Jalen Powell

07/26/2023

COMP 5120

Homework 3

Question 1: Answer the following questions about files and indexes:

1. What alternatives are available for the data entries in an index?

* There are three main alternates for what to store as a data entry in an index.
  + (1) A data entry k\* is an actual data record (with search key value k).
  + (2) A data entry is a 〈 k, rid 〉 pair, where rid is the record id of a data record with search key value k.
  + (3) A data entry is a 〈 k, rid-list 〉 pair, where rid-list is a list of record ids of data records with search key value k.

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 one in which the ordering of data entries is the same as the ordering of data records. We can have at most one clustered index on a data file. An unclustered index is an index that is not clustered. We can have several unclustered indexes on a data file. If the index contains data records as ’data entries’, it means the index uses Alternative (1). By definition of clustered indexes, the index is clustered

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

* At most one, because we want to avoid replicating data records. Sometimes, we may not create any clustered indexes because no query requires a clustered index for adequate performance, and clustered indexes are more expensive to maintain than unclustered indexes.

Question 2: Explain the terms seek time, rotational delay, and transfer time.

* Seek time: It measures the time that head assembly takes on the actuator arm to travel to the disk track on which the data is either read or written.
* Rotational delay: It is the time taken by the required sector in the track to position itself under the read/write head.
* Transfer time: It is when the access time is added to the time that it takes to actually read the data from the disk

Question 3: What is sequential flooding of the buffer pool?

* **Sequential flooding of the buffer pool is when the number of buffer frames is less than the pages in file, this will result in reading every page of the file.**

Question 4: Describe two possible record formats. What are the trade-offs between them?

* Two famous record formats are “fixed length record format” and “variable length record format.” Fixed length record format is used to store text and images and it is platform independent and even it can work in different computer system. Variable length record format is not, much important part of record format because it is difficult to read or write and it uses byte string method to store data.

Question 5: Why do frames in the buffer pool have a pin count instead of a pin flag?

* A page in a frame can be requested by many different users or transactions at the same time.