CSCI3170 Short Assignment #4  
(Deadline: Nov 29 23:59)

Name: Liu Hoi Pan Pass / Fail

Student ID: 1155127464

Consider the following three relations for the database in a company:

**Worker** (*WID*: integer, *Name*: string, *age:* integer, *rating*: integer)

**Project** (*PID*: integer, *project\_name*: string, *budget*: real)

**Work\_in** (*PID*: integer, *WID*: integer, *since:* date)

Assumptions:

* Assume 
* Assume
* The distribution of *age* in **Worker** is uniform.
* The distribution of *rating* in **Worker** is uniform.
* **Worker:** 50 tuples per page, 1000 pages
* **Project**: 60 tuples per page, 2000 pages
* **Work\_in**: 40 tuples per page, 5000 pages
* The buffer size is 20 pages.

Consider the following execution plan. Assume the sizes of T1 and T2 are 20 and 50 pages respectively.

**(Sort-Merge Join)**

**On-the-fly:**

**Write to T1**

**On-the-fly:**

**Write to T2**

**(File scan)**

**(Cluster B+ tree on age)**

1. Calculate the number of page accesses for scanning the table **Work\_in** and writing the matching tuples to **T1**.

Answer:

Scanning the table **Work\_in** access 5000 pages.

Write the tuples to **T1** access 20 pages.

So, the answer is 5020 pages.

1. Assume the height of the B+ tree on age is 3 (i.e. number of levels = 4). Calculate the number of page accesses for selecting the tuples from **Worker** and writing the matching tuples to **T2**.

Answer:

Searching the tree access 4 pages.

Since the B+ tree is clustered, we do not need to scan the leaf nodes.

Note that the ages of workers are uniform, the number of workers with ages between 35-44 is 1/4 of number of workers with ages between 21-60. Totally, we need to access 1/4 \* 1000 = 250 pages

Write the tuples to **T2** access 50 pages.

So, the answer is 304 pages.

1. Calculate the number of page accesses for sorting **T1** and **T2**, and hence calculate the number of page accesses for the Sort-Merge Join of **T1** and **T2**.

[Hint: When M > B, the formula for the cost of sorting is **2\*M\*(⎡logB-1M/B⎤ + 1)** ]

Answer:

T1 contains 20 pages, so M = 20 for sorting T1.

Buffer B is always 20.

Sorting T1 in the join column, the cost is 2\*20\*(⎡log20-120/20⎤ + 1) = 40 pages.

T2 contains 50 pages, so M 0 50 for sorting T1.

Sorting T2 in the join column, the cost is 2\*50\*(⎡log20-150/20⎤ + 1) = 200 pages.

Merge join T1 and T2 requires us to scan T1 and T2, the cost is 20 + 50 = 70 pages

So, the answer is 310 pages.