Ferhat Elmas 214805

Advanced Databases Hw-8

1-) a) Low selectivity factor is the assumption behind of this equation because otherwise, output rate is basically depends on the probing speed of cpu since we will select all tuples we can probe.

b) Therefore, we need a query with low selectivity factor. For example, R(a) and S(b) are tables to be joined by |R.a - S.b| < 2 when columns a and b are int. Moreover, they have a range of i.e [0, 100000].

c) We can utilize of multi core architecture by doing round robin processing of the stream between cores. By this design, processing will seem faster relative to the stream (in some order) then allocation resources starts to make difference.

2) a)

3, 5, 6, 15, 15, 16, 19, 19, |>41, 45, 49, 54, 57, 64, 65, 67, 71<|, 79, 80, 85, 92, 93, 95, 97

if we count values that are in the range, query result is 9.

b) bucket boundaries are : 6, 16, 41, 54, 65, 79, 92, 97
Therefore, this calculation ((41-20)/(41-16))*3+3+3+((71-65)/(79-65))*3 gives 9.81

c) Haar wavelet transform is [56.25, -27, -18.25, -11.25, -5, -6.5, -7, -2.5]
Ignoring the leaves, query result is
((47.5-20)/(47.5-11))*6 + ((71-47.5)/(72-47.5))*6 = 4.52 + 5.76 = 10.28

3) a) After 30

cs100	3.7	90	2.32
cs101	3.3	90	2.32
cs103	3.9	90	2.32

b) After 60

cs100	3.7	90	1.64
cs101	3.6	90	1.64
cs103	3.8	90	1.64

c) After 90

cs100	3.77	90	1.34

cs101	3.57	90	1.34
cs103	3.87	90	1.34

As it is seen from the tables, confidence interval decreases while more tuples are processed and at the end it will be zero and meaningless since all tuples are processed and grand truth is learnt.

4) Here are scanning methods and optimal join types:
Intermediate1 = R(sequential) |X|{MergeJoin} S(sequential)
Intermediate2 = T(indexStride) |X|{NLJ} Intermediate1(sequential)
Result = Intermediate2(indexStride) |X|{NLJ} V(sequential)