

CS 387: Outlab 4

Nimay Gupta - 180050068

Shalabh Gupta - 180050096

Exercise 1:

Schema

```
create table COVID
(
  cdc_report_dt      date,
  pos_spec_dt        date,
  onset_dt           date,
  current_status      text,
  sex                text,
  age_group           text,
  "Race and ethnicity (combined)" text,
  hosp_yn            text,
  icu_yn             text,
  death_yn           text,
  medcond_yn         text
);
```

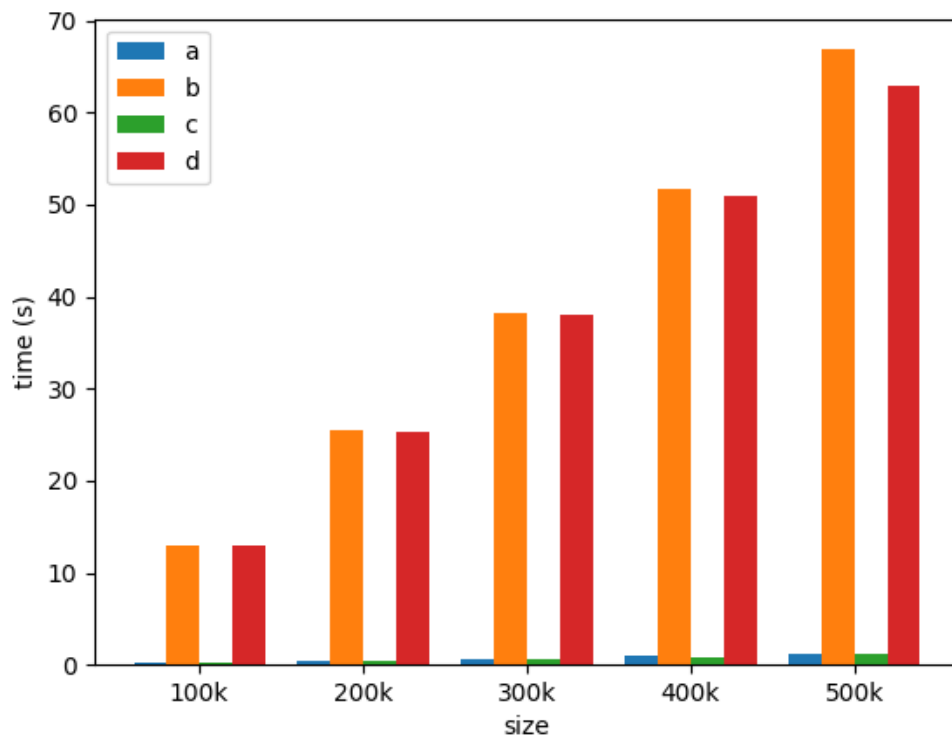
Machine configuration

- CPUs: 16
- RAM: 16 GB

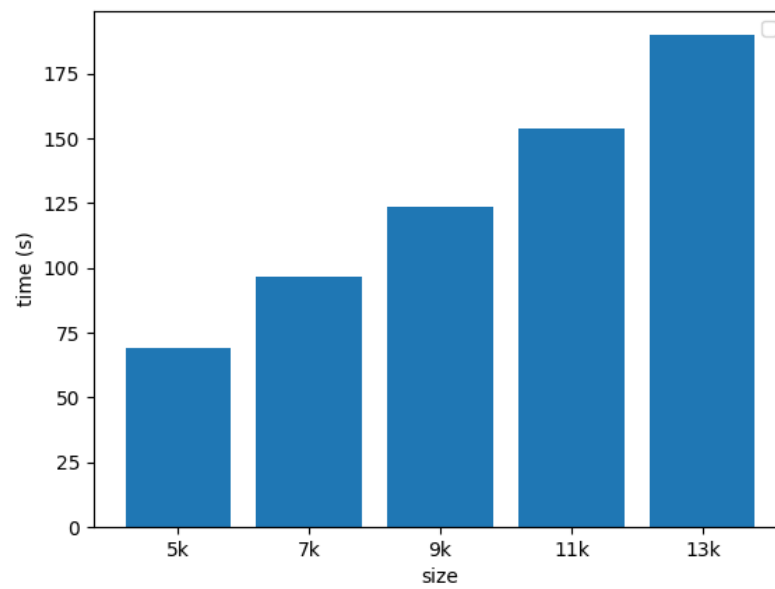
Methods of loading:

- a) Enabling \timing to benchmark and using the “copy from ‘file’ csv header” query in psql
- b) By generating a .sql file containing INSERT queries using python
bash command used to run and time:
time PGPASSWORD=<password> psql -qU postgres -d <db_name> -h 127.0.0.1 -p 5432 -w -f <filename> -q
- c) Bulk loading using copy_expert method of psycopg2 cursors
- d) Creating insert queries from csv files using python and reusing the connection for each execution of insert query
- e) Creating insert queries from csv files using python and creating fresh connection for each execution of insert query

Plot for a-d



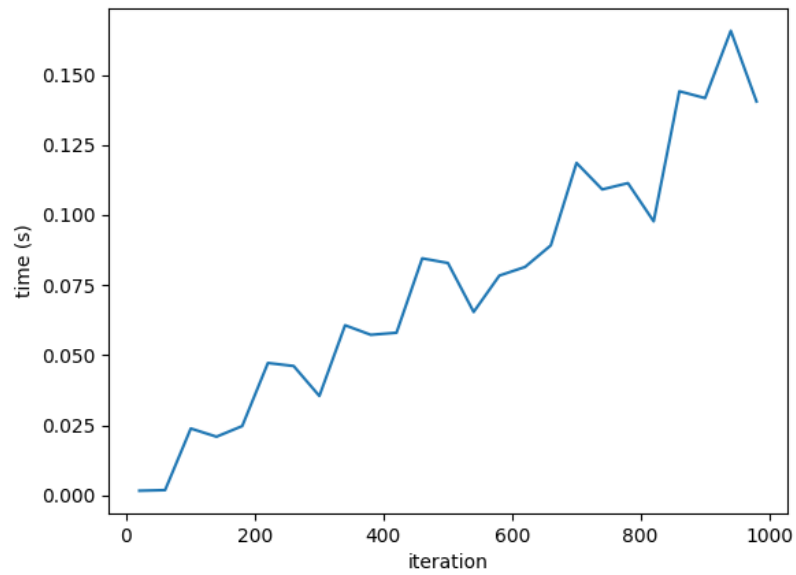
Plot for e



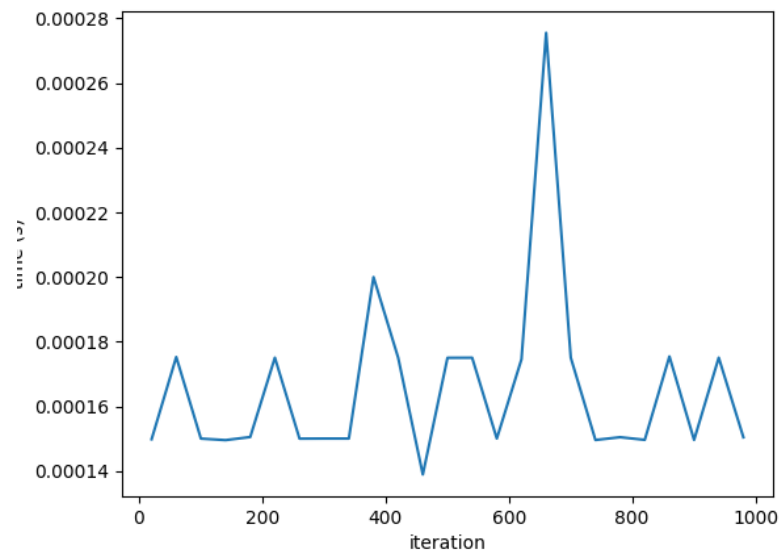
Exercise 2:

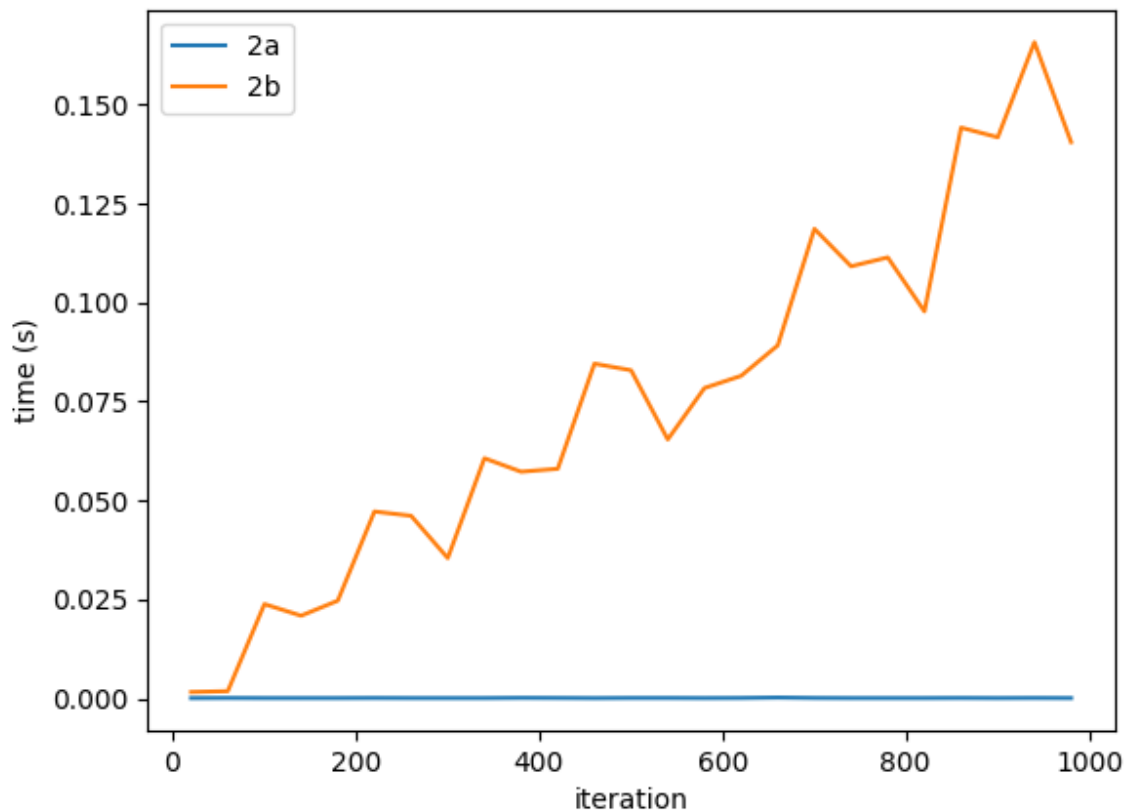
- Time taken by task 2a: 21.375s
- Time taken by task 2b: 1.216s

Plot for 2a



Plot for 2b





The cursor acts as a pointer to the next record we wish to start querying from. While getting the next page of results, the rows that we have already seen are removed. But offset style of querying simply specifies the number of records you wish to skip before selecting records which is slower because every time we read rows up to offset to select the required data.