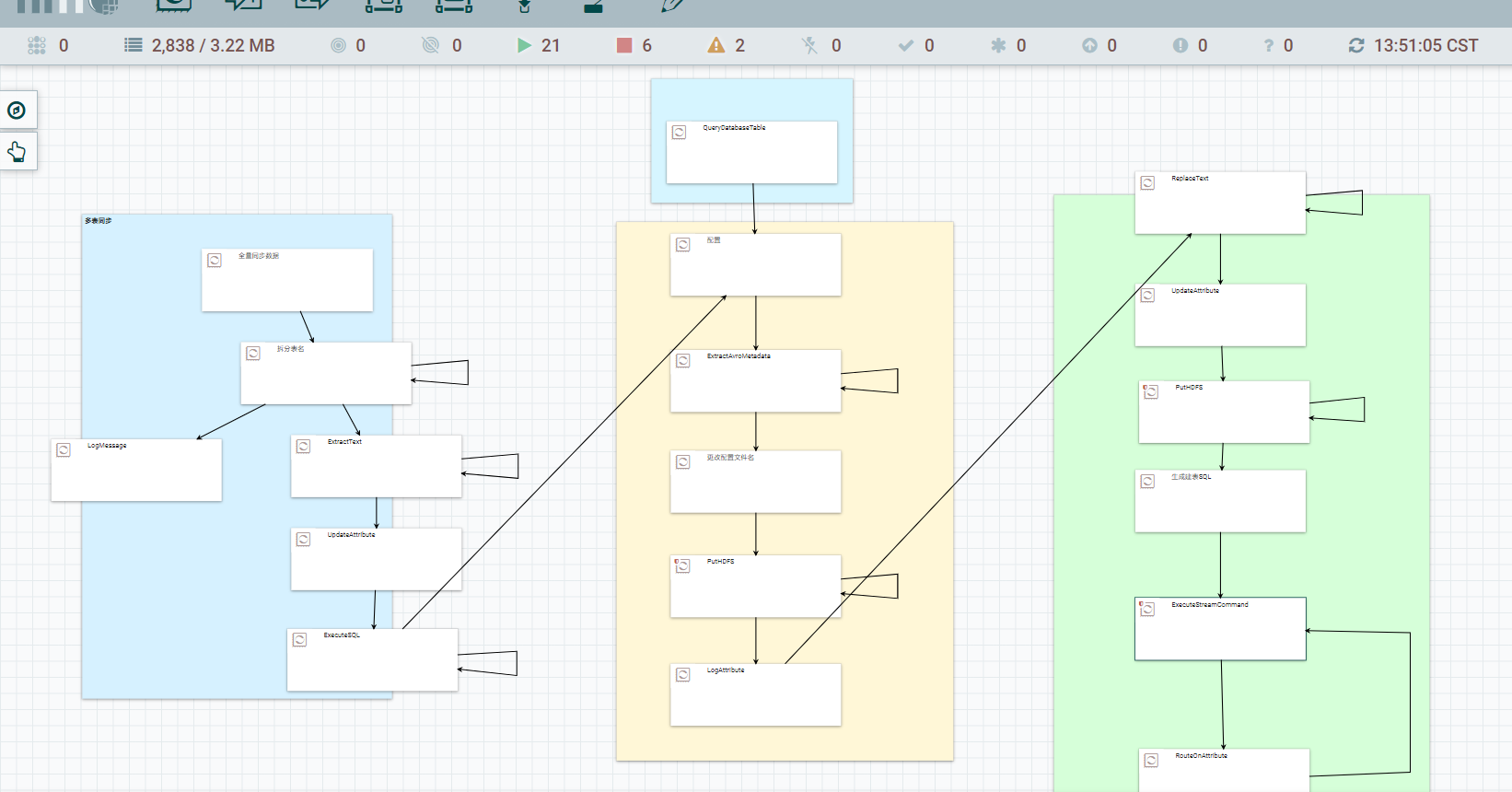
Nifi关系行数据库同步数据到hive

访问nifi web,可以看到打开的界面是一个用于编排数据流的空白画布，可以在空白画布中创建插入处理器组，在组中创建一个数据流，如数据从oracle同步到hive

1. 多表同步
2. 选择处理器，这里是选择GenerateFlowFile 处理器来维护所需要同步的表，
3. 选择文本拆分处理器SlitText来拆分表名，
4. 选择extractText处理器匹配文本内容将表名赋值给变量，
5. 选择updateAtrribute处理器来定义后续流程用到的变量，
6. 选择executeSQL处理器从oracle中开始读取数据
7. 选择updateAtrribute处理器来定义后续流程用到的变量
8. 选择ExtractAvroMetadata处理器提取Avro元数据信息，这里提取的是表字段信息
9. 选择 updateAtrribute处理器来更改配置文件名
10. 选择putHDFS处理器将数据存放到hdfs上
11. 选择replaceText修改元数据信息
12. 选择updateAtrribute处理器定义后续流程所用到的变量
13. 选择putHDFS将元数据厨房到hdfs上
14. 选择updateAtrribute处理器生成创建hive表的sql
15. 选择ExecuteStreamCommand 处理器来执行sql语句
16. 选择RouteOnAttribute处理器查看Execute是否执行成功



二、单表同步

1、选择QueryDatabaseTable处理器从oracle数据库中读取数据

2、选择updateAtrribute处理器来定义后续流程用到的变量

3、选择ExtractAvroMetadata处理器提取Avro元数据信息，这里提取的是表字段信息

4、选择 updateAtrribute处理器来更改配置文件名

5、选择putHDFS处理器将数据存放到hdfs上

6、选择replaceText修改元数据信息

7、选择updateAtrribute处理器定义后续流程所用到的变量

8、选择putHDFS将元数据厨房到hdfs上

9、选择updateAtrribute处理器生成创建hive表的sql

10、选择ExecuteStreamCommand 处理器来执行sql语句

11、选择RouteOnAttribute处理器查看Execute是否执行成功