**分布节点使用说明**

**作者：王龙强**

[**1.   节点设置参数页面 1**](#_Toc425927970)

[**2.   节点使用过程 2**](#_Toc425927971)

[**3.   注意事项 4**](#_Toc425927972)

分布节点是 Data Studio 数据集节点分类中的一个数据统计节点，它可以数据流中各种数据统计信息，比如，数据流中中位数、平均值、最大值、最小值（当然流中字段需数值类型）等。还可以记录数据流中最多出现、最少出现的记录数（流中字段可以不是数值类型）。最多出现前N条记录是什么，分别出现了多少次，最少出现的前N条记录是什么，分别出现了多少次，以上这些内容都有统计。节点有一个输出，输出包含了所有统计信息

# 1.     节点设置参数页面

如图 1 所示的分布节点设置界面设置项意义如下：

1. 计算中位数：计算流中整型、数值类型字段的中位数，由于计算中位数时需要将大量数据放入内存，此选项要慎选，会影响效率
2. 显示出现次数最多最少的标称值数量：即在视图中每个字段显示前N个最多的记录数 ，前N个最少的记录数。
3. 输出列中包含的标称值最大数量：即输出的列中，每一列所能包含的最大标称值的数量，超出则会在视图展现统计信息时给出错误提示
4. 标称字段：需要进行统计的字段



图1分布设置界面

# 2.    节点使用过程

如上图，选中“计算中位数”，显示出现次数最多最少的标称值数量设置20，输出列中包含的标称值最大数量设置1000，标称字段全选，确定后右键执行。如图2，右键“预览：数据分布”



图2 预览:数据分布

查看数据统计信息如下：图3，图4，图5



图3 数值统计



图4 标称值统计

  

图5 最多/最少统计

在图2右键菜单中选择输出结果集，可以查看分布节点的输出数据，输出数据如下图6



图6 输出数据

# 3.    注意事项

1. 当数据记录数过多时不建议勾选“计算中位数”，有可能会