**Kate Moreland**

**kem0149**

**904239271**

**COMP 5120**

**Homework 3**

1. Answer the following questions about files and indexes:
   1. What alternatives are available for the data entries in an index?

There are 3 alternatives for data entries in an index. They can be either actual data records, rid of data record, and list of rids of data records. The actual data records takes use of index structure. Rid of data records is normally smaller than the first alternative. Lists of rids of data records is even smaller than the second alternative.

* 1. What is the difference between a clustered index and an unclustered index? If an index contains data records as ‘data entries’, can it be unclustered?

A clustered index is when the order of data records is very similar to the order of data entries. An unclustered index is when the index order is different than the physical storage order. If there are data records as data entries in an index, then it has to be clustered and cannot be unclustered.

* 1. How many clustered indexes can you create on a file? Would you always create at least one clustered index for a file?

You can only use one cluster index on each file. There is only one way to store the records. There is not always a need to create a clustered index. If there is a need for a workload that is insert-based, then it would cost more to take advantage of a clustered index.

1. Explain the terms seek time, rotational delay, and transfer time.

The term seek time refers to the time it takes to move the disk arm to the correct track. Rotational delay refers to the time that it takes to wait for the disk to rotate to the correct track under the read and write head. Finally, the transfer time refers to the time that it takes to move the data from the track to the head.

1. What is sequential flooding of the buffer pool?

A sequential flooding is when a sequential scan of a fairly large file will replace all buffer pages. This will result in all of the correct pages being pushed out. This will flood the buffer, meaning that no local storage can be stored there.

1. Describe two possible record formats. What are the trade-offs between them?

The two possible record formats are fixed length and variable length records. Fixed length records are much simpler and have a quicker access time. However, they result in a greater waste of space. Variable length records are a better use of space and allow for more flexibility. They are more difficult to manage.

1. Why do frames in the buffer pool have a pin count instead of a pin flag?

A pin count allows to keep track of the number of users on the page and return the address. If the pin flag would have been used it would have failed it multiple users were to access the same page. This allows for more accuracy and correctness.