Министерство образования Республики Беларусь

Учреждение образования БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра электронных вычислительных машин

ОТЧЕТ по лабораторной работе №3 на тему

Освоение прикладного интерфейса СУБД BerkeleyDB. Разработка конвертора базы данных PostgreSQL в набор баз данных Berkeley DB. Адаптация спецификаций приложения

Вариант 3 (Аэропорт)

Студент:	И.И. Божко
Преподаватель:	Ю.Ю. Желтко

1 ЦЕЛЬ РАБОТЫ

- 1. Научиться преобразовывать реляционные базы данных (PostgreSQL) в формат ключ-значение (Berkeley DB).
- 2. Освоить процесс сериализации и десериализации данных для хранения в нереляционной базе данных.
- 3. Выполнить адаптацию существующих спецификаций приложения для работы с Berkeley DB.

2 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Извлечение схемы и данных из PostgreSQL:

1. Используя подключение к PostgreSQL, извлеките структуру (схему) и данные из таблиц базы данных PostgreSQL. Для этого используйте SQL-запросы и выполните сериализацию данных.

2. Конвертация данных в Berkeley DB:

- 1. Для каждой таблицы в PostgreSQL создайте соответствующую базу данных Berkeley DB.
- 2. Используйте первичные ключи таблиц PostgreSQL в качестве ключей для Berkeley DB, а значения столбцов в виде сериализованных структур

3. Запись данных в Berkeley DB:

1. Реализуйте запись данных в формате ключ-значение в Berkeley DB. Для этого используйте соответствующие функции библиотеки Berkeley DB.

4. Адаптация спецификаций приложения:

- 1. Проанализируйте спецификации приложения, которые ранее работали с PostgreSQL, и адаптируйте их для работы с Berkeley DB.
- 2. Убедитесь, что операции вставки, обновления, удаления и поиска данных выполняются корректно с использованием нового формата хранения (ключ-значение).

3 СТРУКТУРА ДАННЫХ

3.1 Модель данных в postgresql

Модель данных "Аэропорт" содержит сущности "Airline" – авиакомпания, "Destination" – направление, "Flight" – рейс, "Airplane" – самолёт, "Passenger" – пассажир.

В таблице 3.1 представлено описание сущности "Airline".

Таблица 3.1 – Описание сущности "Airline".

Название поля	Описание поля	Ключ
code	Код компании	Первичный ключ
name	Название компании	_
country	Страна	_

В таблице 3.2 представлено описание сущности "Destination".

Таблица 3.2 – Описание сущности "Destination".

Название поля	Описание поля	Ключ
code	Код аэропорта	Первичный ключ
name	Название аэропорта	_
country	Страна	_

В таблице 3.3 представлено описание сущности "Destination".

Таблица 3.3 – Описание сущности "Flight".

Название поля	Описание поля	Ключ
flight_number	Номер рейса	Первичный ключ
airline_code	Код авиакомпании	Внешний ключ 1
airplane_code	Код аэропорта	Внешний ключ 2
days	Дни недели	_
arrival_time	Время прибытия	_
departure_time	Время отправления	_

В таблице 3.4 представлено описание сущности "Airplane".

Таблица 3.4 – Описание сущности "Airplane".

	1	
Название поля	Описание поля	Ключ
code	Код самолёта	Первичный ключ
name	Название самолёта	_
manufacturer	Производитель	_
capacity	Вместимость	_
width	Фюзеляж	_
	(широкий/узкий)	

В таблице 3.5 представлено описание сущности "Passenger".

Таблица 3.5 – Описание сущности "Passenger".

Название поля	Описание поля	Ключ
passport number	Номер паспорта	Первичный ключ
first_name	Имя	_
last_name	Фамилия	_
email	E-mail адрес	_

В таблице 3.6 представлено описание связей в модели данных.

Таблица 3.6 – Описание связей.

Название связи	Связываемые	Промежуточная
	таблицы	таблица
Авиакомпания,	Рейс, Авиакомпания	_
выполняющая рейс		
Самолёт, выполняющий	Рейс, Самолёт	_
рейс		
Направление	Рейс, Направление	Flights-Destinations
(направления) рейса		
Пассажир (пассажиры)	Рейс, Пассажир	Flights-Passengers
рейса		

3.2 Хранение данных в berkeleydb

Каждая таблица postgresql преобразуется в отдельную базу данных. В качестве ключей баз данных используются первичные ключи postgresql.

4 ПРИМЕРЫ SQL-ЗАПРОСОВ ДЛЯ ИЗВЛЕЧЕНИЯ СХЕМЫ И ДАННЫХ ИЗ POSTGRESQL

Пример запроса для получения названий столбцов таблицы:

```
SELECT column_name
FROM information_schema.columns
WHERE table_name = '{table_name}';
```

Пример запроса для получения первичных ключей таблицы:

```
SELECT kcu.column_name
FROM information_schema.table_constraints tco
JOIN information_schema.key_column_usage kcu
ON kcu.constraint_name = tco.constraint_name
WHERE tco.constraint_type = 'PRIMARY KEY' AND
tco.table_name = '{table_name}';
```

5 КОД КОНВЕРТОРА ДАННЫХ

Для сериализации данных получаются названия таблиц, столбцов и первичных ключей каждой таблицы. Первичный ключ обязательно переносится в начало таблицы. Затем для каждой таблицы создаётся JSON файл, содержащий все данные из таблицы.

Код для сериализации данных:

```
import psycopg2
import datetime
import json
import os
def get_all_tables(cursor):
    try:
        cursor.execute("""
            SELECT table name
            FROM information_schema.tables
            WHERE table_schema = 'postgres';
        """)
        tables = cursor.fetchall()
        return [table[0] for table in tables]
    except Exception as e:
        print(f"Error getting tables: {e}")
        return []
def get_primary_keys(cursor, table_name):
    try:
        cursor.execute(f"""
            SELECT kcu.column_name
            FROM information_schema.table_constraints tco
            JOIN information_schema.key_column_usage kcu
            ON kcu.constraint_name = tco.constraint_name
            WHERE tco.constraint_type = 'PRIMARY KEY' AND tco.table_name
= '{table_name}';
        """)
        primary_keys = cursor.fetchall()
        return [pk[0] for pk in primary_keys]
    except Exception as e:
```

```
print(f"Error getting primary keys for {table_name}: {e}")
        return []
def get column names(cursor, table name):
    try:
        cursor.execute(f"""
            SELECT column_name
            FROM information_schema.columns
            WHERE table_name = '{table_name}';
        """)
        columns = cursor.fetchall()
        return [column[0] for column in columns]
    except Exception as e:
        print(f"Error getting column names for {table_name}: {e}")
        return []
import datetime
def get_all_data(cursor, table_name, cnfk):
    try:
        cursor.execute(f"SELECT * FROM {table_name};")
        data = cursor.fetchall()
        column_names = [description[0] for description in
cursor.description]
        cnfk_index = column_names.index(cnfk)
        rearranged_column_names = [column_names[cnfk_index]] + [
            column_names[i] for i in range(len(column_names)) if i !=
cnfk index
        ]
        filtered_data = []
        for row in data:
            new_row = (row[cnfk_index],) + tuple(
                item.strftime('%H:%M:%S') if isinstance(item,
datetime.time) else item
                for i, item in enumerate(row) if i != cnfk_index
            )
            filtered_data.append(new_row)
        return rearranged_column_names, filtered_data
    except Exception as e:
```

```
print(f"Error getting data from {table name}: {e}")
        return [], []
def create json from tuples(tuples array, file path, column names):
    data_list = [
        {column_names[i]: value for i, value in enumerate(tup)}
        for tup in tuples_array
    ]
    with open(file_path, 'w') as json_file:
        json.dump(data_list, json_file, indent=4)
def main():
    conn = None
    try:
        conn = psycopg2.connect(host='localhost', database='newairport',
user='postgres', password='admin')
    except Exception as e:
        print(f"Error connecting to database: {e}")
        return
    if conn is not None:
        cursor = conn.cursor()
        tables = get all tables(cursor)
        print("Tables in the database:", tables)
        if tables:
            try:
                os.makedirs('./dumps', exist ok=True)
            except Exception as e:
                pass
            for table name in tables:
                columns = get_column_names(cursor, table_name)
                primary_key = get_primary_keys(cursor, table_name)
                columns, data = get_all_data(cursor, table_name,
primary_key[0])
                print(f"Data from {table_name}:", columns, data)
                create_json_from_tuples(data,
f'dumps/{table_name}.json', columns)
        conn.close()
```

```
if __name__ == "__main__":
    main()
```

Для десериализации данных получаются данные из JSON файлов и для каждого файла создаётся база данных, в качестве ключа используется первое значение из json.

Код для десериализации данных:

```
import os
import json
import bsddb3
def process_json_files(directory):
    for filename in os.listdir(directory):
        base = os.path.splitext(filename)[0]
        new_filename = f"./databases/{base}.db"
        print(f'{filename} -> {new filename}')
        try:
            os.makedirs('./databases', exist_ok=True)
        except Exception as e:
            pass
        if filename.endswith('.json'):
            file_path = os.path.join(directory, filename)
            db = bsddb3.db.DB()
db.open(new_filename, None, bsddb3.db.DB_BTREE, bsddb3.db.DB_CREATE)
            try:
                with open(file_path, 'r') as json_file:
                     data = json.load(json_file)
                     if isinstance(data, list):
                         for item in data:
                             if isinstance(item, dict):
f'{list(item.values())[0]}'.encode('utf-8')
                                 value_data = {k: item[k] for k in item
if k != list(item.keys())[0]}
value_structure =
json.dumps(value_data).encode("utf-8")
                                 print(key, value structure)
                                 db.put(key, value_structure)
                             else:
                                 print("Item is not a dictionary:", item)
                     else:
                         print(f"Data in {filename} is not a list:
{data}")
```

На рисунке 5.1 показан процесс десериализации данных.

```
| Disapport number": HBA37HM, "first name": "Krista", "last name": "Viola", "email": "kristabloom@gmail.com")
| Disapport number": BBA37HM, "first name": "Timothy", "last name": "Braksn', "email": "thaaJul@mail.be")
| Disapport number": "BA37HM, "first name": "Inonifer", "last name": "Erikson", "email": "beapport number": "BA307HM, "first name": "Inonifer", "last name": "Frikson", "email": "kroa@mail.uk")
| Disapport number": "BYA37", "first name": "Achistopher", "last name": "Friola", "email": "kroa@mail.uk")
| Disapport number": "BYA37", "first name": "Lee", "last name": "Triola", "email": "kroa@mail.uk")
| Disapport number": "BYA37", "first name": "Lee", "last name": "Friola", "email": "kroa@mail.uk")
| Disapport number": "BYA37", "first name": "Lee", "last name": "Eexi", "email": "kroa@mail.uk")
| Disapport number": "BYA37", "first name": "Country": "Sarael")
| Disapport number": "BYA37", "country": "Portugal")
| Disapport number": "BYA37", "country": "Portugal")
| Disapport number": "BYA37", "country": "Portugal")
| Disapport number: "BYA37", "name: "Beoing 737-880", "manufacturer": "Boeing", "capacity": 188, "width": 1)
| Disapport number: "BYA37", "name: "Beoing 737-880", "manufacturer: "Ribraer", "capacity": 180, width": 1)
| Disapport number: "Lister, "name: "Ribraer: "Ribraer", "capacity": 180, width": 2)
| Disapport number: "Lister, "name: "Ribraer: "Ribraer: "Capacity: 180, width": 2)
| Disapport number: "Lister, "name: "Ribraer: "Ribraer: "Ribraer: "capacity: 180, width": 2)
| Disapport number: "Lister, "name: "Ribraer: "Ribraer: "capacity: "Alo, width": 2)
| Disapport number: "Lister, "name: "Ribraer: "Ribraer: "Ribraer: "Capacity: "Alo, width
```

Рисунок 5.1 – Десериализация данных

Для просмотра содержимого баз данных используется следующий код:

```
import bsddb3
import os

def print_all_records(db_path):
    try:
        db = bsddb3.btopen(db_path, 'r')
    except Exception as e:
        print(f"Error opening database: {e}")
        return

    try:
        for key in db.keys():
```

На рисунке 5.2 показан процесс просмотра содержимого баз данных.

Рисунок 5.2 – Просмотр данных

Для добавления данных используется следующий код:

```
import bsddb3
import os
```

```
maxkey = 1
def print_all_records(db_path):
   global maxkey
   try:
        db = bsddb3.btopen(db_path, 'r')
    except Exception as e:
        print(f"Error opening database: {e}")
        return
   try:
        for key in db.keys():
            value = db[key]
            key = key.decode("utf-8")
            ikey = int(key)
            value = value.decode("utf-8")
            print(f"Key: {key}, Value: {value}")
            if ikey > maxkey:
                maxkey = ikey
   except Exception as e:
        print(f"Error reading records: {e}")
    finally:
        db.close()
def add_new_record(db_path, data):
    global maxkey
    try:
        db = bsddb3.btopen(db_path, 'c')
    except Exception as e:
        print(f"Error opening database: {e}")
        return
```

```
try:
       key = f'{maxkey + 1}'.encode("utf-8")
       value = data.encode("utf-8")
       db[key] = value
   except Exception as e:
       print(f"Error adding record: {e}")
   finally:
       db.close()
if __name__ == "__main__":
   user_input = input("Enter table name: ")
   if os.path.exists(f'./databases/{user_input}.db'):
       print_all_records(f'./databases/{user_input}.db')
       print('=======')
       new_data = input("Enter new data: ")
       add_new_record(f'./databases/{user_input}.db', new_data)
       print('========')
       print_all_records(f'./databases/{user_input}.db')
   else:
       print(f"The file ./databases/'{user_input}'.db does not exist.")
```

На рисунке 5.3 показан процесс добавления данных.

Рисунок 5.3 – Добавление данных

Для удаления данных используется следующий код:

```
import bsddb3
import os
maxkey = 1
def print_all_records(db_path):
    global maxkey
   try:
        db = bsddb3.btopen(db_path, 'r')
   except Exception as e:
        print(f"Error opening database: {e}")
        return
   try:
        for key in db.keys():
            value = db[key]
            key = key.decode("utf-8")
            ikey = int(key)
            value = value.decode("utf-8")
            print(f"Key: {key}, Value: {value}")
            if ikey > maxkey:
                maxkey = ikey
    except Exception as e:
        print(f"Error reading records: {e}")
    finally:
        db.close()
def delete_record(db_path, orkey):
    global maxkey
   k = None
```

```
try:
       db = bsddb3.btopen(db_path, 'c')
   except Exception as e:
       print(f"Error opening database: {e}")
       return
   try:
       k = f'{orkey}'.encode("utf-8")
       value = db[k]
   except Exception as e:
       print(f"Record with key '{orkey}' does not exist.")
       return
   try:
       del db[k]
   except Exception as e:
       print(f"Error editing record: {e}")
   finally:
       db.close()
if __name__ == "__main__":
   user_input = input("Enter table name: ")
   if os.path.exists(f'./databases/{user_input}.db'):
       print_all_records(f'./databases/{user_input}.db')
       print('=======')
       orig_key = input("Enter the key: ")
       delete_record(f'./databases/{user_input}.db', orig_key)
       print('=======')
       print_all_records(f'./databases/{user_input}.db')
   else:
       print(f"The file ./databases/'{user_input}'.db does not exist.")
```

На рисунке 5.4 показан процесс удаления данных.

Рисунок 5.4 – Удаление данных

Для редактирования данных используется следующий код:

```
import bsddb3
import os
maxkey = 1
def print_all_records(db_path):
    global maxkey
    try:
        db = bsddb3.btopen(db_path, 'r')
    except Exception as e:
        print(f"Error opening database: {e}")
        return
    try:
        for key in db.keys():
            value = db[key]
            key = key.decode("utf-8")
            ikey = int(key)
            value = value.decode("utf-8")
            print(f"Key: {key}, Value: {value}")
            if ikey > maxkey:
                maxkey = ikey
    except Exception as e:
        print(f"Error reading records: {e}")
```

```
finally:
       db.close()
def edit record(db path, data, orkey):
   global maxkey
   k = None
   try:
       db = bsddb3.btopen(db_path, 'c')
   except Exception as e:
       print(f"Error opening database: {e}")
       return
   try:
       k = f'{orkey}'.encode("utf-8")
       value = db[k]
   except Exception as e:
       print(f"Record with key '{orkey}' does not exist.")
       return
   try:
       value = data.encode("utf-8")
       db[k] = value
   except Exception as e:
       print(f"Error editing record: {e}")
   finally:
       db.close()
if __name__ == "__main__":
   user_input = input("Enter table name: ")
   if os.path.exists(f'./databases/{user_input}.db'):
       print_all_records(f'./databases/{user_input}.db')
       print('=======')
       orig_key = input("Enter the key: ")
       new_data = input("Enter new data: ")
       edit_record(f'./databases/{user_input}.db', new_data, orig_key)
       print('======')
       print_all_records(f'./databases/{user_input}.db')
   else:
       print(f"The file ./databases/'{user_input}'.db does not exist.")
```

На рисунке 5.5 показан процесс редактирования данных.

Рисунок 5.5 – Редактирование данных

6 ИЗМЕНЕНИЯ СПЕЦИФИКАЦИИ ПРИЛОЖЕНИЯ

В связи с заменой postgresql на berkeleydb приложению необходимо выполнять дополнительные функции:

- 1) Контроль и генерация первичных ключей
- 2) Обработка внешних ключей
- 3) Валидация ввода
- 4) Приведение данных к единому формату

ЗАКЛЮЧЕНИЕ

В ходе лабораторной работы были выполнены ключевые шаги по преобразованию данных из реляционной модели postgreSQL в формат ключ-значение, используемый в berkeleydb.

Были разработаны скрипты для сериализации и десериализации данных, просмотра, вставки, удаления, редактирования данных в berkeleydb.

Работа показала возможности и ограничения Berkeley DB в сравнении с PostgreSQL, а также расширила представление о выборе архитектуры хранения данных для разных типов приложений.