The assignment should be done in groups of two students. You must turn in the source code of your program through Canvas. Each group must submit only one file that contains the full name, OSU email, and ONID of every member of the group.

1: External Memory Sorting (12 points)

The objective of this assignment is to learn the implementation of sorting algorithms on external memories.

Consider the following relation:

Emp (eid (integer), ename (string), age (integer), salary (double))

Fields of types integer, double, and string occupy 4, 8, and 40 bytes, respectively. Each page can fit at most one tuple (record) of the input relation. Implement the two-pass multi-way sorting for the file Emp.csv in C/C++ using the skeleton code posted with this assignment. The sorting should be based on the attribute *eid*. There are at most 22 pages available to the sort algorithm in the main memory, i.e., the size of the buffer (M) is 22.

- **Input File:** The input relation is stored in a CSV file, i.e., each tuple is in a separate line and fields of each record are separated by commas. The file that store relation *Emp* is *Emp.csv*. Your program must assume that the input file is in the current working directory, i.e., the one from which your program is running.
- **Final Output:** The program must store the result in a new CSV file with the name *EmpSorted.csv* in the current working directory.
- **Main Memory Limitation:** Your program can keep up to 22 pages in main memory at any time. The control, local, or temporary variables that you use in your program are excluded from this limit. The submitted solutions that use more main memory will *not* get any points.
- **Page Format:** You can use the data structure of your choosing to represent and store pages in main memory and files.
- **Types of Temporary Files (runs):** You can use the type (text or binary) of your choosing for the temporary files (runs).
- Each student has an account on hadoop-master.engr.oregonstate.edu server, which is a Linux machine. You must ensure that your program can be compiled and run on this machine. You can use the following bash command to connect to it:

  ```
  > ssh your_onid_username@hadoop-master.engr.oregonstate.edu
  ```

  Then it asks for your ONID password and probably one another question. To access this server, you must be on campus or connected to the Oregon State VPN.
- You can use the following commands to compile and run C++ code:

  ```
  > g++ -std=c++11 main.cpp -o main.out

  > main.out
  ```

- **Grading Criteria:** The programs that implement the correct algorithm, return correct answers, and do not use more than allowed buffers will get the perfect score. The ones that use more buffer than allowed will not get any points. The ones that implement the right algorithm but return partially correct answers will get partial credits.