Build Cube流程主要分为四个阶段：

1. 根据用户的cube信息计算出多个cuboid文件
2. 根据cuboid文件生成htable
3. 更新cube信息
4. 回收临时文件

1.流程一：作业整体描述

把构建Cube的来源表总行数写到指定的HDFS文件中

2.流程二：生成中间临时数据

这一步的操作是根据Cube设计中的定义生成原始数据，这里会新创建一个Hive外部表，然后再根据Cube中定义的星型模型，查询出维度（对于Derived类型的维度使用的是外键）和度量值并插入到新建的表中，标的数据文件（存储在HDFS）作为下一个子任务的输入，它首先根据维度中的列和度量中作为参数的列得到需要出现在该表中的列，然后执行三步hive操作。

（1）如果临时表存在就删除

（2）创建外部表

（3）插入数据

3.流程三：创建事实表的Distinct Columns文件

根据流程二中生成的hive临时表作为输入，计算出表中的每一个出现在事实表中的维度和度量的distinct值，并写入到以列命名的文件中。如果某一个维度列的distinct值比较大，

那么可能导致mapreduce任务执行过程中OOM

4.流程四：构建维度词典

根据流程三生成的distinct colum文件和维度表计算出所有维度的字典。字典是为了节约存储而设计，事实表的每个维度组合都转换成唯一整数而减少内存占用。维度表的的每一行每一列都会得到一个编码id.然后存储id组成的新维表,id和值得映射关系。

5.流程五：保存Cuboid的统计信息

6.流程六：创建HTable

创建Htable时候需要考虑

1. 列族的设置
2. 每个列族的压缩方式
3. 部署coprocessor
4. HTable中每一个region的大小

hbase中存储的数据key是维度成员的组成，value是对应聚合函数的结果。

7.流程七：计算生成Base Cuboid数据文件

计算BaseCuboid。

8.流程八：计算N层的Cuboid文件

每一层一个mapreduce程序，计算每一程的coboid。

9.流程九：基于内存构建Cube

如果选择In-mem cubing的方式这一步会代替前两步生成cube。

10.流程十：将Cuboid转换成HFile

11.流程十一：将HFile导入到Hbase表中

12.流程十二：更新Cube信息

13流程十三：清理中间表