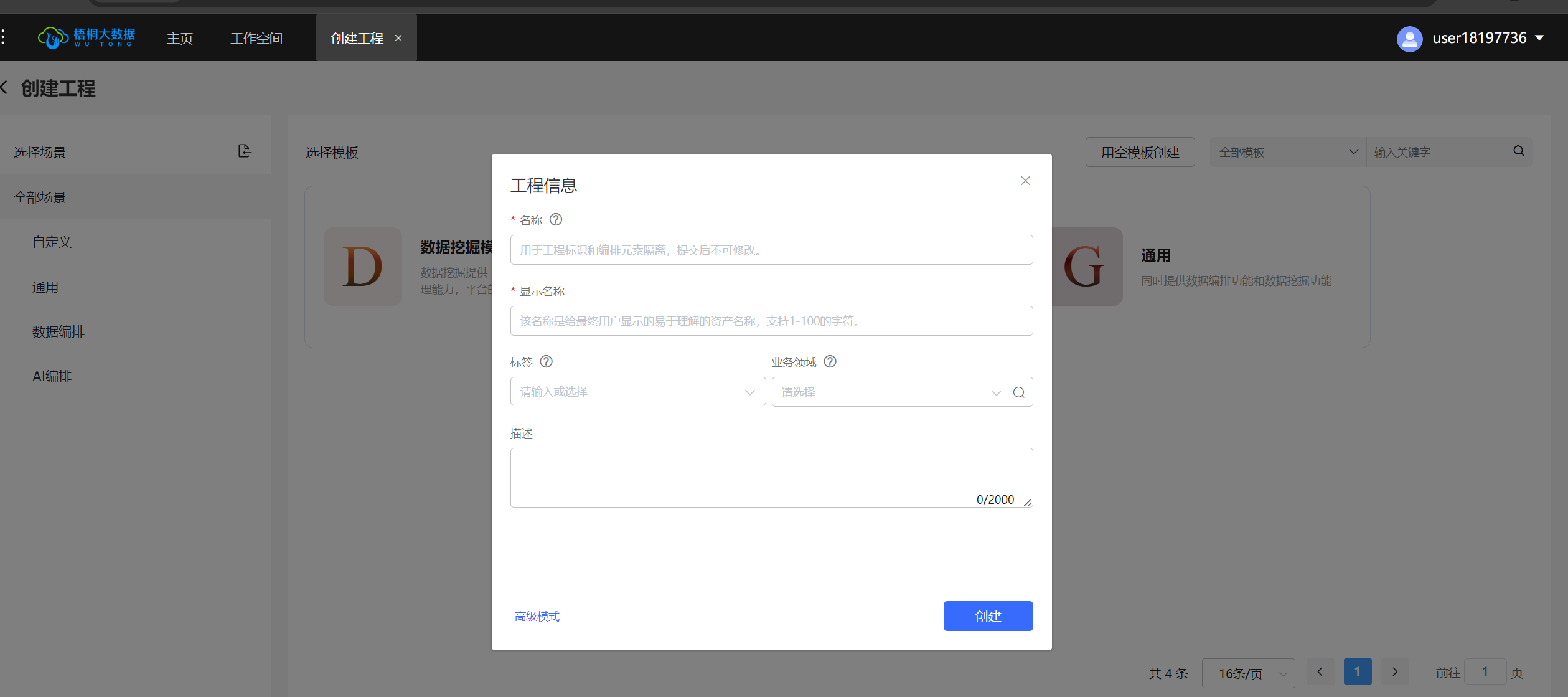
在移动·梧桐大数据平台——校企版——实验中心——大数据实操平台

工作空间——创建工程——通用模板



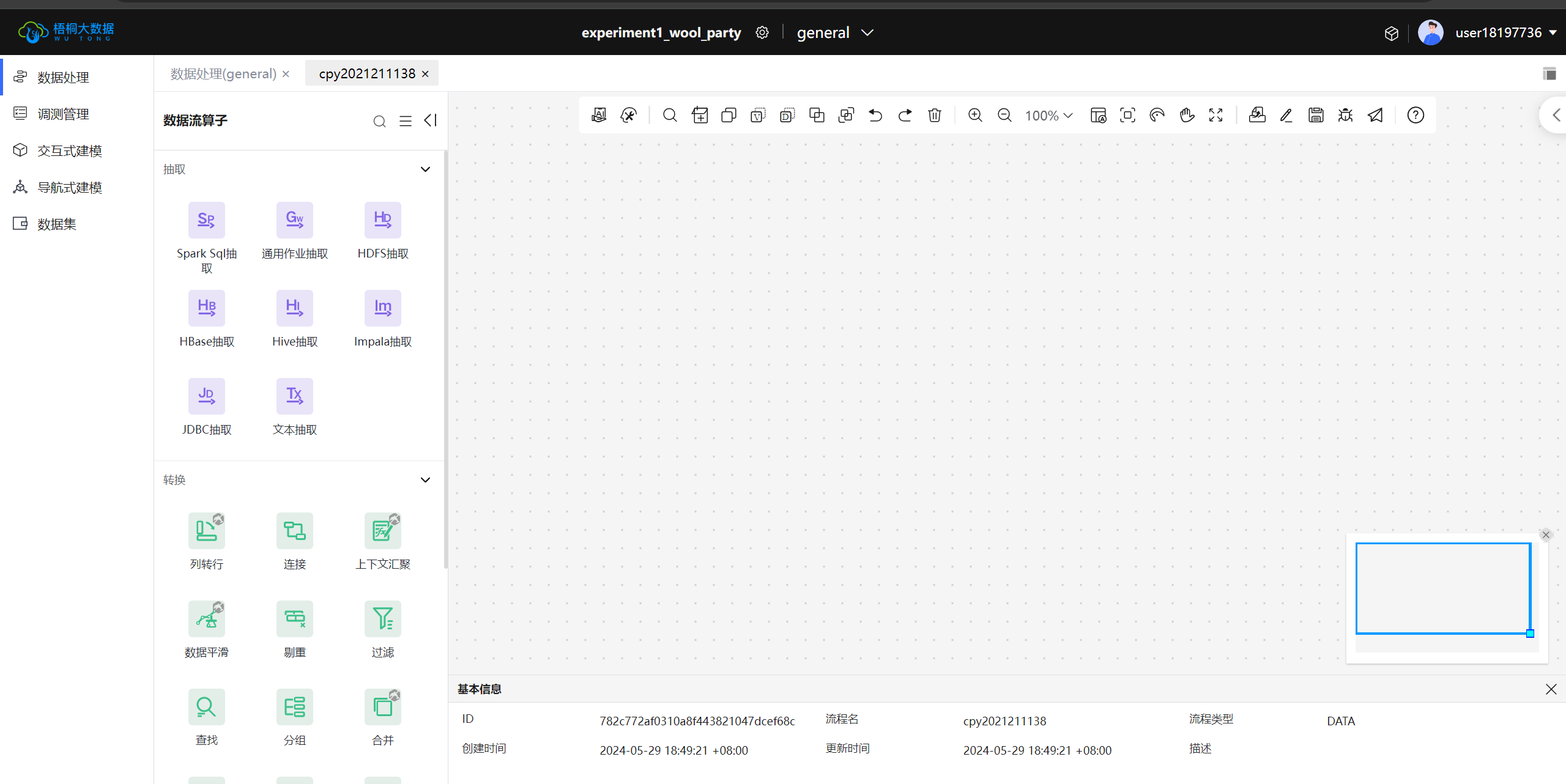
写上名称，点击创建



数据流程编排

可通过图形化界面按钮进行数据加工。打开步骤一创建的工程，在导航栏单击“数据处理”进入数据处理界面，点击“数据流”，然后在右侧点击“新建流程”，如下图所示，弹出新 建数据流对话框，输入名称[名称要求为姓名拼音+学号]即可完成数据流的创建。

创建完后给出清晰的数据流程截图，要有数据流程名称：



在数据流画布中进行算子的编排

编排包括3个阶段：

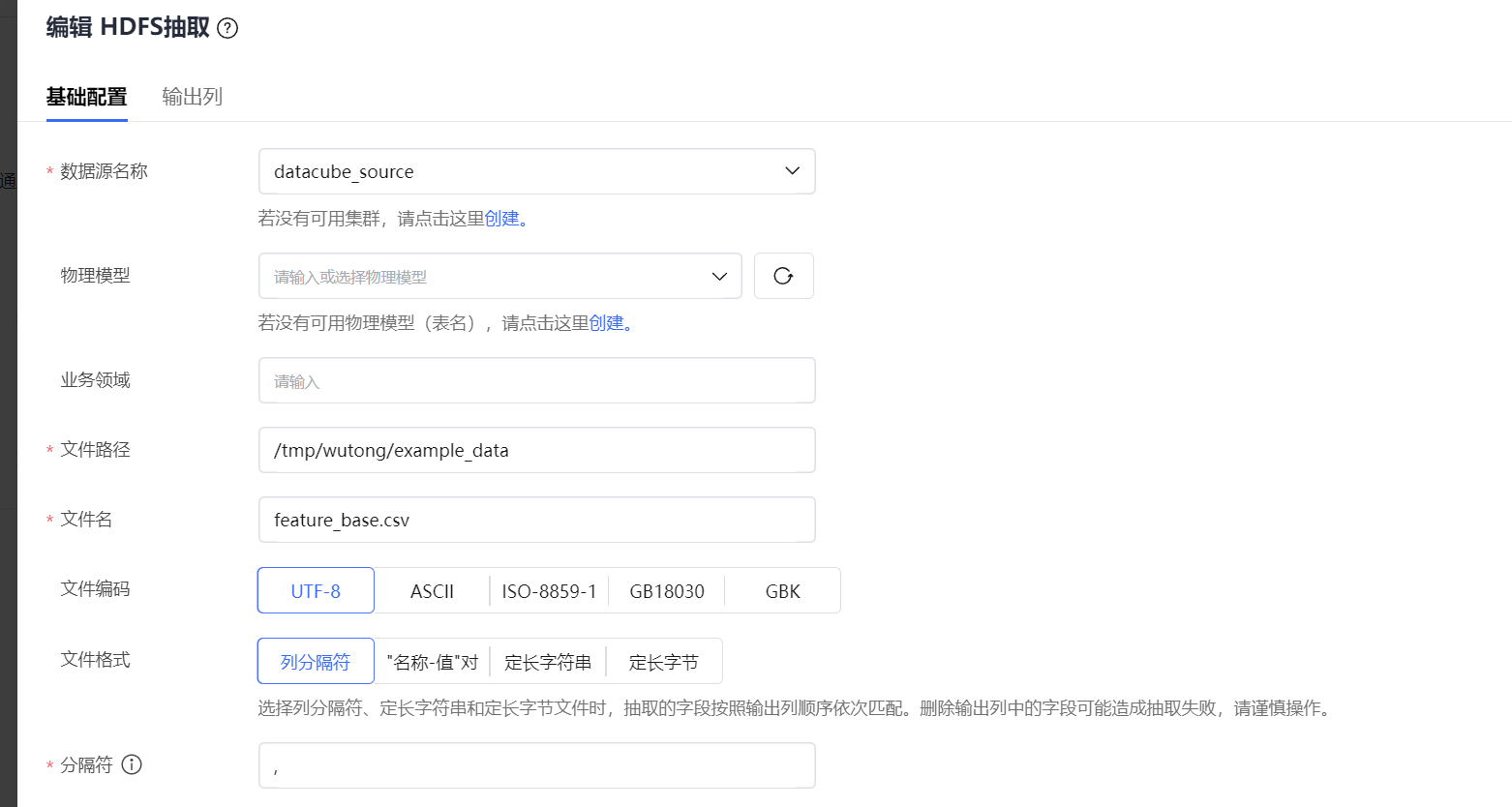
第一阶段是抽取数据，即从HDFS 中抽取本案例的数据到编排的 数据流中；

第二阶段进行数据的处理，即根据实际需要进行表关联或字段计算统计等；

第三 阶段是将处理完成后的数据加载成文件并存放到 HDFS 中。数据流整体算子编排如下图所 示。接下来对整个流程进行详细介绍。

第一阶段算子编排

1



并添加输出列

第一个有三行

第二个有十行

转换操作中phone和month字段保留原有字段输出，其余所有字段均使用 doubleconvert 函数进行转换，例如填写 chrg\_cnt 字段表达式为 doubleconvert(chrg\_cnt)，如 下图所示。