

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

**Гуменюк Андрій Борисович**

УДК 004.65

**Графова база даних Neo4j, для представлення мови Cypher**

124 – Системний аналіз

Автореферат  
магістерської наукової роботи на здобуття освітньої кваліфікації  
«Магістр системного аналізу»

Миколаїв – 2019

Магістерська наукова робота є рукопис.

Робота виконана в Чорноморському національному університеті імені Петра Могили Міністерства освіти і науки України на кафедрі інтелектуальних інформаційних систем

Науковий керівник: к.ф-м.н., доцент Кулаковська Інесса Василівна

Рецензент: д.ф.н, професор Фісун Миколай Тихонович

Захист відбудеться «\_26\_» лютого 2019 р. о 9<sup>30</sup> год. на засіданні екзаменаційної комісії (ауд. 2-403) у Чорноморському національному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

З магістерською науковою роботою можна ознайомитися в бібліотеці Чорноморського національного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

Автореферат представлений «\_25\_» лютого 2019 р.

Секретар  
екзаменаційної комісії,  
к.пед.н., доцент

Н. М. Болюбаш

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** З 1980-х років реляційні системи управління базами даних (СКБД) стали займати домінуюче становище серед засобів зберігання даних. Незважаючи на те що реляційні сховища забезпечують оптимальне поєднання простоти, стійкості, гнучкості, продуктивності, масштабованості і сумісності, їх показники по кожному з цих пунктів не обов'язково вище, ніж у аналогічних систем, орієнтованих на якусь одну особливість. Однак універсальність реляційних СУБД переважає будь-які інші недоліки.

Сьогодні ситуація дещо інша. З'явилися в останні роки так звані NoSQL (Not only SQL, не тільки SQL) сховища реалізують моделі даних, що мають суттєві відмінності від традиційної реляційної моделі. Основна їх мета – розширити можливості баз даних (БД) в тих областях, де реляційна модель і SQL недостатньо гнучкі, і не витіснити їх там, де вони справляються зі своїми завданнями. Творці таких БД серед безлічі переваг використання NoSQL-рішень називають високу продуктивність при використанні специфічних моделей даних і легкість роботи з ними.

У магістерській роботі, необхідно розглянути питання щодо графових баз даних які значно переважають реляційні в специфічних задачах методом порівняння запитів та їй швидкодії на прикладі графової бази даних Neo4j.

**Метою магістерської наукової роботи** є дослідження області застосування графових моделей баз даних і сучасні тенденції по вирішенню деяких завдань на основі зберігання інформації у вигляді графів.

**Об'єкт досліджень** – системи управління і способи використання графових баз даних.

**Предмет досліджень** – функції і можливості мови запитів Cypher для графової бази даних Neo4j, особливості використання запитів для порівняльного аналізу ефективності роботи.

### **Етапи дослідження:**

- з'ясувати поняття графової бази даних, призначення та класифікація систем управління в графових базах даних;
- визначити принципи проектування графових баз даних, особливості використання Neo4j;
- проаналізувати існуючі методи та технології зберігання інформації;
- продемонструвати можливості графової бази даних Neo4j та мови Cypher, на прикладі аналізу запитів SQL та Cypher.

**Методи дослідження.** У процесі виконання поставлених у роботі завдань були використані методи аналізу літературних джерел, аналіз існуючих методів та технологій зберігання інформації, практична робота з графовими базами даних для перевірки і засвоєння набутих знань і навичок.

**Практичне значення отриманих результатів.** Аналіз взаємозв'язків між об'єктами бази даних у багаторівневих системах мають значну перевагу перед реляційними, що показано на прикладах.

**Апробація результатів магістерської наукової роботи:** Під час «XXI Всеукраїнської науково-методичної конференції «Могилянські читання – 2018: Досвід та тенденції суспільства в Україні: глобальний, національний та регіональний аспекти», яка проходила в Чорноморському університеті імені Петра Могили, м. Миколаїв, 12-17 листопада, була представлена доповідь по результатам розгляду графових баз даних та існуючих досліджень в цій сфері.

**Публікації.** Гуменюк А. Б., Кулаковська І. В. «Графова база даних Neo4j, для представлення мови Cypher» / XXI Всеукраїнській науково-методична конференція «Могилянські читання – 2018: Досвід та тенденції суспільства в Україні: глобальний, національний та регіональний аспекти». Миколаїв, 12-17 листопада, 2018 - с. 20.

**Структура магістерської наукової роботи.** Магістерська наукова робота складається із вступу, 5 розділів, висновків, додатків. Загальний обсяг роботи складає 89 сторінки, 21 рисунків, 6 таблиць та 43 посилань на літературні джерела.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

Потреби сучасного бізнесу дуже сильно впливають на ІТ індустрію. Все складніше зберігати та маніпулювати великим обсягом даних зі складною структурою. Для цих потреб сховища основані на реляційних баз даних не підходять для цього.

Одним з найбільш популярних і актуальних сховищ є графові БД . Як зрозуміло з назви, основна модель даних в них - класичний математичний граф. Проекти в області графових БД почали з'являтися з кінця 1980-х років, однак більшою мірою носили академічний характер. Останнім часом спостерігається бурхливий ріст інтересу до графові БД в зв'язку з тим, що така система представлення даних виявилася природною і затребуваною в сучасному світі різних соціальних зв'язків (Інтернет, соціальні мережі і т. д.).

Перший розділ описує реляційні та графові бази даних, їх недоліки та переваги. Також в розділі розглядається особливості графових баз даних.

Другий розділ присвячено інформаційній моделі вимог до мови запитів. Інформаційна модель - модель об'єкта, представлена у вигляді інформації, яка описує суттєві для даного розгляду параметри і змінні величини об'єкта, зв'язку між ними, входи і виходи об'єкта і дозволяє шляхом подачі на модель інформації про зміни вхідних величин моделювати можливі стану об'єкта.

Інформаційна модель (в широкому, загальнонауковому сенсі) - сукупність інформації, що характеризує суттєві властивості і стану об'єкта, процесу, явища, а також взаємозв'язок із зовнішнім світом.

Реляційна модель бази даних була введена Едгаром Коддом. Вона висуває на перший план концепцію абстрактних рівнів, вводячи ідею поділу між фізичним і логічним рівнями. Завдяки простоті використання, вона отримала широку популярність серед бізнес-додатків. На відміну від попередніх моделей, семантичні моделі бази дозволяють розробникам бази даних представляти об'єкти і їх відносини в природною і зрозумілою формі, надаючи користувачам такі інструменти, які дозволяють звертатися до необхідних частинах даними. Добре

відомим прикладом є модель сутність-зв'язок. Об'єктно-орієнтовані моделі баз даних з'явилися у вісімдесятих роках, коли більша частина досліджень була пов'язана з так званими «просунутими системами для нових типів додатків». Ці моделі баз даних засновані на об'єктно-орієнтованій парадигми, і їх мета полягає в тому, щоб представити дані у вигляді колекції об'єктів, які організовані в класи, яким призначені комплексні значення.

Графові моделі баз даних з'явилися поряд з об'єктно моделями. Ці моделі намагаються подолати обмеження, що накладаються традиційними моделями. Мотивацією для розробки таких систем стала графова структура даних, що з'являються в таких додатках, як гіпертекстові або географічні інформаційні системи, де взаємозв'язок між даними є важливим аспектом. Полуструктурованого моделями називають такі концепції, які призначені для зберігання даних з гнучкою структурою, наприклад, документів і веб-сторінок/

В третьому розділі розглядається розробка функціональних вимог. Мова запитів Cypher, та її структура.

Будучи декларативною мовою, Cypher фокусується на ясності висловлювання того, що треба зробити з графа, а не в тому, як отримати це. Це відрізняє його від імперативних мов або скриптових сценаріїв. Такий підхід суттєво спрощує процес оптимізації, не обтяжуючи користувача інформацією про структуру бази даних і не змушуючи оновлювати код запиту тільки тому, що логічна структура бази даних змінилася (нові індекси і т.д.) [8], [17].

Приклади позначення шуканих даних:

- (n) -> (m) - все спрямовані ребра з вершини n в вершину m;
- (n: Holder) - все вершини з міткою Holder;
- (n: Holder: Account) - все вершини, що мають обидві мітки Holder і Account;
- (n: Holder {name: {Mark}}) - все вершини з міткою Holder і відфільтровані по додатковому властивості name;
- (n: Holder) -> (m) - ребра між вершинами n з міткою Person і m;

— `() - [: IS HOLDER] -> ()` - відобразити вершини, які пов'язані між собою іменованої зв'язком;

— `p = (n: Holder) - [* 3..5] - (m: Holder)` - відобразити всі вершини, які з'єднані між собою мінімум через 3 інші вершини і максимум через 5, ігноруючи напрямки зв'язків.

Cypher підтримує запити з параметрами. Це означає, що розробники не повинні вдаватися до строковому побудови. Більш того, параметри полегшують операцію кешування виконання планів для Cypher [14], [17].

Параметри можуть бути використані для літералів і виразів в реченні WHERE, для значення індексу в реченні START, для індексованих запитів і для самих вузлів або відносин. Пункти, які не можуть бути використані в якості імен властивостей, типів відносин або міток.

Cypher використовує підхід SQL-запитів, складених з різних пропозицій. Пропозиції пов'язані і передають проміжні підсумкові набори один одному. Наприклад, відповідні ідентифікатори з одного речення MATCH будуть контекстом, в якому існує наступна пропозиція.

Мова запитів складається з декількох різних пропозицій.

Пропозицій, використовуваних для читання з графа.

**MATCH:** Зразок (патерн), якому відповідає граф. Це найбільш загальний спосіб отримання даних з графа.

**WHERE:** Не цілком самостійне пропозицію, скоріше, частина пропозицій MATCH, OPTIONAL MATCH і WITH. Додає обмеження в патерн або фільтрує проміжний результат на виході пропозиції WITH.

**RETURN:** Що має бути отримано на виході.



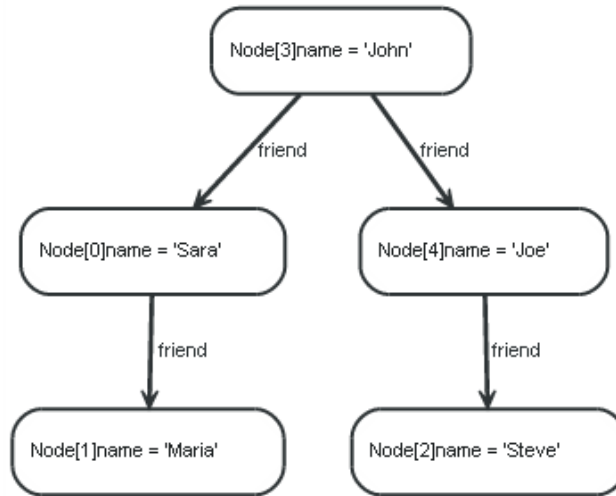


Рис.3.1. Приклад структури графа

Запит, який знаходить користувача з ім'ям Джон (John) і його непрямих друзів (тобто друзів його друзів) і повертає як Джона, так і будь-якого знайденого друга його друзів (табл 3.1).

```
MATCH (john {name: 'John'})-[:friend]->()-[:friend]->(fof)
```

```
RETURN john, fof
```

Таблиця 3.1. Результат запитів знаходження друзів

John	fof
Node [3]{name:" John"}	Node [1]{name:" Maria"}
Node [3]{name:" John"}	Node [2]{name:" Steve"}

У четвертому розділі було розроблено модель аналізу створення запитів на SQL та Cypher та порівняння їх між собою.

JOIN. Перерахуємо всіх осіб і фільми, в яких вони грали роль. Результати запитів відображено в таблиці 4.1.

**SQL:**

```
SELECT person.name, movie.title
```

```
FROM person
```

```
JOIN acted_in AS acted_in ON acted_in.person_id = person.id
```

```
JOIN movie ON acted_in.movie_id = movie.id;
```

**Cypher:**

```
MATCH (person:Person)-[:ACTED_IN]->(movie:Movie)
```

```
RETURN person.name, movie.title;
```

Таблиця 4.1. Перелік осіб та фільмів

Person.Name	Movie.Title
Laurence Fishburne	The Matrix
Keanu Reeves	The Devil's Advocate
Keanu Reeves	The Matrix
Hugo Weaving	The Matrix
Charlize Theron	Monster
Charlize Theron	The Devil's Advocate
Carrie-Anne Moss	The Matrix
Al Pacino	The Devil's Advocate

Щоб зробити речі дещо складнішими, можна знайти співавторів Кіану Рівса.

У SQL ми використовуємо самоз'єднання на таблицю осіб і приєднуємося до таблиці acted\_in один раз для Кіану і один раз для співавторів.

**SQL:**

```
SELECT DISTINCT co_actor.name
```

```
FROM person AS Keanu
```

```
JOIN acted_in AS acted_in1 ON acted_in1.person_id = keanu.id
```

```
JOIN acted_in AS acted_in2 ON acted_in2.movie_id = acted_in1.movie_id
JOIN person AS co_actor
  ON acted_in2.person_id = co_actor.id AND co_actor.id <> keanu.id
WHERE keanu.name = 'Keanu Reeves';
```

У **Cypher** використовується шаблон з двома шляхами, які націлені на той самий вузол Movie (табл 4.2).

```
MATCH (keanu:Person)-[:ACTED_IN]->(movie:Movie),
      (coActor:Person)-[:ACTED_IN]->(movie)
WHERE keanu.name = 'Keanu Reeves'
RETURN DISTINCT coActor.name;
```

Таблиця 4.2. Пошук співавторів

CoActor.Name
Al Pacino
Charlize Theron
Hugo Weaving
Laurence Fishburne
Carrie-Anne Moss

П'ятий розділ присвячено розгляду питання охорони праці та безпеці у надзвичайних ситуаціях. Перший підрозділ містить інформацію щодо забезпечення санітарно-гігієнічних умов праці персоналу ІТ сфери. Розташування робочої техніки та меблів при роботі з нею. У другому підпункті розроблено інструкцію для персоналу, щодо дії у випадку виникнення пожежних ситуацій.

## **ЗАГАЛЬНІ ВИСНОВКИ**

Магістерська наукова робота присвячена ознайомленню з графовою СУБД Neo4j та мовою запитів Cypher, їх властивостями та особливостями. Для розкриття всього потенціалу бази даних та мови запитів, було приведено декілька прикладів застосування її на практиці. Підсумовуючи результати отримані в ході дослідницько-наукової роботи, можна зробити висновок, що для задач з великим обсягом даних та складною структурою графова база даних Neo4j та мова запитів Cypher дуже добре для них підходить.

## АНОТАЦІЯ

**Гуменюк А.Б.** Графова база даних **Neo4j**, для представлення мови **Cypher**. – На правах рукопису.

Магістерська наукова робота на здобуття освітньої «Магістр системного аналізу». – Чорноморський національний університет імені Петра Могили, Миколаїв, 2019.

**Мета роботи** – вивчити, як працюють графові моделі баз даних і область їх застосування, розібрати СУБД Neo4j і мову запитів Cypher, розробити запити і виконати порівняльний аналіз ефективності та проаналізувати висновки на основі отриманих даних.

**Об’єкт дослідження** – системи управління і способи використання графових баз даних.

**Предмет дослідження** – функції і можливості мови запитів Cypher для графової бази даних Neo4j, особливості використання запитів для порівняльного аналізу ефективності роботи.

**Завдання дослідження:** вивчення графової моделі бази даних, вивчення області застосування графових баз даних, розроблення кількох запитів відповідно до обраної інформаційної системою для СУБД MySQL і СУБД Neo4j, порівняння ефективності запитів.

**Методи дослідження.** У процесі виконання поставлених у роботі завдань були використані методи аналізу літературних джерел, аналіз існуючих методів та технологій зберігання інформації, практична робота з графовими базами даних для перевірки і засвоєння набутих знань і навичок.

Методична частина дипломного проекту включає в себе розроблену практичні роботи для ознайомлення з оболонкою графової бази даних Neo4j та використання REST, індексів і транзакцій.

В спеціальній частині «Охорона праці» розглянуто умови праці на робочих місцях персоналу комп’ютерної фірми, а також розроблено інструкцію дій персоналу у випадку надзвичайній ситуації.

В цілому робота складається зі 89 сторінок, 6 таблиць, 21 рисунки, 1 додаток.

**Ключові слова:** *автострахування, шахраї в страхуванні, шахрайські дії в автострахуванні, шахрайські дії з банківськими картками, системи виявлення шахраїв в страхових.*