行过滤节点使用说明

作者：张青松

目录

[1. 行过滤节点 2](#_Toc426032003)

[2. 设置行过滤参数 2](#_Toc426032004)

[2.1. 基础过滤功能 3](#_Toc426032005)

[2.1.1. 标称量过滤 3](#_Toc426032006)

[2.1.2. 数值类型过滤 4](#_Toc426032007)

[2.1.3. 日期类型过滤 5](#_Toc426032008)

[2.2. 高级过滤功能 6](#_Toc426032009)

[2.3. 查看过滤结果 7](#_Toc426032010)

# 行过滤节点

行过滤节点是一个数据预处理工具，位于数据集目录下。使用行过滤节点，根据指定的过滤规则，将输入数据中不符合规则的数据行去除，得到一个由所有符合规则的行组成的子数据集。

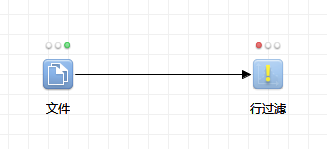
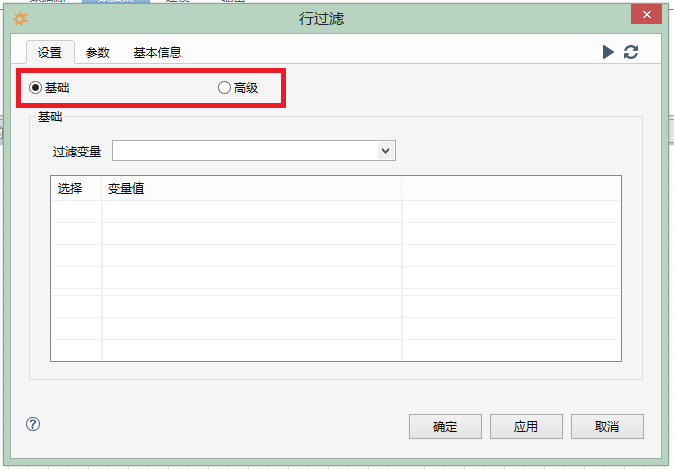


图1 行过滤工作流示例

与其他节点相同，行过滤节点需要有其他节点输入的数据，此处以文件读取节点为例。输入的数据中可以包含数值类型、字符串类型和日期类型等等。

# 设置行过滤参数

打开行过滤节点的设置窗口，根据过滤需求选择窗口上方的基础或者高级过滤设置，如图。

图2 行过滤设置窗口

## 基础过滤功能

选择基础过滤设置。在下方的过滤变量下拉框中，选择用于过滤条件的列名称。

### 标称量过滤

如果选择的列为标称量类型，将在下方的表格中列出该列的所有可能取值。并勾选其中的一部分。如果某一行的该列的值在被勾选的标称量中，则该行被保留，否则该行不会出现在过滤结果集中，如下图。

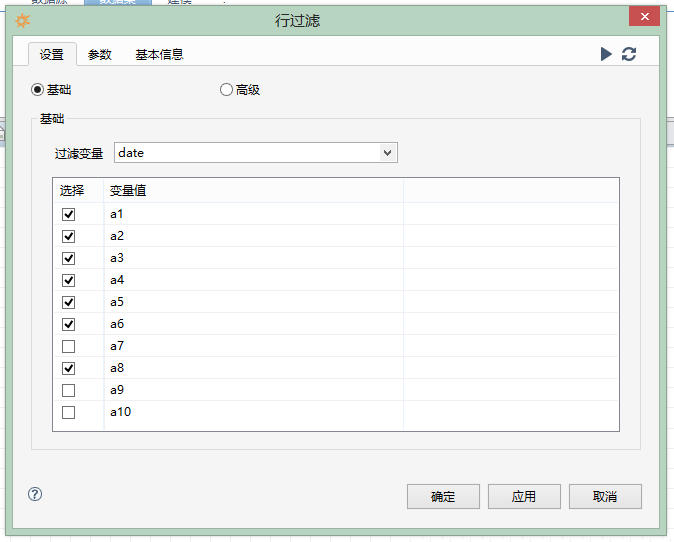


图3 标称量过滤

### 数值类型过滤

如果选择作为过滤条件的列为数值类型，在窗口中设置一个取值范围，只保留所有数据行中，该列的取值在改范围内的，如果选中“取互斥区间”，则相反。如下图。

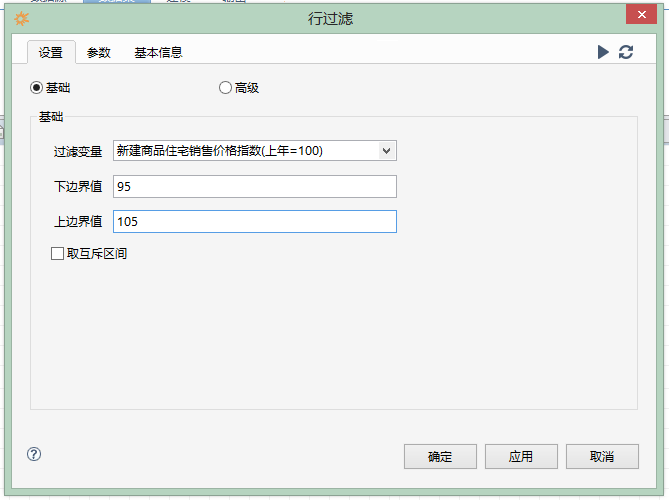


图4 数值类型过滤

### 日期类型过滤

如果选择的列为日期类型，操作与数值类型相似。如下图。

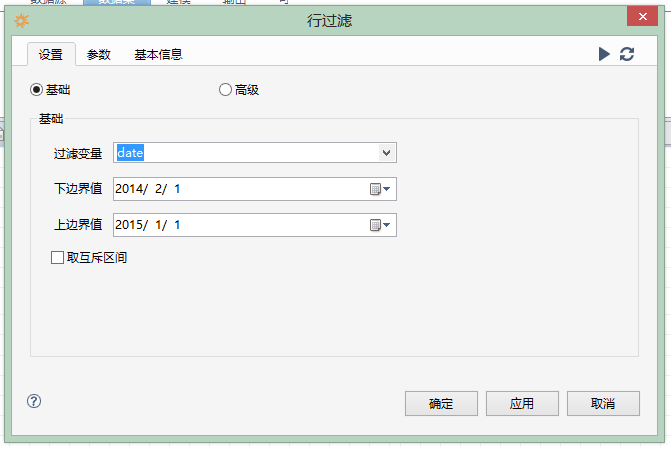


图5 日期类型行过滤设置

## 高级过滤功能

如果基础过滤功能不能满足过滤需求，选择使用高级过滤。如下图。

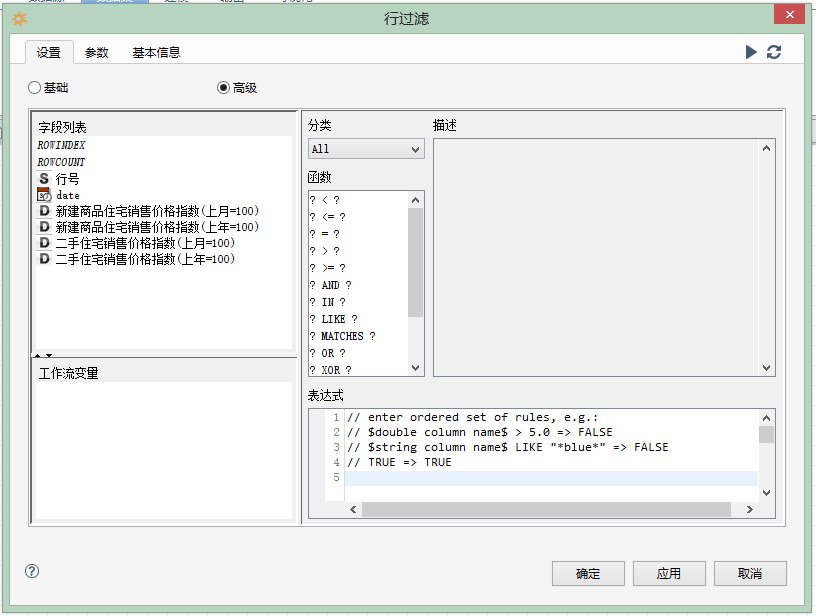


图6 高级过滤设置

根据需要过滤的条件，选择字段，工作流变量以及操作符组成过滤条件。形如：

* $行号$ >= 10 => TRUE
* $新建商品住宅销售价格指数(上年=100)$ >= 100 =>FALSE
* $新建商品住宅销售价格指数(上年=100)$ >= 100 AND $新建商品住宅销售价格指数(上月=100)$ >= 100 =>FALSE

如果同时出现多条形如

$double column name$ > 5.0 => FALSE

$二手住宅销售价格指数(上年=100)$ => TRUE

取各条过滤后的并集。

如果某一行出现红色下划线或者行头有红叉，表示有语法错误，需要修改。

## 查看过滤结果

设置完成后，点击右上角的执行预览，或者保存后执行该节点，并查看过滤后的数据表。