**sql语句**

select

t1.a, count(\*), avg(t1.b)

from

t1 left outer join t2 on t1.a = t2.a

group by t1.a

**解析**：t1左外连t2，等值条件t1.a=t2.a,字段a的行数，字段b的平均值，查询内容按字段a分组

**执行计划**：将t1,t2左外连然后执行聚合计算

**性能分析**：如果t1数据特别大字段特别多，左外连后就会产生一个巨大的结果集，由于这个sql有两个运算，所以会需要把前一个计算的结果落盘一个临时关系做备用，临时关系都需要写入磁盘所以会使执行效率大大降低。由于整个sql语句只用到t1.a,b和t2.a，应该先将表投影获取需要的数据然后在进行连接。在hashjoin中，左外连，t1是outer表，t1越大outer filter计算耗时增加。最后我们应该考虑数据分布的影响，我们应该将计算尽量下推到存储节点，避免不必要的网络传输，例如本句sql中的group by。我们可以进行预聚合只需每个节点返回一个结果，由主服务器一起处理即可。考虑计算与数据之间的依赖关系。

Plan树

1.直接连接

group by t1.a

count(\*)

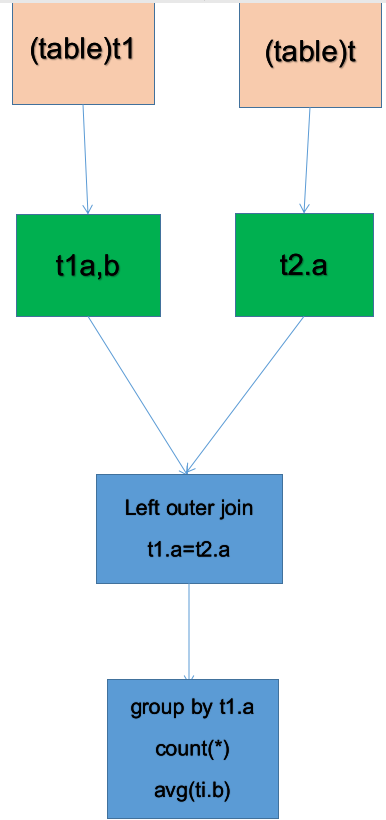
avg(t1.b)

Left outer join t1.a=t2.a

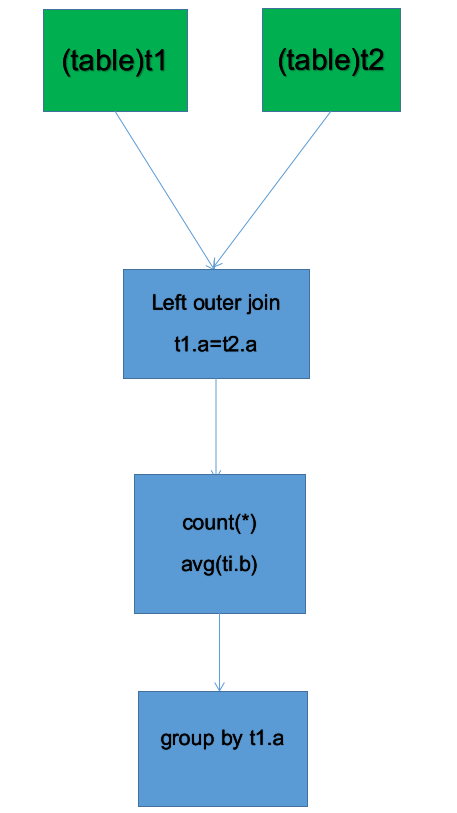
(table)t2

(table)t1

2.先投影再连接



3.直接连接，group by下推



4.先投影再连接，group by下推（和3差不多就不画了）

5.既然所有的聚合函数与group与table2没有关系，那么就将他们都下推到存储节点table1。

