

Database Design

Database Design

- Binary Format
- Store (Origin, Dest, ArrDelay) ^{one after another} tuples
- Ignore other unnecessary attributes
- e.g. (AAA, BBB, 1) becomes :

0x ^{origin}41414 ^{dest}14242420 ^{ArrDelay}0000000

$\text{sizeof(char)} \times 3$ $\text{sizeof(char)} \times 3$ $\text{sizeof(int)} \times 1$

Index Design

Index Design

- Hash Index
- $h: (\text{Origin}, \text{Dest}) \rightarrow \{ \text{position in DB that matches} \}$
- Use `unordered_map<string, vector<long>>` to implement a set of positions
- Key: Simply concatenate origin and Dest
- Values: Point to the first byte of ArrDelay
↓
Already sorted, because we build the index in order

Result w/ Index

```
~/Documents/php-guide/rdbms/practice4 > master make
rm test
g++ -Wall -Wextra -Wpedantic -std=c++11 main.cpp db.cpp -o test
~/Documents/php-guide/rdbms/practice4 > master ./test
Import Time: 20.652s
Index Time: 9.951s

##### IAH -> JFK #####
Average Arrival Delay: 17.20762
Query Time: 0.002s

##### IAH -> LAX #####
Average Arrival Delay: 11.47905
Query Time: 0.003s

##### JFK -> LAX #####
Average Arrival Delay: 11.26123
Query Time: 0.005s

##### JFK -> IAH #####
Average Arrival Delay: 12.64593
Query Time: 0.002s

##### LAX -> IAH #####
Average Arrival Delay: 7.19692
Query Time: 0.003s

##### ATL -> MCI #####
Average Arrival Delay: 9.45637
Query Time: 0.005s
```


Result w/ Index

```
Command being timed: "./test"
User time (seconds): 30.66
System time (seconds): 2.55
Percent of CPU this job got: 81%
Elapsed (wall clock) time (h:mm:ss or m:ss): 0:40.99
Average shared text size (kbytes): 0
Average unshared data size (kbytes): 0
Average stack size (kbytes): 0
Average total size (kbytes): 0
Maximum resident set size (kbytes): 214304
Average resident set size (kbytes): 0
Major (requiring I/O) page faults: 0
Minor (reclaiming a frame) page faults: 55816
Voluntary context switches: 5158
Involuntary context switches: 273
Swaps: 0
File system inputs: 1351648
File system outputs: 412776
Socket messages sent: 0
Socket messages received: 0
Signals delivered: 0
Page size (bytes): 4096
Exit status: 0
```

Result w/o Index

```
~/Documents/php-guide/rdbms/practice4 > master > make
rm test
g++ -Wall -Wextra -Wpedantic -std=c++11 main.cpp db.cpp -o test
~/Documents/php-guide/rdbms/practice4 > master > ./test
Import Time: 19.709s
Index Time: 0.000s

##### IAH -> JFK #####
Average Arrival Delay: 17.20762
Query Time: 3.129s

##### IAH -> LAX #####
Average Arrival Delay: 11.47905
Query Time: 3.202s

##### JFK -> LAX #####
Average Arrival Delay: 11.26123
Query Time: 3.118s

##### JFK -> IAH #####
Average Arrival Delay: 12.64593
Query Time: 3.196s

##### LAX -> IAH #####
Average Arrival Delay: 7.19692
Query Time: 3.125s

##### ATL -> MCI #####
Average Arrival Delay: 9.45637
Query Time: 3.206s
```

Result w/o Index

```
Command being timed: "./test"
User time (seconds): 32.35
System time (seconds): 8.23
Percent of CPU this job got: 78%
Elapsed (wall clock) time (h:mm:ss or m:ss): 0:52.02
Average shared text size (kbytes): 0
Average unshared data size (kbytes): 0
Average stack size (kbytes): 0
Average total size (kbytes): 0
Maximum resident set size (kbytes): 2796
Average resident set size (kbytes): 0
Major (requiring I/O) page faults: 0
Minor (reclaiming a frame) page faults: 132
Voluntary context switches: 5528
Involuntary context switches: 425
Swaps: 0
File system inputs: 1581296
File system outputs: 412784
Socket messages sent: 0
Socket messages received: 0
Signals delivered: 0
Page size (bytes): 4096
Exit status: 0
```