COMP2411 Database Systems

Assignment 2

Name: ZHOU Siyu

Question A

1）Hash function advantages: hash function makes it easier for searching point, as hash index is given, the specific key will be found. Time for searching is O(1) as is there's a hash index for an attributes, the point should be found more faster than if there's no hash function realative B-tree indeces.
Hash function's disadvantage: hash function is only suitable for searching one value or one column.
If only secondary index is available in condition, each retrived records should be further tested to determine whther it satisfies the remaining condition.

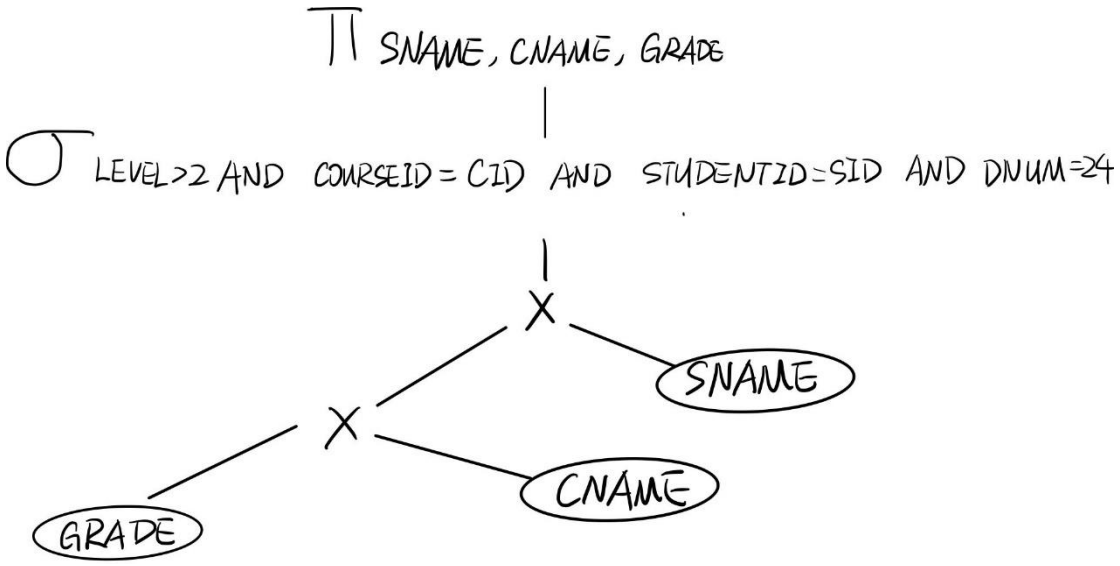
2）Dense has index entries for every search key value in the database file. This makes searching faster but requires more space to store index records itself. In a sparse index, index records are not created for every search key. An index record here contains a search key and an actual pointer to the data on the disk. Then dense index is preferable when user don't care about the cost of storing space, and then we need to query data in shorter time when searching data.

3）Difference between primary index and secondary index
In definition, the primary index is an index on a set of fields that includes the unique primary key, and is guaranteed not contain duplicates, and in contrast, a secondary index may have duplicates.
As for order, the primary index requires the row in data block to be ordered on the index key, while the secondary index don't impacts on how rows are actually organized in data blocks.
For number of indexes, there's only one primary index, while there can be multiple secondary indices.
There's no duplicates in the primary index, but secondary index can have duplicates.

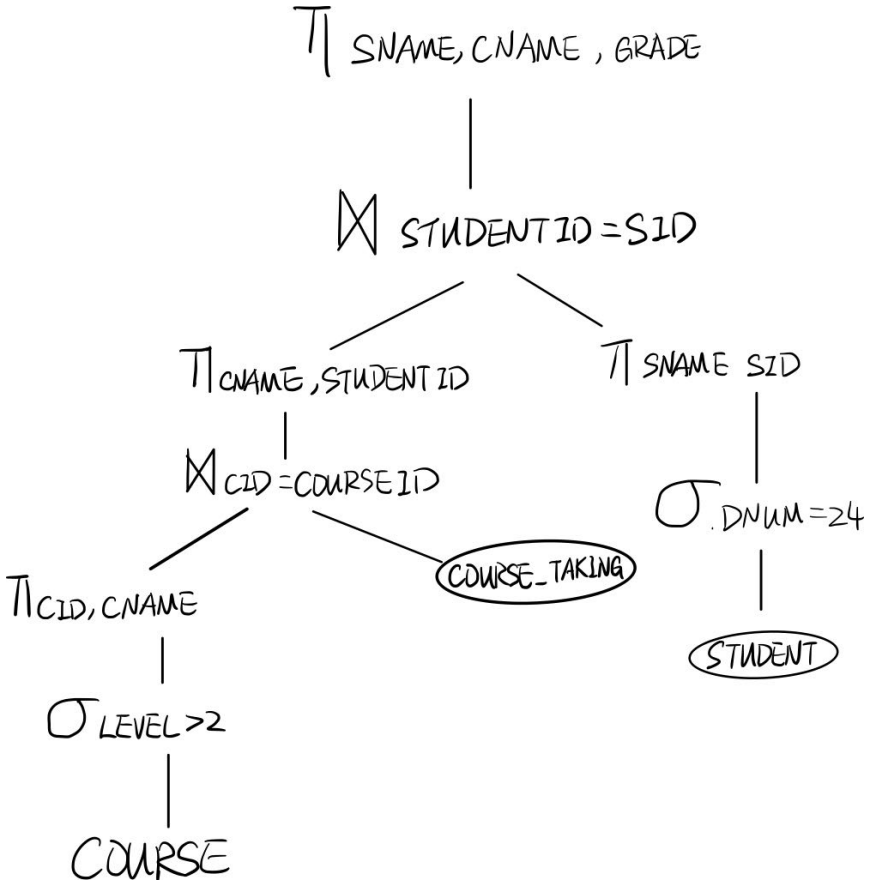
4）Hash function should have distribution characteristics so that the key that are together for the search key are not to be mapped to the same hash value. Also, it makes searching more efficient during time for reducing number of hash value collision.

Question B

a)



b) Assumption: there are less course with LEVEL > 2 compared to the student with DNUM = 24



Question C

1) I agree.
   Reason: By definition, If a transaction reads the value of an operation from an uncommitted transaction and commits before the transaction from where it has read the value, then such a schedule is called nonrecoverable schedule.As for nonrecoverable schedual, we cannot a transaction. Brcause it let the database being inconsitent, and read is not recoverable.

2) Because at the time of a system crash, only the log entries that have been written back to disk are considered in the recovery. The transactions will be undone if they apply to a single operation rather than to whole transaction. The certain transactions will be redone to ensure that all the operations of a committed transaction hav been applied successfully to the database.

3) Crash recovery is the process of rolling back transaction of the database after a system crash. When the system is recovering from a crash then transactions must be undone and redone in the reverse order by undo or redo the write operations from the log file using the redo method. Additionally, restoring the database restores the system files to the earlier backup point without affecting the data.