10 – Sorting & Aggregations

1. External Merge Sort

   A. Many queries need specific order of tuples

   B. Quick-sort can be terrible if DB doesn't fit in memory

   C. process

      i. read pages, sort them individually, write back

      ii. recursively merge pairs of pages
      using 3 buffer frames

   D. double buffering optimization

      i. make double input and double output buffer
      toggle them when each buffer is full

      ii. minimize I/O overhead

2. Aggregations

   A. Hashing can be alternative of sorting
   if query doesn't need the data to be ordered

   B. External Hashing

      i. Partition
      hash tuples at B-1 pages (1 page for input)

      ii. Re-hash
      hash each partition with another hash function
      if partition doesn't fit in memory, partition again and re-hash them