**CSE 510: Database Management Systems Implementation**

**Project Phase 1**

**Submitted by Kyle Stevens (1214241534)**

**Group Members Group 1:**

Aryyama Kumar Jana

Sahil Santosh Patil

Neha Dandekar

Kyle Stevens

Jasmin Kantibhai Trada

Venkataramana Balaji Rajendran

**Abstract**

The purpose of phase 1 of this project is to introduce, reintroduce as the case might be, students to the idea of a DBMS, specifically in the case of this course the java-based DBMS Minibase. This gives students an opportunity to ensure their understanding of relational database functionality and familiarize themselves with the visualization and interface of Minibase. Through the analysis of the included Minibase tests the user also can gain an understanding of the limitations and capabilities of the Minibase DBMS and DBMS systems in general. Analysis of the Buffer Manager Test allows the user to focus on learning the operations that can be done on the buffer and what limitations exist on the buffer. This is also true of the Heap File Test, Disk Manager Test, and B+ Tree Test. The phase also provides the user with an introduction to B+ tree data structures if they are unfamiliar with it, or a refresh for those who have encountered the data structure before. This phase serves to test to ensure that the user can implement basic tests, has a general understanding of DBMS capabilities, and is able to use both the Java programming environment and a Unix-based Linux environment.

**Introduction**

Phase 1 of this project covers various test cases for features found in a database management system (DBMS); presenting both a live teaching tool and opportunity to refine or refresh basic database knowledge and the component of a DBMS system. Not only does this phase serve as a method of introduction to DBMS systems, but also to a specific version used for education and visualization in a universal platform in the form the Minibase java DBMS.

Minibase is a Java based single user DBMS, intended primarily for use in education and teaching. Research provides some basic knowledge of the system at a surface level, single user with no concurrency control or recovery support. It is under development by the University of Konstanz and based on the Minibase of Java written by Chris Mayfield, which was an extended and redesigned version of the initial Java port of the first C/C++ version of Minibase developed by Raghu Ramakrishnan to accompany the book Database Management Systems.

The buffer manager is responsible in this system for a few primary tasks. First, the database buffer is temporary storage in memory that is quick to access, but very limited in size. The manager is responsible for allocating space in the buffer to store data, providing block addresses when they are available and moving the block to the buffer if it is not. When there is no free space in the buffer the manager will use one of a series of replacement algorithms to decide which blocks are replaced in the buffer; least recently used(LRU) is one such common algorithm often used for buffer management. The buffer is also a key component in the systems recovery and concurrency of data in the event of a system crash.

The disk space manager works closest to the operating system of all the DBMS components since it is responsible for the management, creation, and manipulation of disk space to write the data for the database. In databases the concept of a “page” is created as a unit of data. Generally, the page size as less than that of the block size mentioned before to facilitate retrieval in one disk I/O operation. The disk manager handles the allocation of disk space, as well as certain standard practices for data storage, such as storing data accessed sequentially in contiguous block sequences. In terms of continued use, the disk manager handles the manipulation of the blocks and pages as data is created and removed from the database.

Heap file organization, at the most basic level means that files are unordered and inserted at the file’s end. Records are not sorted or ordered. When a data block is full in heap file organization the new record is stored in a different block, and due to the structure, this does not have to be the next sequential block. While this works fine for small databases, such as those that Minibase mostly handles, for large databases this quickly becomes extremely inefficient; while insertion is not an extraordinary issue retrieving datafiles and records is time consuming as the entire file must be searched from the beginning for the record’s unique id.

Minibase does utilize a B+ tree structure which allows for efficient insertion, deletion, and search operations. Extending the functionality of a B Tree, the B+ Tree stores records exclusively in the leaf nodes of internal nodes which contain key values. Like a directory at the mall that tells you what each section of the mall caters to. If you are shopping for clothes at Kohls, the directory tells you where you would find Kohls, but not the specific location of Kohls. Like this, a B+ Tree’s internal node stores key values with pointers to the leaf nodes containing the actual records. So if you were to search for the key value 55, and your internal nodes contains the values 50 and 75, the DBMS would search the branch that exists between the 50 and 75 key nodes to find the leaf which contains that record. Insertion and deletion require some manipulation of tree structure in certain cases and this is done to maintain a key B Tree property; that a B Tree is always balanced.

**Description of Tests**

**Buffer Manager Test(BMTest):**

There are three buffer manager tests included in the automated tests of Minibase. Test 1 tests simple operations as described by the system. Analyzing the test shows that the first test allocates a number of pages equal to the number of unpinned buffers plus one, so that it can test the unpinning of all buffers, writing new data, and then unpinning the ‘dirty’ buffers to write back to the disk. It then finishes operations by freeing all pages that were allocated for the test.

Test 2 expects failures by trying to perform illegal operations on the buffer, including loading, or pinning, more pages to the buffer than exist free buffers. This should fail as there are not enough locations in the buffer to hold those pages. After testing the buffer bounds this test also attempts to free a pinned page with the expectation of failure as the DBMS should prevent a page that has not been written back to disk and unpinned from being freed. The final portion of this test is to try and unpin a page not held in the buffer pool; this should fail as pages that are pinned are defined as being pages that are in use and any pages not in the buffer are not in use. Operations of the test are then finished, and the system cleaned up for the next test.

Test 3 begins with a test of internals by first allocating new pages and altering them in the buffer so that they are ‘dirty’; the system then attempts to unpin the pages left pinned and with dirty bits. This should take these pages and overwrite them on the disk with there new modified value. The second part of this test attempts to read data back from pages and ensure that the data is correct before unpinning the pages and completing the test.

**Disk Space Management Tests(DBTest):**

The tests performed for Disk Space Management build off the previous ones by generally testing in these ways. Test 1 is a test of allocating pages and files, writing to some of them, and then deallocating some pages.

Test 2 uses the database created in the previous test and performs deletions of some file entries before then looking up the file entries that should remain. Test 2 ends by reading data written to the pages in test 1.

Test 3 tests for error conditions in the disk manager, beginning with searching for deleted file entries. Next the test tries to delete entries already removed, and entries that did not exist in the database at any time. There are also testing of the look up for nonexistent files, the addition of already existing files, and boundary conditions on page allocation in terms of both name length and page runs.

Test 4 also tests boundary conditions in a very implementation specific way, beginning with attempting to allocate more pages than is space available on the disk. The tests continue to manipulate the number of entries and pages to test the boundary conditions of the systems disk space manager until another final check is done to ensure that no allocation of pages beyond those available is allowed; this includes testing boundaries at the maximum and minimum capacity on the disk.

**Heap File Tests(HFTest):**

The heap file is tested in this test, beginning with test 1 which performs some standard heap file operations. A heap file is first created with a set size before records are added to the file. This test ends after scanning the records and ensuring that they were all created. Test 2 begins by opening the file created prior and deleting half of the records before scanning the remaining records on the file. Test 3 finally tests the manipulation of data on the heap file by changing the records stored on the heap file and scanning to ensure that the record values were updated.

Test 4 tests error conditions for the heap file by attempting to change the size of the records without adding or deleting records; the records are first shortened, then lengthened. The final test is then ended after the system checks the error condition for record length by attempting to insert a record that is too long.

**B+ Tree Test(BTTest):**

The B+ Tree Test is the only test which required interaction by the user, this was done as the B+ tree is visualized by Minibase, and the user is allowed to determine the number of records to allocate and delete, while giving them experience with the concept of B+ trees. Test 1 in the system allows the user to insert **n** records in order, and this allocates n records beginning at 1 and iterating to **n**. Test 2 and test 3 follow the same format as test 1 but allocate from **n** to 1, and at random in the range (1..**n**) instead. These tests are meant to show the B+ tree format as the number of records are increased and to test and ensure proper sorting and ordering within the data structure. Test 4 and 5 are also similar in that they randomly add **n** records and randomly delete **m** records; ensuring proper addition and deletion of records which also test the reallocation of pages as the number of records are mutated. The significant difference between these two tests being that test 5 utilizes string-based keys as compared to the integer keys used previously. The utilization of string-based keys changes the way in which the records should be sorted and tests the DBMS management of keys of this type.

**Index Test(IndexTest):**

The index tests are simple tests to ensure that the DBMS is properly indexing records and data. Test 1 checks that the index scan returns the index sorted in order, while also ensuring that there are the proper number of records existing in the database. Throwing an erro if too few or too many records are found. Test 2 scans for specific indexes and a range of indexes to ensure that both functionalities are properly scanning the database. Test 3 does an index scan on integer keys to ensure that these are also properly found.

**Join and Sort-Merge Tests( JoinTest & SM\_JoinTest):**

The join tests are defined as a set of queries, and as are the sort-merge tests. In the case of the join tests, there exists sailors, boats, and reserves groups and these are joined in the same database. The queries search this joined database and checks that all expected values are returned. Query 1 searches for all sailors who have reserved boat number 1 printing out their reservation dates, query 2 searches for all sailors who had reserved a boat that was red and returns them in alphabetical order. Queries 3 and 4 both search for all sailors who have reserved a boat, however query 4 prints that names of each unique sailor. Query 5 finds the names of old sailors and those with poor ratings, a score under 7, from the database. Query 6 is the last query in the join test and finds the name of all sailors who have reserved a red boat and have a rating greater than 7.

The queries for the sort-merge tests are identical to those of join, but for two exceptions; the sort-merge test does not implement queries 2 or 6.

**Sort Tests(SortTest):**

The sort tests are used to check the sorting capabilities and algorithms of the DBMS. The first two tests sort the records and run a check to ensure that the sorted records are the same as the expected records order. Test 3 sorts in descending order the float field, and in ascending order the int field. Test 4 tests the sorting algorithm on multiple files rather than the previous tests which only sorted one file.

**Conclusion**

Phase 1 of the project covered specific uses of the Minibase DBMS and tests for those specific functionalities and components of a DBMS; these included the buffer manager, disk manager, heap file management, sorting functions, and both join and sort-merge join functionalities. Beyond merely running these tests and familiarizing the user with the interface of Minibase and its visualization functions, this phase also introduced major topics and ideas that are included in the standard DBMS and that are important and key topics to understand when dealing with databases. It forced research into topics that while specific to Minibase and this project, also are of importance to dealing with databases in general. The research done and study done when viewing the various tests also served to present the limitations and capabilities of the various features of not only Minibase, but DBMS software in general.

**Bibliography**

References

Authors, Title, Publication Venue, Volume, Pages, Year

JavaTPoint, B+ Tree, JavaTPoint, 2018

https://www.javatpoint.com/b-plus-tree/

JavaTPoint, DBMS B+ Tree, JavaTPoint, 2018

https://www.javatpoint.com/dbms-b-plus-tree/

Sagnik Adhikary, Introduction of B Tree, GeeksForGeeks, 2020

https://www.geeksforgeeks.org/introduction-of-b-tree/

JavaTPoint, DBMS Heap file organization, JavaTPoint, 2018

https://www.javatpoint.com/dbms-heap-file-organization/

JavaTPoint, Database Buffer, JavaTPoint, 2018

https://www.javatpoint.com/database-buffer/

DBMSInternals, Disk Space Management And Pages, DBMSInternals, 2020

<http://www.dbmsinternals.com/database-fundamentals/data-storage/disk-space-management-pages/>