spark.sql.catalyst模块

1. 类型types：

1最抽象：AbstractDataType

2继承类1：DataType

3 继承类2：如DoubleType,StringType等

4 用户自定义UDT（自定义类型，如使用javaBean），继承类1

1. Antlr——词法语法解析器：<https://blog.csdn.net/duan_zhihua/article/details/74853103>

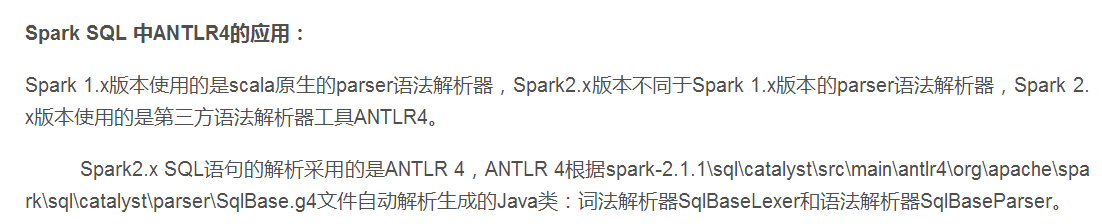
作用：将文本文件，解析成java语法或代码（也可转成其它语言）

spark2中的sql语句解析，采用Antlr4解析器（有对应jar包）,将解析语法都放在了

sql\catalyst\src\main\antlr4\org\apache\spark\sql\catalyst\parser\SqlBase.g4文件中

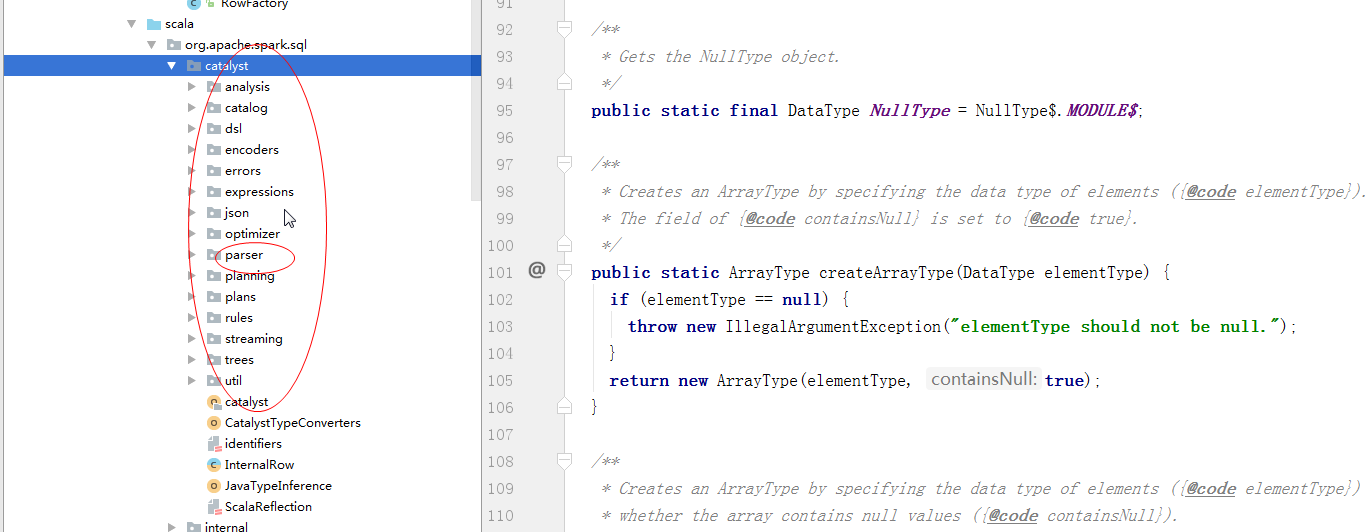
hive应该也是用的Antlr4作的词法语法解析（将sql语句转成MapReduce任务）

而spark则是将该g4文件转成java会scala类，某个Parser类





spark sql的源代码在catalyst里，一连串计划



构建逻辑如下图：

都在上述源码中有对应的包及类

