

Benchmark baz danych

Aleks Zieliński, Filip Kalinowski

June 18, 2025

Rozdziały

1	Informacje wstępne	2
1.1	Założenia projektu	2
1.2	Użyte technologie i sprzęt	2
1.3	Schemat logiczny bazy danych	3
1.4	Przykładowe dane z plików JSON	3
2	Instalacja baz danych	4
2.1	PostgreSQL	4
2.2	MongoDB	6
3	Benchmark	8
3.1	Zapytania	8
3.2	Zapytania w PostgreSQL	8
3.3	Zapytania w PostgreSQL w JSONB	8
3.4	Zapytania w MongoDB	9
3.5	Zapytanie xyz	9
3.6	Zapytanie abc	9

1 Informacje wstępne

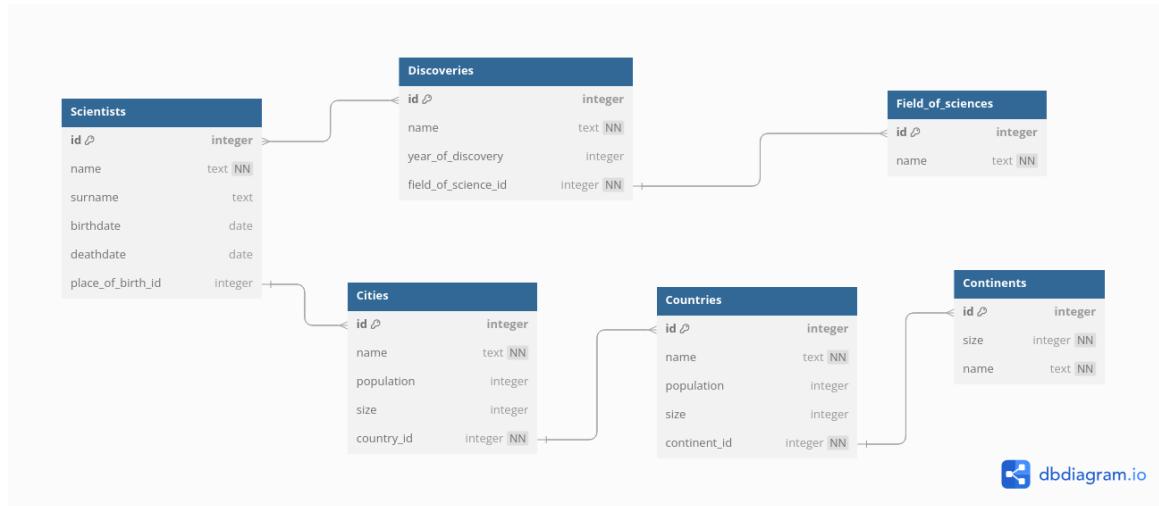
1.1 Założenia projektu

- Zaprojektować schemat dokumentów JSON dla wybranego tematu. Należy mieć min 2 typy dokumentów na osobę w projekcie. Minimum jeden typ musi mieć dokumenty zagnieżdżone
- Opracowany model, wraz z propozycją najważniejszych/najbardziej rozbudowanych poleceń należy zatwierdzić u wykładowcy (ew. poprawić wg. sugestii)
- Dla każdego typu dokumentu wygenerować min 100 sensownych instancji. Jeden z typów powinien mieć minimum 500 instancji. Dane można generować automatami typu generatedata.com bądź napisać 4 skrypt w pythonie (dodatkowe punkty).
- Tak przygotowane dokumenty należy zainportować do:
 1. Wybranej bazy NoSQL
 2. PostgreSQLa do kolumn typu JSONB
 3. PostgreSQLa przy jednoczesnej konwersji JSON na tabele
- Wszystkie 3 procesy importowania należy udokumentować screenshotami i zapisanymi wykorzystanymi poleceniami
- Należy utworzyć min 4 zapytania na osobę realizujące najważniejsze problemy wyszukiwania w tworzonej bazie.
- Każde z zapytań musi mieć 3 wersje dla 3 sytuacji z pkt 4.
- Należy przeprowadzić eksperyment pomiaru czasu wykonania tych zapytań. Aby eksperyment był rzetelnie przeprowadzony, zarówno baza NoSQLowa, jak i PostgreSQL muszą operować w podobnych warunkach: albo obie bazy są postawione w osobnych dockerach, albo zainstalowane w tym samym OS, ale uruchamiane jedna na raz (proces instalacji musi być udokumentowany). Wyniki czasowe należy przedstawić w tabelce

1.2 Użyte technologie i sprzęt

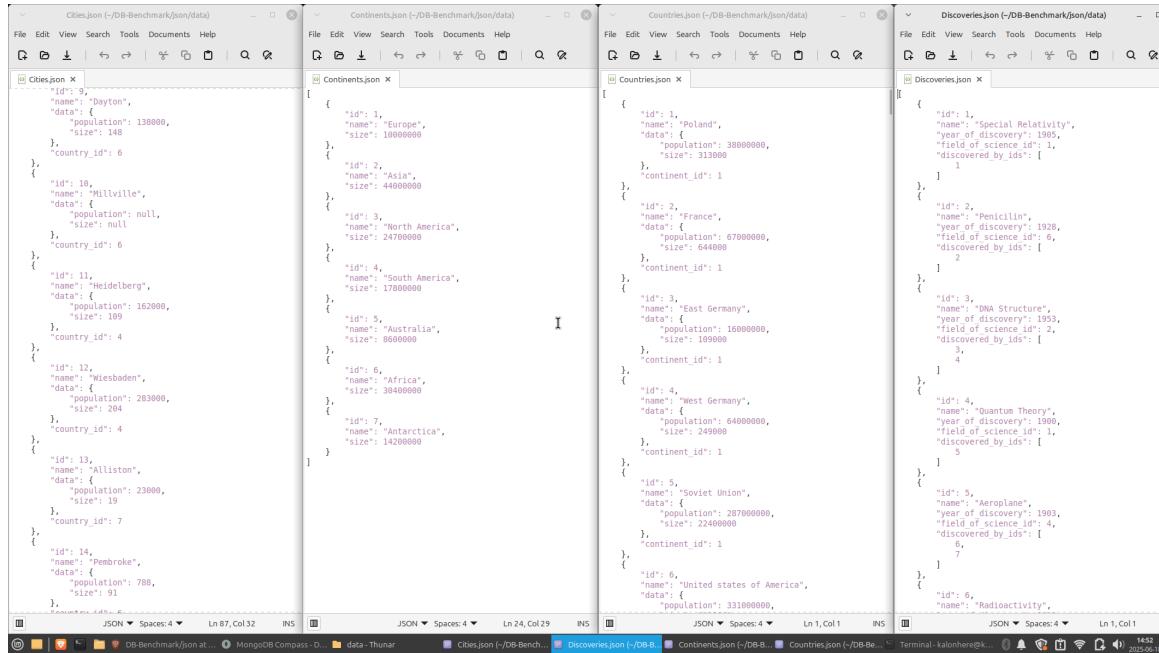
- PostgreSQL - wersja 16.9
- MongoDB - wersja 8.0.10
- Laptop:
 - System operacyjny - Linux Mint 22.1
 - CPU - Intel i7-9750H
 - GPU - Nvidia GTX 1660
 - RAM - 16GB 2667 MHz

1.3 Schemat logiczny bazy danych



Zdjęcie. 1: Schemat logiczny zrobiony w serwisie dbdiagram.io

1.4 Przykładowe dane z plików JSON



Zdjęcie. 2: Pliki JSON od lewej: Cities, Continents, Countries, Discoveries

Zdjęcie. 3: Pliki JSON od lewej: Field_of_sciences, Scientists

2 Instalacja baz danych

2.1 PostgreSQL

Zdjęcie. 4: Instalacja za pomocą apt

```

fixing permissions on existing directory /var/lib/postgresql/16/main ... ok
creating subdirectories ... ok
selecting dynamic shared memory implementation ... posix
selecting default max_connections ... 100
selecting default shared_buffers ... 128MB
selecting default time zone ... Europe/Warsaw
creating configuration files ... ok
running bootstrap script ... ok
performing post-bootstrap initialization ... ok
syncing data to disk ... ok
Getting up postgresql-contrib (16+257build1.1) ...
Setting up postgresql (16+257build1.1) ...
Processing triggers for man-db (2.12.0-4ubuntu2) ...
Processing triggers for libc-bin (2.39-0ubuntu0.4) ...
kalonhere@kalonhere-laptok:~$ sudo systemctl status postgresql
● postgresql.service - PostgreSQL RDBMS
   Loaded: loaded (/usr/lib/systemd/system/postgresql.service; enabled; prese>
   Active: active (exited) since Tue 2025-06-17 13:42:44 CEST; 11s ago
     Main PID: 8053 (code=exited, status=0/SUCCESS)
       CPU: 999us

Jun 17 13:42:44 kalonhere-laptok systemd[1]: Starting postgresql.service - Post>
Jun 17 13:42:44 kalonhere-laptok systemd[1]: Finished postgresql.service - Post>
Lines 1-8/8 (END)

```

```

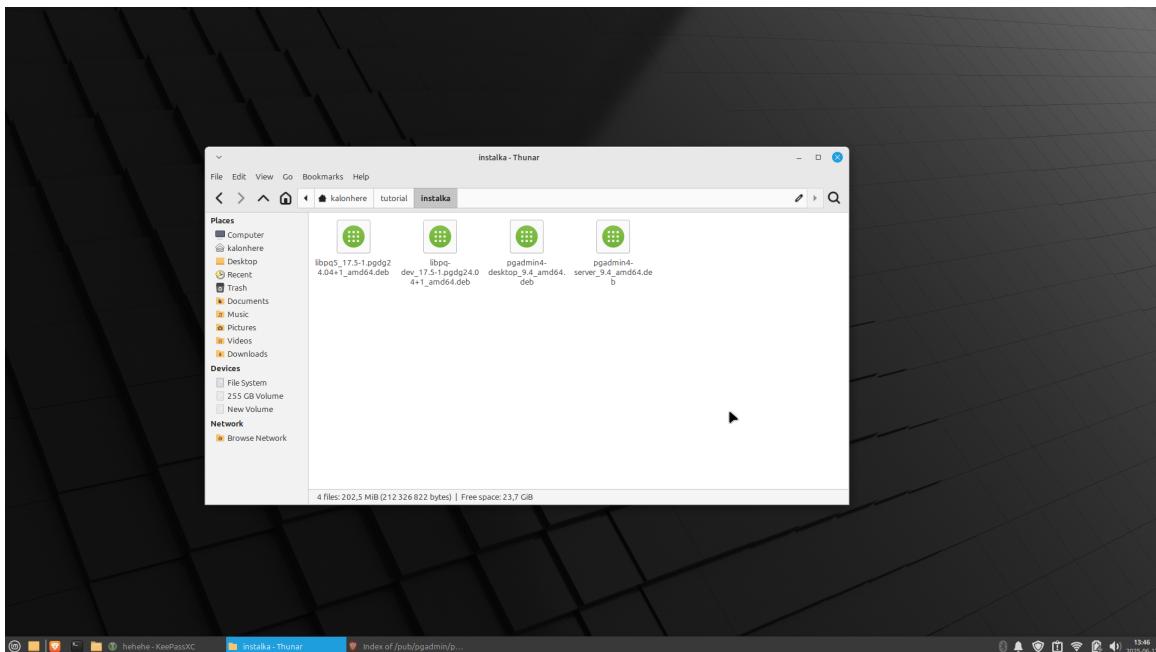
Terminal - kalonhere@kalonhere-laptok:~$ sudo -i -u postgres
[sudo] password for kalonhere:
kalonhere@kalonhere-laptok:~$ psql
psql (16.9 (Ubuntu 16.9-0ubuntu0.24.04.1))
Type "help" for help.

postgres=# select 'Pozdrawiamy' as essa;
      essa
-----
Pozdrawiamy
(1 row)

postgres# 

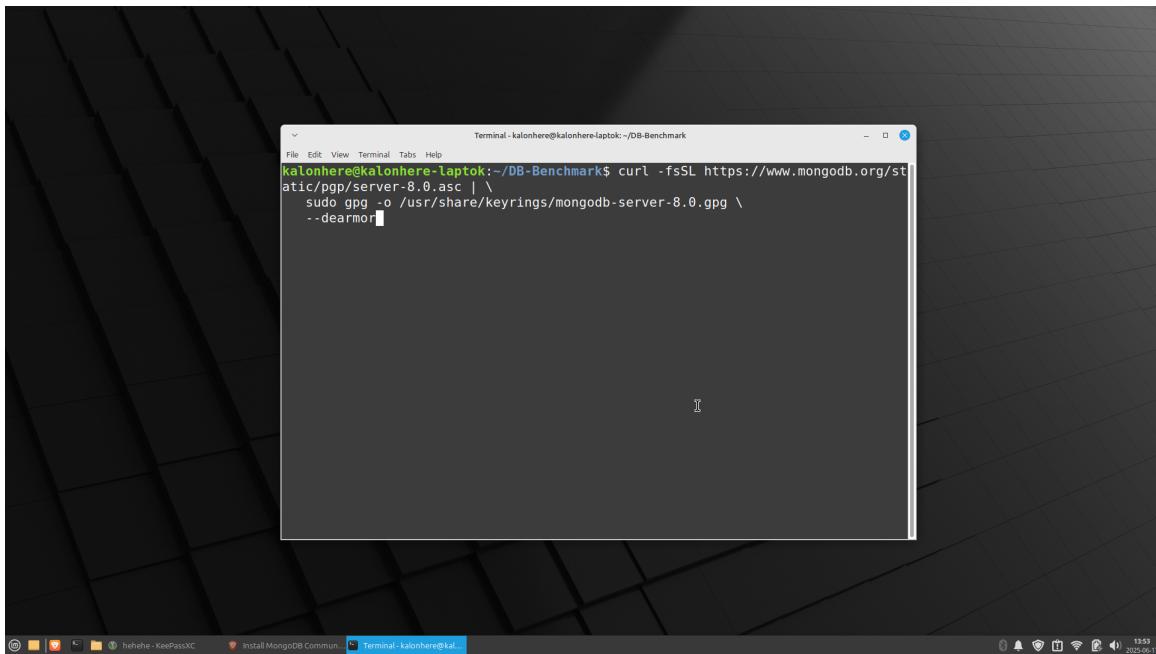
```

Zdjęcie. 5: Po lewej - komenda `systemctl` aby sprawdzić czy działa, po prawej - logowanie na user, uruchomienie `psql` i test działania

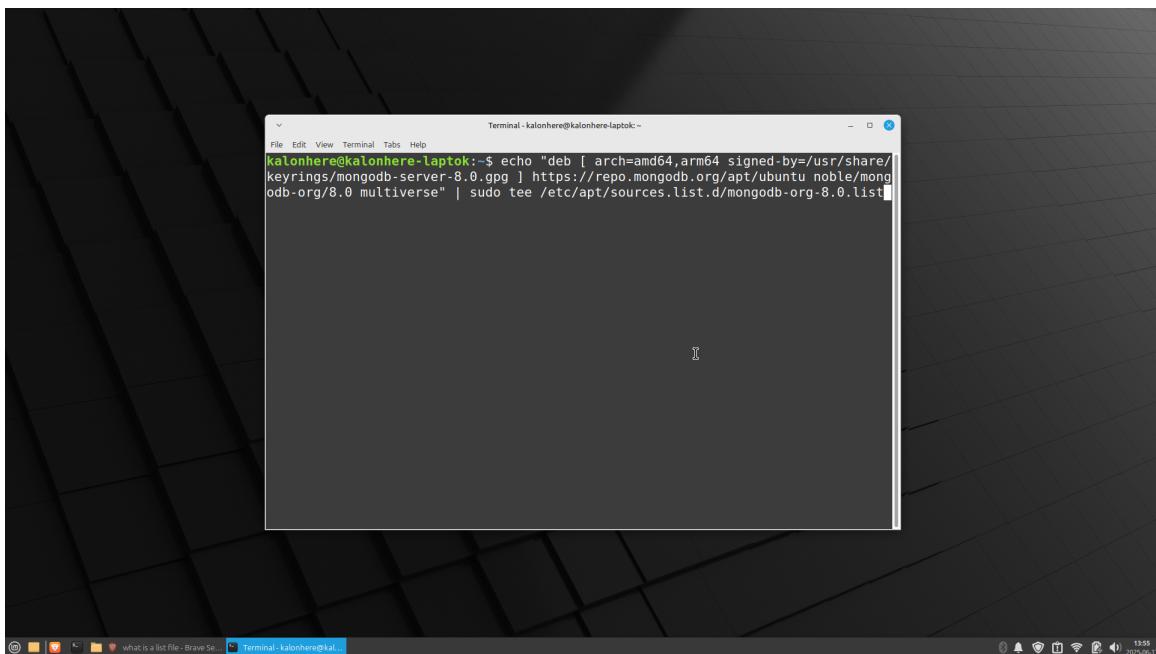


Zdjęcie. 6: Instalacja pgadmina z serwera ftp PostgreSQL

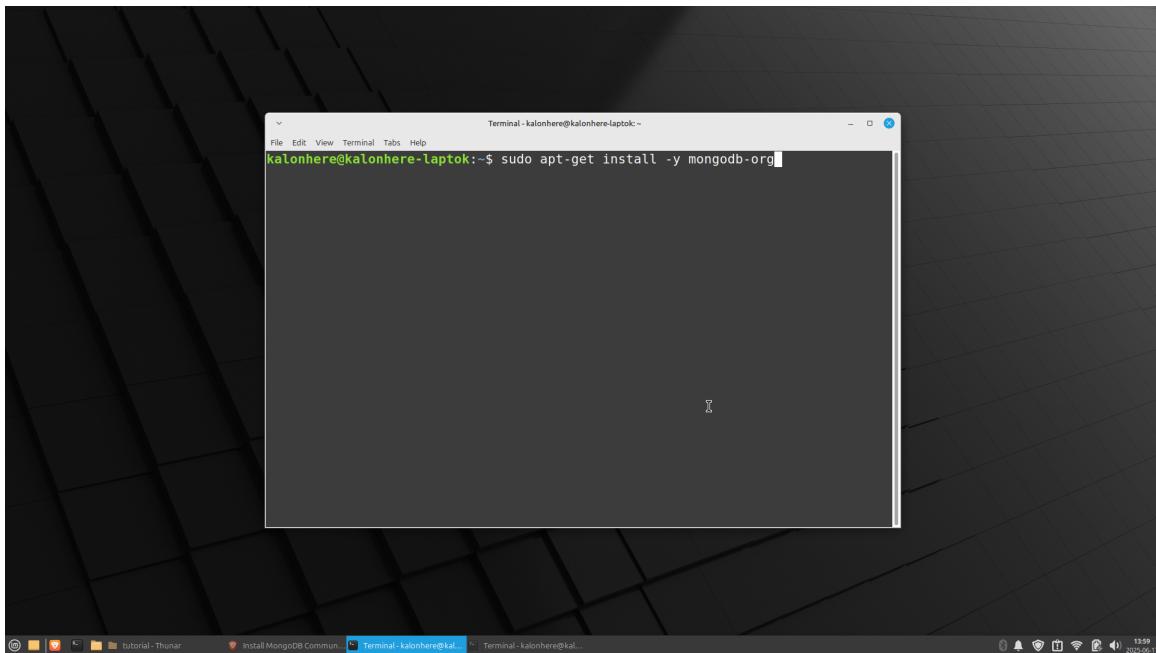
2.2 MongoDB



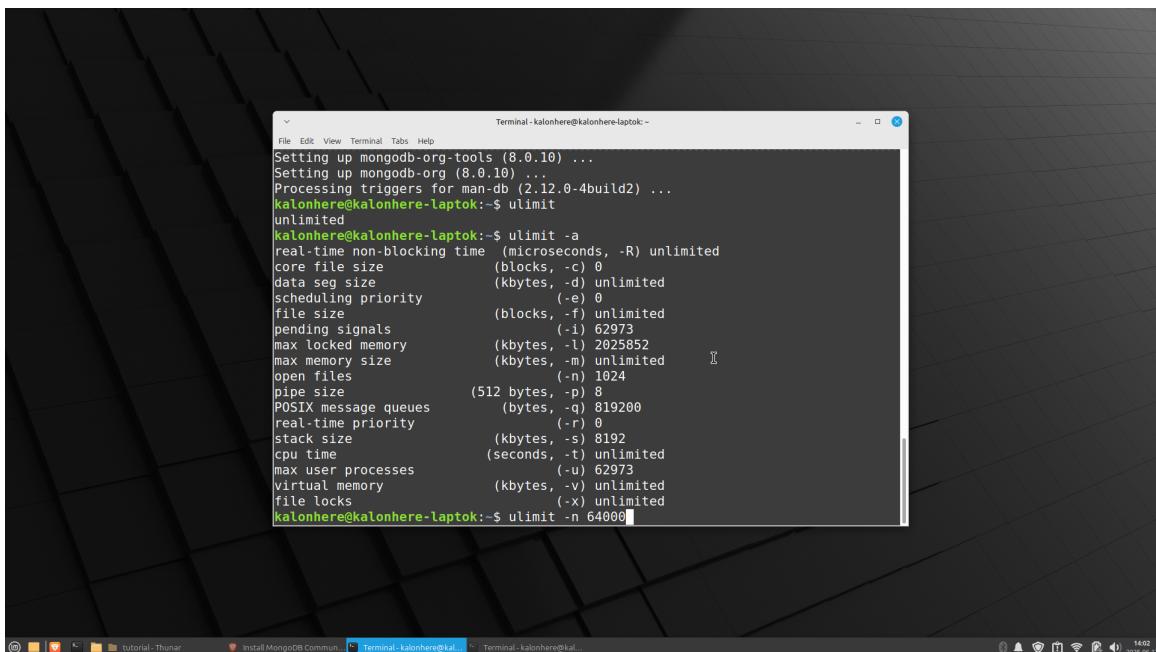
Zdjęcie. 7: Importowanie kluczy gpg do MongoDB



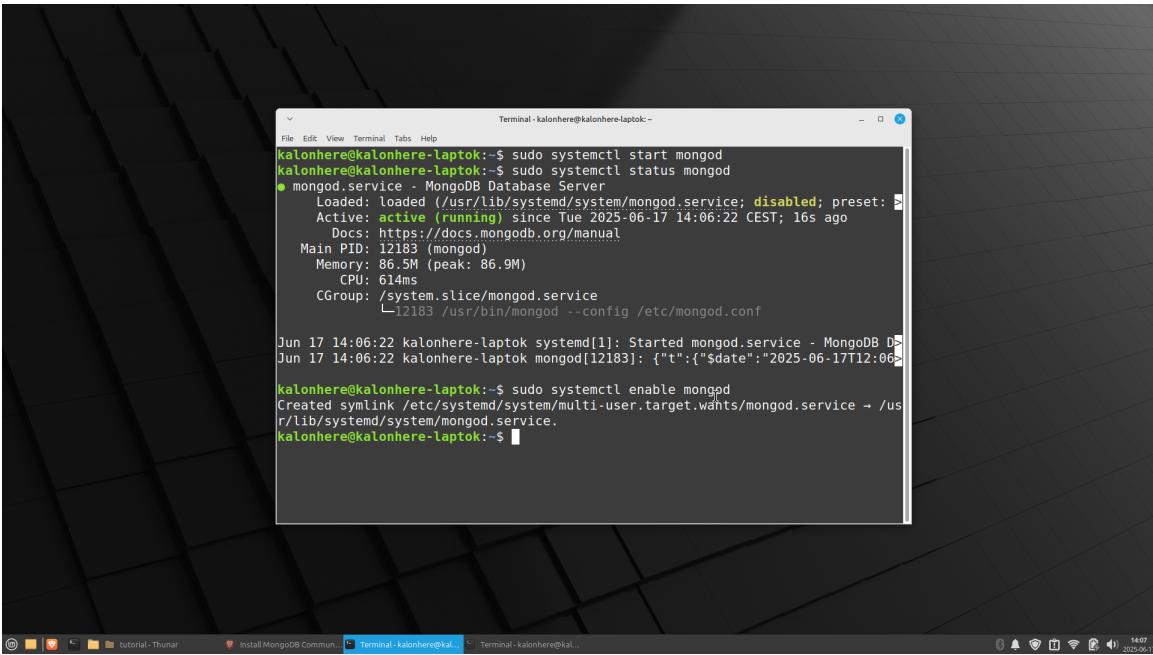
Zdjęcie. 8: Stworzenie pliku z listą źródeł do pobierania repo/package



Zdjęcie. 9: Instalacja MongoDB



Zdjęcie. 10: Zwiększenie limitu otwartych plików z 1024 do 64000



A screenshot of a terminal window titled "Terminal - kalonhere@kalonhere-laptop: ~". The window shows the following command-line session:

```
File Edit View Terminal Tabs Help
kalonhere@kalonhere-laptop:~$ sudo systemctl start mongod
kalonhere@kalonhere-laptop:~$ sudo systemctl status mongod
● mongod.service - MongoDB Database Server
   Loaded: loaded (/usr/lib/systemd/system/mongod.service; disabled; preset: )
   Active: active (running) since Tue 2025-06-17 14:06:22 CEST; 16s ago
     Docs: https://docs.mongodb.org/manual
   Main PID: 12183 (mongod)
      Memory: 86.5M (peak: 86.9M)
        CPU: 614ms
      CGroup: /system.slice/mongod.service
              └─12183 /usr/bin/mongod --config /etc/mongod.conf

Jun 17 14:06:22 kalonhere-laptop[1]: Started mongod.service - MongoDB D...
Jun 17 14:06:22 kalonhere-laptop mongod[12183]: {"t": {"$date": "2025-06-17T12:06:22.000+02:00"}, "v": 1, "s": "main", "c": "initandlisten", "m": "MongoDB starting in background mode on port 27017 with pid 12183. Version 5.2.2, build 2025-06-17T12:06:22.000+02:00, running on Linux 5.15.0-102-generic #103-Ubuntu SMP Mon Jun 12 10:00:00 UTC 2023 (Ubuntu 23.04), data directory /var/lib/mongo, lock file /var/run/mongodb/mongod.lock, using port 27017, waiting for connections on unix socket /var/run/mongodb/mongodb.sock", "err": null, "ns": null}
kalonhere@kalonhere-laptop:~$ sudo systemctl enable mongod
Created symlink /etc/systemd/system/multi-user.target.wants/mongod.service → /usr/lib/systemd/system/mongod.service.
kalonhere@kalonhere-laptop:~$
```

Zdjęcie. 11: Uruchomienie MongoDB, sprawdzenie statusu i włączenie autouruchamiania

3 Benchmark

3.1 Zapytania

1. Wybierz państwo które ma najwięcej odkryć z fizyki oraz nazwa państwa kończy się na "a"
2. Znajdź najstarsze i najmłodsze odkrycie z USA z fizyki
3. Znajdź państwa które mają odkrycia z biologii po 1960 roku
4. Znajdź średnią liczbę wszystkich odkryć na rok w każdym państwie w europie
5. Wypisz dziedzine w której każde państwo z europy dokonało najwięcej odkryć
6. Wypisz średnią liczbę odkryć na naukowca z każdego kontynentu
7. Znajdź państwo z największą średnią liczbą odkryć na osobę
8. Znajdź miasto z którego pochodzi najwięcej fizyków

3.2 Zapytania w PostgreSQL

mnkln

3.3 Zapytania w PostgreSQL w JSONB

faafe

3.4 Zapytania w MongoDB

cxzkm

3.5 Zapytanie xyz

iohntr

3.6 Zapytanie abc

opnfj