

Федеральное государственное автономное образовательное учреждение
высшего образования

«Университет ИТМО»

Факультет ПИиКТ

Дисциплина: Хранилища и базы данных

Лабораторная работа 1

Выполнили: Гурин Евгений Иванович, Камышанская Ксения Васильевна

Преподаватель: Королёва Юлия Александровна

Группа: Р4116

Санкт-Петербург 2023г.

Цель работы

Выбрать любую нереляционную базу данных и изучить язык запросов этой БД.

Выполнение

Для выполнения задания была выбрана БД Tarantool

СУБД Tarantool – это ПО с открытым исходным кодом, разработанное в России. Он совмещает в себе СУБД и сервер приложений на языке Lua, что делает данную СУБД легко расширяемой за счёт lua скриптов и создания переиспользуемых модулей. Также такое решение называют NewSQL. Данная СУБД позиционируется как многопоточная, готовая к высокой нагрузке (ориентируется на множественное чтение) и с возможностью объединения в кластер.

Для работы с данными используется in-memory подход. Данный факт позволяет СУБД ориентироваться на высоконагруженные приложения. При этом хранение подразумевает 2 способа:

- memtx — хранение данных в оперативной памяти, с дисковыми снимками данных и логом транзакций.
- vinyl — хранение данных на жестком диске, с оптимизацией для быстрой вставки данных.

В данной СУБД есть аналоги таблицам и строкам, привычным для реляционных БД:

- space – таблицы
- tuple (кортежи) – строки

Возможности БД:

- Имеется поддержка ACID транзакций
- Поддерживается Шардирование – возможность масштабирования с использованием разделения хранимых данных по различным серверам базы данных
- Расширение за счёт модулей на Lua

- Совместимость с SQL
- Асинхронная репликация

Язык запросов

Для написания запросов и работы с базой данных предоставляется интерфейс работы с базой данных на основе языка Lua. Для начала работы с БД необходимо создать space

```
s = box.schema.space.create('some_space')
```

Для указания структуры спейса (аналогия с созданием таблицы в реляционной БД) используется функция `format`. В примере ниже задаются поля `id`, `name` и `count`

```
box.space.some_space:format({  
  {name = 'id', type = 'unsigned'},  
  {name = 'name', type = 'string'},  
  {name = 'count', type = 'unsigned'}  
})
```

Для эффективной работы с данными указывается индекс спейса. Данный функционал аналогичен индексам в реляционных БД. Поддерживаются следующие типы индексов:

- tree (по умолчанию)
- hash
- rtree
- bitset

Кроме того тип индекса зависит от типа поля. Например для `datetime` `hash` индекс не доступен.

```
box.space.some_space:create_index('primary', {  
  type = 'tree',  
  parts = {'id'}  
})
```

Для создания данных (кортежей) используется функция insert.
Данные указываются в порядке из определения в функции format.

```
box.space.some_space:insert{1, 'Ivan', 3}  
box.space.some_space:insert{2, 'Igor', 3}
```

Для осуществления замены кортежа используется функция replace

```
box.space.some_space:replace{1, 'Petr', 10}
```

Кроме того есть возможность использования ассоциативных массивов и преобразования их в таблицы и кортежи.

```
form = {{name='id',type='unsigned'},{name='name',type='string'}}  
  
s = box.schema.create_space('person', {format = form})  
  
s:frommap({name = 'Ivan', id = 1})  
# Будет преобразовано в кортеж [1, 'Ivan']
```

Для выборки данных используется команда select. Tarantool производит поиск по ключам индексов, которые создаются функцией create_index. Поиск по нескольким полям возможен если в индексе было использовано несколько полей.

```
box.space.some_space:select{} - получение всех кортежей  
  
box.space.some_space:select{1,'В'} - совпадение с двумя полями  
  
box.space.some_space:select{1} - совпадения с одним полем  
  
box.space.some_space:select({1},{iterator='GT',offer=1,limit=2})  
- в этом случае первое поле кортежа должно быть больше 0, не выводим первый кортеж и на экран выводим максимум 2 кортежа.  
Параметры фильтрации зависят от выбранного индекса
```

Выборка одного элемента. По сути аналог select с получением первого элемента

```
box.space.some_space:get{1}
```

Для подсчёта кортежей используется функция count. В примере происходит подсчёт кортежей с индексом более или равному 2.

```
box.space.some_space:count(2, {iterator='GE'})
```

Кроме того для выборки можно использовать функцию pairs, что может быть также актуально на больших объемах данных, так как является итератором. В примере итерируется 100 значений из спейса.

```
box.space.some_space:pairs():take_n(100):totable()
```

Удаление происходит по созданному на спейсе индексу

```
box.space.some_space:delete(1)
```

У Tarantool также есть продукт Tarantool Data Grid для разработки сложных enterprise приложений и построения доменных моделей. При его использовании появляется возможность построения связей

Вывод

В ходе выполнения лабораторной работы был изучен язык запросов базы данных Tarantool. Эта СУБД показывает хорошие результаты по производительности и с каждым днем становится все популярнее. Опыт в работе с данной СУБД может быть использован при разработке высоконагруженных приложений.