```
COMP9315 Sample Exam, Q3 Sample Solution

(a)

space for tuples = S
space for directory = D = ceil(S/8)
S*32 + ceil(S/8) <= 4096

C = 127

(b)

total pages = ceil(1234567/127)

b = 9721

(c)

first pass: read 101 pages at a time and sort

gives ceil(9721/101) = 97 sorted chunks

subsequent passes: use 100-way merge

number of passes  = ceil(log_100(#chunks)) = ceil(log_100(97)) = 1

sort cost = 2b for first pass + 2b * number of later passes
          = 2b + 2b

could eliminate duplicates on final sort pass,
in which case we save one writing of all b pages
because we don't count the cost of writing the final result
=> total cost = 3b + = 3*9721 = 29163

if we didn't do the above optimisation
=> sorting cost = 4b = 4*9721 = 38884 to produce a sorted file
then, one more pass is needed to remove duplicates
=> total cost = sorting cost + duplicate elim cost
              = 4*9721 + 9721 = 48605

All the more reason to eliminate duplicates in the final sorting
pass.

(d)

first pass: partitions data into 100 files, each containing 98 pages

second pass: handles each partition to eliminate duplicates

cost = 3b = 3*9721 = 29163
```

(e)

With 91 buffers, there would be fewer/larger partitions and so the partitions would no longer all fit in memory. Thus, some partitions would need rescanning. Overall cost would increase.