CSCI3170 Short Assignment/Classwork #5  
(Issued in the lecture on April 11)

Name: Pass / Fail

Student ID:

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**.
2. 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**.
3. 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)** ]