Question 1 (20 points)

1.1) (6 points)
Block Nested Loops Join
(A) Reads = |R| + |S| * ceil(|R|/(B-2)) = 5,000 + 15,000 * ceil(5,000/(800-2)) = 110,000
(B) Writes = 0 (no writes are required for nested loops joins)

1.2) (7 points)
Grace Hash Join
(A) Reads in Partition phase = |R| + |S| = 20,000
Reads in Probe phase = |R| + |S| = 20,000
Total reads = 2 * (|R| + |S|) = 40,000
(B) Total writes = |R| + |S| = 20,000

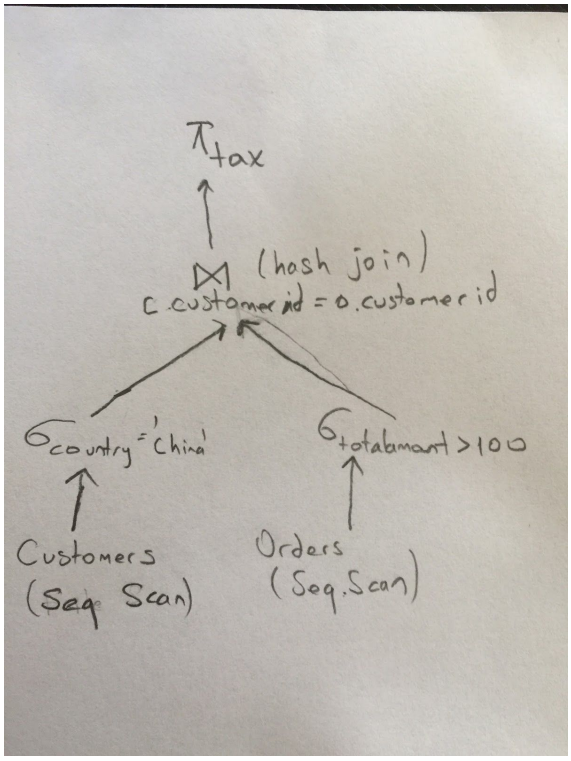
1.3) (7 points)
BNJ = 110,000 * 1 [reads; no writes] = 110,000
GHJ = 40,000 * 1 [reads] + 20,000 * X [writes]
From BNJ = GHJ ⟶ 110,000 = 40,000 + 20,000 * 3.5,
we get X = 3.5
So we should choose BNJ if X > 3.5, but GHJ if X < 3.5

Question 2 (20 points)

2.1) (10 points)
4 Join Orders:
Join(Join(S, R), B)
Join(Join(R, S), B)
Join(Join(B, R), S)
Join(Join(R,B), S)

2.2) (10 points)
5 of the following attributes:
S.sid
S.age
S.rating
B.bid
B.color
R.bid
R.sid

Question 3 (10 points)



Question 4 (30 points)

4.1) (6 points)
Size of each B+ tree index entry = size of key + size of RID = 16 + 32 = 48 bytes
# of entries per index page = fill factor * entries per page = 0.71 * 4096/48 = 60 entries per page
Total entries = # of index pages * # of entries per index page = 64 * 60 = 3840 = total tuples in S

4.2) (6 points)
# of pages in S = ceil(# of tuples in S / floor(page size / size of S tuple)) = ceil(3840 / floor(4096 / 48)) = 46 pages

4.3) (6 points)
# of pages in T = ceil(# of tuples in T / floor(page size / size of T tuple)) = ceil((240 * 1,024) /floor (4,096 / 32)) = 1,920 pages

4.4) (6 points)
Read cost = ceil(.5/4 * # of leaf pages) + ceil(.5/4 * # of pages in S) = ceil(.5/4 * 64) + ceil(.5/4 * 46) = 14

Write cost =
# of tuples selected = ceil(0.5 / 4 * # of tuples of S) = ceil(.125 * 3840) = 480
Only S.sid and S.sname are required for later operations, so each tuple is 32B.
Thus write cost = ceil(# of tuples selected / (page size / tuple size)) = ceil(480 / (4096 / 32)) = 4

4.5) (6 points)
(A) 2 (|S| + |T|) = 2 * (4 + 1,920) = 3848
(B) |S| + |T| = 4+ 1,920 = 1924

Question 5 (20 points)

(A) (8 points) 3 non-leaf index nodes + 400 leaf pages + 8,000 data record pages = 8,403
(B) (8 points) 3 non-leaf index nodes + 400 leaf pages + (100 tuples per page * 8,000 data record pages) = 800,403
(C) (4 points) An external sort will perform better than using the index when...
Answer 1: When the index is unclustered, OR
Answer 2: When the selectivity is high