

# НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені Ігоря Сікорського»

# ФАКУЛЬТЕТ ПРИКЛАДНОЇ МАТЕМАТИКИ

# Кафедра системного програмування та спеціалізованих комп'ютерних систем

Лабораторна робота №3

з дисципліни Бази даних і засоби управління

на тему: "Засоби оптимізації роботи СУБД PostgreSQL"

Виконала:

студентка III курсу

групи КВ-94

Iyc I. O.

Перевірив:

Петрашенко А. В.

### Постановка задачі

 $Mетою poботи \in 3$ добуття практичних навичок використання засобів оптимізації СУБД PostgreSQL.

Завдання роботи полягає у наступному:

- 1. Перетворити модуль "Модель" з шаблону MVC лабораторної роботи №2 у вигляд об'єктно-реляційної проекції (ORM).
- 2. Створити та проаналізувати різні типи індексів у PostgreSQL.
- 3. Розробити тригер бази даних PostgreSQL.
- 4. Навести приклади та проаналізувати рівні ізоляції транзакцій у PostgreSQL.

№ варіанта	Види індексів	Умови для тригера	
8	BTree, GIN	After, insert, update	

Посилання на репозиторій у GitHub з вихідним кодом програми та звітом: https://github.com/van2ivan/KV-94-Ivan-Ius

# Завдання №1

Обрана предметна галузь передбачає ведення обліку оцінок у навчальному закладі

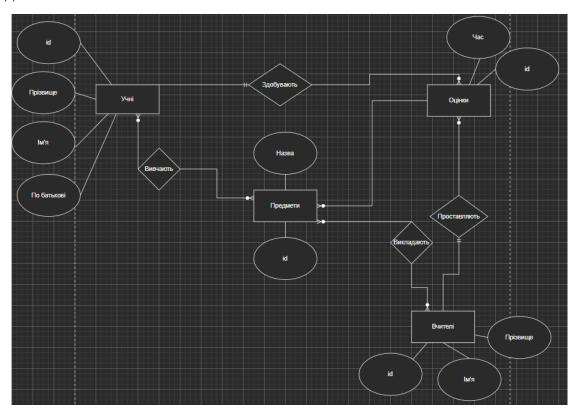


Рисунок 1. ER-діаграма

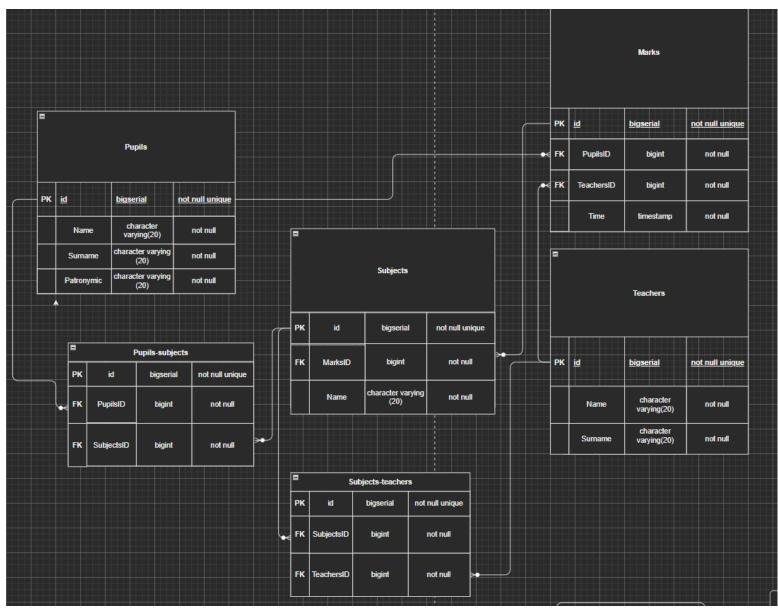


Рисунок 2. Схема бази даних

#### Класи ORM у реалізованому модулі Model

```
from sqlalchemy import BigInteger, Column, DateTime, Float, ForeignKey, Numeric, String, Table, Text, text
from sqlalchemy.dialects.postgresql import OID
from sqlalchemy.orm import relationship
from sqlalchemy.ext.declarative import declarative_base
Base = declarative base()
metadata = Base.metadata
class Pupil(Base):
    __tablename__ = 'Pupils'
    Id = Column(BigInteger, primary_key=True,
server_default=text("nextval('\"Pupils_Id_seq\"'::regclass)"))
    Name = Column(String(20), nullable=False)
    Patronymic = Column(String(20), nullable=False)
    Surname = Column(String(20), nullable=False)
class Teacher(Base):
    __tablename__ = 'Teachers'
    id = Column(BigInteger, primary_key=True, unique=True,
server_default=text("nextval('\"Teachers_id_seq\"'::regclass)"))
    name = Column(String(20), nullable=False)
    surname = Column(String(20), nullable=False)
class Mark(Base):
    __tablename__ = 'Marks'
    id = Column(BigInteger, primary_key=True, unique=True,
server default=text("nextval('\"Marks id seq\"'::regclass)"))
    time = Column(DateTime, nullable=False)
    pupilsid = Column(ForeignKey('Pupils.Id'), nullable=False)
    teachersid = Column(ForeignKey('Teachers.id'), nullable=False)
    Pupil = relationship('Pupil')
    Teacher = relationship('Teacher')
class Subject(Base):
    tablename = 'Subjects'
    id = Column(BigInteger, primary_key=True, unique=True,
server_default=text("nextval('\"Subjects_id_seq\"'::regclass)"))
    name = Column(String(20), nullable=False)
    marksid = Column(ForeignKey('Marks.id'), nullable=False)
    Mark = relationship('Mark')
class PupilsSubject(Base):
    __tablename__ = 'PupilsSubjects'
    id = Column(BigInteger, primary_key=True, unique=True,
server_default=text("nextval('\"PupilsSubjects_id_seq\"'::regclass)"))
    pupilsid = Column(ForeignKey('Pupils.Id'), nullable=False)
    subjectsid = Column(ForeignKey('Subjects.id'), nullable=False)
```

```
Pupil = relationship('Pupil')
Subject = relationship('Subject')

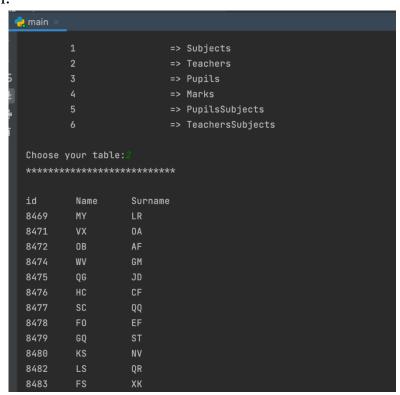
class TeachersSubject(Base):
    __tablename__ = 'TeachersSubjects'

    id = Column(BigInteger, primary_key=True, unique=True,
server_default=text("nextval('\"TeachersSubjects_id_seq\"'::regclass)"))
    teachersid = Column(ForeignKey('Teachers.id'), nullable=False)
    subjectsid = Column(ForeignKey('Subjects.id'), nullable=False)

Subject = relationship('Subject')
Teacher = relationship('Teacher')
```

Запити у вигляді ORM

Продемонструємо вставку, виучення, редагування даних. Початковий стан:



#### Видалення запису:

#### Вставка запису:

```
Choose your table: 2
Name = Ins
Surname = Ert
'added'
```

#### Редагування запису:

Запити пошуку та генерації рандомізованих даних також було реалізовано, логіку пошуку було змінено у порівнянні з лабораторною роботою №2 (усі дані для пошуку передвизначено, тепер вони не вводяться з клавіатури). Запити на пошук ті самі, що і л.р. №2.

```
Choose your table:2
Row to update where id = 8469
Name = Ann
Surname = Us
'updated'
1 => Continue update, 2 => Stop update => 2
  8488
            Petro
                       Poroshenko
  8494
            sql
                    alchemy
  8495
            Ins
                       Ert
                       'Us'
  8469
            'Ann'
  *********
```

# Завдання №2

Для тестування індексів було створено окремі таблиці у базі даних з 1000000 записів.

#### Hash

Хеш-індекси в PostgreSQL використовують форму структури даних хештаблиці (використовують хеш-функцію). Хеш-коди поділені на обмежену кількість комірок. Коли до індексу додається нове значення, PostgreSQL застосовує хешфункцію до значення і поміщає хеш-код і вказівник на кортеж у відповідну комірку. Коли відбувається запит за допомогою індексу хешування, PostgreSQL бере значення індексу і застосовує хеш-функцію, щоб визначити, яка комірка може містити потрібні дані.

# Стверення таблиці БД:

```
DROP TABLE IF EXISTS "hash_test";
CREATE TABLE "hash_test"("id" bigserial PRIMARY KEY, "time" timestamp);
INSERT INTO "hash_test"("time") SELECT (timestamp '2021-01-01' + random() * (timestamp)
```

```
'2020-01-01' - timestamp '2022-01-01')) FROM
(VALUES('qwertyuiopasdfghjklzxcvbnmQWERTYUIOPASDFGHJKLZXCVBNM')) as
symbols(characters), generate series(1, 1000000) as q;
Запити для тестування:
SELECT COUNT(*) FROM "hash_test" WHERE "id" % 2 = 0;
SELECT COUNT(*) FROM "hash_test" WHERE "time" >= '20191001';
SELECT AVG("id") FROM "hash_test" WHERE "time" >= '20191001' AND "time" <= '20211207';</pre>
SELECT SUM("id"), MAX("id") FROM "hash_test" WHERE "time" >= '20200505' AND "time" <=
'20210505' GROUP BY "id" % 2;
Створення індексу:
DROP TABLE IF EXISTS "hash_test";
```

```
CREATE INDEX "time_hash_index" ON "hash_test" USING hash("id");
```

### Результати і час виконання на скріншотах з psql

#### Запити без індексування:

```
postgres=# DROP INDEX IF EXISTS "time_hash_index";
ПОВІДОМЛЕННЯ: індекс "time_hash_index" не існує, пропускається
Время: 2,492 мс
postgres=# SELECT COUNT(*) FROM "hash_test" WHERE "id" % 2 = 0;
500000
(1 строка)
Время: 111,332 мс
postgres=# SELECT COUNT(*) FROM "hash_test" WHERE "time" >= '20191001';
count
626919
(1 строка)
Время: 163,147 мс
.
postgres=# SELECT AVG("id") FROM "hash test" WHERE "time" >= '20191001' AND "time" <= '20211207';
500254.786621557171
(1 строка)
Время: 176,904 мс
postgres=# SELECT SUM("id"), MAX("id") FROM "hash_test" WHERE "time" >= '20200505' AND "time" <= '20210505' GROUP BY "id" % 2;
            max
82418773060 | 999999
 82457311990 | 999994
(2 строки)
Время: 120,725 мс
```

Запити з індексуванням:

```
postgres=# CREATE INDEX "time_hash_index" ON "hash_test" USING hash("id");
CREATE INDEX
Время: 3745,561 мс (00:03,746)
postgres=# SELECT COUNT(*) FROM "hash_test" WHERE "id" % 2 = 0;
count
500000
(1 строка)
Время: 95,619 мс
postgres=# SELECT COUNT(*) FROM "hash_test" WHERE "time" >= '20191001';
count
626336
(1 строка)
Время: 100,444 мс
.
postgres=# SELECT AVG("id") FROM "hash_test" WHERE "time" >= '20191001' AND "time" <= '20211207';
500044.170913056251
(1 строка)
Время: 155,144 мс
postgres=# SELECT SUM("id"), MAX("id") FROM "hash_test" WHERE "time" >= '20200505' AND "time" <= '20210505' GROUP BY "id" % 2;
    sum
            max
82587011394 | 999994
82348391481 | 999989
(2 строки)
Время: 155,314 мс
```

Очевидно, що індексування за допомогою hash не значно пришвидшує пошук даних у табиці, а іноді навіть показує гірші результати, ніж запити без ідексування. Це випливає з того, що це один із найпримітивніших методів індексування і для пошуку потрібних даних алгоритм все одно проходить через усі записи у таблиці (на відміну від GIN). Він ефективний при застосуванні до поля числового типу. GIN

GIN призначений для обробки випадків, коли елементи, що підлягають індексації, є складеними значеннями (наприклад - реченнями), а запити, які обробляються індексом, мають шукати значення елементів, які з'являються в складених елементах (повторювані частини слів або речень). Індекс GIN зберігає набір пар (ключ, список появи ключа), де список появи — це набір ідентифікаторів рядків, у яких міститься ключ. Один і той самий ідентифікатор рядка може знаходитись у кількох списках, оскільки елемент може містити більше одного ключа. Кожне значення ключа зберігається лише один раз, тому індекс GIN дуже швидкий для випадків, коли один і той же ключ з'являється багато разів. Цей індекс може взаємодіяти тільки з полем типу tsvector.

## Стверення таблиці БД:

Створення індексу:

DROP INDEX IF EXISTS "gin\_index";

CREATE INDEX "gin\_index" ON "gin\_test" USING gin("gin\_vector");

```
DROP TABLE IF EXISTS "gin_test";
CREATE TABLE "gin_test"("id" bigserial PRIMARY KEY, "string" text, "gin_vector" tsvector);
INSERT INTO "gin_test"("string") SELECT substr(characters, (random() * length(characters) + 1)::integer, 10) FROM
(VALUES('qwertyuiopasdfghjklzxcvbnmQWERTYUIOPASDFGHJKLZXCVBNM')) as symbols(characters), generate_series(1, 1000000) as q;
UPDATE "gin_test" set "gin_vector" = to_tsvector("string");

3aпити для тестування:

SELECT COUNT(*) FROM "gin_test" WHERE "id" % 2 = 0;
SELECT COUNT(*) FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('bnm'));
SELECT SUM("id") FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('QWERTYUIOP')) OR ("gin_vector" @@ to_tsquery('bnm'));
SELECT MIN("id"), MAX("id") FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('bnm'))
GROUP BY "id" % 2;
```

# Результати і час виконання на скріншотах з psql

#### Запити без індексування:

```
postgres=# DROP INDEX IF EXISTS "gin_index";
ПОВІДОМЛЕННЯ: індекс "gin_index" не існує, пропускається
Время: 1,634 мс
postgres=# SELECT COUNT(*) FROM "gin_test" WHERE "id" % 2 = 0;
count
500000
(1 строка)
Время: 203,518 мс
postgres=# SELECT COUNT(*) FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('bnm'));
19142
(1 строка)
Время: 474,229 мс
postgres=# SELECT SUM("id") FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('QWERTYUIOP')) OR ("gin_vector" @@ to_tsquery('bnm'));
   sum
23943769938
(1 строка)
Время: 1188,034 мс (00:01,188)
postgres=# SELECT MIN("id"), MAX("id") FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('bnm')) GROUP BY "id" % 2;
100 | 999994
45 | 999937
(2 строки)
Время: 1120,586 мс (00:01,121)
```

#### Запити з індексуванням:

```
postgres=# CREATE INDEX "gin_index" ON "gin_test" USING gin("gin_vector");
CREATE INDEX
Время: 355,983 мс
oostgres=# SELECT COUNT(*) FROM "gin_test" WHERE "id" % 2 = 0;
count
500000
(1 строка)
Время: 156,321 мс
postgres=# SELECT COUNT(*) FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('bnm'));
count
19142
(1 строка)
Время: 25,425 мс
postgres=# SELECT SUM("id") FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('QWERTYVIOP')) OR ("gin_vector" @@ to_tsquery('bnm'));
23943769938
(1 строка)
Время: 243,217 мс
postgres=# SELECT MIN("id"), MAX("id") FROM "gin_test" WHERE ("gin_vector" @@ to_tsquery('bnm')) GROUP BY "id" % 2;
min | max
45 | 999937
100 | 999994
(2 строки)
Время: 13,533 мс
```

З отриманих результатів бачимо, що в усіх заданих випадках пошук з індексацією відбувається значно швидше, ніж пошук без індексації (окрім першого, оскільки на перший запит дана індексація не впливає). Це відбувається завдяки головній особливості індексування GIN: кожне значення шуканого ключа зберігається один раз і запит іде не по всій таблиці, а лише по тим даним, що містяться у списку появи цього ключа. Для даних типу numeric даний тип індексування використовувати недоцільно і неможливо.

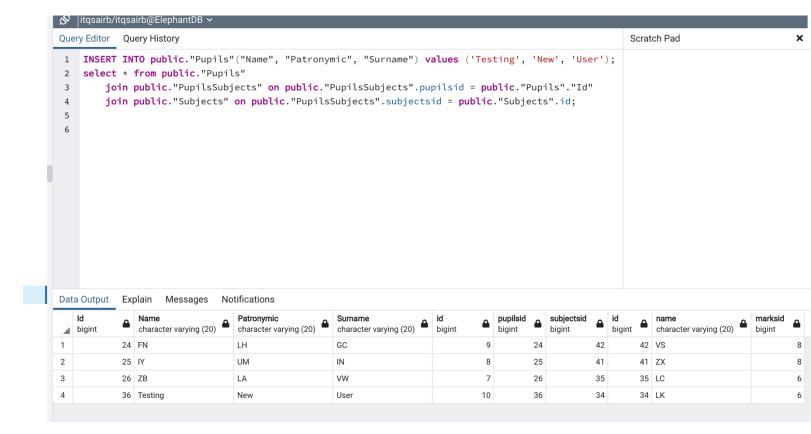
#### Завдання №3

Тригер створений для таблиці Pupils. Під час додання нового учня він прив'язується до викладача та предмета.

Команди, що ініціюють виконання тригера:

```
CREATE TRIGGER new_pupil_arrived
AFTER INSERT ON public. "Pupils"
FOR EACH ROW EXECUTE PROCEDURE assignPupilToSubject();
Текст тригера:
CREATE OR REPLACE FUNCTION assignPupilToSubject() RETURNS TRIGGER AS $$
declare
 teacherID bigint;
 subjectID bigint;
BEGIN
       select id into teacherID from public."Teachers" order by random() limit 1;
 select id into subjectID from public."Subjects" order by random() limit 1;
 insert into public. "Pupils Subjects" (pupils id, subjects id) values (NEW. "Id",
subjectID);
 insert into public. "TeachersSubjects" (teachersid, subjectsid) values (teacherID,
subjectID);
    RETURN NULL;
END
$$ LANGUAGE 'plpgsql'
```

### Результат роботи тригера



За результатами видно що після додання нового студента з'явилися відповідні записи у Subjects та PupilsSubjects

# Завдання №4

Для цього завдання також створювалась окрема таблиця з деякими початковими

#### даними:

```
DROP TABLE IF EXISTS "transactions";

CREATE TABLE "transactions"(
   "id" bigserial PRIMARY KEY,
   "numeric" bigint,
   "text" text
);

INSERT INTO "transactions"("numeric", "text") VALUES (111, 'string1'), (222, 'string2'), (333, 'string3');

ДНОГО ЧИТАННЯ.
```

#### REPEATABLE READ

На цьому рівні ізоляції Т2 не бачитиме змінені дані транзакцією Т1, але також не зможе отримати доступ до тих самих даних.

Тут видно, що друга не бачить змін з першої:

А тут, що отримуємо помилку при спробі доступу до тих самих даних:

```
postgres=# START TRANSACTION;
START TRANSACTION
postgres=*# SET TRANSACTION ISOLATION LEVEL REPEATABLE READ READ WRITE;
SET
postgres=*# UPDATE "transactions" SET "numeric" = "numeric" + 1;
UPDATE 3
postgres=# COMMIT;
COMMIT
postgres=#

COMM
```

Бачимо, що не виникає читання фантомів та повторного читання, а також заборонено одночасний доступ до незбережених даних. Хоча класично цей рівень ізоляції призначений для попередження повторного читання.

#### **SERIALIZABLE**

На цьому рівні транзакіції поводять себе так, ніби вони не знають одна про одну. Вони не можуть вплинути одна на одну і одночасний доступ строго заборонений.

```
ostgres=*# SET TRANSACTION ISOLATION LEVEL SERIALIZABLE READ WRITE;
                                                                                                                postgres=*# SET TRANSACTION ISOLATION LEVEL SERIALIZABLE READ WRITE;
oostgres="# UPDATE "transactions" SET "numeric" = "numeric" + 1;
                                                                                                                 postgres=*# SELECT * FROM "transactions";
                                                                                                                 id | numeric | text
ostgres=*# INSERT INTO "transactions"("numeric", "text") VALUES (444, 'strin
                                                                                                                              111 | string1
222 | string2
333 | string3
oostgres=*# DELETE FROM "transactions" WHERE "id"=1;
ostgres=*# SELECT * FROM "transactions";
id | numeric | text
                                                                                                                 postgres=*# SELECT * FROM "transactions";
            223 | string2
334 | string3
444 | string4
                                                                                                                 id | numeric | text
                                                                                                                              111 | string1
222 | string2
333 | string3
 ostgres=*# SELECT * FROM "transactions";
                                                                                                                postgres=*# UPDATE "transactions" SET "numeric" = "numeric" + 1;
                                                                                                                postgres="# orbArte transactions Set numeric = numeric +1;
ПОМИЛКА: не вдалося серіалізувати доступ через паралельне оновлення
postgres=!# INSERI INTO "transactions"("numeric", "text") VALUES (444, 'string4');
ПОМИЛКА: поточна транзакція перервана, команди до кінця блока транзакції пропускаються
postgres=!# DELETE FROM "transactions" WHERE "id"=1;
            223 | string2
334 | string3
             444 | string4
                                                                                                                ПОМИЛКА: поточна транзакція перервана, команди до кінця блока транзакції пропускаються
postgres=!# COMMIT
ostgres=*# COMMIT;
                                                                                                                 postgres-!#
                                                                                                                postgres-!# ROLLBACK
postgres-!# COMMIT
TIMMO
ostgres=# SELECT * FROM "transactions";
```

У попередньому випадку вдалось "відкатити" другу тразакцію і це не вплинуло на подальшу можливість роботи в терміналі. На цьому ж рівні навіть після завершення першої не вдалося зробити ні COMMIT ні ROLLBACK для другої транзакції. Взагалі, в класичному представленні цей рівень призначений для недопущення явища читання фантомів. На цьому рівні ізоляції ми отримуємо максимальну узгодженість даних і можемо бути впевнені, що зайві дані не будуть зафіксовані.

#### READ COMMITTED

На цьому рівні ізоляції одна транзакція не бачить змін у базі даних, викликаних іншою доки та не завершить своє виконання (командою COMMIT або ROLLBACK).

Дані після вставки та видалення так само будуть видні другій тільки після

завершення першої.

```
oostgres=*# SET TRANSACTION ISOLATION LEVEL READ COMMITTED READ WRITE;
                                                                                                  postgres=*# SET TRANSACTION ISOLATION LEVEL READ COMMITTED READ WRITE;
                                                                                                 postgres=*# SELECT * FROM "transactions";
id | numeric | text
oostgres=*# UPDATE "transactions" SET "numeric" = "numeric" + 1;
 ostgres=*#
                                                                                                             111 | string1
222 | string2
333 | string3
 ostgres=*#
ostgres=*#
                                                                                                   2 |
 ostgres=*#
 ostgres=*#
                                                                                                  (3 строки)
 ostgres=*#
 ostgres=*#
                                                                                                  postgres=*#
postgres=*# INSERT INTO "transactions"("numeric", "text") VALUES (444, 'strin
                                                                                                  postgres=*# SELECT * FROM "transactions";
                                                                                                   id | numeric | text
INSERT 0 1
postgres=*#
                                                                                                   1 |
2 |
3 |
                                                                                                             111 | string1
222 | string2
333 | string3
                                                                                                  (3 строки)
                                                                                                  postgres=*# _
postgres= #
postgres=*# DELETE FROM "transactions" WHERE "id"=1;
DELETE 1
                                                                                                ostgres=*# SELECT * FROM "transactions";
 ostgres=*#
                                                                                                1 | 2 | 3 |
                                                                                                         111 | string1
222 | string2
333 | string3
postgres="#
postgres=*# COMMIT;
                                                                                                        postgres=*# SELECT * FROM "transactions";
                                                                                                         id | numeric | text
postgres=#
                                                                                                                     223 | string2
334 | string3
444 | string4
                                                                                                          3 |
                                                                                                        (3 строки)
                                                                                                        postgres=*#
```

На цьому знімку також бачимо, що друга транзакція (справа) не може внести дані у базу, доки не завершилась попередня.

```
postgres=# START TRANSACTION;
START TRANSACTION
postgres=# START TRANSACTION;
START TRANSACTION
postgres=# SET TRANSACTION
postgres=# SET TRANSACTION ISOLATION LEVEL READ COMMITTED READ WRITE;
SET
postgres=*# UPDATE "transactions" SET "numeric" = "numeric" + 1;
UPDATE 3
postgres=*#

postgres=#

post
```

А тут бачимо, що після завершення першої, друга транзакція виконала запит, змінивши вже ті дані, що були закомічені першою транзакцією

```
postgres=# START TRANSACTION;
START TRANSACTION
                                                                                                            postgres=# START TRANSACTION;
START TRANSACTION
postgres=*# SET TRANSACTION ISOLATION LEVEL READ COMMITTED READ WRITE;
                                                                                                             ostgres=*# SET TRANSACTION ISOLATION LEVEL READ COMMITTED READ WRITE;
postgres=*# UPDATE "transactions" SET "numeric" = "numeric" + 1;
                                                                                                            postgres=*# UPDATE "transactions" SET "numeric" = "numeric" + 1;
DPDATE 3
postgres=*# SELECT * FROM "transactions";
id | numeric | text
                                                                                                            postgres=*#
postgres=*#
postgres=*#
           112 | string1
223 | string2
334 | string3
                                                                                                            postgres=*#
                                                                                                            postgres=*#
                                                                                                             postgres=*#
                                                                                                            postgres=*#
                                                                                                             postgres=*#
ostgres=*# commit;
                                                                                                             postgres=*#
                                                                                                             postgres=*#
                                                                                                             postgres=*#
postgres=#
postgres=# SELECT * FROM "transactions";
id | numeric | text
                                                                                                            postgres=*#
                                                                                                             postgres=*# SELECT * FROM "transactions";
                                                                                                             id | numeric | text
                                                                                                            1 |
2 |
3 |
(3 строки)
            113 | string1
224 | string2
335 | string3
                                                                                                                        113 | string1
224 | string2
335 | string3
 3 строки)
                                                                                                           postgres=*# commit;
COMMIT
postgres=#
 ostgres=#
```

Коли Т2 бачить дані Т1 запитів UPDATE, DELETE виникає феномен повторного читаня, а коли бачить дані запиту INSERT – читання фантомів.