**分区节点使用说明**

**作者：王龙强**

[1.     节点设置参数页面 1](#_Toc431212471)

[2.    节点使用过程 3](#_Toc431212472)

[3.    注意事项 5](#_Toc431212473)

分区节点是 Data Studio 数据集节点分类中的一个数据统计节点，它拥有一个输入端口，两个输出端口，它可以将数据流中数据以指定的规则切分成两部分，比如，将数据流中前100行作为第一部分输出，其余的作为后一部分输出等。

分区节点支持各种各样的规则（包括正则表达式，数学表达式等），足以满足用户的各种要求。

# 1.     节点设置参数页面

如图1所示的分区节点基础设置界面设置项意义如下：

1. 数量：分区数据行数。
2. 百分比：分区数据行数（总行数的百分比）。
3. 从第一个开始计算：从数据流中第一行数据开始计算。
4. 线性抽样：以同等间隔取指定数量的数据。
5. 随机抽样：随机抽取指定数量的数据。
6. 分层抽样：选一非数值列，DataStudio会以该数值列分层取指定数量的数据。

如图2所示的分区节点高级设置界面设置项意义如下：

1. 属性值：根据属性值进行数据分区，首先需要选择过滤字段，然后可以根据模式匹配（精确匹配、模糊匹配、正则表达式匹配）， 数值范围，仅缺失值抽取指定数量的数据。
2. 数值范围：根据指定规则从数据流中选择满足指定条件的数据行。
3. RowID：根据RowID从数据流中选择满足指定条件的数据行。

如图3所示的分区节点规则设置界面设置项意义如下：

1. 表达式：用户可以根据字段列表、工作流变量、函数来自定义表达式，从数据流中选择满足指定条件的数据行。

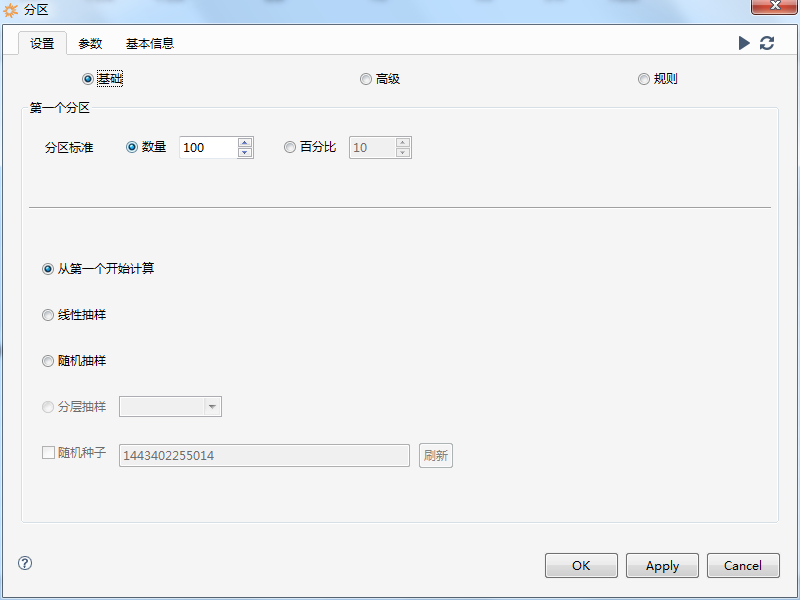


图1分区设置界面（基础）

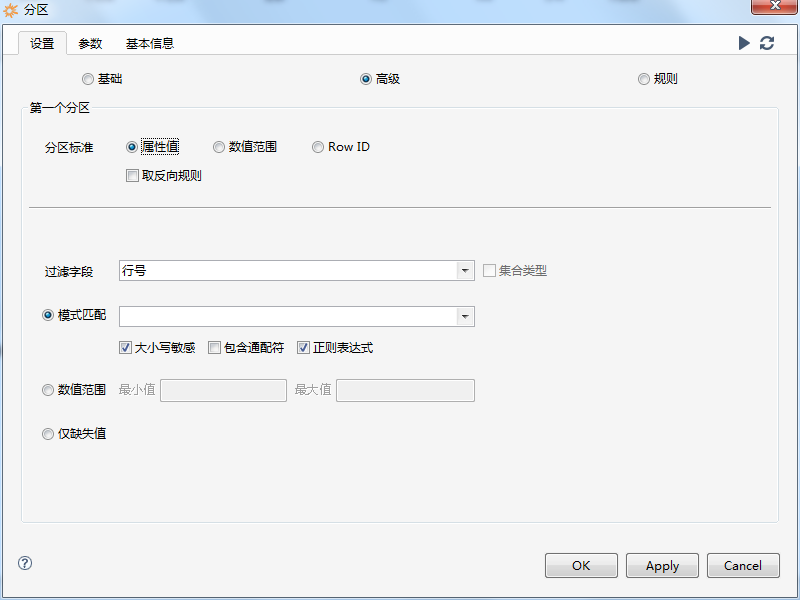


图2分区设置界面（高级）

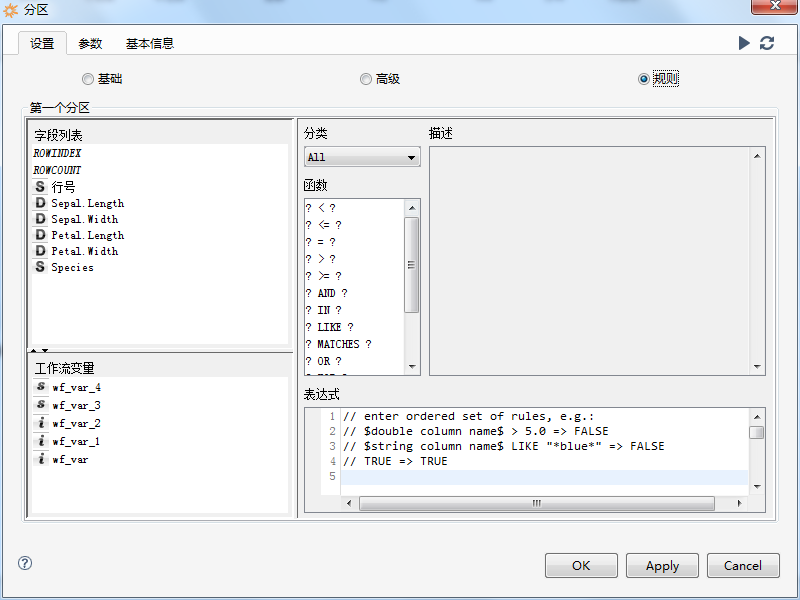
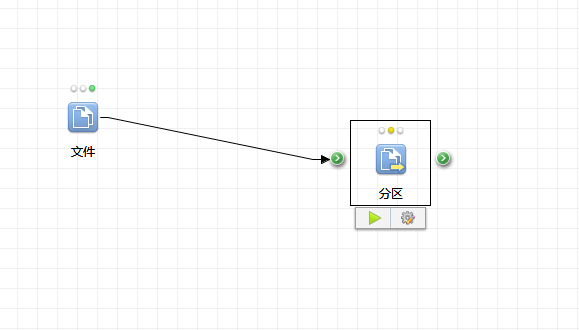


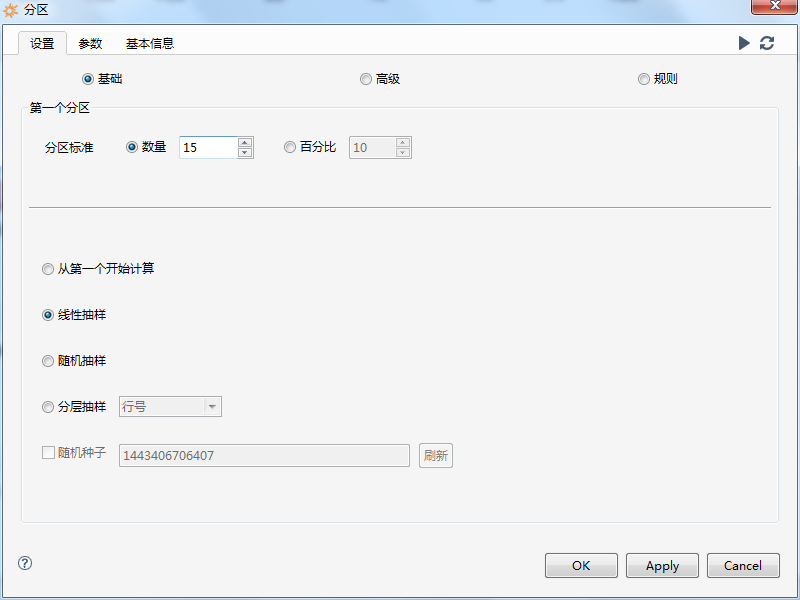
图3分区设置界面（规则）

# 2.    节点使用过程

以基础分区为例，如下图，连接数据输入

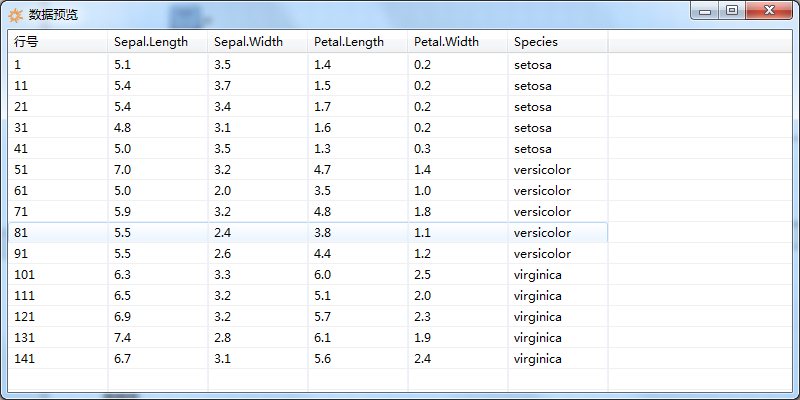


配置分区节点如下图：

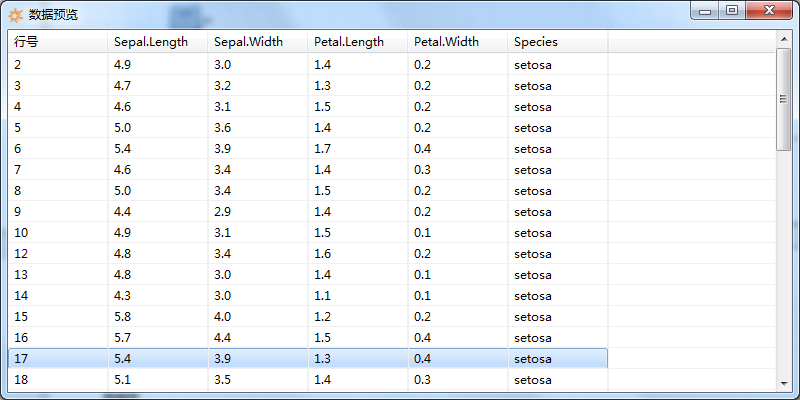


确定右键执行，执行结果如下图：

输出数据集一：



输出数据集二：



可以看出DataStudio根据指定规则，将数据集成功划分。

# 3.    注意事项

“高级”中有个取反向规则的复选框，顾名思义，即取相反的规则对数据集进行分区。