

Smartbi V10-数据准备

注意：（新特性列表中：+表示新增；^表示增强）

- +【数据准备】新增数据模型，将所有查询结果归集并基于CUBE重新构建数据结构
- +【数据准备】自助ETL新增Python脚本节点
- +【数据准备】新增作业流功能，实现多ETL实验之间执行依赖关系
- +【数据准备】新增多表JOIN节点，支持同时接入多个数据源并设置表关系
- +【数据准备】新增组合查询节点，支持去重、聚合、计新建算字段等高级功能
- ^【数据准备】数据模型新建或编辑计算度量、成员、命名集支持参数设置
- <【数据准备】取消支持Infobright作为高速缓存库

V10版本数据准备的改进点如下：

新增	增强	变更
+【数据准备】新增数据模型，将所有查询结果归集并基于CUBE重新构建数据结构 +【数据准备】自助ETL新增Python脚本节点 +【数据准备】新增作业流功能，实现多ETL实验之间执行依赖关系 +【数据准备】新增多表JOIN节点，支持同时接入多个数据源并设置表关系 +【数据准备】新增组合查询节点，支持去重、聚合、计新建算字段等高级功能	^【数据准备】新建或编辑计算度量、成员、命名集支持参数设置	<【数据准备】取消支持Infobright作为高速缓存库

+【数据准备】新增数据模型，将所有查询结果归集并基于CUBE重新构建数据结构

背景介绍

随着市场竞争的日趋激烈，企业的决策更加强调及时性和准确性，越来越多的用户需要多维度、更加灵活的方式观察数据变化，以获得对数据更加深入的了解，从而更好的促进企业的发展。为了满足用户的需求，新版本产品新增数据模型，可实现将所有查询结果归集后，基于CUBE模型重新构建数据结构。CUBE模型主要以“维度”和“度量”进行构建，同时增加了“成员”和“命名集”的创建，实现了数据模型构建的灵活性及应用广泛性。

功能概述

新版本在 **数据准备>数据集** 中，新增“数据模型”类型。



包含了以下功能：

私有查询

数据模型可实现将所有数据集作为私有查询结果归集。

支持数据来源有：数据源表、数据文件、SQL查询、即席查询、脚本查询、存储过程查询和ETL高级查询。



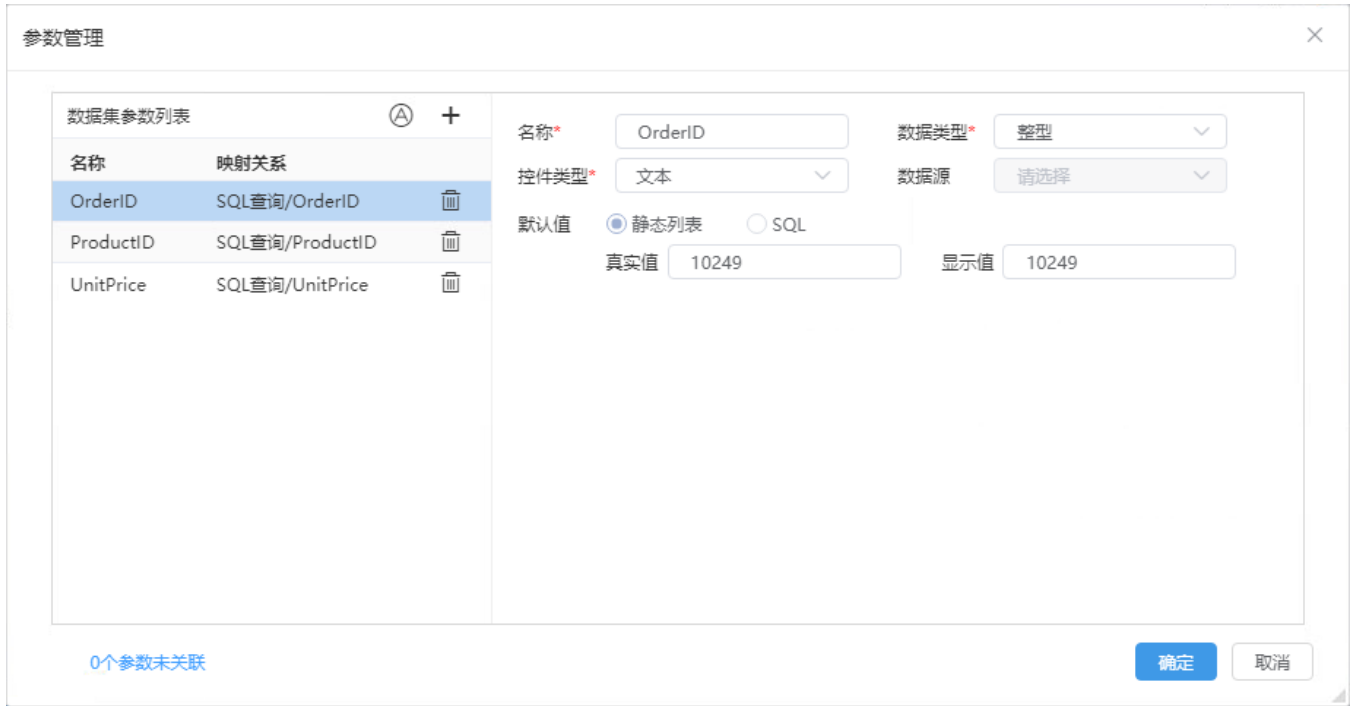
重构数据结构

数据模型基于CUBE模型重新构建了数据结构，CUBE模型主要以“维度”和“度量”进行构建，支持使用MDX表达式创建计算度量、计算成员、命名集。



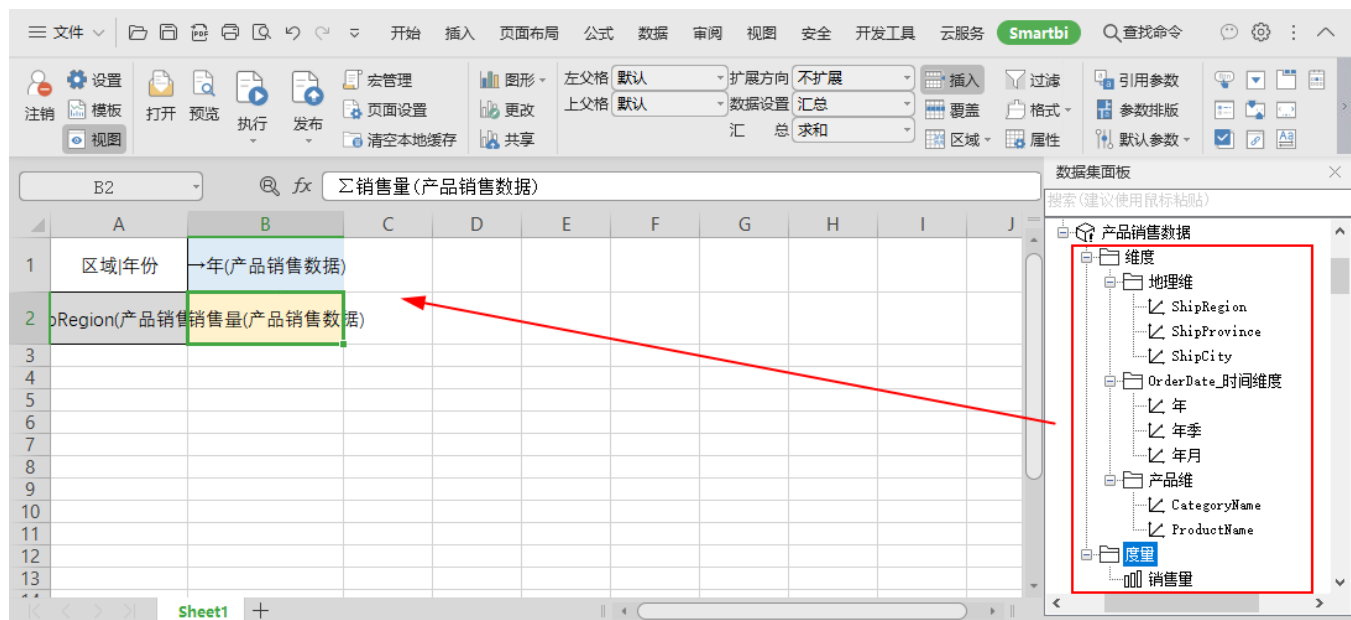
参数及参数管理

数据模型支持设置每个私有查询的参数，可对整个数据集包含的参数进行管理。



应用于报表

数据模型作为数据来源，可用于定制自助仪表盘和电子表格。



参考文档

关于数据模型的功能，详情请参考 [数据模型](#)。

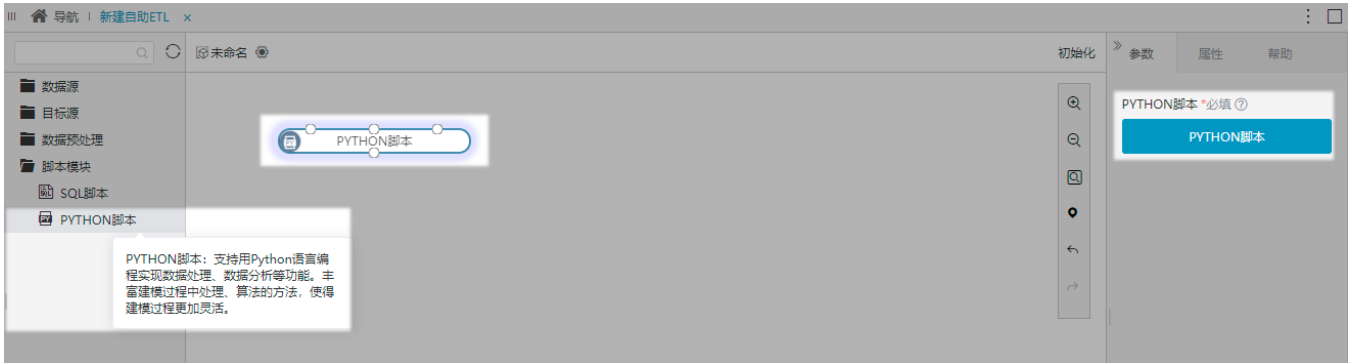
+【数据准备】自助ETL新增Python脚本节点

背景介绍

随着大数据时代的到来，Python已经在数据处理、数据可视化、机器学习等领域受到广泛应用。V10版本的ETL功能支持用户编写Python脚本，利用其丰富的科学计算扩展库，满足更多数据处理和分析场景，提高生产效率。

功能概述

Python脚本节点，支持用户编写Python代码，以便实现更自由、更高效的数据处理、数据分析操作，丰富了数据建模过程。



参考文档

关于Python脚本的功能，详情请参考 [Python脚本](#) 。

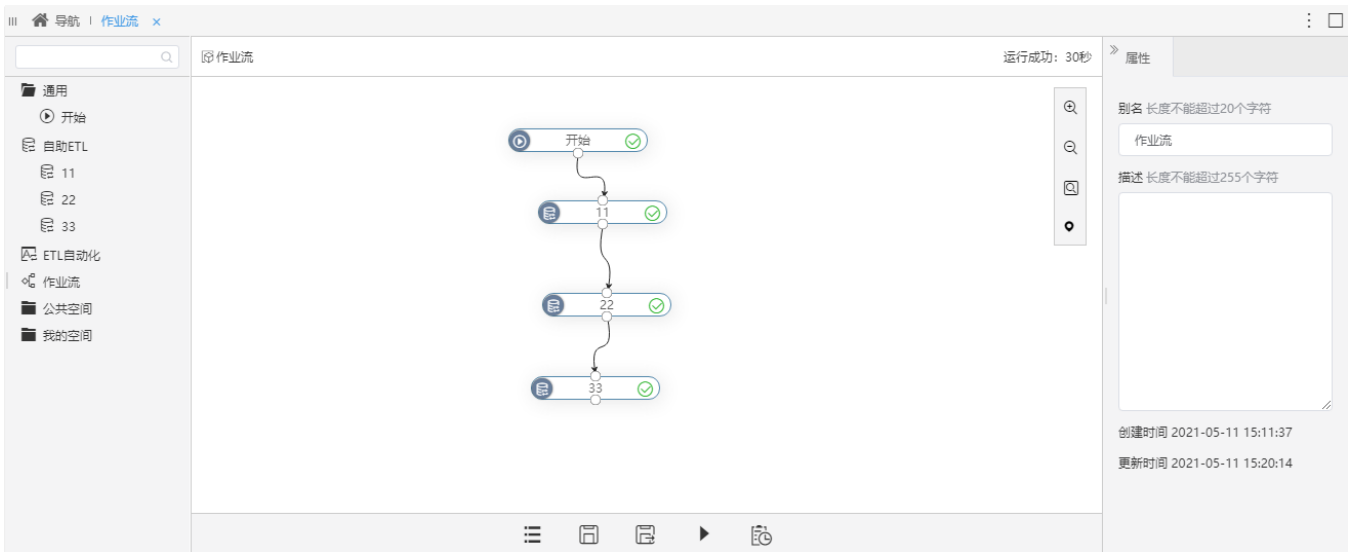
+【数据准备】新增作业流功能，实现多ETL实验之间执行依赖关系

背景介绍

在数据准备中，从原始数据到最终展现数据需要经过多个ETL实验处理，这些ETL实验的执行调度存在着依赖关系，也就是多个ETL实验需要按照指定的顺序和条件调度执行。新版本，产品在自助ETL中新增作业流功能，是一个轻量级的调度功能，多个ETL实验可按照指定的顺序和条件调度执行，实现了多ETL之间执行依赖关系。

功能简介

新版本，产品在自助ETL中新增作业流功能，可将自助ETL、作业流实验当做作业执行。



并支持设置分支条件，作业只有满足设置的条件才会继续执行。

编辑条件

依赖作业22

目标作业33

条件类型

作业参数

作业状态

条件

请选择参数

等于

值

请输入参数值

+

依赖作业	目标作业	条件类型	条件	操作
22	33	作业状态	成功	

系统运维人员可通过作业流监控，对作业流的运行情况进行监控。

系统

线程

性能

网络

日志

会话

缓存

垃圾回收活动

堆打印

对象池

类查找

扩展包

所有字体

wsdl地址

log4j信息

SQL监控

作业流监控

实验监控

服务监控

计算节点

系统补丁

作业流监控

最大运行作业流数: 8

等待队列长度: 0

最大等待作业流数: 10

查看日志

在线作业流

作业流名称	开始时间	持续时间 (秒)	操作
作业流	2021-05-11 15:11:51	5	节点列表

共 1 条 5条/页 < 1 > 前往 1 页

排队作业流

作业流名称	提交时间
暂无数据	

共 0 条 5条/页 < > 前往 1 页

历史作业流

参考文档

关于作业流的功能，详情请参考 [作业流](#) 。

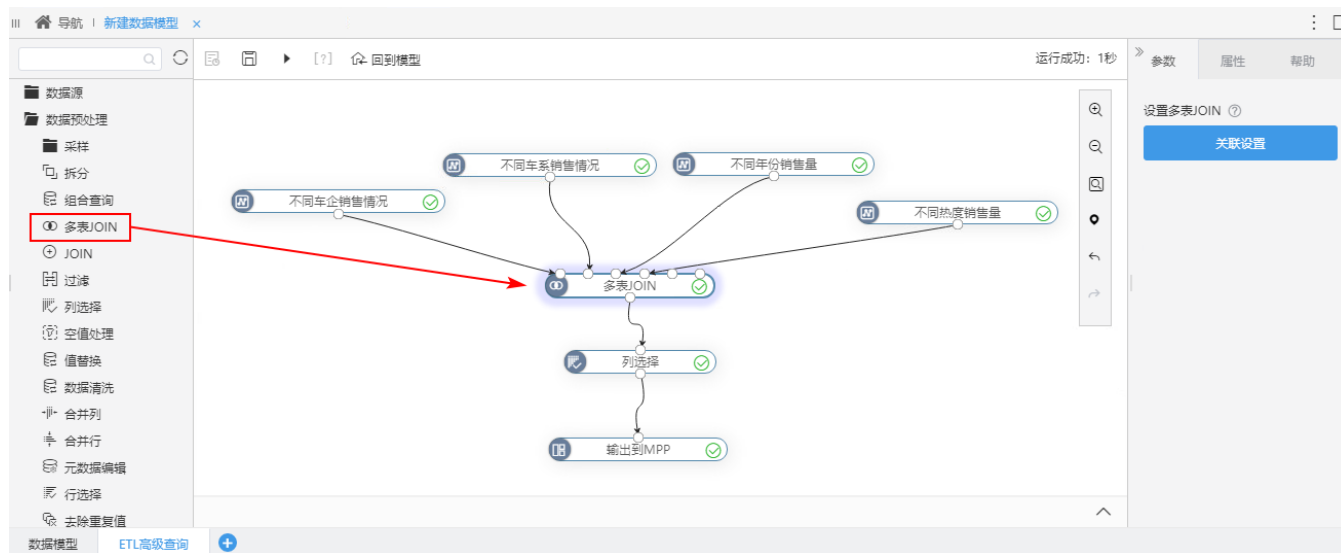
+【数据准备】新增多表JOIN节点，支持同时接入多个数据源并设置表关系

背景介绍

在实际应用中，有的用户业务复杂需要多张表的多个指标实现，使用节点数量较多，用户操作重复性高、灵活性较差。为了解决以上问题，新版本产品新增多表JOIN节点，支持同时接入多张数据源表并设置表关系，将多张表合并为一张表，减少产品与数据库的交互，提高产品性能。

功能简介

新版本，在数据模型、自助ETL、数据挖掘中，数据预处理下新增多表JOIN节点，支持同时接入多个数据集并设置表关系。



参考文档

关于多表JOIN节点功能，详情请参考 [数据挖掘-多表JOIN](#) 。

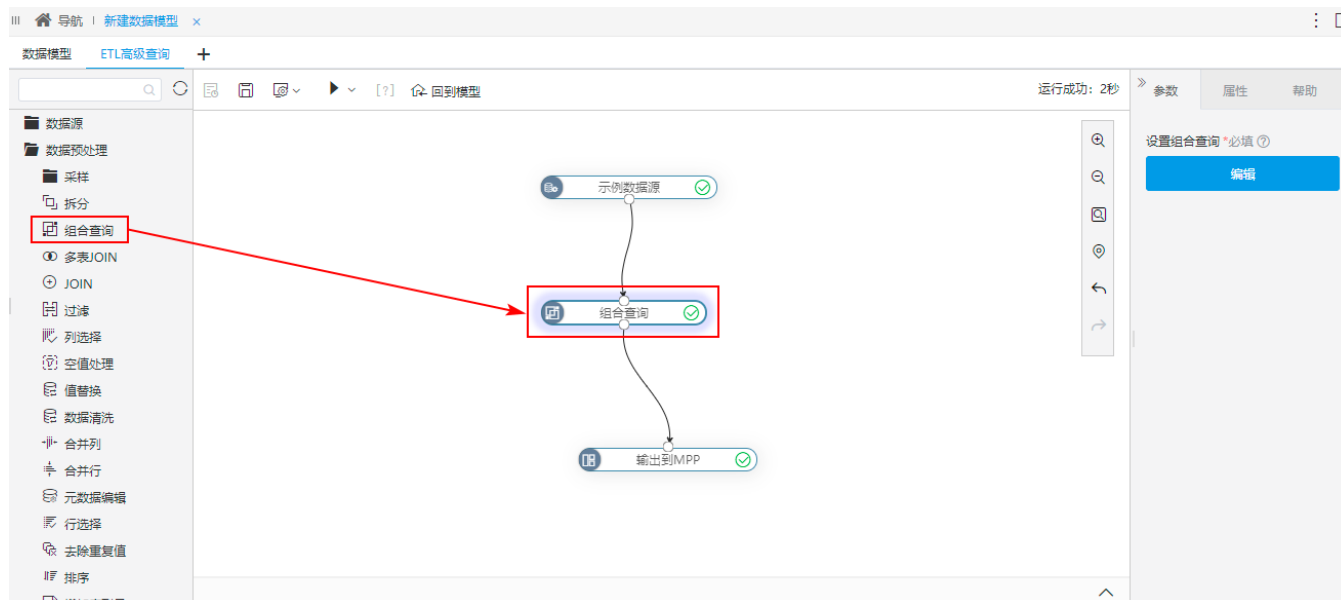
+【数据准备】新增组合查询节点，支持去重、聚合、计新建算字段等高级功能

背景介绍

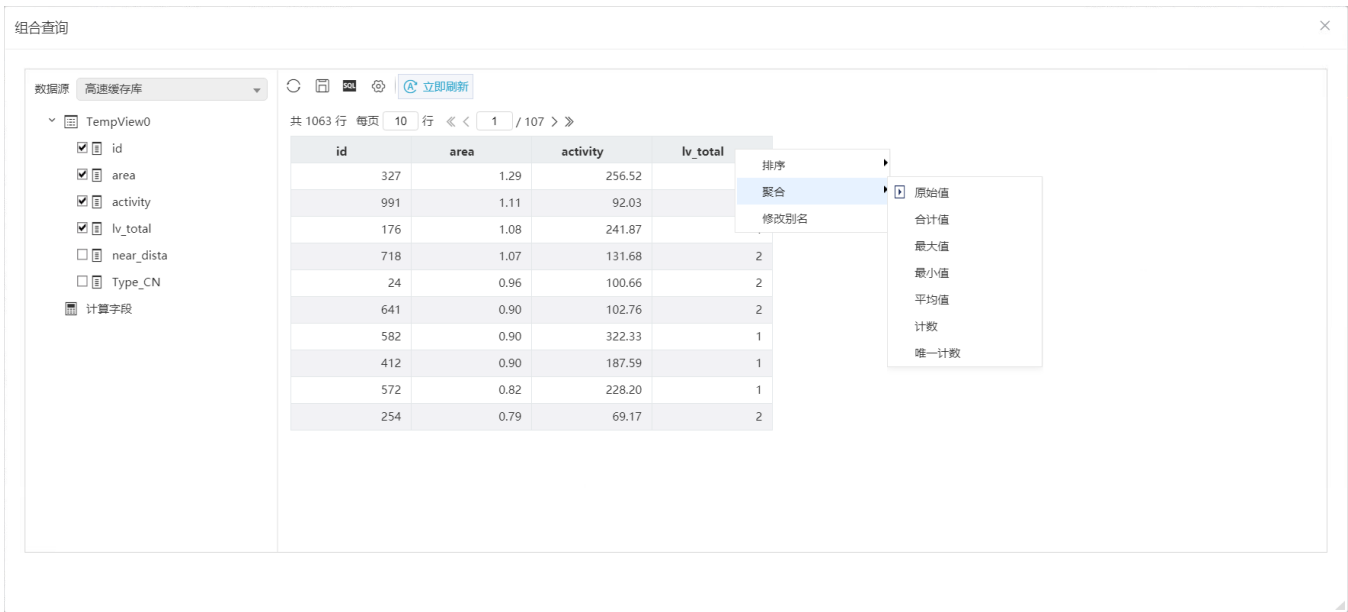
在实际场景中，有的用户业务复杂需要多张表的多个指标实现，且在数据处理时进行聚合、排序、条件筛选操作等，需要通过多个节点处理，操作既麻烦又复杂，也增加了实验运行的时间。为了解决以上问题，新版本产品新增组合查询节点，支持灵活选择查询的字段和去重、聚合、新建计算字段等高级功能，提升了数据处理效率与性能。

功能简介

新版本在自助ETL、ETL高级查询、数据挖掘中，新增组合查询节点。



支持灵活的选择字段和去重、条件筛选、聚合筛选、聚合、排序、修改别名、新建计算字段、批量选择列、取消选择列等高级功能。



注意事项

组合查询节点只支持使用ClickHouse21及以上版本作为高速缓存库。

参考文档

关于组合查询节点，详情请参考 [数据挖掘-组合查询](#) 。

^【数据准备】数据模型新建或编辑计算度量、成员、命名集支持参数设置

功能简介

为了满足用户更多需求，进一步扩大参数的应用范围，新版本支持用户在数据模型中新建或编辑度量、成员、命名集时设置参数，可实现动态获取结果。

参数管理

数据模型参数列表

参数名称

映射关系

销售区域

即席查询/ShipRegion

名称*

销售区域

数据类型*

字符串

控件类型*

下拉 (单选)

数据源*

northwind

默认值*

静态列表

SQL

真实值*

华北

显示值

华北

备选值*

静态列表

SQL

1

select distinct ShipRegion from orders

真实值 *

ShipRegion

显示值

ShipRegion

0个参数未关联

确定

取消

注意事项

由于清单表不支持度量，因此不支持新建命名集。

<【数据准备】取消支持Infobright作为高速缓存库

功能简介

由于Infobright性能较差，不适合作为用户生产环境的缓存库，因此新版本产品取消支持Infobright作为高速缓存库。