```
# COMP9315 20T1 Final Exam Q5

Type your answer(s) to replace the xxx's
Submit this file as your answer for Q5

a.
- Buffer usage:
  - 18 buffers used for input from R
  - 1 buffer used for input from S
  - 1 buffer used for output

- #blocks of R = ceil(100 / 18) = 6
- Cost involves reading in every page of R,  and reading each page
  of S for every block of R

- Cost = 100 + 6 * 20
       = 100 + 120
       = 220

b.
Sort phase
- Each table is sorted on x

- Sorting R
  - Uses external merge sort
  - Produces sorted version of R with 100 pages
  - Cost = 2 * b_R * (1 + ceil(log(B - 1, ceil(b_R / B))))
         = 2 * 100 * (1 + ceil(log(19, ceil(100 / 20))))
         = 2 * 100 * (1 + ceil(log(19, 5)))
         = 2 * 100 * (1 + 1)
         = 400

- Sorting S
  - Requires only one pass, as there are exactly enough buffers to
    store the entire table in memory
  - Produces sorted version of S with 20 pages
  - Cost = 20 (reads) + 20 (writes)
         = 40

- Cost of sort phase = 400 + 40
                     = 440

Join phase
- Buffer usage:
  - 1 buffer used for output
  - 19 buffers used for input from sorted R and sorted S
- Cost involves reading in sorted version of R and S from previous
  phase (cost of writing final result is ignored)
- Assume that there are no long runs of equal x values which would
  requires us to re-read pages from disk
- Cost of join phase = 100 + 20
```

```
                        = 120

In summary
- Total cost = 440 + 120
             = 560

c.
Partition phase
- Partioning R
  - Buffer usage:
    - 1 buffer used for input from R
    - 19 buffers used for partitions of R

  - Cost involves reading all pages of R and writing partitions
  - Assuming that hash functions distribute tuples uniformly, each
    of the 19 partitions will contain ceil(100 / 19) = 6 pages, so
    partitions will occupy 19 * 6 = 114 pages in total.

  - Cost = 100 (reads) + 114 (writes)
         = 214

- Partitioning S
  - Buffer usage:
    - 1 buffer used for input from S
    - 19 buffers used for partitions of S

  - Cost involves reading all pages of S and writing partitions
  - Assuming that hash functions distribute tuples uniformly, each
    of the 19 partitions will contain ceil(20 / 19) = 2 pages,  so
    partitions will occupy 19 * 2 = 38 pages in total.

  - Cost = 20 (reads) + 38 (writes)
         = 58

- Cost of partition phase = 214 + 58
                          = 272

Join phase
- Buffer usage:
  - 1 buffer used for input from R partition
  - 1 buffer used for input from S partition
  - 1 buffer used for output
  - 17 buffers used for in-memory hash table

- Cost involves reading in partitions of R and S from the previous
  phase (cost of writing final result is ignored)
- Assuming that all partitions of R will fit in the in-memory hash
  table, no partition will need to be read more than once

- Cost of join phase = 114 + 38
                     = 152
```

```
In summary
- Total cost = 272 + 152
             = 424
```