Problems

1.
A. Estimate the number of tuples and blocks in the selection $\sigma_{h=723}(X)$.
Select * from x where h = 723

$T(\sigma_{h=723}(X)) \approx T(X) / V(X, h) = 100,000 / 1,000 = 100$
$B(\sigma_{h=723}(X)) \approx B(X) / V(X, h) = 10,000 / 1,000 = 10$

B. Estimate the number of tuples in the join $R \bowtie S$.

This is a HARD question. Cardinality estimation of joins has been an active research topic for many years. We don't have enough statistics on R and S to make a good estimate. Here is a common estimate:

$\approx (T(R) * T(S)) / \max\{V(R, f), V(S, f)\} = (1,000 * 5,000) / \max\{100,1000\} = 5,000$

2.

a) (2 points) Clustered index scan, Cost = X*B(R) = 4000 / 2×10^3 = 2
Cardinality, T(a) = X*T(R) = 6×10^5 / 2×10^3 = 300)

b) (3 points) Clustered index join, Cost = B(a) + T(a)*X*B(S) = 0 + 300 * 9×10^4 / 3×10^4 = 900

c) (2 points) On-the-fly, cost = 0

d) (2 points) Unclustered scan, cost = X*T(S) = 3×1000000 / 1000 =

3000

e) (3 points) In memory join, pipelined, cost = 0

f) (3 points) What is the total IO cost of each plan? Which plan would you choose to minimize the total IO cost?

plan i cost = a + b + c = 2 + 900 + 0 = 902.
plan ii cost = a + d + e = 2 + 3000 + 0 = 3002. Pick plan i since lower cost.