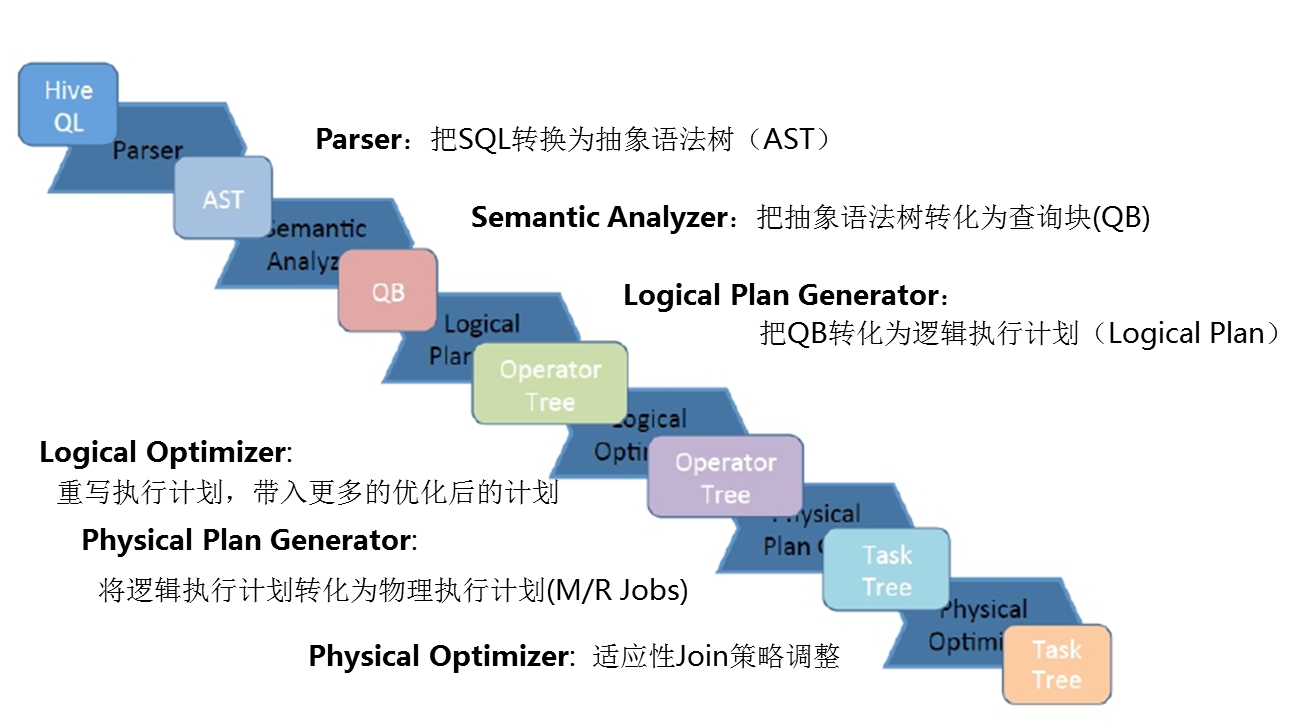
Apache Catalyst

# Hive SQL执行过程

Hive是基于Hadoop的一个数据仓库系统，通过Hive SQL执行分析任务。首先要将SQL编译为MapReduce程序，过程如下图所示：



执行步骤：

1. Antlr定义了SQL的语法规则，完成SQL词法及语法解析，将SQL转化为抽象语法树AST Tree。
2. 遍历AST Tree，抽象出查询的基本组织单元QueryBlock
3. 遍历QueryBlock，将其翻译为执行操作树Operator Tree，即逻辑执行计划（Logical Plan）
4. 逻辑层优化器进行OperatorTree变换，合并不必要的ReduceSinkOperator，减少Shuffle数据量
5. 遍历OperatorTree，翻译为MapReduce任务
6. 物理层优化器进行MapReduce任务的变化，生成最终的执行计划。

https://www.programcreek.com/java-api-examples/index.php?api=org.apache.hadoop.hive.ql.parse.ParseDriver

http://www.jianshu.com/p/7cd2afacc9bb

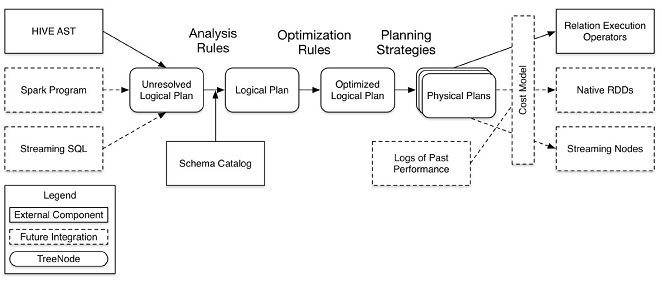
https://tech.meituan.com/hive-sql-to-mapreduce.html

<http://lengbingteng-163-com.iteye.com/blog/1344669>

<http://lxw1234.com/archives/2015/09/476.htm>

# Spark Catalyst

Spark SQL的执行流程如下：



Spark SQL源码分析：

http://blog.csdn.net/column/details/sparksql.html

http://www.cnblogs.com/shishanyuan/p/4723604.html?utm\_source=tuicool