**INSTALLATION**

Extract .zip and add as a project in Eclipse. Run the server startup as usual. Performance tests can be run from simpledb.datagenerator. Otherwise, the default studentClient database functions just as well.

**TESTING QUERIES**

The test queries from the previous project (pokemonTest.simpledb.\*, studentClient.simpledb.\*) as well as the performance testing queries were used in order to test whether the database still functioned properly.

To ensure that the Extensible Hashing index worked properly, a toString() method, in conjunction with various print statements throughout, print information during each transaction. A modifed version of the provided test program was run with the output as follows:

new transaction: 1

creating new database

...

new transaction: 18

Opening Extensible Hash Index

EH: INITIALIZING BUCKETS

EH: UPDATING BITMASK TO 2

EH: SEARCHKEY = 985, BUCKET ID = 1

EH: BUCKET 1 TOTAL = 0

EH: UPDATING BITMASK TO 2

EH: ADDING 985 TO BUCKET 1

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

BUCKETS (GlobalDepth = 2):

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

0: LocalDepth = 2

-----------

============

1: LocalDepth = 2

-----------

985

============

2: LocalDepth = 2

-----------

============

3: LocalDepth = 2

-----------

============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

transaction 18 committed

new transaction: 19

Opening Extensible Hash Index

EH: UPDATING BITMASK TO 2

EH: SEARCHKEY = 847, BUCKET ID = 3

EH: BUCKET 3 TOTAL = 0

EH: UPDATING BITMASK TO 2

EH: ADDING 847 TO BUCKET 3

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

BUCKETS (GlobalDepth = 2):

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

0: LocalDepth = 2

-----------

============

1: LocalDepth = 2

-----------

985

============

2: LocalDepth = 2

-----------

============

3: LocalDepth = 2

-----------

847

============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

transaction 19 committed

... // Ommitted several transactions for space

transaction 25 committed

new transaction: 26

Opening Extensible Hash Index

EH: UPDATING BITMASK TO 2

EH: SEARCHKEY = 592, BUCKET ID = 0

EH: BUCKET 0 TOTAL = 0

EH: UPDATING BITMASK TO 2

EH: ADDING 592 TO BUCKET 0

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

BUCKETS (GlobalDepth = 5):

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

0: LocalDepth = 2

-----------

592

============

1: LocalDepth = 3

-----------

569 985

============

2: LocalDepth = 5

-----------

434

============

18: LocalDepth = 5

-----------

978 562

============

3: LocalDepth = 2

-----------

847

============

5: LocalDepth = 3

-----------

317

============

6: LocalDepth = 3

-----------

254

============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

transaction 26 committed

new transaction: 27

Opening Extensible Hash Index

EH: UPDATING BITMASK TO 2

EH: SEARCHKEY = 596, BUCKET ID = 0

EH: BUCKET 0 TOTAL = 1

EH: UPDATING BITMASK TO 2

EH: ADDING 596 TO BUCKET 0

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

BUCKETS (GlobalDepth = 5):

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

0: LocalDepth = 2

-----------

592 596

============

1: LocalDepth = 3

-----------

569 985

============

2: LocalDepth = 5

-----------

434

============

18: LocalDepth = 5

-----------

978 562

============

3: LocalDepth = 2

-----------

847

============

5: LocalDepth = 3

-----------

317

============

6: LocalDepth = 3

-----------

254

============

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**DESIGN**

Index

The Index data structure was updated throughout simpleDB in order to accommodate the index type based on user input; “bt” and “eh” will create B-tree and extensible hash indices respectively, and any other input (or lack thereof) will make a static hash index.

Extensible Hashing

The Extensible Hash index (simpledb.index.exhash) expands on the functionality of the original hash index class. Initially, 4 buckets are created and the global depth is set to 2 (2-bit checks). Buckets are now a new class rather than an abstract integer and hold an ID number, local depth, contents, and capacity. The contents can not be larger than 4. Buckets are held in a hashmap to match ID numbers to buckets during hashing.

A bitmask is used to determine the bucket to send a record to (2 => b11, 3 => b111, etc). For each record, the bitmask is initalized based on the global depth for safety and a bucket number is generated with the search key's hash and the bitmask. If the local depth of this bucket is not equal to the global depth, the bitmask and bucket number are updated before continuing. This is how our design handles multiple local depths; only the necessary number of bits are checked.

If a bucket is not full, the record is added to its contents and the index is updated. Otherwise, the local depth (and the global depth if need be) are incremented. The hashed, unmasked value of the record, as well as this bucket, are then passed to a bucket updater for splitting and proper hashing.

The bucket updater first applies the bitmask to the hashed value, then uses this to make a new bucket. The contents of the old bucket are removed and rehashed to update these two buckets, and finally, the record that caused this mess is put into its proper bucket and the index is updated.

SmartMergeJoin

ExploitSortQueryPlanner

**TESTING**

Performance tests were run with 5,000 queries per table (anything higher did not complete after several hours, despite running on a very powerful machine).

Extensible Hashing

SmartMergeJoin

**BUGS**

We are unaware of any bugs in this application. However, we are not sure if our IO counter for select statements works as intended.