

INF-2700 Exercise 4

15.09.2022

1 C gdb

This part helps students debug with `gdb`.

- In directory `db2700/`, run `gdb ./run_test`
- In the `gdb` session:
 - `h`
get help
 - `b get_page_record`
to set a breakpoint
 - `r -md`
to run the program at the DEBUG message level
 - `l`
to list the source code at the break point
 - `bt`
to backtrace
 - `p s`
to print the `s` value
 - `p r`
to print the `r` value
 - `call put_schema_info (BEDUG, s)`
to display `s` in a better way
 - `n` several times to step forward

- call `put_record_info (DEBUG, r, s)`
- and so on
- `man gdb`

2 Relational algebra and SQL

Work with question 1 in the 2011 re-exam.

3 Data on disk

Review of some key points:

- Disk IO is the most important factor of database performance.
- Disk blocks are the smallest unit we can address and access data on disk.
- Random disk access costs much more than sequential disk access.

Recall that the following terms mean more or less the same things in different context (sometimes we use different terms even in the same context):

relational algebra	SQL	physical data
relation	table schema	file descriptor
relation instance	table	file
tuple	row	record
attribute	attribute	field

4 Database file organization —Ordered files

For an *ordered file* (ordered on a given *search key*), we can find a record with $\lg N$ disk reads (where N is the file size in terms of number of disk blocks). We can use binary search. Beware that disk access is random in binary search.

Can you achieve a record insertion with the following (Big-O) complexity?

(in terms of number of *disk writes*, i.e. after you have located the location for the insertion)

- N^2

- $N \lg(N)$
- N
- $\lg N$
- 1