

中国最专业软件开发培训机构

HIVE

加入尚学堂，一起进步！

SXT 尚学堂
中国IT教育高端领先品牌



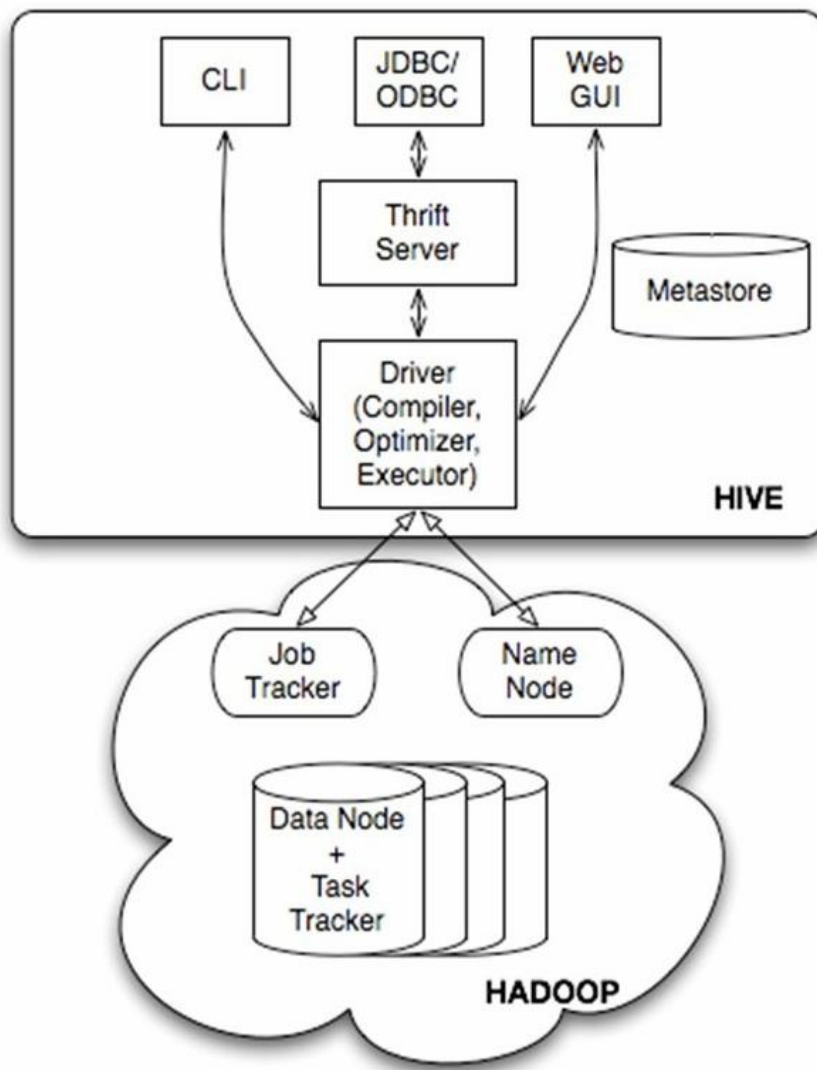
- Hive的产生：
 - 非java编程者对hdfs的数据做mapreduce操作



- **Hive:数据仓库。**
- **Hive：解释器，编译器，优化器等。**
- **Hive运行时，元数据存储的关系型数据库里面。**



- **Hive的架构**

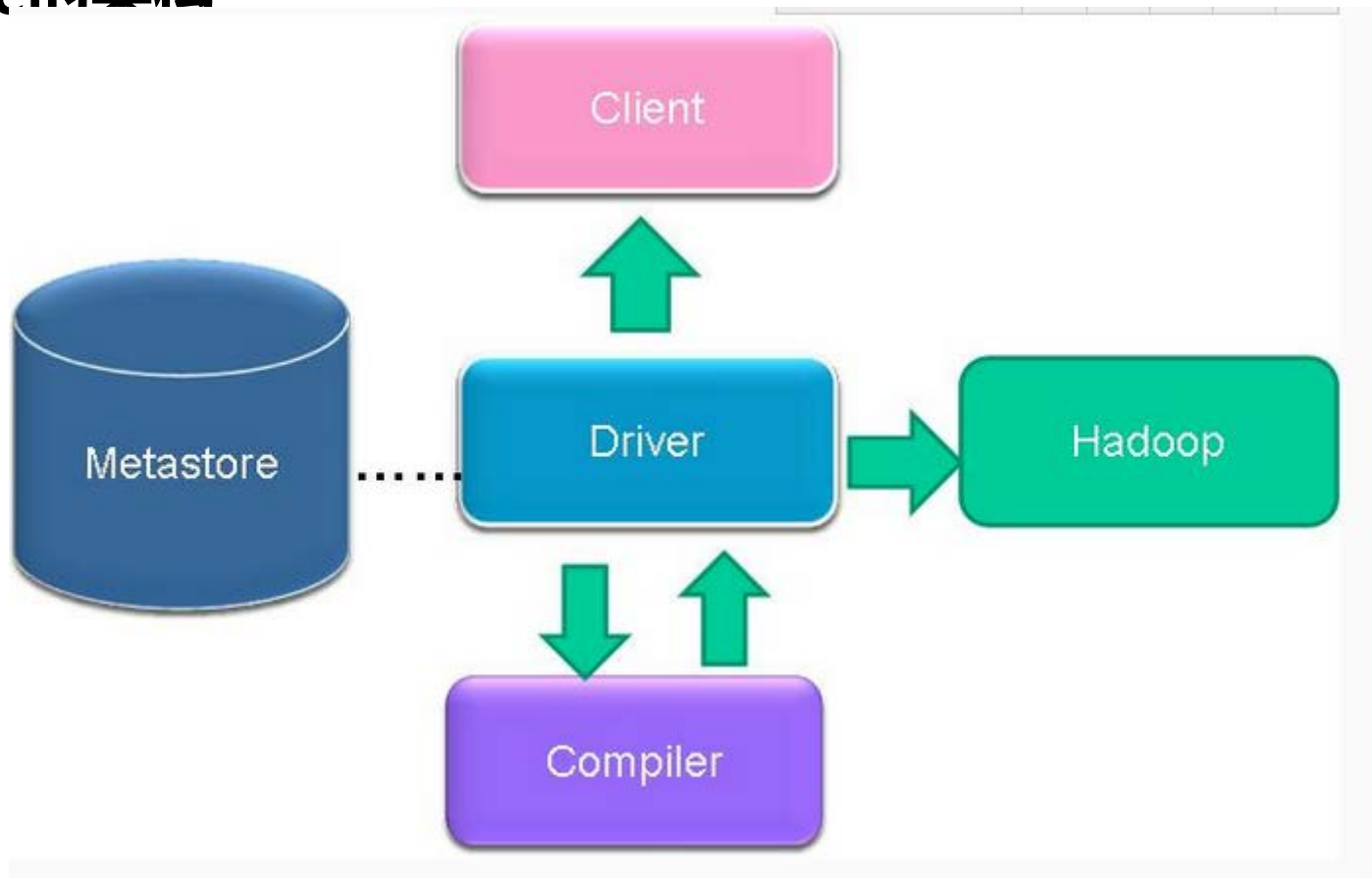


• Hive的架构

- (1) 用户接口主要有三个：CLI，Client 和 WUI。其中最常用的是CLI，Cli启动的时候，会同时启动一个Hive副本。Client是Hive的客户端，用户连接至Hive Server。在启动 Client模式的时候，需要指出Hive Server所在节点，并且在该节点启动Hive Server。WUI是通过浏览器访问Hive。
- (2) Hive将元数据存储在数据库中，如mysql、derby。Hive中的元数据包括表的名字，表的列和分区及其属性，表的属性（是否为外部表等），表的数据所在目录等。
- (3) 解释器、编译器、优化器完成HQL查询语句从词法分析、语法分析、编译、优化以及查询计划的生成。生成的查询计划存储在HDFS中，并在随后有MapReduce调用执行。
- (4) Hive的数据存储在HDFS中，大部分的查询、计算由MapReduce完成（包含*的查询，比如select * from tbl不会生成MapRedcue任务）。



- **Hive的架构**



- **Hive的架构**

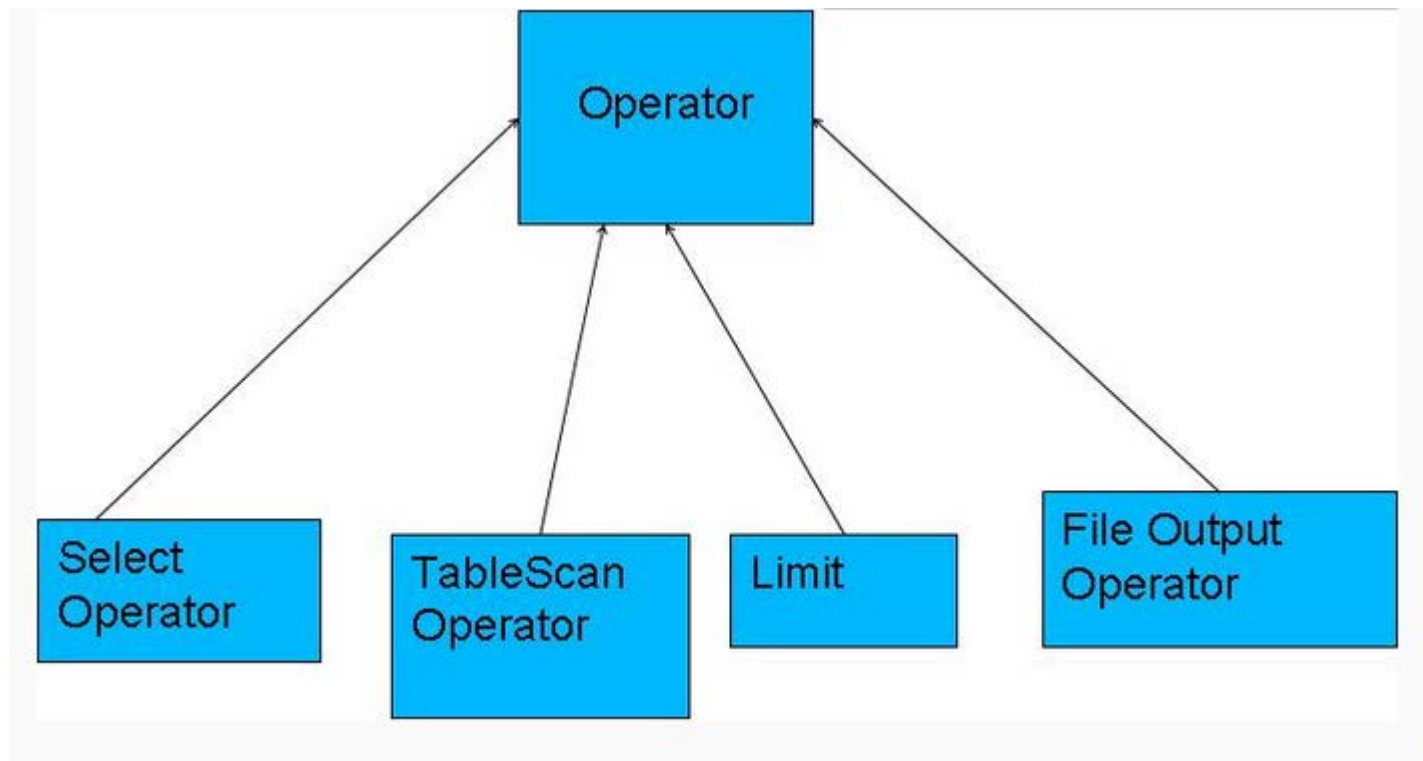
- 编译器将一个Hive QL转换操作符
- 操作符是Hive的最小的处理单元
- 每个操作符代表HDFS的一个操作或者一道MapReduce作业

- **Operator**

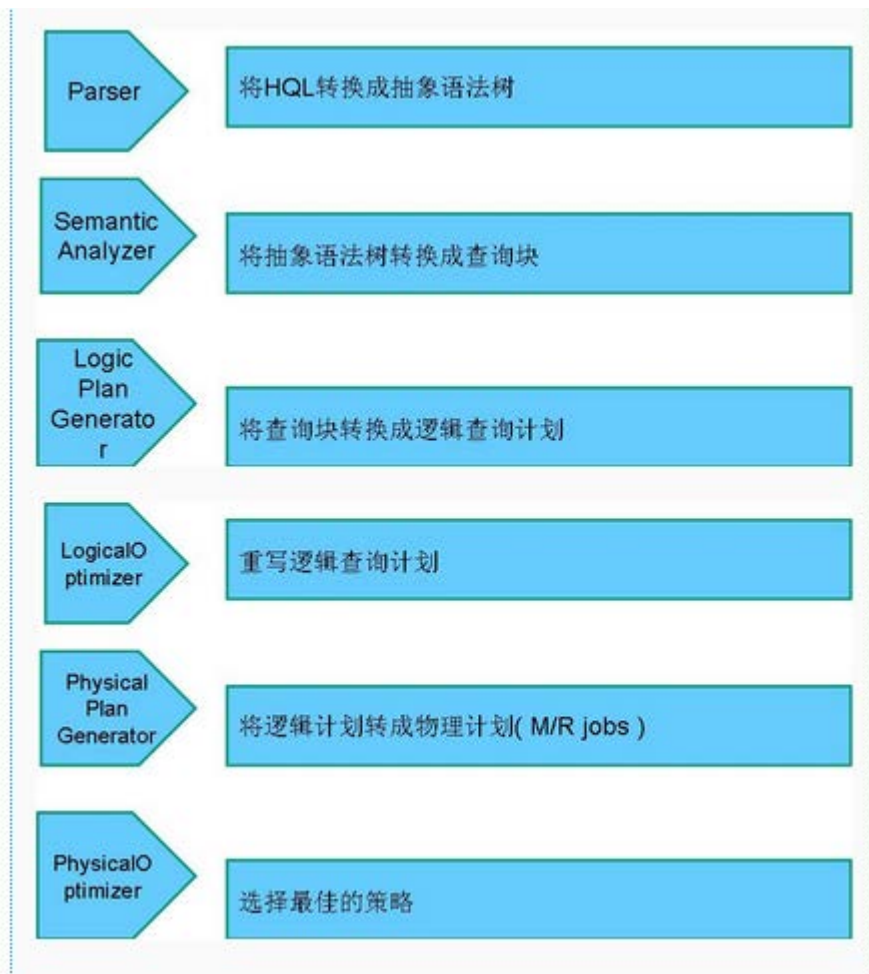
- Operator都是hive定义的一个处理过程
- Operator都定义有:
- `protected List <Operator<? extends Serializable >> childOperators;`
- `protected List <Operator<? extends Serializable >> parentOperators;`
- `protected boolean done; // 初始化值为false`



- **Hive的架构**

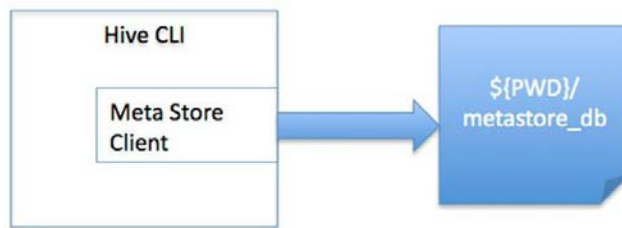


- **ANTLR词法语法分析工具解析hql**



• Hive的三种模式

- 单用户模式。此模式连接到一个In-memory 的数据库Derby，一般用于Unit Test。

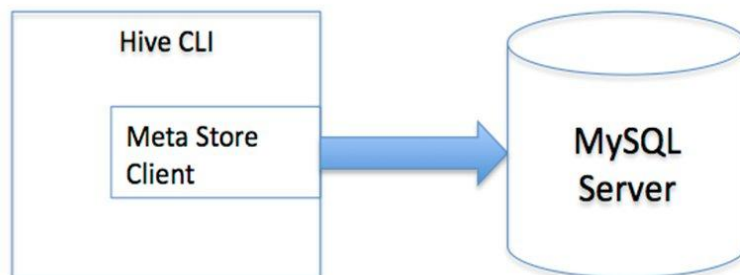


Parameter	Description	Example
javax.jdo.option.ConnectionURL	JDBC connection URL along with database name containing metadata	jdbc:derby;;databaseName=metastore_db;create=true
javax.jdo.option.ConnectionDriverName	JDBC driver name. Embedded Derby for Single user mode.	org.apache.derby.jdbc.EmbeddedDriver
javax.jdo.option.ConnectionUserName	User name for Derby database	APP
javax.jdo.option.ConnectionPassword	Password	mine



- **Hive单用户模式**

- 单用户模式。通过网络连接到一个数据库中，是最经常使用到的模式。



Parameter	Description	Example
javax.jdo.option.ConnectionURL	JDBC connection URL along with database name containing metadata	jdbc:mysql://<host name>/<database name>? createDatabaseIfNotExist=true
javax.jdo.option.ConnectionDriverName	Any JDO supported JDBC driver.	com.mysql.jdbc.Driver
javax.jdo.option.ConnectionUserName	User name	
javax.jdo.option.ConnectionPassword	Password	



• Hive多用户模式

- 远程服务器模式。用于非Java客户端访问元数据库，在服务器端启动MetaStoreServer，客户端利用Thrift协议通过MetaStoreServer访问元数据库



- Server Configuration same as multi user mode client config (prev slide). To run server
`$JAVA_HOME/bin/java -Xmx1024m -Dlog4j.configuration=file://$HIVE_HOME/conf/hms-log4j.properties
 -Djava.library.path=$HADOOP_HOME/lib/native/Linux-amd64-64/ -cp $CLASSPATH
 org.apache.hadoop.hive.metastore.HiveMetaStore`
- Client Configuration

Parameter	Description	Example
hive.metastore.uris	Location of the metastore server	thrift://<host_name>:9083
hive.metastore.local		false

