Write-up

In Lab-1, we implemented different components of SimpleDb, which helped us understand how data is stored in files, organized in pages, maintained in a buffer pool for quick access, and sequentially scanned for every execution.

The components we implemented are TupleDesc, Tuple, BufferPool, HeapFile, HeapPage, SeqScan, PageID, and RecordID.

In SimpleDB, each table is associated with a single HeapFile, which is saved as a single file. HeapFile comprises several pages, each of which is referred to as a HeapPage and stores a certain number of bytes of data in it. The data stored in the HeapPage is in the form of tuples, which means a single tuple of data in the table. Each of the tuples corresponds to a TupleDesc, which is a schema of the table. BufferPool acts as a cache for the pages read from the disk. However, BufferPool depends on the HeapFile to read the data from the disk and store each page for quicker access. It can only hold a fixed number of pages currently. SeqScan operator takes tableId as one of the parameters, then SeqScan retrieves the corresponding file for that table and calls its iterator. To return the next tuple, SeqScan requests the HeapFile to get one more tuple. The HeapFile has an internal iterator that keeps track of the different HeapPages and sequentially gets all the tuples in the table. The HeapFile stores the iterator's instance on the HeapPage it is currently examining. When the HeapPage does not have any tuples to return, then the HeapFile moves to the next page. The HeapPage header in the HeapPage keeps track of all the valid tuples in the page. The HeapPage iteratively returns all the tuples until no more is left on the page. The SeqScan operator retrieves all the tuples for a specific file as the heap file associated with the table keeps returning tuples one by one.

Design Decisions: We used 'ConcurrentHashMap' in 'BufferPool' for the thread-safe operation; although threading is not essential for this class, we included it since it's suitable for concurrent access.

Unit Test Addition: One example of the unit test could be to check whether the 'getNumEmptySlots()' in 'HeapPage' gives the correct number of empty slots.

API changes: No changes.

All elements in my code are present.

No Feedback.