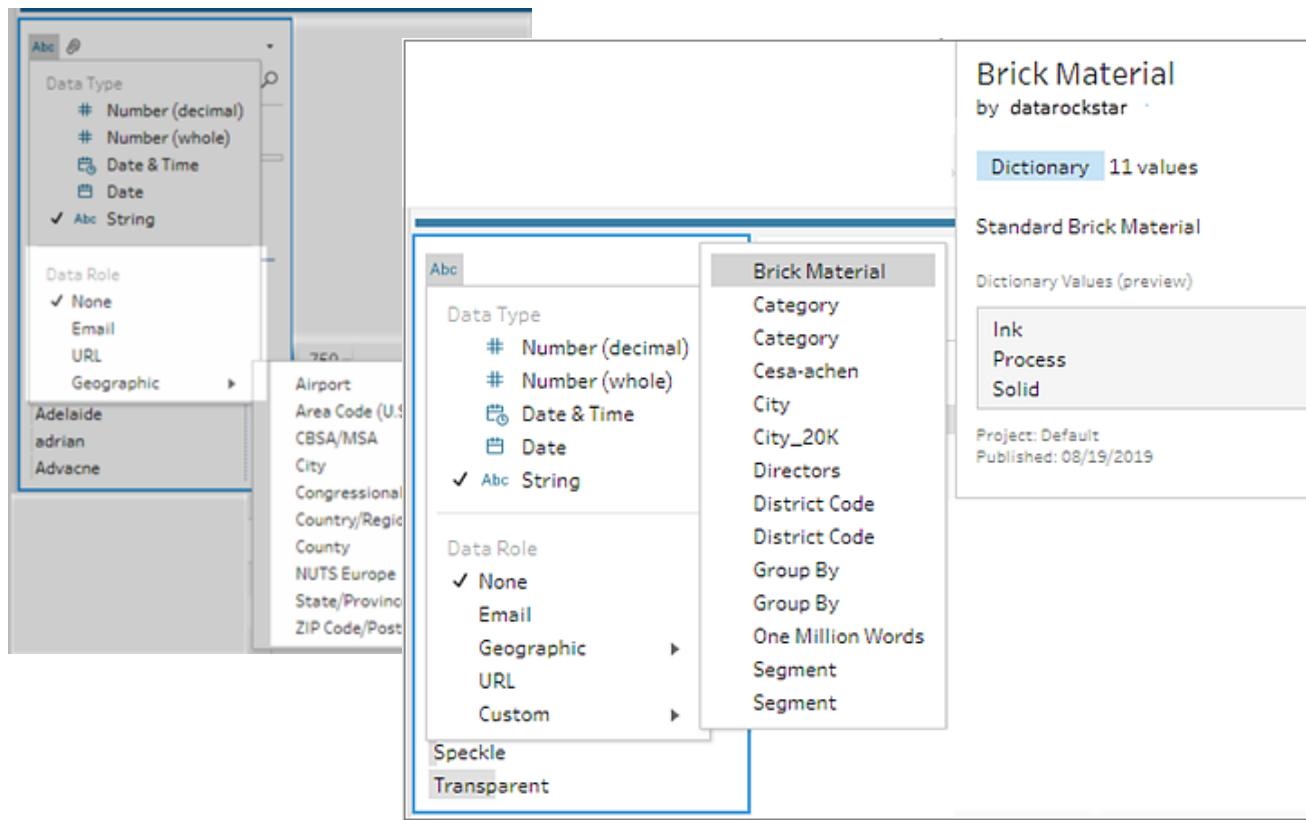
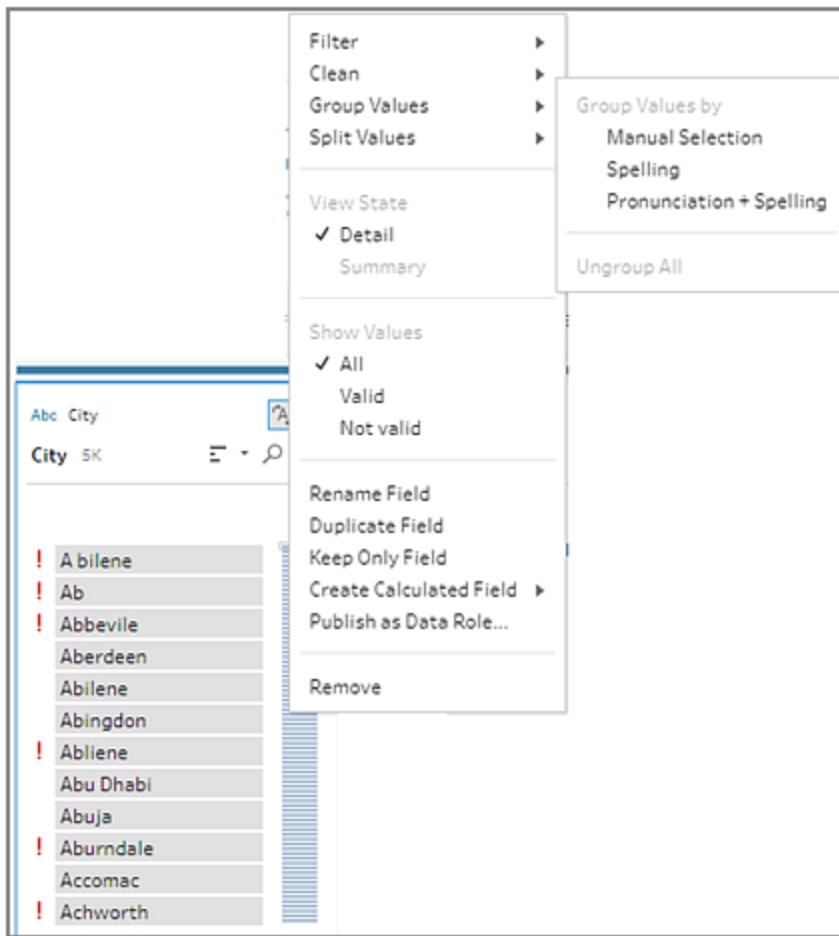


Tableau Prep Builder 会将字段的数据值与所选数据角色的已知域值进行比较，并用红色感叹号标记任何不匹配的值。

3. 单击“更多选项”…，选择“对值进行分组”(在以前的版本中为“分组和替换”)，然后选择以下选项之一：
 - **拼写**: 将无效值与因为添加、移除或替换了字符而不同的最接近有效值匹配。
 - **发音 + 拼写**: 根据拼写和发音将无效值与最相似的有效值匹配。

注意: 在 Tableau Prep Builder 版本 2019.1.4 或 2019.2.1 中，此选项称为“数据角色匹配”。



您还可以单击字段上的“建议”图标将建议应用于组，并将无效值替换为有效的值。此选项使用“发音 + 拼写”分组和替换选项。

Tableau Prep Builder 按拼写或者拼写和发音比较各个值，然后依据数据角色的标准化值对类似的值进行分组。如果数据集中没有标准化值，则系统会添加该值，并使用红点加以标记。

Group Values by Pronunciation + Spelling

Done

City 5K

Albany
Albemarle
! Albermale
@ Albertville
! albrightsfield
Albrightsville
@ Albuquerque
! Albququerque
! alburn
@ Alburquerque
ALCOA
Aldan

Albuquerque 4 members

! Albuquerque
 Albuquerque
 ! Albuquerue
 ! Abuqurque

! Abilene
 ! Ab
 ! Abbeville
 Abbeville •
 Aberdeen
 Abilene
 Abingdon
 ! Abliene
 Abu Dhabi

清理和调整数据

Tableau Prep Builder 提供了您可立即用来清理和调整数据的各种清理操作。通过清理不良数据，可以更轻松地合并和分析数据，或可以让其他人在共享您的数据集时更轻松地理解您的数据。

您还可以使用转置步骤或脚本步骤来清理数据，以将 R 或 Python 脚本应用于流程。有关详细信息，请参见[转置您的数据 在本页 273](#)或在流程中使用 R 和 Python 脚本 在本页 281。

关于清理操作

您通过应用诸如筛选、添加、重命名、拆分、分组或移除字段等清理操作来清理数据。在以前的版本中，清理操作只能在“清理步骤”类型中执行。在 Tableau Prep Builder 版本 2018.2.1 及更高版本中，您还可以在清理步骤的数据网格中执行清理操作。

在 Tableau Prep Builder 版本 2018.3.1 及更高版本中，您可以在流程中的大多数步骤类型中执行清理操作。您可以在输入步骤中应用有限的清理操作，并且不能在输出步骤中应用清理操作。有关在输入步骤中应用清理操作的详细信息，请参见[在输入步骤中应用清理操作 在本页 164](#)。

可用的清理操作

下表显示了在每个步骤类型中可以执行哪些清理操作：

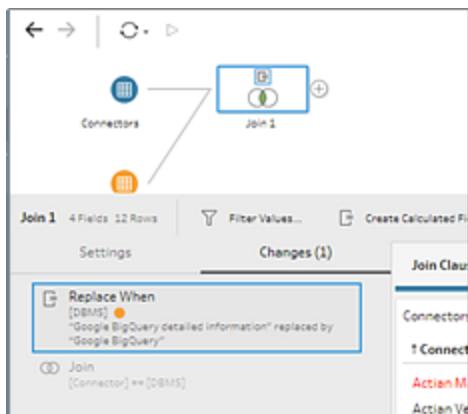
	输入	清理	聚合	转置	联接	合并	输出
筛选	X	X	X	X	X	X	
对值进行分组		X		X		X	
清理		X		X	X	X	
转换日期		X	X	X	X	X	
拆分值		X		X	X	X	
重命名字段	X	X	X	X	X	X	
复制字段		X		X	X	X	

仅保留字段	X	X	X	X	X	X	
移除字段	X	X	X	X	X	X	
创建计算字段		X		X	X	X	
编辑值		X		X	X	X	
更改数据类型	X	X	X	X	X	X	

对数据进行更改时，将会向“流程”窗格中的对应步骤中添加注释，并会在“更改”窗格中添加一个条目来跟踪您的操作。如果在“输入”步骤中进行更改，则注释会显示在“流程”窗格中步骤的左侧，并也会显示在字段列表的“输入配置”中。

您应用更改的顺序很重要。在“聚合”、“转置”、“联接”和“合并”步骤类型中进行的更改会在这些调整操作之前或之后执行，具体情况取决于您进行更改时字段的位置。进行更改的位置显示在步骤的“更改”窗格中。

以下示例显示了联接步骤中单个表的字段更改。将在联接操作之前执行更改以生成更正的结果。



操作顺序

下表显示了“聚合”、“转置”、“联接”和“合并”步骤类型中清理操作的执行位置，具体情况取决于字段在步骤中的何处。

操作 步骤 类型:	聚合	聚合	转置	转置	联接	联接	合并	合并

字段位置：	分组字段	聚合字段	不在转置中	通过转置创建	包含在一个表中*	同时包含在两个表中*	不匹配的字段	组合字段
筛选		聚合之前	聚合之后	转置之前	转置之后	联接之前	联接之后	合并之前 合并之后
对值进行分组		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前 合并之后
清理		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前 合并之后
转换日期		聚合之前	聚合之后	转置之前	转置之后	联接之前	联接之后	合并之前 合并之后
拆分值		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前 合并之后
重命名字段		聚合之前	聚合之后	转置之前	转置之后	联接之前	联接之后	合并之前 合并之后
复制字段		NA	NA	转置之前	转置之后	联接之前	联接之后	合并之前 合并之

									后
仅保 留字 段		聚合 之后	聚合 之后	转置 之前	转置 之后	联接之 前	联接之后	合并 之前	合 并之 后
移除 字段		从聚 合中 移除	从聚 合中 移除	转置 之前	转置 之后	联接之 前	联接之后	合并 之前	合 并之 后
创建 计算 字段		NA	NA	转置 之前	转置 之后	联接之 后	联接之后	合并 之前	合 并之 后
编辑 值		NA	NA	转置 之前	转置 之后	联接之 前	联接之后	合并 之前	合 并之 后
更改 数据类 型		聚合 之前	聚合 之后	转置 之前	转置 之后	联接之 前	联接之前	合并 之前	合 并之 后

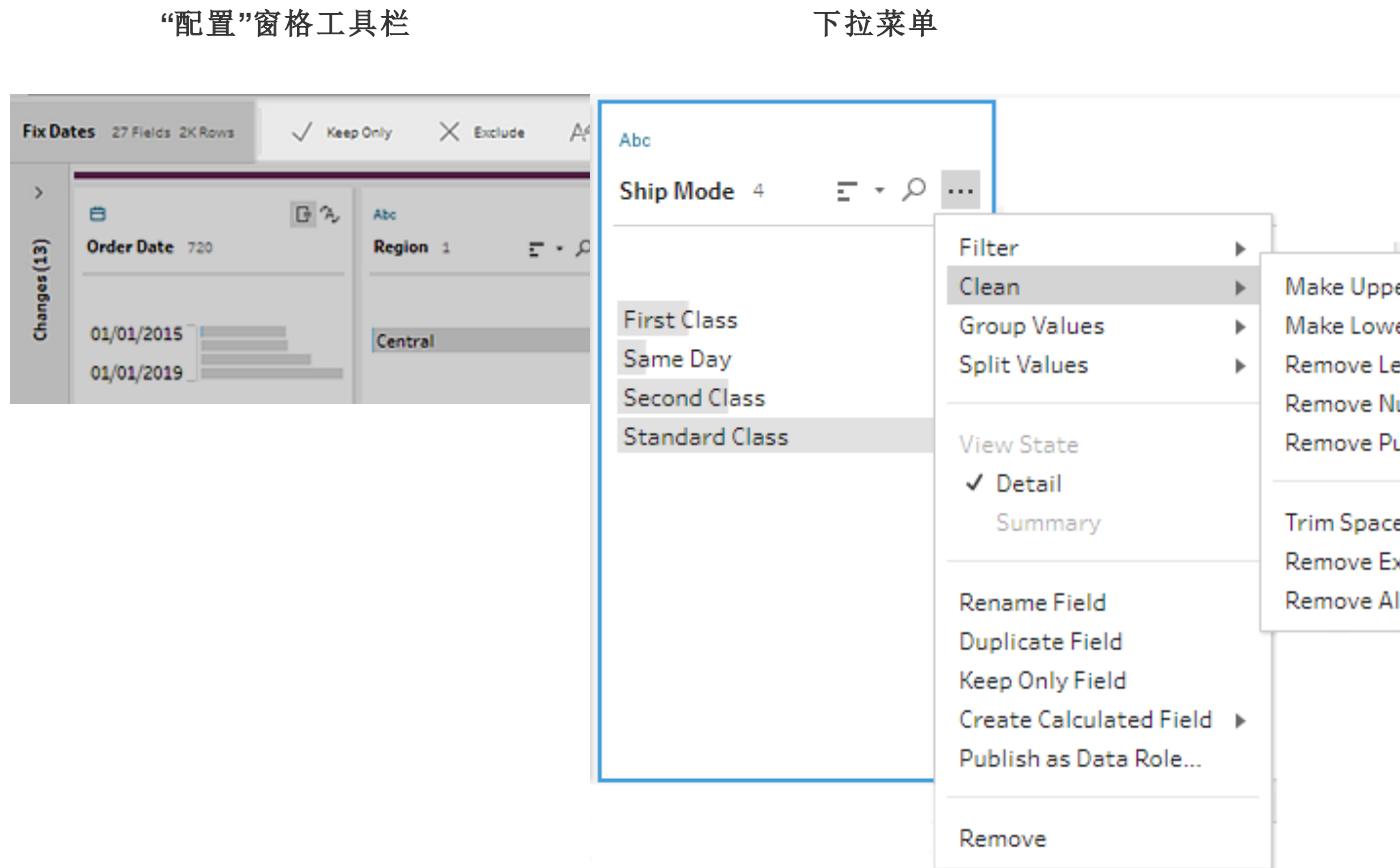
注意:对于联接,如果字段是使用一个表中的字段创建的计算字段,则在联接之前应用更改。如果字段是同时使用两个表中的字段创建的,则在联接之后应用更改。

应用清理操作

若要向字段应用清理操作,请使用工具栏选项,或在字段配置卡、数据网格或“结果”窗格上单击**“更多选项”…** 打开菜单。

在“聚合”、“转置”、“联接”和“合并”步骤类型中,可以在“结果”窗格和对应数据网格中的配置卡上找到**“更多选项”…** 菜单。如果您在整个流程中反复执行相同的清理操作或操作,则可

以复制和粘贴步骤、操作甚至字段。有关详细信息，请参见[复制步骤、操作和字段](#)在本页 252。

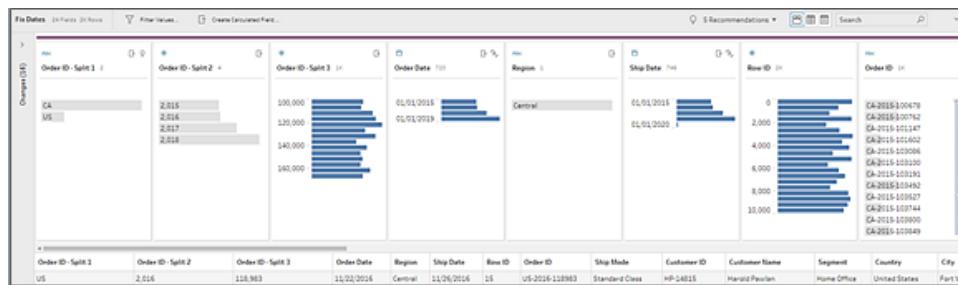


选择视图

您可以在数据网格或列表视图中的配置窗格或结果窗格外部中执行清理操作。使用视图工

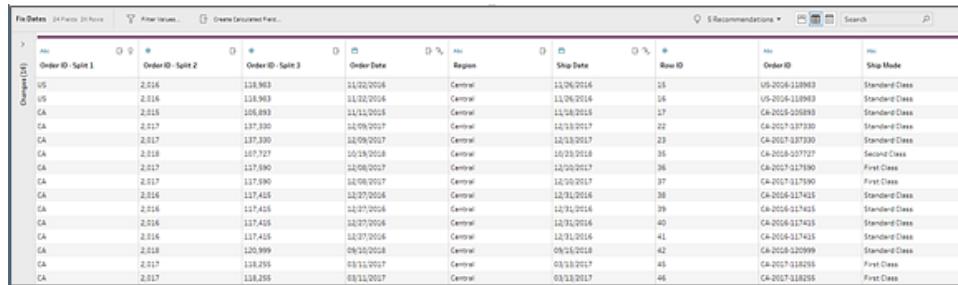
具栏 (版本 2019.3.2 及更高版本) 更改视图, 然后在字段上单击“更多选项”... 打开清理菜单。

- 显示配置窗格 : 这是默认视图。选择此按钮可返回配置窗格或结果窗格视图。



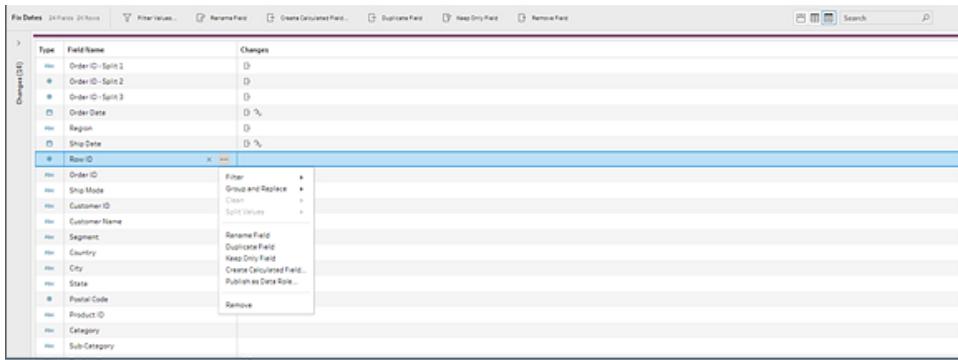
- 显示数据网格 : 折叠配置窗格或结果窗格以展开并仅显示数据网格。此视图提供了更详细的数据视图，在您需要处理特定字段值时非常有用。选择此选项后，此视图状态将在流程的所有步骤中保持不变，但您可以随时对其进行更改。

注意: 并非所有清理操作都可在数据网格中使用。举例来说，如果对值进行内联编辑，则必须在“配置”窗格中执行此操作。



- 显示列表视图 (版本 2019.3.2 及更高版本) : 将配置窗格或结果窗格转换为列表。选择此选项后，此视图状态将在流程的所有步骤中保持不变，但您可以随时对其进行更改。

您可以选择多个行并单击 **X** 将其移除，或者使用“更多选项” 菜单将操作应用于所选字段。如果为字段分配数据角色，或选择“筛选”、“分组和替换”、“清理”或“拆分值”，则将返回到“配置”或“结果”视图来完成这些操作。可以在列表视图中执行所有其他选项。



版本 2019.3.1 及更低版本

使用视图工具栏 隐藏“配置”窗格，并只显示数据网格。然后在数据网格中的某个字段上单击**“更多选项”**… 以打开清理菜单。此视图提供了更详细的数据视图，在您需要处理特定字段值时非常有用。选择此选项后，此视图状态将在流程的所有步骤中保持不变，但您可以随时对其进行更改。

注意：并非所有清理操作都可在数据网格中使用。举例来说，如果对值进行内联编辑，则必须在“配置”窗格中执行此操作。

	Order ID - Split 1	Order ID - Split 2	Order ID - Split 3	Order Date	Region	Ship Date	Row ID	Order ID	Ship Mode
VS	2,016	128,963	11/20/2008	Central	11/26/2016	15	US-2020-1110963	Standard Class	
VS	2,016	128,963	11/20/2008	Central	11/26/2016	16	US-2020-1110963	Standard Class	
CA	2,015	106,893	11/11/2005	Central	11/16/2015	17	CA-2015-1050893	Standard Class	
CA	2,017	137,330	12/01/2015	Central	12/13/2017	22	CA-2017-137330	Standard Class	
CA	2,017	137,330	12/01/2017	Central	12/13/2017	23	CA-2017-137330	Standard Class	
CA	2,018	137,727	10/19/2018	Central	10/23/2018	35	CA-2018-137727	Second Class	
CA	2,017	137,390	12/01/2017	Central	12/05/2017	36	CA-2017-137390	First Class	
CA	2,017	137,390	12/06/2017	Central	12/05/2017	37	CA-2017-137390	First Class	
CA	2,016	137,415	12/27/2016	Central	12/31/2016	38	CA-2016-137415	Standard Class	
CA	2,016	137,415	12/27/2016	Central	12/31/2016	39	CA-2016-137415	Standard Class	
CA	2,016	137,415	12/27/2016	Central	12/31/2016	40	CA-2016-137415	Standard Class	
CA	2,018	120,999	09/20/2018	Central	09/15/2018	41	CA-2018-120999	Standard Class	
CA	2,017	128,295	03/11/2017	Central	03/13/2017	45	CA-2017-128295	First Class	
CA	2,017	128,295	03/11/2017	Central	03/13/2017	46	CA-2017-128295	First Class	

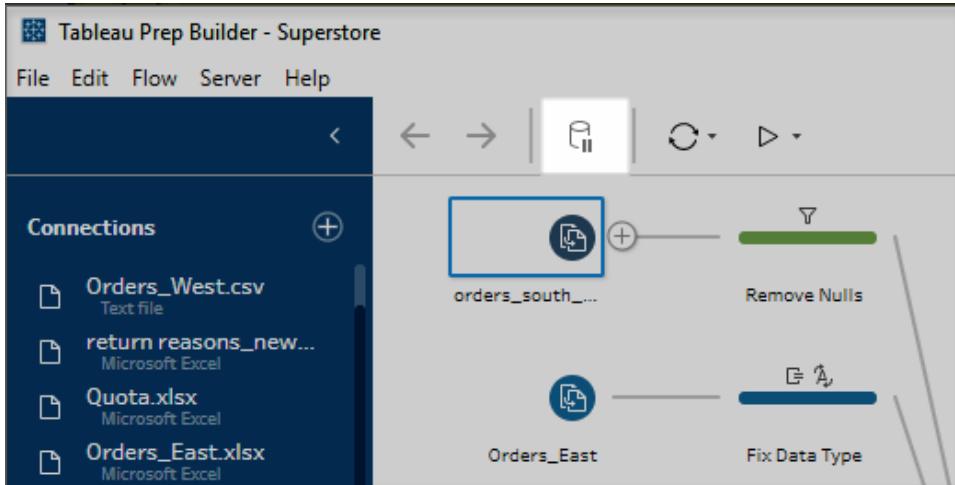
暂停数据更新以提高性能

对数据执行清理操作时，Tableau Prep Builder 会随着进展应用更改，以便立即向您显示结果。但是，如果您确切知道要进行哪些更改，并且不需要在进行每次更改时立即获得反馈，或者只想提高性能，则可以暂停数据更新以节省宝贵的处理时间。暂停数据更新时，可以一

次进行所有更改，然后恢复更新以查看结果。您可以随时恢复数据更新并启用所有可用操作。

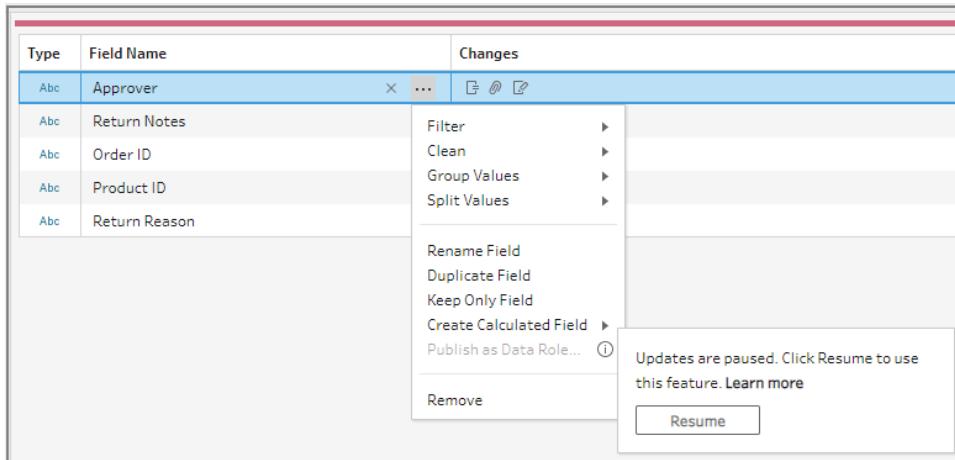
注意：暂停数据更新时，将禁用任何要求您查看值的操作。举例来说，如果要将筛选器应用于所选值，则需要查看要排除的值。

1. 在顶部菜单中，单击“暂停数据更新”以暂停更新。

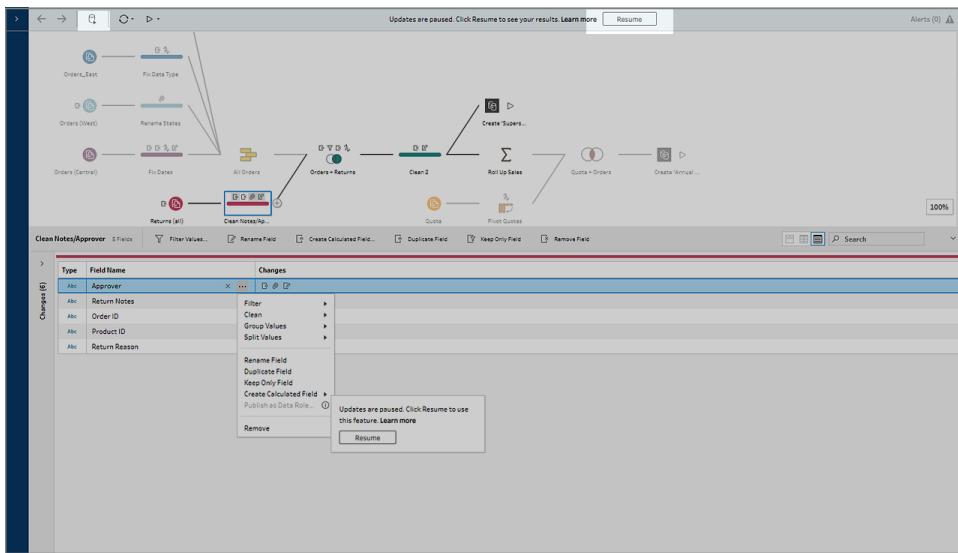


2. Tableau Prep Builder 会将“配置”窗格转换为“列表”视图。在“列表”视图中，使用“更多选项”... 菜单将操作应用于所选字段。如果操作要求您查看值，它将被禁用。若要启用该操作，您将需要恢复数据更新。

有关使用“列表”视图模式的详细信息，请参见[选择视图 在本页 227](#)。



3. 若要查看更改的结果或启用禁用的功能，请恢复数据更新。单击“恢复数据更新”按钮，单击菜单对话框中或“流程”窗格顶部消息横幅中的“恢复”按钮。



应用清理操作

若要将清理操作应用于字段，请执行以下操作：

注意:从版本 2019.3.2 开始，您可以在列表视图中执行清理操作。

1. 在“配置”窗格、数据网格、“结果”窗格或列表视图中，选择要进行更改的字段。
2. 从工具栏或字段的“更多选项”... 菜单中，从以下选项中进行选择：
 - **筛选或筛选值:**选择筛选器选项之一，右键单击或按住 Ctrl 并单击 (MacOS) 字段值以保留或排除值。您也可以使用“选定值”筛选器来选取和选择要筛选的值，包括流程样本中没有的值。有关筛选器选项的详细信息，请参见[筛选您的数据 在本页 202](#)。
 - **对值进行分组(以前的版本中为“分组和替换”):**手动选择值或使用自动分组。您也可以在“配置”卡中选择多个值，并右键单击或按住 Ctrl 并单击 (MacOS) 以对值进行分组或取消分组，或者编辑组值。有关使用“对值进行分组”的详细信息，请参见[使用模糊匹配将值自动映射到标准值 在本页 248](#)。
 - **清理:**从快速清理操作列表中进行选择，以便应用于字段中的所有值。

- **转换日期**(版本 2020.1.4 及更高版本):对于分配给“日期”或“日期和时间”数据类型的字段，从 DATEPART 快速清理操作列表中选择，将日期字段值转换为表示年份、季度、月份、周、日或日期和时间值的整数值。
- **拆分值**:基于常见分隔符自动拆分值，或使用自定义拆分来指定要如何拆分字段值。

自动拆分和自定义拆分的工作方式与它们在 Tableau Desktop 中的工作方式相同。有关详细信息，请参见 Tableau Desktop 和 Web 制作帮助中的[将字段拆分为多个字段](#)。

- **重命名字段**:编辑字段名称。
- **复制字段**(版本 2019.2.3 及更高版本):创建字段和值的副本。
- **仅保留字段**(版本 2019.2.2 及更高版本):仅保留所选字段并排除步骤中的所有其他字段。
- **创建计算字段**:在计算编辑器中编写自定义计算，或使用可视化计算编辑器(版本 2020.1.1 及更高版本)创建详细级别、排名或行号计算。有关详细信息，请参见[创建详细级别和排名计算 在本页 261](#)。
- **移除**(在以前的版本中为“**移除字段**”):从流程中移除字段。

注意:从版本 2019.3.1 开始，您可以使用“**以数据角色形式发布**”选项创建随后可应用于字段以在清理数据时验证字段值的自定义数据角色。有关此选项的详细信息，请参见[创建自定义数据角色\(版本 2019.3.1 及更高版本\) 在本页 212](#)。

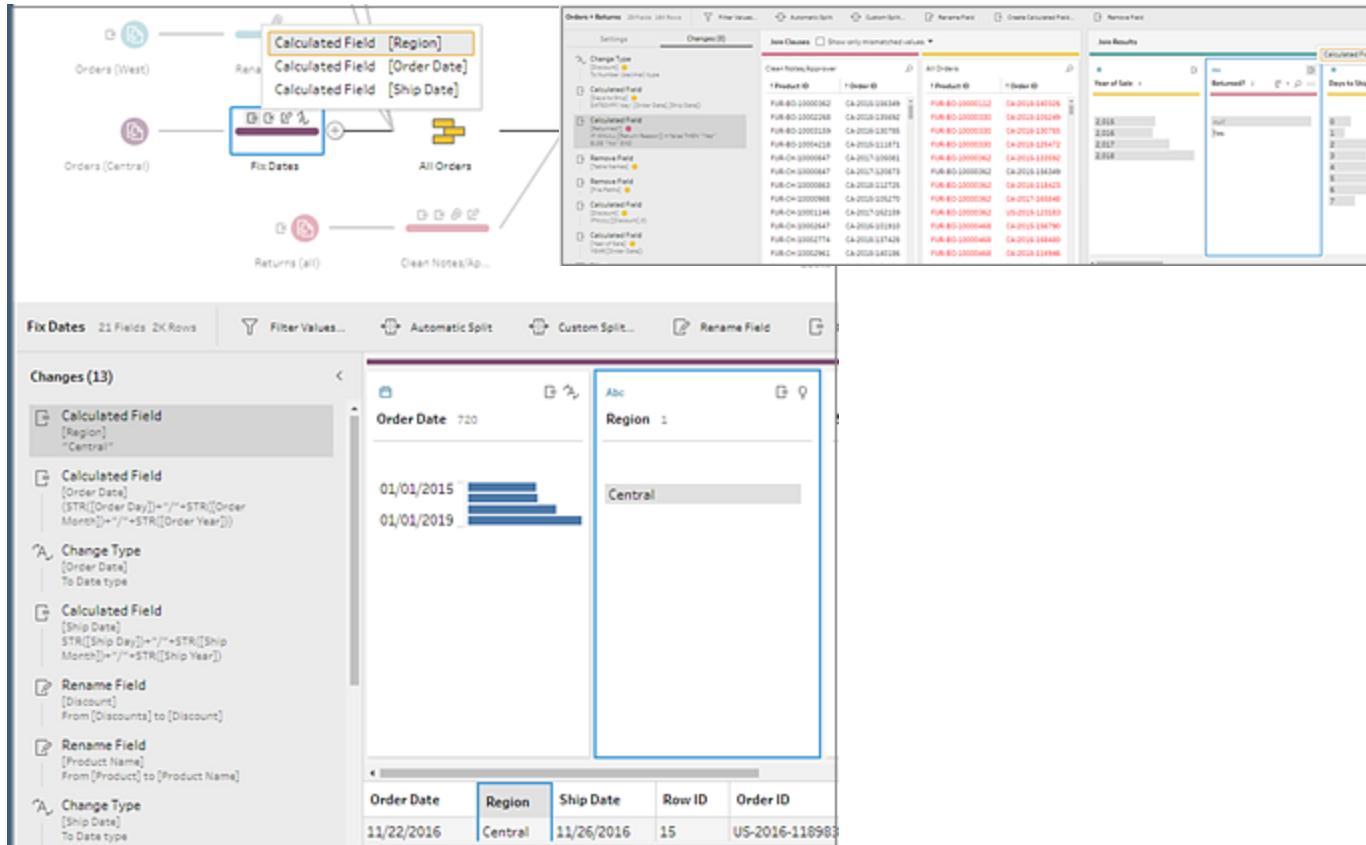
3. 若要编辑值，请右键单击或按住 Ctrl 并单击 (MacOS) 一个或多个值，选择“**编辑值**”，然后输入新值。您也可以选择“**替换为 Null**”将值替换为 Null 值，或在单个字段中双击以直接对其进行编辑。有关编辑字段值的详细信息，请参见[编辑字段值 在本页 239](#)。
4. 在“**配置**”窗格、“**摘要**”窗格或数据网络中查看这些操作的结果。

查看您所做的更改

不同类型的清理操作由流程中步骤上的图标表示。如果向步骤应用了超过四种类型的操作，则步骤上会显示省略号。将光标悬停在这些图标上即可查看显示所应用的操作及其执行顺序的注释。

Tableau Prep 帮助

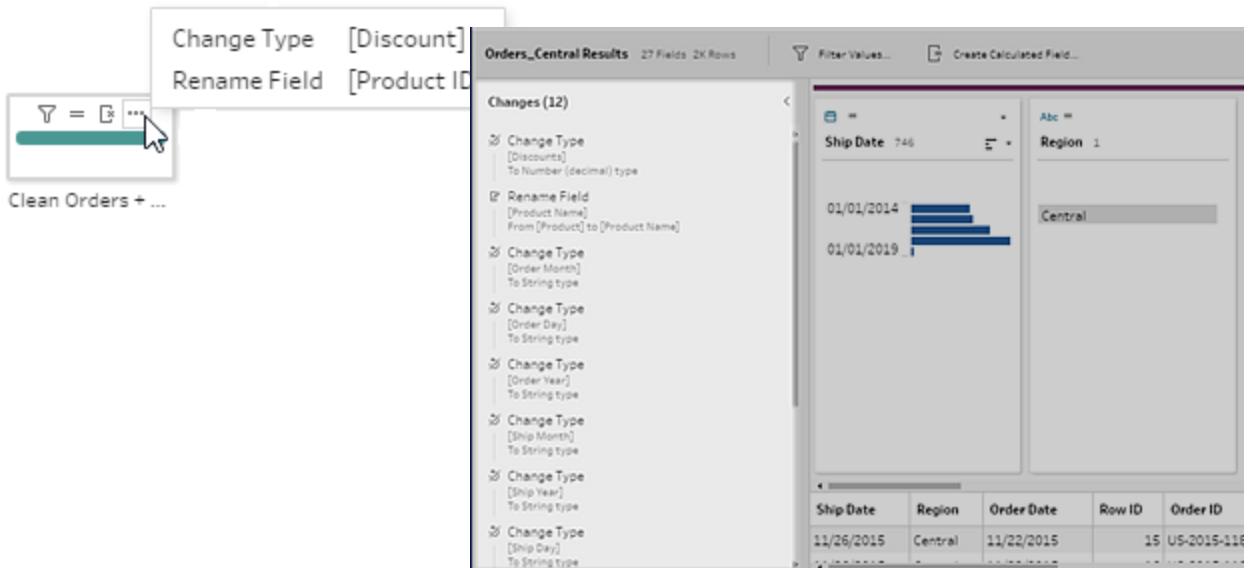
从 Tableau Prep Builder 版本 2019.1.3 及更高版本开始，您可以在“流程”窗格中某个步骤上单击更改图标上的注释，或者单击“配置”或“结果”窗格中的某个配置卡，更改及其影响的字段将会在“更改”窗格以及“配置”或“结果”窗格中突出显示。



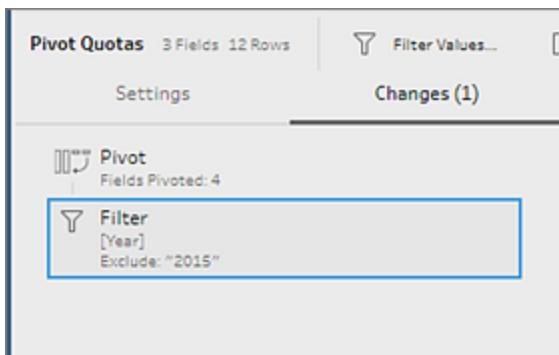
您也可以选择步骤，然后展开“更改”窗格以查看每个更改的详细信息、编辑或移除更改，或者将更改向上或向下拖动以更改其应用顺序，并且，在 Tableau Prep Builder 版本 2019.1.1 中，您可以添加说明，以便为其他用户提供上下文。有关为更改添加说明的详细信息，请参见[为流程步骤和清理动作添加说明 在本页 186](#)

清理注释

“更改”窗格



在“聚合”、“转置”、“联接”或“合并”步骤中查看更改时，更改的应用顺序会在调整操作之前或之后显示。这些更改的顺序由系统应用，无法更改。您可以编辑和移除更改。



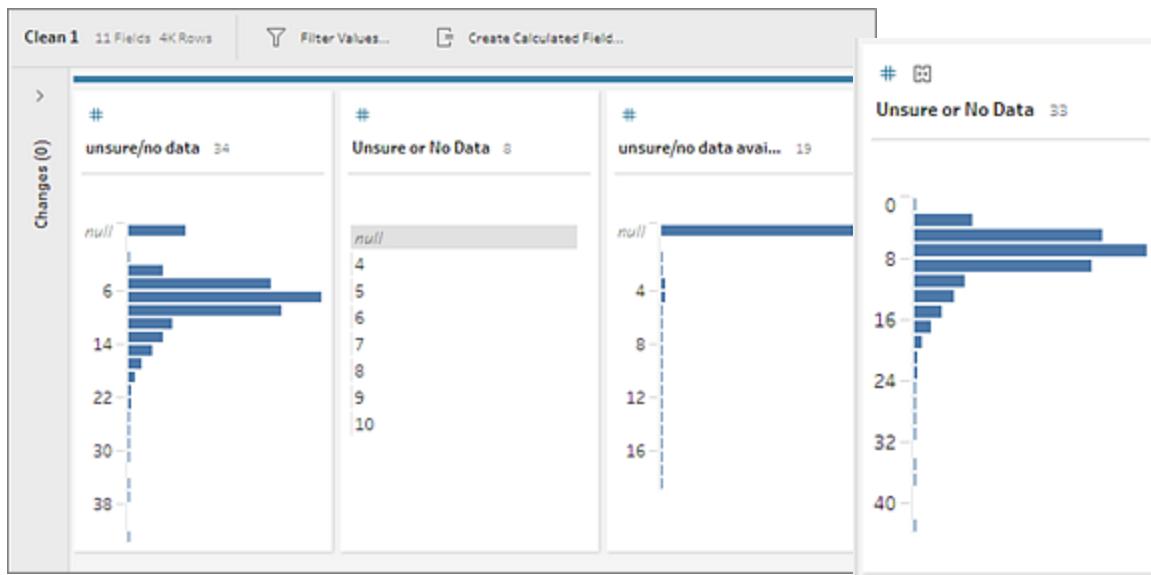
合并字段

如果字段包含的值相同，但名称不同，您可以通过将一个字段拖到另一个字段上，轻松地将它们合并为单个字段。在合并字段时，目标字段将成为主字段，并且目标字段的字段名称保持不变。合并到目标字段的字段会被移除。

示例：

通配符并集生成 3 个具有相同值的字段

将 3 个字段合并为 1 个字段



合并字段时，Tableau Prep Builder 会保留目标字段中的所有字段，并将该字段中的任何 null 值替换为您与目标字段合并的源字段中的值。源字段会被移除。

示例

Name	Contact_Phone	Business_Phone	Cell_Phone	Home_Phone
Bob	123-4567	123-4567	null	null
Sally	null	null	456-7890	789-0123
Fred	null	null	null	567-8901
Emma	null	234-5678	345-6789	null

如果将 **Business_Phone**、**Cell_Phone** 和 **Home_Phone** 字段与 **Contact_Phone** 字段合并，则会移除其他字段并生成以下内容：

Name	Contact_Phone
Bob	123-4567
Sally	456-7890
Fred	567-8901

Emma	234-5678
------	----------

若要合并字段，请执行以下操作之一：

- 将一个字段拖放到另一个字段上。将显示“放置以合并字段”指示符。
- 选择多个字段，并在所选范围内右键单击以打开上下文菜单，然后单击“合并字段”。
- 选择多个字段，然后在工具栏上单击“合并字段”。

有关如何修复合并所产生的不匹配字段的信息，请参见[修复不匹配的字段 在本页 306](#)。

使用建议应用清理操作(版本 2018.3.2 及更高版本)

有时可能很难确定您需要使用哪些清理操作来修复数据中的问题。Tableau Prep Builder 可对数据进行分析并提出清理操作建议，您可以自动应用这些操作来快速修复数据字段中的问题，或者帮助确定问题以便能对其进行修复。除“输入”、“输出”和“联接”步骤类型外的所有步骤类型中提供了此功能。

注意：如果不想使用此功能，您可以将其关闭。从顶部菜单中，转到“帮助”>“设置和性能”。然后单击“启用建议”以清除设置旁边的复选标记。

建议类型包括：

- 数据角色
- 筛选
- 对值进行分组(从版本 2019.2.3 开始，也适用具有数据角色的字段)
- 将列转置为行(2019.4.2 及更高版本)
- 将值替换为 Null 值
- 移除字段
- 拆分(版本 2019.1.1 及更高版本)

注意：此选项专用于处理固定宽度类型文本文件中的数据。若要将拆分建议与此文件类型结合使用，请在连接到数据源后，在“输入”步骤的“文本设置”选项

卡中选择一个数据中未使用的“字段分隔符”字符，使数据以单一字段形式加载。

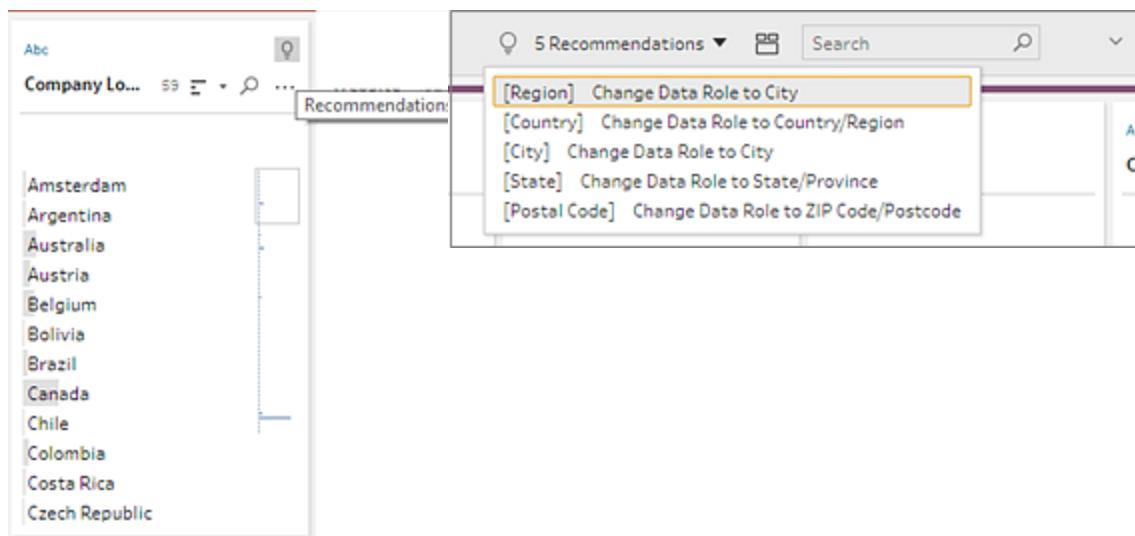
- 剪裁空格

应用建议

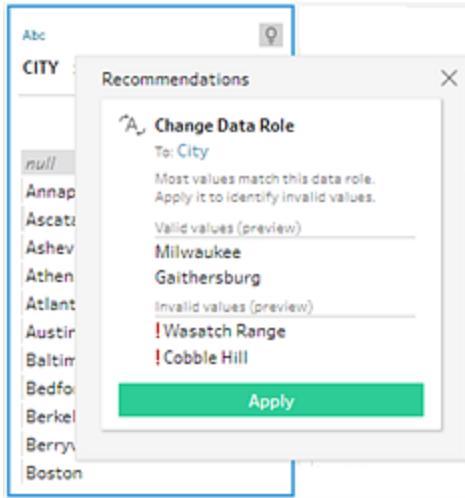
1. 执行以下操作之一：

- 单击配置卡右上角中的灯泡  图标。
- 从工具栏中单击“建议”下拉箭头以查看适用于您的数据集的所有建议，并从列表中选择一个建议。

只有在 Tableau Prep Builder 确定了建议的更改时，此选项才会出现。



2. 若要应用建议，请将鼠标指针悬停在“建议”卡上，然后单击“应用”。

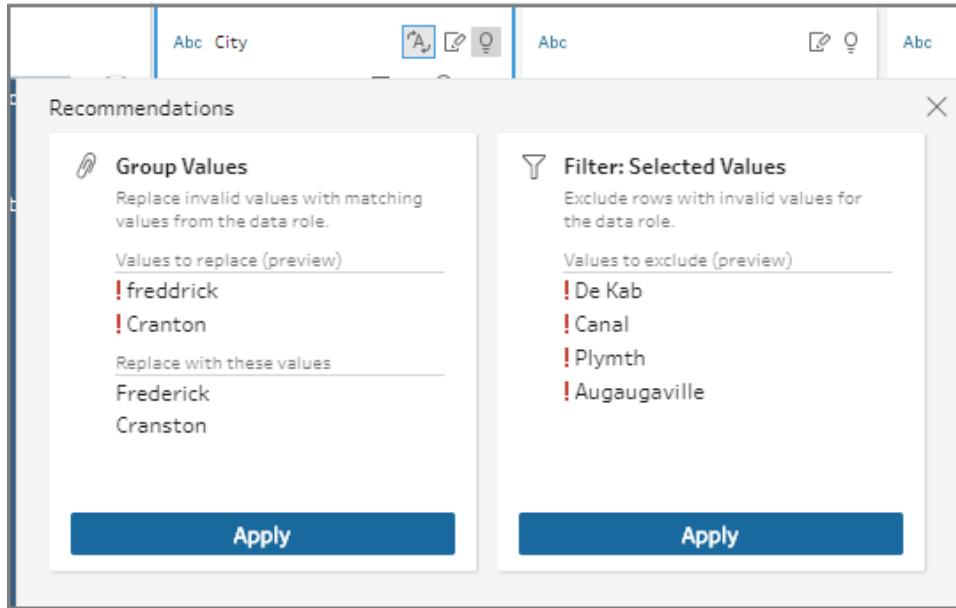


系统会自动应用更改，并向“**更改**”窗格中添加一个条目。若要移除更改，请在顶部菜单中单击“**撤消**”，或将鼠标指针悬停在“**更改**”窗格中的更改上，并单击“**X**”将其移除。

如果将建议应用于转置字段(版本 2019.4.2 及更高版本)，将自动创建一个“转置”步骤，您可以在其中执行任何其他转置操作，如重命名转置字段或对其他字段进行转置。

3. 如果 Tableau Prep Builder 因为所做更改而确定了进一步的建议，则灯泡图标会一直显示在“配置”卡上，直至找不到进一步的建议为止。

重复上面的步骤以应用任何其他更改或忽略建议的更改，并使用 Tableau Prep Builder 中的其他清理工具来解决数据问题。



编辑字段值

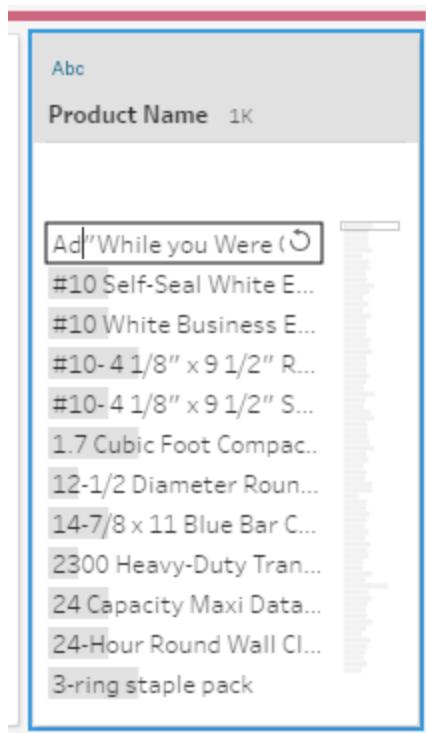
相同值的多个变体可能会让您无法准确地汇总数据。可以在 Tableau Prep Builder 中使用以下选项快捷轻松地纠正这些变化。

注意: 对值进行的任何编辑都必须与字段数据类型兼容。

编辑单个值

1. 在“配置”卡中，单击要编辑的值，并输入新值。值旁边将显示一个组图标 。

或者，右键单击值并单击“编辑值”。更改记录在屏幕左侧的“更改”窗格中。



2. 在“配置”窗格和数据网络中查看结果。

编辑多个值

可通过多个选项来一次性编辑多个值。例如，使用快速清理操作来移除字段中所有值的标点符号、使用多选手动将值分组、使用可找到类似值的模糊匹配算法自动将值分组在一起，或者选择多个值并将它们替换为 Null。

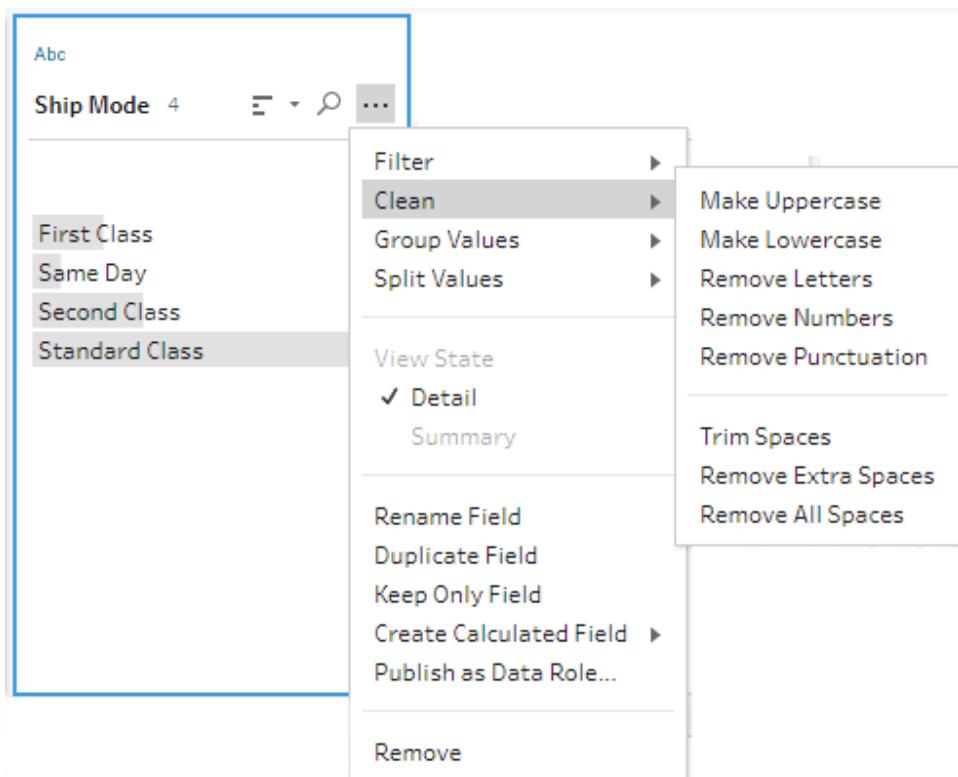
注意: 将多个值映射到单个值时，原始字段会在值旁边显示一个分组图标 ，显示哪些值分组在一起。

使用快速清理操作编辑多个值

此选项仅适用于文本字段。

1. 在“配置”窗格、“结果”窗格或数据网格中，选择要编辑的字段。
2. 单击“更多选项”…，选择“清理”，然后选择以下选项之一：

- **设为大写**: 将所有值更改为大写文本。
- **设为小写**: 将所有值更改为小写文本。
- **移除字母**: 移除所有字母并仅保留其他字符。
- **移除数字**: 移除所有数字并保留字母和其他字符。
- **移除标点符号**: 移除所有标点符号。
- **剪裁空格**: 移除前导和尾随空格。
- **移除额外的空格(版本 2018.3.3 及更高版本)**: 移除前导和尾随空格，以及将字符之间的额外空格替换为单一空格。
- **移除所有空格(版本 2018.3.3 及更高版本)**: 移除所有空格，包括前导和尾随空格以及字符之间的任何空格。



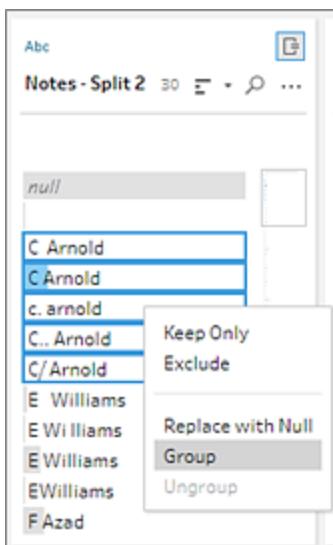
您可以累加操作，以将多个清理操作应用于字段。例如，首先选择“清理”>“移除数字”，然后选择“清理”>“移除标点符号”以从字段值中移除所有数字和标点符号。

3. 若要撤消更改，请单击“流程”窗格顶部的“撤消”箭头，或从更改列表中移除更改。

以内联方式编辑分组或多个值

使用此选项可手动选择多个值，并将它们分组在配置卡中的某个标准值下。若要使用其他方法对值进行分组，请参见[将多个值手动映射到标准值 对页](#)和[使用模糊匹配将值自动映射到标准值 在本页 248](#)。

1. 在“**配置**”卡中，选择要编辑的字段。
2. 按 **Ctrl** 或按住 **Shift** 并单击，或者按 **Command** 或按住 **Shift** 并单击 (MacOS)，并选择要分组的值。
3. 右键单击，并从上下文菜单中选择“**分组**”。您右键单击的所选内容中的值将成为新组的默认名称，但您可以对此名称进行内联编辑。



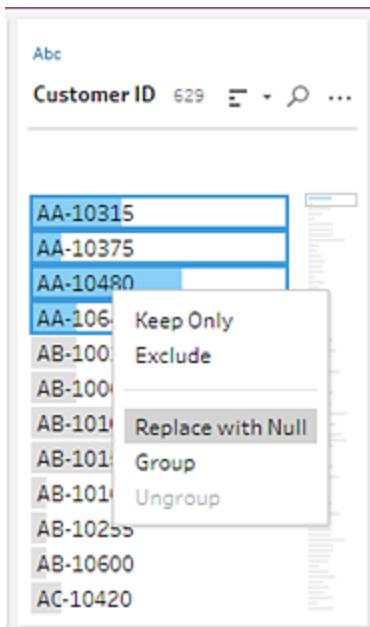
4. 若要编辑组名称，请选择分组的字段并编辑值，或者右键单击或按住 **Ctrl** 单击 (Mac) 分组的字段并从上下文菜单中选择“**编辑值**”。
5. 若要取消分组已分组的字段值，请右键单击分组的字段，并从上下文菜单中选择“**取消分组**”。

将一个或多个值替换为 Null

如果有要包括在分析中的数据行，但想要排除特定字段值，您可以将这些值更改为 **Null** 值。

1. 在“**配置**”卡中，按 **Ctrl** 或按住 **Shift** 并单击，或者按 **Command** 或按住 **Shift** 并单击 (Mac 上)，并选择要更改的值
2. 右键单击或按住 **Ctrl** 单击 (Mac)，并从菜单中选择“**替换为 Null**”。值将更改为 **Null**，

并且值旁边会显示分组图标 。



将多个值手动映射到标准值

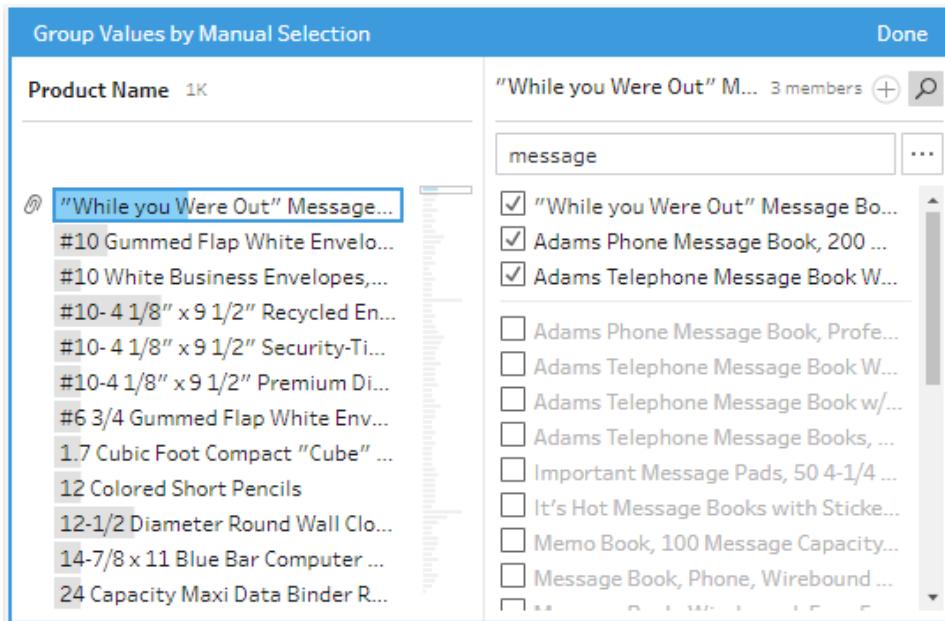
使用“对值进行分组”(以前的版本中为“分组和替换”)将字段的值从一个值映射到另一个值，或者手动选择多个值以对它们进行分组。您甚至可以添加新值，设置映射关系来组织数据。

例如，假设字段中有三个值：“My Company”、“My Company Incorporated”和“My Company Inc”。所有这些值都表示同一公司，即“My Company”。您可以使用“对值进行分组”将值“My Company Incorporated”和“My Company Inc”映射到“My Company”，以便所有三个值在字段中都显示为“My Company”。

将多个值映射到单个所选字段

1. 在“配置”窗格或“结果”窗格中，选择要编辑的字段。
2. 单击“更多选项”…，并从菜单中选择“对值进行分组”(以前的版本中为“分组和替换”) >“手动选择”。
3. 在“对值进行分组”编辑器的左侧窗格中，选择要用作分组值的字段值。此值现在显示在右侧窗格的顶部。

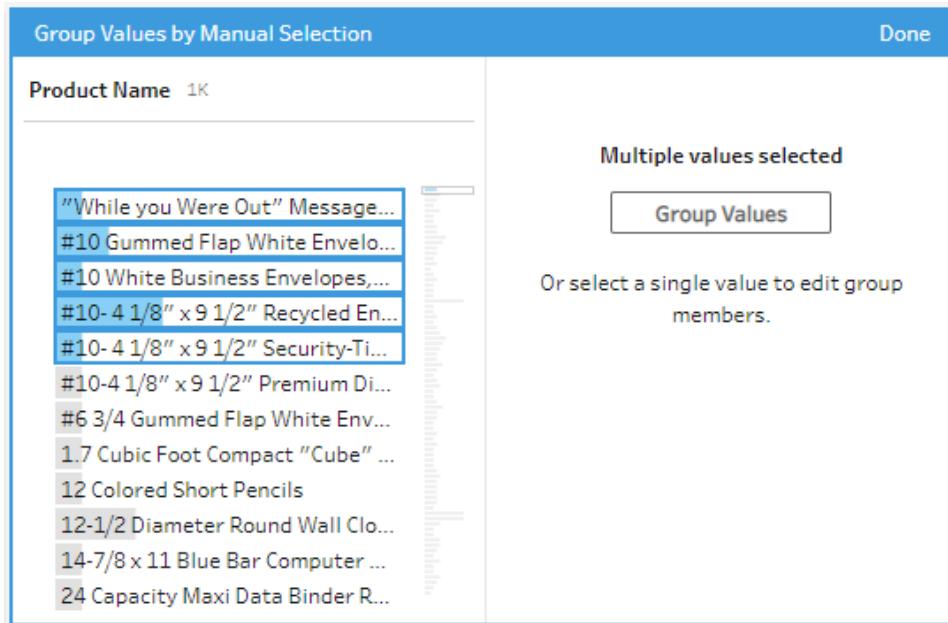
- 在“对值进行分组”编辑器右侧窗格的下半部分中，选择要添加到组的值。



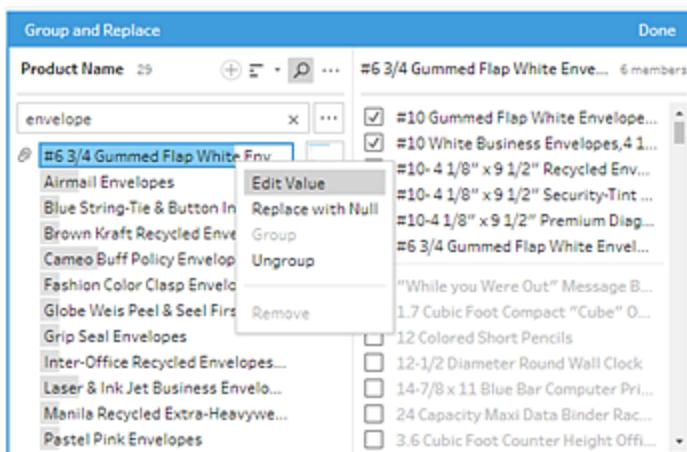
若要从组中移除值，请在“对值进行分组”编辑器右侧窗格的上半部分中，清除值旁边的复选框。

通过选择多个值来创建组

- 在“配置”窗格或“结果”窗格中，选择要编辑的字段。
- 单击“更多选项”…，并从菜单中选择“对值进行分组”(以前的版本中为“分组和替换”) >“手动选择”。
- 在“对值进行分组”编辑器的左侧窗格中，选择要分组的多个值。
- 在“对值进行分组”编辑器的右侧窗格中，单击“对值进行分组”。



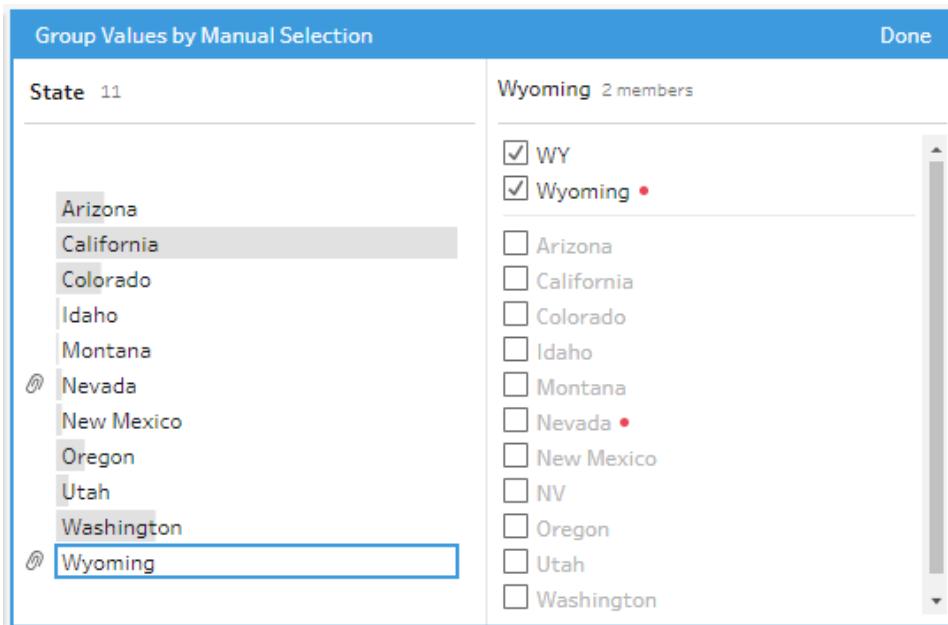
将会使用最后一个所选的值作为组名称创建一个新组。若要编辑组名称，请选择分组的字段并编辑值，或者右键单击或按住 Ctrl 单击 (MacOS) 分组的字段并从菜单中选择“**编辑值**”。



添加和标识不在数据集中的值

如果要将数据集中的值映射到不存在的新值，您可以使用“**对值进行分组**”(以前的版本中为“**分组和替换**”)来添加该值。为了轻松识别数据集中没有的任何值，这些值在“**对值进行分组**”编辑器中的值名称旁边带有红点标记。

例如在下图中，Wyoming 和 Nevada 不在数据集中。

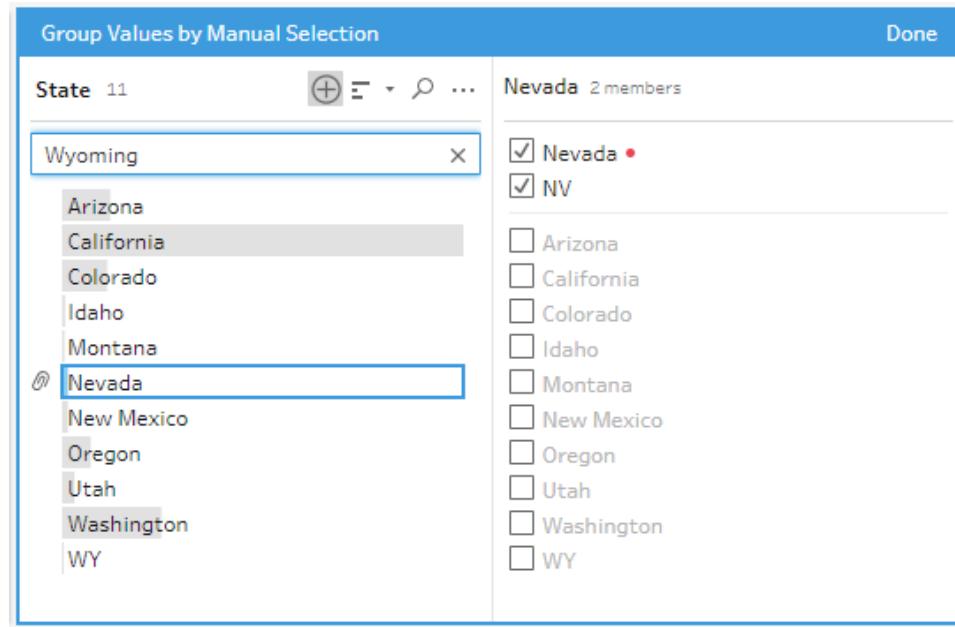


数据集中之所以可能没有某个值，可能有以下一些原因：

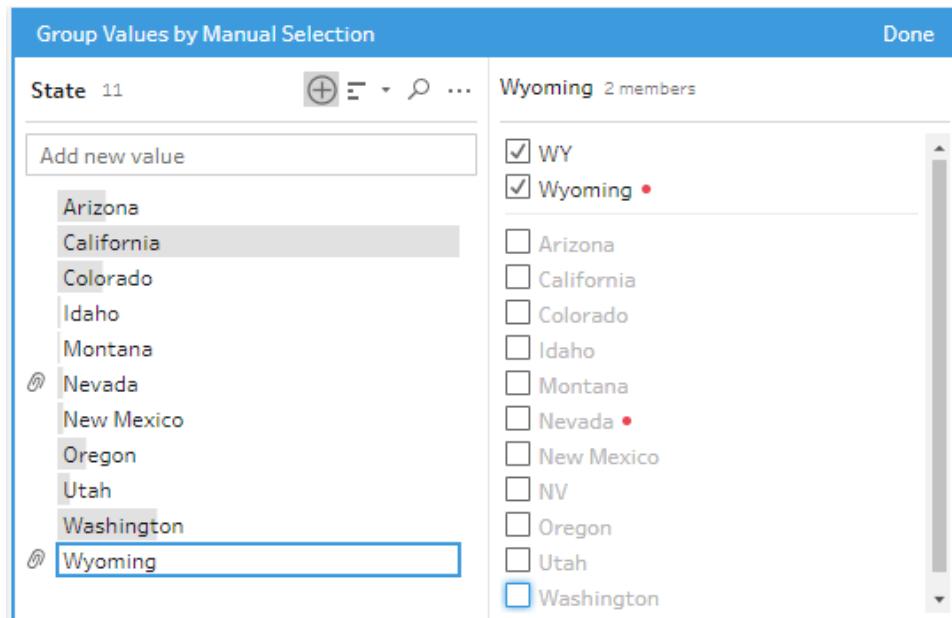
- 您刚刚手动添加了新值。
- 值不再在数据中。
- 值在数据中，但不在抽样数据集中。

添加新值：

1. 在“配置”窗格或“结果”窗格中，选择要编辑的字段。
2. 单击“更多选项”…，并从上下文菜单中选择“对值进行分组”(以前的版本中为“分组和替换”) >“手动选择”。
3. 在“对值进行分组”编辑器的左侧窗格中，单击加号 以添加新值。
4. 在字段中键入新值，并按 Enter 添加该值。



5. 在右侧窗格中，选择要映射到新值的值。



6. (可选) 若要向映射的值中添加其他新值，请单击“对值进行分组”编辑器右侧窗格中的加号 按钮。

使用模糊匹配将值自动映射到标准值

若要搜索类似的值并对其进行自动分组，请使用模糊匹配算法之一。字段值依据最频繁出现的值进行分组。查看分组的值，并根据需要在组中添加或移除值。

如果使用数据角色来验证字段值，则可以使用“**对值进行分组**”(以前的版本中为“**分组和替换**”)选项将无效值与有效值匹配。有关详细信息，请参见[按数据角色对类似值进行分组 在本页219](#)。

选择以下选项之一对值进行分组：

- **发音**: 查找发音类似的值并对其进行分组。此选项使用 **Metaphone 3** 算法，该算法按单词的发音对单词编制索引，最适合于英语单词。许多常用的拼写检查器都使用这种类型的算法。此选项对于数据角色不可用。
- **常用字符**: 查找具有共同字母或数字的值并对其进行分组。此选项使用 **ngram** 指纹算法，该算法按单词移除标点符号、重复字符和空格之后的唯一字符对单词编制索引。此算法适合于任何支持的语言。此选项对于数据角色不可用。

例如，此算法将匹配表示为“John Smith”和“Smith, John”的名称，因为它们均生成键“hijmnost”。由于此算法不考虑发音，因此值“Tom Jhinois”将具有相同的键“hijmnost”，并也会包括在组中。

- **拼写**: 查找拼写类似的值并对其进行分组。此选项使用 **Levenshtein** 距离算法，使用固定默认阈值来计算两个文本值之间的编辑距离。当编辑距离小于阈值时，则将这两个值分组在一起。此算法适合于任何支持的语言。

从 Tableau Prep Builder 版本 2019.2.3 开始，此选项可在应用了数据角色之后使用。在这种情况下，它会使用编辑距离将无效值与最近的有效值匹配。如果数据集示例中没有标准值，Tableau Prep Builder 会自动添加该值，并将值标记为不在原始数据集中。

- **发音 + 拼写**:(版本 2019.1.4 及更高版本) 如果为字段分配数据角色，您可以使用该数据角色，通过数据角色定义的标准值来匹配值和对值进行分组。此选项随后将根据拼写和发音将无效值与最相似的有效值匹配。如果数据集示例中没有标准值，Tableau Prep Builder 会自动添加该值，并将值标记为不在原始数据集中。此选项最适合于英文单词。

有关详细信息，请参见[清理和调整数据 在本页223](#)。

注意:在 Tableau Prep Builder 版本 2019.1.4 和 2019.2.1 中,此选项标为“**数据角色匹配**”。

使用模糊匹配分组类似值

1. 在“配置”窗格或“结果”窗格中,选择要编辑的字段。
2. 单击“更多选项”… 并选择“对值进行分组”,然后选择以下选项之一:
 - 发音
 - 常用字符
 - 拼写

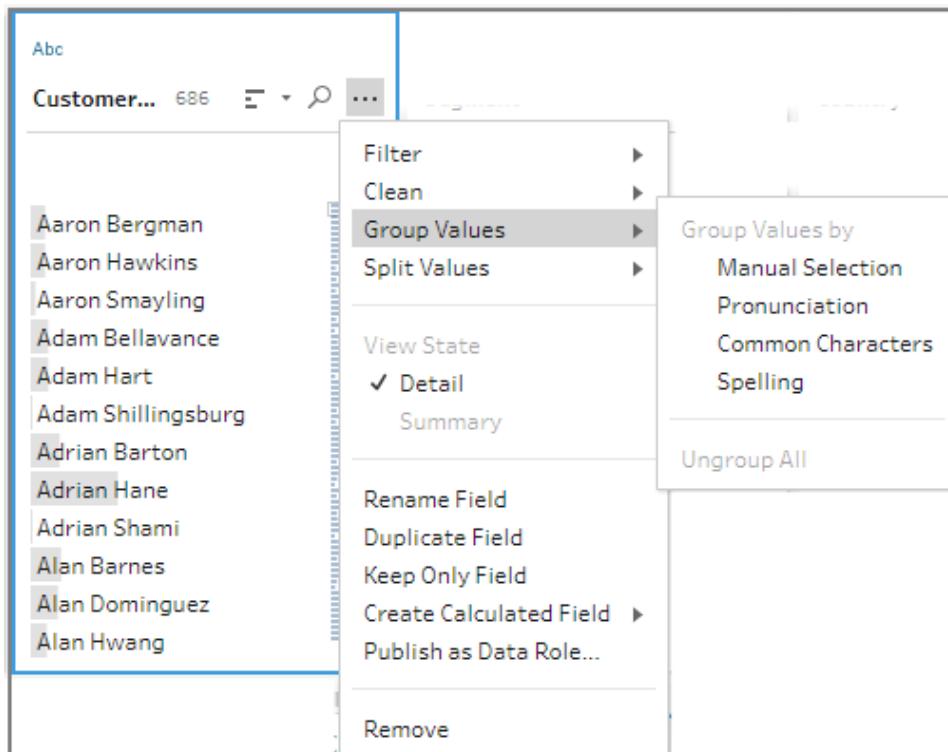
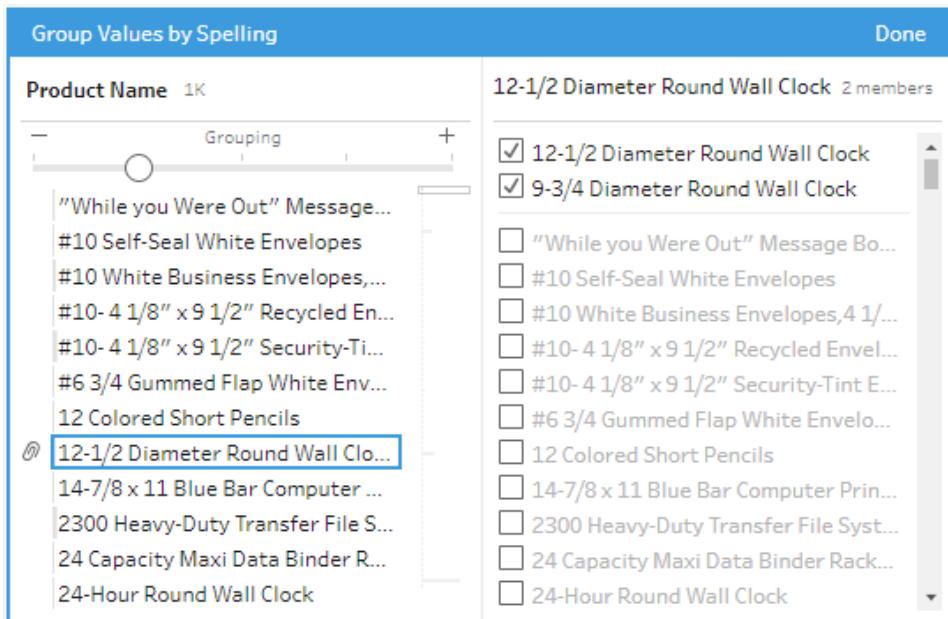


Tableau Prep Builder 将查找匹配的值并对其进行分组,并将它们替换为组中出现最频繁的值。

- 查看分组，并根据需要手动添加或移除值或对其进行编辑。然后单击“完成”。



在对字段值进行分组时调整结果

如果按“**拼写**”或“**发音**”对类似值进行分组，您可以使用字段上的滑块来调整分组参数的严格程度，从而更改结果。

根据滑块的设置方式，您可以更大程度地控制组中包括的值数量以及创建的组数量。默认情况下，Tableau Prep Builder 会检测最佳分组设置，并将滑块显示在该位置。

更改阈值时，Tableau Prep Builder 会对值的样本进行分析来确定新的分组。依据设置生成的组保存并记录在“**更改**”窗格中，但不会保存阈值设置。下次通过编辑现有更改或进行新更改打开“**对值进行分组**”编辑器时，阈值滑块会显示在默认位置，使您能够根据当前数据集进行任何调整。

- 在“**配置**”窗格或“**结果**”窗格中，选择要编辑的字段。
- 单击“**更多选项**”…，并选择“**对值进行分组**”(在前的版本中为“**分组和替换**”)，然后选择以下选项之一：
 - **发音**
 - **拼写**

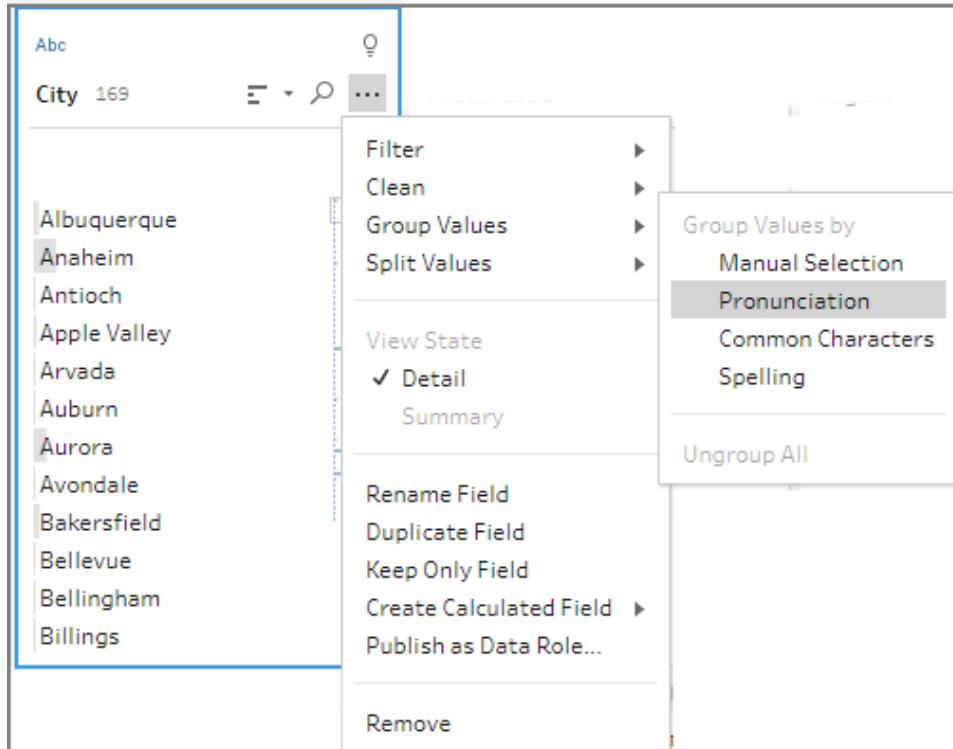
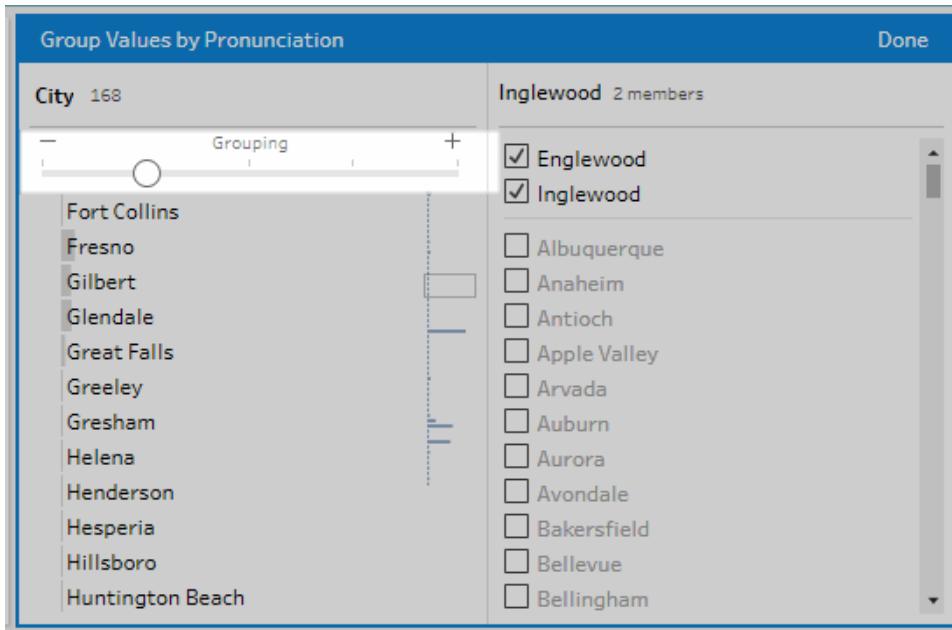


Tableau Prep Builder 将查找匹配的值并对其进行分组，并将它们替换为组中出现最频繁的值。

3. 在“对值进行分组”编辑器的左侧窗格中，将滑块拖到 5 个阈值级别之一以更改结果。



若要设置更严格的阈值，请将滑块向左移动。这会使匹配数更少，并创建更少的组。

若要设置更宽松的阈值，请将滑块向右移动。这会使匹配数更多，并创建更多的组。

4. 单击“完成”保存您所做的更改。

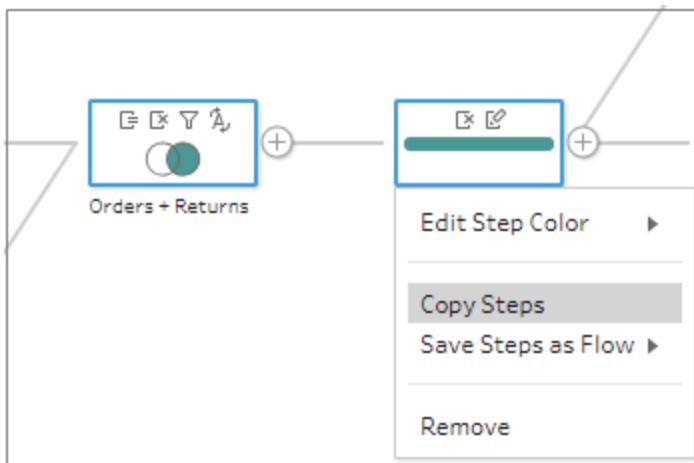
复制步骤、操作和字段

在清理数据时，您通常会在整个流程中反复执行相同的清理操作或动作。为了使数据的清理和调整更加高效，您可以在整个流程中复制和粘贴这些操作或动作，或者甚至复制选定的步骤并保存它们，这样您就能执行清理操作或动作一次，然后在需要的地方重用该操作或动作。您甚至可以复制字段来体验不同的清理操作。

复制并粘贴步骤

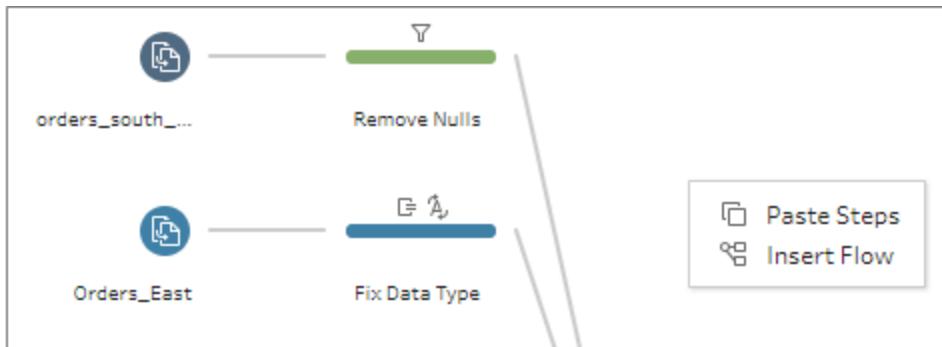
复制一个或多个步骤以在同一流程的另一个区域中使用它们。此选项不可用于包括通配符并集的输入步骤。

1. 在“流程”窗格中，选择流程中的一个或多个步骤。
2. 右键单击或按住 Ctrl 单击 (MacOS) 选定的步骤，然后选择“复制”。



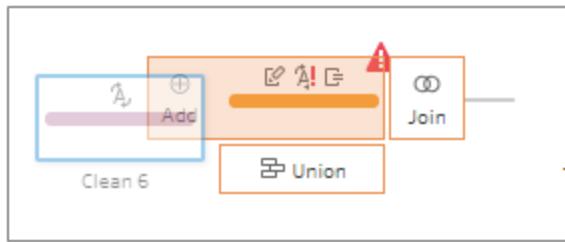
3. 若要粘贴复制的步骤，请执行下列操作之一：

- **版本 2019.4.2 及更高版本:** 将鼠标悬停在步骤或流程行上，直到出现加号 $(+)$ 图标，然后单击图标并选择“插入流程”
- **所有版本:** 在画布的任何空白处右键单击或按住 Ctrl 单击 (MacOS)，并单击“粘贴”或“粘贴步骤”。



4. 将步骤拖放到流程中要放置步骤的位置。如果将步骤添加到流程步骤的末尾，则这些步骤将自动添加到步骤的末尾。如果在现有流程步骤之间插入步骤。将步骤移到流程中所需的位置并修复任何错误。

如果需要，您可以移除流程行或四处移动步骤。例如，若要将某个步骤连接到复制的步骤，请移除现有流程行(如果有)，然后将现有步骤拖到新步骤，并放在“添加”上。



有关组织流程的详细信息，请参见[重新组织流程的布局\(版本 2019.2.2 及更高版本\)](#)在本页 190。

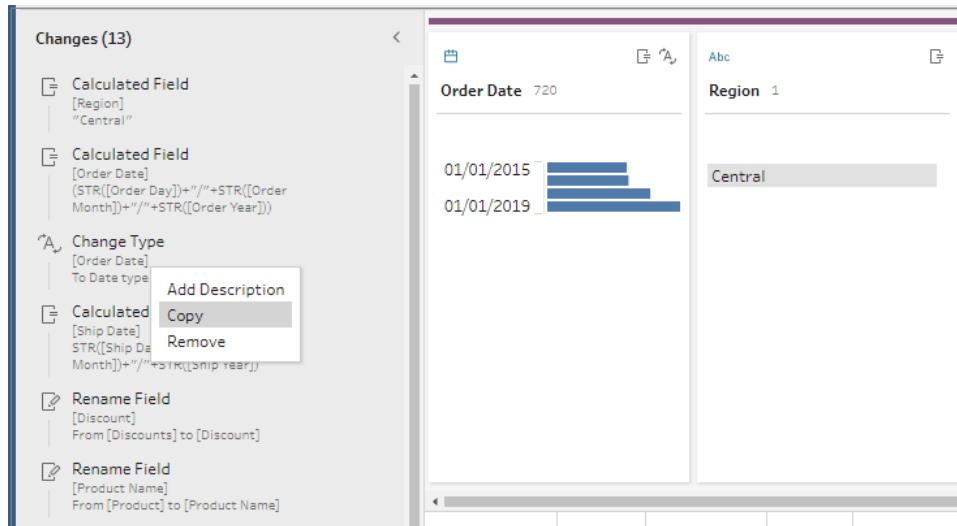
复制并粘贴清理操作

您可以在同一流程中复制和粘贴清理操作，以便通过以下选项之一来重用操作：

- 从一个步骤的“更改”窗格中复制操作，然后将其粘贴在同一步骤或另一个步骤的“更改”窗格中，从而在该步骤中应用该相同操作。
- 将一个操作从“更改”窗格拖放到“配置”窗格中的其他字段，以将该步骤应用于多个字段。此选项不可用于影响多个字段(例如计算字段)的操作。

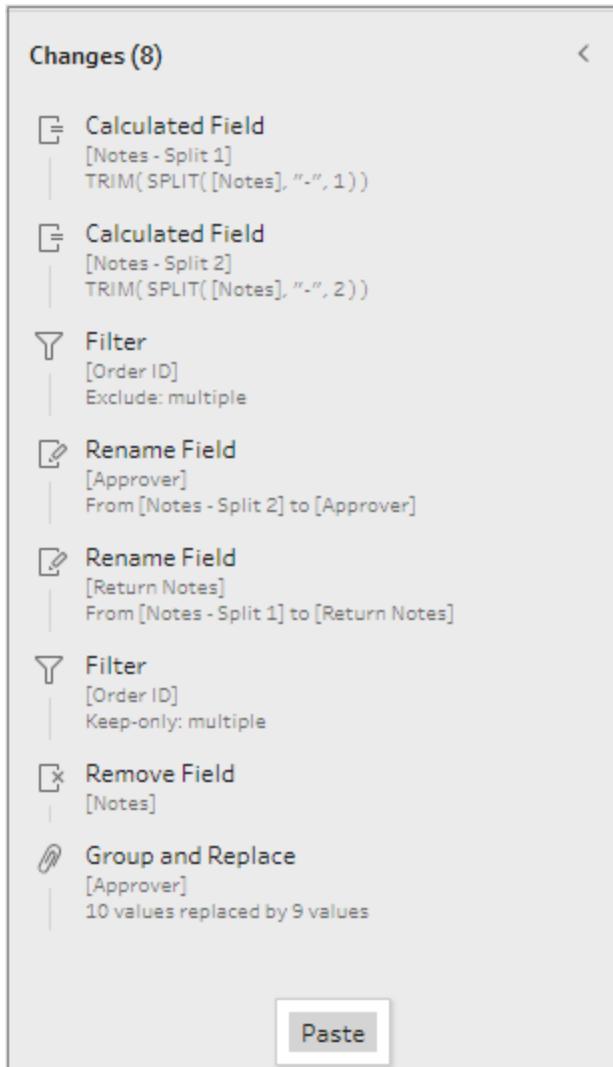
若要将步骤中的更改复制并粘贴到同一步骤或其他步骤，请执行以下操作：

1. 在“更改”窗格中，选择要复制的更改。
2. 在更改项上右键单击或按住 Ctrl 并单击 (MacOS)，然后从菜单中选择“**复制**”。



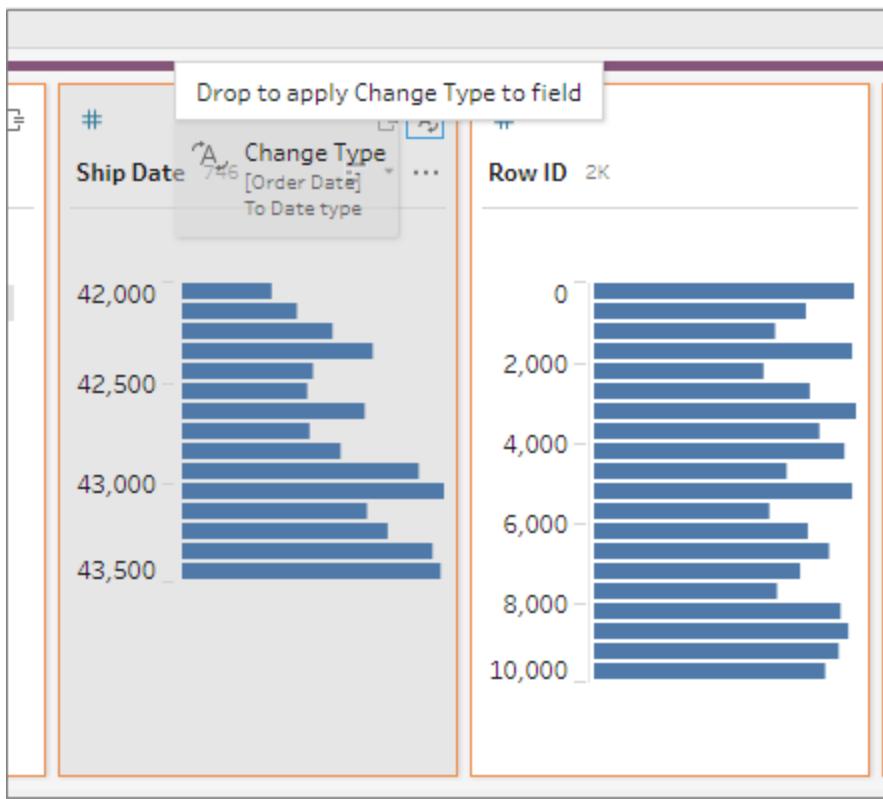
3. 在要在其中粘贴更改的“更改”窗格中，右键单击或按住 Ctrl 并单击 (MacOS)，并选择

“粘贴”。选择更改，并根据需要单击“编辑”进行任何调整。



若要将更改拖放到步骤中的其他字段，请执行以下操作：

1. 在“更改”窗格中，选择要复制的更改。
2. 将更改拖放到要应用更改的字段上。根据需要重复此操作。

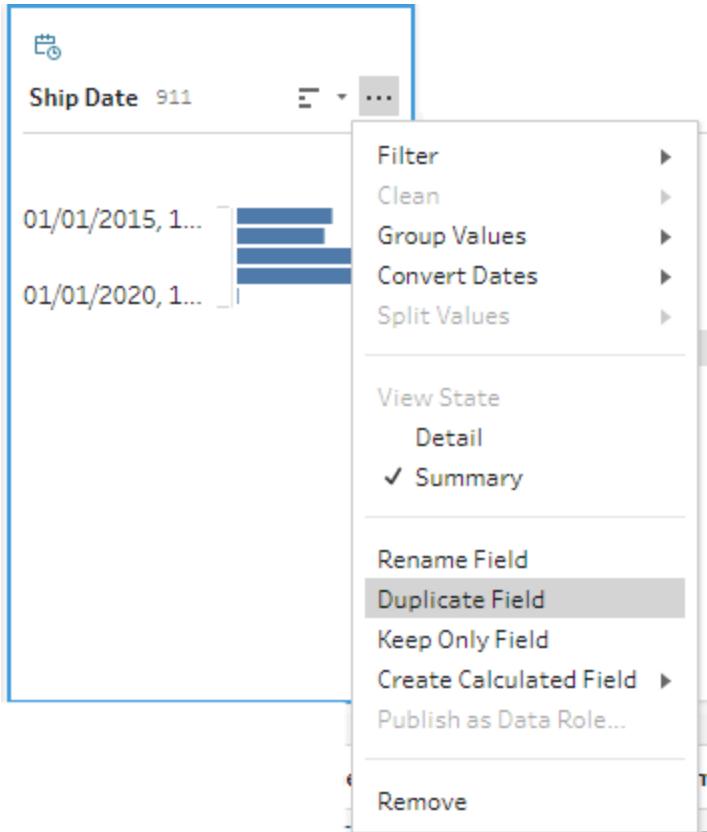


复制字段(版本 2019.2.3 及更高版本)

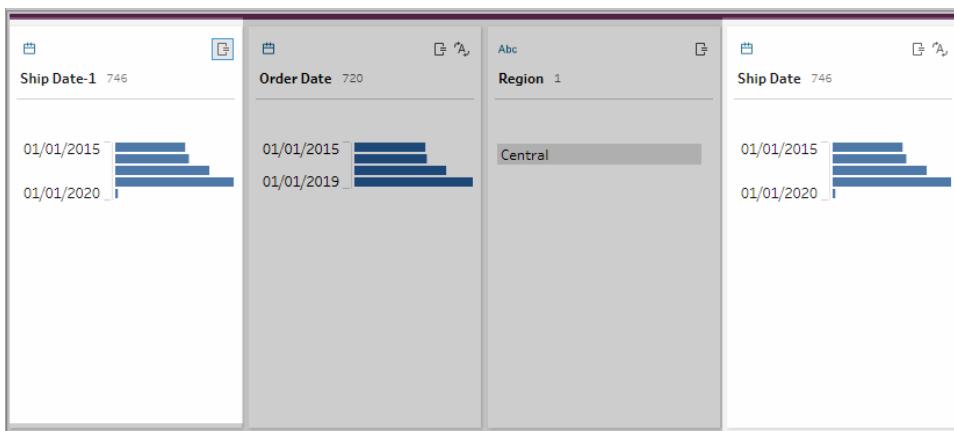
如果想要在一个字段上试验清理操作但不想更改原始数据，则可以复制字段。

Tableau Prep 帮助

1. 在“配置”窗格、数据网格、“结果”窗格或列表视图中，选择要复制的字段。
2. 从“更多选项”菜单中，选择“复制字段”。



系统将使用相同的名称和修饰符创建一个新字段。例如，“Ship Date -1”(发货日期 -1)。

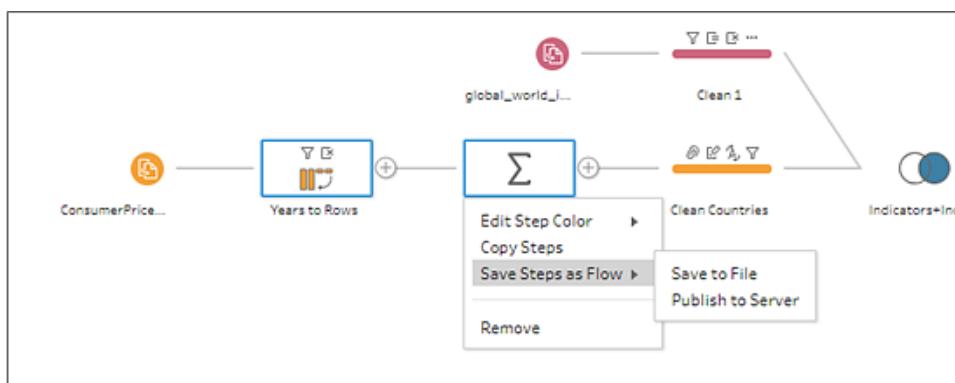


创建可重用的流程步骤(版本 2019.3.2 及更高版本)

如果您通常反复地对数据执行相同的操作，并且希望在其他流程中应用这些相同的步骤，则可以选择一个或多个流程步骤及其关联的操作或整个流程，并将其以本地方式保存到计算机上，或将其发布到 Tableau Server 或 Tableau Online 与他人分享。将流程步骤发布到服务器时，将自动添加“已保存的步骤”标记，以便在将这些步骤添加到流程中时可以轻松地搜索和找到它们。

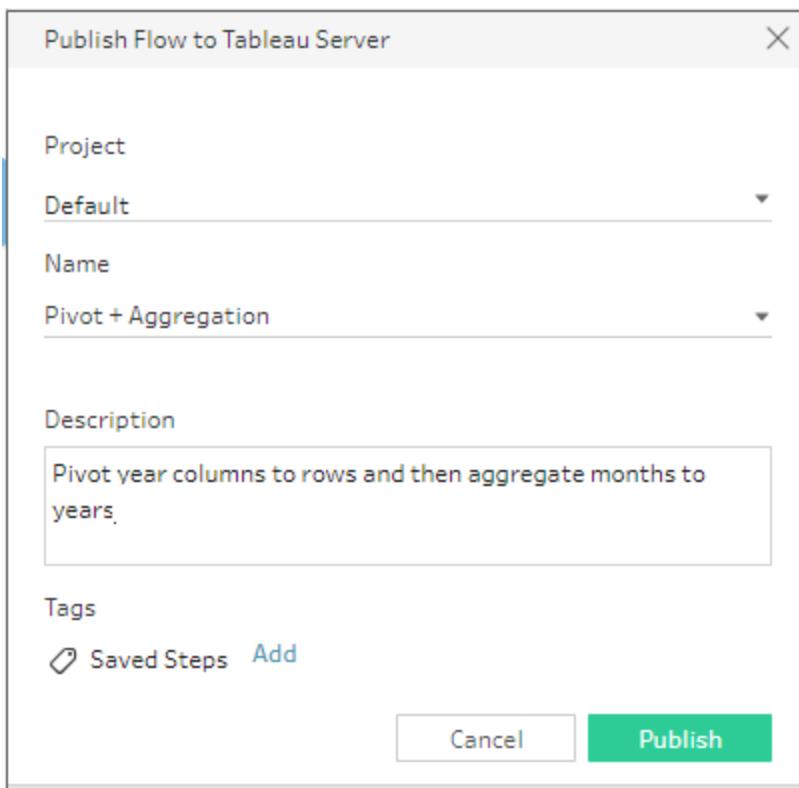
创建可重用的步骤

1. 选择一个或多个步骤。
2. 右键单击或按住 Ctrl 单击 (MacOS) 选定步骤，并选择“将步骤另存为流程”。



3. 选择“保存到文件”以本地方式保存流程，或选择“发布到服务器”以将流程发布到 Tableau Server 或 Tableau Online。
4. 如果将流程发布到 Tableau Server 或 Tableau Online，请登录到服务器(如果需要)，

然后填写“发布流程”对话框中的字段，再单击“发布”。

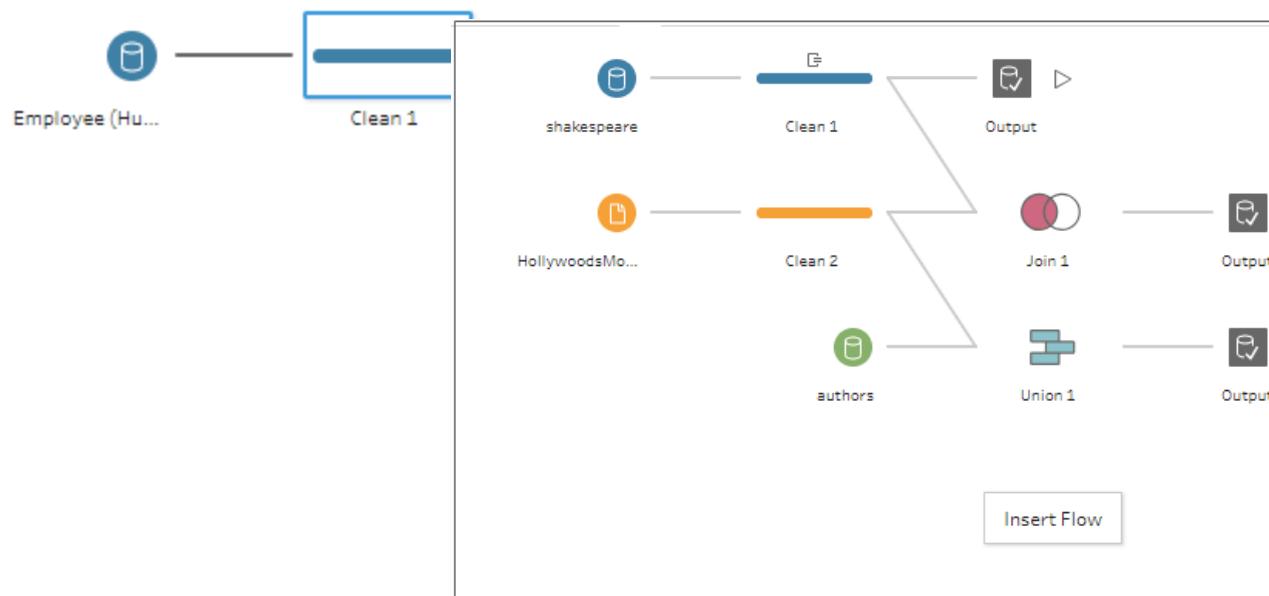


在流程中插入可重用的步骤

1. 打开一个流程。
2. 在流程窗格中，执行以下操作之一：
 - 版本 2019.4.2 及更高版本：将鼠标悬停在步骤或流程行上，直到出现加号⁽⁺⁾图标，然后单击图标并选择“插入流程”
 - 版本 2019.3.2 及更高版本：在画布的白色区域中，右键单击或按住 Ctrl 单击(MacOS)，然后单击“插入流程”，或者从顶部菜单中单击“编辑”>“插入流程”。

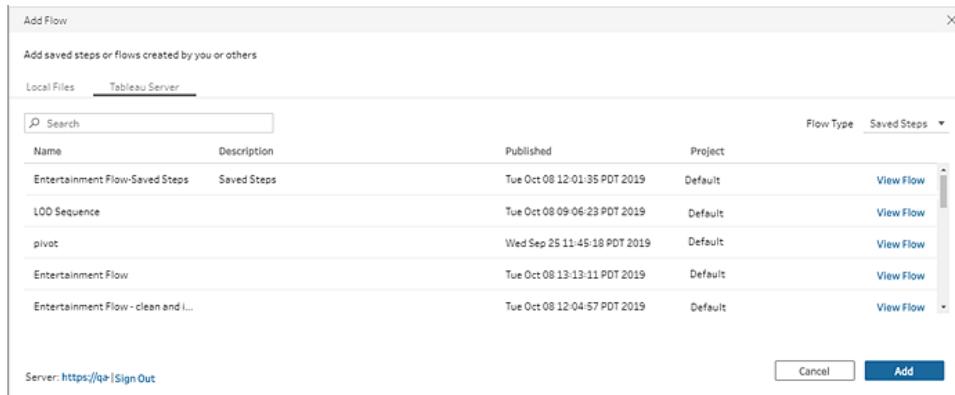
“流程步骤”菜单(版本 2019.4.2 及
更高版本)

“画布”菜单(版本 2019.3.2 及更高版本)

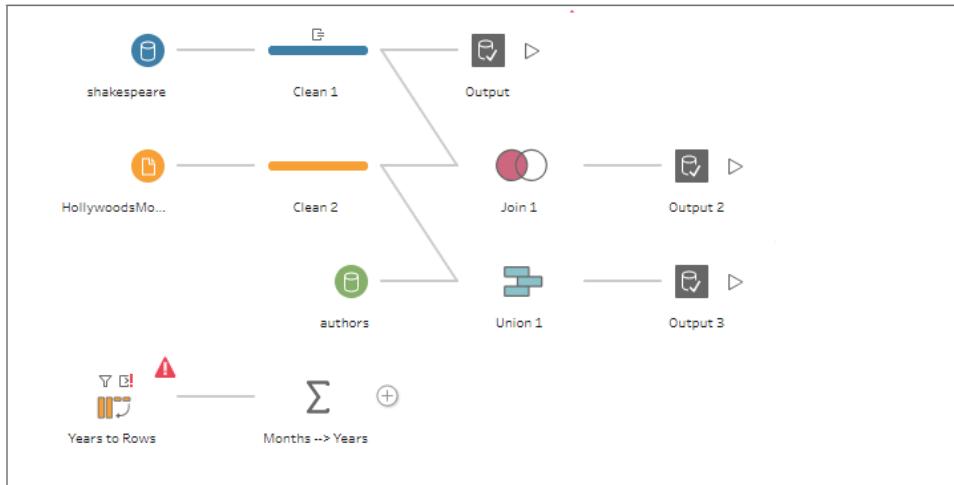


3. 在“添加流程”对话框中，从保存到本地文件或服务器的流程中进行选择，然后单击“添加”。系统将自动筛选流程列表，以显示带有“已保存的步骤”标记的流程。若要插入其他流程，请将“流程类型”更改为“所有流程”。

在版本 2019.4.2 中，可以单击“查看流程”以打开并查看已登录服务器中的已发布流程。



4. 流程已添加到流程窗格中。如果将流程添加到流程步骤的末尾，则流程步骤将自动添加到步骤的末尾。如果在现有流程步骤之间插入流程步骤。将步骤移到流程中所需的位置并修复任何错误。



创建详细级别和排名计算

可以使用计算字段，通过数据源中已存在的数据来创建新数据。Tableau Prep Builder 支持的许多计算类型与 Tableau Desktop 相同。有关创建计算的一般信息，请参见 [Tableau 中的计算入门指南](#)。

注意: Tableau Desktop 中支持的某些函数在 Tableau Prep Builder 中可能不受支持。
若要查看可用于 Tableau Prep Builder 的函数，请查看计算编辑器中的函数列表。

从 Tableau Prep Builder 版本 2020.1.3 开始，您可以使用 FIXED 详细级别 (LOD) 以及 RANK 和 ROW_NUMBER 分析函数来执行更复杂的计算。

例如，添加 FIXED LOD 计算来更改表中字段的粒度、使用新的 ROW_NUMBER() 分析函数快速查找重复行，或使用新的 RANK() 函数之一来查找具有类似数据的所选行的前 N 个或后 N 个值。如果希望在构建这些类型的表达式时获得更多指引，您可以使用新的可视化计算编辑器。

计算详细级别

当您需要在同一表中以多个粒度级别计算数据时，可以编写详细级别 (LOD) 表达式来执行此操作。举例来说，如果要查找每个区域的总销售额，可以编写类似于 {FIXED [Region] : SUM ([Sales])} 的计算。

Tableau Prep Builder 支持 FIXED 详细级别表达式，并使用语法 {FIXED [Field1], [Field2] : Aggregation ([Field])}。

LOD 表达式有两个用冒号分隔的等式部分。

- **FIXED [字段](必需)**: 这是您要为其计算值的一个或多个字段。举例来说, 如果要查找客户和区域的总销售额, 请输入 `FIXED [Customer ID], [Region] :`。如果未选择任何字段, 这相当于执行在冒号右侧定义的聚合, 并为每一行重复该值。
- **聚合([字段])(必需)**: 选择要计算的内容以及所需的聚合级别。举例来说, 如果要查找总销售额, 请输入 `SUM([Sales])`。

在 Tableau Prep Builder 中使用此功能时, 以下要求适用:

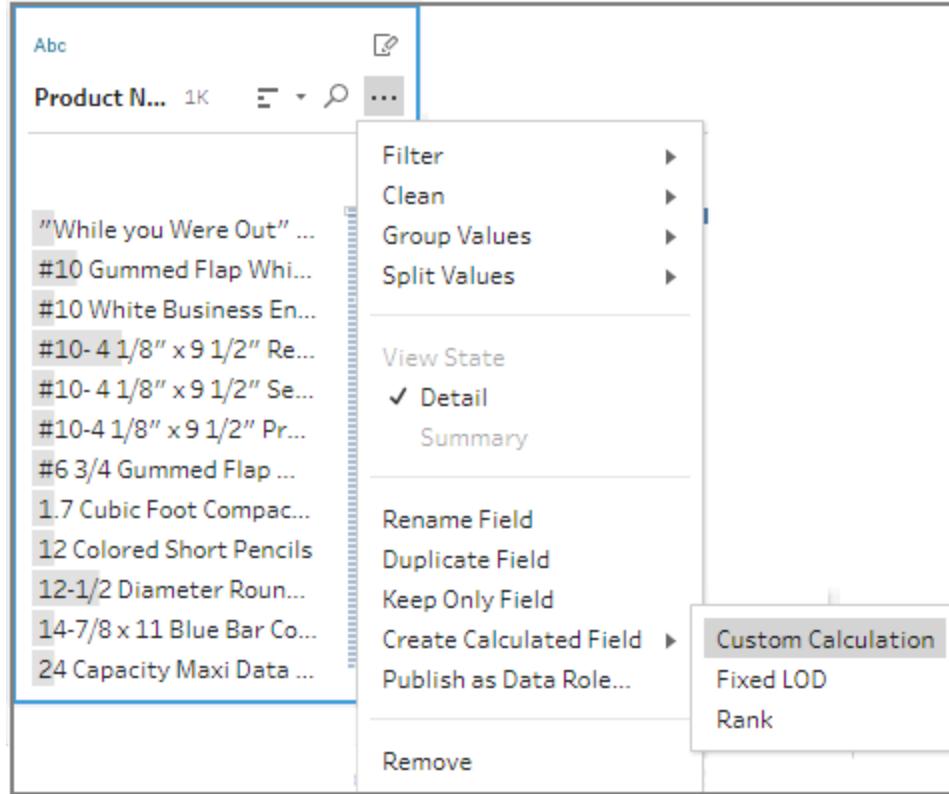
- 不支持 `INCLUDE` 和 `EXCLUDE` LOD 表达式。
- 仅支持在 LOD 表达式内部使用聚合计算。例如, `SUM([Sales])` 将无效, 但 `{FIXED [Region] : SUM([Sales])}` 有效。
- 不支持在 LOD 表达式内嵌套表达式。例如, 不支持 `{ FIXED [Region] : AVG([Sales]) / SUM([Profit]) }`。
- 不支持将 LOD 表达式与其他表达式结合使用。例如, `[Sales]/{ FIXED [Country / Region]:SUM([Sales])}` 不受支持。

创建详细级别 (LOD) 计算

若要创建详细级别计算, 可以使用计算编辑器自行编写计算, 或者, 如果您希望获得更多指引, 则可以使用可视化计算编辑器, 您可以在其中选择字段, Tableau Prep Builder 将为您编写计算。

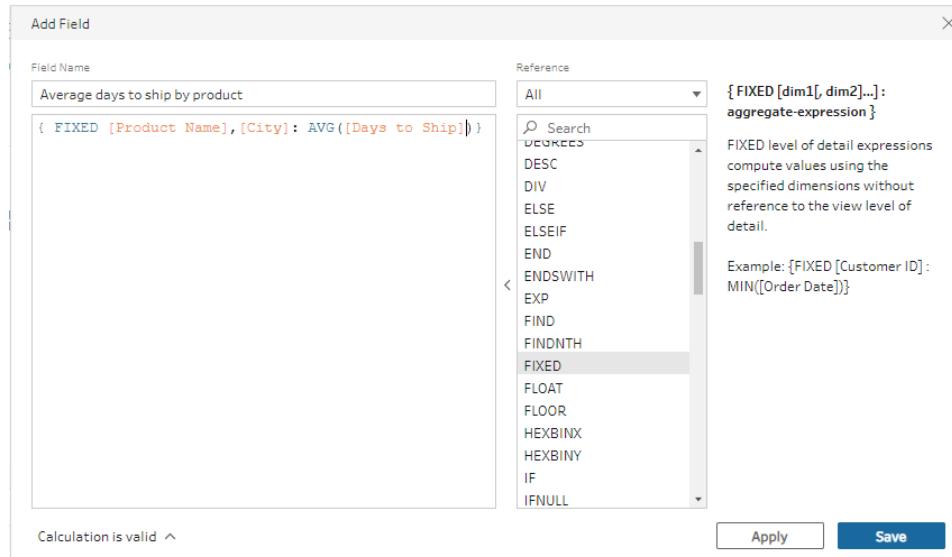
计算编辑器

1. 在“配置”窗格工具栏中单击“**创建计算字段**”, 或在配置卡或数据网格中单击“**更多选项**”… 菜单, 并选择“**创建计算字段**”>“**自定义计算**”。



2. 在计算编辑器中，输入计算的名称，并输入表达式。

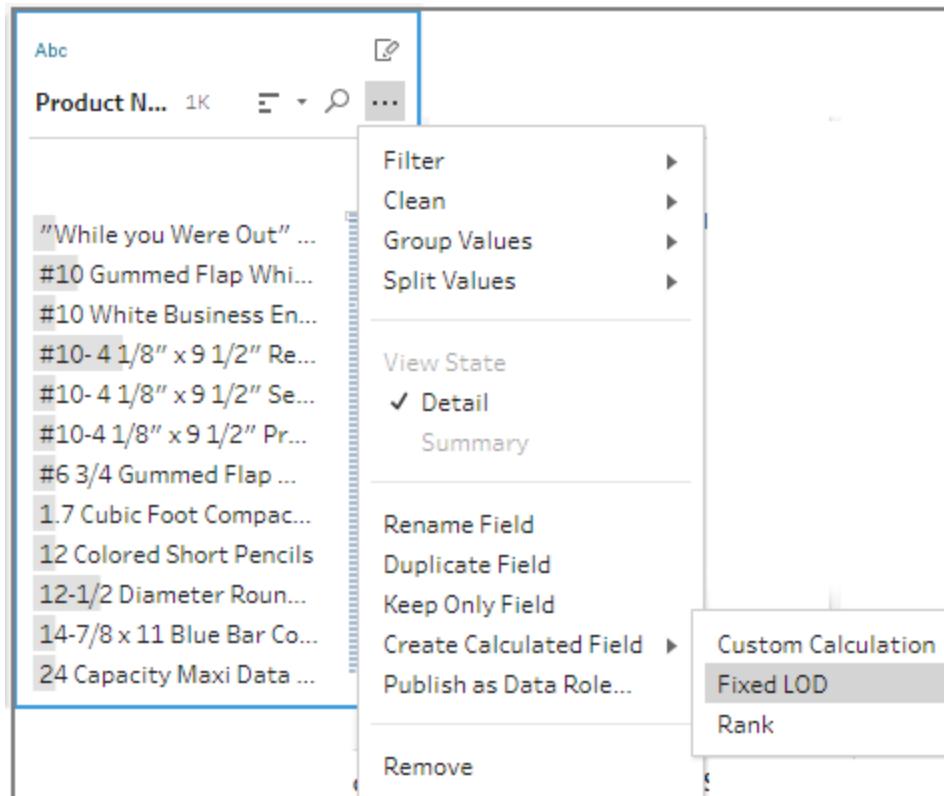
例如，若要按城市查找产品平均发货天数，请创建如下所示的计算。



可视化计算编辑器

从列表中选择字段，Tableau Prep Builder 将在您进行选择时为您构建计算。结果的预览显示在左侧窗格中，这样您就能够随时查看所选内容的结果。

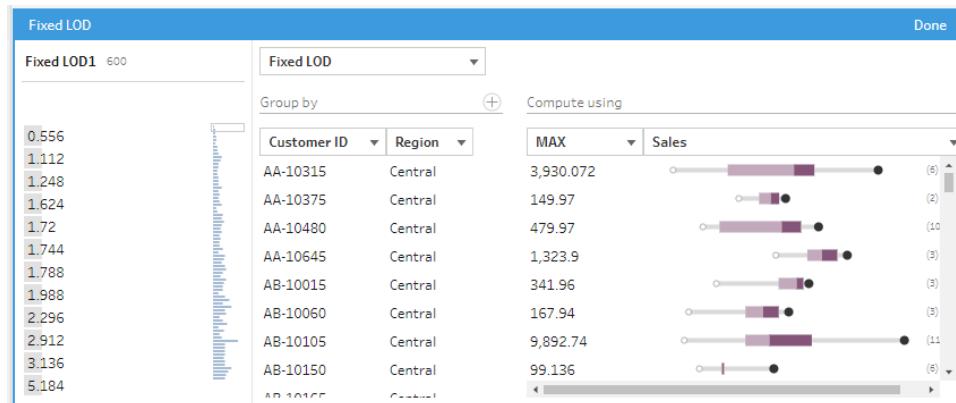
1. 在配置卡或结果窗格中，单击“更多选项”… 菜单并选择“创建计算字段”>“**FIXED LOD**”。



2. 在可视化计算编辑器中，执行以下操作：

- 在“分组依据”部分，选择要为其计算值的字段。默认情况下，将添加您在其中选择了“创建计算字段”>“**FIXED LOD**”菜单选项的字段。单击加号 图标以向计算中添加任何其他字段。这将填充等式的左侧，即 `{FIXED [Field1], [Field2] :}`。
- 在“计算依据”部分，选择要用于计算新值的字段。然后选择您的聚合。这将填充等式的右侧，即 `Aggregation([Field])`。

字段下方的图形显示值的分布和每个值组合的总计数。根据数据类型，这可以是盒形图、值范围或实际值。



注意: 可用聚合值因分配给字段的数据类型而异。

- 若要移除字段，请在“**分组依据**”部分的字段下拉框中右键单击或按住**Cmd**单击(MacOS)，然后选择“**移除字段**”。
- 在左侧窗格中，双击字段标题并输入用于计算的名称。

- 单击“**完成**”以添加新的计算字段。在“**更改**”窗格中，您可以看到 Tableau Prep Builder 生成的计算。单击“**编辑**”打开可视化计算编辑器以进行任何更改。



计算排名或行号

分析函数(有时称为窗口计算)使您能够对整个表或数据集中的所选行(分区)执行计算。例如，在将排名应用于所选行时，可以使用以下计算语法：

```
{PARTITION [field]: {ORDERBY [field]: RANK()}}
```

- PARTITION(可选):** 指定要对其执行计算的行。可以指定多个字段，但如果要使用整个表，请省略函数的此部分，Tableau Prep Builder 会将所有行视为分区。例如，
`{ORDERBY [Sales] : RANK() }.`
- ORDERBY(必需):** 指定要用于生成排名序列的一个或多个字段。

- **Rank ()(必需)**: 指定要计算的排名类型或 ROW_NUMBER()。Tableau Prep Builder 支持 RANK()、RANK_DENSE()、RANK_MODIFIED()、RANK_PERCENTILE() 和 ROW_NUMBER() 函数。
- **DESC 或 ASC(可选)**: 表示降序 (DESC) 或升序 (ASC) 顺序。默认情况下，排名按降序排序，因此无需在表达式中指定此项。如果要更改排序顺序，请向表达式中添加 **ASC**。也可以在函数中同时包括这两个选项。举例来说，如果要对所选行进行排名，但希望按升序对行进行排序，然后按降序应用排名，则应在表达式中包括这两个选项。例如：{PARTITION [Country], [State]: {ORDERBY [Sales] ASC, [CUSTOMER] DESC: RANK()}}

在 Tableau Prep Builder 中使用此功能时，以下要求适用：

- 不支持在 RANK() 函数中嵌套表达式。例如，不支持 [Sales]/ {PARTITION [Country]: {ORDERBY [Sales]: RANK()}} / SUM([Profit])。
- 不支持将 RANK() 函数与其他表达式结合使用。例如，[Sales]/ {PARTITION [Country]: {ORDERBY [Sales]: RANK()}} 不受支持。

支持的分析函数

函数	说明	示例																								
RANK()	从 1 开始按升序或降序为每一行分配整数排名。如果行具有相同的值，则它们共享分配给该值的第一个实例的排名。在计算下一行的排名时，将添加具有相同排名的行数，因此您可能无法获得连续排名值。	<table border="1"> <thead> <tr> <th>Sales Person</th><th>Region</th><th>Commission</th><th>Rank</th></tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td><td>Central</td><td>7000</td><td>1</td></tr> <tr> <td>Lisa Martin</td><td>Central</td><td>15000</td><td>2</td></tr> <tr> <td>Sarah Scott</td><td>Central</td><td>15000</td><td>2</td></tr> <tr> <td>Barbara Davis</td><td>East</td><td>13000</td><td>4</td></tr> <tr> <td>Charles Lee</td><td>East</td><td>9000</td><td>5</td></tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	2	Sarah Scott	Central	15000	2	Barbara Davis	East	13000	4	Charles Lee	East	9000	5
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	2																							
Sarah Scott	Central	15000	2																							
Barbara Davis	East	13000	4																							
Charles Lee	East	9000	5																							
RANK_DENSE()	从 1 开始按升序或降序为每一行分配整数排名。如果行具有相同的值，则它们共享分配给该值的第一个实例的排名，但不会	<table border="1"> <thead> <tr> <th>Sales Person</th><th>Region</th><th>Commission</th><th>Rank</th></tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td><td>Central</td><td>7000</td><td>1</td></tr> <tr> <td>Lisa Martin</td><td>Central</td><td>15000</td><td>2</td></tr> <tr> <td>Sarah Scott</td><td>Central</td><td>15000</td><td>2</td></tr> <tr> <td>Barbara Davis</td><td>East</td><td>13000</td><td>3</td></tr> <tr> <td>Charles Lee</td><td>East</td><td>9000</td><td>4</td></tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	2	Sarah Scott	Central	15000	2	Barbara Davis	East	13000	3	Charles Lee	East	9000	4
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	2																							
Sarah Scott	Central	15000	2																							
Barbara Davis	East	13000	3																							
Charles Lee	East	9000	4																							

	跳过任何排名值，因此您会看到连续排名值。																									
RANK_MODIFIED()	从 1 开始按升序或降序为每一行分配整数排名。如果行具有相同的值，则它们共享分配给该值的最后一个实例的排名。不会跳过排名值。Rank_Modified 的计算方式为 $\text{Rank} + (\text{Rank} + \text{Number of duplicate rows} - 1)$ 。	<table border="1"> <thead> <tr> <th>Sales Person</th><th>Region</th><th>Commission</th><th>Rank</th></tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td><td>Central</td><td>7000</td><td>1</td></tr> <tr> <td>Lisa Martin</td><td>Central</td><td>15000</td><td>3</td></tr> <tr> <td>Sarah Scott</td><td>Central</td><td>15000</td><td>3</td></tr> <tr> <td>Barbara Davis</td><td>East</td><td>13000</td><td>4</td></tr> <tr> <td>Charles Lee</td><td>East</td><td>9000</td><td>5</td></tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	3	Sarah Scott	Central	15000	3	Barbara Davis	East	13000	4	Charles Lee	East	9000	5
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	3																							
Sarah Scott	Central	15000	3																							
Barbara Davis	East	13000	4																							
Charles Lee	East	9000	5																							
RANK_PERCENTILE()	从 0 到 1 按升序或降序为每一行分配百分比排名。RANK_PERCENTILE 的计算方式为 $(\text{Rank}-1) / (\text{Total rows}-1)$ 。	<table border="1"> <thead> <tr> <th>Sales Person</th><th>Region</th><th>Commission</th><th>Rank</th></tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td><td>Central</td><td>7000</td><td>0.0</td></tr> <tr> <td>Charles Lee</td><td>East</td><td>9000</td><td>0.25</td></tr> <tr> <td>Barbara Davis</td><td>East</td><td>13000</td><td>0.5</td></tr> <tr> <td>Lisa Martin</td><td>Central</td><td>15000</td><td>0.75</td></tr> <tr> <td>Sarah Scott</td><td>Central</td><td>15000</td><td>0.75</td></tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	0.0	Charles Lee	East	9000	0.25	Barbara Davis	East	13000	0.5	Lisa Martin	Central	15000	0.75	Sarah Scott	Central	15000	0.75
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	0.0																							
Charles Lee	East	9000	0.25																							
Barbara Davis	East	13000	0.5																							
Lisa Martin	Central	15000	0.75																							
Sarah Scott	Central	15000	0.75																							
ROW_NUMBER()	为每个唯一行分配连续的行 ID。如果具有重复行并使用此计算，则每次运行流程时，如果行的顺序发生更改，结果可能会发生变化。	<table border="1"> <thead> <tr> <th>Sales Person</th><th>Region</th><th>Commission</th><th>ROW ID</th></tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td><td>Central</td><td>7000</td><td>1</td></tr> <tr> <td>Lisa Martin</td><td>Central</td><td>15000</td><td>2</td></tr> <tr> <td>Sarah Scott</td><td>Central</td><td>15000</td><td>3</td></tr> <tr> <td>Barbara Davis</td><td>East</td><td>13000</td><td>4</td></tr> <tr> <td>Charles Lee</td><td>East</td><td>9000</td><td>5</td></tr> </tbody> </table>	Sales Person	Region	Commission	ROW ID	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	2	Sarah Scott	Central	15000	3	Barbara Davis	East	13000	4	Charles Lee	East	9000	5
Sales Person	Region	Commission	ROW ID																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	2																							
Sarah Scott	Central	15000	3																							
Barbara Davis	East	13000	4																							
Charles Lee	East	9000	5																							

创建排名或行号计算

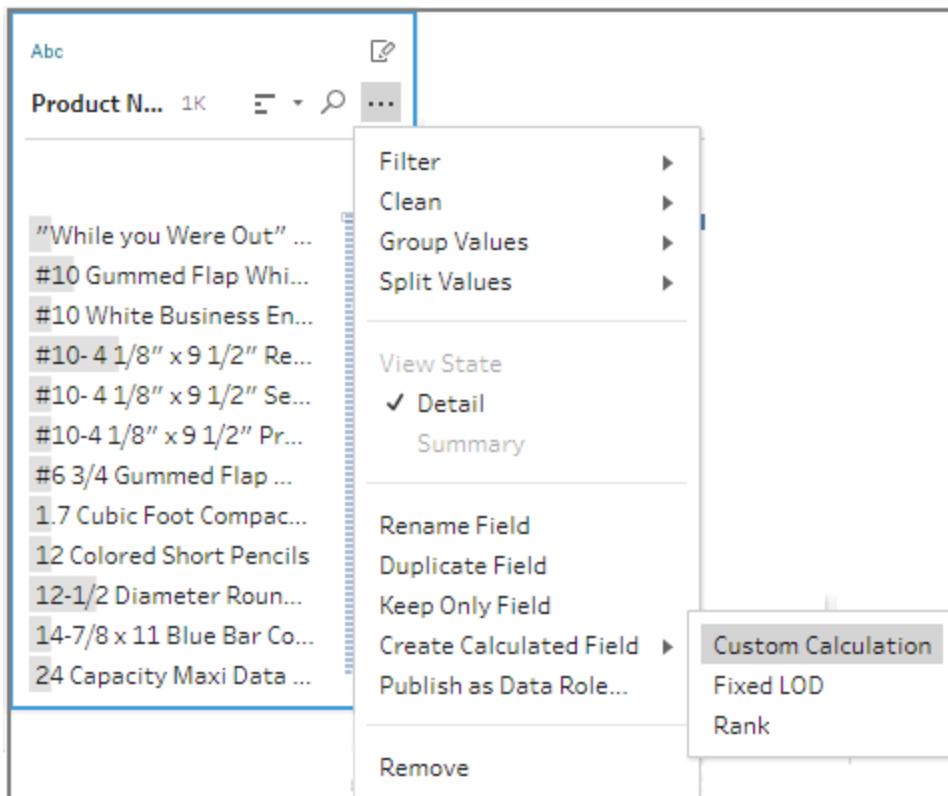
若要创建排名或行号计算，可以使用计算编辑器自行编写计算，或者，如果您希望获得更多指引，则可以使用可视化计算编辑器，您可以在其中选择字段，Tableau Prep Builder 将为您编写计算。

注意: RROW_NUMBER() 计算在可视化计算编辑器中不可用。

计算编辑器

使用计算编辑器创建任何支持的 RANK() 或 ROW_NUMBER() 计算。支持的分析计算列表显示在计算编辑器中“分析”下的“参考”下拉列表中。

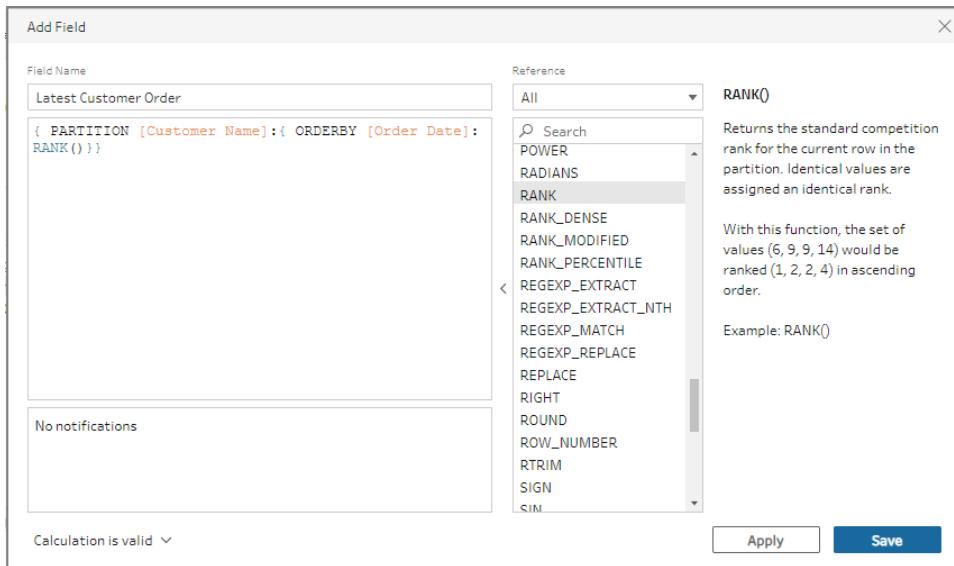
1. 在“配置”窗格工具栏中单击“创建计算字段”，或在配置卡或数据网格中单击“更多选项”… 菜单，并选择“创建计算字段”>“自定义计算”。



2. 在计算编辑器中，输入计算的名称，并输入表达式。

例如，要查找最新的客户订单，请创建如下所示的计算，然后仅保留使用数字 1 排名

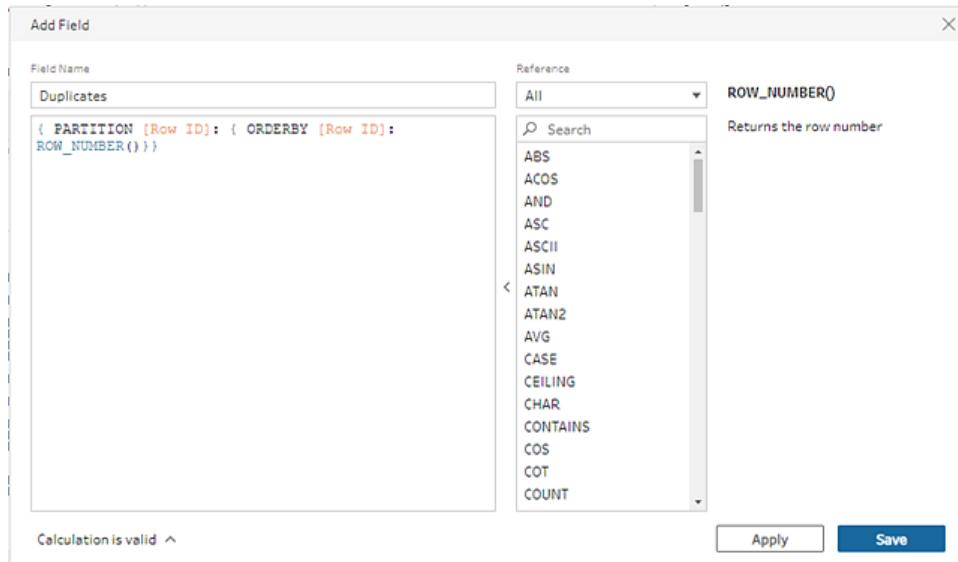
的客户订单行。



示例：使用 ROW_NUMBER 查找和移除重复值

本示例使用 Superstore 示例数据集，使用 ROW_NUMBER 函数查找和移除字段“**Row ID**”（**行 ID**）的确切重复值。

1. 打开“**Sample Superstore**”（示例超市）流程。
2. 在“流程”窗格中，对于“输入”步骤“**Orders West**”（对西部排序），单击“清理”步骤“**Rename States**”（重命名状态）。
3. 在工具栏中，单击“创建计算字段”。
4. 在计算编辑器中，使用 ROW_NUMBER 函数，通过 {PARTITION [Row ID] : {ORDERBY [Row ID] : ROW_NUMBER ()} } 表达式向字段“**Row Id**”（**行 ID**）中添加行号，然后单击“保存”。



5. 在新的计算字段中，在字段值“1”上右键单击或按住 Cmd 单击 (MacOS)，然后从菜单中选择“只保留”。

#

Duplicates 11

1

2

3

4

5

6

7

8

9

10

11

Keep Only

Exclude

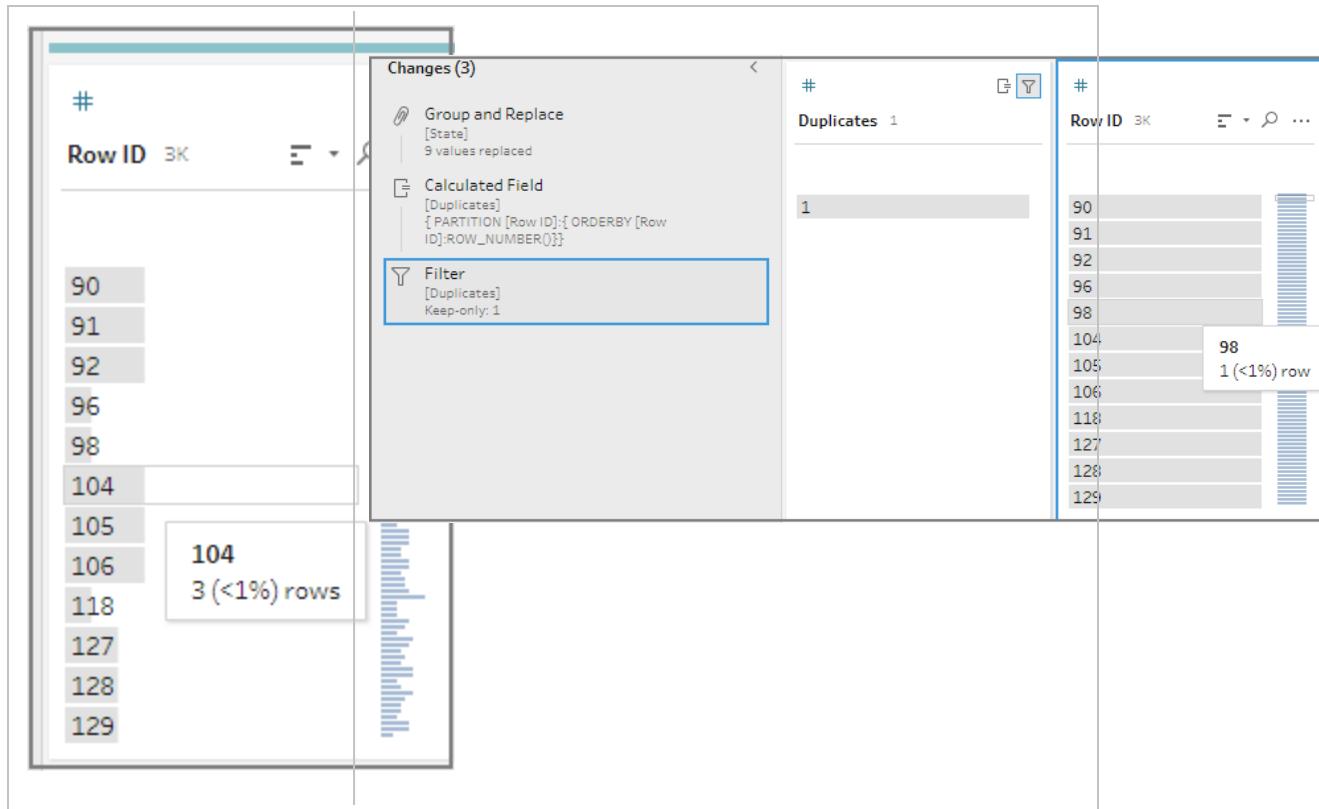
Edit Value

Replace with Null

Group

Ungroup

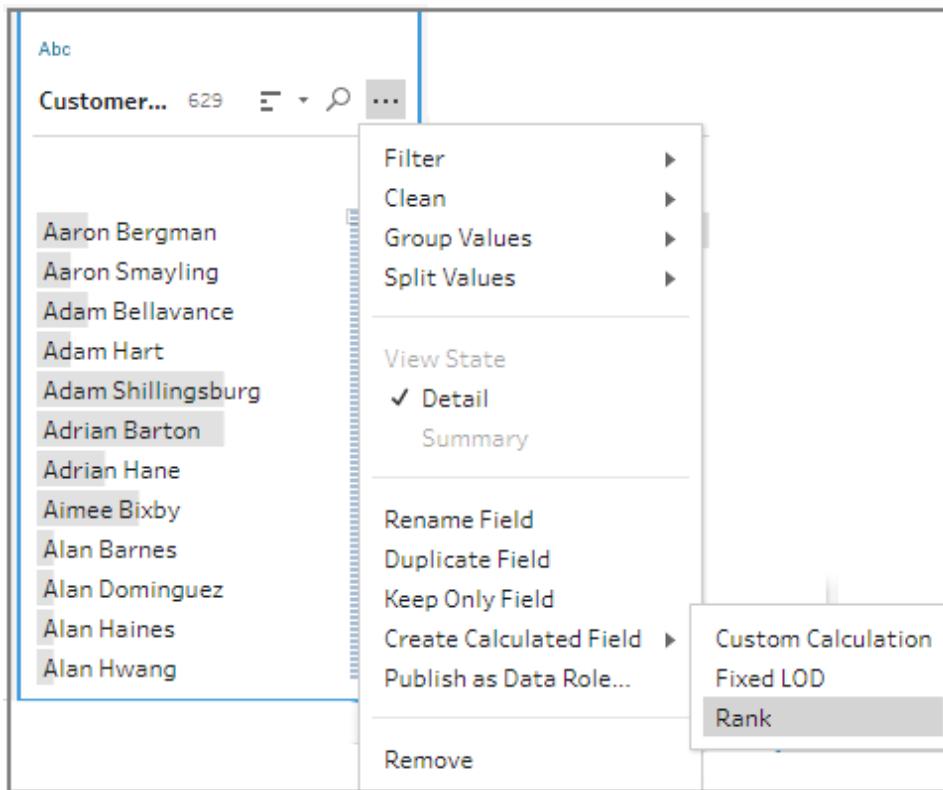
之前 | 之后



可视化计算编辑器

就像创建详细级别计算时一样，您可以使用可视化计算编辑器来生成排名计算。选择要在计算中包括的字段，然后选择要用于对行进行排名的字段，以及要计算的排名类型。结果的预览显示在左侧窗格中，这样您就能够随时查看所选内容的结果。

- 在配置卡或结果窗格中，单击“更多选项”… 菜单并选择“创建计算字段”>“排名”。



- 在可视化计算编辑器中，执行以下操作：

- 在“分组依据”部分，选择要针对其各个行计算值的字段。这将创建计算的 Partition 部分。

单击加号 图标以向计算中添加任何其他字段。如果要包括所有行或移除所选行，请在“分组依据”部分的字段下拉框中右键单击或按住 Cmd 单击 (MacOS)，然后选择“移除字段”。

- 在“排序依据”部分，选择要用于对新值进行排名的字段。默认情况下，将添加您在其中选择了“创建计算字段”>“排名”菜单选项的字段。

单击加号 图标以向计算中添加任何其他字段，然后选择“排名”类型。单击

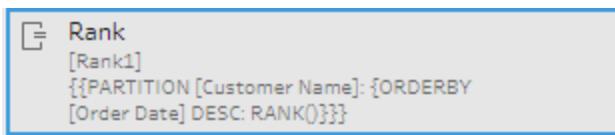
排序 图标将排名顺序从降序 (DESC) 更改为升序 (ASC)。

注意: 排名值因分配给字段的数据类型而异。

- 在左侧窗格中，双击字段标题并输入用于计算的名称。

Customer Name	Rank	Order Date
Aaron Bergman	1	03/07/2015, 12:00:00 AM
Aaron Hawkins	1	12/27/2016, 12:00:00 AM
Aaron Hawkins	5	10/25/2015, 12:00:00 AM
Aaron Hawkins	6	05/13/2015, 12:00:00 AM
Aaron Smayling	1	10/03/2018, 12:00:00 AM
Aaron Smayling	2	09/04/2018, 12:00:00 AM
Aaron Smayling	3	09/25/2017, 12:00:00 AM
Adam Bellavance	1	11/06/2018, 12:00:00 AM
Adam Bellavance	10	05/07/2018, 12:00:00 AM

- 单击“完成”以添加新的计算字段。在“更改”窗格中，您可以看到 Tableau Prep Builder 生成的计算。单击“编辑”打开可视化计算编辑器以进行任何更改。



转置您的数据

有时候，使用 Tableau 分析电子表格或者横向表格(交叉数据)的数据会遇到一些困难。Tableau 更倾向于数据是一行一行的“高高的数据”(行数据)，而不是一列一列的“很宽很长的数据”(列数据)，这就需要我们把列数据转置为行数据。

但是，可能还会有一些数据表既长又窄，并且太过标准化而无法正常分析的情况。例如，销售部门在两个列中跟踪广告支出，一个列名为“Advertising”(广告)，其中包含表示“radio”(广播广告)、“television”(电视广告)和“print”(印刷广告)的行，而另一个列则表示总支出。在这种情况下，若要将此数据作为单独的度量进行分析，您将需要将该行数据转置为列。

但如果转置较大的数据集或在一段时间内频繁更改的数据，会怎么样？您可以使用通配符模式匹配来搜索与模式匹配的字段，并自动转置数据。

在转置数据时，使用以下选项之一：

- 将列转置为行
- 使用通配符搜索，根据模式匹配即时转置字段(Tableau Prep Builder 版本 2019.1.1 及更高版本)。
- 将行转置为列(Tableau Prep Builder 版本 2019.1.1 及更高版本)。

不管如何转置字段，您都可以直接与结果交互，并执行任何额外的清理操作，让数据看起来恰如您所需。您还可以使用 Tableau Prep Builder 的智能默认命名功能来自动重命名已转置的字段和值。

注意: Tableau Prep Builder 2018.1.1 中不支持在多个字段组上将列转置为行。如果转置包括 Tableau Prep Builder 版本 2018.1.1 中多个字段组上的已转置列，则编辑该转置会导致错误或意外结果。

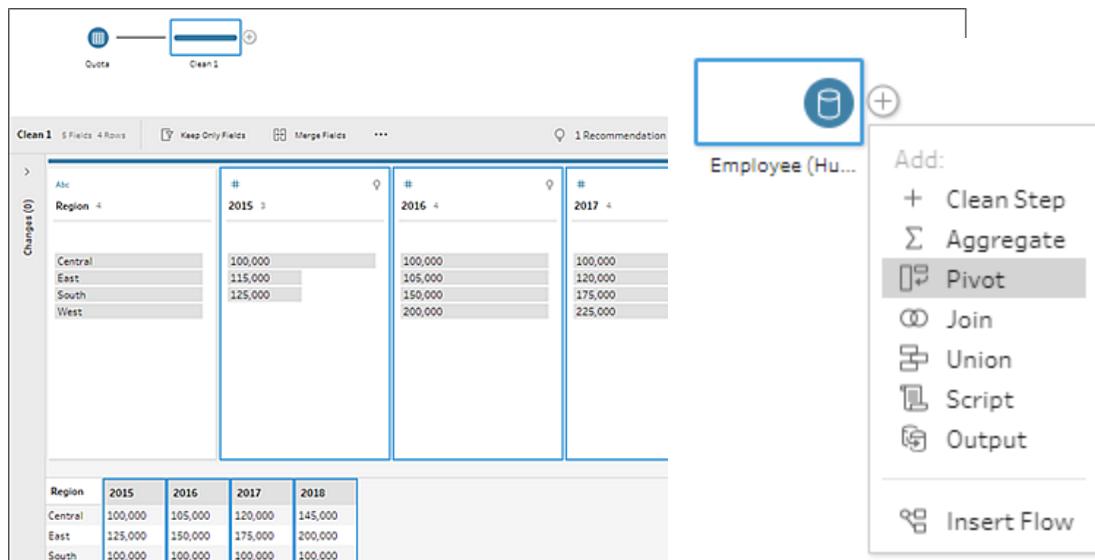
将列转置为行

使用此转置选项将宽数据转换为高数据。在一个或多个字段组上将列转置为行。选择要处理的字段，并将数据从列转置为行。

1. 连接到您的数据源。
2. 将要进行转置的表拖到“流程”窗格。
3. 执行以下操作之一：
 - **版本 2019.4.2 及更高版本**：在“配置”窗格中，选择要转置的字段，然后右键单击或按住 Ctrl 单击 (MacOS)，并从菜单中选择“将列转置为行”。如果使用此选项，请跳到步骤 7。
 - **所有版本**：单击加号  图标，并从上下文菜单中选择“添加转置”。

选择字段(版本 2019.4.2 及更高版本)

流程步骤菜单(所有版本)



4. (可选) 在“字段”窗格的“搜索”字段中输入一个值，以在字段列表中搜索要转置的字段。
5. (可选) 选中“自动重命名转置的字段和值”复选框，使 Tableau Prep 能够使用数据中的常用值重命名新的转置字段。如果找不到常用值，则使用默认名称。
6. 从左侧窗格中选择一个或多个字段，并将这些字段拖到“转置的字段”窗格中的“转置 1 值”列。
7. (可选) 在“转置的字段”窗格中，单击加号 图标添加要进行转置的更多列，然后重复上面的步骤以选择要进行转置的更多字段。结果将立即出现在“转置结果”窗格和数据窗格中。

注意: 您选择的字段数必须与在步骤 5 中选择的字段数相同。举例来说，如果选择了 3 个字段对其进行最初转置，则您进行转置的每个后续列也必须包含 3 个字段。

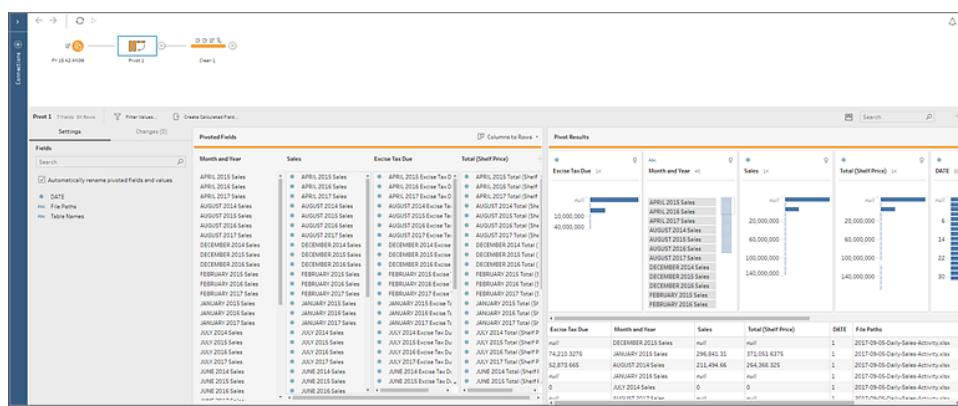
8. 如果未启用默认命名选项或者 Tableau Prep 无法自动检测名称，请编辑字段的名称。您也可以在此窗格中编辑原始字段的名称，以对数据进行最恰当地描述。
9. (可选) 重命名新的“转置”步骤以跟踪更改。例如“转置月份”。
10. 若要在数据更改时刷新转置数据，请运行流程。如果将需要添加到转置的新字段添加到数据源，请将它们手动添加到转置。

示例: 对多个字段进行转置

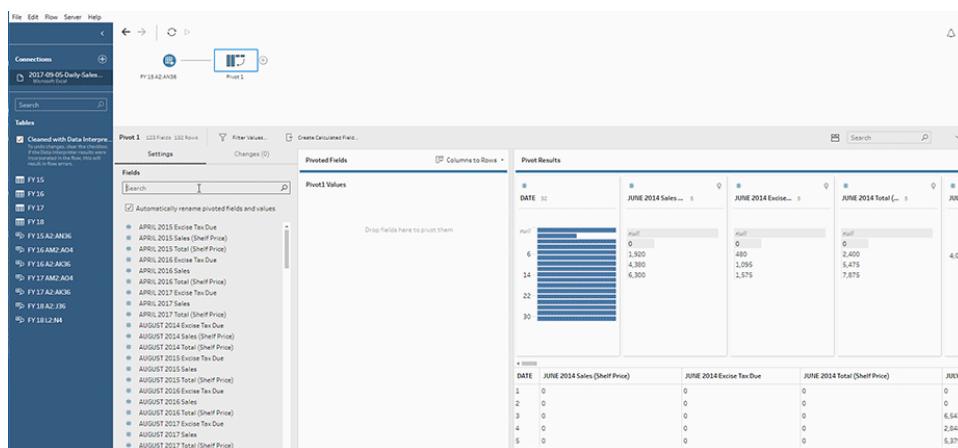
此示例按月和年显示药品销售额、税款和总计的电子表格。

A	T	U	V	W	X	Y	Z	AA	AB
DATE	DECEMBER 2014			JANUARY 2015			FEBRUARY 2015		
	Sales (\$ Shelf Price)	Excise Tax Due	Total (\$ Shelf Price)	Sales (\$ Shelf Price)	Excise Tax Due	Total (\$ Shelf Price)	Sales (\$ Shelf Price)	Excise Tax Due	Total (\$ Shelf Price)
4	1 \$ 448,111 \$ 112,026 \$	\$ 560,139 \$	\$ 296,841 \$	\$ 74,210 \$	\$ 371,052 \$	\$ 212,919 \$	\$ 53,230 \$	\$ 266,148	
5	2 \$ 425,472 \$ 106,368 \$	\$ 531,849 \$	\$ 754,061 \$	\$ 188,515 \$	\$ 942,577 \$	\$ 449,897 \$	\$ 112,474 \$	\$ 562,371	
6	3 \$ 435,525 \$ 108,881 \$	\$ 544,406 \$	\$ 482,497 \$	\$ 120,624 \$	\$ 603,121 \$	\$ 627,711 \$	\$ 156,928 \$	\$ 784,639	
7	4 \$ 634,765 \$ 158,691 \$	\$ 793,456 \$	\$ 332,228 \$	\$ 83,057 \$	\$ 415,284 \$	\$ 688,263 \$	\$ 172,066 \$	\$ 860,329	
8	5 \$ 695,425 \$ 173,856 \$	\$ 869,282 \$	\$ 601,529 \$	\$ 150,382 \$	\$ 751,912 \$	\$ 789,233 \$	\$ 197,308 \$	\$ 986,541	
9	6 \$ 436,720 \$ 109,180 \$	\$ 545,899 \$	\$ 527,374 \$	\$ 131,843 \$	\$ 659,217 \$	\$ 867,501 \$	\$ 216,875 \$	\$ 1,084,377	
10	7 \$ 238,481 \$ 59,620 \$	\$ 298,101 \$	\$ 560,102 \$	\$ 149,026 \$	\$ 706,128 \$	\$ 554,459 \$	\$ 138,615 \$	\$ 693,074	
11	8 \$ 421,422 \$ 105,356 \$	\$ 526,778 \$	\$ 539,974 \$	\$ 134,993 \$	\$ 674,967 \$	\$ 448,846 \$	\$ 112,211 \$	\$ 561,057	
12	9 \$ 543,816 \$ 135,954 \$	\$ 679,778 \$	\$ 683,408 \$	\$ 170,854 \$	\$ 854,260 \$	\$ 768,266 \$	\$ 192,067 \$	\$ 960,333	
13	10 \$ 616,271 \$ 154,068 \$	\$ 770,339 \$	\$ 442,352 \$	\$ 110,588 \$	\$ 552,940 \$	\$ 719,637 \$	\$ 179,909 \$	\$ 899,546	
14	11 \$ 756,542 \$ 189,135 \$	\$ 945,677 \$	\$ 288,605 \$	\$ 72,151 \$	\$ 360,756 \$	\$ 1,154,919 \$	\$ 288,730 \$	\$ 1,443,649	
15	12 \$ 726,270 \$ 907,837 \$	\$ 674,121 \$	\$ 168,530 \$	\$ 842,651 \$	\$ 1,019,936 \$	\$ 254,984 \$	\$ 1,274,921		
16	13 \$ 477,208 \$ 119,302 \$	\$ 596,510 \$	\$ 526,451 \$	\$ 131,613 \$	\$ 658,064 \$	\$ 951,242 \$	\$ 237,811 \$	\$ 1,189,053	
17	14 \$ 245,898 \$ 61,475 \$	\$ 307,373 \$	\$ 573,842 \$	\$ 143,461 \$	\$ 717,393 \$	\$ 798,392 \$	\$ 199,598 \$	\$ 997,991	
18	15 \$ 456,254 \$ 114,064 \$	\$ 570,318 \$	\$ 658,952 \$	\$ 164,738 \$	\$ 823,690 \$	\$ 453,091 \$	\$ 113,273 \$	\$ 566,364	

通过对数据进行转置，您可以为每个月份和年度创建行，并为销售额、税款和总计创建单独的列，以便 Tableau 能够更轻松地解读此数据来进行分析。



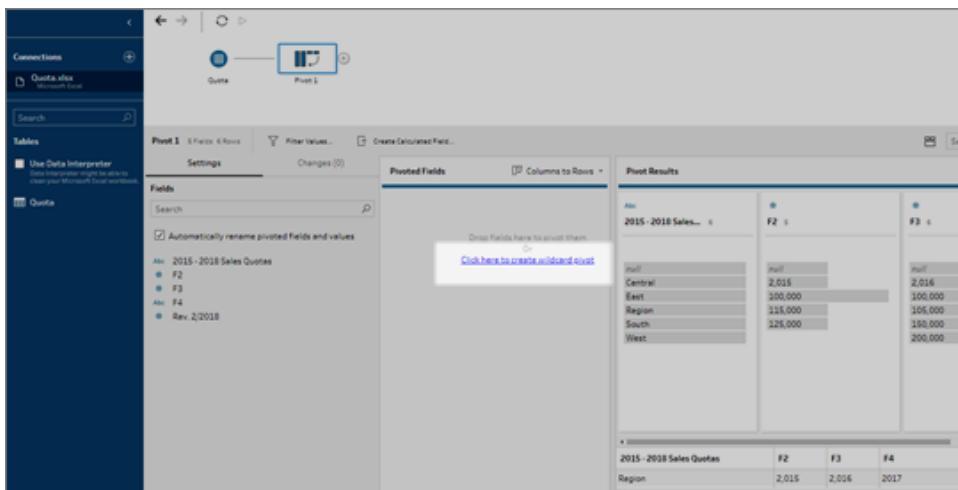
观看“对多个字段进行转置”的实际运行效果。



使用通配符搜索进行转置(版本 2019.1.1 及更高版本)

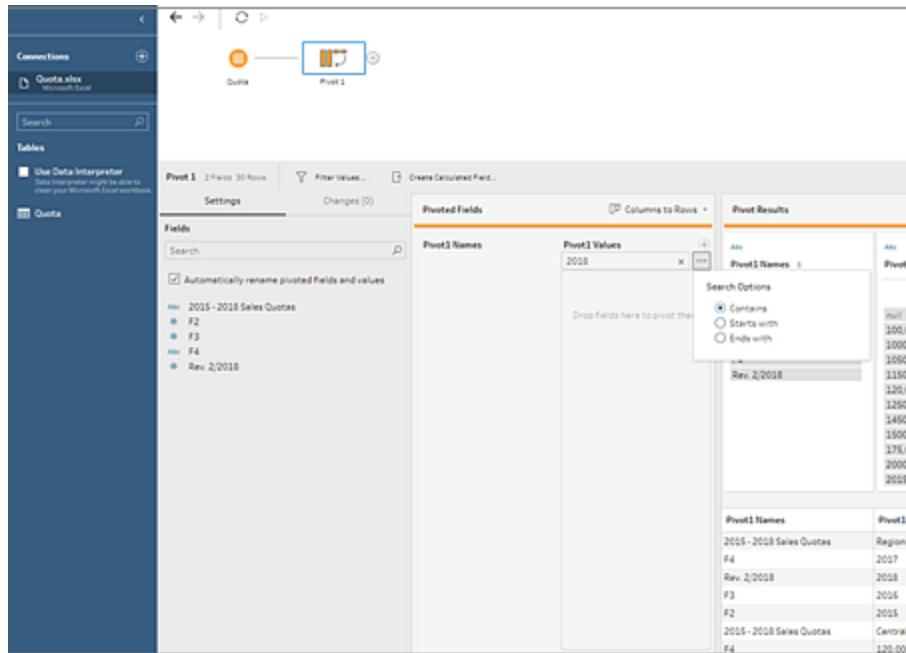
如果使用大数据集或者数据在一段时间内频繁更改，您可以在将列转置为行时使用通配符搜索，以便根据通配符模式匹配即时转置数据。如果添加或移除了与模式匹配的新字段，Tableau Prep Builder 会在运行流程时检测到架构更改，并且转置结果会自动更新。

1. 连接到您的数据源。
2. 将要进行转置的表拖到“流程”窗格。
3. 单击加号  图标，并从上下文菜单中选择“添加转置”。
4. 在“转置的字段”窗格中，单击“单击此处创建通配符转置”链接。



5. 输入要搜索的值或部分值。例如，输入“**Sales_**”以匹配标为“**sales_2017**”(2017 年销售额)、“**sales_2018**”(2018 年销售额) 和“**sales_2019**”(2019 年销售额) 的字段。

不要使用星号来匹配模式，除非星号是所搜索的字段值的一部分。相反，请单击“**搜索选项**” 按钮来选择所需的值匹配方式。然后按 **Enter** 应用搜索并转置匹配的值。



6. (可选) 在“转置的字段”窗格中, 单击加号 图标添加要进行转置的更多列, 然后重复上面的步骤以选择要进行转置的更多字段。
7. 如果未启用默认命名选项或者 Tableau Prep Builder 无法自动检测名称, 请编辑字段的名称。
8. 若要在数据更改时刷新转置数据, 请运行流程。系统会自动检测到添加到数据源并与通配符模式匹配的任何新字段, 并将其添加到转置。
9. 如果结果不是预期结果, 请尝试以下选项之一:
 - 在“搜索”字段中输入其他值模式, 并按 **Enter**。转置将自动刷新并显示新结果。
 - 将其他字段手动拖到“转置的字段”窗格中的“转置 1 值”列。您也可以移除手动添加的字段, 方法是将它们拖离“转置 1 值”列并放到“字段”窗格中。

注意:对于依据通配符搜索结果添加的字段, 无法通过将它们拖离“转置 1 值”列来移除。作为替代, 请尝试使用更具体的模式来匹配所查找的搜索结果。

将行转置为列(版本 2019.1.1 及更高版本)

如果您的数据太过标准化，并且您需要创建新列，请使用此选项 - 从高数据转换为较宽的数据。

举例来说，假设您有每个月的广告成本数据，其中每个月都在一列中包括了所有广告类型。如果您将数据从行转置为列，则每种广告类型会变为拥有单独的一列，从而使数据更易于分析。



Month	Cost	Advertising Type
January	12187	Radio
January	45698	Television
January	9762	Print
February	4598	Print
February	23987	Television
February	8907	Radio
March	10987	Print

Month	Radio	Television	Print
January	12187	45698	9762
February	8907	23987	4598
March	Null	Null	10987

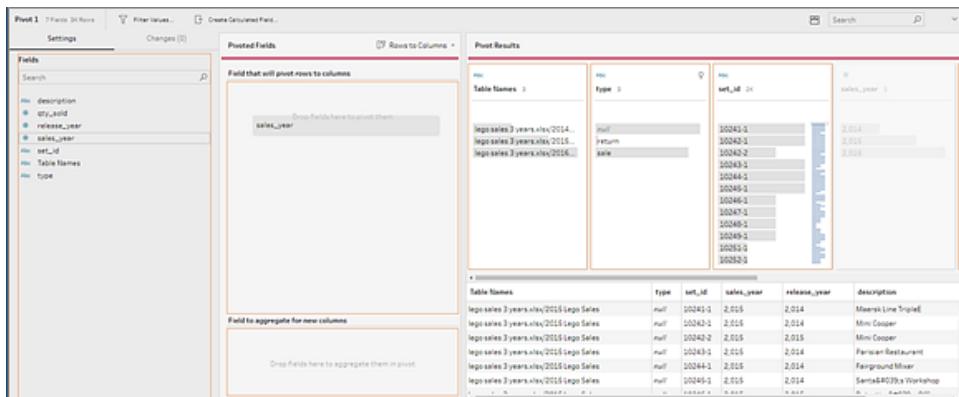
您可以选择一个要转置的字段。随后将使用该字段的字段值来创建新列。然后，选择一个要用于填充新列的字段。这些字段值是聚合的，您可要选择要应用的聚合类型。

由于应用了聚合，因此将列转置回行将不会反向进行此转置操作。若要逆转行到列转置类型，您将需要撤消操作。在顶部菜单中单击“**撤消**”按钮、从“**转置的字段**”窗格中移除字段，或删除转置步骤。

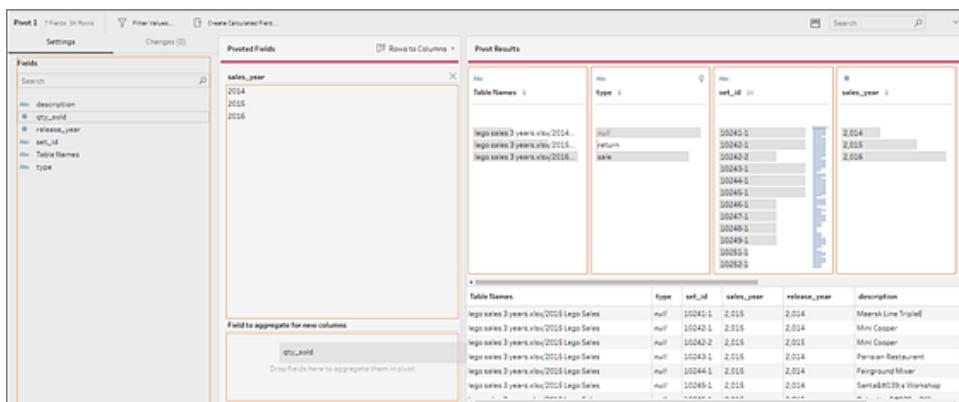
1. 连接到您的数据源。
2. 将要进行转置的表拖到“**流程**”窗格。
3. 单击加号  图标，并从上下文菜单中选择“**添加转置**”。
4. 在“**转置的字段**”窗格中，从下拉列表中选择“**行转置为列**”。
5. (可选)在“**字段**”窗格的“**搜索**”字段中输入一个值，以在字段列表中搜索要转置的字段
6. 从左侧窗格中选择一个字段，并将其拖到“**转置的字段**”窗格中的“**将从行转置为列的字段**”部分。

注意:如果要转置的字段具有日期或日期时间数据类型，则需要将数据类型更改为字符串才能对其进行转置。

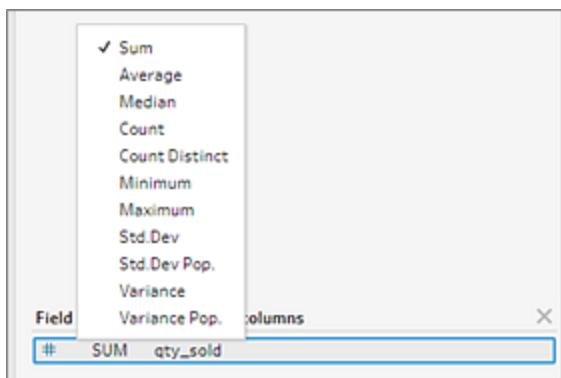
此字段中的值将用于创建和命名新列。您可以稍后在“**转置结果**”窗格中更改列名称。



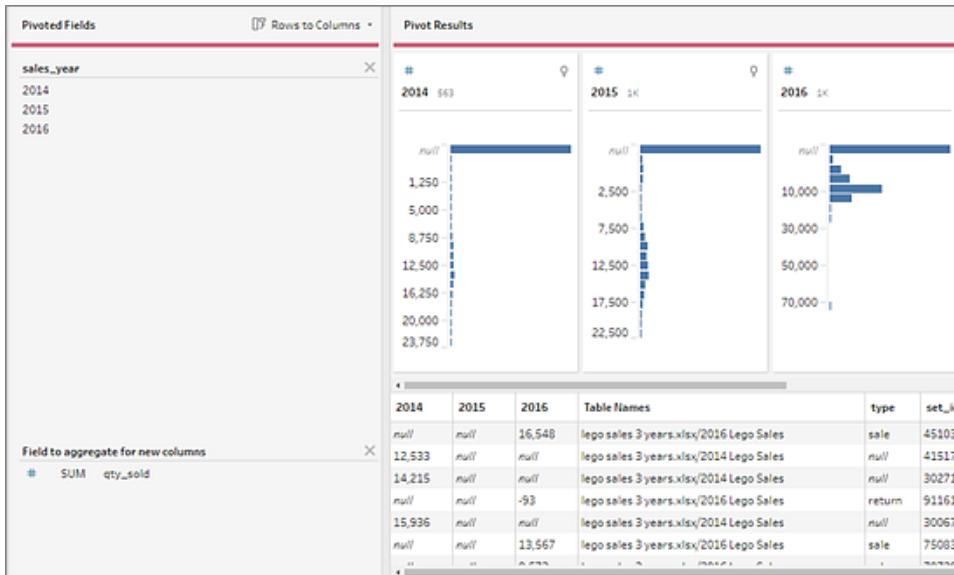
7. 从左侧窗格中选择一个字段，并将其拖到“转置的字段”窗格中的“要为新列聚合的字段”部分。此字段中的值用于填充通过上一步创建的新列。



系统会为字段分配一个默认聚合类型。单击聚合类型可对其进行更改。



8. 在“转置结果”窗格中，查看结果并向您创建的新列应用任何清理操作。

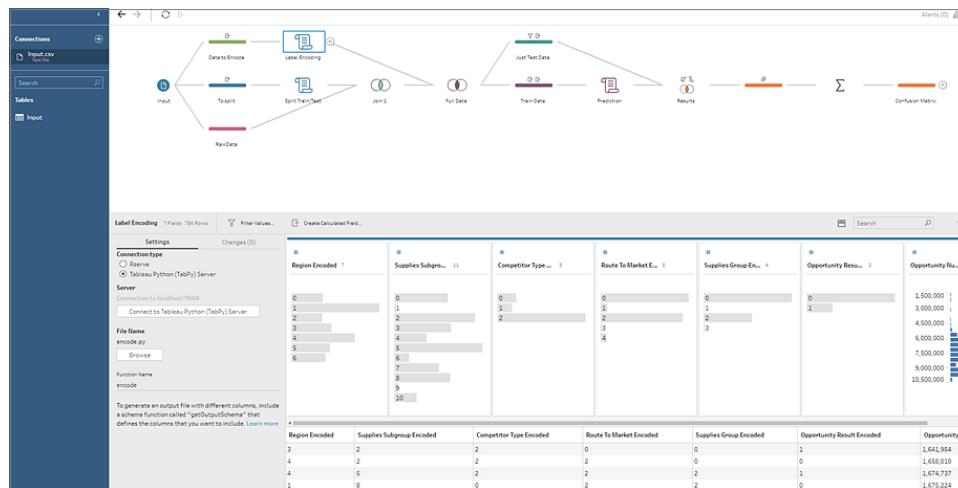


9. 如果正在转置的字段的行数据发生变化，请在流程窗格中右键单击或按住 Ctrl 并单击 (Mac)“转置”步骤，并选择“刷新”。

在流程中使用 R 和 Python 脚本

从版本 2019.3.1 开始，您可以使用 R 和 Python 脚本来执行更复杂的清理操作，或者在流程中并入预测建模数据。数据通过 R 或 Python 脚本步骤以输入形式从流程中传递，然后以您可以使用 Tableau Prep Builder 的功能和函数继续清理的输出数据的形式返回。

注意：尚不支持以流程输入步骤的形式连接到脚本。



向流程中添加脚本步骤并输入所选服务的配置详细信息时，Tableau Prep Builder 会将数据传递到 R 的 Rserve 或 Python 的 Tableau Python 服务器 (TabPy)，并将生成的数据以表形式返回到流程。您可以继续将清理操作应用到结果并生成分析输出。

创建脚本时，需要包含一个函数，该函数将数据框架指定为函数的参数。如果要返回不同于输入的字段，则需要在脚本中包含定义输出和数据类型的 `getOutputSchema` 函数。否则，输出将使用输入数据中的字段。

有关如何配置要在流程中使用的 R 或 Python 以及如何创建脚本的信息，请参见[在您的流程中使用 R \(Rserve\) 脚本 向下](#)或[在您的流程中使用 Python 脚本 在本页 289](#)。

在您的流程中使用 R (Rserve) 脚本

免责声明：本主题包括有关第三方产品的信息。请注意，尽管我们努力保持第三方内容的准确性，但我们在此处提供的信息可能会在 R 和 Rserve 发生变化时更改，恕不另行通知。有关最新信息，请查阅 R 和 Rserve 文档及支持。

R 是一种开源软件编程语言和用于统计计算及图形的软件环境。为了扩展 Tableau Prep Builder 的功能，您可以用 R 创建脚本以在通过 Rserve 服务器运行的流程中使用，生成可在流程中进一步处理的输出。

例如，您可能想要使用 R 语言脚本向流程中已有的数据中添加统计建模数据或预测数据，然后使用 Tableau Prep Builder 的功能清理生成的数据集以进行分析。

若要在流程中包括 R 脚本，您需要在 Tableau Prep Builder 和 Rserve 服务器之间配置连接。然后，您可以使用 R 脚本，通过 R 表达式将支持的函数应用于流程中的数据。输入配置详细信息并将 Tableau Prep Builder 指向要使用的文件和函数后，系统会将数据安全地传递到 Rserve 服务器，应用表达式，并以可根据需要清理或输出的表 (R `data.frame`) 的形式返回结果。

注意：若要在 Tableau Server(版本 2019.3 及更高版本) 上运行包括脚本步骤的流程，Tableau Server 还必须连接到 Rserve 服务器。

先决条件

若要在流程中包括 R 脚本步骤，请安装 R 并配置与 Rserve 服务器的连接。

资源

- **下载并安装 R。** 下载并安装适用于 Linux、Mac 或 Windows 的最新版本的 R。
- **R 实施说明(社区帖子)。** 针对 Windows 安装并配置与 R 和 Rserve 的连接。
- **安装并配置 Rserve:** 适用于所有平台的常规安装和配置说明。
 - **Rserve for Windows(版本说明):** 本主题讨论在 Windows 上以本地方式安装 Rserve 时的限制。

针对 Tableau Server 配置 Rserve Server

如果要发布包括脚本步骤的流程并在 Tableau Server 上运行它们，您将需要在 Rserve 服务器和 Tableau Server 版本 2019.3 或更高版本之间配置连接。当前不支持在 Tableau Online 中运行包含脚本步骤的流程。

1. 打开 TSM 命令行。
2. 输入以下命令以设置主机地址、端口值和连接超时：

```
tsm security maestro-rserve-ssl enable --connection-type
{maestro-rserve-secure/maestro-rserve} --rserve-host <Rserve
IP address or host name> --rserve-port <Rserve port> --
rserve-username <Rserve username> --rserve-password <Rserve
password> --rserve-connect-timeout-ms <RServe connect
timeout>
```

- 选择 {maestro-rserve-secure} 以启用安全连接，或选择 {maestro-rserve} 以启用不安全的连接。
- 如果选择 {maestro-rserve-secure}，请在命令行中指定证书文件 - cf<certificate file path>。
- 以毫秒为单位指定 --rserve-connect-timeout-ms <RServe connect timeout>。例如，--rserve-connect-timeout-ms 900000。

3. 若要禁用 Rserve 连接，请输入以下命令

```
tsm security maestro-rserve-ssl disable
```

其他 Rserve 配置(可选)

您可以创建一个名为 Rserv.cfg 的文件以设置默认配置值来自定义 Rserve，并将其放在 /etc/Rserve.conf 安装位置中。为了提高 Rserve 服务器和 Tableau Prep Builder 的稳定性，您可以向 Rserve 配置中添加其他值。启动 Rserve 时，您可以引用此文件来应用配置选项。例如：

- **Windows:** Rserve (args="--RS-conf C:\\\\folder\\\\Rserv.cfg")
- **MacOS 和 Linux:** Rserve (args=" --no-save --RS-conf ~/Documents/Rserv.cfg")

下面的示例演示了一些可包括在 Rserve.conf 配置文件中的其他选项：

```
# If your data includes characters other than ASCII, make it
explicit that data should be UTF8 encoded.
encoding utf8

# Disable interactive behavior for Rserve or Tableau Prep Builder
will stall when trying to run the script as it waits for an input
response.

interactive no
```

有关设置 **Rserve.conf** 文件的信息，请参见 [R 实施说明](#)(社区帖子)中的“高级 Rserve 配置”部分。

创建 R 脚本

创建脚本时，请包括一个指定数据框架作为函数参数的函数。这将从 Tableau Prep Builder 中调用您的数据。您还需要使用支持的数据类型在数据框架中返回结果。

例如：

```
postal_cluster <- function(df) {
  out <- kmeans(c(df$Latitude, df$Longitude), 3, iter.max=10)
  return(data.frame(Latitude=df$Latitude, Longitude=df$Longitude,
Cluster=out$cluster))
}
```

支持以下数据类型：

Tableau Prep Builder 中的数据 类型	R 中的数据类型
字符串	标准 UTF-8 字符串
十进制	双精度

整数	整数
布尔值	逻辑
日期	ISO_DATE 格式“YYYY-MM-DD”的字符串，带有可选区域偏移。例如，“2011-12-03+01:00”是有效日期。
日期时间	ISO_DATE_TIME 格式“YYYY-MM-DDT:HH:mm:ss”的字符串，带有可选区域偏移。例如，“2011-12-03T10:15:30+01:00”是有效日期。

注意：日期和日期时间必须始终以有效字符串形式返回。不支持将 R 中的本机日期（日期时间）类型作为返回值，但可在脚本中使用。

如果要返回不同于输入的字段，则需要在脚本中包含定义输出和数据类型的 `getOutputSchema` 函数。否则，输出将使用输入数据中的字段，这些字段是从流程中脚本步骤之前的步骤中获取的。

在 `getOutputSchema` 中指定字段的数据类型时，请使用以下语法：

R 中的函数	生成的数据类型
<code>prep_string()</code>	字符串
<code>prep_decimal()</code>	十进制
<code>prep_int()</code>	整数
<code>prep_bool()</code>	布尔值
<code>prep_date()</code>	日期
<code>prep_datetime()</code>	日期时间

以下示例显示 `postal_cluster` 脚本的 `getOutputSchema` 函数：

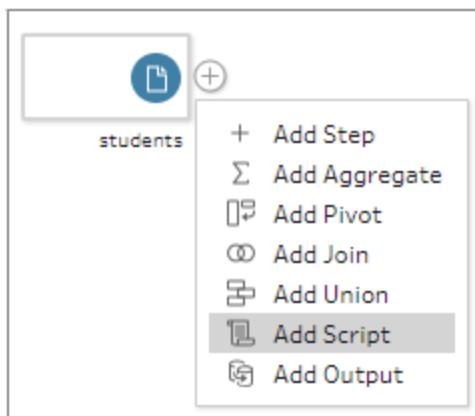
```
getOutputSchema <- function() {
  return (data.frame (
    Latitude = prep_decimal (),
    Longitude = prep_decimal (),
```

```
Cluster = prep_int ());  
}
```

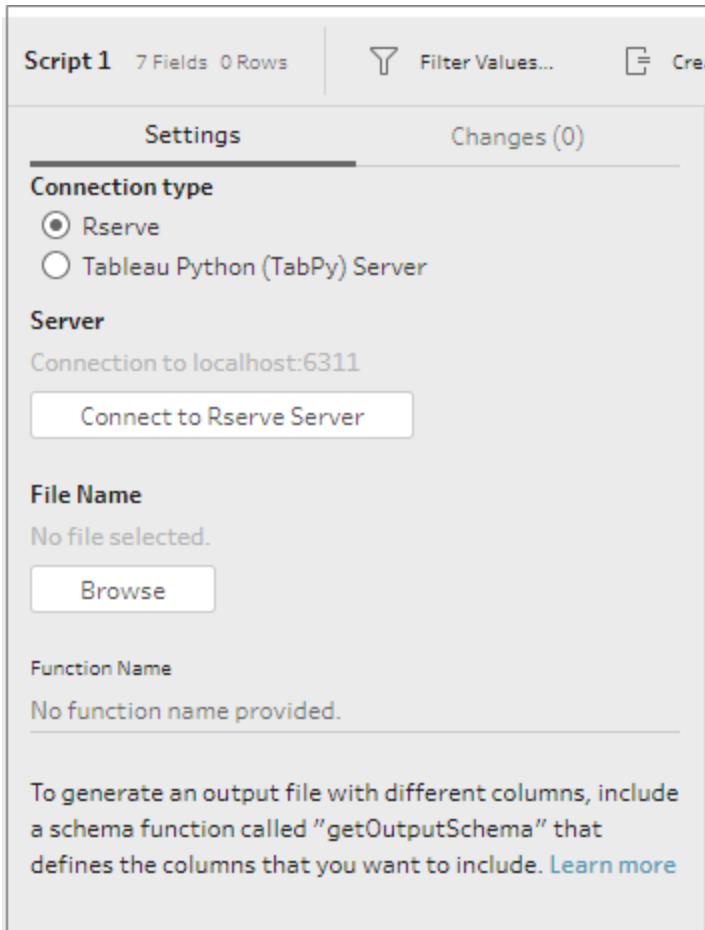
向流程中添加脚本

启动 **Rserve** 服务器，然后完成以下步骤：

1. 打开 Tableau Prep Builder 并单击“添加连接” 按钮。
2. 从连接器列表中，选择文件类型或托管您的数据的服务器。如果出现提示，请输入所需的信息以登录并访问数据。
3. 单击加号  图标，并从上下文菜单中选择“添加脚本”。

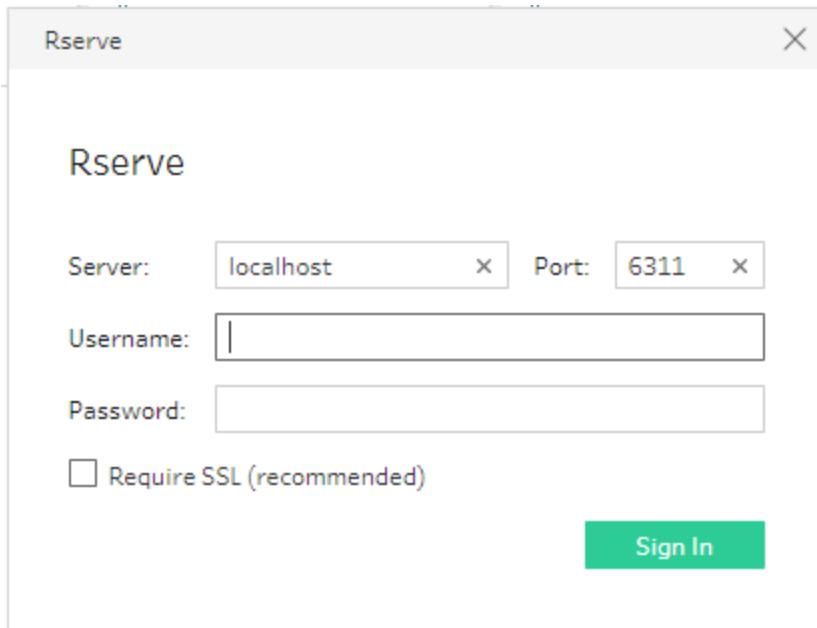


4. 在“脚本”窗格中的“连接类型”下，选择“**Rserve**”。



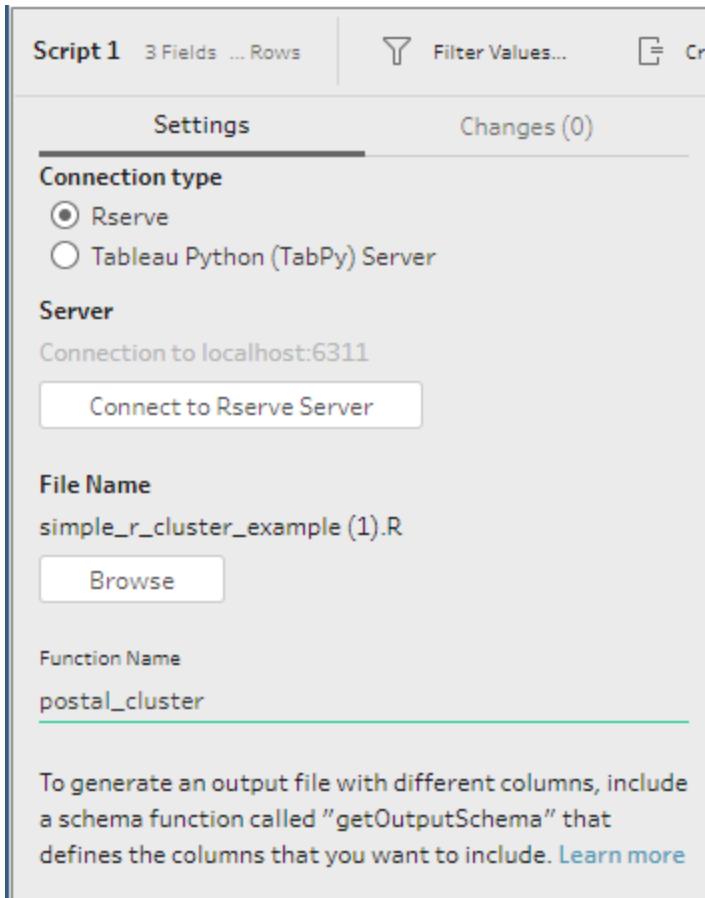
5. 单击“连接到 Rserve 服务器”，并输入连接详细信息。

- 端口 6311 是纯文本 Rserve 服务器的默认端口。
- 端口 4912 是 SSL 加密 Rserve 服务器的默认端口。
- 如果服务器需要凭据，请输入用户名和密码。
- 如果服务器使用 SSL 加密，请选中“**需要 SSL**”复选框，然后单击“**自定义配置文件**”链接，为连接指定证书。



注意: Tableau Prep Builder 未提供测试连接的方式。如果连接存在问题，则会在您尝试并运行流程时显示一条错误消息。

6. 在“文件名”部分，单击“浏览”选择您的脚本文件。
7. 输入“函数名称”，然后按 **Enter** 运行脚本。



在您的流程中使用 Python 脚本

免责声明: 本主题包括有关第三方产品的信息。请注意，尽管我们努力保持第三方内容的准确性，但我们在此处提供的信息可能会在 Python 发生变化时更改，恕不另行通知。有关最新信息，请查阅 Python 文档和支持。

Python 是广泛使用的高级编程语言，用于一般用途编程。通过借助 Tableau Prep Builder 将 Python 命令发送到外部服务，您可以执行诸如添加行号、进行字段排名、填写字段，以及执行可使用计算字段以其他方式执行的清理操作，从而轻松地扩展数据准备选项。

若要在流程中包括 Python 脚本，您需要在 Tableau Prep Builder 和 TabPy 服务器之间配置连接。然后，您可以使用 Python 脚本，通过 pandas 数据框架将支持的函数应用于流程中的数据。向流程中添加脚本步骤并指定要使用的配置详细信息、文件和函数时，系统会将数据安全传递给 TabPy 服务器，应用脚本中的表达式，并以可根据需要清理或输出的表的形式返回结果。

注意: 若要在 Tableau Server(版本 2019.3 及更高版本) 上运行包括脚本步骤的流程, Tableau Server 还必须连接到 TabPy 服务器。

先决条件

若要在您的流程中包括 Python 脚本, 请完成以下设置。

1. 下载并安装 **Python**。下载并安装适用于 Linux、Mac 或 Windows 的最新版本的 Python。
2. 下载并安装 Tableau Python 服务器 (**TabPy**)。按照用于安装 TabPy 的安装和配置说明进行操作。Tableau Prep Builder 使用 TabPy 以输入形式通过 TabPy 传递流程中的数据, 应用您的脚本, 然后将结果返回到流程。
3. 安装 **Pandas**。运行 `pip3 install pandas`。您必须在脚本中使用 **pandas** 数据框架来与 Tableau Prep Builder 集成。

为 Tableau Server 配置 Tableau Python (TabPy) 服务器

如果要发布包括脚本步骤的流程并在 Tableau Server 上运行它们, 您将需要在 TabPy 服务器和 Tableau Server 版本 2019.3 或更高版本之间配置连接。当前不支持在 Tableau Online 中运行包含脚本步骤的流程。

1. 打开 TSM 命令行/shell。

2. 输入以下命令以设置主机地址、端口值和连接超时:

```
tsm security maestro-tabpy-ssl enable --connection-type
{maestro-tabpy-secure/maestro-tabpy} --tabpy-host <TabPy IP
address or host name> --tabpy-port <TabPy port> --tabpy-
username <TabPy username> --tabpy-password <TabPy password>
--tabpy-connect-timeout-ms <TabPy connect timeout>
```

- 选择 {maestro-tabpy-secure} 以启用安全连接, 或选择 {maestro-tabpy} 以启用不安全的连接。
- 如果选择 {maestro-tabpy-secure}, 请在命令行中指定证书文件 - cf<certificate file path>。
- 以毫秒为单位指定 --tabpy-connect-timeout-ms <TabPy connect timeout>。例如, --tabpy-connect-timeout-ms 900000。

3. 若要禁用 TabPy 连接, 请输入以下命令

```
tsm security maestro-tabpy-ssl disable
```

创建 Python 脚本

创建脚本时，请包括一个指定 `pandas (pd.DataFrame)` 作为函数参数的函数。这将从 Tableau Prep Builder 中调用您的数据。您还需要使用支持的数据类型在 `pandas (pd.DataFrame)` 中返回结果。

例如，若要向流程中的一组字段添加编码，可以编写以下脚本：

```
def encode(input):
    le = preprocessing.LabelEncoder()
    Return pd.DataFrame({
        'Opportunity Number' : input['Opportunity Number'],
        'Supplies Subgroup Encoded' : le.fit_transform(input
        ['Supplies Subgroup']),
        'Region Encoded' : le.fit_transform(input['Region']),
        'Route To Market Encoded' : le.fit_transform(input['Route To
        Market']),
        'Opportunity Result Encoded' : le.fit_transform(input
        ['Opportunity Result']),
        'Competitor Type Encoded' : le.fit_transform(input
        ['Competitor Type']),
        'Supplies Group Encoded' : le.fit_transform(input['Supplies
        Group']),
    })
```

支持以下数据类型：

Tableau Prep Builder 中的数据 类型	Python 中的数据类型
字符串	标准 UTF-8 字符串
十进制	双精度
整数	整数
布尔值	布尔值
日期	<code>ISO_DATE</code> 格式“YYYY-MM-DD”的字符串，带有可选区域偏移。例

	如，“2011-12-03”是有效日期。
日期时间	ISO_DATE_TIME 格式“YYYY-MM-DDT:HH:mm:ss”的字符串，带有可选区域偏移。例如，“2011-12-03T10:15:30+01:00”是有效日期。

注意：日期和日期时间必须始终以有效字符串形式返回。

如果要返回不同于输入的字段，则需要在脚本中包含定义输出和数据类型的 `get_output_schema` 函数。否则，输出将使用输入数据中的字段，这些字段是从流程中脚本步骤之前的步骤中获取的。

在 `get_output_schema` 中指定字段的数据类型时，请使用以下语法：

Python 中的函数	生成的数据类型
<code>prep_string()</code>	字符串
<code>prep_decimal()</code>	十进制
<code>prep_int()</code>	整数
<code>prep_bool()</code>	布尔值
<code>prep_date()</code>	日期
<code>prep_datetime()</code>	日期时间

下面的示例演示添加到字段编码 Python 脚本的 `get_output_schema` 函数：

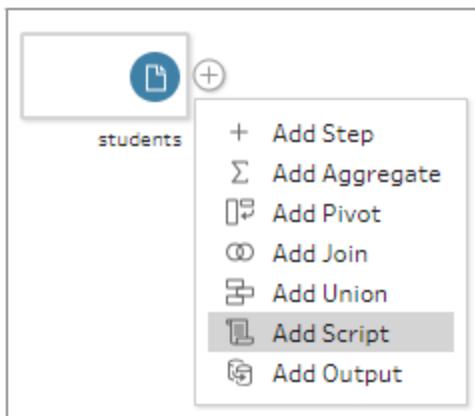
```
def get_output_schema():
    return pd.DataFrame({
        'Opportunity Number' : prep_int(),
        'Supplies Subgroup Encoded' : prep_int(),
        'Region Encoded' : prep_int(),
        'Route To Market Encoded' : prep_int(),
        'Opportunity Result Encoded' : prep_int(),
        'Competitor Type Encoded' : prep_int(),
        'Supplies Group Encoded' : prep_int()
    })
```

向流程中添加脚本

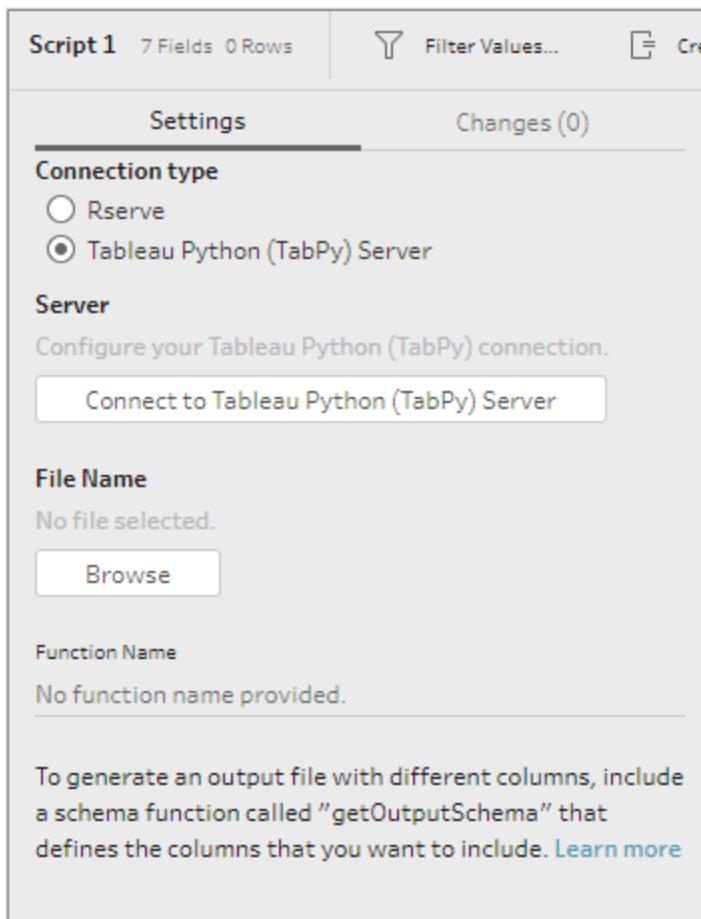
启动 TabPy 服务器，然后完成以下步骤：

注意：TabPy 需要 tornado 程序包版本 5.1.1 才能运行。如果在尝试启动 TabPy 时遇到“tornado.web' has no attribute 'asynchronous'”(“tornado.web”没有属性“asynchronous”)错误，请从命令行中运行 pip list，检查已安装的 tornado 的版本。如果安装了其他版本，请下载 **tornado 程序包版本 5.1.1**。运行 pip uninstall tornado 卸载当前版本，然后运行 pip install tornado==5.1.1 安装所需的版本。

1. 打开 Tableau Prep Builder 并单击“添加连接” 按钮。
2. 从连接器列表中，选择文件类型或托管您的数据的服务器。如果出现提示，请输入所需的信息以登录并访问数据。
3. 单击加号  图标，并从上下文菜单中选择“添加脚本”。

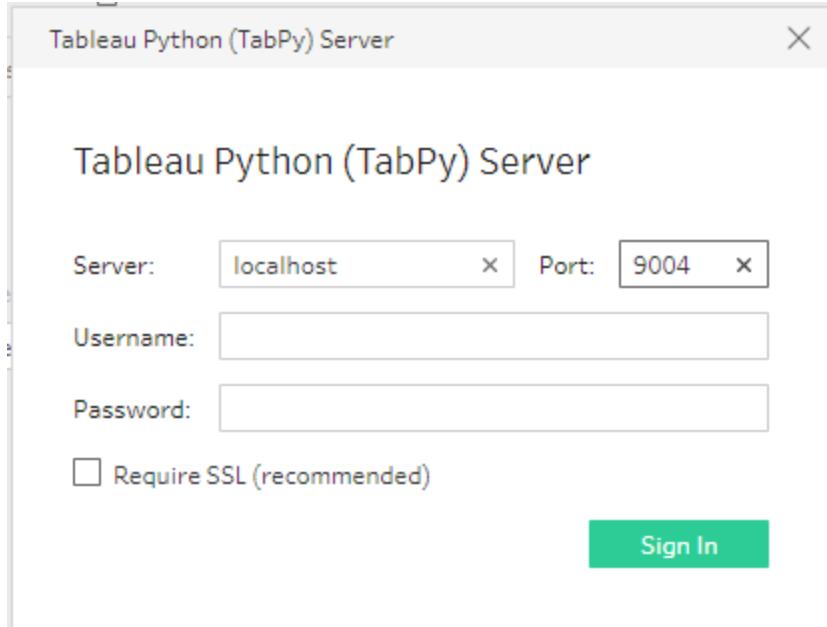


4. 在“脚本”窗格中的“连接类型”部分，选择“Tableau Python (TabPy) 服务器”。



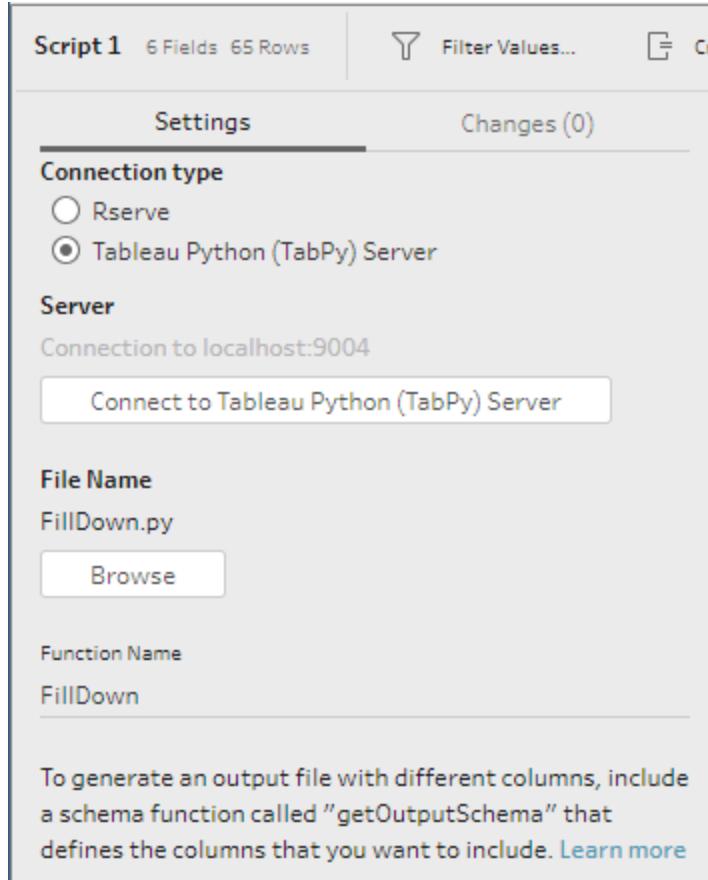
5. 单击“连接到 Tableau Python (TabPy) 服务器”，并输入连接详细信息。

- 端口 9004 是 TabPy 的默认端口。
- 如果服务器需要凭据，请输入用户名和密码。
- 如果服务器使用 SSL 加密，请选中“需要 SSL”复选框，然后单击“自定义配置文件...”链接，为连接指定证书。



注意: Tableau Prep Builder 未提供测试连接的方式。如果连接出现错误，将会显示一条错误消息。

6. 在“文件名”部分，单击“浏览”选择您的脚本文件。
7. 输入“函数名称”，然后按 **Enter** 运行脚本。



聚合、联接或合并数据

对数据进行聚合、联接或合并，以便将数据分组或合并来进行分析。

对值进行聚合和分组

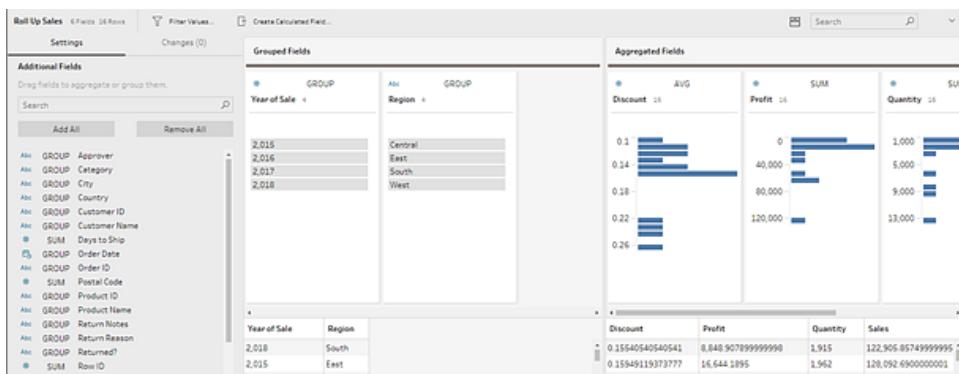
有时，您将需要调整某些数据的粒度，以便减少从流程中生成的数据量，或将数据与可能需要联接或合并在一起的其他数据对齐。例如，在将销售额表与客户表联接之前，您可能需要按客户聚合销售额数据。

如果需要调整数据的粒度，请使用“添加聚合”选项创建用于聚合或分组数据的步骤。对数据进行聚合还是分组取决于数据类型(字符串、数字或日期)。

1. 在“流程”窗格中，单击加号  图标，并选择“添加聚合”。一个新的聚合步骤将显示在“流程”窗格中，并且“配置”窗格将更新以显示聚合和分组配置。
2. 若要对字段进行分组或聚合，请将字段从左侧窗格拖到右侧窗格中的其中一个列。

您还可以：

- 在两个窗格之间中拖放字段。
- 在列表中搜索字段，并仅选择要包括在聚合中的字段。
- 双击字段，将其添加到左侧或右侧窗格。
- 更改字段的函数以将其自动添加到适当的窗格。
- 单击“全部添加”或“全部移除”以批量应用或移除字段。
- 将特定清理操作应用于字段。有关可以使用哪些清理选项的详细信息，请参见[关于清理操作 在本页 223](#)。



字段基于其数据类型分布在“分组字段”和“聚合字段”列之间。单击分组或聚合类型（例如，AVG 或 SUM）标题以更改分组或聚合类型。

在聚合和分组配置下方的数据网格中，您可以看到分组或聚合成员的示例。

系统会在“更改”窗格中跟踪对字段执行的所有清理操作。

联接数据

要分析的数据通常由与特定字段相关的表的集合组成。联接是一种在这些公共字段上合并相关数据的方法。使用联接合并数据后会产生一个通常通过添加数据字段横向扩展的表。

联接是一种可在流程中的任何地方执行的操作。在流程前期进行联接可帮助您了解您的数据集，并揭示出需要立即关注的区域。

若要创建联接，请执行以下操作：

1. 使用以下方法之一联接两个表：

- 在流程窗格中至少添加两个表，然后选择相关表并将其拖到另一个表上，直至显示“联接”选项，或者，单击 \oplus 图标，并从菜单中选择“添加联接”。



- 单击 \oplus 图标并从菜单中选择“添加联接”，然后将其他输入手动添加到联接，并添加联接子句。

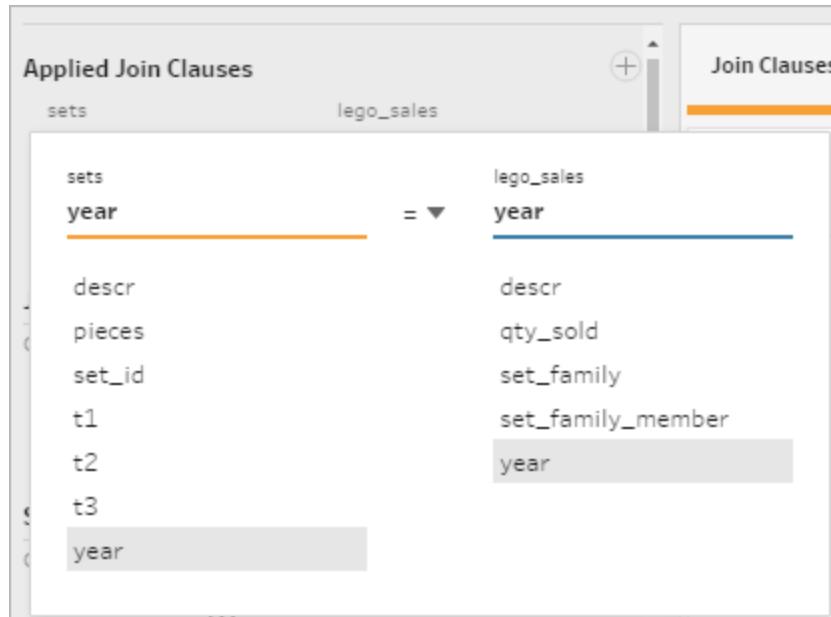
注意: 在 Tableau Prep Builder 版本 2019.1.3 及更高版本中, 如果连接到表关系已定义并包含相关字段的表, 则可以选择“添加联接”, 并从相关表列表中进行选择。Tableau Prep Builder 基于组成两个表之间的关系的字段创建联接。

有关包含表关系的连接器的详细信息, 请参见在“输入”步骤中联接数据(版本 2019.1.3 及更高版本)在本页 176。

将会向流程中添加一个新联接步骤, 并且配置窗格将更新以显示联接配置。

2. 若要查看和配置联接, 请执行以下操作:

- a. 查看“联接结果汇总”, 查看由于联接类型和联接条件的原因而包括和排除的字段数。
- b. 在“联接类型”下的维恩图中单击以指定所需联接的类型。
- c. 在“已应用联接子句”下, 单击加号 图标, 或在为默认联接条件选择的字段上指定或编辑联接子句。您在联接条件中指定的字段是联接中各表之间的共同字段。



- d. 或者, 您可以单击“联接子句建议”下显示的建议联接子句, 将子句添加到已应

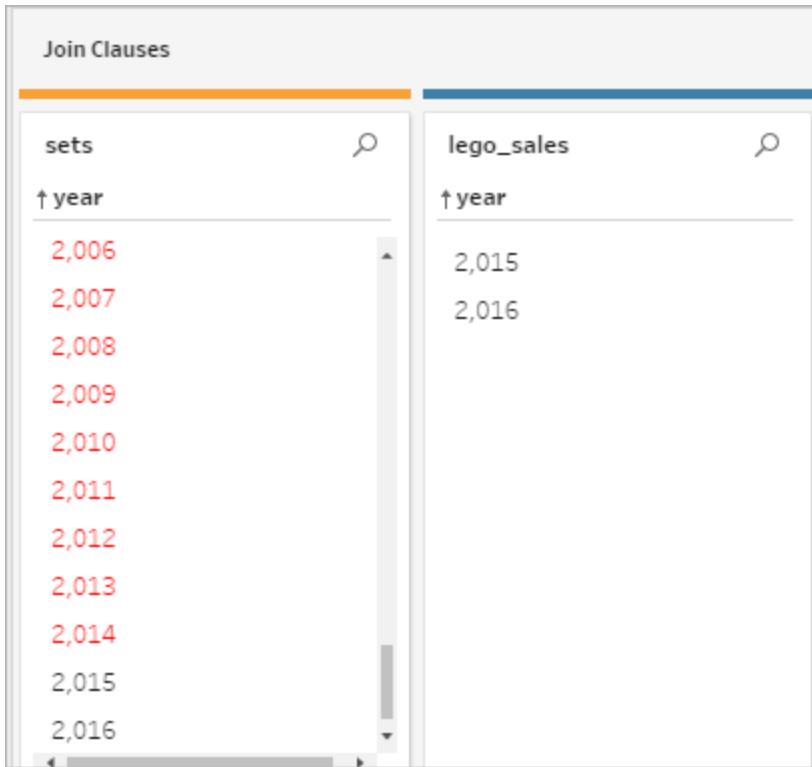
用联接子句的列表。

The screenshot shows the Tableau Prep Builder interface with the 'Join Clauses' pane open. It displays a 'Clean Notes_Approver' table and an 'All Orders' table being joined on 'Product ID'. The 'Join Type' is set to 'Right join'. The 'Summary of Join Results' section indicates 16,302 included values and 0 excluded values. The 'Join Clause Recommendations' pane lists various join clauses for the 'Product ID' field. To the right, the 'Join Results' pane shows the combined data from both tables.

检查联接的结果

联接配置中的汇总显示有关联接的元数据，帮助您验证联接是否包括预期的数据。

- 已应用联接子句:** 默认情况下，Tableau Prep Builder 基于所联接的表中的共同字段名称定义第一个联接子句。根据需要添加或移除联接子句。
- 联接类型:** 默认情况下，当您创建联接时，Tableau Prep Builder 在表之间使用内部联接。根据您连接到的数据，您也许能够使用左联接、内部联接、右联接或外部联接。
- 联接结果汇总:** “联接结果汇总”显示包括在联接的表中和从中排除的值的分布。
 - 单击每个“**已包括**”条形以隔离并查看包括在联接中的联接配置中的数据。
 - 单击每个“**已排除**”条形以隔离并查看从联接中排除的联接配置中的数据。
 - 单击**“已包括”**和**“已排除”**条形的任何组合以查看数据的累积性洞察。
- 联接子句建议:** 单击建议联接子句旁边的加号 图标以将其添加到“**已应用联接子句**”列表。
- “联接子句”窗格:** 在“**联接子句**”窗格中，您可以看到联接子句中每个字段中的值。不符合联接子句条件的值将以红色文本显示。



- “**联接结果**”窗格：如果在“**联接结果**”窗格中看到要更改的值，您可以在此窗格中编辑值。

常见联接问题

如果在联接数据后未看到预期结果，您可能需要对字段值执行一些额外的清理操作。以下问题会导致 Tableau Prep Builder 将值读取为非匹配值，并将其从联接中排除：

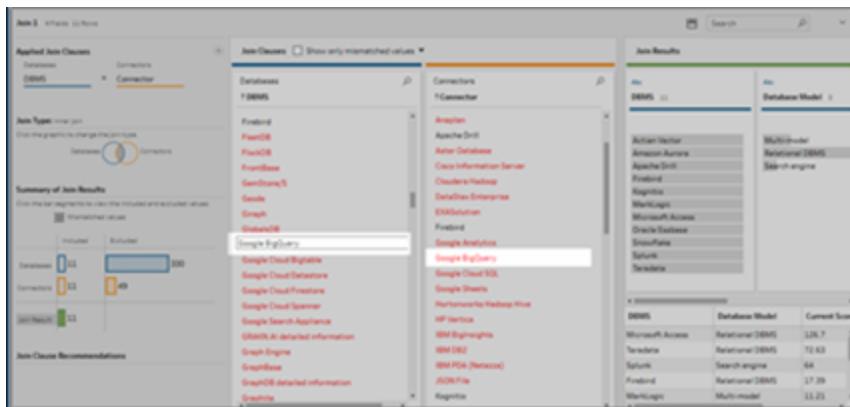
- 大小写不同：“My Sales”和“my sales”
- 拼写不同：“Hawaii”和“Hawai'i”
- 拼写错误或打字错误：“My Company Health”和“My Company Heath”
- 名称变化：“John Smith”和“John Smith Jr.”
- 缩写：“My Company Limited”和“My Company Ltd”
- 额外的分隔符：“Honolulu”和“Honolulu (Hawaii)”
- 额外的空格：这包括字符之间的额外空格、制表符空格或者额外的前导或尾随空格
- 句点的使用不一致：“Returned, not needed”和“Returned, not needed.”

好消息是，如果您的字段值存在任何这些问题，您可以直接在“联接子句”中以内联方式修复这些字段值，或者通过在“联接结果汇总”中单击“已排除”栏使用已排除的值，并在配置卡菜单中使用清理操作。

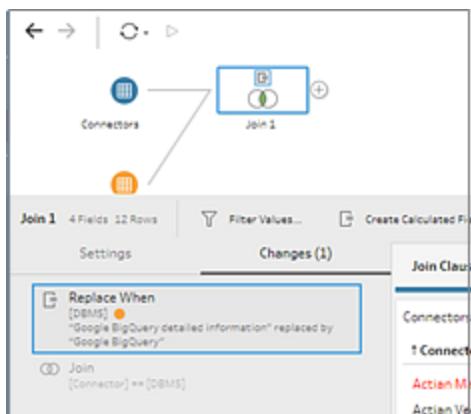
有关可在“联接”步骤中使用的不同清理选项的详细信息，请参见[关于清理操作 在本页 223](#)。

修复不匹配的字段以及更多操作

您可以直接在联接子句中修复不匹配的字段。双击或右键单击值，从要修复的字段上的上下文菜单中选择“编辑值”，并输入一个新值。系统会对您的数据更改进行跟踪，并将更改添加到“联接”步骤中的“更改”窗格。



您也可以在“联接子句”窗格中选择要保留、排除或筛选的多个值，或在“联接结果”窗格中应用其他清理操作。根据您更改的字段以及这些字段位于联接过程中的何处，将会在联接之前或之后应用您所做的更改，以便为您提供更正的结果。



有关清理字段的详细信息，请参见[应用清理操作 在本页 226](#)。

合并数据

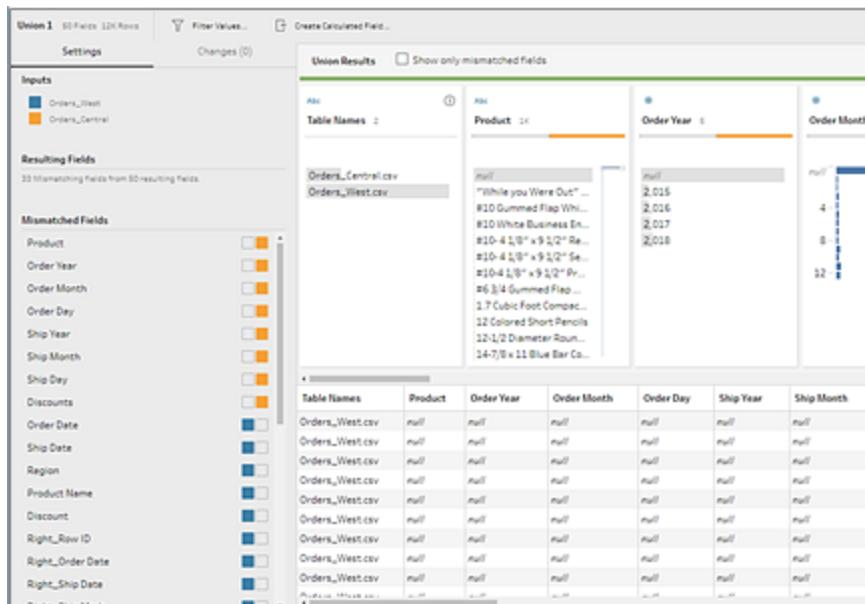
合并是一种通过将一个表的行附加于另一个表来合并数据的方法。例如，您可能需要将一个表中的新事务添加到另一个表中的过去的事务列表。确保您合并的表具有相同的字段数、相同的字段名称，并且字段的数据类型相同。

提示：为了最大程度地提升性能，一个并集最多可以有 10 个输入。如果需要合并超过 10 个文件或表，请尝试在输入步骤中合并文件。有关这种类型的合并的详细信息，请参见[在输入步骤中合并文件和数据库表 在本页 170](#)。

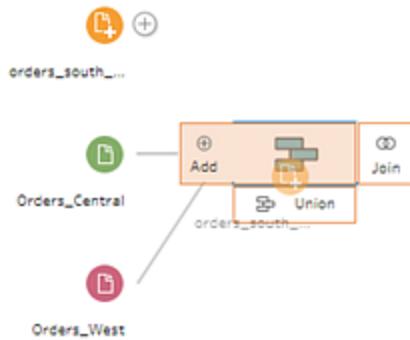
与联接类似，您可以在流程中的任何地方使用合并操作。

若要创建并集，请执行以下操作：

1. 将至少两个表添加到流程窗格之后，选择并将相关的表拖到其他表上，直至显示“并集”选项。您也可以单击 图标，并从菜单中选择“添加并集”。将会在“流程”窗格中添加一个新的并集步骤，并且“配置”窗格将更新以显示并集配置。



2. 通过将表拖到合并的表，直至看到“添加”选项，向并集中添加其他表。

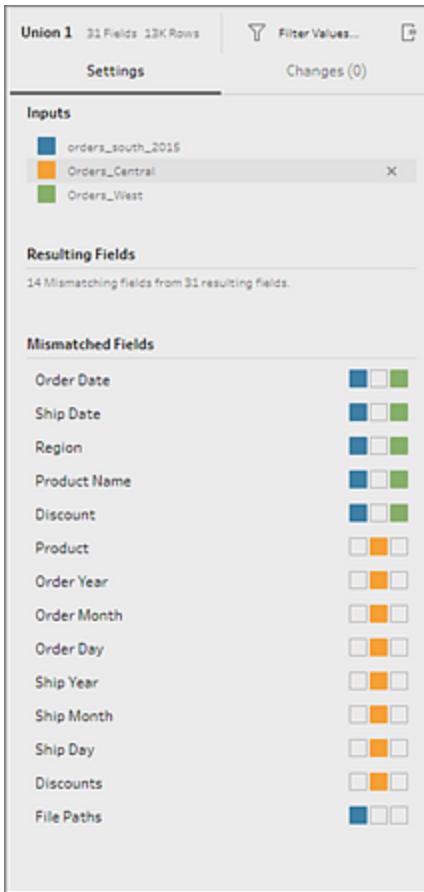


3. 在并集配置中，查看有关并集的元数据。您可以从并集中移除表，以及查看有关任何不匹配字段的详细信息。

检查并集的结果

创建并集之后，请检查并集的结果以验证并集中的数据符合预期。您可以对并集配置中的许多方面进行检查，以帮助验证并集中的数据。

- **查看并集元数据：**并集配置显示有关并集的一些元数据。在这里您可以看到组成并集的表、生成的字段数以及任何不匹配的字段。



- 查看每个字段的颜色:** 并集摘要中列出的每个字段的旁边以及并集配置中每个字段的上方有一组颜色。这些颜色与并集中中的每个表对应。

如果为该字段显示所有表颜色，则该字段的合并操作执行正确。缺少一种表颜色表明有不匹配的字段。



不匹配的字段是指可能包含类似数据但在某些方面有所不同的字段。您可以在并集摘要中看到不匹配字段的列表，以及它们所来自的表。如果想要更仔细地查看字段中的数据，请选中“仅显示不匹配字段”复选框来隔离并集配置中的不匹配字段。



若要修复这些字段，请按照下面“修复不匹配的字段”部分中的建议之一进行操作。

修复不匹配的字段

当并集中的表不匹配时，合并操作将会生成额外的字段。这些额外字段是从其对应上下文中排除的有效数据。

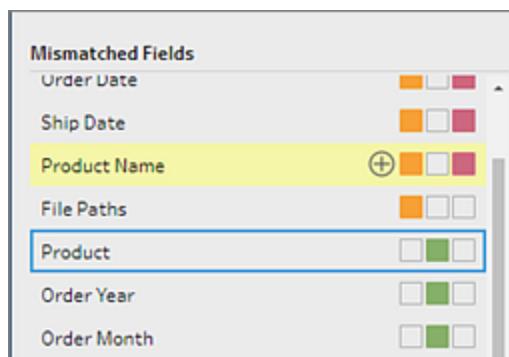
为了解决字段不匹配问题，您必须将不匹配的字段合并在一起。

字段之所以可能不匹配，有很多原因。

- 对应的字段名称不同：**如果表之间的对应字段的名称不同，您可以使用并集建议手动合并“**不匹配字段**”列表中的字段，或者重命名并集配置中的字段，将不匹配的字段合并在一起。

若要使用并集建议，请执行以下操作：

- 在“**不匹配字段**”列表中，单击某个不匹配的字段。如果存在建议的匹配项，则匹配字段将以黄色突出显示。
建议的匹配项基于具有类似数据类型和字段名的字段。
- 将光标悬停在突出显示的字段上，并单击加号图标来合并字段。

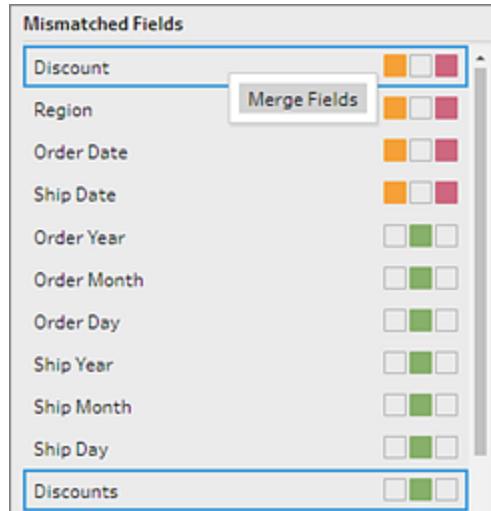


若要手动合并“**不匹配字段**”列表中的字段，请执行以下操作：

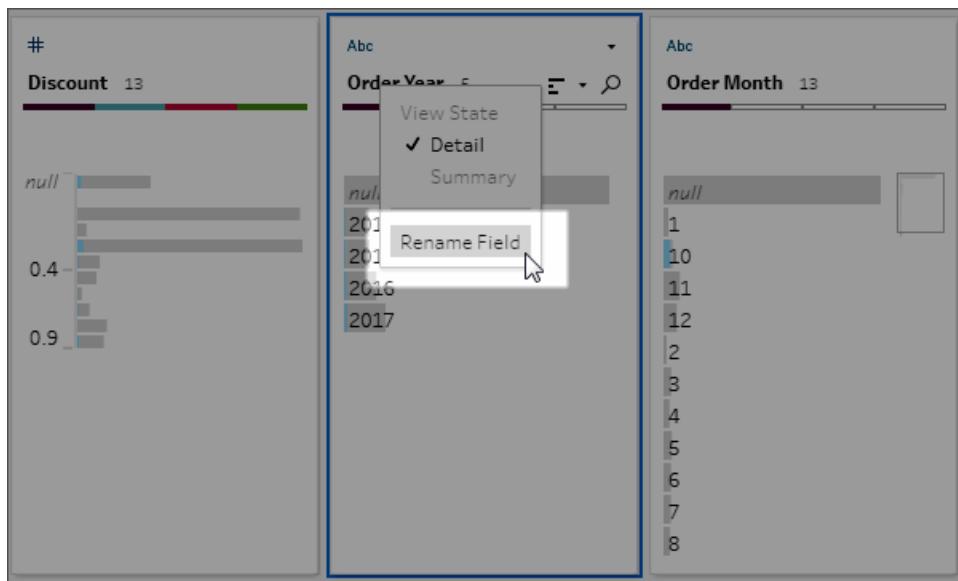
- 在列表中选择一个或多个字段。
- 右键单击所选字段，如果合并有效，“**合并字段**”菜单选项将出现。

如果在右键单击字段时看到“**无可用选项**”，这是因为字段不符合合并条件。例如，尝试合并来自同一输入的两个字段就不符合条件。

3. 单击“**合并字段**”来合并所选字段。

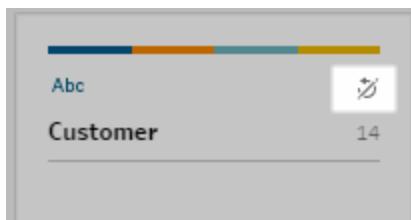


若要在并集配置窗格中重命名字段，请右键单击字段名，并单击“**重命名字段**”。



- 对应的字段名称相同，但类型不同：默认情况下，如果对应字段的名称匹配，但字段的数据类型不匹配，Tableau Prep Builder 将会更改其中一个字段的数据类型，使它们相互兼容。如果 Tableau Prep Builder 进行此更改，则会在合并的字段上用一个“更改”

“数据类型”图标加以注明。



某些情况下，Tableau Prep Builder 可能无法选取正确的数据类型。如果发生这种情况，并且您想要撤消合并，请右键单击“更改数据类型”图标，并选择“**分隔具有不同类型的数据输入**”。



然后，您可以通过以下方式再次合并字段：首先更改其中一个字段的数据类型，然后按照[其他合并字段选项 向下](#)中的建议进行操作。

- **对应的表的字段数不同：**为了合并表，并集中的每个表必须包含相同数量的字段。如果并集产生了额外的字段，请将字段合并到一个现有字段中。

其他合并字段选项

除了上面有关合并字段的部分中介绍的方法外，您还可以使用以下方法之一来合并字段。您可以在除“输出”步骤之外的任何步骤中合并字段。

有关如何合并同一文件中的字段的信息，请参见[合并字段 在本页 234](#)。

若要合并字段，请执行以下操作之一：

- 将一个字段拖放到另一个字段上。将显示“**放置以合并字段**”指示符。
- 选择多个字段，并在所选范围内右键单击以打开上下文菜单，然后单击“**合并字段**”。
- 选择多个字段，然后在上下文相关工具栏上单击“**合并字段**”。

保存和共享工作

您可以在流程中的任何位置保存流程、在 Tableau Desktop 中预览流程中的数据，或创建包括已应用于流程的所有操作的数据的数据提取。您也可以随流程一起打包您的数据，以便与共享人共享，或将数据提取以数据源形式发布到 Tableau Server 或 Tableau Online。

若要使数据保持最新状态，您可以从 Tableau Prep Builder 内或通过命令行运行单独的流程。若要按计划自动运行流程，您可以将流程发布到 Tableau Server 或 Tableau Online。有关运行流程的详细信息，请参见 [通过命令行刷新流程输出文件 在本页 321](#) 或 [将流程发布到 Tableau Server 或 Tableau Online 在本页 378](#)。

保存您的流程

保存流程以在执行任何其他操作之前备份您的工作。流程以 Tableau Prep Builder 流程 (.tfl) 文件格式保存。

您也可以随流程一起打包本地文件 (Excel、文本文件和 Tableau 数据提取) 以与其他人共享，就像在 Tableau Desktop 中打包工作簿用于共享一样。只能随流程一起打包本地文件。举例来说，不会包括来自数据库连接的数据。

在保存打包流程时，流程将保存为打包 Tableau 流程文件 (.tflx)。

- 若要保存流程，请从顶部菜单中选择“文件”>“保存”。
- 若要随流程一起打包数据文件，请从顶部菜单中执行以下操作之一：
 - 选择“文件”>“导出打包流程”
 - 选择“文件”>“另存为”。然后，在“另存为”对话框中，从“另存为类型”下拉菜单中选择“打包 Tableau 流程文件”。

在 Tableau 中查看数据样本

当您清理数据时，有时您可能希望通过在 Tableau Desktop 中查看数据来检查进度。当流程在 Tableau Desktop 中打开时，Tableau Prep Builder 会创建一个永久的 Tableau 数据提取 (.tde 或 .hyper，具体取决于 Tableau 的版本) 和一个 Tableau 数据源 (.tds) 文件。这些文件保存在 Tableau 存储库的“数据源”文件中，以便您能够随时试验数据。

注意: 尽管您可以体验数据, 但 Tableau 只会向您显示数据示例, 您将无法将工作簿另存为打包工作簿 (.twbx)。当您准备在 Tableau 处理自己的数据时, 请为流程创建一个输出步骤, 并将输出保存到文件或另存为发布的数据源, 然后在 Tableau 中连接到完整数据源。

在 Tableau Desktop 中打开流程时, 您可以看到流程中您正在处理的数据样本, 其已应用到所选步骤为止的操作。

若要在 Tableau Desktop 中查看数据样本, 请执行以下操作:

1. 右键单击要在其中查看数据的步骤, 并从上下文菜单中选择“在 Tableau Desktop 中预览”。



2. Tableau Desktop 将打开“工作表”选项卡。

创建和发布数据提取及数据源

若要创建数据提取, 请运行流程。运行流程时, 您所做的更改将应用于整个数据集。运行流程会生成 Tableau 数据源 (.tds) 和 Tableau 数据提取 (.tde 或 .hyper) 文件。您可以依据流程输出创建数据提取文件, 以便在 Tableau Desktop 中使用, 或与第三方共享数据。

从版本 2020.2.1 开始, 可以使用增量刷新运行流程。此选项可以节省时间和资源, 并使您能够仅刷新新数据, 而不是刷新完整数据集。有关如何使用增量刷新配置和运行流程的信息, 请参见[使用增量刷新来刷新流程数据](#) 在本页 314。

注意: 您可以将数据提取或数据源发布到 Tableau Server 版本 10.0 及更高版本, 以及发布到 Tableau Online。

您可以采用以下格式创建数据提取文件:

- **Tableau 数据提取 (.tde):** 数据提取保存为 Tableau 数据提取 (.tde) 和 Tableau 数据源 (.tds) 文件。如果使用 Tableau Desktop 或 Tableau Server 版本 10.0 至 10.4, 请使用此文件类型。
- **Hyper 数据提取 (.hyper):** 这是新的 Tableau 数据提取文件类型, 但只能由 Tableau Desktop 或 Tableau Server 版本 10.5 及更高版本使用。
- **逗号分隔值 (.csv):** 将数据提取保存到 .csv 文件以与第三方共享数据。导出的 CSV 文件的编码将为带 BOM 的 UTF-8。

您还可以将数据输出以数据源形式发布到 Tableau Server 或 Tableau Online, 以便共享数据并提供对您已清理、调整和合并的数据的集中访问。

注意: 为了将输出发布到 Tableau Server, 必须启用 Tableau Server REST API。有关详细信息, 请参见 Tableau REST API 帮助中的 [REST API 要求](#)。若要发布到使用安全套接字层 (SSL) 加密证书的服务器, 还要在运行 Tableau Prep Builder 的计算机上执行一些额外配置步骤。有关详细信息, 请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的 [安装之前](#)。

创建数据提取文件

1. 单击步骤上的加号图标 , 并选择“添加输出”。

如果以前运行过流程, 请单击输出步骤上的运行流程 ▷ 按钮。这将运行流程并更新输出。

“输出”窗格将打开, 并显示数据的快照。

2. 在左侧窗格中，选择“保存到文件”。
3. 单击“浏览”按钮，然后在“将数据提取另存为”对话框中输入文件的名称，并单击“接受”。
4. 在“输出类型”字段中，选择输出类型。根据您使用的 Tableau Desktop 的版本，您可以从以下选项中进行选择：
 - Tableau 数据提取 (.hyper)，适用于 Tableau Desktop 版本 10.5 及更高版本。
 - Tableau 数据提取 (.tde)，适用于 Tableau Desktop 版本 10.0 至 10.4。
 - 逗号分隔值 (.csv)(如果想要与第三方共享数据提取)。
5. (版本 2020.2.1 及更高版本) 在“写入选项”部分，查看用于将新数据写入文件并根据需要进行任何更改的默认写入选项。有关详细信息，请参见[配置写入选项 在本页 317](#)。
 - **创建表：**此选项创建新表，或将现有表替换为新输出。
 - **追加到表：**此选项将新数据添加到现有表中。如果表不存在，将创建一个新表，后续运行会将新行添加到此表。

注意：对于 .csv 和 .tde 输出类型，不支持“追加到表”。有关支持的刷新组合的详细信息，请参见[流程刷新选项 在本页 314](#)。

6. 单击“运行流程”以运行流程并生成数据提取文件。

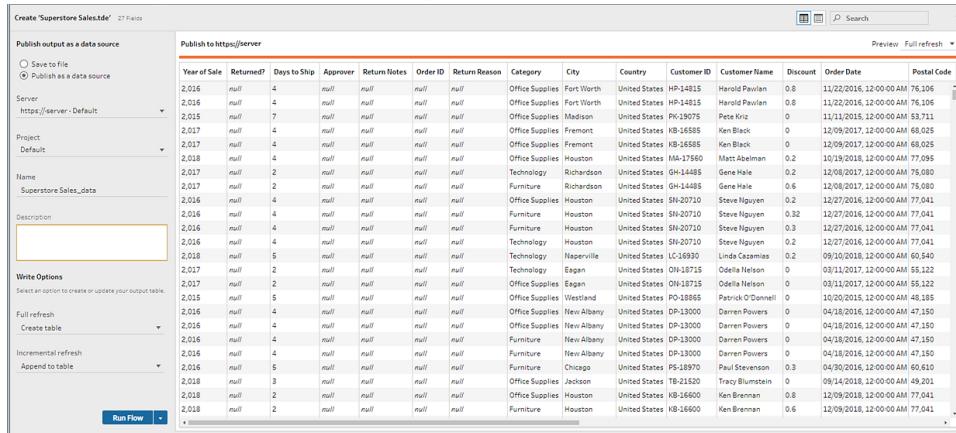
以数据源形式发布

- 单击步骤上的加号图标 ，并选择“添加输出”。

如果以前运行过流程，请单击输出步骤上的运行流程  按钮。这将运行流程并更新输出。

注意: 在版本 2020.1.1 中，Tableau Prep Builder 将刷新以前发布的数据源，并保留数据源中可能包含的任何数据建模(例如计算字段、数字格式设置等)。如果无法刷新数据源，则将改为替换数据源。

- “输出”窗格将打开，并显示数据的快照。



The screenshot shows the 'Create Superstore Sales.tde' dialog with 27 fields. The 'Publish output as a data source' section is selected. Under 'Server', 'https://server-Default' is chosen. Under 'Project', 'Default' is selected. Under 'Name', 'SuperstoreSales_data' is entered. A large preview window shows the first 20 rows of the Superstore Sales data, which includes columns like Year of Sale, Returned?, Days to Ship, Approver, Return Notes, Order ID, Return Reason, Category, City, Country, Customer ID, Customer Name, Discount, Order Date, and Postal Code. The preview also includes a 'Run Flow' button at the bottom.

- 选择“作为数据源发布”单选按钮，并填写以下字段：

- 服务器:** 选择要在其中发布数据源和数据提取的 Tableau 服务器。如果未登录到服务器，则会提示您登录。

注意: 从 Tableau Prep Builder 版本 2020.1.4 开始，一旦登录到服务器，Tableau Prep Builder 在关闭应用程序时会记住您的服务器名称和登录信息，以便下次打开应用程序时您已经登录到服务器。

在 Mac 上，可能会提示您提供对 Mac 密钥链的访问权限，以便 Tableau Prep Builder 能够安全地使用 SSL 证书来连接到 Tableau Server 或 Tableau Online 环境。

如果要输出到 Tableau Online，请在“serverUrl”中包括您的站点托管于其上的 pod。例如，“<https://eu-west-1a.online.tableau.com>”，而不是“<https://online.tableau.com>”。

- **项目**: 选择要在其中加载数据源和数据提取的项目。
- **名称**: 输入文件名
- **说明**: 输入数据源的说明。

4. (版本 2020.2.1 及更高版本) 在“**写入选项**”部分，查看用于将新数据写入文件并根据需要进行任何更改的默认写入选项。有关详细信息，请参见[配置写入选项 在本页 317](#)

- **创建表**: 此选项创建新表，或将现有表替换为新输出。
- **追加到表**: 此选项将新数据添加到现有表中。如果表不存在，将创建一个新表，后续运行会将新行添加到此表。

5. 单击“**运行流程**”以运行流程并发布数据源。

使用增量刷新来刷新流程数据

从版本 2020.2.1 开始，您可以将流程中的选定输入和输出配置为以增量方式刷新，以便在流程运行时仅检索和处理新行，从而节省时间和资源。举例来说，如果流程包含每天更新的事务数据，则可以设置增量刷新以便每天仅检索和处理新事务，然后每周或每月运行一次完全刷新以刷新所有流程数据。

若要使用增量刷新来运行流程，Tableau Prep Builder 需要以下信息：

- 用于检测输入表中的新行的字段。
- 用于将流程输出中最后一个处理的值与输入中的值进行比较以确定哪些行是新行的字段。
- 您希望如何将新数据写入表。您可以将新数据添加到现有表，也可以用新数据覆盖表数据。

流程刷新选项

Tableau Prep Builder 使您能够选择如何刷新数据，以及如何使用流程输出更新表。下表描述了不同的选项及其优点。

刷新组合	处理	表更新	优点
-------------	-----------	------------	-----------

的数据			
完全刷新 + 创建表	所有	使用完整数据集创建或覆盖现有表。	在每次流程运行时跟踪所有数据。
完全刷新 + 追加到表	所有	将新数据添加到现有表。	在每次流程运行时跟踪新数据和现有数据。 “追加到表”不可用于 .csv 或 .tde 输出类型。
增量刷新 + 创建表	仅新行	使用新数据创建或覆盖现有表。	仅保留最新数据作为完整数据集。
增量刷新 + 追加到表	仅新行	将新数据添加到现有表。	将新数据添加到现有增量数据集进行分析。 “追加到表”不可用于 .csv 或 .tde 输出类型。

配置增量刷新

若要将流程配置为使用增量刷新，需要在要使用此选项的“输入”步骤和“输出”步骤上同时指定设置。在“输入”步骤中，指定 Tableau Prep Builder 如何查找新行。在“输出”步骤中，指定如何将新行写入表。运行流程时，可以选择完全或增量刷新类型。

1. 在流程窗格中，选择要为增量刷新配置的输入步骤。
2. 在“设置”选项卡的“输出”窗格中，在“设置设置增量刷新”部分下设置以下选项：
 - 选择“启用”。
 - 使用字段标识新行：选择要在输入数据中刷新的字段。必须为此字段分配“数字(整数)”、“日期”或“日期或时间”数据类型。目前，您只能选择单个字段。

注意：您可以稍后在流程中移除或重命名此字段，只要在“输出中的字段名称”中指定的字段可用于将此字段与最新输出进行比较来查找新行。

- **输出：**选择与您的输入相关的输出，其中包括将用于比较的字段。
- **输出中的字段名称：**选择要用于将流程输出中最后处理的值与输入中的值进行比较来查找新行的字段。此字段的数据类型必须与在“使用字段标识新行”字段

中指定的字段的数据类型相同。

The screenshot shows the 'Input' settings page in Tableau Prep. At the top, there are tabs: 'Settings' (which is selected), 'Multiple Files', 'Data Sample', and 'Changes (0)'.
Connection: Textfile, Orders_Central.csv, Edit, Original Table Name: Orders_Central.
Text Options: First line contains header (radio button selected), Generate field names automatically (radio button unselected).
Field Separator: Comma.
Text Qualifier: Automatic.
Character Set: (dropdown menu).
Locale: English (United States).
Set up Incremental Refresh: A section with a note: 'Get the latest rows for a specific field value when the values have changed since the flow was last run.' It includes an 'Enable' checkbox (checked), 'Identify new rows using field' dropdown set to '# Row ID', and 'Output' dropdown set to 'Create 'Superstore Sales''. Below this, a note says: 'Select the output and field with the last processed value for field "Row ID".'
Field name in output: '# Row ID'.

配置写入选项

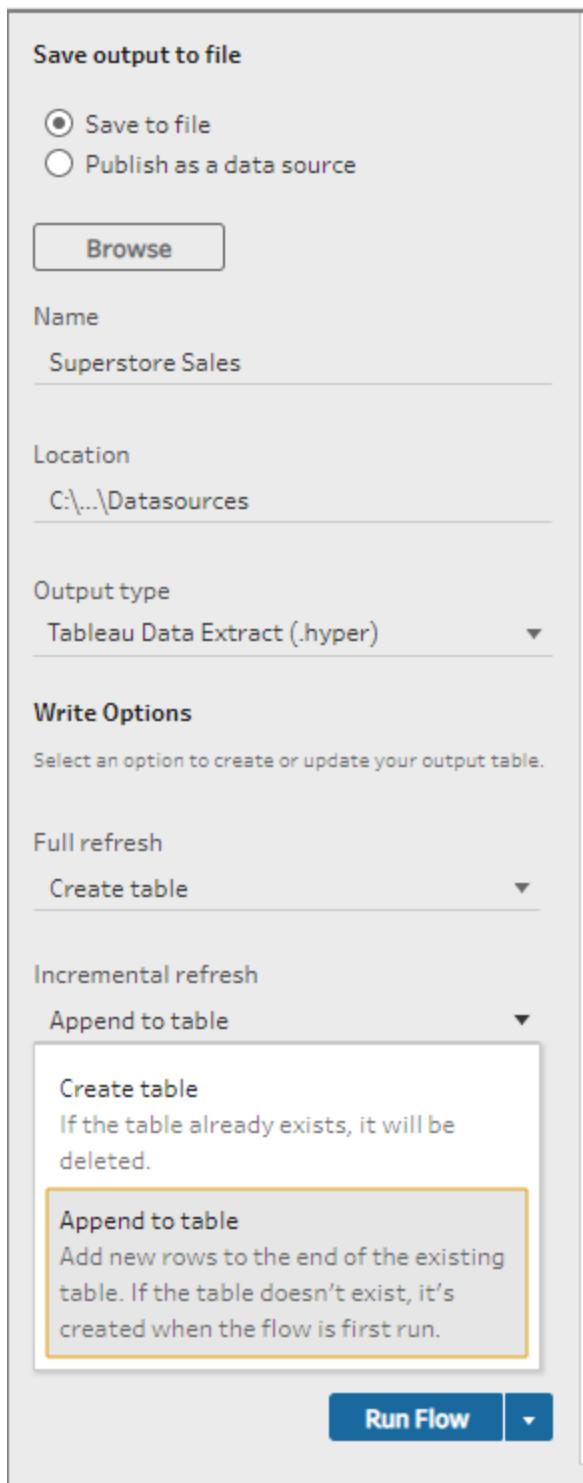
若要完成增量刷新设置，请设置输出“**写入选项**”以指定如何将新行写入表。与配置的输入步骤相关的所有输出都选择了默认写入选项，但您可以将其更改为支持的选项。

默认情况下，针对本地或已发布的 .hyper 数据提取的输出设置为“**追加到表**”。针对 .csv 或 .tde 文件的输出设置为“**创建表**”。

1. 在流程窗格中，选择要为增量刷新配置的输出步骤。
2. 在“输出”窗格中的“**写入选项**”部分，查看默认写入选项并根据需要进行任何更改。
 - **创建表**: 此选项创建新表，或将现有表替换为新输出。
 - **追加到表**: 此选项将新数据添加到现有表中。如果表不存在，则在首次运行流程时创建一个新表，后续运行会将新行添加到此表。

注意: 对于 .csv 和 .tde 输出类型，不支持“**追加到表**”。有关支持的刷新组合的

详细信息，请参见[流程刷新选项](#) 在本页314。



运行流程

您可以在 Tableau Prep Builder 中或通过命令行使用增量刷新来运行单个流程。有关通过命令行运行流程的信息，请参见[在启用增量刷新的情况下运行流程\(版本 2020.2.1 及更高版本\) 在本页 327](#)。

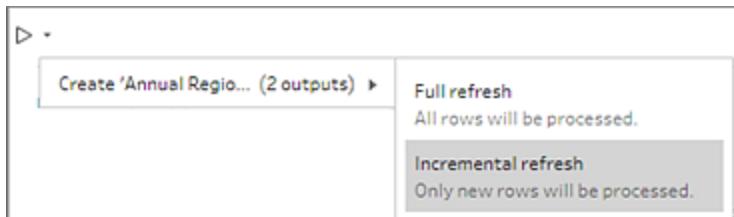
如果在服务器上启用了 Tableau Prep Conductor (数据管理加载项的一部分)，则可以通过在 Tableau Server 或 Tableau Online 中设置计划来使用增量刷新运行流程。有关按计划运行流程的信息，请参见[计划流程任务 在本页 396](#)。

注意: 写入选项在 Tableau Prep Builder 中设置，在 Tableau Server 或 Tableau Online 中运行流程时，无法对其进行更改。

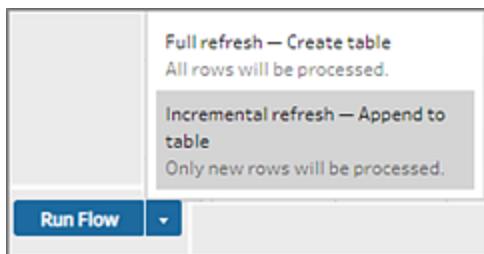
如果找不到现有输出，无论您选择什么运行选项，Tableau Prep Builder 都会对所有输出运行完全刷新。后续流程运行使用增量刷新过程，并仅检索和处理新行，除非缺少增量刷新配置数据，或者现有输出已移除。

若要使用增量刷新在 Tableau Prep Builder 中运行流程，请从以下位置之一选择“**增量刷新**”：

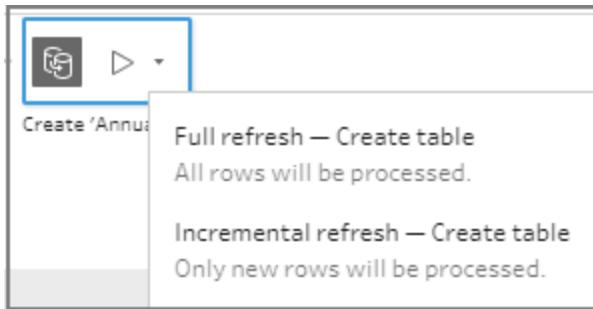
- 从顶部菜单中，单击“运行”按钮上的下拉选项。



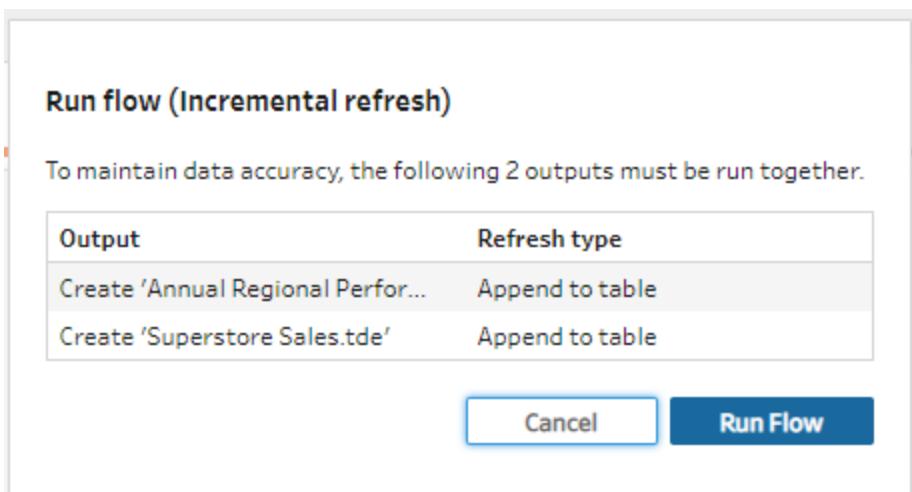
- 从“输出”窗格中，单击“运行流程”按钮上的下拉选项。



- 从“流程”窗格中，单击“输出”步骤旁边“运行”按钮上的下拉选项。



如果启用了增量刷新的一个输入与多个输出关联，这些输出必须一起运行，并且必须使用相同的刷新类型。在 Tableau Prep Builder 中运行刷新时，将显示一个对话框，告知您必须同时运行这两个输出。



通过命令行刷新流程输出文件

可以通过命令行运行流程以刷新流程输出，而不是从 Tableau Prep Builder 中运行流程。可以使用此方法一次运行一个流程。此选项在安装了 Tableau Prep Builder 的 Windows 和 Mac 计算机上都可用。

连接器限制：

- JDBC 或 ODBC 连接器：从版本 2019.2.3 开始，包含这些连接器的流程可以通过命令行运行。
- 云连接器：包含云连接器(例如 Google BigQuery)的流程无法通过命令行运行。作为替代，请手动运行流程，或者在 Tableau Server 或 Tableau Online 中使用 Tableau Prep

Conductor 按计划运行流程。有关详细信息，请参见[使流程数据保持最新 在本页 345。](#)

对于 Windows 计算机，您也可以使用 Windows 任务计划程序来计划此过程。有关详细信息，请参见 Microsoft 联机帮助中的[任务计划程序](#)。

当您通过命令行运行流程时，Tableau Prep Builder 将使用 Tableau Prep Builder 中指定的输出步骤设置刷新流程的所有输出。有关如何指定输出位置的信息，请参见[创建和发布数据提取及数据源 在本页 310](#)。有关设置写入选项(版本 2020.2.1 及更高版本)的信息，请参见[配置写入选项 在本页 317](#)。

运行流程之前

若要通过命令行运行流程，您需要运行流程的计算机上的管理员权限，并且将需要以下信息：

- Tableau Prep Builder 的安装路径。
- 如果连接到数据库并将输出文件发布到服务器，则需要一个包含所有必需凭据的凭据.json 文件。
- Tableau 流程(.tfl)文件所在的路径。

凭据.json 文件要求

注意：如果流程连接到并输出至本地文件、网络共享上存储的文件或使用 Windows 身份验证 (SSPI) 的输入文件，则不需要凭据.json 文件。有关 Windows 身份验证的详细信息，请参见 Microsoft 联机帮助中的[SSPI 模型](#)。

下表列出了您需要包括在凭据.json 文件中的凭据。

输入连接	输出位置
<ul style="list-style-type: none">• hostname(服务器名称)• contentUrl(已发布数据源)站点 ID。当您登录到 Tableau Server 或 Tableau Online 时，此信息出现在 URL 中的 /site/ 之后• port(端口 ID)• username	<ul style="list-style-type: none">• serverUrl• contentUrl (站点 ID)。当您登录到 Tableau Server 或

<ul style="list-style-type: none">• Password	<p>Tableau Online 时, 此信息出现在 URL 中的 /site/ 之后)</p> <ul style="list-style-type: none">• username• password
--	---

- 如果计划重用数据, 请将其放在不会被 Tableau Prep Builder 安装过程覆盖的文件夹中。
- 如果要运行连接到数据库文件、已发布数据源, 或将输出发布到服务器的流程, 则 .json 文件必须包括连接到这些位置所需的凭据。
- 运行该过程时, 主机名、端口和用户名用于在 Tableau 流程文件 (.tfl) 中查找匹配的连接, 并在运行该过程之前更新。“端口 ID”和“站点 ID”是可选的(如果连接不需要此信息)。
- 如果连接到已发布数据源, 请在输入连接中包括主机名、连接 URL 和端口(对于 http 为 80, 对于 https 为 443)。需要主机名以在 Tableau 流程文件 (.tfl) 中查找匹配的连接, 并且内容 URL 和端口用于建立与服务器的连接。
- 如果连接到 Tableau Online, 请在要连接到的 pod 的输入连接中包括端口(80 或 443)。有关 Tableau Online 的详细信息, 请参见 Tableau Online 帮助中的 [Tableau Bridge 与 Tableau Online 的连接](#)。

凭据 .json 文件示例

此部分显示可以使用凭据 json 要求创建的凭据文件的两个不同示例。

连接到已发布数据源

此示例显示一个 .json 凭据文件, 该文件连接到已发布数据源并将数据输出到包含站点 ID 的服务器

注意: 如果 inputConnection 或 outputConnection 使用“默认”站点, 则将 ContentUrl 设置为空白。例如, "contentUrl": ""

```
{
  "inputConnections": [
    {
      "username": "jsmith",
      "contentUrl": "my-testing-site",
      "hostname": "https://my-server",
      "port": 443,
      "password": "passw0rd$"
    }
  ],
  "outputConnections": [
    {
      "serverUrl": "https://my-server",
      "contentUrl": "my-testing-site",
      "username": "jsmith",
      "password": "passw0rd$"
    }
  ]
}
```

连接到两个数据库

此示例显示一个 `.json` 凭据文件，该文件连接到 MySQL 和 Oracle，并将数据输出到包含站点 ID 的服务器。

```
{
  "inputConnections": [
    {
      "username": "jsmith",
      "hostname": "mysql.example.lan",
      "port": 1234,
      "password": "passw0rd"
    },
    {
      "username": "jsmith",
      "hostname": "Oracle.example.lan",
      "port": 5678,
      "password": "passw0rd"
    }
  ]
}
```

```
    "password": "passw0rd"
  }
],
"outputConnections": [
{
  "serverUrl": "http://MyServer",
  "contentUrl": "FinanceTeam",
  "username": "jsmith",
  "password": "passw0rd$"
}
]
```

创建凭据文件的相关提示

如果使用远程连接, Tableau Prep Builder 将使用流程文件和凭据 .json 文件中的信息来运行流程。例如, 远程连接的数据库名称和输出文件的项目名称来自于流程, 服务器名称和登录凭据来自于凭据 .json 文件。

为了避免在运行流程时出现错误, 请确保凭据文件遵循以下准则:

- 如果使用 Tableau Prep Builder 版本 2018.2.2 至 2018.3.1, 则始终包括“inputConnections”和“outputConnections”数组, 即使流程不为输入或输出使用远程连接。只需将这些数组保留为空白。

如果使用 Tableau Prep Builder 版本 2018.3.2 及更高版本, 则无需包括空白数组。

- 没有远程输入连接? 在 .json 文件的顶部包括此语法

```
{
  "inputConnections": [
  ],
}
```

- 没有远程输出连接? 在 .json 文件的底部包括此语法

```
  "outputConnections": [
  ]
}
```

- 输入连接是否没有端口 ID, 或者在服务器名称中指定了端口?

如果连接没有端口 ID，请不要 .json 文件中包括 "port":xxxx, .引用，甚至不要包括 "port": ""。如果端口 ID 包含在服务器名称中，请在主机名中包括端口 ID。例如, "hostname": "mssql.example.lan,1234"

- 引用 "serverUrl": 时，不要在地址结尾包括"/"。例如，使用此 "serverUrl": "http://server"，而不是此 "serverUrl": "http://server/"。
- 如果有多个输入或输出连接，请为文件中的每个连接包括凭据。
- 如果连接到已发布的数据源，请确保在输入连接中包括 hostname 和 contentUrl。

运行流程

重要提示: 下面的示例将“Tableau Prep”版本 2019.1.2 的名称更改为“Tableau Prep Builder”。如果您使用的是产品的更低版本，请改用“Tableau Prep”。

1. 以管理员身份打开命令提示符或终端命令提示符 (MacOS)。
2. 使用如下所示语法运行以下命令之一。
 - 流程连接到本地文件或网络共享上存储的文件，并发布至本地文件、网络共享上存储的文件，或使用 Windows 身份验证：

注意: 如果连接到或输出至网络共享上存储的文件，请为路径使用 UNC 格式:\服务器\路径\文件名。它不能通过密码保护。

Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep  
Builder <version>\scripts"\tableau-prep-cli.bat -t  
"path\to\[your flow file name].tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\[Tableau Prep  
Builder version].app/Contents/scripts./tableau-prep-cli  
-t path/to/[your flow file name].tfl
```

- 流程连接到数据库或发布至服务器：

Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep  
Builder <version>\scripts"\tableau-prep-cli.bat -c  
"path\to\[your credential file name].json" -t "path\to\  
[your flow file name].tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\[Tableau Prep  
Builder version].app/Contents/scripts./tableau-prep-cli  
-c path/to/[your credential file name].json -t path/to/  
[your flow file name].tfl
```

- 流程文件或凭据文件存储在网络共享上(为路径使用 UNC 格式:\服务器\路径\文件名):

Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep  
Builder <version>\scripts"\tableau-prep-cli.bat -c  
"\server\path\[your credential file name].json" -t  
"\server\path\[your flow file name].tfl"
```

Mac: 将网络共享映射到 Finder 中的 **/Volumes** 以使其保持持久, 然后使用 **/Volumes/.../[your file]** 指定路径:

```
/Applications/Tableau\ Prep\ Builder\[Tableau Prep  
Builder version].app/Contents/scripts./tableau-prep-cli  
-c /Volumes/.../[your credential file name].json -t  
path/to/[your flow file name].tfl
```

有关示例命令, 请参见[语法示例 在本页 330](#)。有关常见错误和解决方法, 请参见[使用命令行运行流程时的常见错误 在本页 471](#)。

在启用增量刷新的情况下运行流程(版本 2020.2.1 及更高版本)

如果在服务器上未启用 Tableau Prep Conductor 来计划流程运行, 则可以通过命令行中使用增量刷新来运行流程。只需在命令行中包括 **--incrementalRefresh** 参数, 如下例所示。

Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep Builder
<version>\scripts"\tableau-prep-cli.bat --incrementalRefresh -t
"path\to\[your flow file name].tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\[Tableau Prep Builder
version].app/Contents/scripts/.tableau-prep-cli --
incrementalRefresh -t path/to/[your flow file name].tfl
```

如果流程中的输入步骤启用了增量刷新，并且增量刷新参数配置正确，Tableau Prep Builder 将执行以下操作：

- 流程中启用了增量刷新的所有输入都将使用增量刷新运行所有相应的输出。
- 如果流程中没有输入启用了增量刷新，则将使用完全刷新运行所有输出。一条消息将显示刷新方法详细信息。
- 如果流程中的某些输入启用了增量刷新，则相应的输出将使用增量刷新运行。其他输出将使用完全刷新运行，一条消息将显示刷新方法详细信息。

有关配置流程以使用增量刷新的详细信息，请参见[使用增量刷新来刷新流程数据](#) 在本页 314

命令选项

如果要查看帮助选项，请在命令行中包括 -h。

命令选项	说明	说明
-c, -- connecti- ons <arg>	凭 据 文 件 的 连 接 路 径- 。	需要凭据文件所在的路径。

Tableau Prep 帮助

-d, --debug	调试 流 程 进 程 。	包括此选项以查看更多信息，来帮助调试刷新流程的问题。日志文件存储位置:My Tableau Prep Builder Repository\Command Line Repository\Logs
-dsv, --disableSslValidation	禁用 SS- L 验 证 (M- ac- O- S)	使用 MacOS 上的命令行运行流程时，可能会显示一个对话框，询问钥匙串用户和密码。从 Tableau Prep Builder 版本 2019.3.2 开始，您可以传入此附加参数来禁用此钥匙串对话框。例如:/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder version].app/Contents/scripts./tableau-prep-cli -dsv -c path/to/[your credential file name].json -t path/to/[your flow file name].tfl
-h, --help	查看 语法 选项 的 帮 助 。	此帮助选项或语法错误显示以下信息: usage: tableau-prep-cli [-c <arg>] [-d] [-h] [-t <arg>] -c,--connections <arg> Path to a file with all connection information -d,--debug This option is for debugging -dsv,--disableSslValidation Disable SSL validation -h,--help Print usage message -inc,--incrementalRefresh Run incremental refresh for all outputs that are configured to

		<p>support it</p> <p><code>-t, --tf1File <arg></code> The Tableau Prep Builder flow file</p>
<code>-inc, -- incrementalRefre- sh</code>	为配 置 为使 用 它 的所 有输 出运 行增 量刷 新- 。	<p>包括此选项，以便为配置为使用它的所有输入运行增量刷新。增量刷新使 Tableau Prep Builder 能够仅检索和处理新行，而不是流程中的所有行。</p> <p>输入步骤上的增量刷新配置设置确定哪些流程输出可以增量方式运行。所有其他输出将使用完全刷新运行，一条消息将显示刷新方法详细信息。</p> <p>有关使用增量刷新运行流程的详细信息，请参见使用增量刷新来刷新流程数据在本页314。</p>
<code>-t, -- tf1File <arg></code>	.tf1 流 程 文 件	需要 .tf1 流程文件所在的路径。

语法示例

下面的命令行显示使用以下条件运行流程的四个不同示例：

- **Tableau Prep Builder 版本: 2020.2.2**

重要信息: 以下示例将 Tableau Prep 版本 2019.1.2 的名称更改为 Tableau Prep Builder。如果您使用的是产品的更低版本, 请改用“Tableau Prep”。

- **流程名称:** Flow1.tfl
- **流程位置:** C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows
- **凭据文件名:** Flow 1.json
- **凭据文件位置:** C:\Users\jsmith\Desktop\Flow credentials
- **存储在网络共享上的凭据文件位置:** \tsi.lan\files\Flow credentials

流程连接到并发布至本地文件

Windows

```
"\Program Files\Tableau\Tableau Prep Builder  
2020.2.2\scripts"\tableau-prep-cli.bat -t  
"\C:\Users\jsmith\Documents\My Tableau Prep Builder  
Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\  
2020.2.2.app/Contents/scripts/. ./tableau-prep-cli -t  
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\  
Repository/Flows.Flow1.tfl
```

流程连接并发布到本地文件, 并为增量刷新使用短形式

Windows

```
"\Program Files\Tableau\Tableau Prep Builder  
2020.2.2\scripts"\tableau-prep-cli.bat -inc -t  
"\C:\Users\jsmith\Documents\My Tableau Prep Builder  
Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\  
2020.2.2.app/Contents/scripts/. ./tableau-prep-cli -inc -t
```

```
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\  
Repository/Flows.Flow1.tfl
```

流程连接到数据库并发布至服务器

Windows

```
"\Program Files\Tableau\Tableau Prep Builder  
2020.2.2\scripts"\tableau-prep-cli.bat -c  
"\C:\Users\jsmith\Desktop\Flow credentials\Flow1.json" -t  
"\C:\Users\jsmith\Documents\My Tableau Prep Builder  
Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\  
2020.2.2.app/Contents/scripts/.tableau-prep-cli -c  
/Users/jsmith/Desktop/Flow\ credentials/Flow1.json -t  
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\  
Repository/Flows.Flow1.tfl
```

流程发布至服务器，并且凭据文件存储在网络共享上

Windows

```
"\Program Files\Tableau\Tableau Prep Builder  
2020.2.2\scripts"\tableau-prep-cli.bat -c "\\\tsi.lan\files\Flow  
credentials\Flow1.json" -t "\C:\Users\jsmith\Documents\My Tableau  
Prep Builder Repository\Flows\Flow1.tfl"
```

Mac

```
/Applications/Tableau\ Prep\ Builder\  
2020.2.2.app/Contents/scripts/.tableau-prep-cli -c  
/Volumes/files/Flow\ credentials/Flow1.json -t  
/Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\  
Repository/Flows.Flow1.tfl
```

与 Tableau Prep 的版本兼容性

Tableau Prep Builder 经常会发布新版本，通常每月发布一次。如果新版本中引入了新功能或连接器，并且您在旧版 Tableau Prep Builder 中工作，则尝试打开流程时可能会出现兼容性问题。

同样，如果将流程发布到 Tableau Server 或 Tableau Online 以将它们计划为使用 Tableau Prep Conductor 运行，并且流程包括的新功能或连接器在您的 Tableau Server 或 Tableau Online 版本中不受支持，则可能会遇到兼容性问题，使您无法在计划和运行流程。

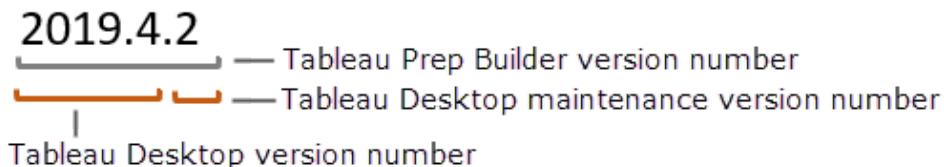
注意: 若要下载特定版本的 Tableau Prep Builder，请打开[产品下载和版本说明](#)页面，并从页面左侧的列表中选择“**Tableau Prep Builder**”。

查找我们的版本

按照以下说明查找产品版本号。

Tableau Prep Builder

Tableau Prep Builder 发布版本号类似于 Tableau Desktop 和 Tableau Server 维护版本，例如 2019.4.2，只是 Tableau Prep Builder 不会发布维护版本。



每个发布版本表示产品的最新版本，包括任何新功能和更新。因此，Tableau Prep Builder 版本 2019.4.2 可能与 Tableau Server 版本 2019.4.2 不完全兼容，但与 Tableau Server 的下一个主要版本(例如 2020.1)兼容。

若要查找产品的发布版本，请打开 Tableau Prep Builder，然后在顶部菜单中执行以下操作之一：

- **Windows:** 在顶部菜单中，单击“帮助”>“关于 Tableau Prep Builder”或“关于 Tableau Prep”，具体情况视您的版本而定。

- Mac: 在顶部菜单中, 单击“**Tableau Prep Builder**”>“关于 Tableau Prep Builder”或“**Tableau Prep**”>“关于 Tableau Prep”, 具体取决于您的版本。

发布版本号显示在对话框的左下角。

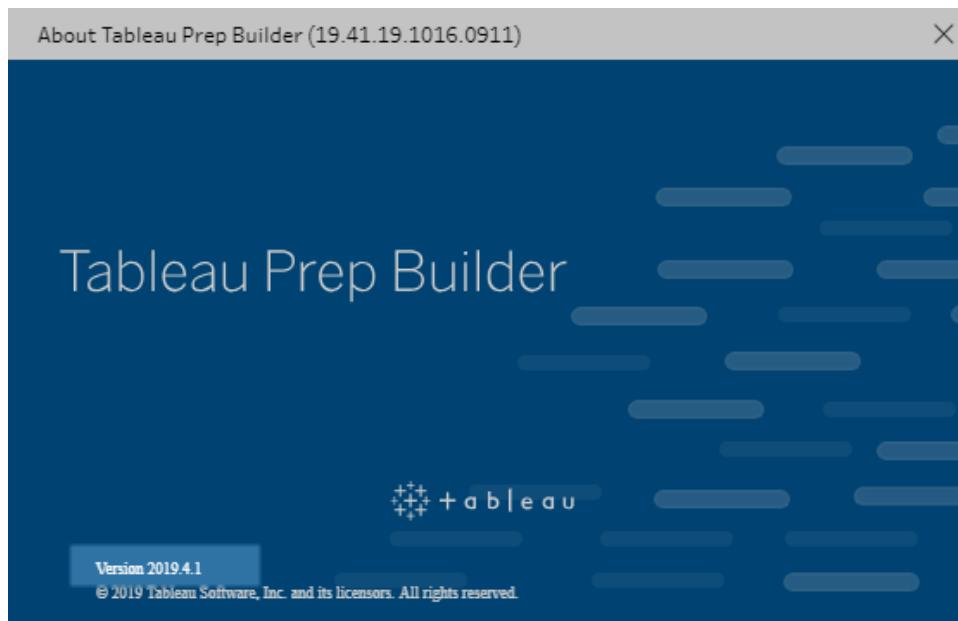


Tableau Server

Tableau Prep Conductor 是作为 Tableau Server 版本 2019.1 的数据管理加载项的一部分引入的。若要发布流程并将流程计划为在 Tableau Server 上运行, 您必须使用 Tableau Server 版本 2019.1 或更高版本, 并且必须启用 Tableau Prep Conductor。

若要查找 Tableau Server 的版本, 请在 Web 浏览器中打开 Tableau Server。在顶部菜单栏

中, 单击右上角的信息图标 , 并选择“关于 Tableau Server”。将打开一个对话框, 告知您使用的 Tableau Server 的版本。有关如何启用 Tableau Prep Conductor 的信息, 请参见[步骤 2: 配置 Tableau Server 的流程设置](#) 在本页 369。

Tableau Prep 帮助

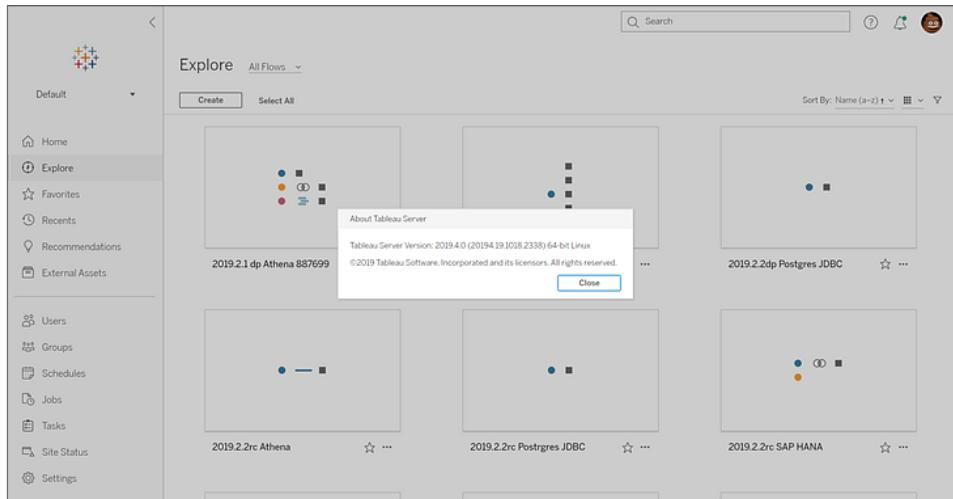


Tableau Online

Tableau Prep Conductor 是作为 Tableau Online 版本 2019.3 的数据管理加载项的一部分引入的。若要发布流程并将流程计划为在 Tableau Online 上运行，您必须使用 Tableau Online 版本 2019.3 或更高版本，并且必须启用 Tableau Prep Conductor。

若要查找 Tableau Online 的版本，请在 Web 浏览器中打开 Tableau Server。在顶部菜单栏中，单击右上角的信息图标，并选择“关于 Tableau Online”。将打开一个对话框，告知您使用的 Tableau Online 的版本。有关启用 Tableau Prep Conductor 的信息，请参见在 [Tableau Online 站点上启用 Tableau Prep Conductor 在本页 377](#)。

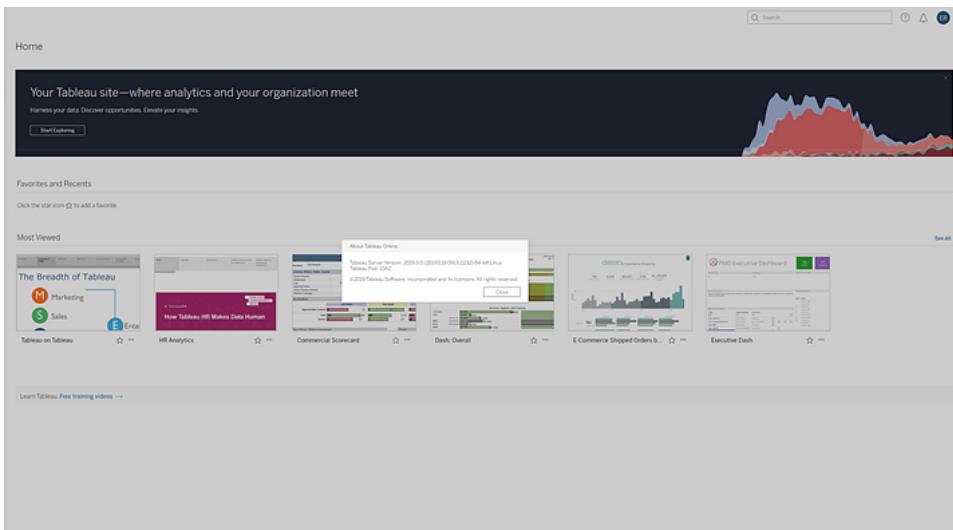


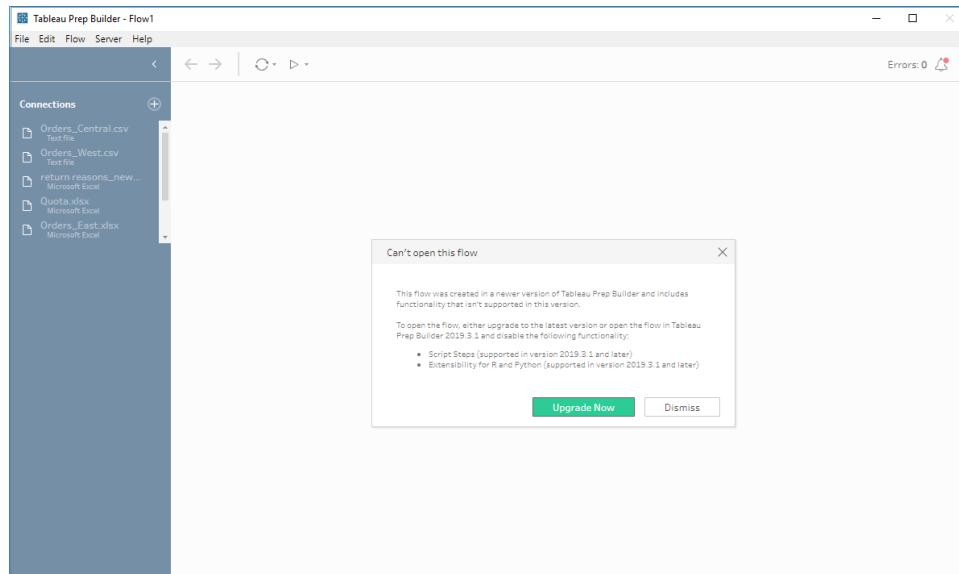
Tableau Prep Builder 不同版本之间的兼容性

通常，Tableau Prep Builder 的新版本可打开在旧版本中创建的流程。但是，当您尝试在新版本的 Tableau Prep Builder 中打开旧版本的流程，以及尝试在旧版本中打开新版本的流程时，或者甚至在使用不同的计算机在相同版本的 Tableau Prep Builder 中打开流程时，都可能会影响兼容性。

例如：

- 流程包括的输入连接器或功能在打开流程的版本中不受支持。
- 用于打开流程的计算机没有安装所需的输入连接器，或其具有的连接器驱动程序版本不兼容。Tableau Prep Builder 需要安装 64 位驱动程序才能处理流程输入连接器。

如果在尝试打开流程时出现兼容性问题，流程可能会打开，但包含错误，或者流程根本无法打开，并且您收到错误消息。在下面的示例中，流程不会打开，并且会显示一条错误消息，并列出不兼容的功能以及用于解决问题的选项。



修复 Tableau Prep Builder 的兼容性问题

若要修复兼容性问题，请尝试以下方法之一：

- 升级到 Tableau Prep Builder 的最新版本。

单击“发现”窗格底部的更新按钮，下载产品的最新版本，并按照 Tableau Desktop 和 Tableau Prep Builder 部署指南中的说明[安装 Tableau Prep Builder](#)。如果您对“发现”

窗格上的更新按钮没有访问权限，则有关如何下载产品最新版本的说明包括在“**安装 Tableau Prep Builder**”主题中。

- 确保您的计算机与 Tableau Prep Builder 兼容。例如，确保您为流程使用的连接器安装 64 位驱动程序。若要安装驱动程序，请参见[驱动程序下载](#)页面。
- 打开移除了不兼容功能的流程副本。

不同版本的 Tableau Prep Builder 和 Tableau Server 之间的兼容性

从较新版本的 Tableau Prep Builder 发布到较旧版本的 Tableau Server 可能会导致兼容性问题。例如，Tableau Prep Builder 版本 2019.4.2 中添加的新功能可能与 Tableau Server 版本 2019.4.2 不兼容，但与 Tableau Server 版本 2020.1 和任何以后的 Tableau Server 主要版本（例如版本 2020.2）兼容。

在 Tableau Server 中，Tableau Prep Conductor 会在发布流程时检测流程中包括的功能。如果它发现不支持的功能，则流程仍然可发布到 Tableau Server，但无法运行、计划流程或将该功能添加到任务。Tableau Online 会定期自动更新，因此通常与 Tableau Prep Builder 的所有版本兼容。

如果有较旧版本的 Tableau Server，则仍可以在 Tableau Prep Builder 中手动运行或使用命令行运行不兼容的流程。有关使用此过程的详细信息，请参见[通过命令行刷新流程输出文件 在本页 321](#)。

检测不兼容的功能

根据所使用的 Tableau Prep Builder 版本，您可以通过不同方式找出不兼容的功能。

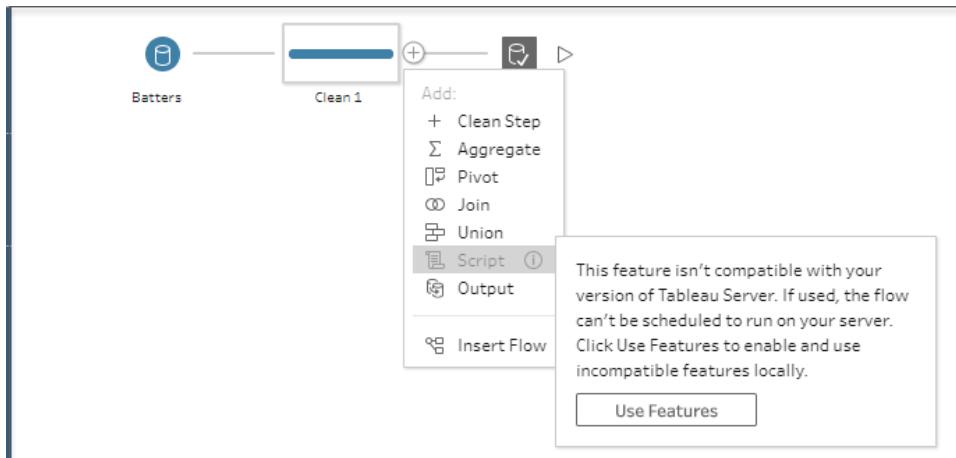
Tableau Prep Builder(版本 2020.1.1 及更高版本)

登录到 Tableau Server，Tableau Prep Builder 将检测并为您禁用不兼容的功能。任何不兼容的功能都将显示为灰色。如果仍要使用该功能并手动或通过命令行运行流程，您可以通过菜单启用它。

注意: 从 Tableau Prep Builder 版本 2020.1.4 开始，一旦登录到服务器，Tableau Prep Builder 在关闭应用程序时会记住您的服务器名称和登录信息，以便下次打开应用程序时您已经登录到服务器。

- 将鼠标悬停在禁用的功能上以查看它是不是因为与服务器版本不兼容而禁用的，然后单击“使用功能”按钮。此选项可在“流程”窗格以及“配置”窗格、“结果”窗格和数据网格内的菜单中找到。

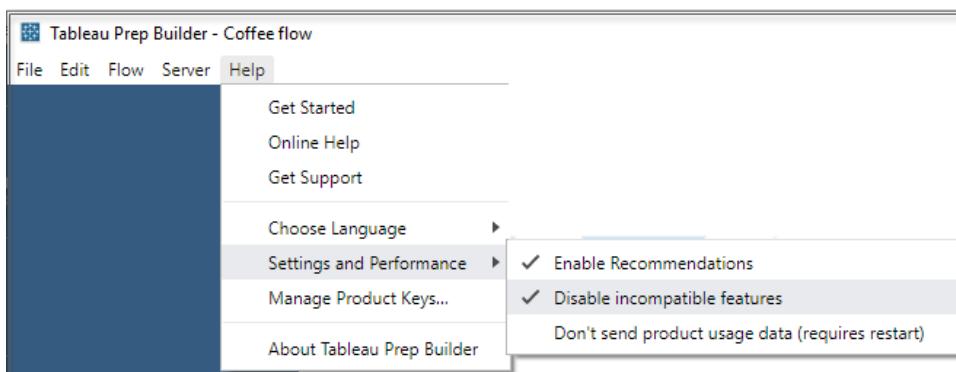
注意：功能可能会由于其他原因而被禁用，例如数据更新被暂停，或者该选项对于特定步骤或数据类型不可用。



- 将应用所选功能，并且所有不兼容的功能将启用并可供使用。不兼容的功能带有警告标志，这样，如果您希望使用您的 Tableau Server 版本中的计划运行流程，就能轻松地找到并移除它们。

若要完全禁用此功能并启用所有不兼容的功能，请执行以下操作：

- 从顶部菜单中，选择“帮助”>“设置和性能”>“禁用不兼容的功能”。



- 选择“禁用不兼容的功能”以清除此选项旁边的复选标记。若要再次启用该功能，请选择“禁用不兼容的功能”。默认情况此选项应已启用。

Tableau Prep Builder(版本 2019.3.1 及更高版本)

生成流程时, Tableau Prep Builder可以在添加不兼容功能时检测到它们, 并使用通知图标标记这些功能。您必须登录到服务器才能看到这些通知。这种新的通知系统可帮助您快速识别流程中不兼容的功能, 以便能够决定是将该功能保留在流程中, 还是将其删除。

将鼠标悬停在“流程”窗格中的通知上以查看有关不兼容功能的信息, 或使用新的通知中心来查看更多详细信息。在通知中心中, 单击“**在流程中查看**”链接直接导航到触发警告的步骤、注释、字段或变更。

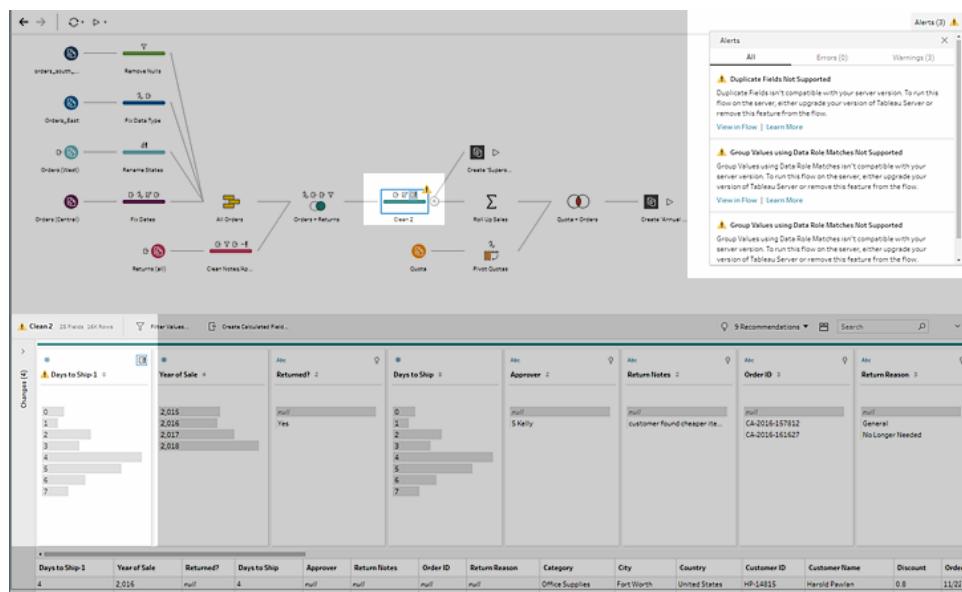
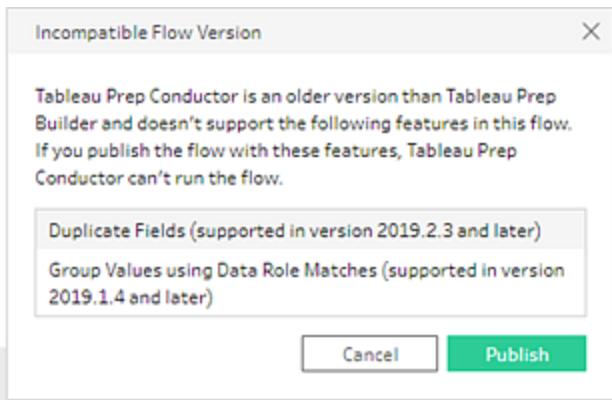


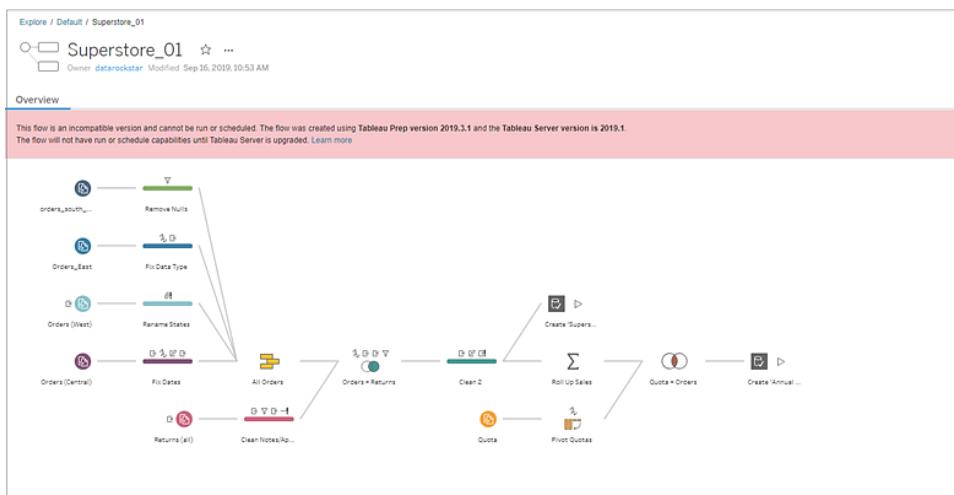
Tableau Prep Builder(所有版本)

如果发布包含不兼容功能的流程, 则会显示以下消息, 并列出您登录到的 Tableau Server 中不支持的功能。在 Tableau Prep Builder 版本 2019.2.3 及更低版本中, 这是确定哪些功能在流程中不兼容的唯一方式。

注意: 错误消息列出了引入该功能时的 Tableau Prep Builder 版本。Tableau Prep Builder 不会发布维护版本, 因此, 为了使功能兼容, 或 Tableau Server 必须运行下一个主要发布版本。在下面的示例中, “复制字段”功能是在 Tableau Prep Builder 版本 2019.2.3 中引入的, 因此它与 Tableau Server 的维护发布版本 2019.2.3 不兼容。相反, 它将与 Tableau Server 的下一个主要版本 2019.3 兼容。



如果继续发布流程，发布将成功完成。但是，在 Tableau Server 或 Tableau Online 中打开流程时，您将看到以下消息：



若要在 Tableau Server 中计划并运行流程，您可以执行以下操作之一：

- 查找与您正在使用的 Tableau Prep Builder 版本兼容的 Tableau Server 的最新主要版本。例如，Tableau Prep Builder 每月会发布新版本。Tableau Prep Builder 版本 2019.2.3 中发布的功能在 Tableau Server 版本 2019.2.2 或 2019.2.3 中可能不受支持，但在 Tableau Server 的下一个主要版本版本 2019.3 中受支持。

Tableau Online 会定期自动更新，通常每季度更新一次。在发布之前首先测试您的流程，确保其与当前版本的 Tableau Online 兼容。

- 在发布流程之前，从流程中移除不兼容的功能，然后发布流程。
- 如果已将流程发布到 Tableau Server，请下载流程并移除功能，或者在较旧版本的 Tableau Prep Builder 中创建只使用该版本中可用功能的流程。

注意: 若要下载特定版本的 Tableau Prep Builder, 请打开[产品下载和版本说明](#)页面, 并从页面左侧的列表中选择“**Tableau Prep Builder**”。

修复兼容性问题

如果已将流程发布到 Tableau Server, 请尝试下面的步骤, 使用当前版本的 Tableau Prep Builder 移除不兼容的功能。移除了相关功能并且不再看到版本不兼容消息或警告后, 请将流程重新发布到 Tableau Server 或 Tableau Online, 并使用 Tableau Prep Conductor 来计划流程。

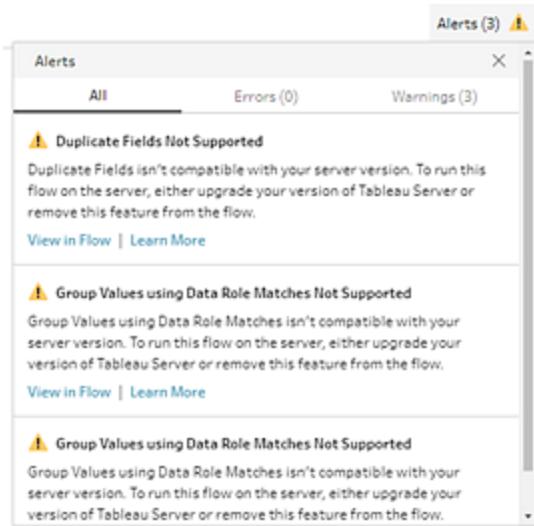
注意: Tableau Prep Conductor 是数据管理加载项的一部分。必须在 Tableau Server 或 Tableau Online 中启用它才能使用计划功能运行流程。有关数据管理加载项的详细信息, 请参见[Tableau 数据管理加载项](#)。有关在 Tableau Server 或 Tableau Online 中启用 Tableau Prep Conductor 的详细信息, 请参见[步骤 2: 配置 Tableau Server 的流程设置](#)在本页 369 或在[Tableau Online 站点上启用 Tableau Prep Conductor](#)在本页 377。

确定不兼容的功能

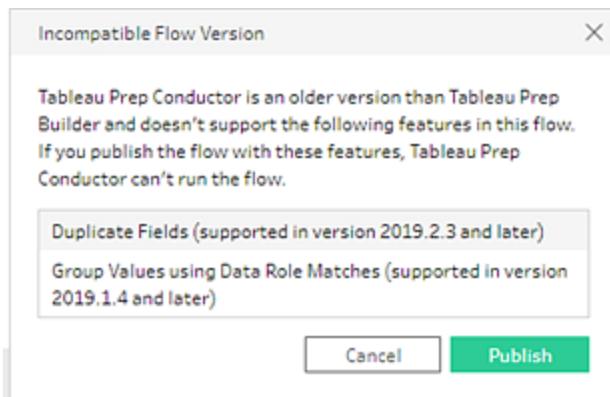
如果在 Tableau Server 中工作, 它当前不会列出流程中不兼容的功能。若要确定要从流程中移除的功能的列表, 您需要在 Tableau Prep Builder 中打开流程, 然后在流程中查找并移除这些功能。

1. 打开流程。如果您在 Tableau Prep Conductor 中, 请从“**更多操作**”... 菜单中单击“**下载**”, 以下载并在 Tableau Prep Builder 中打开流程, 或者只是在 Tableau Prep Builder 中打开流程。
2. 如果下载流程, 请单击下载的流程将其打开。
3. 根据您的版本, 执行以下操作之一:
 - 版本 **2019.3.1 及更高版本**: 从顶部菜单中选择“**服务器**”>“**登录**”。确保选择与流程不兼容的同一服务器。任何不兼容的步骤、注释、字段或变更应标有通知图标。

在流程窗格的右上角, 单击“**通知**”以查看每个不兼容功能的详细信息。单击“**在流程中查看**”, 导航到不兼容的功能以采取措施。



- **版本 2019.2.3 及更低版本:** 从顶部菜单中选择“服务器”>“发布流程”。如果需要再次登录服务器，请确保选择与流程不兼容的相同服务器。将打开一个警告对话框，其中列出与服务器版本不兼容的功能。记下这些功能，以便可以从流程中识别并删除它们。然后单击“取消”关闭该对话框。



4. 从顶部菜单中，单击“文件”>“另存为”以保存流程的副本。使用以下部分中的选项从流程中移除不兼容的功能。

从流程中移除不兼容的功能

您可以使用各种方法来查找和移除流程中的功能。本部分介绍了一些选项来帮助您解决不兼容性错误。

不兼容的数据源

如果数据源不兼容，例如，添加了 Tableau Prep Conductor 中尚不支持的新连接器，则您需要连接到支持的数据源。

若要更改数据连接，请执行以下操作：

1. 在“**连接**”窗格中，单击连接器上的下拉箭头，并选择“**移除**”以移除数据源连接。
这会使您的流程暂时处于错误状态。
2. 连接到兼容的数据源，并将其拖到“流程”窗格。若要将新的“输入”步骤与流程连接，请将该步骤拖到流程中要连接该步骤的第一步，并放在“**添加**”选项上的连接步骤上。



3. 重复这些步骤以替换任何其他连接器。然后保存您的流程并重新发布。

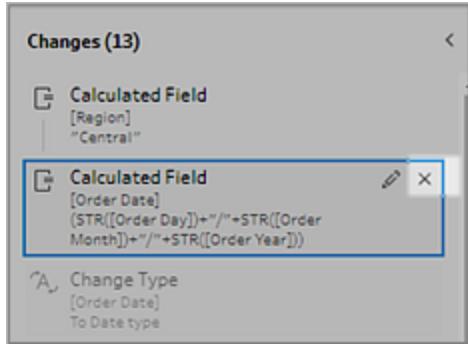
不兼容的功能

若要移除不兼容的功能，您需要查找使用了这些功能的步骤，并将其移除。您可以按照[确定不兼容的功能](#)在本页341中的说明来查找不兼容的功能。

1. 如果该功能是一种步骤类型，请在“流程”窗格中单击使用了该功能的步骤。右键单击或按住 Ctrl 并单击 (MacOS) 步骤，然后选择“**移除**”。
2. 如果该功能是清理操作，请在“流程”窗格中单击使用了该功能的步骤。您可以将指针悬停在“流程”窗格或者“**配置**”或“**结果**”窗格中的批注上，以查看更改的列表。

注意: 在 Tableau Prep Builder 版本 2019.1.3 及更高版本中，您可以将指针悬停在表示所查找的更改的图标上，或悬停在“流程”窗格或“配置”卡中的步骤上，然后从更改列表中选择批注。更改将在“**更改**”窗格、“**配置**”或“**结果**”窗格以及数据网格中突出显示。

3. 打开“**更改**”窗格(如果需要)，并选择与需要移除的功能匹配的更改。单击更改将其选中，并单击“**移除**”将其从流程中移除。



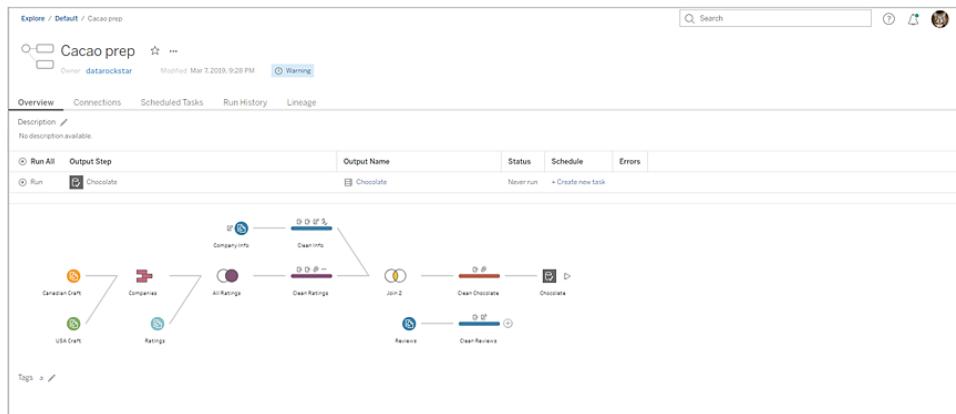
4. 重复这些步骤以替换任何其他功能。然后保存您的流程并重新发布。

使流程数据保持最新

您已经构建了流程并清理了数据，但现在想要与其他人共享数据集，并希望保持使这些数据保持最新。您始终可以在 Tableau Prep Builder 中运行流程，并将数据提取发布到 Tableau Server，但现在有更好的方法。

使用 Tableau Server 版本 2019.1 及更高版本全新的数据管理加载项的一部分，即 Tableau Prep Conductor。如果您将此选项添加到 Tableau Server 或 Tableau Online 安装，则可以使用 Tableau Prep Conductor 自动运行流程并使流程数据保持最新。

注意：如果已安装 Tableau Catalog，您可能还会看到有关流程输入数据的数据质量警告，并在新的“世系”选项卡中查看流程中字段的上游和下游影响。有关 Tableau Catalog 的详细信息，请参见 Tableau Desktop 联机帮助中的[关于 Tableau Catalog](#)。



使用 Tableau Prep Conductor，您可以执行以下操作：

- 将服务器或站点配置为使用 Tableau Prep Conductor
 - 为单独的站点启用或禁用 Tableau Prep Conductor
 - 为按需或使用计划运行的流程设置流程失败电子邮件通知
 - 配置流程超时设置
- 将流程从 Tableau Prep Builder 发布到 Tableau Server 或 Tableau Online
 - 上载数据文件或直接连接到文件或数据库。如果连接到数据库，您可以嵌入数据库凭据或要求用户提示。

注意: 如果通过直接连接方式连接到数据文件或将流程输出发布到文件共享，则文件需要位于 Tableau Server 可以访问的位置。有关详细信息，请参见[步骤 4: 安全列表输入和输出位置](#)在本页 375。

- 发布流程时从项目层次结构中选择
- 输入标记和说明以帮助其他人查找您的流程
- 管理流程
 - 设置权限
 - 将流程移到其他项目
 - 更改流程所有者
 - 添加或编辑标记
 - 查看版本历史记录并从列表中选择以将流程还原到早期版本
 - 将流程标记为收藏夹并将其添加到收藏夹列表
 - 编辑输入连接和更新凭据
 - 查看通过流程创建的数据源并链接回创建它的流程
- 创建计划以运行流程
- 添加计划任务以运行流程并选择要更新的流程输出
- 在没有计划的情况下按需运行流程
- 监视流程
 - 设置电子邮件通知
 - 查看错误
 - 监视并重新启动已挂起的流程
 - 查看运行历史记录
 - 使用管理视图

关于 Tableau Prep Conductor

利用 Tableau Prep Conductor，您可以无缝地将流程从 Tableau Prep Builder 发布到 Tableau Server 或 Tableau Online，然后利用这些应用程序中的计划和跟踪功能自动运行流程来更

新流程输出。Tableau Prep Conductor 是 Tableau Server 版本 2019.1 中引入的 Tableau 数据管理加载项的一部分，必须启用才能使用这些功能。

发布流类似于使用 Tableau Desktop 发布数据源和工作簿。您可以将文件与流程打包在一起，或者指定与数据源的直接连接，以便在数据更改时更新流程输入。如果流程连接到数据库，请指定身份验证类型并设置凭据以访问数据。

您还可以发布流程以与他人共享。例如，将不完整的流程发布到 Tableau Server 或 Tableau Online，然后将其下载到另一台计算机以继续对其进行处理。创建一个仅包含正确配置的输入步骤的流程，并将其与同事共享，后者随后可以将流程下载到其计算机，并创建和发布他们自己的流程。

如果流程包括输出步骤，并且流程没有错误或不兼容的功能，则流程只能在 Tableau Server 或 Tableau Online 中运行。有关发布流程的详细信息，请参见[将流程发布到 Tableau Server 或 Tableau Online 在本页 378](#)。有关不兼容性的详细信息，请参见[与 Tableau Prep 的版本兼容性 在本页 333](#)。

跟踪流程的运行状况很容易。使用 Tableau Prep Conductor，您可以：

- 在“**内容**”页面中查看和监视有关流程的详细信息，包括最近的活动。
- 在“**运行历史记录**”选项卡中查看流程运行的结果及任何错误。
- 使用**管理视图**来监视服务器和网站活动，包括用于跟踪流程性能历史记录的新视图。
- 查看失败流程运行的详细通知。
- 设置电子邮件通知以流程所有者发送电子邮件，以便在流程运行失败时通知他们，并说明相关原因。

有关设置通知的详细信息，请参见[监视流程运行状况和性能 在本页 407](#)。

如果流程由于连接错误而无法运行，您可以通过编辑连接，直接在 Tableau Server 或 Tableau Online 中修复错误。如果某个流程由于一个或多个流程步骤存在错误(例如无效的计算)而无法运行，则只需将流程下载到 Tableau Prep Builder，修复问题，然后将流程重新发布到 Tableau Server 或 Tableau Online，从您离开的位置继续操作。

启用 Tableau Prep Conductor

Tableau Prep Conductor 可与数据管理加载项一起使用，并且可随 Tableau Server 或 Tableau Online 一起安装，或添加到 Tableau Server 或 Tableau Online 版本 2019.1 或更高版本的现有安装。有关如何安装和启用 Tableau Prep Conductor 的详细信息，请参见[在 Tableau Server 上启用和配置 Tableau Prep Conductor 在本页 352](#)或[在 Tableau Online 站点上启用 Tableau Prep Conductor 在本页 377](#)。

Tableau Prep Conductor 设计为可与 Tableau Prep Builder 版本 2019.1.2 及更高版本配合使用。您必须安装或升级 Tableau Prep Builder 以使用版本 2019.1.2 或更高版本，才能将流程发布到 Tableau Server 或 Tableau Online。有关如何安装 Tableau Prep Builder 的信息，请参见[通过用户界面安装 Tableau Desktop 或 Tableau Prep Builder](#)。

了解流程工作区

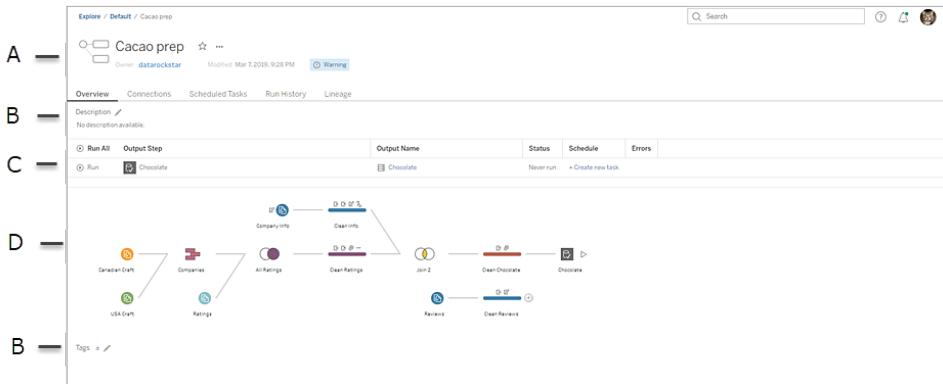
发布流程后，您可以在 Tableau Server 或 Tableau Online 中计划任务，以便定期自动运行流程以使输出数据保持最新。

在 Tableau Server 或 Tableau Online 中，Tableau Prep Conductor 利用的功能与通过 Tableau Desktop 管理工作簿或数据源时您可能看到的用于管理流程的功能大致相同。例如，就像数据提取刷新一样，计划的流程任务和按需流程运行将以后台任务形式排队。但谈到处理流程，则存在一些不同之处。

“流程概述”页面

“**流程概述**”页面是主登录页面，您可以在其中查看有关流程和计划的数据，以及监视和维护流程运行。

通过单击列表中的流程来打开“**流程概述**”页面。您可以通过“**内容**”>“**浏览**”>“**所有流程**”或打开包含流程的项目导航到该处。



- A. 标题列出流程的名称、流程所有者以及流程的上次修改日期。

将流程添加到收藏夹中，或者可通过“**更多操作**”... 菜单下载流程、设置权限、更改流程所有者、还原以前的流程版本等等。

- B. 查看和编辑流程说明和设置标记，以帮助其他人找到他们所查找的流程。
- C. 查看流程的输出步骤以及上次更新的状态、为其分配输出的任何计划，以及上次流程

Tableau Prep 帮助

运行中的任何错误。您还可以单击“**运行**” 按钮以运行所有输出步骤，或按需运行单独的输出步骤。

如果流程有错误，流程运行将失败。通过导航到流程的“**连接**”选项卡并编辑输入连接，可以直接在 Tableau Server 或 Tableau Online 中解决连接错误。若要解决任何其他错误，请从“**更多操作**”... 菜单中选择“**下载**”，下载流程并将其在 Tableau Prep Builder 中打开。编辑流程然后将其重新发布，并再次尝试运行该流程。

在成功运行流程之后，将输出数据源，这些数据源将变为链接形式，您可以单击以打开“**数据源**”页面来查看有关数据源的详细信息，或编辑流程输入连接。

在“**计划**”字段中，查看为其分配了输出步骤的计划任务。可以将流程输出分配给一个或多个任务。

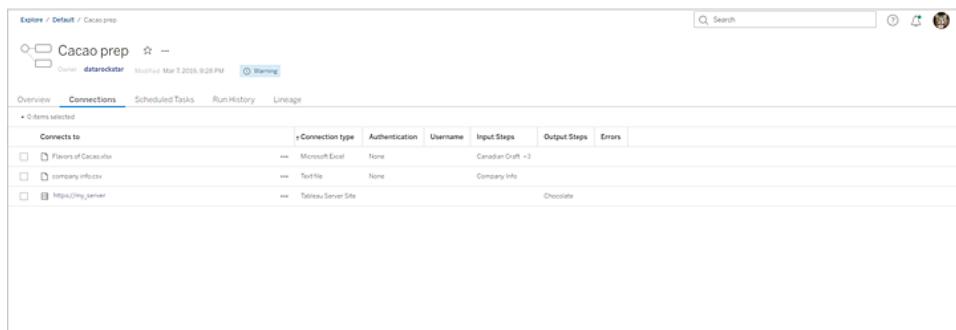
如果尚未分配计划，请单击“**创建新任务**”，将输出步骤添加到计划中。若要立即运行流程以更新特定输出步骤，请单击行左侧的“**运行**” 按钮。

D. 查看流程的图像。

“流程连接”页面

查看流程的输入和输出位置、连接类型、身份验证设置、输入和输出步骤以及任何连接错误。您可以在发布流程时设置身份验证设置。有关详细信息，请参见[发布流程 在本页 383](#)。

对于数据库输入类型，请单击输入连接的“**更多操作**”... 菜单，以编辑连接并更改服务器名称、端口、用户名和密码。



“流程计划任务”页面

查看为其分配了流程的任何计划以及这些计划中包括的输出。单击任务卡顶部的计划链接以打开“**计划**”页面，并查看分配给该计划的流程列表。有关为计划分配流程的详细信息，请

参见[计划流程任务 在本页 396。](#)

您还可以通过此页面添加新任务或管理现有任务。若要对现有任务执行操作，请选中任务卡上的复选框，然后单击“**操作**”下拉菜单以运行、编辑或删除任务。

在 Tableau Server 中，如果在配置的连续尝试次数后流程未能运行，则会自动挂起流程。在 Tableau Online 中，如果流程任务在连续尝试 5 次后未能运行，则会自动挂起流程。

您可以在“**概览**”选项卡以及此选项卡上看到该状态。您可以通过此菜单恢复挂起的任务。

有关如何为挂起的流程任务设置阈值的信息，请参见[步骤 5：可选服务器配置 在本页 377](#)。有关挂起的流程任务的详细信息，请参见[查看和解决错误 在本页 408](#)。

The screenshot shows the 'Scheduled Tasks' tab of the Tableau Prep interface. It lists three scheduled tasks:

- Run Flow - Weekdays 4:00PM**: Weekly at 4:00 PM on Monday, Tuesday, Wednesday, Thursday, and Friday. Status: Succeeded (next run at: Sep 11, 2019 4:00 PM). Contains one output step: pivot (output name: resa_pivot_out, type: Tableau Server Site, location: https://my_server).
- Fast Flow Refresh**: Every 15 minutes. Status: Scheduled (next run at: Sep 11, 2019 9:15 AM). Contains three output steps: pivot, pivot (output name: resa_pivot_out, type: Tableau Server Site, location: https://my_server), and pivot (output name: resa_pivot_out, type: Tableau Server Site, location: https://my_server).
- Run Flow - Every night - 11:00PM**: Daily at 11:00 PM. Status: Succeeded (last run at: Sep 11, 2019 11:00 PM). Contains one output step: pivot (output name: resa_pivot_out, type: Tableau Server Site, location: https://my_server).

“计划”页面

在“**计划**”页面上，您可以查看分配给计划的流程，以及有关流程运行的详细信息。可以按需运行计划并运行分配给它的所有流程。您还可以选择一个或多个流程，然后使用“**操作**”菜单更改流程计划或优先级、从计划中删除选定的流程，或者恢复挂起的流程。

有关如何设置计划的信息，请参见[步骤 3：为流程任务创建计划 在本页 373](#)。

Tableau Prep 帮助

The screenshot shows the 'Run History' section of the Tableau Prep Conductor. It displays a table with two rows of data. The columns are: Flow, Output Step, Priority, Status, Last Run, Next Scheduled Run, and Errors. The first row shows a flow named 'Another DAI Flow' with an output step 'Output [http://DAIServer01]'. The second row shows a flow named 'My Superstore Sales' with an output step 'Create Annual Regional Performance TSV [http://DAIServer01]-1'. Both rows show a priority of 50, a status of 'Succeeded' or 'Scheduled', and a last run date of April 25, 2020.

流程运行历史记录

查看、搜索流程的历史运行列表并对其进行排序。此页面还包括有关流程运行的详细信息，例如运行类型、持续时间以及生成的行数。

如果流程输出有错误，请将鼠标指针悬停在错误上以查看消息。如果适用，请单击错误消息中的“转到连接”链接，导航到“连接”页面来修复连接错误。如果显示，请单击“下载流程”链接下载流程，以在 Tableau Prep Builder 中修复流程错误，然后重新发布流程以继续使用 Tableau Prep Conductor 进行对其进行管理。

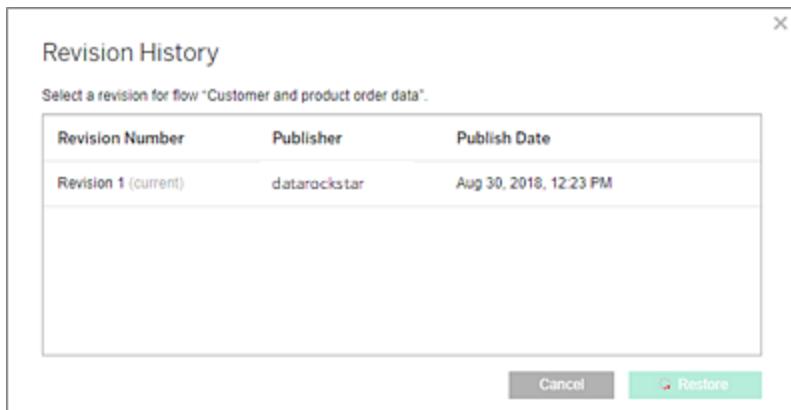
注意：除非删除流程，否则流程的运行历史记录将保留。

The screenshot shows the 'Run History' section for a workflow named 'Coffee Flow_incremental'. The table lists three runs: 'Output 2' (Incremental refresh), 'Output' (Incremental refresh), and 'Output' (Full refresh). The 'Full refresh' run failed with an error message: 'Outputs that share an input and are set up to refresh incrementally must use the same run setting. The following outputs must all be run using the incremental refresh settings or must all be run as a full refresh. Outputs: "Output, Output 2"'.

Output Step	Run Type	Run Start	Run End	Duration	Status	Rows Generated	Errors
Output 2	Incremental refresh	Apr 30, 2020, 11:01 PM	Apr 30, 2020, 11:01 PM	00:00:10	✓ Succeeded	4,248	
Output	Incremental refresh	Apr 30, 2020, 11:01 PM	Apr 30, 2020, 11:01 PM	00:00:09	✓ Succeeded	4,248	
Output	Full refresh	Apr 21, 2020, 11:49 AM	Apr 21, 2020, 11:49 AM	00:00:03	! Failed	0	1 error

流程修订历史记录

如果需要将流程恢复为以前的版本，请从该流程的“更多操作”...菜单中选择“修订历史记录”。在“修订历史记录”对话框中，从列表中选择要恢复到的流程版本。



谁可以执行此操作

服务器管理员可以激活 数据管理加载项 许可证密钥。

服务器管理员可以启用 Tableau Prep Conductor。

许可 Tableau Prep Conductor

Tableau Prep Conductor 通过数据管理加载项获得许可。

有关详细信息，请参见：

- 为 Tableau Server 数据管理加载项授予许可
- 为 Tableau Online 数据管理加载项授予许可

在 Tableau Server 上启用和配置 Tableau Prep Conductor

仅 Tableau Server 2019.1 或更高版本支持 Tableau Prep Conductor。

Tableau Prep Conductor 通过数据管理加载项在每次部署时获得许可，采用基于用户或基于内核的形式。部署包括许可的生产 Tableau Server 安装和许可的非生产 Tableau Server 安装(支持生产安装)。有关部署的详细信息，请参见 [EULA 文档](#)。

有关 Tableau Prep Conductor 许可工作方式的详细信息，请参见[针对 Tableau Server 为 Tableau Prep Conductor 授予许可](#)。

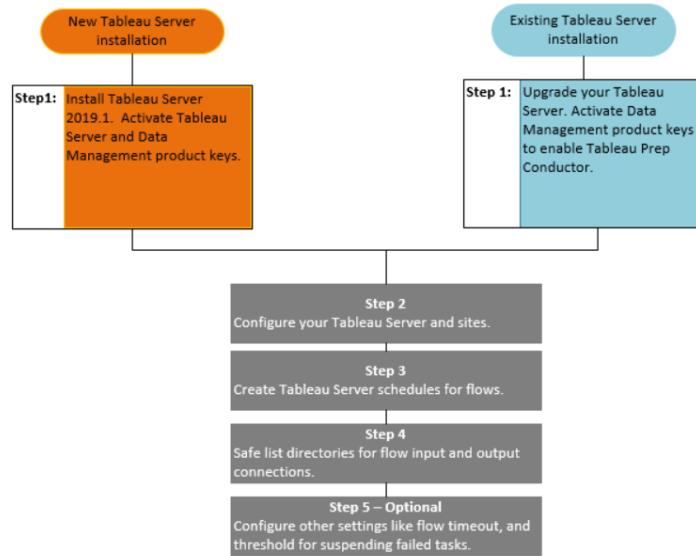
服务器拓扑

使用数据管理产品密钥安装 Tableau Server 并启用 Tableau Prep Conductor 时，默认情况下安装程序会自动启用 Tableau Prep Conductor。

对于多节点安装，默认情况下，会在安装了后台程序的任何节点上启用 Tableau Prep Conductor 的一个实例。在下面的示例中，已在节点 2 和 3(其中也启用了后台程序) 上启用了 Tableau Prep Conductor，但未在节点 1、4 和 5 上启用。

Process	node1	node2	node3	node4	node5
Gateway	✓				
Application Server	✓				
Interactive Microservice Container	✓				
VizQL Server	✓ ✓ ✓ ✓				
Cache Server	✓ ✓				
Cluster Controller	✓	✓	✓	✓	✓
Search & Browse	✓				
Backgrounder		✓ ✓ ✓ ✓	✓ ✓ ✓ ✓		
Background Microservice Container		✓ ✓ ✓ ✓	✓ ✓ ✓ ✓		
Data Server	✓ ✓				
Data Engine	✓	✓	✓		
File Store	✓				
Repository	✓				
Tableau Prep Conductor		✓	✓		
Ask Data	✓				
Elastic Server	✓				

下面是工作流的直观表示形式：



后续步骤：

新 Tableau Server 安装：[步骤 1\(新安装\)：安装包含 Tableau Prep Conductor 的 Tableau Server 向下](#)

现有 Tableau Server 安装：[步骤 1\(现有安装\)：启用 Tableau Prep Conductor 在本页 359](#)

谁可以执行此操作

服务器管理员可以安装 Tableau Server 并启用 Tableau Prep Conductor。

服务器级别设置可由 Tableau Server 管理员配置，而站点级别设置可由 Tableau Server 和站点管理员配置。

步骤 1(新安装)：安装包含 Tableau Prep Conductor 的 Tableau Server

本主题介绍如何在新安装的 Tableau Server 上安装 Tableau Prep Conductor。

仅 Tableau Server 2019.1 或更高版本支持 Tableau Prep Conductor。

Tableau Prep Conductor 通过数据管理加载项在每次部署时获得许可。部署包括许可的生产 Tableau Server 安装和许可的非生产 Tableau Server 安装(支持生产安装)。有关部署的详细信息，请参见 [EULA 文档](#)。

安装之前

用于生产 Tableau Server 安装的推荐拓扑是用于运行流程的专用节点。如果当前打算安装单节点 Tableau Server，则建议您添加第二个节点并将其专用于运行流程。

- 查看 Tableau Server 和 Tableau Prep Conductor 的硬件建议。
 - 在 Windows 上安装 [Tableau Server 的最低硬件要求和建议](#)。
 - 在 Linux 上安装 [Tableau Server 的最低硬件要求和建议](#)。

安装 Tableau Server 并启用 Tableau Prep Conductor

使用以下主题中提供的说明安装 Tableau Server。

[Windows: 安装 Tableau Server](#)

[Linux: 安装 Tableau Server 主题](#)

进入“激活”步骤时，使用 Tableau Server 产品密钥来激活 Tableau Server。

所有产品密钥均通过[客户门户](#)提供。

配置公共网关设置

如果您的 Tableau Server 设置具有以下各项之一：

- 负载平衡器，用于跨网关分发请求。
- 反向代理，用于验证外部 (Internet) 客户端请求和卸载基于 SSL 的加密。

您必须配置以下公共网关设置：

```
tsm configuration set -k gateway.public.host -v <name> (这应该是用户  
用于访问 Tableau Server 的 URL)
```

```
tsm configuration set -k gateway.public.port -v 443
```

有关配置网关设置的详细信息，请参见[为 Tableau Server 配置代理](#)。

启用 Tableau Prep Conductor

使用以下步骤将数据管理产品密钥添加到 Tableau Server：

注意：此过程需要重新启动 Tableau Server。

1. 如果运行 Tableau Server 的计算机已配置为通过转发代理连接到 Internet, 请在继续之前按照[使用转发代理配置产品密钥操作](#)主题中的过程进行操作。
2. 在浏览器中打开 TSM:
`https://<tsm-computer-name>:8850`
3. 在“配置”选项卡上单击“许可”, 并单击“激活许可证”。
4. 输入或粘贴您的[数据管理产品密钥](#), 并单击“激活”:
5. 在“注册”页面上的字段中输入您的信息, 然后单击“注册”。
6. 注册完成后, 按照提示进行操作并重新启动 Tableau。

验证 Tableau Prep Conductor 是否已启用并正在运行

激活数据管理产品密钥后, 将在启用了后台程序的任何节点上自动启用 Tableau Prep Conductor 的单个实例。

使用以下步骤验证是否其已启用并正在运行:

1. 打开浏览器并输入 Tableau Server URL, 然后附加专用 TSM Web UI 端口。URL 看起来可能与下面一些示例类似:

`https://localhost:8850/`(如果直接在服务器计算机上工作)

`https://MarketingServer:8850/`(如果您知道服务器名称)

`https://10.0.0.2:8850/`(如果您知道服务器的 IP 地址)

在出现的登录页中, 输入您的管理员用户名和密码。

注意: Tableau Server 会在安装过程中创建和配置自签名证书。此证书用于对进入 TSM Web UI 的流量进行加密。因为它是自签名证书, 所以默认情况下浏览器不会信任它。因此, 在允许您连接之前, 浏览器会显示有关证书可信度的警告。

2. 在 Tableau 服务管理器 Web 界面中, 单击“状态”选项卡以查看状态。
 - 如果 Tableau Prep Conductor 已启用并正在运行, 您应会在进程列表中看到 Tableau Prep Conductor, 并至少在一个节点上显示为“活动”。如果未启用 Tableau Prep Conductor, 您将在进程列表中看到 Tableau Prep Conductor, 但没有任何节点的状态信息。

Tableau Prep Conductor 未启用:

Tableau Prep 帮助

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VizQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor			
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

Tableau Prep Conductor 已启用并正在运行。在下图中，**Tableau Prep Conductor** 已在节点 1 和节点 3 上启用：

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VizQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor	✓		✓
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

将节点专用于 Tableau Prep Conductor

在打算专用于运行流程的节点上，如果尚未启用后台程序进程，请启用该进程。建议您不要在此节点上运行诸如 VizQL 服务器之类的其他进程。

由于您将此节点专用于运行流程，因此必须配置后台程序以仅运行流程任务。默认情况下，后台程序进程运行所有类型的任务，包括流程、数据提取刷新和订阅。有关详细信息，请参见 [Tableau Server 上的节点角色](#)。

在专用节点上运行以下 `tsm` 命令以仅运行流程任务：

1. 运行以下命令以允许此节点上的后台程序只运行流程任务。

```
tsm topology set-node-role -n node1 -r flows
```

2. 将初始节点上的节点角色设置为无流程。此节点上的后台程序将运行除流程之外的所有作业：

```
tsm topology set-node-role -n node1 -r no-flows
```

3. 应用更改并重新启动 Tableau Server：

```
tsm pending-changes apply
```

多节点安装

如果 Tableau Server 安装中有 2 个以上的节点，则可以选择将其他节点配置为运行除流程之外的所有任务：

1. 限制节点以不允许流程。此命令从此节点移除 Tableau Prep Conductor，此节点上的后台程序将不会运行流程任务。

```
tsm topology set-node-role -n node1 -r no-flows
```

2. 应用更改并重新启动 Tableau Server：

```
tsm pending-changes apply
```

后续步骤

[步骤 2：配置 Tableau Server 的流程设置](#) 在本页 369

谁可以执行此操作

服务器管理员可以安装 Tableau Server 并启用 Tableau Prep Conductor。

步骤 1(现有安装)：启用 Tableau Prep Conductor

本主题介绍如何在现有的 Tableau Server 安装上启用 Tableau Prep Conductor。

仅 Tableau Server 2019.1 或更高版本支持 Tableau Prep Conductor。如果您使用的是 Tableau Server 2018.3 或更低版本，则必须先将 Tableau Server 升级为 2019.1，然后才能在 Tableau Server 安装上启用 Tableau Prep Conductor。

Tableau Prep Conductor 通过数据管理加载项在每次部署时获得许可，采用基于用户或基于内核的形式。部署包括许可的生产 Tableau Server 安装和许可的非生产 Tableau Server 安装（支持生产安装）。有关部署的详细信息，请参见 [EULA 文档](#)。

本主题介绍如何在现有的 Tableau Server 安装上启用 Tableau Prep Conductor。

升级之前

准备升级：

- [升级前须知事项](#)
- [许可 Tableau Prep Conductor](#)
- [Tableau Server 硬件要求和建议](#)

配置公共网关设置

如果您的 Tableau Server 设置具有以下各项之一：

- 负载平衡器，用于跨网关分发请求。
- 反向代理，用于验证外部 (Internet) 客户端请求和卸载基于 SSL 的加密。

您必须配置以下公共网关设置：

```
tsm configuration set -k gateway.public.host -v <name> (这应该是用户  
用于访问 Tableau Server 的 URL)
```

```
tsm configuration set -k gateway.public.port -v 443
```

有关配置网关设置的详细信息，请参见[为 Tableau Server 配置代理](#)。

使用基于用户的许可证的 Tableau Server 安装

用于生产 Tableau Server 安装的推荐拓扑是用于运行流程的专用节点。有关详细信息，请参见[Tableau Server 的最低硬件要求和建议](#)。

Tableau Server 单节点安装

如果当前有一个单节点 Tableau Server 安装，则建议您添加第二个节点并将其专用于运行流程。

1. 使用以下主题中的信息在当前 Tableau Server 安装上运行升级：

- [Windows](#)
- [Linux](#)

进入“激活”步骤时，使用 Tableau Server 产品密钥来激活 Tableau Server。

所有产品密钥均通过[客户门户](#)提供。

2. 完成安装后，添加数据管理产品密钥以在节点上启用 Tableau Prep Conductor。像其他服务器密钥一样，数据管理产品密钥通过[客户门户](#)提供。

- 在 Tableau 服务管理器 Web 界面中，单击“配置”选项卡上的“许可”，然后单击“激活许可证”。
- 输入或粘贴您的产品密钥，并单击“激活”。
- 在“注册”页面上的字段中输入您的信息，然后单击“注册”。

3. 系统将提示您重新启动服务器。重新启动服务器并验证 Tableau Prep Conductor 是否已启用并正在运行。

- 在 Tableau 服务管理器 Web 界面中，单击“状态”选项卡以查看状态。如果 Tableau Prep Conductor 已启用并正在运行，您应会在进程列表中看到 Tableau Prep Conductor，显示为“活动”。如果 Tableau Prep Conductor 未启用，您将在进程列表中看到 Tableau Prep Conductor，但没有状态信息。

Tableau Prep Conductor 未启用：

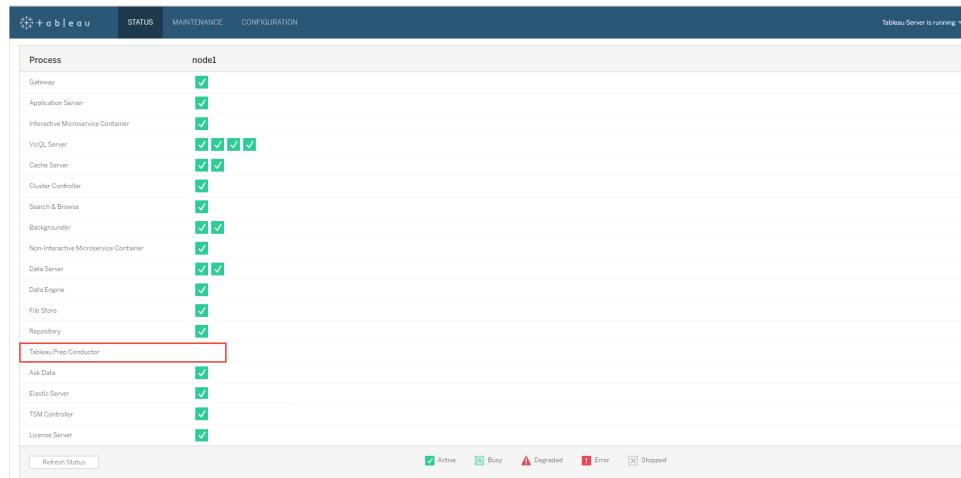
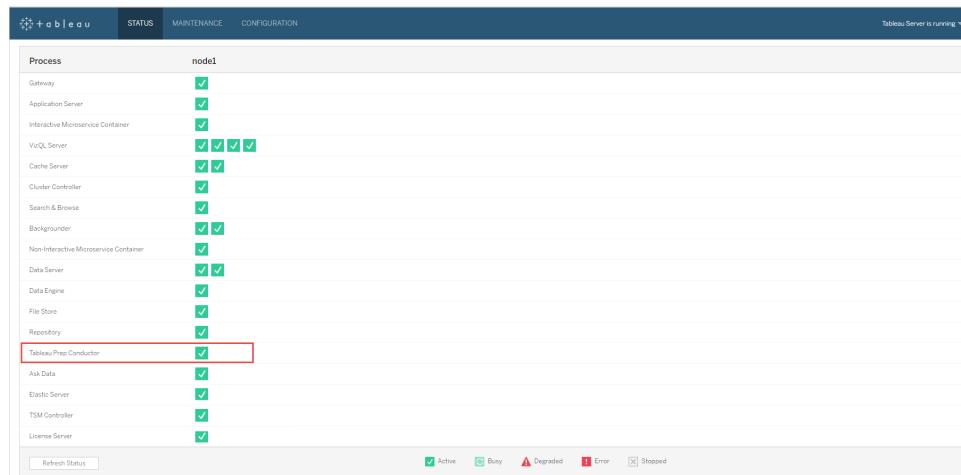


Tableau Prep Conductor 已启用并正在运行：



4. 向 Tableau Server 安装中添加第二个节点。安装程序将启用某些必需的进程，如群集控制器。在运行计划的流程任务需要后台程序进程时启用该进程。启用后台程序进程时，安装程序会自动在节点上启用一个数据引擎实例和 Tableau Prep Conductor。不要在此节点上添加任何其他进程。
5. 运行以下命令以使此节点专用于仅执行流程任务。有关节点角色的详细信息，请参见 [Tableau Server 中的节点角色](#)。
 - 获取专用节点的节点 ID，以查看每个节点上的服务列表：


```
tsm topology list-nodes -v.
```
 - 使用通过运行上述命令获得的节点 ID 为专用节点设置节点角色：

```
tsm topology set-node-role -n <nodeID> -r flows.
```

- 应用更改并重新启动服务器:

```
tsm pending-changes apply.
```

- 查看状态以确保所有进程均已启动并运行，并且配置正确:

```
tsm status -v.
```

您已成功将 Tableau Prep Conductor 添加到 Tableau Server 安装中。

Tableau Server 多节点安装

- 使用以下主题中的信息在当前 Tableau Server 安装上运行升级:

- [Windows](#)
- [Linux](#)

进入“激活”步骤时，使用 Tableau Server 产品密钥来激活 Tableau Server。

所有产品密钥均通过[客户门户](#)提供。

- 完成安装后，添加数据管理产品密钥以启用 Tableau Prep Conductor。在您已启用后台程序进程的节点上，Tableau Prep Conductor 会自动启用。像其他服务器密钥一样，数据管理产品密钥通过[客户门户](#)提供。

- 在 Tableau 服务管理器 Web 界面中，单击“配置”选项卡上的“许可”，然后单击“激活许可证”。
- 输入或粘贴您的产品密钥，并单击“激活”。
- 在“注册”页面上的字段中输入您的信息，然后单击“注册”。

- 系统将提示您重新启动服务器。重新启动服务器并验证 Tableau Prep Conductor 是否已启用并正在运行。

- 在 Tableau 服务管理器 Web 界面中，单击“状态”选项卡以查看所有进程的状态。如果 Tableau Prep Conductor 已启用并正在运行，您应会在进程列表中看到 Tableau Prep Conductor，显示为“活动”。如果 Tableau Prep Conductor 未启用，您将在进程列表中看到 Tableau Prep Conductor，但没有状态信息。

Tableau Prep Conductor 未启用：

Tableau Prep 帮助

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VizQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor			
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

Tableau Prep Conductor 已启用并正在运行：

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VizQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor	✓	✓	
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

- 向 Tableau Server 安装中添加新节点。安装程序将启用某些必需的进程，如群集控制器。在运行计划的流程任务需要后台程序进程时启用该进程。启用后台程序进程时，

安装程序会自动在节点上启用一个数据引擎实例和 Tableau Prep Conductor。不要在此节点上添加任何其他进程。

注意: 专用节点将计入协调服务整体的总数。您可能需要在新节点上部署协调服务，具体取决于您的群集中包含的节点总数，包括新的专用节点。有关详细信息，请参见[部署协调服务整体](#)。

5. 运行以下命令以仅将此节点专用于仅执行流程相关操作。有关节点角色的详细信息，请参见[Tableau Server 中的节点角色](#)。

- 获取专用节点的节点 ID，以查看每个节点上的服务列表：
- `tsm topology list-nodes -v`。
- 使用通过运行上述命令获得的节点 ID 为专用节点设置节点角色：
- `tsm topology set-node-role -n <nodeID> -r flows`。
- 应用更改并重新启动服务器：
- `tsm pending-changes apply`。
- 查看状态以确保所有进程均已启动并运行，并且配置正确：
- `tsm status -v`。

6. 在此阶段，您可能在其他节点上启用了 Tableau Prep Conductor。默认情况下，节点上的后台程序进程执行所有类型的任务，包括流程任务。若要隔离 Tableau Prep Conductor 和流程任务以便仅在某些节点上运行，您可以配置后台程序以执行以下操作之一：

- 若要运行流程任务：`tsm topology set-node-role -n <nodeID> -r flows`。
- 若要运行除流程之外的所有其他任务：`tsm topology set-node-role -n <nodeID> -r no-flows`。

您已成功将 Tableau Prep Conductor 添加到 Tableau Server 安装中。

使用基于内核的许可证的 Tableau Server 安装

用于生产 Tableau Server 安装的推荐拓扑是用于运行流程的专用节点。有关详细信息，请参见[Tableau Server 的最低硬件要求和建议](#)。

适用于基于内核的许可证的数据管理加载项包括用于为 Tableau Server 启用 Tableau Prep Conductor 的产品密钥，以及 Tableau Prep Conductor 内核(以 4 为单位)。应将 Tableau Prep Conductor 内核应用于专用于运行流程的节点。像其他服务器密钥一样，这些产品密钥通过[客户门户](#)提供。

若要详细了解 Tableau Prep Conductor 许可，请参见[针对 Tableau Server 为 Tableau Prep Conductor 授予许可](#)。

Tableau Server 单节点安装

如果当前有一个单节点 Tableau Server 安装，则建议您添加第二个节点并将其专用于运行流程。

1. 使用以下主题中的信息在当前 Tableau Server 安装上运行升级：
 - [Windows](#)
 - [Linux](#)
2. 激活产品密钥。这将在已启用后台程序进程的节点上启用 Tableau Prep Conductor。
 - 在 Tableau 服务管理器 Web 界面中，单击“配置”选项卡上的“许可”，然后单击“激活许可证”。
 - 输入或粘贴您的产品密钥，并单击“激活”。
 - 在“注册”页面上的字段中输入您的信息，然后单击“注册”。
3. 系统将提示您重新启动服务器。重新启动服务器并验证 Tableau Prep Conductor 是否已启用并正在运行。
 - 在 Tableau 服务管理器 Web 界面中，单击“状态”选项卡以查看状态。如果 Tableau Prep Conductor 已启用并正在运行，您应会在进程列表中看到 Tableau Prep Conductor，显示为“活动”。如果 Tableau Prep Conductor 未启用，您将在进程列表中看到 Tableau Prep Conductor，但没有状态信息。

Tableau Prep Conductor 未启用：

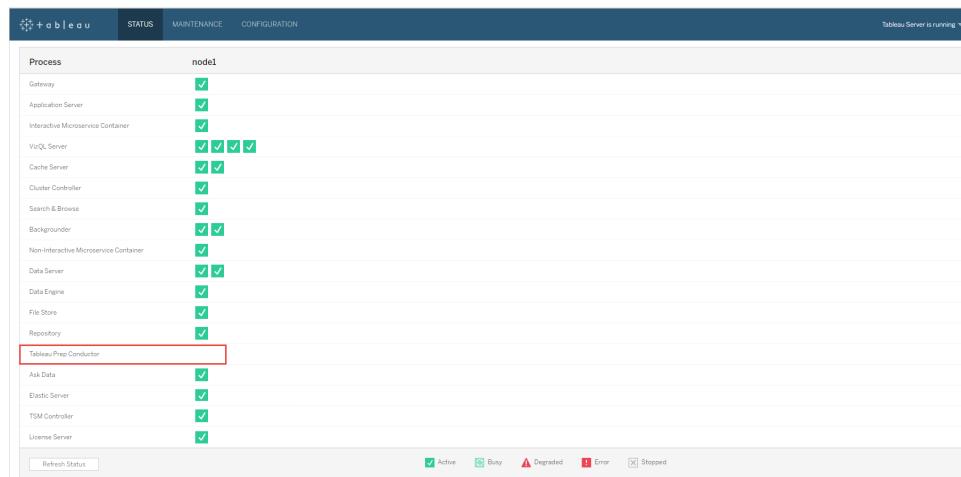
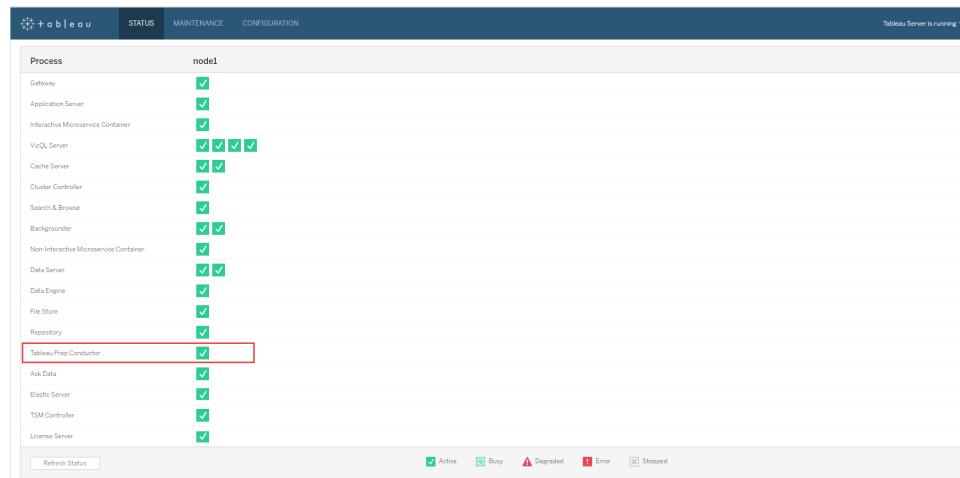


Tableau Prep Conductor 已启用并正在运行：



- 向 Tableau Server 安装中添加第二个节点。安装程序将启用某些必需的进程，如群集控制器。在运行计划的流程任务需要后台程序进程时启用该进程。启用后台程序进程时，安装程序会自动在节点上启用一个数据引擎实例和 Tableau Prep Conductor。不要在此节点上添加任何其他进程。

重要信息：此计算机上的物理内核数必须等于或小于您购买的 Tableau Prep Conductor 内核数。例如，如果购买了 4 个 Tableau Prep Conductor 内核，则节点最多只能有四个物理内核。若要了解 Tableau Prep Conductor 许可的工作方式，请参见[针对 Tableau Server 为 Tableau Prep Conductor 授予许可](#)。

- 运行以下命令以将此节点专用于仅执行流程任务。有关节点角色的详细信息，请参见[Tableau Server 中的节点角色](#)。

- 获取专用节点的节点 ID，以查看每个节点上的服务列表：

```
tsm topology list-nodes -v.
```

- 使用通过运行上述命令获得的节点 ID 为专用节点设置节点角色：

```
tsm topology set-node-role -n <nodeID> -r flows.
```

- 应用更改并重新启动服务器：tsm pending-changes apply。

- 查看状态以确保所有进程均已启动并运行，并且配置正确：

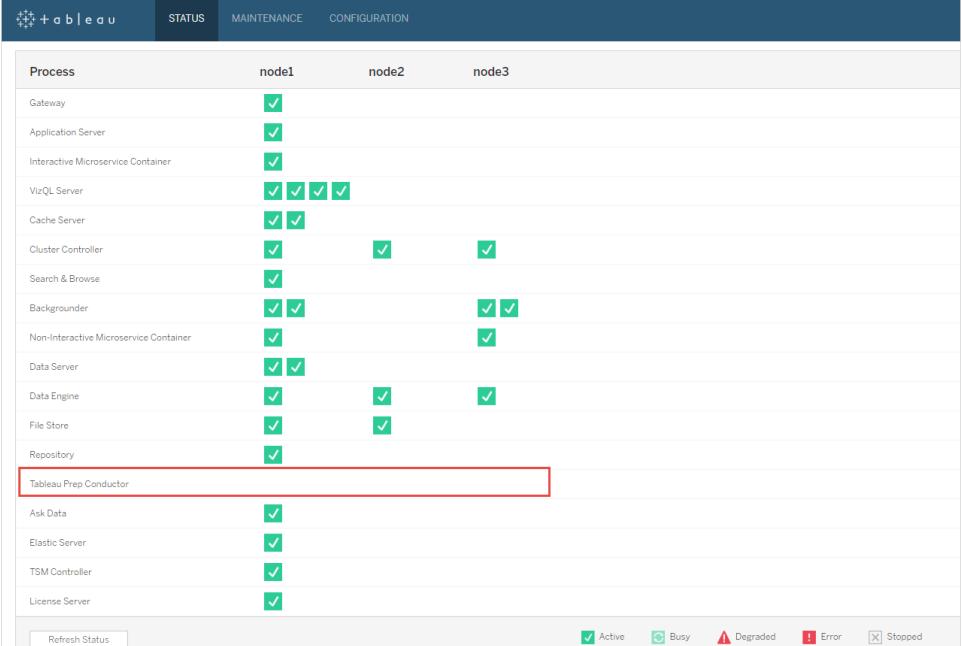
```
tsm status -v.
```

您已成功将 Tableau Prep Conductor 添加到 Tableau Server 安装中。

Tableau Server 多节点安装

- 使用以下主题中的信息在当前 Tableau Server 安装上运行升级：
 - [Windows](#)
 - [Linux](#)
- 激活产品密钥。这将在已启用后台程序进程的节点上启用 Tableau Prep Conductor。
 - 在 Tableau 服务管理器 Web 界面中，单击“配置”选项卡上的“许可”，然后单击“激活许可证”。
 - 输入或粘贴您的产品密钥，并单击“激活”。
 - 在“注册”页面上的字段中输入您的信息，然后单击“注册”。
- 系统将提示您重新启动服务器。重新启动服务器并验证 Tableau Prep Conductor 是否已启用并正在运行。
 - 在 Tableau 服务管理器 Web 界面中，单击“状态”选项卡以查看状态。如果 Tableau Prep Conductor 已启用并正在运行，您应会在进程列表中看到 Tableau Prep Conductor，显示为“活动”。如果 Tableau Prep Conductor 未启用，您将在进程列表中看到 Tableau Prep Conductor，但没有状态信息。

Tableau Prep Conductor 未启用：



Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VizQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor			
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

Refresh Status Active Busy Degraded Error Stopped

Tableau Prep Conductor 已启用并正在运行：

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VizQL Server	✓✓✓✓		
Cache Server	✓✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓✓		✓✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor	✓	✓	
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

Refresh Status ✓ Active 🕒 Busy ⚠ Degraded ❗ Error ✗ Stopped

- 向 Tableau Server 安装中添加新节点。建议为生产 Tableau Server 安装使用运行流程相关操作的专用节点。安装程序将启用某些必需的进程，如群集控制器。在运行计划的流程任务需要后台程序进程时启用该进程。启用后台程序进程时，安装程序会自动在节点上启用一个数据引擎实例。不要在此节点上添加任何其他进程。

注意：专用节点将计入协调服务整体的总数。您可能需要在新节点上部署协调服务，具体取决于您的群集中包含的节点总数，包括新的专用节点。有关详细信息，请参见[部署协调服务整体](#)。

重要信息：

此计算机上的物理内核数必须等于或小于您购买的 Tableau Prep Conductor 内核数。例如，如果购买了 4 个 Tableau Prep Conductor 内核，则节点最多只能有四个物理内核。若要了解 Tableau Prep Conductor 许可的工作方式，请参见[针对 Tableau Server 为 Tableau Prep Conductor 授予许可](#)。

- 运行以下命令以将此节点专用于仅执行流程任务。这将在新节点上启用 Tableau Prep Conductor。有关详细信息，请参见[Tableau Server 上的节点角色](#)。

- 获取专用节点的节点 ID，以查看每个节点上的服务列表：

```
tsm topology list-nodes -v.
```

- 使用通过运行上述命令获得的节点 ID 为专用节点设置节点角色：

```
tsm topology set-node-role -n nodeID -r flows.
```

- 应用更改并重新启动服务器：

```
tsm pending-changes apply.
```

- 查看状态以确保所有进程均已启动并运行，并且配置正确：

```
tsm status -v.
```

6. 在此阶段，您可能在具有后台程序进程的其他节点上启用了 Tableau Prep Conductor。

默认情况下，节点上的后台程序进程执行所有类型的任务，包括流程任务。要隔离 Tableau Prep Conductor 和流程操作以便仅在某些节点上运行，您可以配置后台程序以执行下列操作之一：

- 若要仅运行流程任务：

```
tsm topology set-node-role -n <nodeID> -r flows.
```

- 若要运行除流程之外的所有其他任务：

```
tsm topology set-node-role -n <nodeID> -r no-flows.
```

后续步骤

步骤 2: 配置 Tableau Server 的流程设置 向下。

谁可以执行此操作

Tableau Server 管理员可以安装或升级 Tableau Server，并在 Tableau Server 上启用 Tableau Prep Conductor。

步骤 2: 配置 Tableau Server 的流程设置

本主题介绍您可以为 Tableau Server 配置的各种流程设置。

发布、计划和凭据设置

当您使用数据管理产品密钥激活 Tableau Prep Conductor 时，会为整个 Tableau Server 安装启用 Tableau Prep Conductor。您可以进一步修改和自定义站点的设置。

按照以下说明配置与所有站点或单独站点的流程相关的设置：

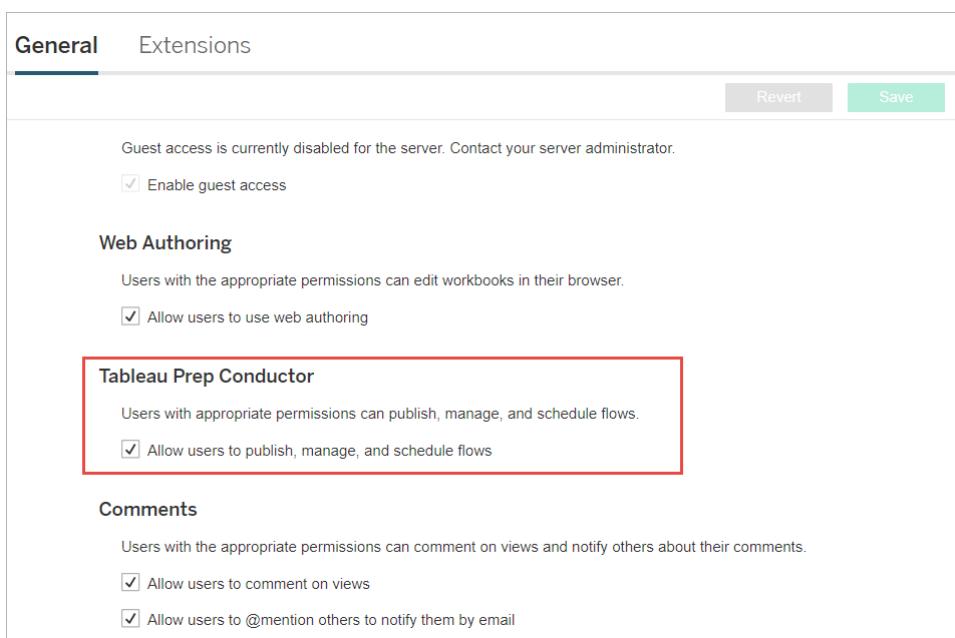
使用以下说明登录到 Tableau Server 管理页面：

- Windows: [Tableau Server 管理页面](#)。
- Linux: [Tableau Server 管理页面](#)。

配置是否允许站点发布和计划流程：

1. 允许用户发布和计划流程：在启用 Tableau Prep Conductor 时默认启用此设置。如果您有多个站点，则可以有选择地为单独的站点关闭 Tableau Prep for Server。如果您为曾经允许流程的站点禁用此设置，请参见[禁用 Tableau Prep Conductor 的含义对页](#)了解详情。

在“设置”下的“常规”页面上，滚动到“流程”设置，并清除“允许用户发布和计划流程”复选框，为站点关闭 Tableau Server。



2. 嵌入凭据

- 允许发布者在数据源、流程或工作簿中嵌入凭据：此设置允许发布者将密码附加到将自动验证 Web 用户身份的已发布流程。
- 允许发布者计划流程运行和数据提取刷新：仅当启用上述设置时，此选项才可用。启用此设置后，发布者将在“发布”对话框中看到计划选项。

禁用 Tableau Prep Conductor 的含义

如果在使用 Tableau Prep Conductor 一段时间后将其禁用，您将无法看到与流程、计划、任务以及与流程相关的其他内容。下表提供了您在完全禁用 Tableau Prep for Server 或仅针对特定站点将其禁用的情况下可看到以及无法看到的内容的相关详细信息。

	在服务器级别未启用 Prep	已在服务器级别启用 Prep , 但对站点禁用	为服务器和站点均启用了 Prep
显示流程	否	否	是
在“服务器”视图中显示任务/计划	否	是	是
在“站点”视图中显示任务/计划	否	否	是
显示站点设置(仅适用于服务器管理员)	是(禁用)	是	是
显示 TSM 状态	是(不显示 Tableau Prep Conductor)	是	是
显示 TSM 设置	是(禁用)	是	是

重要提示: 即使为该站点禁用 Tableau Prep Conductor, 计划任务仍将继续运行, 但将失败。

为流程失败配置通知

您可以配置 Tableau Server 以便针对流程运行失败发送电子邮件通知。将为通过计划任务运行流程或使用“立即运行”菜单选项手动运行时发生的失败情况发送通知。必须首先启用服务器范围设置, 然后在站点级别配置。

启用服务器范围电子邮件通知

您可以使用 Tableau 服务管理器 (TSM) Web 界面或 TSM CLI, 如下所述:

使用 TSM Web 界面

1. 在浏览器中打开 TSM:

[https://<tsm-computer-name>:8850。](https://<tsm-computer-name>:8850)

2. 在“配置”选项卡上单击“通知”，然后单击“电子邮件服务器”。

3. 输入电子邮件服务器信息。

4. 单击“事件”选项卡。

5. 在“内容更新”下，选择“为数据提取刷新失败和流程运行失败发送电子邮件”(如果默认情况尚未启用)。

6. 输入配置信息之后，单击“保存待定更改”。

7. 单击“应用更改并重新启动”。

使用 TSM CLI

可以使用 tsm configuration set 命令单独设置通知值：

Windows: [tsm configuration](#)。

Linux: [tsm configuration](#)。

设置通知值

将 tsm configuration set 命令与以下语法结合使用

若要启用流程失败通知，请运行以下命令：

```
tsm configuration set -k backgrounder.notifications_enabled -v  
true
```

注意：这将为数据提取刷新失败和流程失败启用电子邮件通知。

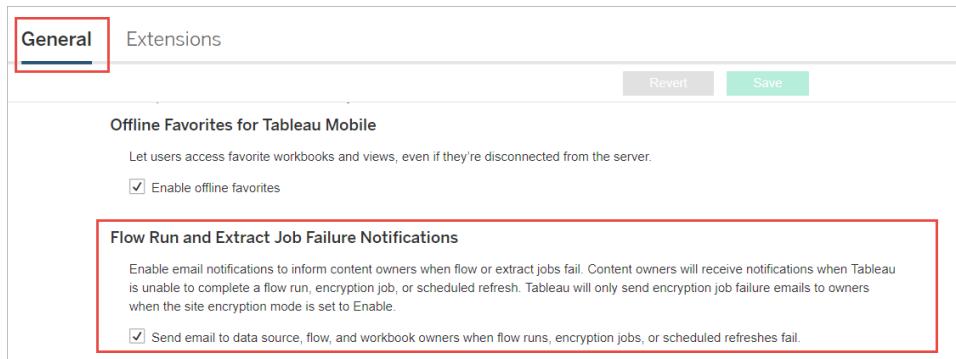
设置完值后，必须运行以下命令：

```
tsm pending-changes apply
```

`pending-changes apply` 命令将显示一条提示，告知您此命令将重新启动 Tableau Server(如果服务器正在运行)。即使服务器已停止，提示也会显示，但在这种情况下不会重新启动。您可以使用 `--ignore-prompt` 选项隐藏提示，但这样做不会改变重新启动行为。

为站点配置电子邮件通知：

在“设置”下的“常规”页面上，滚动到“流程运行”和“刷新失败通知”设置，清除“在流程运行和计划的刷新失败时向数据源、流程和工作簿发送电子邮件通知”复选框，以关闭刷新失败和流程运行通知。



后续步骤

步骤 3: 为流程任务创建计划 向下

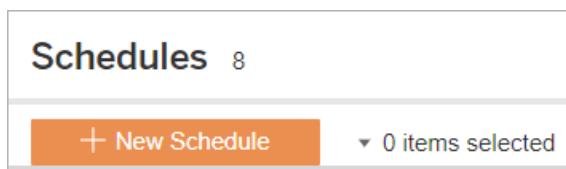
谁可以执行此操作

Tableau Server 管理员可以配置服务器和站点级别设置。Tableau 站点管理员可以配置站点级别设置。

步骤 3: 为流程任务创建计划

创建新计划：

1. 在“计划”选项卡上，单击“新建计划”。



2. 在“新建计划”对话框中输入以下信息，然后单击“创建”。

- 名称:** 输入计划的描述性名称。通常，这包括计划频率的说明。
- 任务类型:** 选择“流程”作为任务类型。
- 默认优先级:** 您可以定义从 1 到 100 的默认优先级，其中 1 为最高优先级。默认情况下，将为任何分配此值。如果队列中有两个任务待定，则优先运行具有较高优先级的任务。
- 执行:** 选择计划是并行运行还是按顺序运行。并行运行的计划在所有可用后台程序进程上运行，因此可以更快地完成。
- 频率:** 您可以定义每小时、每日、每周或每月计划。

New Schedule

Create a schedule users can choose for running extract refreshes, flows, or subscriptions.

Name	<input type="text" value="Run flows every hour"/>
Task type	<input style="width: 100px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;" type="text" value="Flow"/>
Default priority	<input type="text" value="50"/>
Tasks are executed in priority order from 1 to 100	
Execution	<input checked="" type="radio"/> Parallel: Use all available background processes for this schedule <input type="radio"/> Serial: Limit this schedule to one background process
Frequency	<input checked="" type="radio"/> Hourly every <input style="width: 100px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;" type="text" value="1 hour"/> from <input style="width: 40px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-right: 10px;" type="text" value="12"/> : <input style="width: 40px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-right: 10px;" type="text" value="00"/> AM <input style="width: 40px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-right: 10px;" type="text" value="AM"/> to <input style="width: 40px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-right: 10px;" type="text" value="12"/> : <input style="width: 40px; height: 20px; border: 1px solid #ccc; padding: 2px; margin-right: 10px;" type="text" value="00"/> AM <input style="width: 40px; height: 20px; border: 1px solid #ccc; padding: 2px;" type="text" value="AM"/>

后续步骤

步骤 4: 安全列表输入和输出位置 对页

谁可以执行此操作

Tableau Server 管理员可以创建和修改计划。计划在服务器级别创建的，并应用于服务器上的所有站点。

步骤 4: 安全列表输入和输出位置

本主题介绍适用于此功能的规则，以及如何将网络上的目录放入安全列表中。

流程输入和输出连接可能需要连接到网络上目录中的数据库或文件。必须将要允许访问的目录放入安全列表中。只允许输入和输出连接连接到安全列表的位置中的数据。默认情况下不允许连接。

注意: 您仍可以将流程文件 (tflx) 中嵌入的流程和任何数据发布到 Tableau Server，仍然可以将流和嵌入在流文件 (tflx) 中的任何数据发布到 Tableau Server，但如果目录未包含在组织的安全列表中，则流程将无法运行。

如何将输入和输出位置放入安全列表

以下规则适用，在配置此设置时必须加以考虑：

- Tableau Server 应可访问这些目录路径。这些路径在服务器启动期间和流程运行时进行验证，在将流程发布到 Tableau Server 时不会进行验证。
- 网络目录路径必须是绝对路径，并且不能包含通配符或其他路径遍历符号。例如，`\myhost\myShare*` 或 `\myhost\myShare*` 是无效路径，并将导致所有路径都不被允许。将 `myShare` 下的任何文件夹放入安全列表的正确方法是 `\myhost\myShare` 或 `\myhost\\myShare\`。

注意: `\myhost\myShare` 配置将不允许 `\myhost\myShare1`。为了将这两个文件夹都放入安全列表，请在安全列表中将它们以 `\myhost\myShare;` `\myhost\myShare1` 形式列出。

- **Windows:**

- 值可以是 * (例如, `tsm configuration set -k maestro.input.allowed_paths -v "*")` 以允许任何网络目录或指定的网络目录路径列表 (以分号(;)分隔)。
- 如果路径包含空格或特殊字符，则必须使用单引号或双引号。是使用单引号还是双引号取决于您使用的 shell。
- 即使将值设置为 *，也不允许使用本地目录路径。

- **Linux:**

- 值可以表示任何路径的*(例如, tsm configuration set -k maestro.input.allowed_paths -v "*"), 包括本地路径(使用“native_api.internal_disallowed_paths”配置的某些系统路径除外)或路径列表(以分号(;)分隔)。
- 使用的内核版本必须等于或高于4.7。4.7之前的内核版本上不支持安全列表。若要检查内核版本,请在Linux终端中键入命令uname -r。这将显示在Linux计算机上运行的内核的完整版本。

注意:如果路径同时位于允许的列表和internal_disallowed列表中,则internal_disallowed优先。

使用以下命令创建允许的网络目录路径列表:

对于输入连接:

```
tsm configuration set -k maestro.input.allowed_paths -v your_
networkdirectory_path_1;your_networkdirectory_path_2
tsm pending-changes apply
```

对于输出连接:

```
tsm configuration set -k maestro.output.allowed_paths -v your_
networkdirectory_path_1;your_networkdirectory_path_2
tsm pending-changes apply
```

重要信息:

这些命令将覆盖现有信息,并将其替换为您提供的新信息。如果要向现有列表中添加新位置,则必须提供所有位置、现有位置和要添加的新位置的列表。使用以下命令可查看输入和输出位置的当前列表:

```
tsm configuration get -k maestro.input.allowed_paths
tsm configuration get -k maestro.output.allowed_paths
```

后续步骤

步骤 5: 可选服务器配置 对页

谁可以执行此操作

同时作为**tsmadmin**组成员的Tableau Server管理员可以使用`tsm`命令配置设置。

步骤 5: 可选服务器配置

在 Tableau Server 上启用流程发布和计划流程不需要本主题中介绍的选项。这些选项可用于根据您的要求自定义环境。

设置流程的超时期间

您可以为流程可运行多长时间设置时间限制，确保后续任务不会因为任务停滞而挂起。以下两个 **tsm** 命令选项确定在取消流程后台任务之前流程任务可运行多长时间。这两个命令共同确定流程任务的总超时值。

使用以下 **tsm** 命令更改默认超时期间：

```
backgrounder.extra_timeout_in_seconds
```

(默认值: 1800 秒或 30 分钟)

```
backgrounder.default_timeout.run_flow
```

(默认值: 14400 秒或 4 小时)

为挂起的流程任务设置阈值

默认情况下，流程任务会在连续失败 5 次后挂起。若要更改流程任务在挂起之前可失败的阈值次数，请使用以下 **tsm configuration set** 命令：

```
tsm configuration set -k backgrounder.flow_failure_threshold_for_run_prevention -v <number>
```

这会为挂起任务之前流程任务必须连续失败的次数设置阈值。这是一项服务器范围设置。

谁可以执行此操作

Tableau Server 管理员可对服务器配置进行更改。

在 Tableau Online 站点上启用 Tableau Prep Conductor

注意: 当前不支持 Tableau Bridge。您只能直接连接到云托管数据源。如果连接到本地数据源，您可以将数据源转换为已发布数据源，并且，如果针对数据源配置了 Tableau

Bridge，可以使用 Tableau Bridge 客户端连接到数据。有关详细信息，请参见[发布流程在本页383。](#)

购买数据管理加载项后，Prep Conductor 会自动在您的 Tableau Online 站点上启用。有关如何购买数据管理加载项的详细信息，请与您的客户经理联系。

步骤 1：安装 Tableau Prep Builder

在本地计算机上安装 Tableau Prep Builder。

步骤 2：登录到 Tableau Online

1. 通过“服务器”->“登录”菜单登录到服务器。有关如何登录到 Tableau Online 的详细信息，请参见[登录 Tableau Online](#)。
2. 在“设置”下的“常规”页面上，如果您看到一个显示“**数据提取作业和流程运行的通知**”的部分，则 Tableau Online 已获得数据管理加载项的许可，并且 Prep Conductor 已启用。

注意：如果您看到一个仅显示“**数据提取作业的通知**”的部分，则 Tableau Online 未获得数据管理加载项的许可，并且 Prep Conductor 不可用。请与您的客户经理联系以购买数据管理加载项。

3. 验证是否已为您的站点启用了流程运行失败通知电子邮件。默认情况下，此项应已启用。

将流程发布到 Tableau Server 或 Tableau Online

将您的流程发布到 Tableau Server 或 Tableau Online，以按计划自动运行它们，并使用 Tableau Prep Conductor 刷新流程输出。

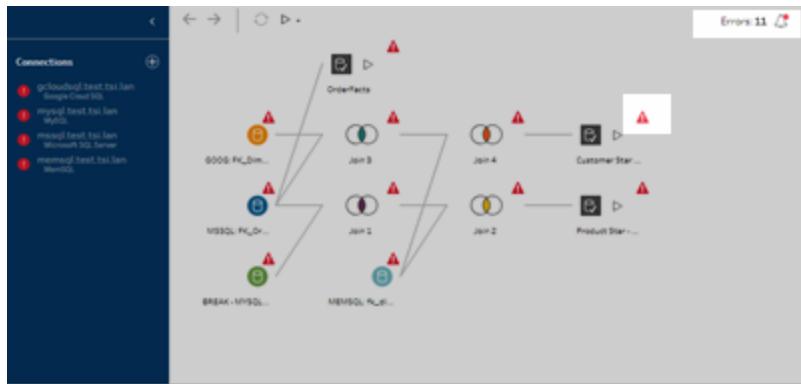
注意：本主题中的内容既适用于 Tableau Server，也适用于 Tableau Online，特别指出了一些例外情况。

发布之前

为了确保可在 Tableau Server 或 Tableau Online 中运行流程，请检查以下各项：

- 验证流程中是否没有错误。

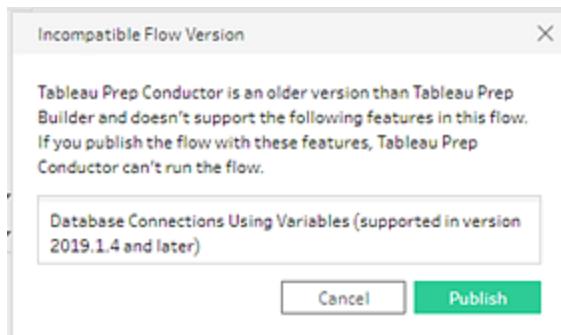
当您尝试在 Tableau Server 或 Tableau Online 中运行流程时，包含错误的流程将失败。流程中的错误由红色感叹号以及画布右上角带有“错误”指示符的红点标识。



- 验证流程未包含与您的 Tableau Server 版本不兼容的输入连接器或功能。Tableau Online 应始终运行最新版本。

如果流程包括与您的 Tableau Server 版本尚不支持的连接器或功能，则可以发布这些流程，但无法将它们安排为在 Tableau Server 中运行。若要发布流程并将流程计划为在 Tableau Server 上运行，您必须使用 Tableau Server 版本 2019.1 或更高版本，并且必须启用 Tableau Prep Conductor。

例如，Tableau Prep Builder 版本 2019.1.4 中引入了 SAP HANA 连接器，但在 Tableau Prep Conductor 的 Tableau Server 版本 2019.2 之前，此连接器不受支持。发布流程时，您会看到类似于如下示例的消息。

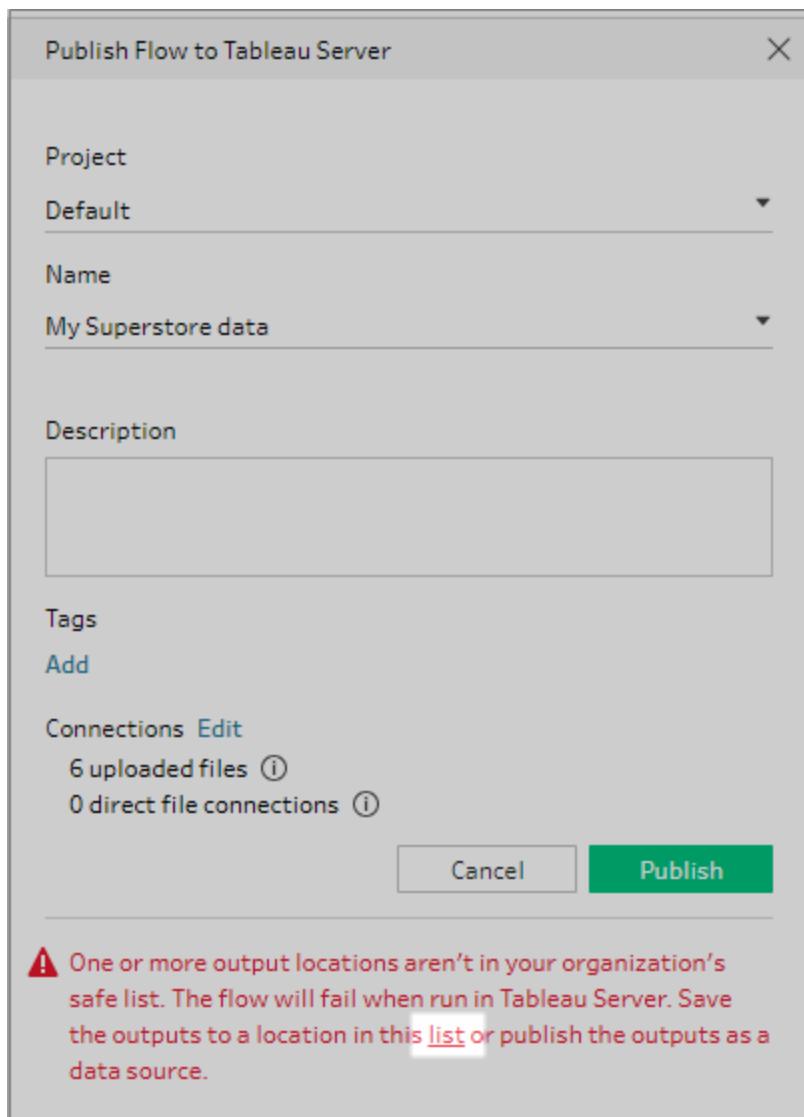


为了在 Tableau Server 中运行流程，您需要采取适当的措施来使流程兼容。有关处理不兼容流程的详细信息，请参见[与 Tableau Prep 的版本兼容性 在本页 333](#)。

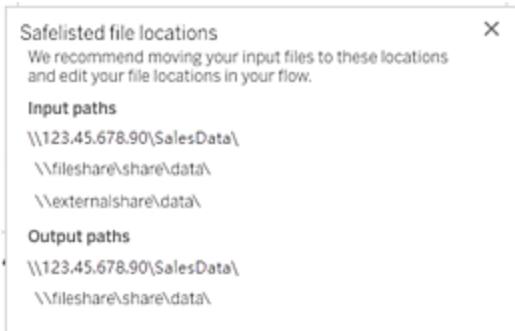
3. 包含与网络共享连接的输入或输出步骤的流程需要安全列表。Tableau Online 不支持此选项，并且必须在发布时将文件与流程打包在一起。

除非服务器可以访问文件和路径，并且文件和路径包括在组织的安全列表中，否则不允许使用指向存储在网络共享(UNC 路径)中的文件的流程输入和输出步骤。如果在没有将文件位置添加到安全列表的情况下发布流程，则流程将发布，但当您尝试并计划或在 Tableau Server 中运行流程时，将会出现错误。

如果文件未存储在安全列出的位置，您将在发布流程时看到一条警告消息。



单击消息中的“列表”链接以查看允许位置的列表。将文件移动到列表中的某个位置，并确保流程指向这些新位置。



在 Tableau Server 中，若要配置允许的网络路径，请使用[步骤 4：安全列表输入和输出位置](#)在本页 375 中描述的 tsm 命令选项。

如果不将文件移到安全列出的位置，则需要将输入文件与流程打包在一起，并将流程输出发布到 Tableau Server 作为已发布数据源。有关设置这些选项的详细信息，请参见本主题中的[发布流程 对页](#)。

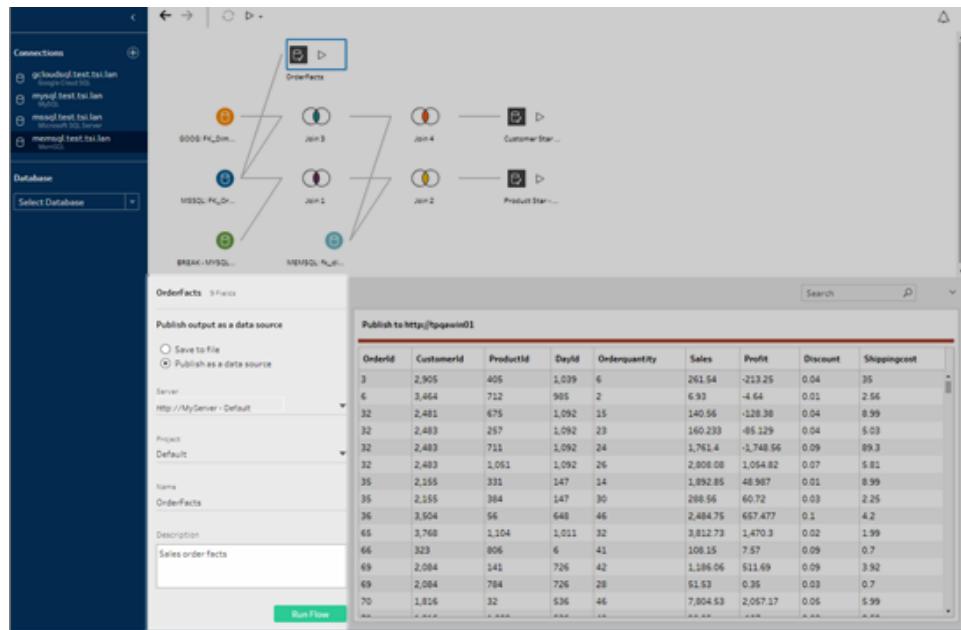
4. 确保每个流程输出步骤设置为“**以数据源形式发布**”。所有流程输出步骤都必须指向在其中发布流程的同一服务器或站点，但可指向该服务器或站点上的不同项目。只能选择一个服务器或站点。

若要为所有输出步骤设置发布位置，请执行以下操作：

- a. 在“流程”窗格中，选择输出步骤。
- b. 在发布窗格中，选择“**以数据源形式发布**”。
- c. 选择要在其中发布流程的服务器或站点以及项目。根据需要登录到服务器或站点。
- d. 输入每个输出的名称和说明。

输出文件名应足够独特，以便运行流程的人员可以轻松确定要刷新的输出文件。在 Tableau Server 或 Tableau Online 中，文件名显示在“**概述**”和“**连接**”页面上。

- e. 保存您的流程。

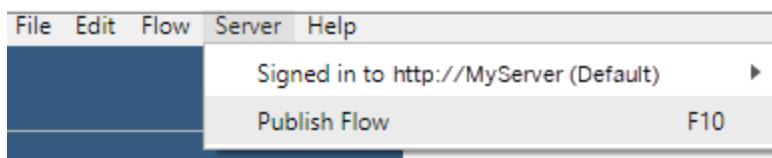


有关如何为发布配置输出步骤的详细信息，请参见[创建和发布数据提取及数据源](#)在本页310。

发布流程

注意: 在发布包括发布的数据源作为输入的流程时，系统会将发布者指派为默认流程所有者。当流程运行时，它将为“**运行身份**”帐户使用流程所有者。有关“**运行身份**”帐户的详细信息，请参见[运行身份服务帐户](#)。只有站点或服务器管理员才能在 Tableau Server 或 Tableau Online 中更改流程所有者。

1. 在 Tableau Prep Builder 中打开流程。
2. 从顶部菜单中选择“服务器”>“发布流程”。



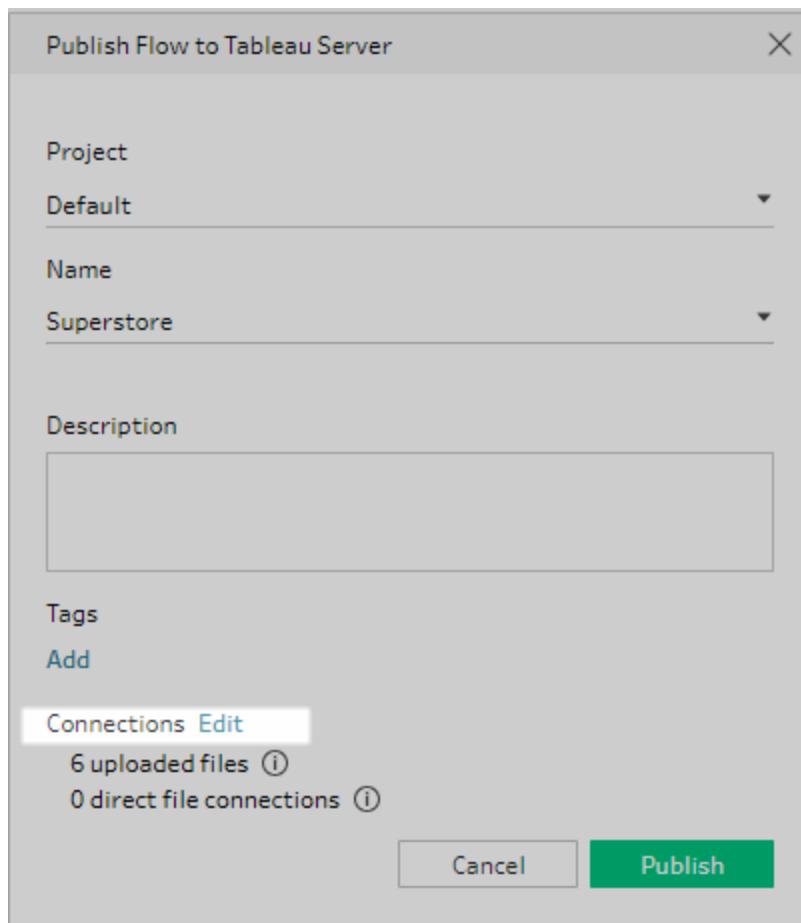
3. 针对您的平台填写字段。然后单击“发布”。Tableau Server 或 Tableau Online 会在默认浏览器中的流程“**概述**”页上自动打开。

Tableau Server

1. 在“发布到 Tableau Server”对话框中，填写以下字段：

- **项目**: 单击下拉选项可从项目分层结构中选择项目。此项目应与向其中发布输出文件的项目相同。
- **名称**: 输入流程名称。此名称显示在服务器的“流程”页面上。如果要覆盖现有流程，请单击下拉选项从列表中选择一个名称。
- **说明(可选)**: 输入流程的说明。
- **标记(可选)**: 单击“添加”键入一个或多个标记来标识流程，以便用户可以轻松找到它。发布之后，还可以在 Tableau Server 的“流程”页面中添加标记。

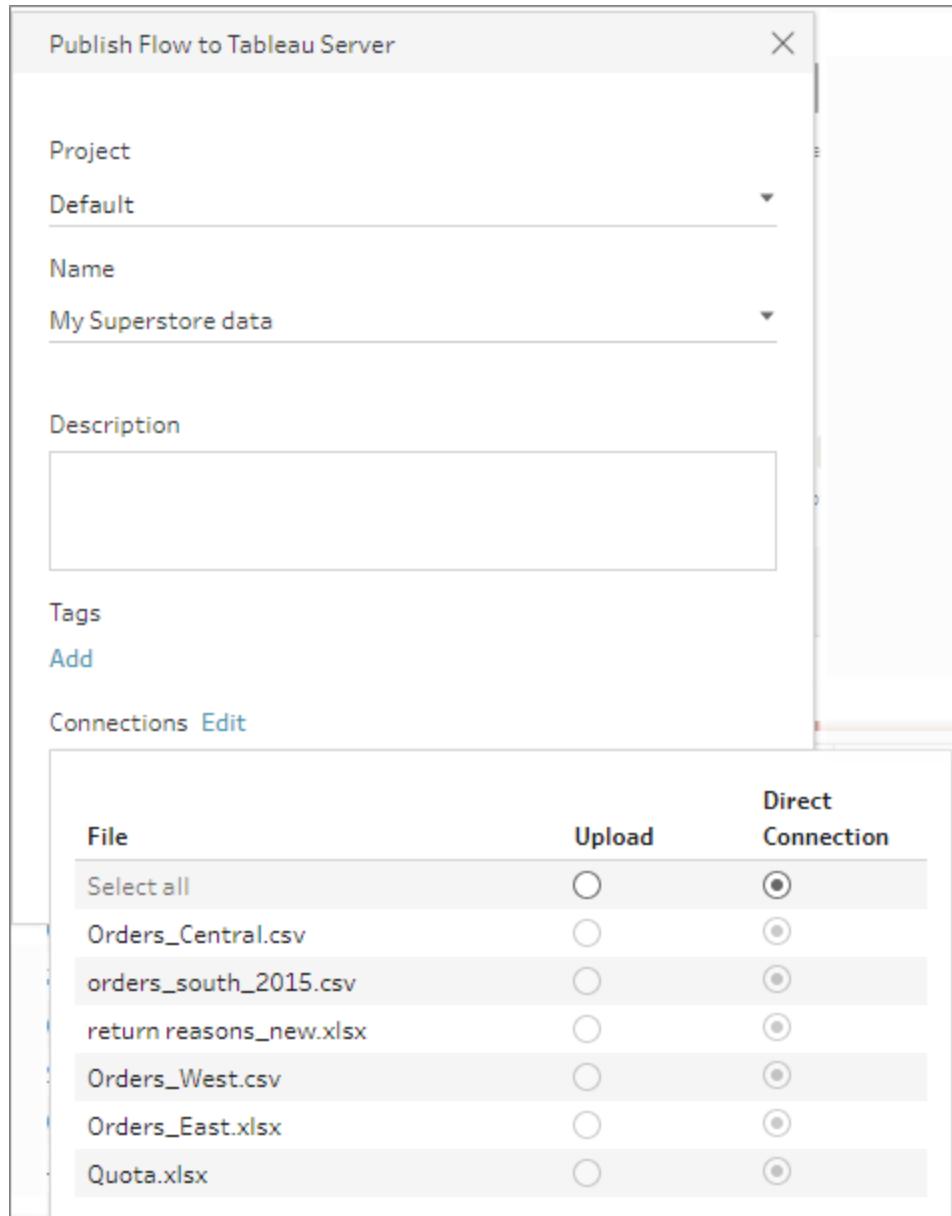
2. 在“连接”部分单击“编辑”，以编辑连接设置或更改身份验证。



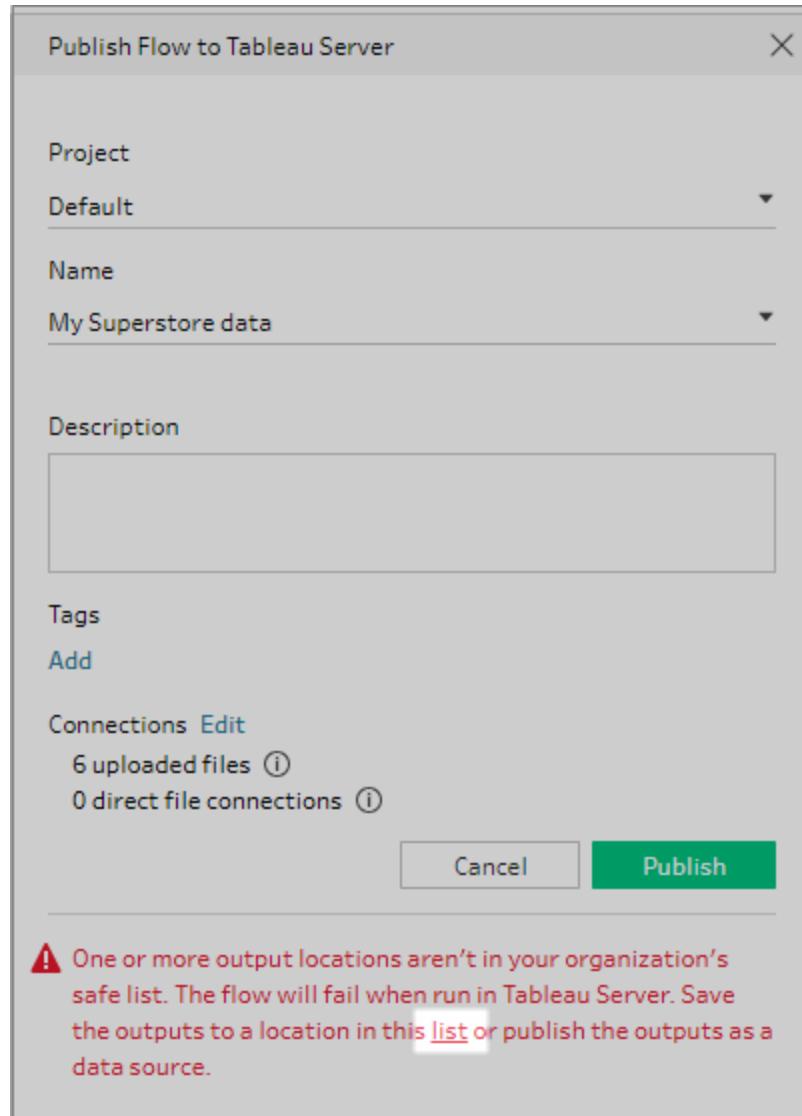
文件

默认情况下，文件输入连接与流程打包在一起。在 Tableau Server 中运行流程时，不会刷新打包的文件。所有文件必须具有相同设置，即“上载”或“直接连接”。

- 若要在刷新输出文件时检索最新数据，如果 Tableau Server 可连接到文件位置，并且该位置包括在组织的安全列表中，请选择“直接连接”。



- 如果输入或输出步骤指向存储在网络共享(UNC 路径)中的文件，并且位置未包括在组织的安全列表中，您将看到一条警告消息。单击消息中的链接以查看安全列出的位置的列表，移动文件并将输入和输出步骤指向新的文件位置。有关详细信息，请参见发布之前 在本页 379 中的“步骤 3”。



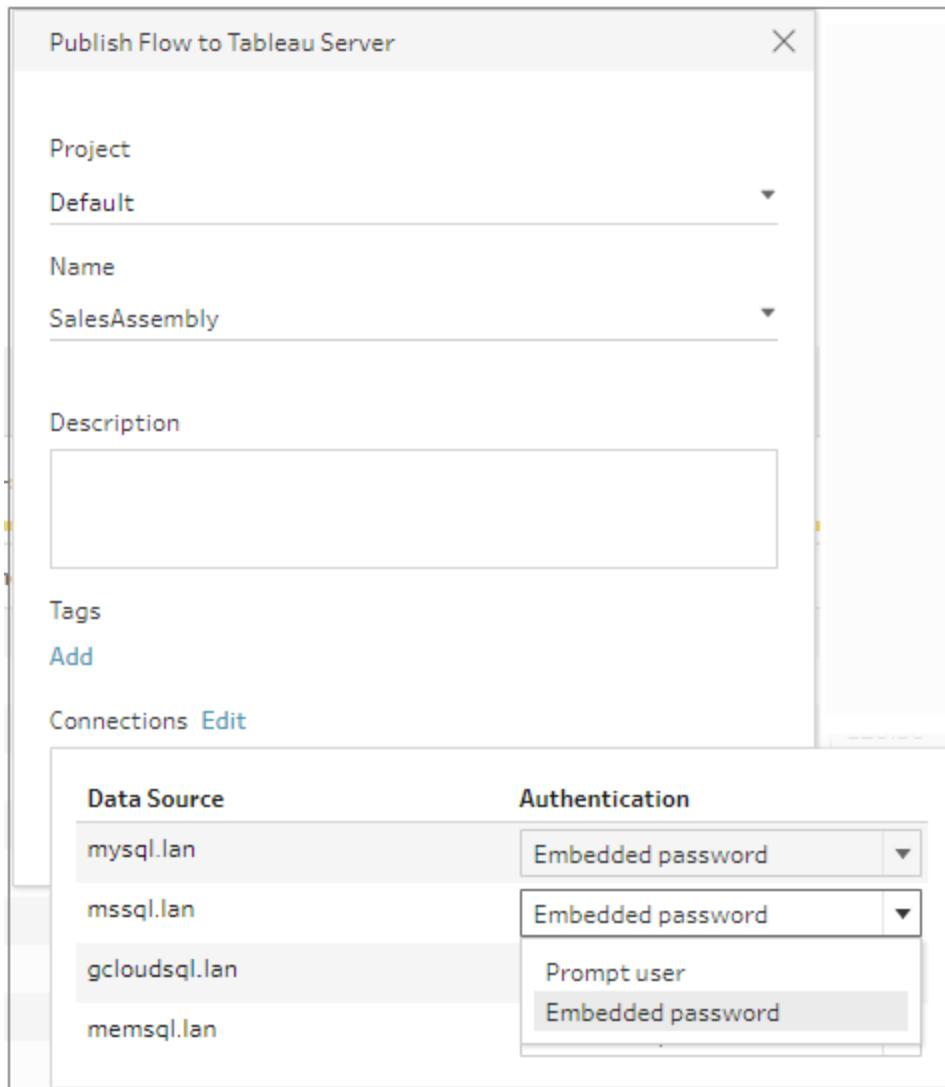
有关如何将位置添加到组织安全列表中的信息，请参见[步骤 4: 安全列表输入和输出位置 在本页 375](#)。

数据库

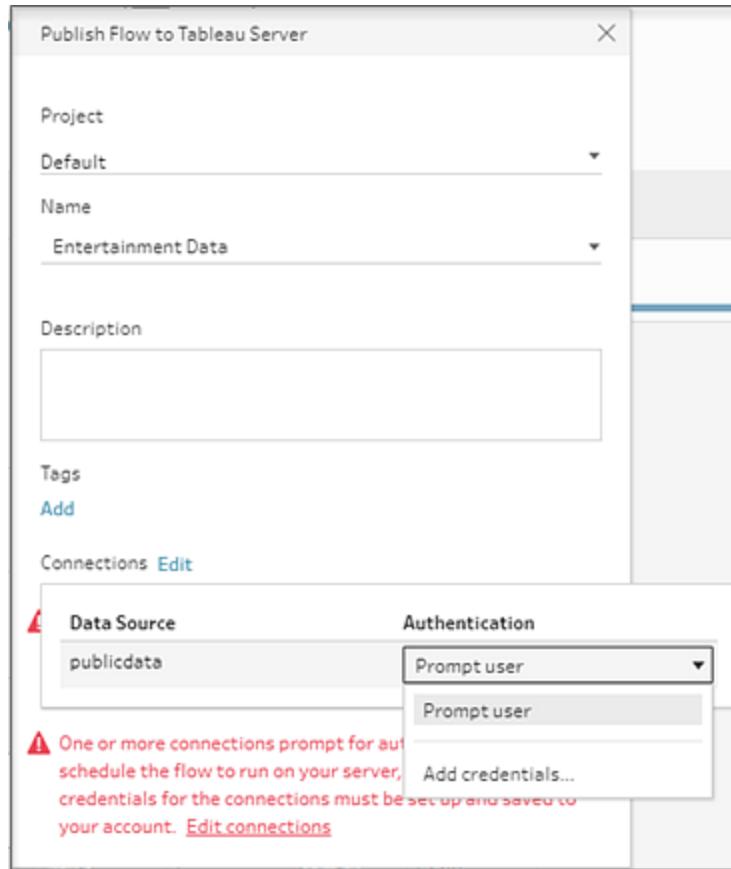
如果您的流程连接到一个或多个数据库，请选择以下身份验证类型之一以用于连接到流程输入数据源。

- **服务器运行身份帐户：**服务器的“用户运行身份”帐户将对所有用户进行身份验证。
- **提示用户：**在运行流程之前，您必须在 Tableau Server 中编辑连接并输入数据库凭据。

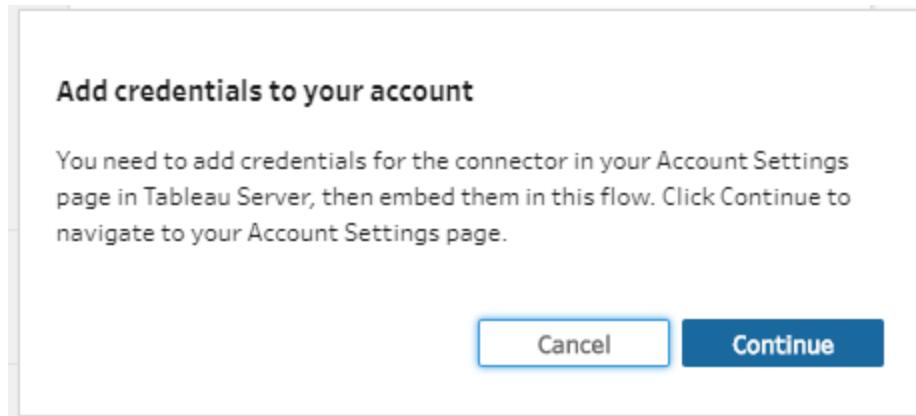
- **嵌入式密码:**用于连接到数据的凭据将随连接一起保存，并由可访问您发布的流程的每个用户使用。



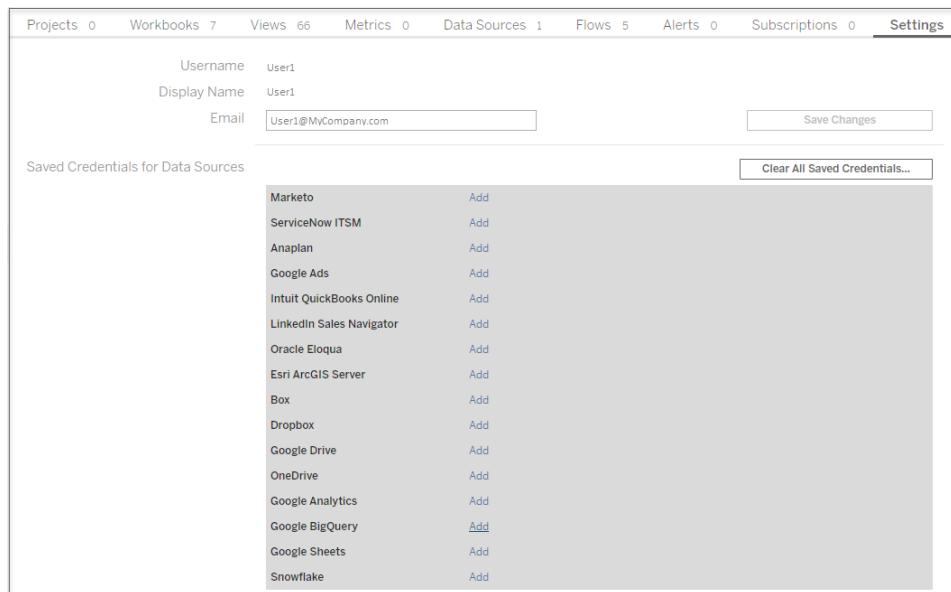
- **添加凭据:**(版本 2020.1.1 及更高版本)如果连接到云连接器，您可以直接通过“发布流程”对话框添加凭据，以将其嵌入流程中。若要添加凭据，请执行以下操作：
 - 在“连接”部分单击**“编辑”**，或从警告消息中单击**“编辑凭据”**。然后从**“身份验证”**下拉菜单中单击**“添加凭据”**。



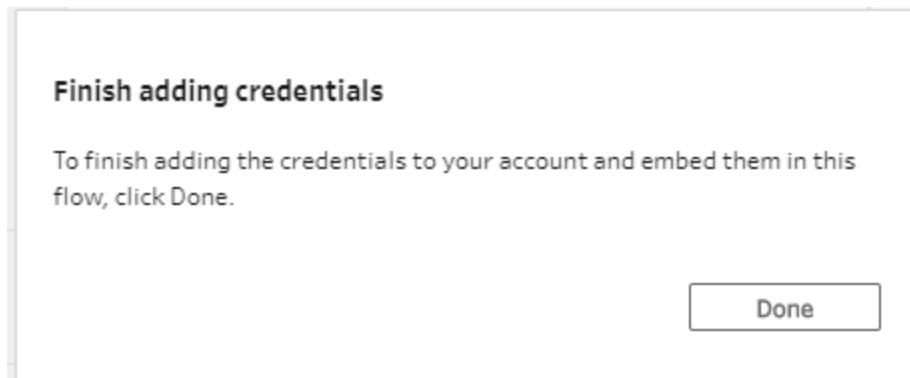
- 在确认对话框中，单击“继续”。Tableau Prep Builder 自动打开您登录的服务器的“帐户设置”页面。



- 添加您的凭据，然后导航回 Tableau Prep Builder。



- 在“完成添加凭据”话框中，单击“完成”。



- 在“连接”部分单击“编辑”，并验证凭据是否已添加并嵌入到流程中。

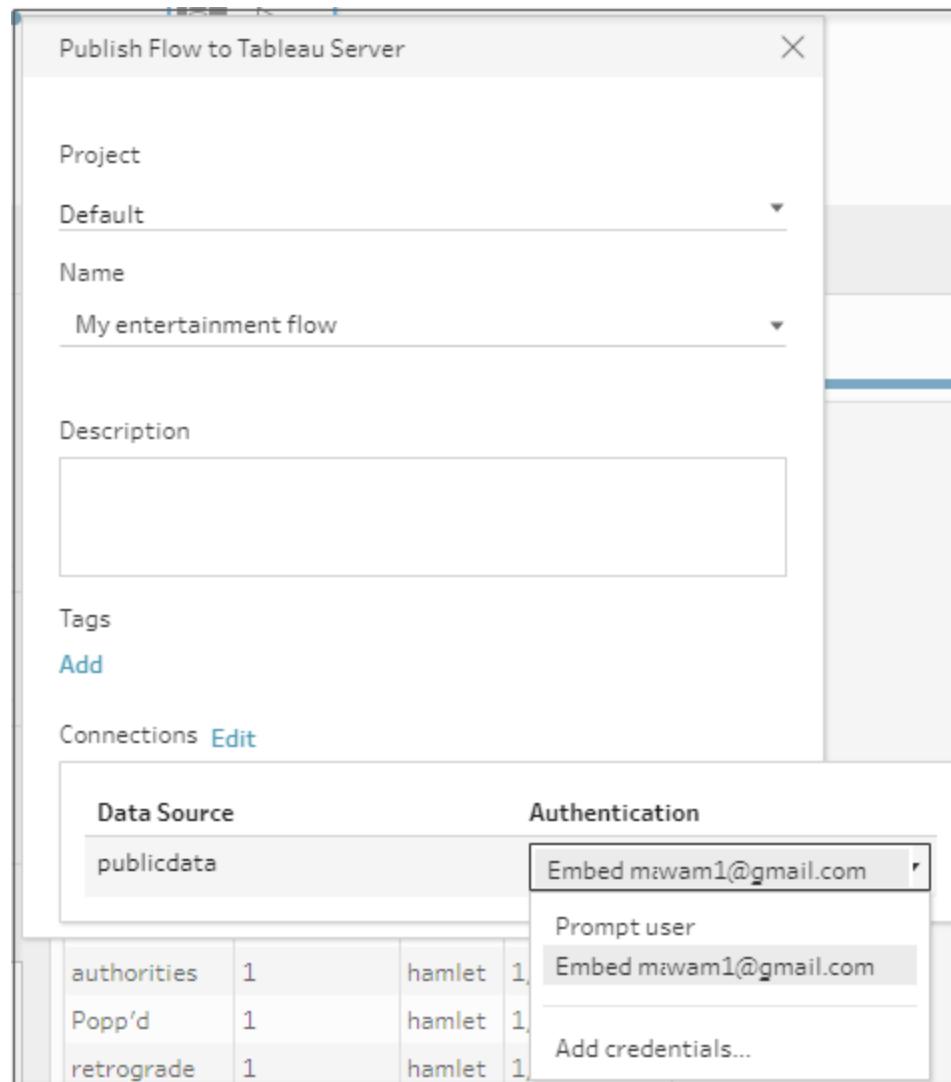
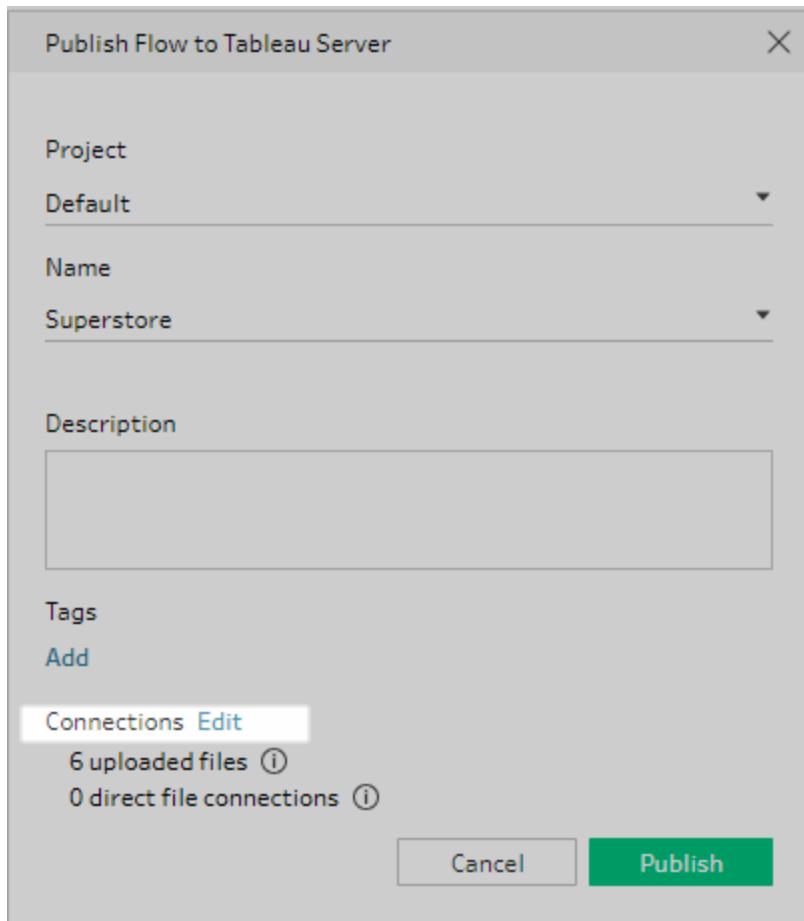


Tableau Online

1. 在“发布到 Tableau Online”对话框中，填写以下字段：

- 项目**: 单击下拉选项可从项目分层结构中选择项目。此项目应与向其中发布输出文件的项目相同。
- 名称**: 输入流程名称。此名称显示在服务器的“流程”页面上。如果要覆盖现有流程，请单击下拉选项从列表中选择一个名称。
- 说明(可选)**: 输入流程的说明。
- 标记(可选)**: 单击“添加”键入一个或多个标记来标识流程，以便用户可以轻松找到它。发布之后，还可以在 Tableau Server 的“流程”页面中添加标记。

2. 在“连接”部分单击“编辑”，以编辑连接设置或更改身份验证。



文件

Tableau Online 不支持输入步骤数据的直接文件连接，您必须将文件与流程打包在一起。在 Tableau Online 中运行流程时，不会刷新打包的文件。

数据库

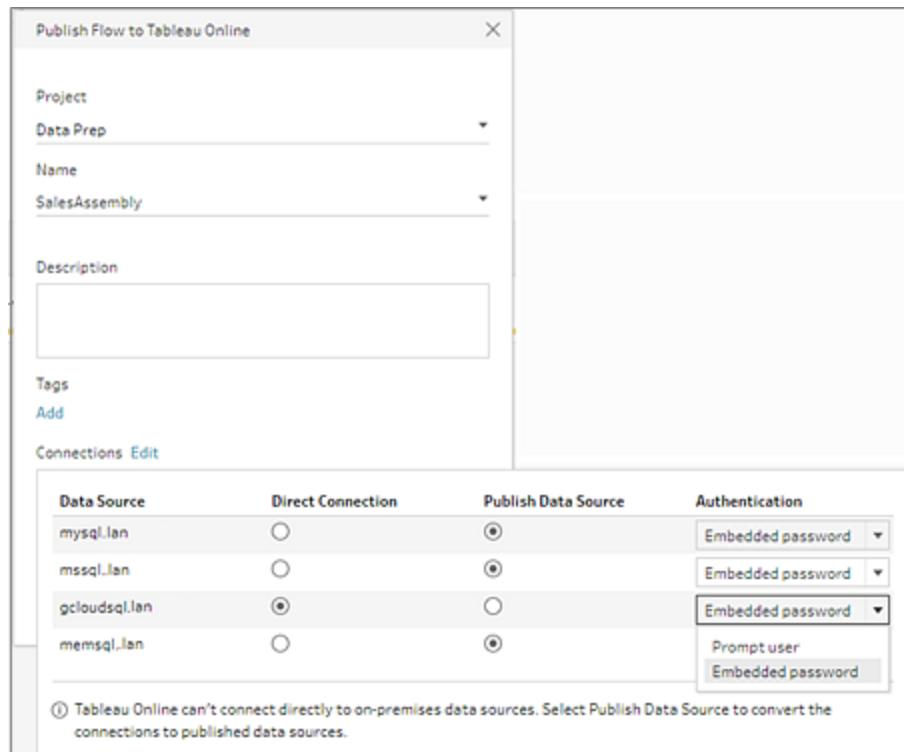
为了在将流程发布到 Tableau Online 时使数据保持最新，您只能直接连接到云托管数据源。但在连接到本地数据源时，您必须将数据源转换为已发布数据源，并且，如果针对数据源配置了 Tableau Bridge，Tableau Online 可以使用 Tableau Bridge 客户端连接到数据。

有关 Tableau Online 支持的直接连接的详细信息，请参见[允许直接连接到云平台上托管的数据](#)。

有关使用 Tableau Bridge 的详细信息，请参见[允许发布者维护本地数据的实时连接](#)。

如果您的流程连接到支持直接连接的基于云的数据源，请选择下列身份验证类型之一以用于连接到流程输入数据源。

- 提示用户:** 在运行流程之前，您必须在 Tableau Online 中编辑连接并输入数据库凭据。
- 嵌入式密码:** 用于连接到数据的凭据将随连接一起保存，并由可访问您发布的流程的每个用户使用。

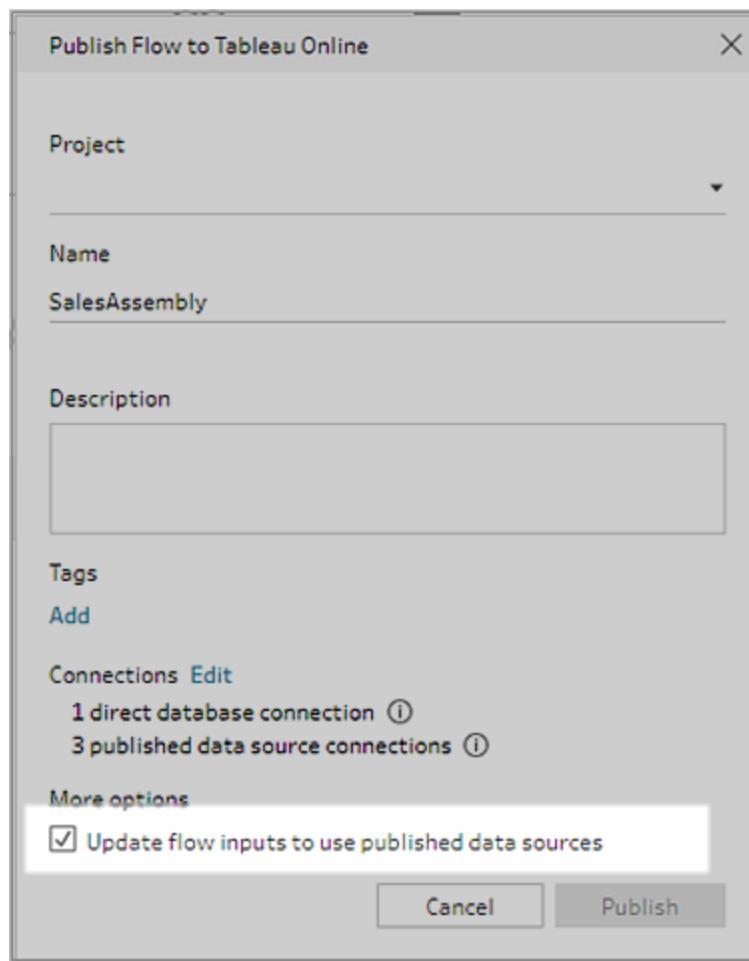


- 为本地数据源选择“**发布数据源**”单选按钮。Tableau Online 无法直接连接到这些数据源来刷新数据。在将流程发布到 Tableau Online 时，选择此选项可将数据源输入连接转换为已发布数据源。

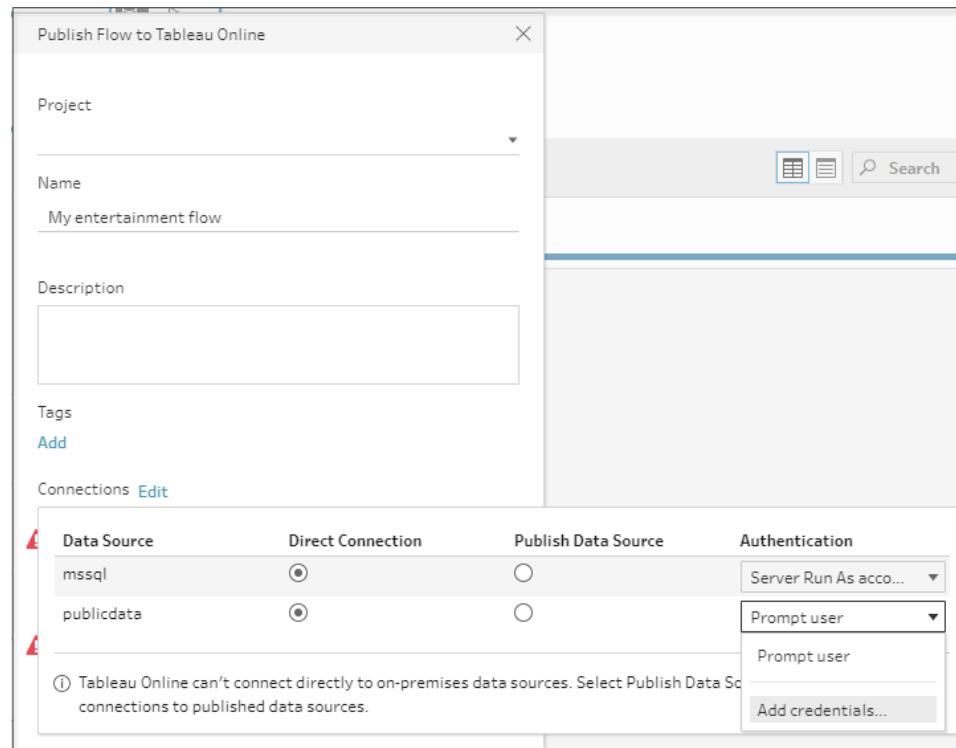
如果为数据源配置了 Tableau Bridge，并且 Tableau Online 支持该数据源，则可以在运行流程时刷新数据。有关详细信息，请参见[允许直接连接到云平台上托管的数据](#)。

- 若要将 Tableau Prep Builder 中流程的本地数据源连接替换为已发布数据源，请在发布流程之前在“更多选项”部分中选择“更新流程输入以使用发布的数据源”。

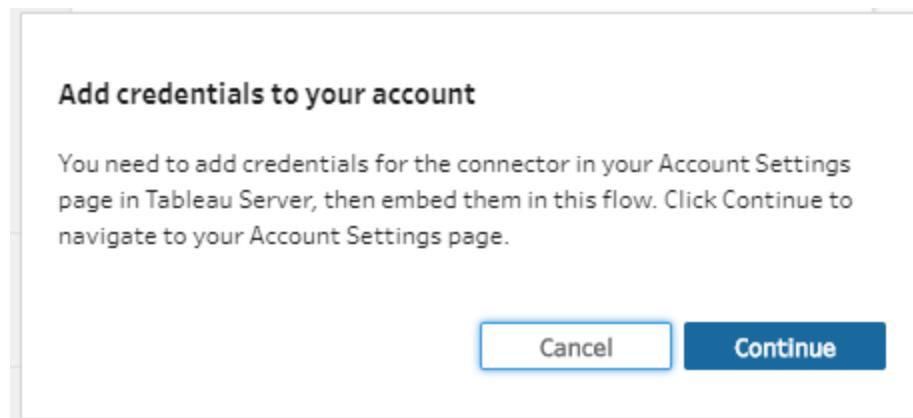
如果未选中该复选框，则 Tableau Prep Builder 中的流程会一直连接到本地数据源，并且 Tableau Prep Builder 中的流程会变得与流程的已发布版本不同步。若要继续使用流程，您将需要从 Tableau Online 中下载流程以对其进行编辑，然后重新发布流程。



- 添加凭据:(版本 2020.1.1 及更高版本)如果连接到云连接器，您可以直接通过“发布流程”对话框添加凭据，以将其嵌入流程中。若要添加凭据，请执行以下操作：
 - 在“连接”部分单击“编辑”，或从警告消息中单击“编辑凭据”。然后从“身份验证”下拉菜单中单击“添加凭据”。

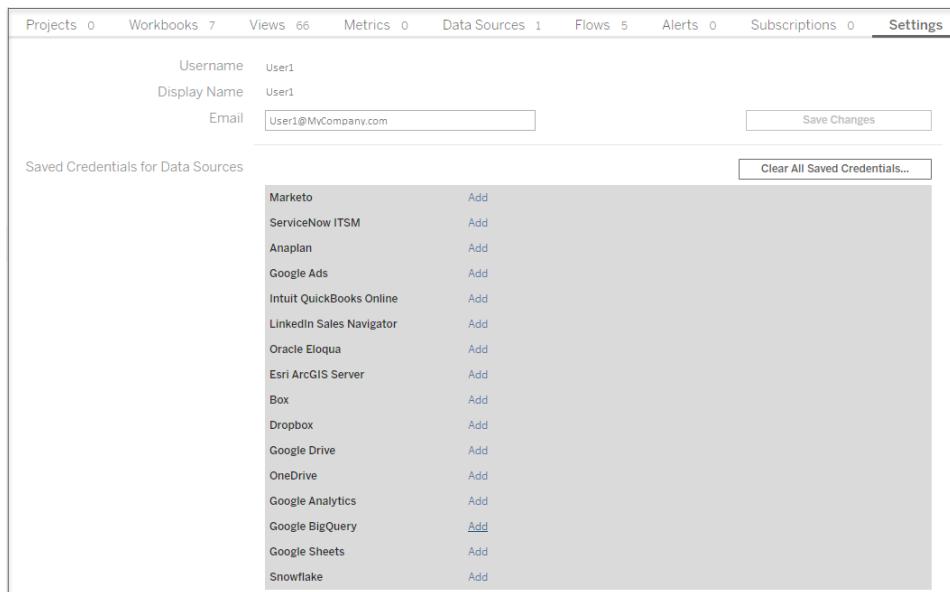


- 在确认对话框中，单击“继续”。Tableau Prep Builder 会自动打开您登录的服务器的“帐户设置”页面。

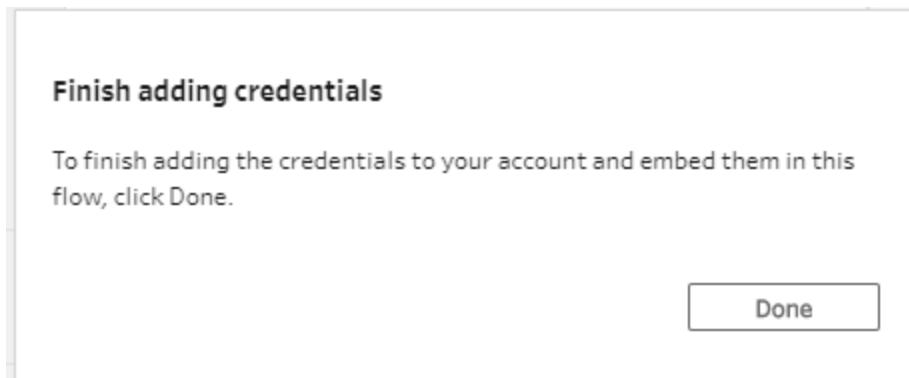


- 添加您的凭据，然后导航回 Tableau Prep Builder。

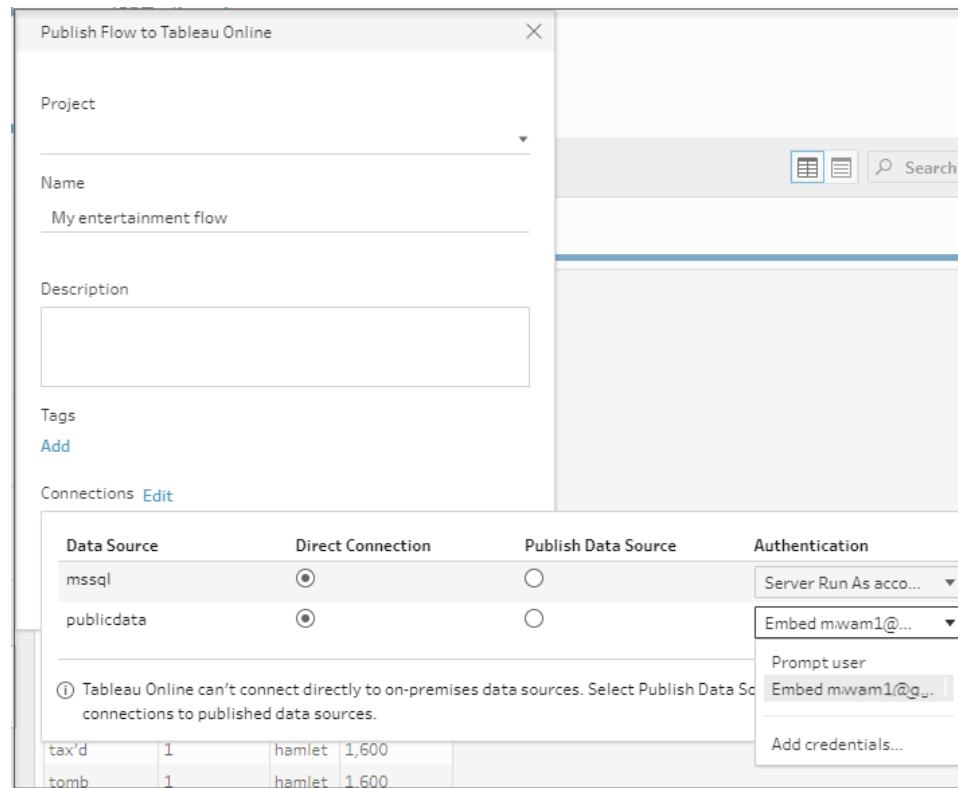
Tableau Prep 帮助



- 在“完成添加凭据”话框中，单击“完成”。



- 在“连接”部分单击“编辑”，并验证凭据是否已添加并嵌入到流程中。



谁可以执行此操作

服务器管理员、站点管理员 Creator 和 Creator 允许完全连接和发布访问权限。

Explorer(可发布) 和站点管理员 Explorer 具有有限的发布能力，如以下主题中所述：

Tableau Server:

- Windows: 允许每个站点角色所具有的一般能力。
- Linux: 允许每个站点角色所具有的一般能力。

Tableau Online: 允许每个站点角色所具有的一般能力。

计划流程任务

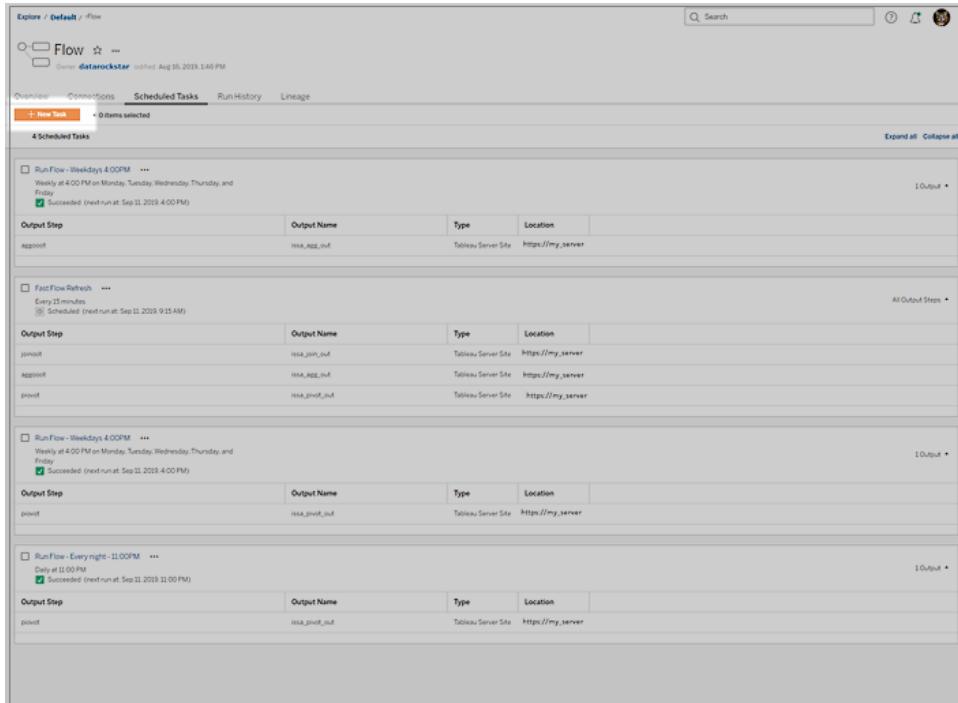
若要将流程安排在特定时间运行或定期运行，您可以创建计划任务。计划任务依赖于预先配置的计划。在 Tableau Server 中。计划由系统管理员创建。有关如何在 Tableau Server 上创建置计划的信息，请参见[步骤 3: 为流程任务创建计划](#) 在本页 373。

在 Tableau Online 中，若要为计划分配流程，请选择随 Tableau Prep Conductor 一起提供的预定义计划之一。无法在 Tableau Online 中创建新计划。

计划流程任务

- 从“计划任务”选项卡中，单击“新建任务”。

如果没有为任务分配输出步骤，您也可以通过“概述”页创建新任务。在该页面上的“计划”字段中，单击“创建新任务”。



- 在“新建任务”对话框中，从下拉列表中选择一个计划。

Tableau Server

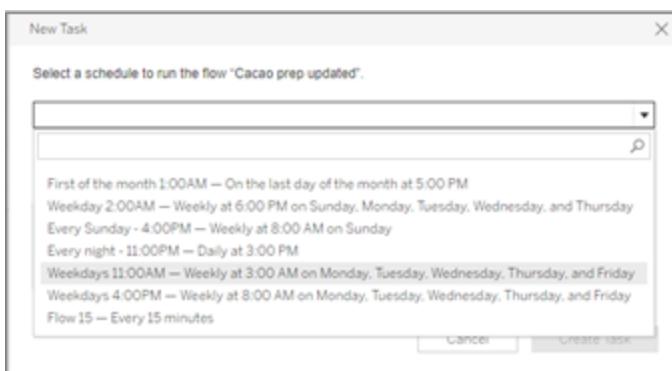
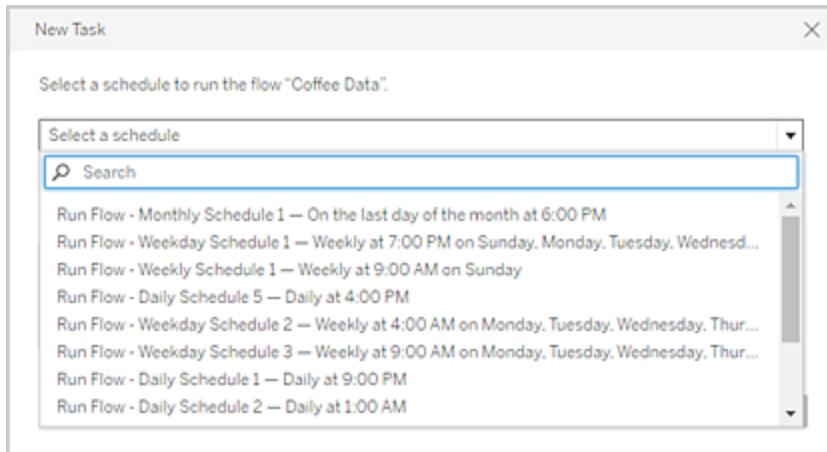


Tableau Online

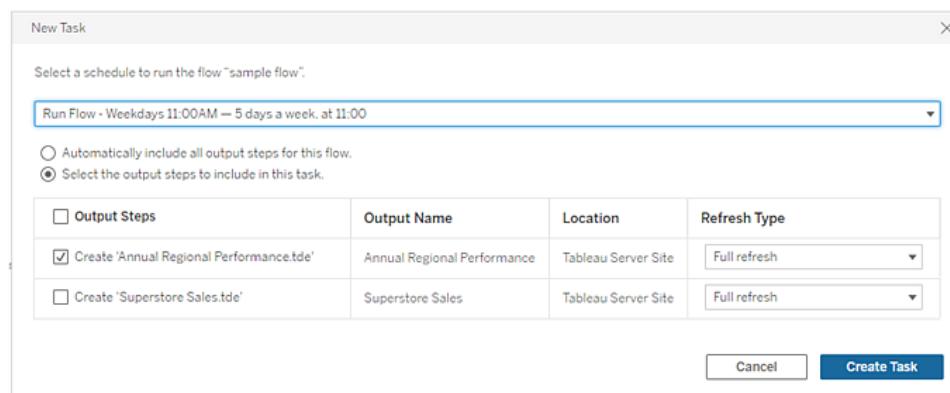


3. 选择以下选项之一：

注意:从版本 2020.2.1 及更高版本开始，如果一个输入配置为使用增量刷新，并且它与多个输出相关联，则这些输出必须一起运行，并且必须使用相同的刷新类型。否则，流程将失败。

- 自动包括此流程的所有输出步骤:**(默认)选择此选项可在计划任务中包含此流程所有当前和将来的输出步骤。随着时间的推移向流程中添加新输出步骤时，这些步骤会在流程运行时自动包括在计划中。
- 选择要包括在此任务中的输出步骤:**选择此选项可手动选择要包括在此计划任务中的输出步骤。然后选择要包括在流程任务中的输出步骤。

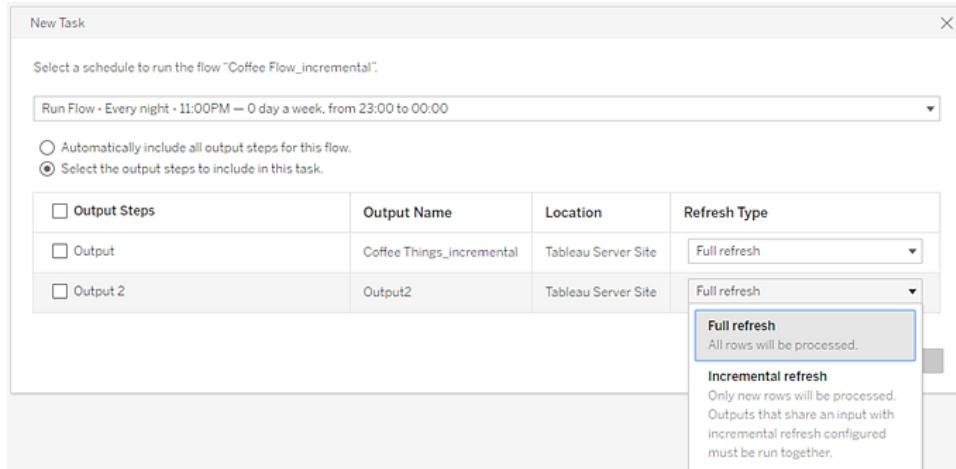
若要在流程任务中包括所有输出步骤，请选中“**输出步骤**”旁的复选框。如果选中了“**为此流程自动包括所有输出步骤**”单选按钮，则无法编辑此区域。选择其他单选按钮以启用此部分。



4. (版本 2020.2.1 及更高版本) 选择“刷新类型”。有关这些设置的详细信息,请参见[使用增量刷新来刷新流程数据 在本页 314](#)。

- **完全刷新(默认)**: 根据 Tableau Prep Builder 中的输出设置, 刷新所有数据并且创建或将数据追加到表中。
- **增量刷新**: 仅刷新新行, 并根据 Tableau Prep Builder 中的输出设置创建或将数据追加到表中。仅当在 Tableau Prep Builder 中将流程配置为使用此刷新类型时, 增量刷新选项才可用。

注意: 如果找不到现有输出, 则无论选择什么运行选项, Tableau Prep Conductor 都会为所有输出运行完全刷新。后续流程运行将使用增量刷新过程, 并仅检索和处理新行, 除非缺少增量刷新配置数据, 或者现有输出已移除。



5. 单击“创建任务”以创建计划任务。

谁可以执行此操作

- 服务器管理员可以在服务器的所有站点上执行此操作。如果启用了允许用户发布和计划任务的站点设置, 则站点管理员可以在他们有权访问的站点上执行此操作。
- 流程所有者和项目主管可以分别为他们自己的流程或项目创建流程任务。

有关详细信息, 请参见以下文章:

- Windows: [设置用户的站点角色以及内容权限和所有权](#)
- Linux: [设置用户的站点角色以及内容权限和所有权](#)

站点管理员、流程所有者、项目主管可为分别为他们拥有的流程和项目创建流程任务。

有关详细信息，请参见[设置用户的站点角色以及内容所有权和权限](#)

管理流程

将流程发布到 Tableau Server 或 Tableau Online 后，您可以管理流程并根据需要对其进行更改。本主题介绍管理流程所需进行的各种操作。

本主题中的内容同时适用于 Tableau Server 和 Tableau Online，并明确指出了例外情况。

管理流程

下面列出了您可执行的流程管理操作：

- 运行流程：**除了创建在特定时间运行的计划流程任务之外，您还可以手动运行流程。
导航到流程列表，选择要运行的一个或多个流程，选择“操作”，并单击“立即运行”。
- 标记：**标记是一些关键字，您可以为流程创建这些关键字来帮助查找、筛选和分类内容。作者可以在发布流程时向其中添加标记。但是，您还可以向允许您访问的任何工作簿、视图或数据源中添加标记，并且可以删除已添加的任何标记。您可以向流程列表添加标记。
导航到流程列表，选择要标记的一项或多项，选择“操作”，并单击“标记”。若要向特定流程中添加标记，您可以通过上述流程列表来执行该操作。从“概述”选项卡中打开“流程”，选择“操作”，然后单击“标记”。
- 更改所有者：**管理员、流程所有者和项目主管可以更改所有者。
- 权限：**可以为用户设置权限，并指定他们是否能执行像保存、下载、移动到不同项目和删除这样的编辑操作。此外，您还可以指定谁可以查看和运行流程。
- 下载：**您可以下载一个流程，以使用 Tableau Prep 进行查看或修改。若要下载流程，您需要下载权限。默认情况下，如果您是所有者，您将拥有下载权限，但您可能需要为其他用户添加该权限。
- 修订历史记录：**当对流程进行更改并使用相同名称将其重新发布到同一项目时，将创建一个新版本的流程。您可以通过从操作菜单中选择修订历史记录来查看修订历史记录。流程所有者具有还原流程的早期版本的权限。

- **移动:**可以在项目之间移动流程。若要移动流程，用户需要移动权限。默认情况下，您对自己拥有的流程具有移动权限，但您可能需要为其他用户添加该权限。
- **重命名:**可以重命名流程。若要重命名流程，用户需要保存权限。默认情况下，您对自己拥有的流程具有保存权限，但您可能需要为其他用户添加该权限。
- **删除:**可以删除流程。若要删除流程，用户需要删除权限。默认情况下，您对自己拥有的流程具有删除权限，但您可能需要为其他用户添加该权限。

谁可以执行此操作

Tableau Server 管理员

可以跨所有站点执行下列任务：

- 运行流程
- 删除流程
- 下载流程
- 更改所有者
- 更改权限
- 更改项目
- 添加/删除标记
- 更改说明
- 更改名称
- 更新流程任务
- 删除流程任务
- 创建流程任务*
- 版本管理*

*某些附加条件适用于这些操作：

- 创建流程任务：
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。

- 流程版本必须与 Tableau Server 版本兼容。
- 版本管理：
 - 必须在站点上启用修订历史记录。
 - 允许用户角色发布到项目。

Tableau 站点管理员

可以对发布到站点(它们是其管理员)的流程执行下列任务：

- 运行流程
- 删除流程
- 下载流程
- 更改所有者
- 更改权限
- 更改项目
- 添加/删除标记
- 更改说明
- 更改名称
- 更新流程任务
- 删除流程任务
- 创建流程任务*
- 版本管理*

*某些附加条件适用于这些操作：

- 创建流程任务：
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。
 - 流程版本必须与 Tableau Server 版本兼容。
- 版本管理：

- 必须在站点上启用修订历史记录。
- 允许用户角色发布到项目。

项目主管

可以对发布到项目(他们在其中具有项目主管权限)的流程执行以下任务:

- 运行流程
- 删除
- 下载
- 更改所有者
- 更改权限
- 更改项目
- 添加/删除标记
- 更改说明
- 更改名称
- 更新流程任务
- 删除流程任务
- 创建流程任务*
- 版本管理*

*某些附加条件适用于这些操作:

- 创建流程任务:
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。
 - 流程版本必须与 Tableau Server 版本兼容。
- 版本管理:
 - 必须在站点上启用修订历史记录。
 - 允许用户角色发布到项目。

项目所有者

可以对发布到他们所拥有项目的流程执行下列任务：

- 运行流程
- 删除
- 下载
- 更改所有者
- 更改权限
- 更改项目
- 添加/删除标记
- 更改说明
- 更改名称
- 更新流程任务
- 删除流程任务
- 创建流程任务*
- 版本管理*

*某些附加条件适用于这些操作：

- 创建流程任务：
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。
 - 流程版本必须与 Tableau Server 版本兼容。
- 版本管理：
 - 必须在站点上启用修订历史记录。
 - 允许用户角色发布到项目。

流程所有者

可以对他们拥有的流程执行下列任务：

- 运行流程
- 删除流程
- 下载流程
- 更改所有者
- 更改权限
- 更改项目
- 添加/删除标记
- 更改说明
- 更改名称
- 更新流程任务
- 删除流程任务
- 创建流程任务*
- 版本管理*

*某些附加条件适用于这些操作：

- 创建流程任务：
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。
 - 流程版本必须与 Tableau Server 版本兼容。
- 版本管理：
 - 必须在站点上启用修订历史记录。
 - 允许用户角色发布到项目。

具有 Creator 许可证的用户

可以执行下列任务：

- 运行流程(使用“运行流程”权限)
- 删除(使用删除权限)
- 下载(使用下载或另存为以及读取权限)

- 更改权限(使用更改权限)
- 更改项目(使用移动权限, 以及对目标项目的写入权限)
- 添加/删除标记(使用读取权限)
- 更改说明(使用保存权限)
- 更改名称(使用保存权限)
- 更新流程任务(使用执行权限)
- 删除流程任务(使用执行权限)
- 创建流程任务*(使用执行权限)
- 版本管理*(使用查看、读取、另存为、下载权限)

*某些附加条件适用于这些操作：

- 创建流程任务：
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。
 - 流程版本必须与 Tableau Server 版本兼容。
- 版本管理：
 - 必须在站点上启用修订历史记录。
 - 允许用户角色发布到项目。

具有 Explorer 许可证的用户

可以执行下列任务：

- 运行流程(使用“运行流程”权限)
- 删除(使用删除权限)
- 下载(使用下载或另存为以及读取权限)
- 更改权限(使用更改权限)
- 更改项目(使用移动权限以及对目标项目的写入权限)
- 添加/删除标记(使用读取权限)
- 更改说明(使用保存权限)
- 更改名称(使用保存权限)

- 更新流程任务(使用执行权限)
- 删除流程任务(使用执行权限)
- 创建流程任务*(使用执行权限)
- 版本管理*(使用查看、读取、另存为、下载权限)

*某些附加条件适用于这些操作：

- 创建流程任务：
 - 流程计划必须可用。只有服务器管理员才能创建计划。
 - 流程至少必须有一个输出步骤。
 - 流程版本必须与 Tableau Server 版本兼容。
- 版本管理：
 - 必须在站点上启用修订历史记录。
 - 允许用户角色发布到项目。

具有 Viewer(查看者) 许可证的用户

Viewer(查看者) 无法管理流程，但可以查看流程和流程的不同版本。

有关可在流程上设置的完整能力的详细信息，请参见[权限能力](#)。

监视流程运行状况和性能

在发布流程并计划定期运行流程后，您希望知道它们正在按预期运行，并在任何问题出现时加以解决。您还需要监视和了解流程的性能。

本主题介绍 Tableau Server 和 Tableau Online 提供的可帮助您监视流程的各种方法。

在问题出现时检测问题并加以解决

您可以设置 Tableau Server，以便在流程失败时发送电子邮件通知，或在 Tableau Server 上使用“通知”菜单或通过查看感兴趣的流程的流程页面来查找和查看错误。此类监视允许您在出现问题时检测到问题。

在流程失败时收到电子邮件通知：

您可以将 Tableau Server 配置为在流程失败时向流程所有者发送电子邮件通知。

若要为 Tableau Server 和站点设置电子邮件通知，请按照[步骤 2: 配置 Tableau Server 的流程设置](#)在本页 369 中的说明进行操作。

Tableau Online 中默认情况下已为您的站点启用电子邮件通知。可通过转到“[设置](#)”>“[常规](#)”页面来验证这一点。

查看和解决错误

运行流程时可能会发生以下错误：

- 连接错误：**当 Tableau Server 或 Tableau Online 无法连接到一个或多个数据输入或无法在一个或多个输出步骤中进行连接时，通常会发生连接错误。
 - 对于输入连接错误，请使用“[连接](#)”选项卡上的“[编辑连接](#)”选项对连接详细信息进行更改，然后再次运行流程。
 - 对于输出连接错误，请下载流程并检查流程输出步骤的输出位置。如果流程输出进入网络共享，请确保输出步骤指向安全列出的位置。进行任何更改后，重新发布流程并尝试再次运行。
- 流程中的错误：**如果流程中一个或多个步骤中存在错误，您将看到一条错误消息。在这种情况下，请将流程下载到 [Tableau Prep Builder](#)，解决错误，将流程重新发布到服务器并再次运行流程。
- 流程任务挂起：**当计划的流程任务运行失败配置的尝试次数后，流程任务将挂起。可以为流程分配多个计划任务，但只会挂起失败的任务。所有其他流程任务将继续运行，除非它们有错误。若要解决挂起的任务，请查看并解决错误，然后按需运行流程，或者让流程基于分配的计划自动运行。

注意：默认情况下，流程任务会在连续失败 5 次后挂起。服务器管理员可以用 `tsm configuration set` 选项来更改此默认设置。有关更多详细信息，请参见[步骤 5: 可选服务器配置](#) 在本页 377。

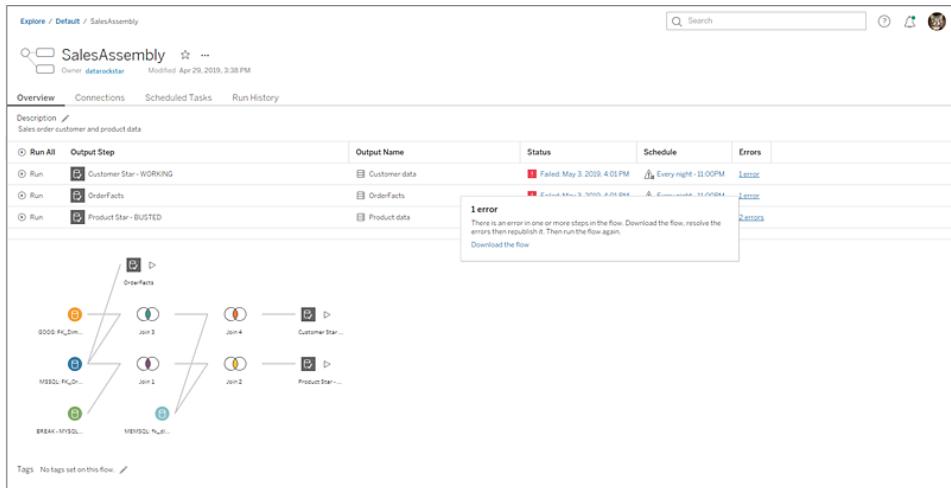
您可以在以下页面上查看错误：

“流程概述”页面

在此页面上，您可以看到最近的流程运行的状态，以及任何错误。将光标悬停在错误文本上可查看错误详细信息。如果计划任务已挂起，则计划旁边会显示一个警告图标。将鼠标指针悬停在图标上以查看状态。

Tableau Prep 帮助

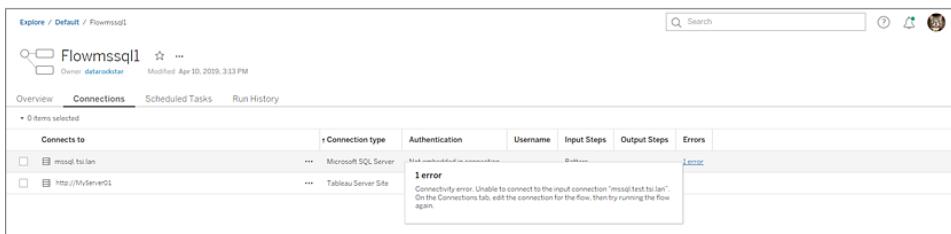
解决了导致流程失败或任务挂起的错误之后，您可以手动运行流程，或让流程基于分配的计划运行。对于挂起的流程任务，在挂起的任务的工具提示上单击“[转到挂起的任务](#)”链接，导航到“[计划任务](#)”页面，并单击“[恢复计划任务](#)”按钮恢复挂起的任务。



“连接”页面

“连接”页面显示最近的状态和任何相关连接错误。若要纠正输入错误，请单击输入连接的“[更多操作](#)”... 菜单，编辑连接并更改服务器名称、端口、用户名和密码。

若要修复输出连接错误，请在 Tableau Prep Builder 中下载流程，更正文件路径，然后重新发布流程以继续运行。



“计划任务”页面

查看分配给流程的计划任务。如果计划任务已挂起，您可以在此查看该任务的状态，并且可以手动从此页面恢复流程任务。恢复挂起的任务之前，请解决流程中的任何错误。此页面上未显示错误，但您可以在“[概述](#)”或“[运行历史记录](#)”页面上查看这些错误。

如果为流程编辑连接或手动运行流程任务，则挂起的计划任务将在重新发布流程时自动恢复。若要手动恢复挂起的任务，请在“**计划任务**”页面上单击“**恢复计划任务**”。这将恢复流程的所有已挂起任务。

若要恢复单独的任务，请单击计划任务的“**更多操作**”... 菜单，然后选择“**恢复**”。也可以单击“**立即运行**”以立即运行流程的所有任务。

Output Step	Output Name	Type	Location
Product Star - BUSTED	Product data	Tableau Server Site	http://Mserver01
Customer Star - WORKING	Customer data	Tableau Server Site	http://Mserver01
OrderFacts	OrderFacts	Tableau Server Site	http://Mserver02

“运行历史记录”页面

“**运行历史记录**”页面显示每个输出已完成或正在进行的所有流程运行的详细信息。通过将鼠标光标悬停在“**错误**”列中的错误上来查看任何错误详细信息。持续时间列显示流程的运行时间。

注意：从版本 2020.2.1 开始，“**运行类型**”字段显示输出的刷新类型。在以前的版本中，此字段显示输出是按计划运行还是按需运行。有关设置输出刷新类型的详细信息，请参见[使用增量刷新来刷新流程数据](#) 在本页 314。

Output Step	Run Type	Run Start	Run End	Duration	Status	Rows Generated	Errors
Output 2	Incremental refresh	Apr 30, 2020, 11:01 PM	Apr 30, 2020, 11:01 PM	00:00:10	Succeeded	4,248	
Output	Incremental refresh	Apr 30, 2020, 11:01 PM	Apr 30, 2020, 11:01 PM	00:00:09	Succeeded	4,248	
Output	Full refresh	Apr 21, 2020, 11:49 AM	Apr 21, 2020, 11:49 AM	00:00:03	Failed	0	1 error

通知

流程失败时，通知菜单中将会出现错误详细信息，并包含用于重新运行流程或下载流程进行故障排查的选项。

注意: 在 Tableau Server 中, 流程所有者、服务器或站点管理员可看到此菜单。在 Tableau Online 中, 流程所有者和站点管理员可看到此菜单。

Scheduled tasks are suspended for flow "Superstore"

Flow: Superstore

Suspended Status: Failed 2 consecutive times: 1 scheduled task is suspended.

[Go to Scheduled Tasks](#)

[Resume Scheduled Tasks](#)

[Download Flow](#)

The flow "Superstore" failed with 1 error

Flow: Superstore

Error: The output file path is not valid or you don't have permissions to write to the file. Check permissions, verify this file path is in the safelist, or use a valid output file path.

[Rerun Output Step "Create 'Superstore Sales.tde'"](#)

[Download Flow](#)

The flow "Superstore" failed with 1 error

Flow: Superstore

Error: The output file path is not valid or you don't have permissions to write to the file. Check permissions, verify this file path is in the safelist, or use a valid output file path.

[Rerun Output Step "Create 'Annual Regional Performance.tde'"](#)

[Download Flow](#)

Scheduled tasks are suspended for flow "SQLServerUserNamePassword Good"

Flow: SQLServerUserNamePassword Good

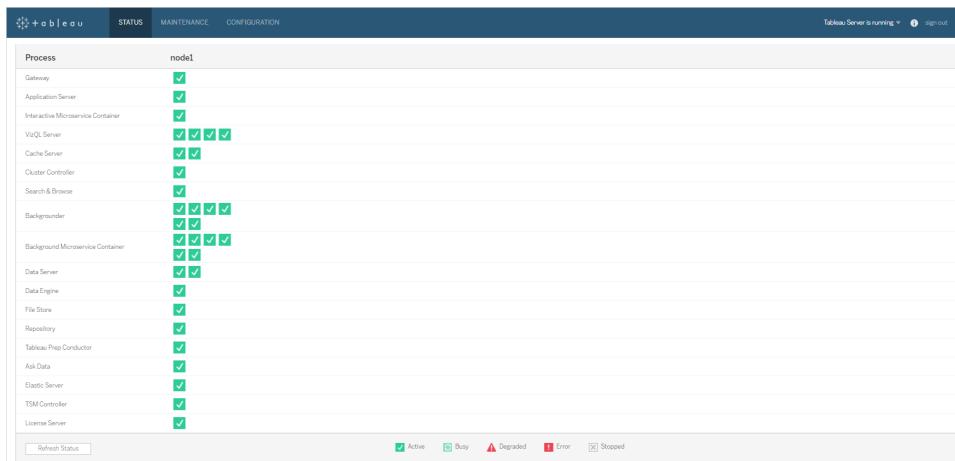
Suspended Status: Failed 2 consecutive times: 2 scheduled tasks are suspended.

[Go to Scheduled Tasks](#)

[Resume Scheduled Tasks](#)

Tableau Prep Conductor 进程状态

- 可在 TSM 中访问 Tableau 服务管理器 (TSM) 状态页面, TSM 管理员可查看该页面。您必须能够登录到 TSM 才能看到此页面。
- Tableau Server 状态页面出现在 Tableau Server Web UI 中, 可由 Tableau Server 管理员访问。如果将鼠标指针悬停在进程的状态指示器上, 工具提示将显示该进程正在其上运行的节点名称和端口。Tableau Server 状态页面不显示 TSM 进程。



当 Tableau Server 正常工作时，Tableau Prep Conductor 将显示为“活动”或“忙碌”：

- **活动** — 进程按预期运行。
- **忙碌** — 进程正在完成一些任务。
- **关闭** — 进程已关闭。根据进程，这种状态的含义有所不同。
- **状态不可用** — Tableau Server 无法确定进程的状态。

使用管理视图监视流程活动和性能

可以使用管理视图来监视流程相关活动、性能历史记录和使用的磁盘空间。“**状态**”页面包含一个嵌入的 Tableau 工作簿，其中包含可用于监视不同类型服务器或站点活动的各种管理视图。

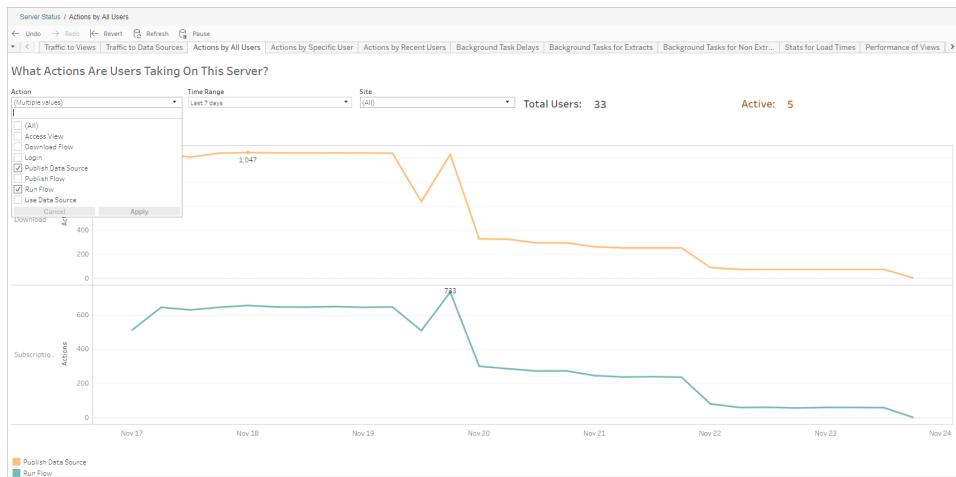
对于 **Tableau Server** 安装，必须先安装 PostgreSQL 驱动程序，然后才能看到管理视图。有关详细信息，请参见[数据库驱动程序 \(Linux\)](#)、[数据库驱动程序 \(Windows\)](#)。服务器管理员可以使用这些视图来查看服务器级别的活动(已为所有站点聚合)或特定站点的活动。只有服务器管理员才能按站点筛选。

谁可以执行此操作？

- **Tableau Server:** Tableau Server 管理员和 Tableau 站点管理员均可查看和使用管理视图。只有服务器管理员才能按站点筛选。
- **Tableau Online:** Tableau 站点管理员可以查看和使用管理视图。

所有用户的操作

使用此视图深入了解所使用流程的情况。其中包括发布、下载和流程运行等操作。可以按操作、站点和时间范围筛选视图。“用户总数”计数显示已执行操作的用户数。此值不受任何筛选的影响。“活动用户”计数显示在选定时间段内处于活动状态并执行所选操作之一的用户数。

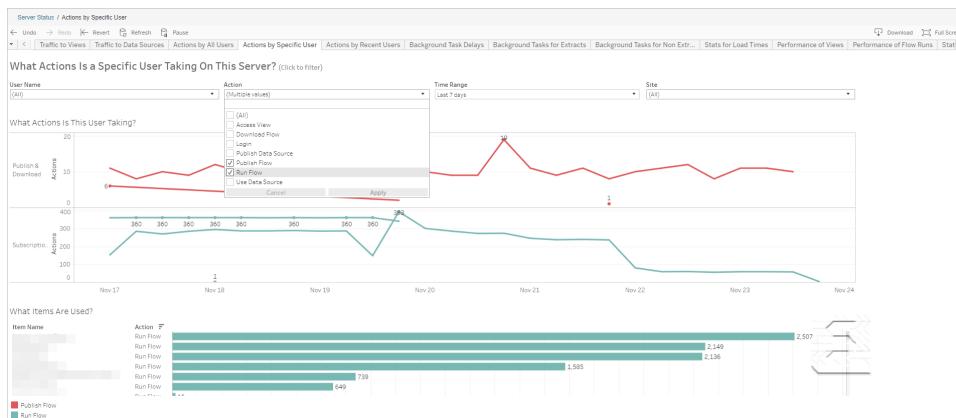


特定用户的操作

使用此视图深入了解单个用户使用流程的情况。

在 Tableau Server 中，可以按用户名、操作类型、时间范围和站点筛选视图。

在 Tableau Online 中，可以按用户名、操作类型和时间范围筛选视图。



最近用户的操作

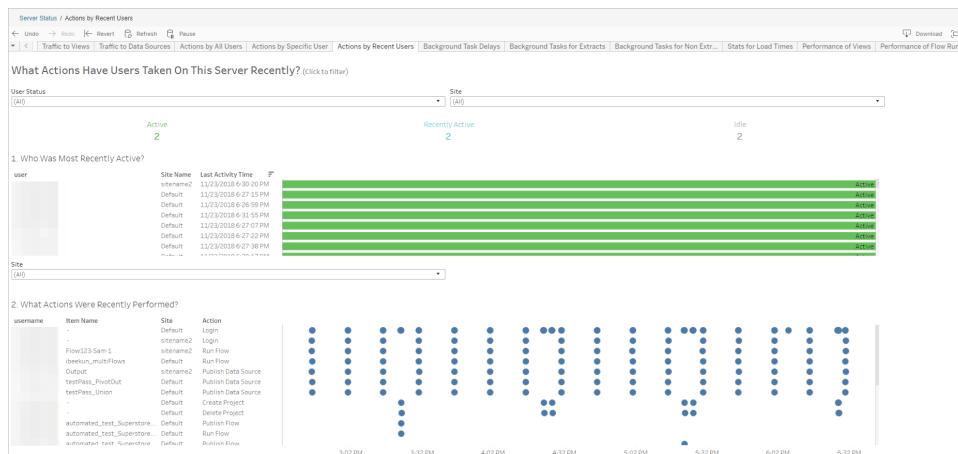
此视图显示过去 24 小时内哪些用户在 Tableau Server 或 Tableau Online 上处于活动状态。

如果您需要在服务器上执行一些维护活动，并希望了解此操作将影响多少用户以及哪些用户以及他们执行的操作，则这将非常有用。

视图显示当前登录到 Tableau Server 或 Tableau Online 的“活动”、“最近活动”和“空闲”用户。

对于此视图，活动用户是最近 5 分钟内执行了操作的用户，最近活动用户是最近 30 分钟内执行了操作的用户，空闲用户是在超过 30 分钟前执行了操作的用户。

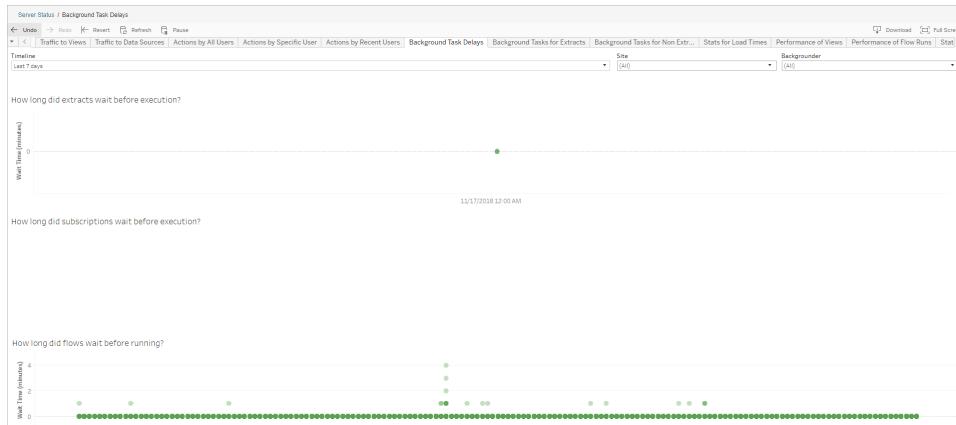
选择用户以仅查看该用户最近执行的操作。将鼠标悬停在操作上可查看操作的详细信息。



后台程序任务延迟

“后台任务延迟”视图显示数据提取刷新任务、订阅和流程任务的延迟，即计划运行时间和实际运行时间相差的时间。您可以使用视图来帮助确定可通过分发任务计划和优化任务来提高服务器性能的地方。

Tableau Prep 帮助

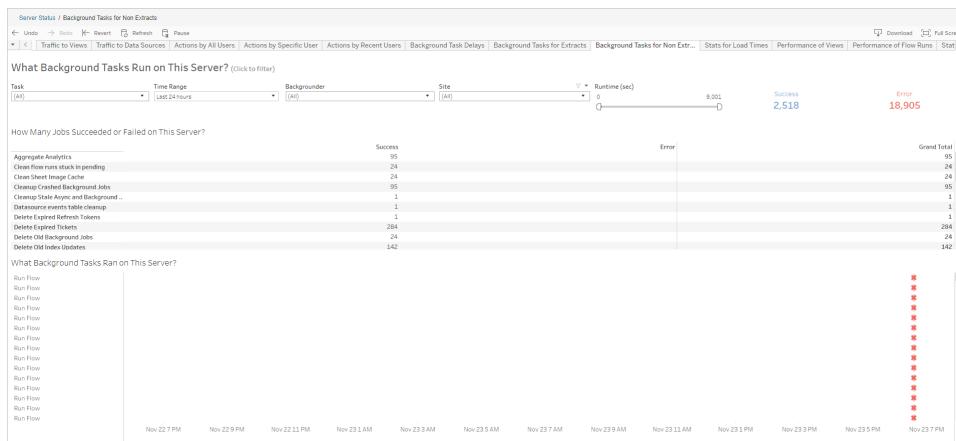


延迟的可能原因以及可能减少延迟的方法包括：

- 同时计划了多个任务。在示例视图中，显示很长延迟的任务都聚集在每天的同一时间，这会导致等待时间中出现峰值。请注意，您可以将时间线筛选器设置为一天，以按小时查看任务延迟，并确定一天中在同一时间计划了许多任务的小时数。此问题的解决方案可将任务分摊到非高峰时段，从而减少服务器负载。
- 在 **Tableau Server** 中，其他服务器进程同时运行并消耗服务器资源，从而导致性能降低。监视服务器进程的 **CPU** 和内存使用情况，查看哪些进程正在消耗大多数资源，然后调整服务器上进程的配置。有关监视进程的详细信息，请参见[使用 Windows 性能监视器收集数据](#)。

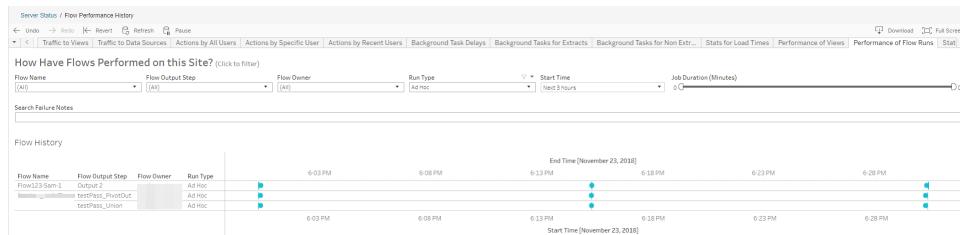
非数据提取的后台任务

创建后台任务的目的是为了运行流程(计划流程和临时流程)。您可以使用此视图查看此站点上成功或失败的流程任务数。若要了解任务的详细信息，请将鼠标指针悬停在其图标上。



流程运行的性能

使用此视图可查看站点上所有流程的性能历史记录。您可以按“流程名称”、“输出步骤名称”、“流程所有者”、“运行类型”(“计划”或“临时”)以及流程的启动时间进行筛选。



以下是您可以使用此视图回答的一些问题：

- 当前计划了哪些流程任务？** -为此，请使用“开始时间”筛选器并选择要查看的时间段。例如，若要查看接下来 3 小时内计划的流程任务，请选择“小时”->“下”，然后输入**3**。
- 流程任务持续多长时间？** - 若要回答此问题，请单击视图中的标记，您应会看到包括任务持续时间在内的详细信息。

有多少流程是临时运行流程，多少流程是计划运行流程？ - 若要回答此问题，请使用“运行类型”筛选器，并选择“临时”或“计划”。

注意: 此版本中该功能不起作用，因此实际上不会筛选数据。

除了上述问题外，下面是您可以收集的一些见解示例：

- 运行最频繁的流程将具有最多标记。
- 若要查看当前有多少流程同时运行，请将鼠标指针悬停在显示“正在运行”或“待定”的标记上，并选择“仅保留”，以筛选当前正在运行的所有流程运行。
- 若要查看在特定时间范围内有多少流程在同时运行，请为“开始时间”筛选器选择一个范围。例如，可以选择“下三个小时”以查看在接下来的三个小时内哪些流程将运行。

空间使用情况统计信息

“空间使用情况统计信息”视图可帮助您确定哪些流程输出正在服务器上占用最大磁盘空间。磁盘空间使用情况按用户、项目和流程输出大小显示，并向下舍入到最接近的数字。

Tableau Prep 帮助



使用“最小大小”筛选器可根据所占用的空间量控制显示的流程输出。为流程使用对象类型筛选器。

- **占用最大空间的用户** – 此部分显示了占用最大空间的流程所有者(针对流程筛选时)。单击用户名可筛选该用户的下两个图形。
- **哪些项目占用最大空间**–此部分显示包含流程的项目，这些项目占用的空间最大(针对流程筛选时)。
- **哪些工作簿、数据源和流程占用最大空间**–此部分显示占用最大空间的流程(针对流程筛选时)。

谁可以执行此操作

- **Tableau Server 管理员:**
 - 在服务器级别设置电子邮件通知
 - 为站点设置电子邮件通知
 - 查看错误
 - 恢复挂起的任务
 - 查看通知
 - 查看进程状态
- **Tableau 站点管理员:**
 - 在站点级别设置电子邮件通知
 - 查看错误
 - 恢复挂起的任务
 - 查看通知
- **流程所有者、项目主管和被授予流程查看权限的任何用户:**
 - 查看错误
 - 恢复挂起的任务
 - 查看通知(流程所有者)

开发人员资源 - REST API

使用 Tableau Server REST API 自动执行操作并将 Tableau 无缝地集成到现有工作流中。Tableau Server REST API 使您能够编程访问您的内容、用户、站点，现在还可以访问流程。通过 HTTP 在 Tableau Server 或 Tableau Online 上管理资源调配、权限和发布。REST API 使您能够访问数据源、项目、工作簿、站点用户、站点以及流程背后的功能。您可以使用此访问权限来创建自定义应用程序，或通过脚本与服务器资源交互。

Tableau REST API 支持以下流程功能：

增加了新的流程端点来支持发布流程、计划流程、按需运行流程、管理权限、下载流程等等。有关流程所有新的 REST API 端点的完整列表，请参见[流程方法](#)。

此外，[现有端点](#)已更新，可支持诸如为流程创建新计划、创建新站点、更新现有站点和管理默认权限等流程功能。

日常生活场景

调整数据的含义是什么？它会对可构建的可视化项以及可执行的分析产生怎样的影响？在下面的教程中，我们将探讨分析和可视化项的场景，确定对我们造成阻碍的数据限制，然后看看 Tableau Prep 如何能帮助我们调整数据来达到预期结果。

下载数据集，并使用 Tableau Prep 和 Tableau Desktop 执行这些日常生活场景。了解如何应用 Tableau Prep 中的特点和功能来准备好数据以在 Tableau Desktop 中进行分析。

为我们提供反馈。我们刚刚开始构建联机帮助的这一部分。如果有您想要在此处看到的特定场景，请不吝告知我们。使用页面顶部的反馈栏来告诉我们更多。

为了完成这些教程中的任务，您需要安装 Tableau Prep 和 Tableau Desktop，并将需要下载并将数据保存到您的计算机。

有关如何安装 Tableau Prep 和 Tableau Desktop 的信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[通过用户界面安装 Tableau Desktop 和 Tableau Prep Builder](#)。否则，您可以下载 Tableau Prep 和 Tableau Desktop 免费试用版。

使用 Tableau Prep 分析医院床位使用情况

医院满员会造成问题，但资源过多也同样会造成问题。务必要从床位即资源的角度来理解医院床位。但是，数据通常是从患者的角度进行存储的。我们如何才能获取相关数据，了解患者何时占用床位并确定床位使用情况？

注意：为了完成这些教程中的任务，您需要安装 Tableau Prep 并根据需要安装 Tableau Desktop：

若要安装 Tableau Prep 和 Tableau Desktop，请参见[Tableau Desktop 和 Tableau Prep 部署指南](#)。否则，您可以下载 Tableau Prep 和 Tableau Desktop 免费试用版。

您还需要下载三个数据文件。建议将这些文件保存在 My Tableau Prep Repository > Datasources 文件夹中。

- [Beds.xlsx](#)
- [Hours.xlsx](#)
- [Patient Beds.xlsx](#)

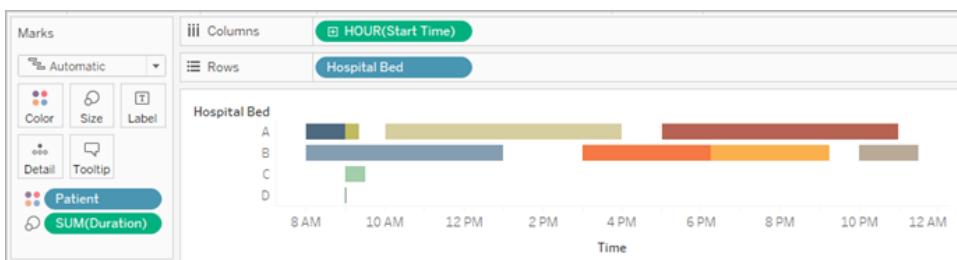
数据

对于我们的四个床位 A、B、C 和 D，我们将跟踪哪个患者占用了床位，以及他们占用床位的开始和结束时间。数据看起如下所示：

	A	B	C	D
1	Hospital Bed	Patient	Start Time	End Time
2	A	Person 1	1/1/2018 8:34	1/1/2018 9:34
3	A	Person 5	1/1/2018 9:55	1/1/2018 10:15
4	A	Person 9	1/1/2018 10:34	1/1/2018 16:34
5	A	Person 8	1/1/2018 17:00	1/1/2018 23:00
6	B	Person 2	1/1/2018 8:45	1/1/2018 13:45
7	B	Person 6	1/1/2018 15:13	1/1/2018 18:27
8	B	Person 7	1/1/2018 18:41	1/1/2018 21:56
9	B	Person 10	1/1/2018 22:13	1/1/2018 23:43
10	C	Person 3	1/1/2018 9:05	1/1/2018 9:35
11	D	Person 4	1/1/2018 9:30	

初步分析

如果我们将此数据引入 Tableau Desktop，我们可以创建一个甘特图来显示患者何时占用床位。



这是一种有用的视觉元素。我们可以看到，床位 A 和 B 的使用间隔时间很短，但床位 C 使用得非常少。床位 D 的患者没有结束时间，但我们将利用一些计算来解决该情况。这为我们提供了一个关于床位使用情况的视觉概览。

但是，如果我们想要计算床位空置的小时数，该怎么办？或者，如果我们要比较实施新政策前后的开放床位时间，该怎么办？通过采用当前结构方式的数据无法轻松地达到目的。

所需的数据结构

通过创建一些非常基本的数据集并在 Tableau Prep 中将其合并，我们可以将此数据集修改为允许我们执行更深入分析并创建更有用的可视化项的形式。

在进入 Tableau Prep 之前，让我们回顾一下，想一想我们需要创建什么才能回答以下问题：“每个床位空置了多少小时？”

我们需要能够查看每小时的每个床位，并了解是否有患者占用床位。目前，数据仅仅是患者何时占用床位；我们还未向 Tableau 提供有关空置小时数的信息。

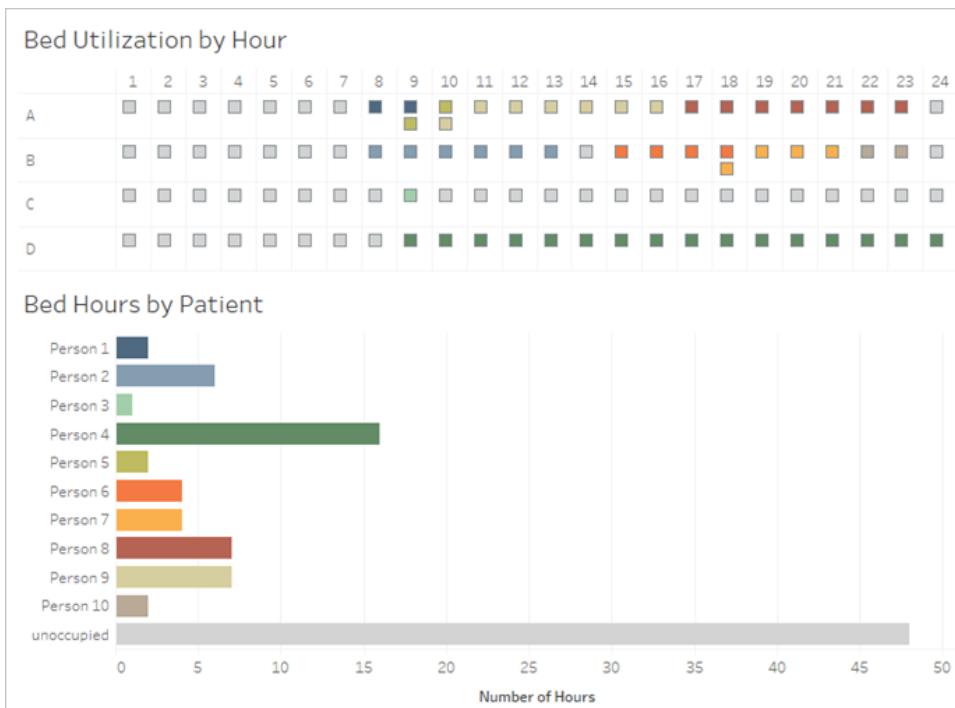
为了创建包含所有床位和所有小时的完整矩阵，我们将创建两个新数据集。一个数据集是床位的列表(A、B、C、D)，另一个数据集是小时的列表(1、2、3、...、23、24)。通过执行交叉联接(将一个数据集中的每一行与另一个数据集中的每一行联接)，我们将最终获得床位和小时的每种可能组合。

Beds.xlsx 数据集看起来如下所示：	Hours.xlsx 数据集看起来如下所示：	而交叉联接的结果看起来如下所示：																																																										
<table border="1"> <thead> <tr> <th></th> <th>A</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Bed</td> </tr> <tr> <td>2</td> <td>A</td> </tr> <tr> <td>3</td> <td>B</td> </tr> <tr> <td>4</td> <td>C</td> </tr> <tr> <td>5</td> <td>D</td> </tr> </tbody> </table>		A	1	Bed	2	A	3	B	4	C	5	D	<table border="1"> <thead> <tr> <th></th> <th>A</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Hour</td> </tr> <tr> <td>2</td> <td>1</td> </tr> <tr> <td>3</td> <td>2</td> </tr> <tr> <td>4</td> <td>3</td> </tr> <tr> <td>5</td> <td>4</td> </tr> <tr> <td>6</td> <td>5</td> </tr> <tr> <td>7</td> <td>6</td> </tr> </tbody> </table>		A	1	Hour	2	1	3	2	4	3	5	4	6	5	7	6	<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> </tr> <tr> <th>1</th> <th>Bed</th> <th>Hour</th> </tr> </thead> <tbody> <tr> <td>2</td> <td>A</td> <td>1</td> </tr> <tr> <td>3</td> <td>B</td> <td>1</td> </tr> <tr> <td>4</td> <td>C</td> <td>1</td> </tr> <tr> <td>5</td> <td>D</td> <td>1</td> </tr> <tr> <td>6</td> <td>A</td> <td>2</td> </tr> <tr> <td>7</td> <td>B</td> <td>2</td> </tr> <tr> <td>8</td> <td>C</td> <td>2</td> </tr> <tr> <td>9</td> <td>D</td> <td>2</td> </tr> </tbody> </table>		A	B	1	Bed	Hour	2	A	1	3	B	1	4	C	1	5	D	1	6	A	2	7	B	2	8	C	2	9	D	2
	A																																																											
1	Bed																																																											
2	A																																																											
3	B																																																											
4	C																																																											
5	D																																																											
	A																																																											
1	Hour																																																											
2	1																																																											
3	2																																																											
4	3																																																											
5	4																																																											
6	5																																																											
7	6																																																											
	A	B																																																										
1	Bed	Hour																																																										
2	A	1																																																										
3	B	1																																																										
4	C	1																																																										
5	D	1																																																										
6	A	2																																																										
7	B	2																																																										
8	C	2																																																										
9	D	2																																																										

接着，我们将引入“**Patient Beds**”(患者床位)信息，以是否有特定患者的形式标记每个床位-小时组合。我们最终会得到一个数据集，该数据集中每个床位-小时组合占一行，并且，如果患者占用了床位，则包含患者的编号及开始和结束时间。Null 值指明床位未被占用。

	A	B	C	D	E
1	Bed	Hour	Patient	Start Time	End Time
29	D	7			
30	A	8	Person 1	1/1/2018 8:34	1/1/2018 9:34
31	B	8	Person 2	1/1/2018 8:45	1/1/2018 13:45
32	C	8			
33	D	8			
34	A	9	Person 5	1/1/2018 9:55	1/1/2018 10:15
35	A	9	Person 1	1/1/2018 8:34	1/1/2018 9:34
36	B	9	Person 2	1/1/2018 8:45	1/1/2018 13:45
37	C	9	Person 3	1/1/2018 9:05	1/1/2018 9:35
38	D	9	Person 4	1/1/2018 9:30	
39	A	10	Person 9	1/1/2018 10:34	1/1/2018 16:34
40	A	10	Person 5	1/1/2018 9:55	1/1/2018 10:15
41	B	10	Person 2	1/1/2018 8:45	1/1/2018 13:45
42	C	10			
43	D	10	Person 4	1/1/2018 9:30	
44	A	11	Person 9	1/1/2018 10:34	1/1/2018 16:34

利用此结构中的数据，我们可以执行像这样的分析，让我们能够像调查患者床位一样轻松地调查未占用的床位。



重构数据

那么如何使用 Tableau Prep 达到该目的？我们将分两个部分构建流程，首先构建“**Bed Hours**”(床位小时) 矩阵，然后将其与“**Patient Beds**”(患者床位) 数据合并。确保下载全部三个 Excel 文件 (**Beds.xlsx**、**Hours.xlsx** 和 **Patient Beds.xlsx**) 以便继续操作。

床位小时矩阵

首先，我们将连接到 **Beds.xlsx** 文件。

1. 打开 Tableau Prep。
2. 从开始屏幕中，单击“连接到数据”。
3. 在“连接”窗格上，单击“**Microsoft Excel**”。导航到 **Beds.xlsx** 的保存位置，并单击“打开”。
4. “**Beds**”(床位) 工作表应会自动显示在“流程”窗格中。

提示：有关连接到数据的详细信息，请参见[连接到数据 在本页 141](#)。

接着，我们需要创建一个可用于与“**Hours**”(小时) 数据集执行交叉联接的字段。我们将添加一个仅仅为值 **1** 的计算。

5. 在“流程”窗格中，选择“**Beds**”(床位)，单击加号  图标，并选择“添加步骤”。
6. 随我们刚刚添加的“清理”步骤一起，“配置”窗格将出现。单击工具栏中的“创建计算字段”。
7. 将字段命名为“**Cross Join**”(交叉联接)，并输入值 **1**。
8. “数据”网格应会更新为显示数据的当前状态。

Cross Join	Bed
1	A
1	B
1	C
1	D

现在，我们将使用“**Hours**”(小时) 数据集重复该过程。

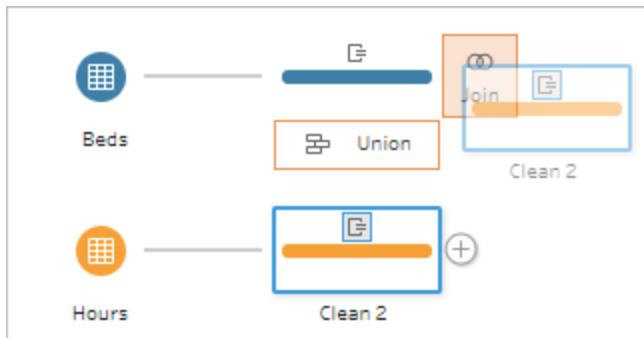
单击查看指引

9. 在“连接”窗格上，单击“添加连接” 按钮以添加另一个数据连接。
10. 选择“Microsoft Excel”，然后选择 **Hours.xlsx** 文件并单击“打开”。
11. 在“流程”窗格中，选择“Hours”(小时)，单击加号 图标，并选择“添加步骤”。
12. 通过“配置”窗格中的工具栏创建一个名为“Cross Join”(交叉联接) 的计算字段，并输入值 1。

Cross Join	Hour
1	1
1	2
1	3
1	4
1	5

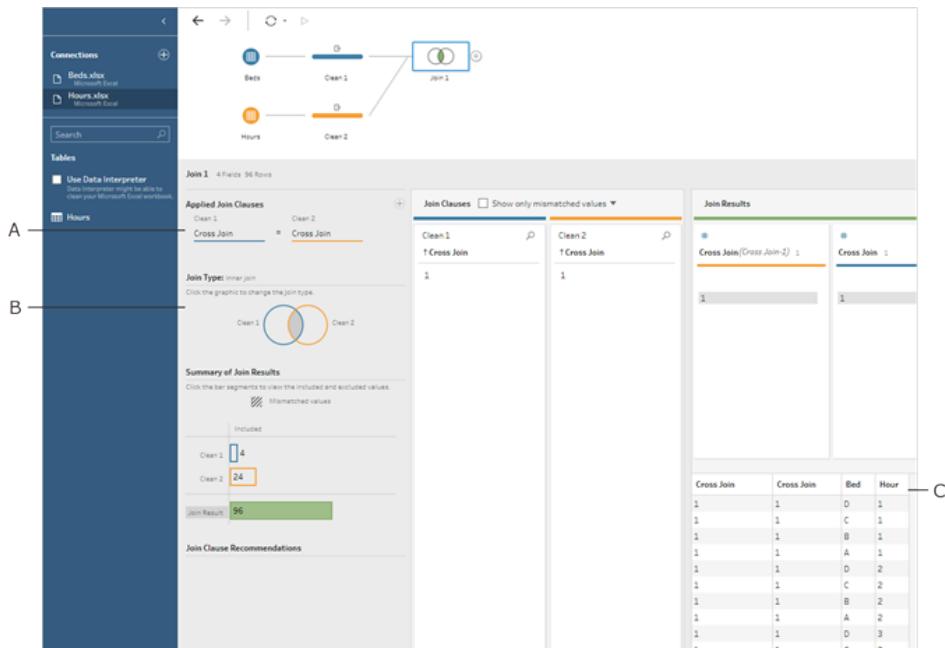
两个数据集现在有一个共同字段“Cross Join”(交叉联接)，因此可以联接。

13. 通过将“Clean 2”(清理 2) 拖到“Clean 1”(清理 1) 并放在“联接”选项上来联接两个清理步骤。



14. 在下面的“联接配置”中，联接配置应会自动填充。
 - 由于我们将两个字段都命名为“Cross Join”(交叉联接)，因此 Tableau Prep 自动将它们识别为共同字段，并创建适当的“已应用联接子句”。
 - 默认“联接类型”为内部联接，正是我们所需要的。

- 此联接会将“**Beds**”(床位) 中的所有行与“**Hours**”(小时) 中的所有行匹配，如“数据”网格中所示。



A. 联接子句, B. 联接类型, C. “数据”网格结果

提示: 有关联接的详细信息，请参见[联接数据 在本页 298](#)。

我们不再需要“**Cross Join**”(交叉联接) 字段，因此可以将这些字段移除。

- 在“流程”窗格中，选择“**Join 1**”(联接 1)，单击加号 图标，并选择“添加步骤”。
- 选择字段“**Cross Join-1**”(交叉联接-1) 和“**Cross Join**”(交叉联接)，并单击“**移除字段**”。
- 双击“**Clean 3**”(清理 3) 标签，并将该步骤重命名为“**Bed Hour Matrix**”(床位小时矩阵)。

我们现在有了“**Bed Hour Matrix**”(床位小时矩阵) 数据集，其中包含所有床位和所有小时，并且我们已完成构建数据集的第一部分。

患者床位使用

第二部分是引入患者床位使用情况。首先，我们将连接到数据。

1. 在“连接”窗格上，单击“添加连接” 按钮以添加另一个数据连接。
2. 选择“Microsoft Excel”，然后选择 **Patient Beds.xlsx** 文件，并单击“打开”。
3. 在“流程”窗格中，选择“**Patient Beds**”(患者床位)，单击加号 图标，并选择“添加步骤”。

由于“床位小时矩阵”文件基于小时，但“患者床位”基于实际时间，因此我们需要从“患者床位”开始和结束时间中提取小时。此外，对于结束时间，我们想要确保，如果患者在一天结束时(午夜，24时)仍然占用床位，则我们指明床位被占用，即使数据集中没有结束时间也是如此。我们将在此新步骤中添加计算字段。

4. 在工具栏中，单击“创建计算字段”。
 5. 将字段命名为“**Start Hour**”(开始小时)。对于计算，输入 `DATEPART('hour', [Start Time])`。
- 这将获取开始时间的小时并将其提取出来。因此，“1/1/18 9:35 AM”将变为“9”。
6. 创建另一个名为“**End Hour**”(结束小时)的计算字段。对于计算，输入 `IFNULL(DATEPART('hour', [End Time]), 24)`。

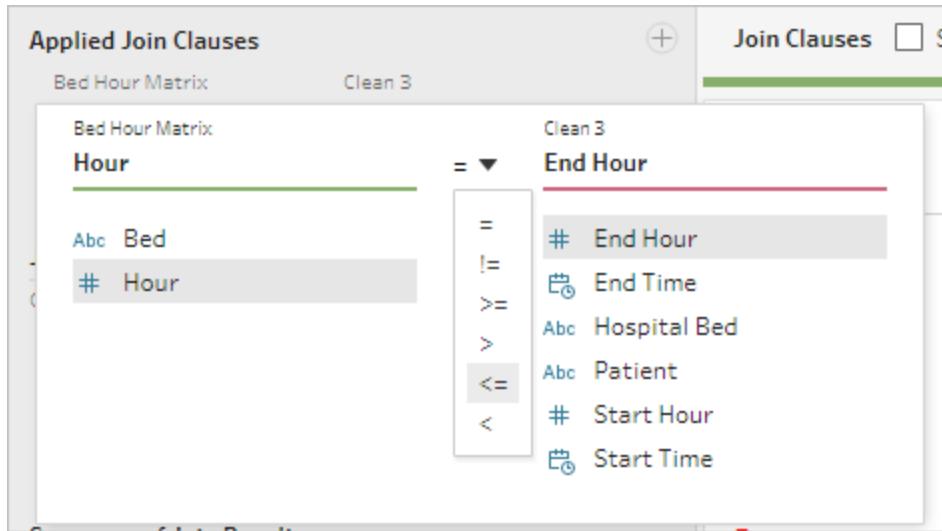
`DATEPART` 部分获取结束时间的小时。`IFNULL` 部分将为任何缺少的结束时间分配结束时间 24 时(午夜)。

现在我们已准备好将患者床位使用情况联接到“**Bed Hour Matrix**”(床位小时矩阵)。与我们之前进行的操作相比，此操作更为复杂一些。内部联接将只会返回两个数据集中都存在的值。由于我们想要确保保留所有床位-小时槽位，而不管患者是否占用床位，因此我们需要执行左联接。此将生成许多 `null` 值，但这是适当的。

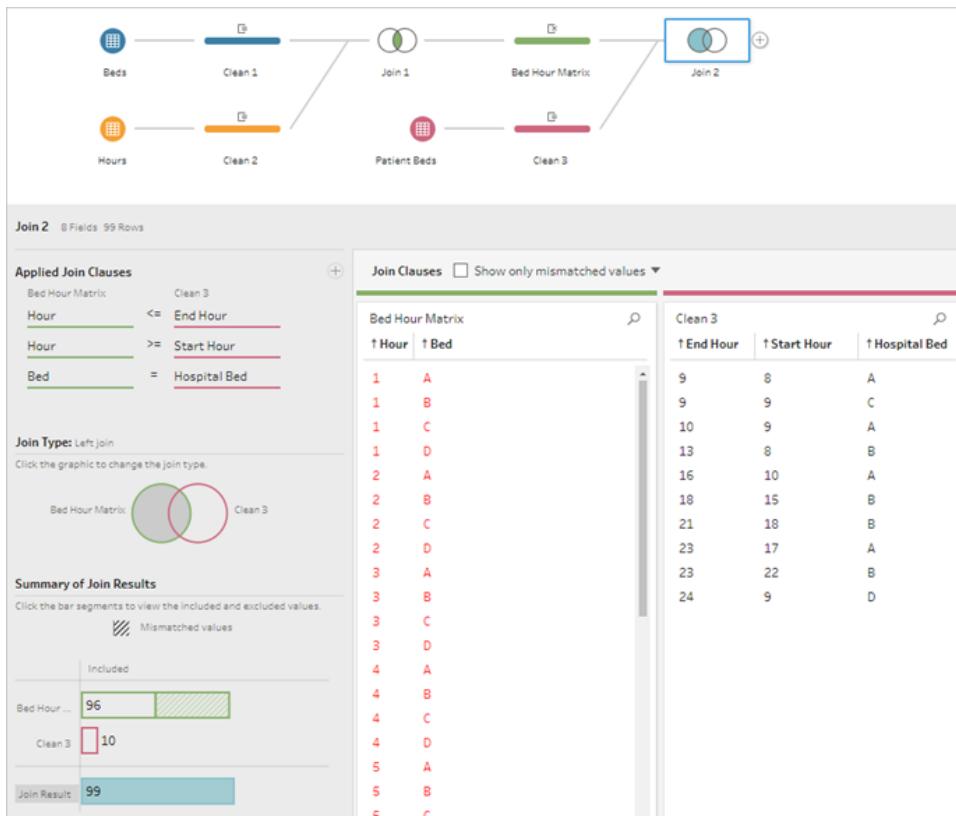
我们还需要匹配床位-小时槽位被一名(或多名)患者占用的时间。除了匹配患者占用的床位外，我们还需要考虑时间。“**Bed Hour Matrix**”(床位小时矩阵)数据集只有“**Hour**”(小时)字段，而“**Patient Beds**”(患者床位)数据集有“**Start Hour**”(开始小时)和“**End Hour**”(结束小时)。我们使用一些基本逻辑来确定是否应将患者分配到给定床位-小时槽位：如果其开始小时小于或等于(\leq)床位-小时槽位，并且其结束小时大于或等于(\geq)床位-小时槽位，则患者被视为占用床位。

因此，需要三个联接子句来将这两个数据集相应地匹配在一起。

9. 将“**Clean 3**”(清理 3) 步骤与“**Bed Hour Matrix**”(床位小时矩阵) 步骤联接。
10. 在“已应用联接子句”区域中，默认值应为“**Hour = End Hour**”。单击联接子句，将运算符从“=”改为“ \leq ”。



11. 单击“已应用联接子句”区域右上角的加号 按钮，以添加另一个联接子句。将其设置为**“Hour >= Start Hour”**
12. 为**“Bed = Hospital Bed”**添加第三个联接子句。
13. 在“联接类型”部分，单击**“Bed Hour Matrix”**(床位小时矩阵)旁边图形的无阴影区域，将联接类型更改为“左”联接。



注意: 如果将“**Bed Hour Matrix**”(床位小时矩阵) 拖到“**Clean 3**”(清理 3), 而不是相反, 则通过使用右联接(而不是左联接)可以获得期望的结果。对于联接的方向, 拖动步骤的顺序很重要。联接子句也将按相反顺序 — 一定要保持比较小时的正确逻辑。

数据现在已联接, 但我们应从联接中清理一些构件, 并确保字段保持整齐。我们不再需要“**Start Hour**”(开始小时)和“**End Hour**”(结束小时)。“**Hospital Bed**”(医院床位)和“**Bed**”(床位)也是多余的。最后, “**Patient**”(患者)字段中的值 null 实际上表示床位未被占用。

14. 在“流程”窗格中, 添加一个清理步骤以便能对联接的数据进行整理。
15. 按住 Ctrl 单击(在 Mac 上按住 Command 单击)以多选方式选择字段“**End Hour**”(结束小时)、“**Start Hour**”(开始小时)和“**Hospital Bed**”(医院床位), 然后在工具栏中单击“**移除字段**”。
16. 在“**Patient**”(患者)字段配置卡上, 双击“null”值并键入“**Unoccupied**”(未占用)。

我们现在有了每个床位-小时占一行的数据结构;如果在该小时期间有患者占用床位, 则我们也有患者信息。余下的工作就是添加输出步骤和生成数据集本身。

17. 在“流程”窗格中，选择“**Clean 4**”(清理 4)，单击加号 \oplus 图标，并选择“添加输出”。
18. 在“输出”窗格中，将“输出类型”更改为.csv，然后单击“浏览”。
19. 为名称输入“**Bed Hour Patient Matrix**”(床位小时患者矩阵)，并在单击“接受”保存之前选择所需的位置。
20. 单击窗格底部的“运行流程” \triangleright 按钮生成输出。在状态对话框中单击“完成”关闭对话框。

提示：有关输出和运行流程的详细信息，请参见[保存和共享工作](#)在本页309。

最终流程应如下所示：

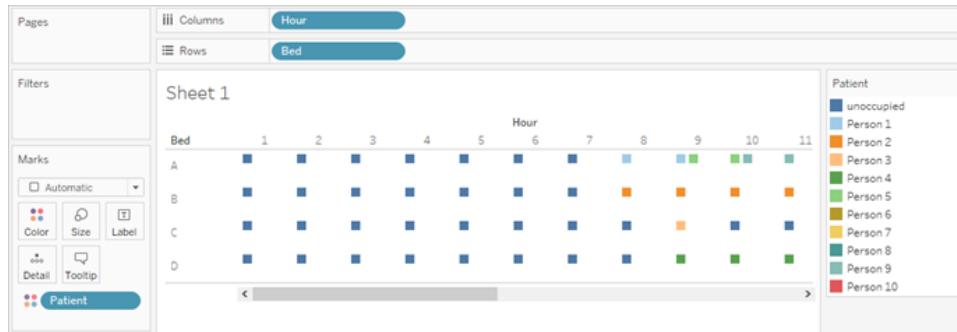


Tableau Desktop 中的分析

若要在继续执行本教程之前安装 Tableau Desktop，您可以下载[免费试用版](#)。

现在我们有了采用所需结构的数据集，我们可以执行比用原始数据更深入的分析。

1. 打开 Tableau Desktop。在“连接”窗格中，选择“文本文件”，导航到 **Bed Hour Patient Matrix.csv** 文件，并单击“打开”。
2. 在“数据源”选项卡上，默认情况下数据应出现在画布中。单击“**Sheet 1**”(工作表 1)。
3. 在“数据”窗格中，将“**Hour**”(小时)从“度量”拖到“维度”，将其设为离散维度。
4. 将“**Bed**”(床位)拖到“行”功能区，并将“**Hour**”(小时)拖到“列”功能区。
5. 将“**Patient**”(患者)拖到“颜色”功能区。

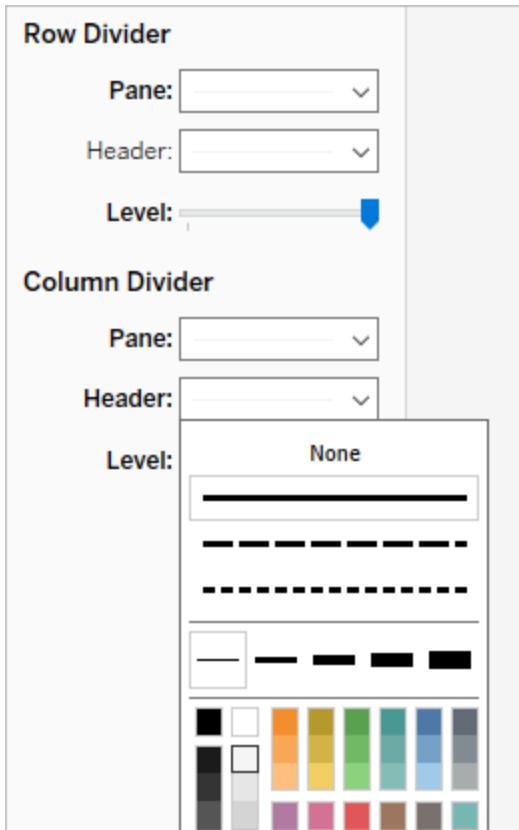


格式设置为可选，但可以帮助使图示更易于理解。

6. 单击“颜色”功能区，并选择“编辑颜色”。
7. 在左侧的区域中，选择“Unoccupied”(未占用)。从右侧的下拉列表中，选择“西雅图灰”调色板。
8. 选择第四个最浅的灰色，并单击“确定”。
9. 再次单击“颜色”功能区，然后单击“边框”下拉列表。选择最右侧的第二个灰色选项。
10. 在工具栏中，通过“大小”下拉列表将“标准”更改为“适合宽度”。

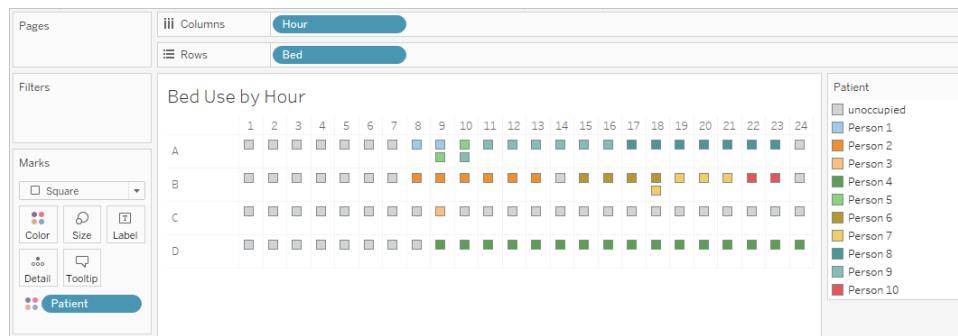


11. 单击“设置格式”菜单，然后单击“边框”。
12. 对于“行分隔符”，单击“区”下拉列表并选择一种非常浅的灰色。
13. 将“级别”滑块调整到第二个刻度标记。
14. 为“列分隔符”重复这些步骤。将“区”颜色设置为浅灰色，并将“级别”调整到第二个刻度标记。



15. 双击底部的工作表标签，并将其重命名为“**Bed Use by Hour**”(按小时列出的床位使用情况)。

此视图让我们能快速查看给定床位何时处于占用或空置状态。

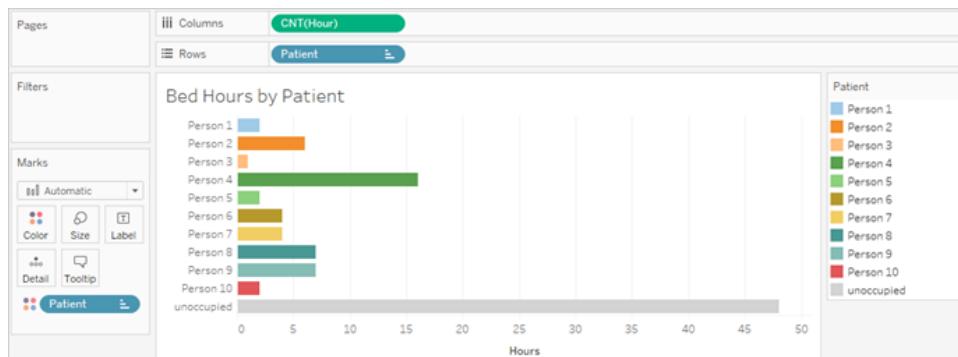


但我们可以更进一步，并计算每个床位未被占用的小时数。

16. 单击底部的新工作表标签 打开一个空白工作表。

17. 将“**Patient**”(患者) 拖到“行”。

18. 将“Hour”(小时) 拖到“列”。右键单击打开菜单。选择“度量”>“计数”。
19. 将“Patient”(患者) 字段的另一个副本从“数据”窗格拖到“颜色”功能区。
20. 右键单击轴，并选择“编辑轴”。将标题更改为“Hours”(时数) 并关闭对话框。
21. 将工作表标签重命名为“Bed Hours by Patient.”(患者占用床位时数)。



此视图使我们能够确定我们有多少未占用床位时数，而这是使用原始数据集做不到的。您可以创建哪些其他图标或仪表板？既然您的数据结构正确，不妨试一下。

总结和资源

为了使用 Tableau Prep 构建此数据结构，我们需要执行以下操作：

1. 针对我们想要分析的各个方面(本例中为床位和小时)构建数据集。
2. 交叉联接这些数据集，创建一个包含床位和小时每种可能组合的“**Bed Hour Matrix**”(床位小时矩阵)数据集。
3. 将“**Bed Hour Matrix**”(床位小时矩阵)与“**Patient Bed**”(患者床位)数据联接，同时确保联接保留所有床位-槽位小时，并且联接子句恰当地将患者床位数据与床位-小时槽位匹配。

我们使用以下计算来创建可以联接的字段。第二个和第三个计算从原始日期时间字段中提取小时信息。

- **Cross Join**(交叉联接)= 1
 - 此计算只是为每一行分配值 1
- **Start Hour**(开始小时)= DATEPART ('hour', [Start Time])
 - 这将获取开始时间的小时并将其提取出来。因此，“1/1/18 9:35 AM”将变为“9”。

- **End Hour**(结束小时) = IFNULL(DATEPART('hour',[End Time]), 24)
 - 我们可以使用 DATEPART('hour',[End Time])，就像我们为“**Start Time**”(开始时间)所做的一样。这将获取结束时间的小时并将其提取出来。因此，“1/1/18 4:34 AM”将变为“4”。
 - 但我们想要指明仍被占用(没有结束时间)的患者床位在使用中，而不是空置。为此，我们将使用 IFNULL 函数为任何缺少的结束时间分配结束时间 24 时(午夜)。如果第一个参数 DATEPART('hour',[End Time]) 为 null，则计算将改为返回“24”。

注意: 想要检查您的工作？下载 Tableau Prep 打包流程文件 ([Hospital Beds.tflx](#)) 和 Tableau Desktop 打包工作簿文件 ([Hospital Beds.twbx](#))。

资源: 需要更多培训？请观看 Tableau Prep 的新[培训视频](#)，并参加[现场培训](#)课程。想知道我们涵盖的功能？请查看 Tableau Prep 联机帮助中的其他主题。在寻找其他资源？[Master Tableau Prep with this list of learning resources](#)(利用此学习资源清单掌握 Tableau Prep) 博文正适合于您。

使用 Tableau Prep 查找第二个日期

分析中的一项常见需求是确定第二个事件(例如客户进行第二次购买并从而成为常客，或者司机第二次交通违章时)的日期。查找第一个事件的日期很容易，它只是最小日期。查找第二个日期则较为复杂。

在这个由两部分组成的教程中，我们将调整交通违章数据并回答以下问题：

1. 每个司机第一次和第二次违章相隔多长时间(以天数为单位)？
2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？
3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？
4. 有多少司机有过多种类型的违章？
5. 从未上过交通法规学习班的司机的平均罚款金额是多少？

在第一个阶段中，我们将使用 Tableau Prep 针对分析重构数据。在第二个阶段 **Tableau Desktop 中包含第二个日期的分析** 在本页 445 中，我们将转而讲述 Tableau Desktop 中的分析。

本教程的目标是在一个真实场景的上下文中呈现各种概念并演练各个选项—而不是规定式地确定哪一个是最好的。最后，您应该会对数据结构对计算和分析的影响有更深的认识，并更加熟悉 Tableau Prep 的各项功能以及 Tableau Desktop 中的计算。

注意:为了完成本教程中的任务，您需要安装 Tableau Prep 并根据需要安装 Tableau Desktop，并下载相关数据。

若要在继续学习本教程之前安装 Tableau Prep 和 Tableau Desktop，请参见 [Tableau Desktop 和 Tableau Prep 部署指南](#)。否则，您可以下载 [Tableau Prep](#) 和 [Tableau Desktop](#) 免费试用版。

数据集为 [Traffic Violations.xlsx](#)。建议将其保存在 My Tableau Prep Repository > Datasources 文件夹中。

数据

对于此示例，我们将探讨交通违章数据。每条违章为一行。将记录司机、日期、违章类型、司机是否需要参加交通法规学习班以及罚款金额。

	A	B	C	D	E
1	Driver ID	Infraction Date	Infraction Type	Traffic School	Fine Amount
2	JO-151451402	1/8/2017	Speeding	Yes	115
3	CM-127151402	3/1/2017	Running a red light	No	55
4	AP-109151404	3/2/2017	Non-moving violation	No	95
5	SH-199751404	3/4/2017	Speeding	Yes	130
6	BT-114401404	3/20/2017	Non-moving violation	No	130
7	MO-175001406	5/30/2017	Speeding	Yes	118
8	RA-1988558	6/2/2017	Speeding	Yes	144
9	BT-1168027	6/5/2017	Speeding	Yes	128
10	MO-175001406	6/18/2017	Speeding	Yes	115
11	MP-174701406	6/19/2017	Speeding	No	125
12	AA-106451404	7/5/2017	Running a red light	No	60
13	RA-199151402	7/20/2017	Speeding	Yes	146
14	SC-202601404	8/31/2017	Running a red light	No	150
15	MO-175001406	9/7/2017	Non-moving violation	No	320
16	AS-100451404	9/15/2017	Running a red light	No	50

所需的数据结构

数据当前的构造方式是每条违章为一行。有多次违章的司机将出现在多行中，并且无法轻松分辨哪条违章是他们的第一次或第二次违章。

为了调查屡屡违章的司机，我们需要一个能够分出第一次和第二次违章日期的数据集、与其中每条违章关联的信息，并且需要每一行为一个司机。

A	B	C	D	E	F	G	H	I
Driver ID	1st Infraction Date	1st Infraction Type	1st Traffic School	1st Fine Amount	2nd Infraction Date	2nd Infraction Type	2nd Traffic School	2nd Fine Amount
1 BD-117701406	12/25/2017	Speeding	Yes	140	2/7/2018	Speeding	Yes	125
2 JO-151451402	1/8/2017	Speeding	Yes	115	11/21/2018	Reckless driving	Yes	550
4 SN-207101402	12/27/2017	Speeding	Yes	280	4/26/2018	Speeding	Yes	130
5 CJ-120101402	11/26/2017	Speeding	Yes	122	3/28/2018	Speeding	Yes	116
6 JR-156701404	12/24/2017	Speeding	No	148	7/28/2018	Speeding	Yes	310
7 AP-109151404	3/2/2017	Non-moving violation	No	95	9/24/2018	Speeding	No	105
8 PC-187451406	11/11/2017	Speeding	Yes	220	12/30/2018	Non-moving violation	No	600
9 TS-214301406	9/13/2018	Speeding	Yes	115	11/10/2018	Non-moving violation	No	95
10 NP-187001404	12/11/2018	Non-moving violation	No	80	12/20/2018	Speeding	No	120
11 DB-129701402	5/13/2018	Running a red light	No	110	11/11/2018	Speeding	Yes	80
12 AJ-107951404	10/15/2017	Speeding	Yes	130	12/31/2017	Running a red light	No	85
13 BT-114401404	3/20/2017	Non-moving violation	No	130	11/13/2018	Speeding	Yes	96
14 AF-108851404	5/9/2018	Non-moving violation	No	200	9/2/2018	Speeding	No	130
15 SC-202601404	8/31/2017	Running a red light	No	150	11/10/2018	Speeding	Yes	50
16 KL-166451406	10/4/2017	Speeding	No	115	11/13/2017	Speeding	Yes	104
17 MO-175001406	5/30/2017	Speeding	Yes	118	6/18/2017	Speeding	Yes	115
18 CM-127151402	3/1/2017	Running a red light	No	55	8/1/2018	Running a red light	No	160
19 KT-164801402	5/31/2018	Non-moving violation	No	190	11/10/2018	Speeding	No	74
20 JB-160001402	11/18/2018	Speeding	Yes	220	12/5/2018	Non-moving violation	No	195
21 LH-170201404	5/6/2018	Running a red light	No	110	9/17/2018	Speeding	Yes	230
22 BG-1103555	12/25/2017	Speeding	Yes	195	12/8/2018	Speeding	Yes	315
23 MP-174701406	6/19/2017	Speeding	No	125	10/12/2017	Running a red light	No	175
24 MV-179051406	10/23/2017	Reckless driving	Yes	200	9/8/2018	Speeding	No	124

重构数据

那么如何使用 Tableau Prep 达到该目的？我们将分阶段构建流程，首先抽取第一次违章日期，然后抽取第二次违章日期，接着根据需要调整最终数据。确保已下载了 Excel 文件 ([Traffic Violations.xlsx](#)) 以便继续操作。

第一次违章日期的初始聚合

首先，我们将连接到 **Traffic Violations.xlsx** 文件。

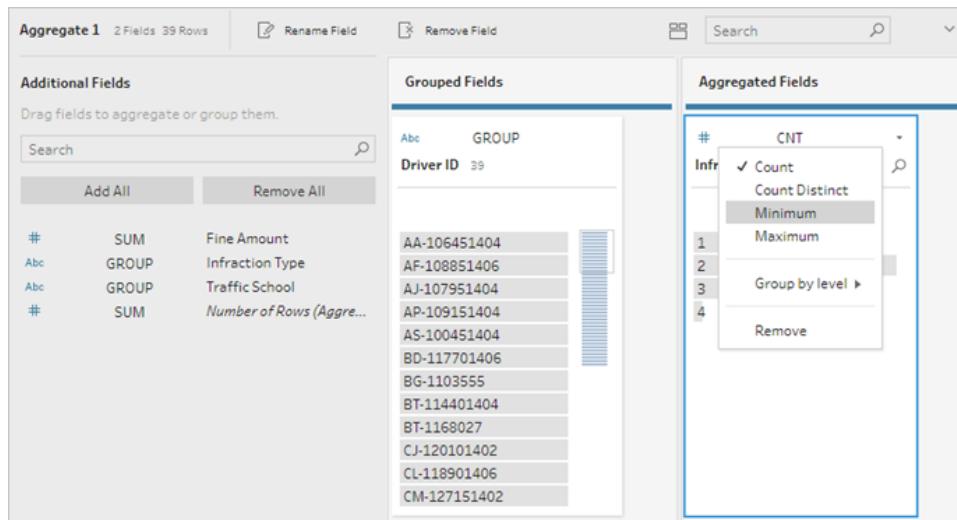
1. 打开 Tableau Prep。
2. 从开始屏幕中，单击“连接到数据”。
3. 在“连接”窗格中，单击“**Microsoft Excel**”。导航到 **Traffic Violations.xlsx** 的保存位置，并单击“打开”。
4. “**Infractions**”(违章) 工作表应会自动显示在“流程”窗格中。

提示：有关连接到数据的详细信息，请参见[连接到数据](#) 在本页 141。

接着，我们需要标识每个司机的第一次违章日期。我们将使用“聚合”步骤来执行此操作，同时创建一个包含“Driver ID”(司机 ID) 和“Minimum Infraction Date”(最早违章日期) 的微型数据集。

注意: 在 Tableau Prep 中使用“聚合”步骤时，任何应定义行组成内容的字段为“分组字段”。(对我们来说，该字段为“Driver ID”(司机 ID)。) 任何将聚合并显示在分组字段级别的字段为“聚合字段”。(对我们来说，该字段为“Infraction Date”(违章日期)。)

5. 在“流程”窗格中，选择“Infractions”(违章)，单击加号  图标，并选择“添加聚合”。
6. 将“Driver ID”(司机 ID) 拖到“分组字段”放置区域。
7. 将“Infraction Date”(违章日期) 拖到“聚合字段”区域。默认聚合为 CNT(计数)。单击“CNT”，并将聚合更改为“Minimum”。



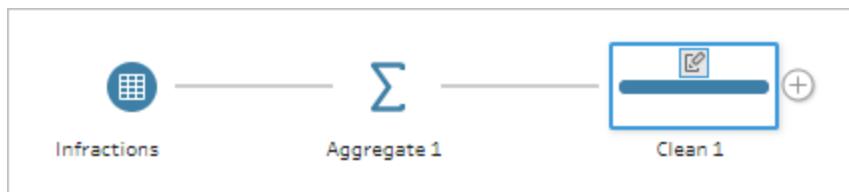
这将标识最小(最早)日期，即每个司机的第一次违章日期。

提示: 有关聚合的详细信息，请参见[清理和调整数据 在本页 223](#)。

8. 在“流程”窗格中，选择“聚合 1”，单击加号  图标，并选择“添加步骤”，以便我们能够清理聚合的输出。

9. 在“配置”窗格中，双击字段名**“Infraction Date”**(违章日期)，并将其更改为**“1st Infraction Date”**(第一次违章日期)。

在此阶段，流程和配置应如下所示：



The screenshot shows the 'Clean 1' configuration window. At the top, it displays 'Clean 1 2 Fields 39 Rows'. Below this are two main sections: 'Changes (1)' on the left and a detailed view on the right.

Changes (1):

- Driver ID 39
- AA-106451404
- AF-108851406
- AJ-107951404
- AP-109151404
- AS-100451404
- BD-117701406
- BG-1103555
- BT-114401404
- BT-1168027
- CJ-120101402
- CL-118901406
- CM-127151402

1st Infraction Date 37:

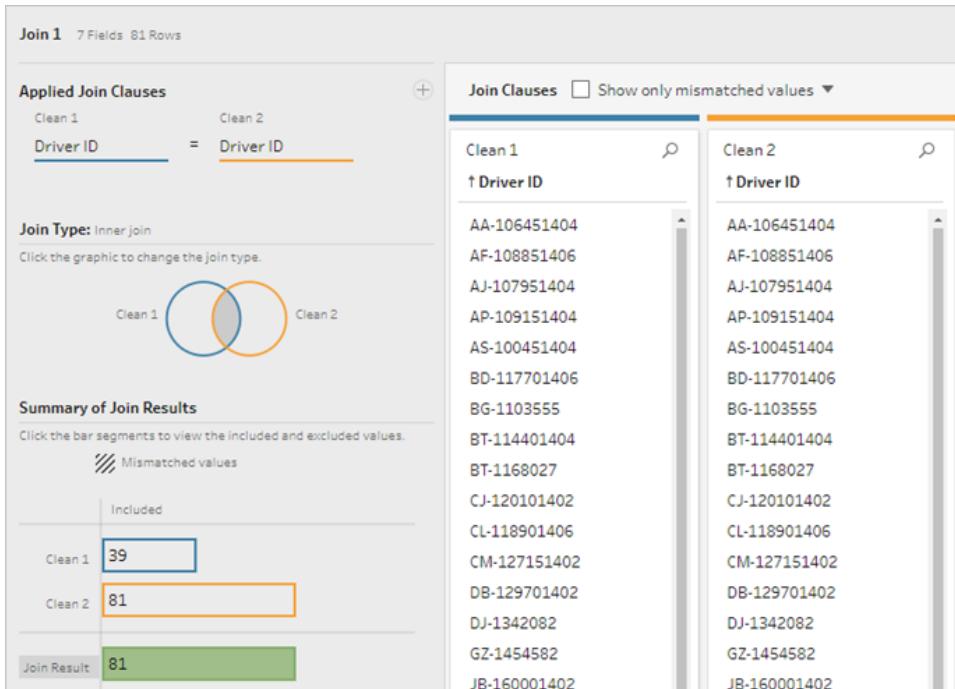
Original Value	New Value
01/01/2017	01/01/2017
01/01/2019	01/01/2019

从此“清理”步骤的“配置”窗格中，我们可以看到数据现在包含 39 行，并且只包含 2 个字段。未用于分组或聚合的任何字段已丢失。但我们想要能够保留一些原始信息。我们可将这些字段添加到分组或聚合中(但这样做会更改详细级别或者需要对字段进行聚合)，或者将此微型数据集联接回原始数据集(本质上是为**“1st Infraction Date”**(第一次违章日期)向原始数据中添加一个新列)。让我们执行联接。

- 在“流程”窗格中，选择“**Violations**”(违章)，单击加号  图标，并选择“添加分支”。

此分支具有所有原始数据。我们会将聚合的结果联接到完整数据的这一副本。通过依据“**Driver ID**”(司机 ID) 进行联接，我们会将聚合中的最小日期添加到原始数据中。

- 选择步骤“清理 2”，并将其拖到步骤“清理 1”上，放在“联接”上。
- 默认联接配置应该是正确的：基于“**Driver ID = Driver ID**”的内部联接。

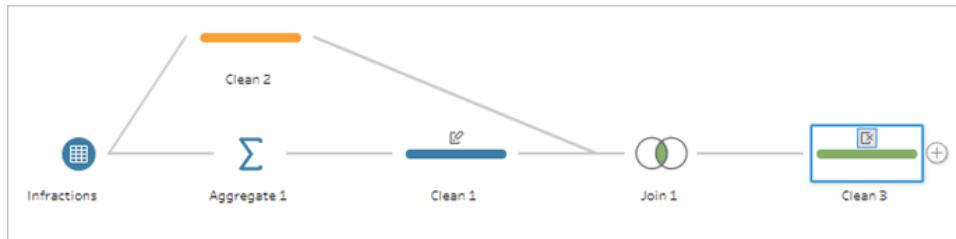


提示：有关联接的详细信息，请参见[联接数据](#) 在本页 298。

由于在联接过程中某些字段可能是重复的，例如联接子句中的字段，因此在执行联接之后清理无关的字段通常是个好主意。

- 在“流程”窗格中，选择“联接 1”，单击加号  图标，并选择“添加步骤”。
- 在“配置”窗格中，单击“**Driver ID-1**”(司机 ID-1) 卡，然后在工具栏中单击“**移除字段**”。
- 若要更改字段顺序，请将“**1st Infraction Date**”(第一次违章日期) 卡拖到“**Driver ID**”(司机 ID) 和“**Infraction Date**”(违章日期) 之间您看到黑色行出现的位置。

在此阶段，流程应如下所示：



查看下面的数据网格，我们可以看到新的合并数据集。我们有了添加到数据集中每一行的每个司机的最小(即第一次)违章日期。

Driver ID	1st Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	01/08/2017	01/08/2017	Speeding	Yes	115
CM-127151402	03/01/2017	03/01/2017	Running a red light	No	55
AP-109151404	03/02/2017	03/02/2017	Non-moving violation	No	95
SH-199751404	03/04/2017	03/04/2017	Speeding	Yes	130
BT-114401404	03/20/2017	03/20/2017	Non-moving violation	No	130
MO-175001406	05/30/2017	05/30/2017	Speeding	Yes	118
RA-1988558	06/02/2017	06/02/2017	Speeding	Yes	144
BT-1168027	06/05/2017	06/05/2017	Speeding	Yes	128
MO-175001406	06/18/2017	06/18/2017	Speeding	Yes	115
MP-174701406	06/19/2017	06/19/2017	Speeding	No	125
AA-106451404	07/05/2017	07/05/2017	Running a red light	No	60
RA-199151402	07/20/2017	07/20/2017	Speeding	Yes	146
SC-202601404	08/31/2017	08/31/2017	Running a red light	No	150
MO-175001406	09/07/2017	09/07/2017	Non-moving violation	No	320
AS-100451404	09/26/2017	09/26/2017	Running a red light	No	50
SH-199751404	09/27/2017	09/27/2017	Speeding	Yes	225
AA-106451404	09/28/2017	09/28/2017	Running a red light	No	195

第二次违章日期的第二个聚合

我们还需要确定第二次违章日期。为此，我们需要筛选出违章日期等于最小日期的任何行 — 从而移除第一个日期。我们随后可以使用另一个聚合步骤获取剩余日期的最小值，剩下的就是第一次违章日期，为了清楚起见，我们将重命名该日期。

注意：由于我们将需要使用当前位于流程后面“清理 3”中的数据，因此我们将添加另一个“清理”步骤来脱离获取第二次违章日期的过程。

16. 在“流程”窗格中，选择“**Clean 3**”(清理 3)，单击加号 图标，并选择“添加步骤”。
17. 在“配置”窗格上的工具栏中，选择“**筛选值**”。创建一个筛选器 [Infraction Date] != [1st Infraction Date]。
18. 移除字段“**1st Infraction Date**”(第一次违章日期)。

19. 在“流程”窗格中，选择“**Clean 4**”(清理 4)，单击加号  图标，并选择“添加聚合”。
20. 将“**Driver ID**”(司机 ID)拖到“分组字段”放置区域。将“**Infraction Date**”(违章日期)拖到“聚合字段”区域，并将聚合更改为“**Minimum**”。
21. 在“流程”窗格中，选择“**Aggregate 2**”(聚合 2)，单击加号  图标，并选择“添加步骤”。将“**Infraction Date**”(违章日期)重命名为“**2nd Infraction Date**”(第二次违章日期)。

在此阶段，流程应如下所示：



我们现在为每个司机标识了第二次违章日期。为了获取与每次违章关联的所有其他信息(类型、罚款、交通法规学习班)，我们需要再次将此项联接回整个数据集。

22. 选择“清理 5”，并将其拖到步骤“清理 3”，放在“联接”上。
23. 同样，默认联接配置应该是正确的：基于“**Driver ID = Driver ID**”的内部联接。
24. 在“流程”窗格中，选择“**Join 2**”(联接 2)，单击加号  图标，并选择“添加步骤”。删除字段“**Driver ID-1**”(司机 ID-1)和“**1st Infraction Date**”(第一次违章日期)，因为不再需要这些字段。

在此阶段，流程应如下所示：



为第一次和第二次违章创建完整数据集

在执行任何进一步操作之前，让我们回顾一下，想一想我们已有的所有内容以及要如何将它们结合在一起。我们期望的最终状态是一个像这样的数据集：包含一个“**Driver ID**”(司机 ID)列，然后包含日期、类型、交通法规学习班列，以及第一次和第二次违章的罚款金额。

Tableau Prep 帮助

A	B	C	D	E	F	G	H	I
Driver ID	1st Infraction Date	1st Infraction Type	1st Traffic School	1st Fine Amount	2nd Infraction Date	2nd Infraction Type	2nd Traffic School	2nd Fine Amount
BD-117701406	12/25/2017	Speeding	Yes	140	2/7/2018	Speeding	Yes	125
JO-151451402	1/8/2017	Speeding	Yes	115	11/21/2018	Reckless driving	Yes	550
SN-207101402	12/27/2017	Speeding	Yes	280	4/26/2018	Speeding	Yes	130
CJ-120101402	11/26/2017	Speeding	Yes	122	3/28/2018	Speeding	Yes	116
JR-156701404	12/24/2017	Speeding	No	148	7/28/2018	Speeding	Yes	310
AP-109151404	3/2/2017	Non-moving violation	No	95	9/24/2018	Speeding	No	105
PC-187451406	11/11/2017	Speeding	Yes	220	12/30/2018	Non-moving violation	No	600
TS-214301406	9/13/2018	Speeding	Yes	115	11/10/2018	Non-moving violation	No	95
NP-187001404	12/11/2018	Non-moving violation	No	80	12/20/2018	Speeding	No	120
DB-129701402	5/13/2018	Running a red light	No	110	11/11/2018	Speeding	Yes	80
AJ-107951404	10/15/2017	Speeding	Yes	130	12/31/2017	Running a red light	No	85
BT-114401404	3/20/2017	Non-moving violation	No	130	11/13/2018	Speeding	Yes	96
AF-108851404	5/9/2018	Non-moving violation	No	200	9/2/2018	Speeding	No	130
SC-202601404	8/31/2017	Running a red light	No	150	11/10/2018	Speeding	Yes	50
KL-166451406	10/4/2017	Speeding	No	115	11/13/2017	Speeding	Yes	104
MO-175001406	5/30/2017	Speeding	Yes	118	6/18/2017	Speeding	Yes	115
CM-127151402	3/1/2017	Running a red light	No	55	8/1/2018	Running a red light	No	160
KT-164801402	5/31/2018	Non-moving violation	No	190	11/10/2018	Speeding	No	74
JB-160001402	11/18/2018	Speeding	Yes	220	12/5/2018	Non-moving violation	No	195
LH-170201404	5/6/2018	Running a red light	No	110	9/17/2018	Speeding	Yes	230
BG-1103555	12/25/2017	Speeding	Yes	195	12/8/2018	Speeding	Yes	315
MP-174701406	6/19/2017	Speeding	No	125	10/12/2017	Running a red light	No	175
AK-179051406	10/23/2017	Reckless driving	Yes	200	9/8/2018	Speeding	No	124

我们怎样从这里实现该目标？

在步骤“**Clean 3**”(清理 3) 中，我们有完整的数据集，其中有一列重复显示每个司机的第一次违章日期。

Driver ID	1st Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	01/08/2017	01/08/2017	Speeding	Yes	115
CM-127151402	03/01/2017	03/01/2017	Running a red light	No	55
AP-109151404	03/02/2017	03/02/2017	Non-moving violation	No	95
SH-199751404	03/04/2017	03/04/2017	Speeding	Yes	130
BT-114401404	03/20/2017	03/20/2017	Non-moving violation	No	130
MO-175001406	05/30/2017	05/30/2017	Speeding	Yes	118
RA-1988558	06/02/2017	06/02/2017	Speeding	Yes	144
BT-1168027	06/05/2017	06/05/2017	Speeding	Yes	128
MO-175001406	05/30/2017	06/18/2017	Speeding	Yes	115
MP-174701406	06/19/2017	06/19/2017	Speeding	No	125

我们想要清除不是第一次违章的司机的所有行，同时构建一个只包含第一次违章的数据集。也就是说，我们只想保留“第一次违章日期”与“违章日期”相等时给定司机的信息。筛选为只保留第一次违章的行之后，我们可以移除“**Infraction Date**”(筛选日期)字段，并整理字段名称。

同样，在进行第二次聚合和联接之后，我们有了完整的数据集，其中有一列表示第二次违章日期。

Driver ID	2nd Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	11/21/2018	01/08/2017	Speeding	Yes	115
CM-127151402	08/01/2018	03/01/2017	Running a red light	No	55
AP-109151404	09/24/2018	03/02/2017	Non-moving violation	No	95
SH-199751404	09/27/2017	03/04/2017	Speeding	Yes	130
BT-114401404	11/13/2018	03/20/2017	Non-moving violation	No	130
MO-175001406	06/18/2017	05/30/2017	Speeding	Yes	118
MO-175001406	06/18/2017	06/18/2017	Speeding	Yes	115
MP-174701406	10/12/2017	06/19/2017	Speeding	No	125
AA-106451404	09/28/2017	07/05/2017	Running a red light	No	60
RA-199151402	12/31/2017	07/20/2017	Speeding	Yes	145
SC-202601404	11/10/2018	08/31/2017	Running a red light	No	150

我们可以执行类似的筛选器“**2nd Infraction Date = Infraction Date**”以仅保留每个司机的第二次违章的信息行。同样，我们也可以移除现在多余的“**Infraction Date**”(违章日期) 并整理字段名称。

我们将从第一个违章数据集开始。

25. 在“流程”窗格中，选择“**Clean 3**”(清理 3)，单击加号  图标，并选择“添加分支”。
26. 在这个新的“清理”步骤处于选定状态的情况下，在“配置”窗格的工具栏中单击“**筛选值**”。创建一个筛选器 [1st Infraction Date] = [Infraction Date]。
27. 移除字段“**Infraction Date**”(违章日期)。
28. 将“**Infraction Type**”(违章类型)、“**Traffic School**”(交通法规学习班) 和“**Fine Amount**”(罚款金额) 字段重命名为以“1st”开头。
29. 在“流程”窗格中的步骤下双击名称“**Clean 7**”(清理 7)，并将其重命名为“**Robust 1st**”。

现在为第二个违章数据集执行操作。

30. 在“流程”窗格中的最后一个联接后选择“**Clean 6**”(清理 6)。
31. 在工具栏中单击“**筛选值**”。创建一个筛选器 [2nd Infraction Date] = [Infraction Date]。
32. 移除字段“**Infraction Date**”(违章日期)。
33. 将“**Infraction Type**”(违章类型)、“**Traffic School**”(交通法规学习班) 和“**Fine Amount**”(罚款金额) 字段重命名为以“2nd”开头。
34. 在“流程”窗格中的步骤下双击名称“**Clean 6**”(清理 6)，并将其重命名为“**Robust 2nd**”。

在此阶段，流程应如下所示：



创建完整数据集

既然我们有了这两个包含各个司机第一次和第二次违章完整信息的两个整齐数据集，我们就可以基于“**Driver ID**”(司机 ID) 将它们重新联接在一起，并最终生成我们期望的数据结构。

35. 选择“**Robust 2nd**”，并将其拖到“**Robust 1st**”，放在“联接”上。
36. 默认联接子句“**Driver ID = Driver ID**”应该是正确的。
37. 由于我们不想删除没有第二次违章的司机，我们需要将此联接设为左联接。在“**联接类型**”区域中，单击“**Robust 1st**”旁边图表的无阴影区域，将其转换为“左”联接。
38. 在“流程”窗格中，选择“**Join 3**”(联接 3)，单击加号 \oplus 图标，并选择“添加步骤”。移除重复字段“**Driver ID-1**”(司机 ID-1)。
- 数据处于期望状态，因此我们可以创建输出并继续进行分析。
39. 在“流程”窗格中，选择新添加的“**Clean 6**”(清理 6)，单击加号 \oplus 图标，并选择“添加输出”。
40. 在“输出”窗格中，将“输出类型”更改为 .csv，然后单击“浏览”。为名称输入“**Driver Infractions**”(司机违章)，并在单击“接受”保存之前选择所需的位置。
41. 单击窗格底部的“运行流程” \triangleright 按钮生成输出。在状态对话框中单击“完成”关闭对话框。

提示: 有关输出和运行流程的详细信息，请参见[保存和共享工作](#) 在本页 309。

最终流程应如下所示：



注意：您可以下载完成的流程文件来检查您的工作：[Driver Infractions.tflx](#)

总结

对于此教程的第一个阶段，我们的目标是获取原始数据集，并准备将其用于涉及第一次和第二次违章日期的分析。该过程由三个阶段组成：

确定第一次和第二次违章日期：

1. 创建用于保留“**Driver ID**”(司机 ID) 和“**MIN Infraction Date**”(最小违章日期) 的聚合。将此聚合与原始数据集联接，创建一个“中间数据集”，其中包含为每一行重复的第一次(最小)违章日期。
2. 在新步骤上，筛选出“**1st Infraction Date**”(第一次违章日期) 与“**Infraction Date**”(违章日期) 相同的所有行。从该筛选的数据集中，创建用于保留“**Driver ID**”(司机 ID) 和“**MIN Infraction date**”(最小违章日期) 的聚合。将此聚合与第一个步骤中的中间数据集联接。这将标识第二次违章日期。

为第一次和第二次违章构建干净的数据集：

3. 返回并依据中间数据集创建一个分支，并筛选为仅包含“**1st Infraction Date**”(第一次违章日期) 与“**Infraction Date**”(违章日期) 相同的行。这将只会为第一次违章构建数据集。通过移除任何不必要的字段对其进行整理，并重命名所有所需字段(“**Driver ID**”(司机 ID) 除外)，以指明它们用于第一次违章。这是“**Robust 1st**”数据集。
4. 针对第二次违章日期整理数据集。通过筛选为仅保留“**2nd Infraction Date**”(第二次违章日期) 与“**Infraction Date**”(违章日期) 相同的行，对步骤 2 中的联接结果进行清理。移除任何不必要的字段，并重命名所有所需字段(“**Driver ID**”(司机 ID) 除外)，以指明它们用于第二次违章。这是“**Robust 2nd**”数据集。

将第一次和第二次违章数据合并为一个数据集：

5. 联接“**Robust 1st**”和“**Robust 2nd**”数据集，同时确保保留“**Robust 1st**”中的所有记录以防止丢失没有第二次违章的任何司机。

接着，我们想要探索如何能在 Tableau Desktop 中分析此数据。

继续阅读 **Tableau Desktop 中包含第二个日期的分析** 向下。

注意: 特别感谢 Ann Jackson 的 Workout Wednesday 主题 [Do Customers Spend More on Their First or Second Purchase?](#)(客户在第一次或第二次购买时是否花费更多？) 以及 Andy Kriebel 的 Tableau Prep 技巧 [Returning the First and Second Purchase Dates](#)(返回第一次和第二次购买日期)，这些文章为本教程提供了最初的灵感。单击这些链接会使您离开 Tableau 网站。Tableau 对外部提供商所维护的页面的准确性或新鲜度不担负任何责任。如果您对其内容有任何疑问，请与所有者联系。

Tableau Desktop 中包含第二个日期的分析

这是教程的第二个阶段，并且假定第一个阶段[使用 Tableau Prep 查找第二个日期](#) 在本页 [433](#) 已完成。

在第一个阶段中，我们采用原始数据集并对其进行调整来回答以下问题：

1. 每个司机第一次和第二次违章相隔多长时间(以天数为单位)？
2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？
3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？
4. 有多少司机有过多种类型的违章？
5. 从未上过交通法规学习班的司机的平均罚款金额是多少？

我们现在探讨这些问题时，可以清晰地发现我们创建的第一个数据结构有一些优缺点。我们将返回到 Tableau Prep 中并进行一些额外的重新调整，然后看看会对 Tableau Desktop 同样的分析产生怎样的影响。最后，我们将探讨 Tableau Desktop 独有的将详细级别 (LOD) 表达式与原始数据结合使用进行分析的方法。

本教程的目标是在一个真实场景的上下文中呈现各种概念并演练各个选项 — 而不是规定式地确定哪一个是最好的。最后，您应该会对数据结构对计算和分析的影响有更深的认识，并更加熟悉 Tableau Prep 的各项功能以及 Tableau Desktop 中的计算。

注意: 为了完成本教程中的任务，您需要安装 Tableau Prep 并根据需要安装 Tableau Desktop，并下载相关数据。

若要在继续学习本教程之前安装 Tableau Prep 和 Tableau Desktop，请参见 [Tableau Desktop](#) 和 [Tableau Prep 部署指南](#)。否则，您可以下载 [Tableau Prep](#) 和 [Tableau Desktop](#)。

Desktop 免费试用版。

数据集是在第一个阶段中构建的 [Driver Infractions.tflx](#) 的输出。

Tableau Desktop 中的分析

既然我们已经配置了数据，我们就将其引入 Tableau Desktop。我们可以轻松地回答一些问题，但其他问题会涉及一些(或很多)计算。请尝试下面的问题；如果遇到困难，您可以展开每个问题来了解有关如何继续的基本信息。

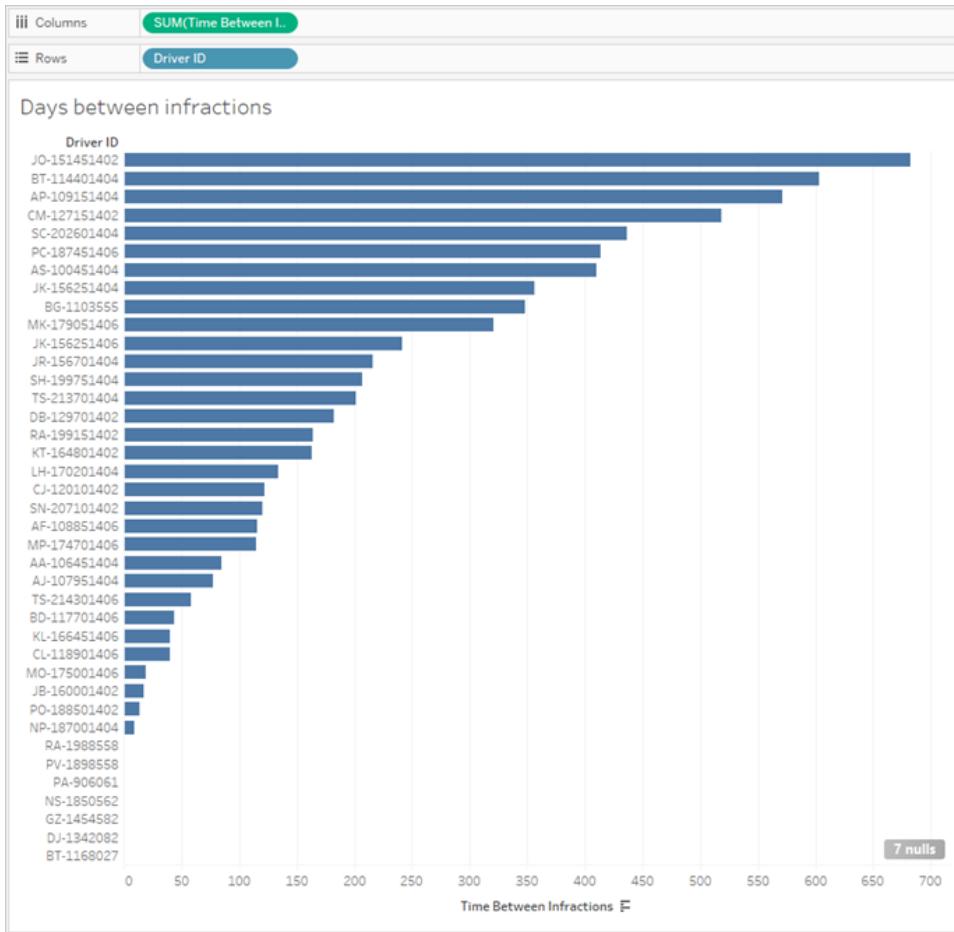
注意：您可以下载工作簿 [Driver Infractions.twbx](#)，在上下文中探讨解决方案。请记住，可以通过一些替代方式来解读分析或寻找答案。

1. 每个司机第一次和第二次违章相隔多长时间(以天数为单位)？

- A. 为了在 Tableau Desktop 中回答此问题，我们将使用 DATEDIFF 函数。此函数采用三个参数—日期部分、开始日期和结束日期。由于我们想知道这些事件的间隔天数，因此将使用日期部分“day”。我们的开始日期和结束日期在数据集中为“**1st Infraction Date**”(第一次违章日期) 和“**2nd Infraction Date**”(第二次违章日期)。
- B. 计算为：

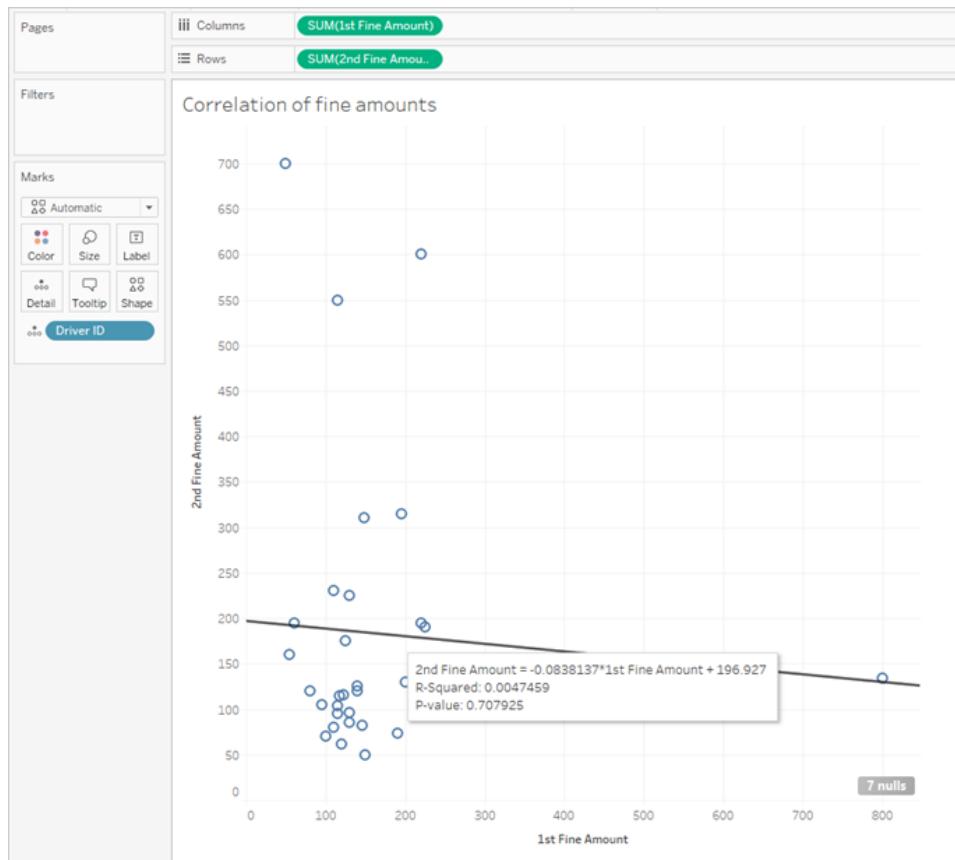
Time Between Infractions(违章间隔时间)=DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])

- C. 我们依据“**Driver ID**”(司机 ID) 将该计算绘制为条形图。请注意，7 名司机没有第二次违章，因此有 7 个 null 值。



2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？

- 为了在 Tableau Desktop 中回答此问题，我们将创建“**1st Fine Amount**”(第一次罚款金额) 和“**2nd Fine Amount**”(第二次罚款金额) 的散点图。通过将“**Driver ID**”(司机 ID) 放到“标记”卡上的“**详细信息**”功能区，我们可以为每名司机创建一个标记。
- 若要添加趋势线，请使用左侧窗格中的“**分析**”选项卡，并显示线性趋势线。将光标悬停在趋势线上，我们可以看到 R 平方值实际上为零，并且 p 值明显高于任何临界值。我们可以确定，第一次罚款金额和第二次罚款金额之间没有相关性。



3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？

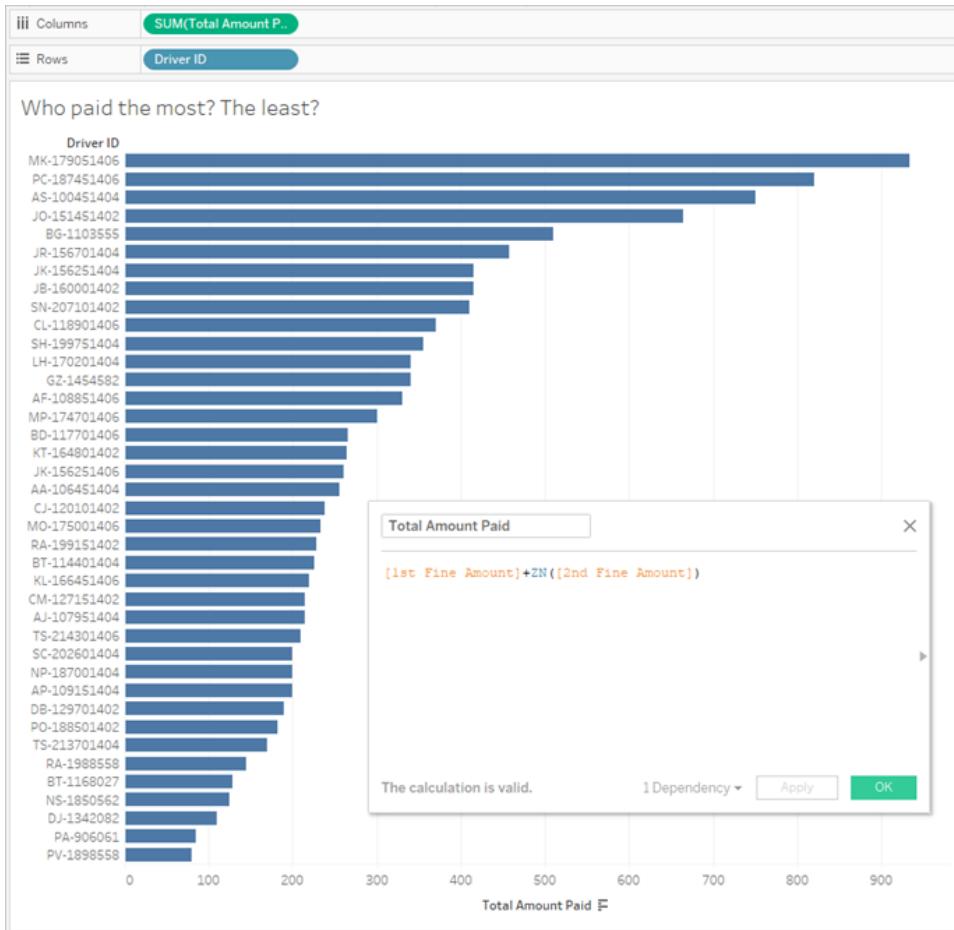
如果想要更深入地进行分析，我们可能需要创建一些计算。

- A. 为了在 Tableau Desktop 中回答此问题，我们需要将两次违章的罚款添加到单一字段中。由于某些司机可能没有第二次违章，我们需要使用零 null ZN 函数将“**2nd Fine Amount**”(第二次罚款金额) 的任何 null 值转换为零。

- B. 计算为：

Total Amount Paid(总支付金额) = [1st Fine Amount] + ZN([2nd Fine Amount])

- C. 我们可以依据“**Driver ID**”(司机 ID) 绘制“**Total Amount Paid**”(总支付金额) 条形图并对其进行排序。



4. 有多少司机有过多种类型的违章？

A. 为了在 Tableau Desktop 中回答此问题，我们需要执行一个更细致的 IF 计算，比较第一次和第二次违章类型是否相同。如果相同，我们想要分配值“1”。如果不同，我们将分配“2”。对于另一种结果（例如第二次违章类型为 null），我们将分配“1”。

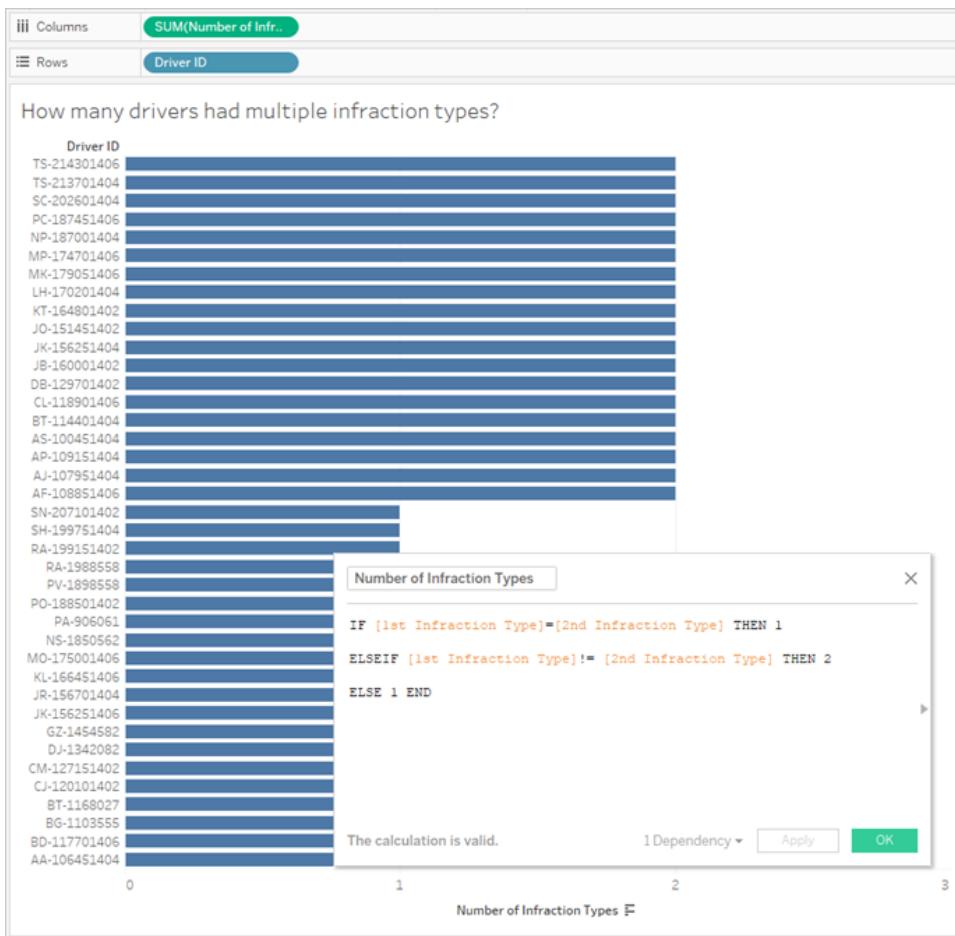
B. 计算为：

Number of Infraction Types(违章类型数量) =

```
IF [1st Infraction Type]=[2nd Infraction Type] THEN 1
ELSEIF [1st Infraction Type] != [2nd Infraction Type] THEN 2
ELSE 1 END
```

C. 然后，我们可以依据“**Driver ID**”(司机 ID) 绘制“**Number of Infraction Types**”(违章类

型数量)条形图并对其进行排序。



5. 从未上过交通法规学习班的司机的平均罚款金额是多少？

- A. 为了在 Tableau Desktop 中回答此问题，我们无法简单地将总罚款金额除以二，因为某些司机只有一次违章。我们也无法计算每个司机的平均罚款并取这些值的平均值，因为对平均值进行平均计算可能会导致不一致。作为替代，我们需要计算从未上过交通法规学习班的司机支付的总金额，然后除以与这些罚款关联的总违章次数。
- 首先，我们需要确定是否每个司机都有第二次违章。如果没有第二次违章，则所有“2nd”字段中的信息都将为 null，我们可以利用这一事实并开始构建计算：

```
IFNULL([2nd Infraction Type], 'no')
```

这将返回违章类型(如果存在)，或者如果没有第二次违章，则返回“no”。

2. 接着，我们需要将此信息转换为违章次数 1 或 2。如果 IFNULL 计算的结果为“no”，则应将司机标记为有一次罚款。对于任何其他结果，则应标记为有两次罚款。计算为：

Number of Infractions(违章次数)=

```
IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1
ELSE 2
END
```

3. 现在，我们需要考虑总罚款金额。与上面的问题 3 类似，我们将添加第一次和第二次罚款金额，并将第二次罚款金额放在 ZN 函数内。但是，由于我们想在整个数据集级别计算此值，因此最佳做法是在计算本身中指定聚合 **SUM**。计算为：

```
SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) )
```

4. 总而言之，我们将利用这个总罚款金额，并将其除以新的**“Number of Infractions”**(违章次数)计算字段，从而确定平均罚款金额：

Average Fine(平均罚款)= (SUM([1st Fine Amount]) + SUM(ZN ([2nd Fine Amount]))) / SUM([Number of Infractions])

- B. 我们还需要筛选出曾经参加过交通法规学习班的司机 —— 但该信息也跨两个字段存储。

1. Tableau 在进行数值计算时非常高效。我们将用尽可能多可帮助提高性能的数字表述这一点。为了合并这两个字段，我们将为每个字段创建一个显示为“**Yes = 1**”和“**No = 0**”的字段(对于没有第二次违章的司机，**null** 也应 = 0)。通过对这些计算的结果求和，总值为 0 的任何司机从未上过交通法规学习班(值 1 或 2 表示他们上过交通法规学习班的次数)。我们随后可以进行筛选，以仅保留值为 0 的司机。
2. 这一次，我们将使用 CASE 语句，而不是 IF。这些语句的功能非常相似，但具有不同的语法。计算的开头应如下所示：

```
CASE [1st Traffic School]
WHEN 'Yes' THEN 1
WHEN 'No' THEN
```

```

ELSE 0
END

```

3. 然后，我们将为“2nd Traffic School”执行相同的操作。通过将每个 CASE 语句放在括号内并在两者之间添加一个加号，我们可以在同一计算中添加这两段语句。移除一些换行符，它看起来如下所示：

Number of Traffic School Attendances(参加交通法规学习班次数)=

```

(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
THEN 0 ELSE 0 END)
+
(CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
THEN 0 ELSE 0 END)

```

4. 如果我们将**“Number of Traffic School Attendances”**(参加交通法规学习班次数) 拖到“数据”窗格的“维度”区域，则值 0–2 将变为离散。
5. 现在，如果我们对**“Number of Traffic School Attendances”**(参加交通法规学习班次数) 进行筛选，我们可以只选择 0，并且知道我们获得了从未参加过交通法规学习班的司机。
- C. 为了回答原来的问题，我们只需将**“Average Fine”**(平均罚款) 放到“标记”卡上的“文本”功能区。由于我们在计算中构建了聚合，因此字段上的聚合将为 **AGG**，我们无法对其进行更改。这是预期行为。

The screenshot shows the Tableau Prep Data Flow interface. On the left, the 'Marks' card has 'Automatic' selected and 'Text' chosen. Below it, 'AGG(Average Fine)' is selected. In the center, the 'Average amount paid with no traffic school' is displayed as 161.8. A tooltip window for the calculated field 'Number of Traffic School Atterr' is open, showing the detailed formula:

```

(CASE [1st Traffic School]
WHEN 'Yes' THEN 1
WHEN 'No' THEN 0
ELSE 0 END)
+
(CASE [2nd Traffic School]
WHEN 'Yes' THEN 1
WHEN 'No' THEN 0
ELSE 0 END)

```

At the bottom of the tooltip, it says "The calculation is valid." and has "1 Dependency" and "OK" buttons.

更进一步 – 转置的数据

尽管我们一直在处理的数据结构良好，可以解决与第一次和第二次违章特定相关的问题，但它并不是我们建议用于 Tableau Desktop 的标准结构。我们的分析背离与违章日期相关的基本问题越远，用于将相关信息合并为可用形式的计算就会变得越复杂。

通常，如果为同一类型的数据使用多个列存储数据(例如两个列用于日期，两个列用于罚款金额等)，并且唯一的信息存储在字段名中(例如是第一次还是第二次违章)，这就表明应该要对数据进行转置。

执行多重转置可以很好地处理这一点。返回到 Tableau Prep 中，并从上面的逐步教程中创建的“**Driver Infraction**”(司机违章) Tableau Prep 流程的结尾开始工作：

1. 从最后的清理步骤中，添加一个“**Pivot**”(转置) 步骤，该步骤将按每个重复的字段进行转置。

提示：使用“**转置的字段**”区域右上角的加号 图标来添加更多“**转置值**”。每组字段(例如第一次和第二次罚款金额)应一起转置。
有关转置的详细信息，请参见[清理和调整数据 在本页 223](#)。

通过移除 null 日期以及重命名字段并对其进行重新排序，可以对结果进行整理。

2. 在转置后面添加清理步骤。在“**Infraction Date**”(违章日期) 列中，右键单击 null 栏并选择“**排除**”。
3. 双击字段名“**Pivot1 Names**”(转置 1 名称)，并将其重命名为“**Infraction Number**”(违章次数)。

4. 视情况而定拖动数据以按如下方式对其进行重新排序：

Driver ID	Infraction Number	Infraction Date	Infraction Type	Traffic School	Fine Amount
MO-175001406	1st	05/30/2017	Speeding	Yes	118
SH-199751404	1st	03/04/2017	Speeding	Yes	130
AA-106451404	1st	07/05/2017	Running a red light	No	60
MP-174701406	1st	06/19/2017	Speeding	No	125
PO-188501402	1st	10/30/2017	Speeding	Yes	120
KL-166451406	1st	10/04/2017	Speeding	No	115
RA-199151402	1st	07/20/2017	Speeding	Yes	146
AJ-107951404	1st	10/15/2017	Speeding	Yes	130
BD-117701406	1st	12/25/2017	Speeding	Yes	140
CJ-120101402	1st	11/26/2017	Speeding	Yes	122
SN-207101402	1st	12/27/2017	Speeding	Yes	280
TS-213701404	1st	10/23/2017	Speeding	Yes	100
JR-156701404	1st	12/24/2017	Speeding	No	148
CM-127151402	1st	03/01/2017	Running a red light	No	55
JK-156251406	1st	12/25/2017	Speeding	Yes	140
AE-109051406	1st	05/04/2019	Non-moving violation	No	200

5. 依据新的转置的数据创建一个名为“**Pivoted Driver Infractions**”(转置的司机违章) 的输出，并将其引入 Tableau Desktop。(在添加“输出”步骤之后，不要忘记运行流程。)

现在，我们可以用这个转置的数据结构再次探讨我们的五个问题；如果遇到困难，您可以展开每个问题来了解有关如何继续操作的基本信息。

注意：您可以下载完成的流程文件 [Pivoted Driver Infractions.tflx](#) 来检查您的工作，或者下载工作簿 [Pivoted Driver Infractions.twbx](#) 在上下文中查看解决方案。请记住，可以通过一些替代方式来解读分析或寻找答案。

1. 每个司机第一次和第二次违章相隔多长时间(以天数为单位)？

- A. 为了在 Tableau Desktop 中回答此问题，就像我们对第一个数据集所做的那样，我们将使用 DATEDIFF 函数。此函数需要一个开始日期和一个结束日期。此信息存在于我们的数据中，但全部在一个字段中。我们需要将其拉出到两个字段中。

1. 创建两个初始计算字段：

```
1st Infraction(第一次违章)=IF [Infraction Number] = "1st"
THEN [Infraction Date] END
```

2nd Infraction(第二次违章)= IF [Infraction Number] = "2nd"
THEN [Infraction Date] END

- 由于我们想要确保这两个字段均可用于为每个司机比较，我们需要将它们固定到“Driver ID”(司机 ID) 级别。

注意: 不相信我？尝试按原样对这两个字段进行 DATEDIFF 计算：**Time**

Between Infractions(违章间隔时间)= DATEDIFF('day', [1st
Infraction], [2nd Infraction])

您将在所有地方都获得 null 结果，因为 Tableau 会尝试跨类似于如下的数据结构进行比较：

Driver ID	1st Infraction Date	2nd Infraction Date	Time between infractions
AA-106451404	Null	9/28/2017	Null
	7/5/2017	Null	Null
AF-108851406	Null	9/2/2018	Null
	5/9/2018	Null	Null
AJ-107951404	Null	12/31/2017	Null
	10/15/2017	Null	Null

在这里，知道第一个日期的行不知道第二个日期，反之亦然。为了解决此问题，我们将使用“FIXED”详细级别表达式将第一个和第二个日期强制与“Driver ID”(司机 ID) 相关。

按如下方式编辑计算：

1st Infraction(第一次违章)= { FIXED [Driver ID] : MIN (IF
[Infraction Number] = "1st" THEN [Infraction Date] END)
}

2nd Infraction(第二次违章)= { FIXED [Driver ID] : MIN (IF
[Infraction Number] = "2nd" THEN [Infraction Date] END)
}

注意: 如果嵌入在 LOD 表达式中，必须对原始计算进行聚合。我们可以使用任何将保留日期值的基本聚合(因此像 SUM、AVG 或 MIN 这样的聚合有用，而 CNT 或 CNTD 则不行)。

- 现在，我们可以按如下方式创建 DATEDIFF 计算：

Time Between Infractions(违章间隔时间)=DATEDIFF('day', [1st Infraction], [2nd Infraction])

4. 如果我们想要查看周或月，只需修改日期部分(当前为 'day')。

结果将与使用未转置数据结构的结果相同。

2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？

A. 为了在 Tableau Desktop 中回答此问题，我们将使用与上一个问题非常相似的逻辑。我们将使用“**Infraction Number**”(违章次数) 来确定给定行是第一次违章还是第二次违章，然后相应拉出罚款金额。

1. 如果我们只需要建立散点图，则可以跳过 LOD 部分，只使用 IF 计算：

1st Fine Amount(第一次罚款金额)=IF [Infraction Number] = "1st" THEN [Fine Amount] END

2nd Fine Amount(第二次罚款金额)=IF [Infraction Number] = "2nd" THEN [Fine Amount] END

2. 但是，如果我们想要进行比较并查看某个司机第一次和第二次罚款之间的金额差异，我们会遇到像日期一样的 null 问题。将这些计算封装在 FIXED LOD 内不会有坏处，因此从一开始这样做或许会不错：

1st Fine Amount(第一次罚款金额)={ FIXED [Driver ID] : MIN (IF [Infraction Number] = "1st" THEN [Fine Amount] END) }

2nd Fine Amount(第二次罚款金额)={ FIXED [Driver ID] : MIN (IF [Infraction Number] = "2nd" THEN [Fine Amount] END) }

3. 创建一个散点图，并像以前一样显示线性趋势线。

结果将与使用未转置数据结构的结果相同。

3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？

- A. 为了在 Tableau Desktop 中回答此问题，转置的数据结构非常理想。我们只需在条形图中显示“**Driver ID**”(司机 ID) 和“**Fine Amount**”(罚款金额)。默认聚合已经是 **SUM**，因此将自动绘制司机支付的总金额。

结果将与使用未转置数据结构的结果相同。

4.有多少司机有过多种类型的违章？

- A. 为了在 Tableau Desktop 中回答此问题，转置的数据结构非常理想。我们只需以条形图形式显示“**Driver ID**”(司机 ID) 以及“**Violation Type**”(违章类型) 的“不重复计数”，就能得到答案。

结果将与使用未转置数据结构的结果相同。

5.从未上过交通法规学习班的司机的平均罚款金额是多少？

- A. 为了在 Tableau Desktop 中回答此问题，我们无法简单地将总罚款金额除以二，因为某些司机只有一次违章。我们也无法计算每个司机的平均罚款并取这些值的平均值，因为对平均值进行平均计算可能会导致不一致。作为替代，我们需要计算从未上过交通法规学习班的司机支付的总金额，然后除以与这些罚款关联的总违章次数。

- 首先，我们需要确定是否每个司机都有第二次违章。如果没有第二次违章，则“**2nd Infraction Date**”(第二次违章日期) 将为 **null**，我们可以利用这一事实并开始构建计算：

```
IFNULL(STR([2nd Infraction]), 'no')
```

这将返回第二次违章的日期(如果存在)，或者如果没有第二次违章，则返回“no”。

注意: 此计算的 STR 部分是必需的，因为 IFNULL 需要在其参数中保持数据类型的一致性。由于我们想要为 **null** 值返回字符串“no”，因此也要将日期转换为字符串。

- 接着，我们需要将此信息转换为违章次数 1 或 2。如果 IFNULL 计算的结果为“no”，则应将司机标记为有一次罚款。对于任何其他结果，则应标记为有两次罚

款。计算为：

Number of Infractions(违章次数) =

```
IF ISNULL(STR([2nd Infraction]), 'no')= 'no' THEN 1
ELSE 2
END
```

3. 现在，我们需要考虑平均罚款金额。我们已经有了一个“**Fine Amount**”(罚款金额) 字段。我们只需将该字段除以新的“**Number of Infractions**”(违章次数) 字段，同时将两者都放在 **SUM** 内：

Average Fine(平均罚款) = (SUM([Fine Amount]) / SUM([Number of Infractions]))

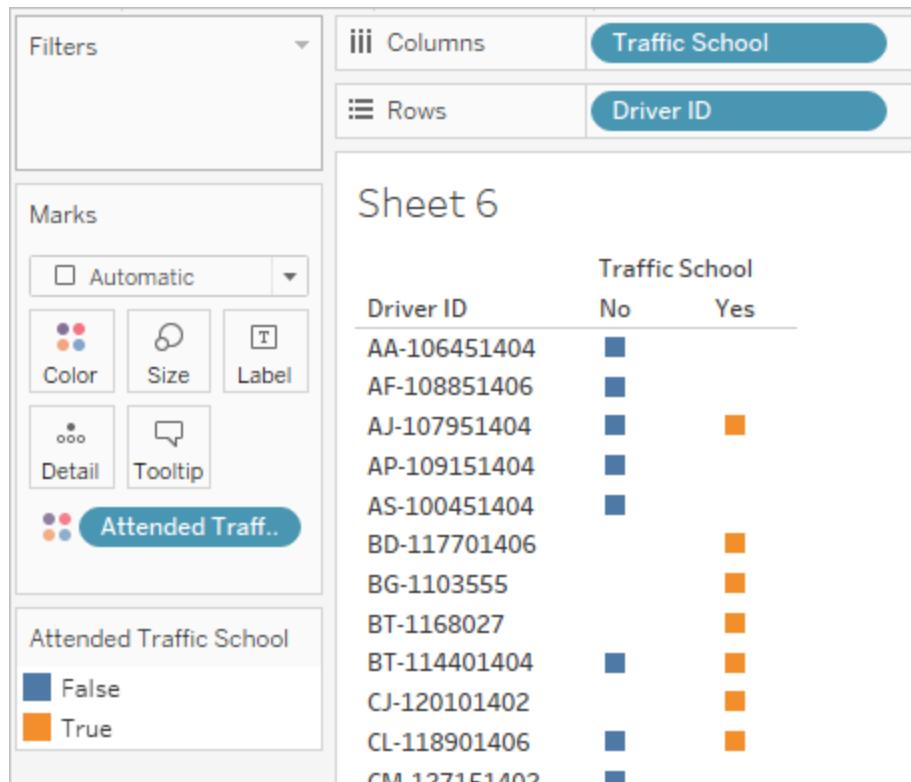
B. 我们还需要筛选出参加过交通法规学习班的司机。看起来我们可以使用“**Traffic School**”(交通法规学习班) 字段，并针对“**Traffic School = no**”进行筛选。但是，这将会对与交通法规学习班无关的违章进行筛选，而不是从未上过交通法规学习班的司机。如果一名司机为一次违章上过交通法规学习班，但没有为第二次上过，则我们不需要在此处考虑第二次违章。

我们需要做的是筛选出参加过交通法规学习班的任何司机。就数据而言，我们想要筛选出任何行上“**Traffic School**”(交通法规学习班) 为“**Yes**”(是) 的任何司机。让我们分阶段构建计算，并使用一个简单的视图来帮助跟踪所发生的情况：

1. 首先，我们想要知道司机的“**Traffic School**”(交通法规学习班) 是否为“**Yes**”(是)。将“**Driver ID**”(司机 ID) 拖到“行”，并将“**Traffic School**”(交通法规学习班) 拖到“列”。我们将获得一个文本表，其中占位符“Abc”文本表示每个司机的相关值。
2. 接着，我们想要构建一个计算，该计算将确定“**Traffic School**”(交通法规学习班) 的值是否为“**Yes**”(是)。计算的第一个阶段是：

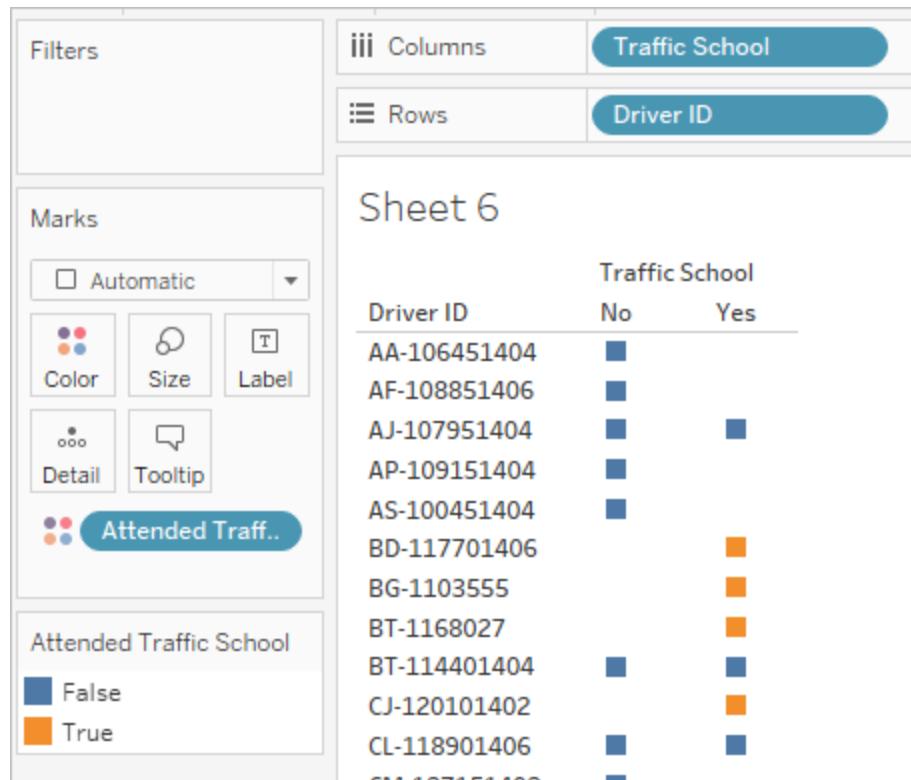
Attended Traffic School(已参加交通法规学习班) = CONTAINS ([Traffic School), 'Yes')

如果我们将“**Attended Traffic School**”(已参加交通法规学习班) 拖到“**标记**”卡上的“**颜色**”功能区，我们会看到它准确地针对“**No**”(否) 列中的每个标记标为“**False**”，并针对“**Yes**”(是) 列中的每个标记标为“**True**”。



3. 但是，我们真正需要的是位于司机(而不是违章)级别的此信息。尝试在与基本数据结构不同的详细级别计算结果时，LOD 表达式非常适合。我们将此表达式设为 FIXED LOD 表达式。但是，正如我们所知，必须对 LOD 的聚合表达式进行聚合。以前，我们使用过 **MIN**。它在这里是否有用？我们将计算修改为：

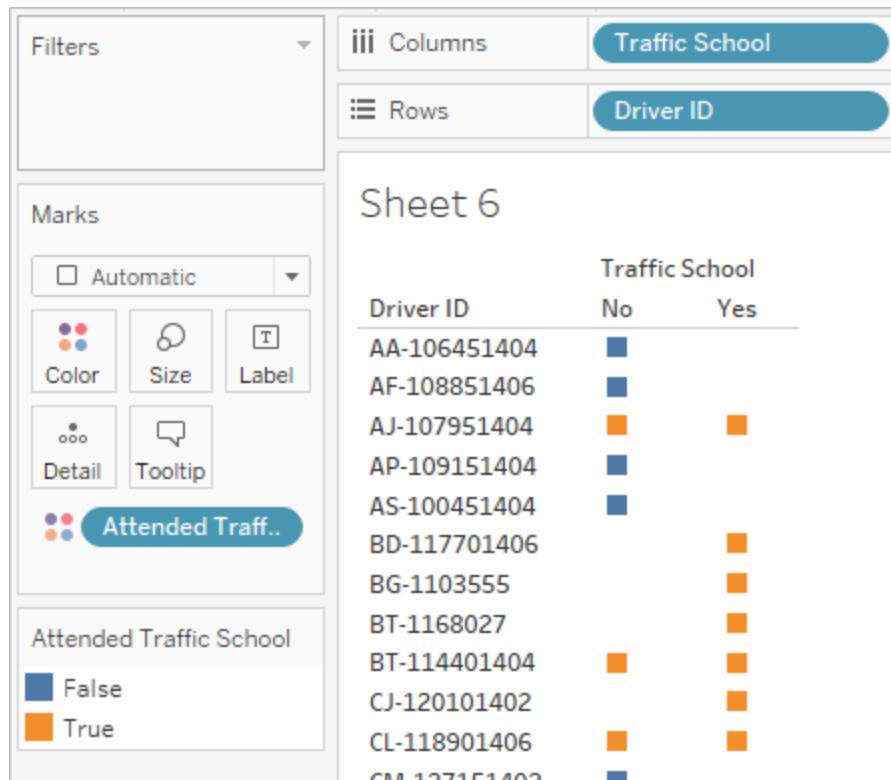
Attended Traffic School(已参加交通法规学习班) = { FIXED [Driver ID] : MIN(CONTAINS([Traffic School], 'Yes')) }



在视图中应用该更改之后，我们看到了与预期相反的结果。具有“**No**”值的任何司机全部被标记为“**False**”。相反，我们希望为该司机的每条记录将“**Yes**”标记为“**True**”。**MIN** 在这里做了什么？它按字母顺序选取第一个响应，即“**No**”。

4. 如果我们将其更改为 **MAX**，将会怎么样？该函数是否会按字母顺序获取最后一个响应？我们将计算修改为：

```
Attended Traffic School(已参加交通法规学习班)) = { FIXED [Driver ID] : MAX ( CONTAINS( [Traffic School], 'Yes') ) }
```



我们得到了想要的结果：如果司机在数据中的任何位置有“**Yes**”，他们会因为参加过交通法规学习班而被标记为“**True**”，即使对于不涉及交通法规学习班的违章也是如此。

5. 如果我们将“**Attended Traffic School**”(已参加交通法规学习班) 拖到“**筛选器**”功能区并仅选择“**False**”，将只会剩下从未参加过交通法规学习班的司机。
- C. 为了回答原来的问题，在筛选器就位的情况下，我们只需将“**Average Fine**”(平均罚款) 放到“**标记**”卡上的“**文本**”功能区。由于我们在计算中构建了聚合，因此字段上的聚合将为 **AGG**，我们无法对其进行更改。这是预期行为。

结果将与使用未转置数据结构的结果相同。

转置数据的优点

如果知道我们只需要回答可以使用教程中的原始数据结构轻松回答的问题，我们可以继续使用该结构。不过，转置的数据格式更加灵活。即使它需要一些计算，但在计算就位之后，生成的数据集将非常适合于回答更广泛的问题。

再进一步 – 只使用计算

如果您无法访问 Tableau Prep, 将会怎么样? 如果被迫使用原始数据, 您是否完全不走运? 根本不会!

Tableau Desktop 和 LOD 表达式可以回答所有分析问题。如果我们连接到原始的 **Traffic Violations.xlsx**, 它看起来与已转置数据集非常相似 — 只是没有重要的“**Violation Number**”(违章次数) 字段。我们将需要通过 LOD 表达式模拟聚合步骤的输出。

注意: 您可以下载工作簿 **LOD Driver Infractions.twbx**, 在上下文中探讨解决方案。请记住, 可以通过一些替代方式来解读分析或寻找答案。

1. 每个司机第一次和第二次违章相隔多长时间(以天数为单位) ?

A. 为了在 Tableau Desktop 中回答此问题, 我们将再次使用 DATEDIFF 函数。此函数需要一个开始日期和一个结束日期。此信息存在于我们的数据中, 但全部在一个字段中。我们需要将其拉出到两个字段中。由于我们想要确保这两个字段均可用于为每个司机比较, 我们需要将它们固定到“**Driver ID**”(司机 ID) 级别。

1. 为了查找第一次违章日期, 我们将使用计算:

1st Infraction(第一次违章)= { FIXED [Driver ID] : MIN ([Infraction Date]) }

2. 我们将分阶段执行第二次违章日期相关操作。

- a. 首先, 我们需要只查看比第一个日期大的日期:

```
IF [Infraction Date] > [1st Infraction] THEN
[Infraction Date] END
```

- b. 但这将列出第一次违章之后每次违章, 而我们只需要第二次违章。因此我们需要这些日期中最小的日期。将整个计算放在 **MIN** 内:

```
MIN( IF [Infraction] : [1st Infraction] THEN
[Infraction Date] END )
```

- c. 我们还想要为每个司机重新计算第二次违章日期。这就是 LOD 表达式的用武之地。我们会将此计算固定到“**Driver ID**”(司机 ID) 级别:

```
2nd Infraction(第二次违章)= { FIXED [Driver ID] : MIN ( IF  
[Infraction Date] > [1st Infraction] THEN  
[Infraction Date] END ) }
```

3. 现在我们可以创建 DATEDIFF 计算:

```
Time Between Infractions(违章间隔时间)=DATEDIFF('day', [1st  
Infraction], [2nd Infraction])
```

结果将与使用其他两个数据结构的结果相同。

2. 比较第一次和第二次违章的罚款金额。这些金额是否相关？

- 为了在 Tableau Desktop 中回答此问题，我们将使用此问题的转置数据版本相似的逻辑。我们将使用为问题 1 创建的“**1st Infraction**”(第一次违章) 和“**2nd Infraction**”(第二次违章) 来确定给定行是第一次违章还是第二次违章，然后相应拉出罚款金额。

1. 如果我们只需要建立散点图，则可以跳过 LOD 部分，只使用 IF 计算:

```
1st Fine Amount(第一次罚款金额)=IF [1st Infraction] =  
[Infraction Date] THEN [Fine Amount] END
```

```
2nd Fine Amount(第二次罚款金额)=IF [2nd Infraction] =  
[Infraction Date] THEN [Fine Amount] END
```

2. 但是，如果我们想要进行比较并查看某个司机第一次和第二次罚款之间的金额差异，我们会遇到像第一个数据结构中一样的 null 问题。将这些计算封装在 FIXED LOD 内不会有坏处，因此从一开始这样做或许会不错：

```
1st Fine Amount(第一次罚款金额)={ FIXED [Driver ID] : MIN ( IF  
[1st Infraction] = [Infraction Date] THEN [Fine  
Amount] END ) }
```

```
2nd Fine Amount(第二次罚款金额)={FIXED [Driver ID] : MIN( IF  
[2ndInfraction] = [Infraction Date] THEN [Fine Amount]  
END ) }
```

结果将与使用其他两个数据结构的结果相同。

3. 总体而言，哪个司机支付的罚款最多？谁支付的最少？

- A. 为了在 Tableau Desktop 中回答此问题，我们需要了解有关仅 LOD 方法的一些信息。使用 Tableau Prep 的两种方法都会为司机筛选出不是第一次或第二次违章的记录。Tableau Desktop 中的 LOD 方法会保留所有记录。这意味着，如果我们按“**Driver ID**”(司机 ID) 创建“**SUM(Amount Paid)**”的可视化项，仅 Tableau Desktop 版本将为违章超过两次的司机显示更高的金额。若要从与其他方法匹配的完整数据中获取“**Total Amount Paid**”(总支付金额) 值，而不是使用原始“**Fine Amount**”(罚款金额) 字段，则我们需要像使用第一个数据结构所做的那样将第一次和第二次罚款加总。
- B. 我们将使用为问题 2 创建的字段添加两个罚款金额。为了防止只有一次违章的任何司机出现 null 结果，ZN 是必需的。计算为：

Total Amount Paid(总支付金额)= [1st Fine Amount] + ZN([2nd Fine Amount])

结果将与使用其他两个数据结构的结果相同。

4. 有多少司机有过多种类型的违章？

- A. 为了在 Tableau Desktop 中回答此问题，我们不能简单地显示“**Driver ID**”(司机 ID) 和“**Infraction Type**”(违章类型) 的不重复计数。由于此数据集有超过第二次的违章，因此某些司机可能有两种以上的违章类型。为了使结果与其他方法相符，我们需要将范围限制为仅包含前两次违章。
- B. 我们可以拉出第一次和第二次违章类型，将这些类型放在 LOD 表达式中，使它们 FIXED (固定) 到司机级别，然后使用 IF 计算来对类型计数：

1. **1st Infraction Type**(第一次违章类型)= { FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END) }

2. **2nd Infraction Type**(第二次违章类型)= { FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END) }

3. **Number of Infraction Types**(违章类型数量)=

```
IF [1st Infraction Type] = [2nd Infraction Type] THEN 1
ELSEIF [1st Infraction Type] != [2nd Infraction Type]
```

```
THEN 2  
ELSE 1 END
```

注意: 也可以通过直接将初始计算嵌入在较大的计算中, 以单一字段的形式创建多个这样的计算。此处的合并计算将如下所示:

```
IF  
  {FIXED [Driver ID] : MIN(IF [1st Infraction]=  
    [Infraction Date] THEN [Infraction Type] END)}  
  =  
  {FIXED [Driver ID] : MIN(IF [2nd Infraction]=  
    [Infraction Date] THEN [Infraction Type] END)}  
THEN 1  
  
ELSEIF  
  {FIXED [Driver ID] : MIN(IF [1st Infraction]=  
    [Infraction Date] THEN [Infraction Type] END)}  
  !=  
  {FIXED [Driver ID] : MIN(IF [2nd Infraction]=  
    [Infraction Date] THEN [Infraction Type] END)}  
THEN 2  
  
ELSE 1  
END
```

理解起来有点困难, 但如果愿意使用的话, 它确实有用。(请注意, 换行符和一些空格不会影响 Tableau 解读计算的方式。)

- C. 然后, 我们可以依据**“Driver ID”**(司机 ID)绘制**“Number of Infraction Types”**(违章类型数量)条形图并对其进行排序。

结果将与使用其他两个数据结构的结果相同。

5.从未上过交通法规学习班的司机的平均罚款金额是多少?

A. 为了在 Tableau Desktop 中回答此问题，我们无法简单地将总罚款金额除以二，因为某些司机只有一次违章。我们也无法计算每个司机的平均罚款并取这些值的平均值，因为对平均值进行平均计算可能会导致不一致。作为替代，我们需要计算从未上过交通法规学习班的司机支付的总金额，然后除以与这些罚款关联的总违章次数。

- 首先，我们需要确定是否每个司机都有第二次违章。如果没有第二次违章，则所有“2nd”字段中的信息都将为 `null`，我们可以利用这一事实并开始构建计算：

```
IFNULL([2nd Infraction Type], 'no')
```

这将返回违章类型(如果存在)，或者如果没有第二次违章，则返回“no”。

- 接着，我们需要将此信息转换为违章次数 1 或 2。如果 `IFNULL` 计算的结果为“no”，则应将司机标记为有一次罚款。对于任何其他结果，则应标记为有两次罚款。计算为：

Number of Infractions(违章次数)=

```
IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1
ELSE 2
END
```

- 对于“Total Amount Paid”(总支付金额)，我们可以使用问题 3 中的计算。总而言之，我们将利用这个总罚款金额，并将其除以新的**“Number of Infractions”**(违章次数)计算字段，从而确定平均罚款金额：

Average Fine(平均罚款)= SUM([Total Amount Paid]) / SUM
([Number of Infractions])

B. 我们还需要筛选出参加过交通法规学习班的司机。由于此数据集包含一些有第三次或第四次违章的司机，因此我们无法使用与转置数据结构相同的方法。相反，我们将采用与未转置数据相同的方法，总结如下：

- 首先，我们需要构建两个计算，确定第一次和第二次违章是否涉及交通法规学习班：

1st Traffic School(第一次交通法规学习班)= { FIXED [Driver ID] :
MIN (IF [1st Infraction] = [Infraction Date] THEN
[Traffic School] END) }

2nd Traffic School(第二次交通法规学习班)= { FIXED [Driver ID] :
MIN (IF [2nd Infraction] = [Infraction Date] THEN
[Traffic School] END) }

2. 然后，我们将添加这些值来获取参加交通法规学习班的总次数：

Number of Traffic School Attendances(参加交通法规学习班次数)=

```
(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'  
THEN 0 ELSE 0 END)  
+  
(CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'  
THEN 0 ELSE 0 END)
```

3. 如果我们将“**Number of Traffic School Attendances**”(参加交通法规学习班次数) 拖到“数据”窗格的“维度”区域，则值 0–2 将变为离散。
4. 现在，如果我们将“**Number of Traffic School Attendances**”(参加交通法规学习班次数) 进行筛选，我们可以只选择 0，并且知道我们获得了从未参加过交通法规学习班的司机。
- C. 为了回答原来的问题，我们只需将“**Average Fine**”(平均罚款) 放到“标记”卡上的“文本”功能区。由于我们在计算中构建了聚合，因此字段上的聚合将为 **AGG**，我们无法对其进行更改。这是预期行为。

结果将与使用其他两个数据结构的结果相同。

请务必记住，此解决方案包含大量的嵌套计算和 LOD 表达式。视数据集的大小以及数据集的复杂性而定，可能会出现性能问题。

对各种方法的反思

那么您应该采用哪种方法呢？这完全取决于您，并且工具任由您随意使用。

- 如果您想要避开 LOD，则可以使用数据调整解决方案，尽管对于某些分析([Tableau Desktop 中的分析 在本页 446](#)) 计算可能是必需的。
- 如果您能够调整数据并且熟悉计算(包括 LOD)，则中间型选项可以提供最好的灵活性([更进一步 — 转置的数据 在本页 453](#))。
- 如果您熟悉 LOD、对性能的影响较小，并且/或者您无法访问 Tableau Prep，则单独使用 LOD 解决此问题是可行的选项([再进一步 — 只使用计算 在本页 462](#))。

至少，它对于了解 Tableau Prep 中的聚合和 Tableau Desktop 中的详细级别表达式如何相互关联以及会对数据分析产生怎样的影响是有价值的。如同 Tableau 中的大多数功能一样，执行任何操作的方法都有多种。探索各种各样的选项可以帮助将概念联系在一起，让您能挑选最适合于自己的解决方案。

使用的计算：

Driver Infractions(司机违章)

- **Time Between Infractions**(违章间隔时间)= DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])
- **Total Amount Paid**(总支付金额)=[1st Fine Amount] + ZN([2nd Fine Amount])
- **Number of Infraction Types**(违章类型数量)=IF [1st Infraction Type]=[2nd Infraction Type] THEN 1 ELSEIF [1st Infraction Type]!= [2nd Infraction Type] THEN 2 ELSE 1 END
- **Number of Infractions**(违章次数)=IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END
- **Average Fine**(平均罚款)=(SUM([1st Fine Amount]) + SUM(ZN([2nd Fine Amount]))) / SUM([Number of Infractions])
- **Number of Traffic School Attendances**(参加交通法规学习班次数)=(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)

Pivoted Driver Infractions(转置的司机违章)

- **1st Infraction**(第一次违章)={FIXED [Driver ID] : MIN(IF [Infraction Number] = "1st" THEN [Infraction Date] END) }
- **2nd Infraction**(第二次违章)={FIXED [Driver ID] : MIN(IF [Infraction Number] = "2nd" THEN [Infraction Date] END) }
- **Time Between Infractions**(违章间隔时间)=DATEDIFF('day', [1st Infraction], [2nd Infraction])
- **1st Fine Amount**(第一次罚款金额)={FIXED [Driver ID] : MIN(IF [Infraction Number] = "1st" THEN [Fine Amount] END) }
- **Number of Infractions**(违章次数)= IF IFNULL(STR([2nd Infraction]),

```
'no') = 'no' THEN 1 ELSE 2 END
```

- **Average Fine**(平均罚款)=SUM([Fine Amount])/SUM([Number of Infractions])
- **Attended Traffic School**(已参加交通法规学习班)={ FIXED [Driver ID] : MAX(CONTAINS([Traffic School], 'Yes')) }

LOD Driver Infractions(LOD 司机违章)

- **1st Infraction**(第一次违章)={FIXED [Driver ID] : MIN([Infraction Date])}
- **2nd Infraction**(第二次违章)={ FIXED [Driver ID] : MIN(IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END) }
- **Time Between Infractions**(违章间隔时间)=DATEDIFF('day', [1st Infraction], [2nd Infraction])
- **1st Fine Amount**(第一次罚款金额)={FIXED [Driver ID] : MIN(IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END) }
- **2nd Fine Amount**(第二次罚款金额)={FIXED [Driver ID] : MIN(IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END) }
- **Total Amount Paid**(总支付金额)=[1st Fine Amount] + ZN([2nd Fine Amount])
- **1st Infraction Type**(第一次违章类型)={FIXED [Driver ID] : MIN(IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END) }
- **2nd Infraction Type**(第二次违章类型)={FIXED [Driver ID] : MIN(IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END) }
- **Number of Infraction Types**(违章类型数量)=IF [1st Infraction Type]=[2nd Infraction Type] THEN 1 ELSEIF [1st Infraction Type] != [2nd Infraction Type] THEN 2 ELSE 1 END
- **Number of Infractions**(违章次数)=IF ISNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END

- **Average Fine(平均罚款)**= SUM ([Total Amount Paid]) / SUM([Number of Infractions])
- **1st Traffic School(第一次交通法规学习班)**= {FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Traffic School] END) }
- **2nd Traffic School(第二次交通法规学习班)**= {FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Traffic School] END) }
- **Number of Traffic School Attendances(参加交通法规学习班次数)**= (CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)

注意: 特别感谢 Ann Jackson 的 Workout Wednesday 主题 [Do Customers Spend More on Their First or Second Purchase?\(客户在第一次或第二次购买时是否花费更多? \)](#) 以及 Andy Kriebel 的 Tableau Prep 技巧 [Returning the First and Second Purchase Dates\(返回第一次和第二次购买日期\)](#), 这些文章为本教程提供了最初的灵感。单击这些链接会使您离开 Tableau 网站。Tableau 对外部提供商所维护的页面的准确性或新鲜度不担负任何责任。如果您对其内容有任何疑问, 请与所有者联系。

Tableau Prep Builder 疑难解答

本文列出了您在使用 Tableau Prep Builder 时可能遇到的问题以及有关如何解决这些问题的建议。

使用命令行运行流程时的常见错误

您可以从命令行中以编程方式运行流程来刷新输出文件，而不是打开 Tableau Prep Builder 以手动方式运行每个流程。尽管此过程可帮助提高流程进程的效率，但如果语法不正确，或者缺少连接或输出位置的凭据，则在运行此过程时会出现错误。

下表描述了常见错误以及如何解决这些错误。有关如何从命令行中运行流程的信息，请参见[通过命令行刷新流程输出文件 在本页 321](#)。

错误	原因	如何修复
“Missing arguments”(缺少参数)	缺少必需的命令行参数之一。	使用“tableau-prep-cli -help”来查看命令行参数的列表。
“Unable to read the connections file.”(无法读取连接文件。)	输入连接的 credentials.json 文件中的语法或格式存在错误。	请检查 .json 文件中输入连接的语法。有关详细信息和示例，请参见 通过命令行刷新流程输出文件 在本页 321 。
“There are errors in the flow.Unable to run the flow. Check that the credentials .json file includes all required credentials.Open the flow in Tableau Prep Builder to view error details.”(流程中存在错误。无法运行流程。请检查 credentials.json 文件是否包括所有必需的凭据。在 Tableau Prep Builder 中打开流程以查看错误详细信息。)	输入连接的 credentials.json 文件中缺少凭据，或者流程包含错误。	请检查 .json 文件是否具有所有连接的凭据，并在 Tableau Prep Builder 中打开流程文件，确定流程中是否存在任何错误。 如果流程有错误，

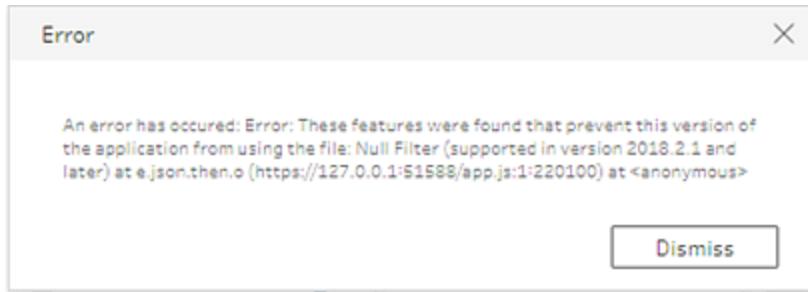
		您必须修复这些错误并将流程重新发布到 Tableau Server, 然后再次尝试运行该过程。
“Could not find match for <hostname of inputConnections>”(找不到 <输入连接的主机名> 的匹配项)	credentials.json 文件缺少主机名(服务器名称)条目。	请确保 credentials.json 文件包括主机名(服务器名称)的正确凭据。 有关详细信息和示例, 请参见 通过命令行刷新流程输出文件 在本页 321 。
“We don't have credentials of all connections in tfl/tflx file. The following connection(s) were not found: <hostname of inputConnections>”(tfl/tflx 文件中没有所有连接的凭据。找不到以下连接:<输入连接的主机名>)	credentials.json 文件缺失, 或包含的凭据对于错误消息中显示的主机名(服务器名称)不正确。	请确保 credentials.json 文件包括错误消息中所列主机名(服务器名称)的正确凭据。 有关详细信息和示例, 请参见 通过命令行刷新流程输出文件 在本页 321 。
“Error signing in server <serverUrl> as a user <userName>. Please check the credentials.”(以用户 <用户名> 身份登录到服务器 <服务器 URL> 时出错。请检查凭据。)	credentials.json 文件包含的凭据对于 Tableau Server 不正确。	请确保 credentials.json 文件包括适用于输出连接的所有正确凭据和元素。 有关详细信息和示

		例, 请参见 通过命令行刷新流程输出文件 在本页321。
“Could not sign in successfully as <userName> to server <serverUrl>(<contentUrl>)”(无法以 <用户名> 身份成功登录到服务器 <服务器 URL>(<内容 URL>))	credentials.json 文件包含的凭据对于 Tableau Server 不正确。	请确保 credentials.json 文件包括适用于输出连接的所有正确凭据和元素。 有关详细信息和示例, 请参见 通过命令行刷新流程输出文件 在本页321。
“We don't have credentials for Tableau Server to publish extract for one or more output nodes in tfl/tflx file.”(tfl/tflx 文件中没有用于为一个或多个输出节点发布数据提取的 Tableau Server 凭据。)	credentials.json 文件未作为命令行参数传入, 或者其缺少输出连接的凭据。	请确保在命令行中包括 credentials.json 文件的路径, 并验证 credentials.json 文件是否包括适用于输出连接的所有正确凭据和元素。 有关详细信息和示例, 请参见 通过命令行刷新流程输出文件 在本页321。
“Loom rest api server not started”(Loom rest api 服务器未启动)	安装或环境设置不正确。	请确保正确安装了 Tableau Prep Builder, 并且您以管理员身份运行命令。

		有关如何安装 Tableau Prep Builder 的信息，请参见 通过用户界面安装 Tableau Desktop 或 Tableau Prep Builder 。
“Error.Flow file does not exist.”(错误。流程文件不存在。)	流程文件的路径不正确。	请确保在命令行中包括流程文件的正确路径。
“Error.Connections file does not exist.”(错误。连接文件不存在。)	credentials.json 文件的路径不正确。	请确保在命令行中包括 credentials.json 文件的正确路径。
“Could not find match for <mapr01:5181>,<mapr02:5181>,<mapr03:5181>”(找不到 <mapr01:5181>、<mapr02:5181>、<mapr03:5181> 的匹配项)	使用 ZooKeeper 连接到 Apache Drill 时，您必须指定具体端口 ID。	在命令行中包括为输入凭据指定“端口”31010 的 credentials.json 文件。

错误：“These features were found that prevent this version of the application from using this file”(发现了使此版本的应用程序无法使用此文件的这些功能)

如果在较早版本的 Tableau Prep Builder 中打开在版本 2018.2.1 或更高版本中创建的流程，您可能会看到以下错误：



包括较早版本中不支持的功能的流程将导致此兼容性错误。为了解决错误，请在更高版本中打开流程，并保存没有所指明功能的流程的副本。在上面的示例中，请从应用了 null 筛选器的字段中将该筛选器移除。

然后在较早版本的 Tableau Prep Builder 中打开已移除该功能的副本。

发布 Tableau Prep 输出时出现“Failed to parse response from Tableau Server”(未能解析来自 Tableau Server 的响应) 错误

为了成功将输出从 Tableau Prep Builder 发布到 Tableau Server，必须在 Tableau Server 上启用 REST API。如果未启用 REST API，您将看到以下错误：

```
Failed to parse response from Tableau server due
to:javax.xml.bind.UnmarshalException - with linked exception:
[org.xml.sax.SAXParseException; lineNumber: 1; columnNumber:
10; DOCTYPE is disallowed when the feature
"http://apache.org/xml/features/disallow-doctype-decl" set to
true.]
```

有关在 Tableau Server 上启用 REST API 的信息，请参见 REST API 帮助中的 **REST API 要求**。

有关从 Tableau Prep Builder 中发布输出的信息，请参见 [创建和发布数据提取及数据源 在本页 310](#)。

使用 Tableau Prep 登录到启用 SSL 的 Tableau Server 时出现“**You are using Server version: null...**”(您正使用 Server 版本: null...) 错误

从 Tableau Prep Builder 中登录到启用 SSL 的 Tableau Server 时，您必须在安装了 Tableau Prep Builder 的计算机上安装根证书。如果未安装证书，您可能会看到以下错误：

You are using Server version: null but the minimum compatible version is: 10.0. Please upgrade to a compatible version(您正使用 Server 版本: null, 但兼容的最低版本为: 10.0。请升级到兼容版本)

如果看到此错误，请与 IT 部门或系统管理员合作，在安装了 Tableau Prep Builder 的计算机上安装所需的根证书。有关详细信息，请参见 Tableau Desktop 和 Tableau Prep Builder 部署指南中的[系统要求](#)。

维护 Tableau Desktop 和 Tableau Prep 的许可证

重要信息: Tableau 许可服务已于 2018 年 10 月 6 日转移到新的数据中心。这意味着，需要特殊配置(例如静态 IP 安全列表)来访问 licensing.tableau.com 或 licensing.tableau.com 的任何环境将需要更新，然后您才能激活、刷新或停用 Tableau 产品密钥。有关详细信息，请参见[Tableau 社区](#)。

Tableau Desktop 和 Tableau Prep 可依据短期许可证模型获得许可。短期许可证(也称为订阅许可证)允许您在一段指定的时间内使用和更新 Tableau Desktop 和 Tableau Prep。

必须续订短期许可证并刷新产品密钥才能提供不中断的服务。您可以在每个指定的期间过期时连续续订短期许可证。如果未续订短期许可证并且期限过期，Tableau 将停止工作，您将不再能访问软件。有关续订许可证的详细信息，请参见[如何预订 Tableau 许可证](#)。

注意: Tableau Desktop 或 Tableau Prep 的试用许可证会在一段设定的时间(通常为 14 天)后过期。试用期过期后，您将需要[购买许可证](#)才能继续使用产品。

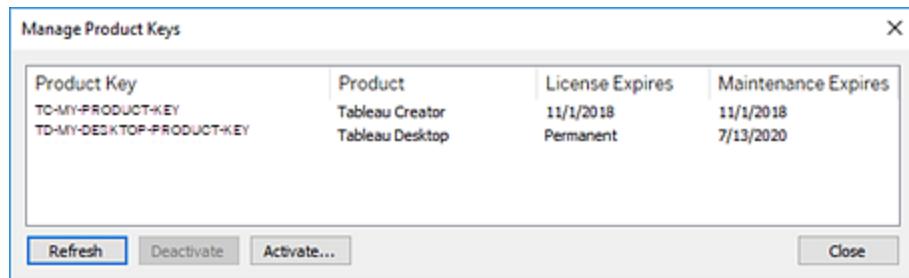
查看有关许可证的数据

安装 Tableau Desktop 或 Tableau Prep 之后，打开应用程序，然后从顶部菜单中导航到“帮助”>“管理产品密钥”，以查看有关您拥有的许可证类型及其何时过期的信息。

您也可以通过此对话框激活或停用产品密钥，或者刷新维护产品密钥。

注意：Tableau 提供具备一定范围能力的短期许可证。您拥有的许可证的类型显示在“产品”字段中。有关不同类型基于用户的可用许可证的详细信息，请参见 Tableau Server 帮助中的[基于用户的许可证](#)。

现有 Tableau Desktop 用户可能拥有永久许可证。永久许可证不会过期。但是，若要访问产品更新和技术支持，您必须购买支持和维护服务。这些服务必须续订才能继续获得服务。不再为 Tableau Desktop 提供永久许可证。



使用以下按钮来对产品密钥执行操作：

- 刷新(仅限 Tableau Desktop)：**单击“刷新”按钮以刷新将要过期的维护许可证，然后关闭并重新启动 Tableau Desktop。如果“维护过期”日期未更新，请咨询许可证管理员，因为密钥或维护协议可能已更改。

若要通过命令行刷新维护密钥，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[刷新产品密钥](#)。

注意：如果 Tableau Desktop 处于脱机状态，则无法刷新产品密钥。如果在脱机模式下激活 Tableau Desktop，则必须从 Tableau 客户门户获取并激活新密钥。

- 停用：**在列表中选择一个产品密钥，然后单击“停用”以停用产品密钥。如果需要将产品密钥转移到另一台计算机或者此计算机上不再需要产品密钥，请将产品密钥停用。

有关停用产品密钥的详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的转移或停用产品密钥。

- **激活:** 安装 Tableau Desktop 或 Tableau Prep 之后，单击“激活”打开激活对话框并输入您的产品密钥。如果遇到错误，并且无法使用产品密钥激活 Tableau Desktop 或 Tableau Prep，请与 [Tableau 支持部门](#) 联系。

有关激活产品密钥的详细信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的激活和注册产品。

跟踪 Tableau Desktop 许可证使用情况和过期数据

如果在 Tableau Server 中要跟踪和查看 Tableau Desktop 的许可证使用情况和过期数据，您必须将 Tableau Desktop 配置为按设定的间隔将许可证数据发送到 Tableau Server，然后在 Tableau Server 上启用报告。

此样，服务器管理员将能够访问两个报告：

- **Desktop 许可证使用情况:** 此报告让服务器管理员能够看到组织中 Tableau Desktop 许可证的使用情况数据。
- **Desktop 许可证过期:** 此报告为服务器管理员提供有关组织中的哪些 Tableau Desktop 许可证已过期或需要维护续订的信息。

如果针对许可证报告配置了 Tableau Desktop 和 Tableau Server，在以管理员身份登录到 Tableau Server 时，您将看到列在“分析”部分的“服务器状态”页面上的这两个报告。

The screenshot shows the 'Server Status' section of the Tableau Server interface. On the left, there's a sidebar with icons for All Sites, Sites, Users, Schedules, Tasks, Server Status (which is selected), and Settings. The main content area is titled 'Server Status' and has a heading 'Analysis'. It lists several reports with their descriptions:

Report	Description
Dashboard	Analysis
Traffic to Views	Usage and users for published views.
Traffic to Data Sources	Usage and users for published data sources.
Actions by All Users	Actions for all users.
Actions by Specific User	Actions for a specific user, including items used.
Actions by Recent Users	Recent actions by users, including last action time and idle time.
Background Tasks for Extracts	Completed and pending task details for extract refresh.
Background Tasks for Non Extracts	Completed and pending background task details for non-extract refresh.
Flow Performance History	View flow run performance.
Stats for Load Times	View load times and performance history.
Stats for Space Usage	Space used by published content, including extracts and live connections.
Login-based License Usage	Usage summary for login-based licenses
Background Task Delays	Difference between scheduled and actual start times of background tasks.
Performance of Views	Overall distribution of view load times and slowest views in a given time period.
Server Disk Space	Current and historical disk space usage, by server node.
Tableau Desktop License Usage	Summary of usage for Tableau Desktop licenses.
Tableau Desktop License Expiration	Expiration information for Tableau Desktop licenses.

如果未看到列出的这些报告，则可能未针对 Tableau Desktop 使用情况报告配置 Tableau Desktop 和 Tableau Server。

有关如何针对使用情况报告配置 Tableau Desktop 和 Tableau Server 的信息，请参见 Tableau Desktop 和 Tableau Prep 部署指南中的[管理 Tableau Desktop 许可证使用情况](#)。

其他资源

有关管理许可证的详细信息，请参阅以下主题：

- 若要查找产品密钥并激活 Tableau Desktop 或 Tableau Prep Builder，请参见[我的产品密钥在何处](#)。
- 若要停用产品密钥或将其转移到另一台计算机，请参见[转移或停用 Tableau Desktop](#)。
- 若要详细了解非持久性虚拟桌面或定期重新映像的计算机的产品密钥，请参见[配置虚拟桌面支持](#)。
- 若要详细了解 Tableau Server 或 Tableau Online 的产品密钥管理，请参见“[许可概述 \(Linux | Windows\)](#)”。
- 若要详细了解许可证续订过程或续订许可证，请参见[如何续订 Tableau 许可证](#)。