Write Up for acmdb-lab4

在lab4中，我们实现了数据库中基于成本的优化器(cost-based optimizer)。

## Exercise 1 (**IntHistogram**)

其实现逻辑大概如下：首先先全表扫描一次，获取每个字段的最大值与最小值（目的是为了获得区间范围）。此时就可以建立直方图的初始坐标轴了，一种简单的方法是使用固定数量的桶NumB，每个桶表示直方图属性域的固定范围内的记录数。例如，如果字段f的范围从1到100，并且有10个存储桶，那么存储桶1可能包含1到10之间的记录数计数，存储桶2可能包含11到20之间的记录计数，依此类推。然后再次扫描表，选择所有元组的所有字段，并使用它们填充每个直方图中的桶计数。

## Exercise 2 (TableStats)

这个Exercise主要是完成TableStats类与每页的IO花销，构造每个字段的直方图，并利用Exercise1的方法，得出每个直方图相关的选择性估算。

## Exercise 3 (Join Cost Estimation)

这个exercise中完成的是进行对连接成本的估算，根据lab中对join成本估算的公式完成。

## Exercise 4 (Join Ordering)

对于Exercise4则是总体生成优化过后的连接顺序。根据计算成本的公式，不同顺序的连接成本也不同，这里使用了动态规划的思想完成。

## 总结

lab4总体来说理解难度比较高，难度上主要集中在queryTest的跑通，有些逻辑要根据测试的逻辑来编写，比较耗费时间。







