

## DB Assignment 6

Brittany Klose

12/07/24

### Tasks

- 1) Similar to what we did in class, use a stored procedure (generate\_accounts)
  - 2) Create timing analysis and populate the table
  - 3) Create **indexes** on the `branch_name` and `account_type` columns.
  - 4) Compare **point queries** and **range queries**.
  - 5) Experiment with the following dataset sizes: 50K, 100K, 150K
- 6) Analyze the results of your timing tests and compare how the execution times change when indexes are applied. Discuss any performance differences in your report.
- My results overall were not surprising. The inclusion of indices proved to lower execution time with every test. There was an especially dramatic difference in execution times between index vs no index for the data sets of size 150k. This was most likely as a result to also not including the primary index; which I dropped because I kept receiving an error that a primary key duplicated after the record 100,000. I decided to simply remove the index as well to get the code to run. However, based on the difference in execution time between 50k and 100k datasizes without indexes, the dataset of 150k without even the primary index was not that much greater than what I was expecting. What was surprising was that my lowest execution time overall was, not just a query including indices, but the one with a data size of 150k records. I'm not sure what could be attributed to this outlier. Even with indexes added in I still didn't include the primary key for this larger dataset due to the error. While I expected time to drop I did not expect it to be the fastest run. Other than this exception my results pointed to lower run times when indexes are included, data size is smaller, and if it's a range query over a point query. I was not surprised that range queries operated somewhat

faster as these queries will have multiple results whereas the point queries are searching for one specific case and that could take longer to find.

8) Summarize the results of the timing experiments

Query Type	Description	Dataset Size	Index Type	Execution time (Microseconds)
Point Query 1	SELECT count(*) FROM accounts WHERE branch_name = 'Redwood' AND account_type = 'Checking';	50k records	Without Indexes	38,874
			With index Branch_name, Account_type	16,358
Point Query 2	SELECT count(*) FROM accounts WHERE branch_name = 'Perryridge' AND account_type = 'Savings'; = 'Perryridge' AND balance = 10000;	100k records	Without Indexes	63,198
			With index Branch_name, Account_type	20,545
Point Query 3	SELECT count(*) FROM accounts WHERE branch_name = 'Mianus' AND balance = 90000;	150k records	Without Indexes (without primary index as well)	85,032
			With index Branch_name, balance	13,010
Range Query 1	SELECT count(*) FROM accounts WHERE branch_name	50k records	Without Indexes	35,132

	= 'Mianus' AND balance BETWEEN 90000 AND 70000;			
			With index Branch_name, balance	11,388
Range Query 2	SELECT count(*) FROM accounts WHERE branch_name = 'Brighton' AND balance BETWEEN 30000 AND 10000;	100k records	Without Indexes	55,304
			With index Branch_name, balance	11,245
Range Query 3	SELECT count(*) FROM accounts WHERE branch_name = 'Downtown' AND balance BETWEEN 10000 AND 5000;	150k records	Without Indexes (without primary index as well)	81,002
			With index Branch_name, balance	8,451

## 7) Create a procedure for average execution time

- Runs each query 10 times, calculates the sum of execution times and the average.

Query	Total Time	Average Time
SELECT count(*) FROM accounts WHERE branch_name = 'Redwood' AND account_type = 'Checking';	184810	18481
SELECT count(*) FROM accounts WHERE branch_name = 'Perryridge' AND account_type = 'Savings'; = 'Perryridge' AND balance = 10000;	194322	19432.2

SELECT count(*) FROM accounts WHERE branch_name = 'Mianus' AND balance = 90000;	195426	19542.6
SELECT count(*) FROM accounts WHERE branch_name = 'Mianus' AND balance BETWEEN 90000 AND 70000;	193937	19393.7
SELECT count(*) FROM accounts WHERE branch_name = 'Brighton' AND balance BETWEEN 30000 AND 10000;	197680	19768.0
SELECT count(*) FROM accounts WHERE branch_name = 'Downtown' AND balance BETWEEN 10000 AND 5000;	193398	19339.8

### **Extra Credit**

- Plot the timing results for each query. Represent execution times (index vs. no index) on the y-axis and num. of records on the x-axis
- View extra credit jupyter notebook file