任务总体要求

整体要求点：

1. 项目以Maven或者SBT构建
2. 语言可以使用scala或者Java
3. 任务1 以 **字段规则转换**、**SQL执行返回统计信息**功能优先级最高
4. 任务2以 **配置规则，不进行硬编码为原则**
5. 第二期任务会比较具体，需要以**两周为周期**，进行相关设计、原型的实现的**工作汇报**
6. 有任何问题可以及时在QQ群中反馈、交流，避免工作进展堵塞

任务1 - 基于Spark 和RESTFul 接口的SQL服务

**需求点：**

1. [总体]一个命令行终端，一个RestFul服务端
2. [命令行]终端可以连接到服务端
3. [命令行]终端可以提交sql语句
4. [命令行]终端可以对SQL语句返回的语句进行自动解析，打印结果及标题
5. [服务端]可以接收、处理客户端RestFul命令
6. [服务端]可以对命令进行队列管理，防止大规模命令同时执行影响集群效率
7. [服务端]可以根据配置建立多个JDBC连接，以增加集群的利用率
8. [功能性]需要完成字段规则转换功能
9. [功能性]需要完成SQL执行的信息统计返回

**字段规则转换功能描述：**

第一点：该功能主要的场景如下，如一个表TableA{name:String, id:String, prop1:String, prop2:String}，这个表第一个字段name为姓名，第二个字段ID为身份证，这样如果针对该表的查询如：Select \* from TableA，不能够将id显示，需要按照既定义规则进行转换，规则比方描述为只显示后四位，则关于id的显示则类似是 “\*\*\*\*\*\*\*\*\*1234” 这样的返回结果。

第二点：身份证字段可以在内部进行表关联的条件判断，所以内部存储就是原始的身份证字段，不能在内部存储中转换来实现；

第三点：可能用户进行表的join或者直接查询存储到另外一张表，所以需要针对这个情况进行识别处理。

1. 需要有定义转换字段的地方，有即可；
2. 需要在执行过程中进行相关转换；
3. 需要有删除规则的地方，有即可；
4. 规则需要灵活性，如定义正则表达式或者其他方式均可。

**SQL执行返回统计信息描述：**

进行sql执行的时候，需要对于hdfs上读写的大小，sql执行时间等进行统计，返回给客户端。

**大体架构可以参考但不限于：**



任务2 - 基于流式处理的规则转换

**需求点-1：**

1. 基于kafka、Spark Streaming开发基于规则的流式处理框架
2. 定义输入结构、输出结构，使用xml或者json等进行描述
3. 定义处理规则，可以针对某个字段，或者针对于某几个字段进行转换
4. 规则中的转换函数的规则定义好后具有通用性，如string2int，sha1等
5. 也可以让用户按照规则进行相关算子的编写
6. 以上均只需要配置，不需要硬编码

**相关结构可以参考但不限于：**



输入输出1对1时是上图，为了解决基于某些该时间段的统计，需要在spark Streaming中嵌套入SQL，产生如下需求点。

**需求点-2：**

1. 参考上图，针对于记录做统计是，希望基于sql
2. 输入输出定义可以使用一套，规则定义可以分类

