

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

Павлович Діана Сергіївна

УДК 004.65

Аналіз страхових випадків з використанням Neo4j

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 9³⁰ 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. Болубаш

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Використання графових баз даних для систем аналізу та попередження шахрайства в страховій сфері. Графові бази даних дають перевагу при роботі з даними, для яких зв'язки відіграють важливу роль, особливо якщо потрібно буде пройти по зв'язках у кілька рівнів вглиб.

Гіганти ІТ, такі як Google, Facebook, LinkedIn і PayPal, активно використовують графові бази даних. Це дозволяє їм повністю розкрити потенціал взаємозв'язків. Графові бази даних значно швидше оперують високо пов'язаними даними, ніж реляційні і NoSQL бази даних, крім того, мають вбудовану підтримку графових алгоритмів.

У магістерській роботі, необхідно розглянути питання поточного стану страхового ринку України, які інформаційні технології використовуються щодо запобіганню шахрайських дій, та реалізувати приклад для автостраховання на базі Neo4j.

Метою магістерської наукової роботи є дослідження основних принципів функціонування графових баз даних, способи роботи з ними та демонстрація можливостей графової бази даних Neo4j, на прикладі виявлення шахрайських дій в сфері страхування.

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

Предмет досліджень – функції і можливості графової бази даних Neo4j, особливості її використання для аналізу шахрайських дій в сфері страхування.

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 виявлення шахраїв;

- проаналізувати можливості і принципи роботи системи управління базами даних на прикладі сфери страхування;
- продемонструвати можливості графової бази даних Neo4j, на прикладі аналізу шахрайських дій в сфері страхування.

Методи дослідження. У процесі виконання поставлених у роботі завдань були використані методи аналізу літературних джерел, аналіз існуючих методів та технологій виявлення шахраїв, 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. Павлович Д. С., Кулаковська І. В. «Аналіз страхових випадків з використанням Neo4j» / 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. 69.

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

ОСНОВНИЙ ЗМІСТ РОБОТИ

За кордоном страховики витрачають величезні суми на боротьбу зі злочинами в сфері страхування. Близько 15% їхнього бюджету йде на запобігання махінацій. І ці витрати себе виправдовують. Американськими страховиками підраховано, що кожен долар, вкладений в антишахрайські заходи, окупається в десяти-, а то й двадцятикратному розмірі.

В Україні за різними даними збиток страхових компаній від дій шахраїв складає 10-15%, а в деяких компаніях до 25% від загальної суми страхового відшкодування, але ці дані приведені експертами. Справжні дані, страхові компанії не оприлюднюють. Так, згідно статистики за 2009 рік, фальсифікація страхових випадків склала 1.5-2% від загальної кількості випадків. Грошовий еквівалент збитків від шахрайства складає 5-10%.

Для запобігання шахрайських дій у сфері страхування в Україні, бажано взяти до уваги методичку запобігання та боротьби з шахрайськими діями наших закордонних сусідів, та багаторічний досвід розвинутих держав.

Перший розділ описує проблеми боротьби з шахрайством в світовій практиці та в Україні. Нижче наведено кроки боротьби з страховими злочинцями.

Шахрайством у страхуванні займається все більше осіб. Починають формуватися організовані злочинні групи, що спеціалізуються на цих злочинах. У зв'язку із цим страховики починають взаємне співробітництво, обмінюються інформацією, ведуть спільні реєстри і бази даних. На цьому етапі починають створюватися групи для дослідження і класифікації страхових злочинів і боротьби з ними. Посилюється корупція в середовищі працівників страхових компаній, поліцейських, медпрацівників, митників.

Другий розділ присвячено огляду графових баз даних, їх властивостей. Представлено розроблену інформаційну модель для подальшої роботи з Neo4j.

Графова база даних — це база даних, яка використовує структури графів для семантичних запитів з вершинами, ребрами та властивостями для

представлення та зберігання даних. Ключовим поняттям системи є «граф» (або «ребро» або «відносини»), який безпосередньо пов'язує елементи даних у сховищі. Відносини дозволяють безпосередньо об'єднувати дані у сховищі, а в багатьох випадках, вибрати однією операцією.

Це відрізняється від реляційних баз даних, які за допомогою реляційних систем управління дозволяють керувати даними без накладання аспектів реалізації; наприклад, зв'язки між даними зберігаються в самій базі даних на логічному рівні, і операції реляційної алгебри (наприклад, join можуть використовуватися для маніпулювання та повернення відповідних даних у відповідний логічний формат. Виконання реляційних запитів можливе за допомогою систем управління базами даних на фізичному рівні (наприклад, за допомогою індексів), що дозволяє підвищити продуктивність без зміни логічної структури бази даних.

Графові бази даних базуються на теорії графів і використовують вершини, ребра та властивості. Вершини або вузли представляють такі об'єкти, як люди, підприємства, облікові записи або будь-який інший об'єкт, який слід відслідковувати. Вони приблизно еквівалентні «запису», «відношенню» або «рядку» в реляційній базі даних або «документу» в документо-орієнтованій базі даних. Ребра, також називаються графи або відносини, це лінії, що з'єднують вузли до інших вузлів; вони представляють відносини між ними. Змістовні моделі виникають при розгляді зв'язків і взаємозв'язків вузлів, властивостей і ребер. Ребра є ключовими поняттями в графовій базі даних, що представляють собою абстракцію, яка безпосередньо не реалізована в інших системах.

Властивості це важлива інформація, яка стосується вузлів. Наприклад, якщо Вікіпедія була б одним із вузлів, вона могла б бути прив'язана до таких властивостей, як веб-сайт, довідковий матеріал або слово, яке починається з букви «в», залежно на яких аспектах «Вікіпедія» знаходиться в певній базі даних.

На відміну від реляційних баз, графові бази даних безпосередньо зберігають відносини між записами.

В третьому розділі розглядаються приклади з Neo4j в автострахованні, махінаціях з банківськими картками. На відміну від реляційних баз даних, Neo4j зберігає пов'язані між собою дані, які не є ні чисто лінійними, ані суто ієрархічними, що полегшує виявлення схем шахрайської діяльності незалежно від глибини або форми даних.

Універсальна модель графів властивостей Neo4j полегшує організаціям розробку моделей даних виявлення шахрайства, допомагаючи командам з безпеки працювати з темпами постійних шахраїв, наприклад у випадку аналізу моделей аварій та практичної реалізації для знаходження аналогій.

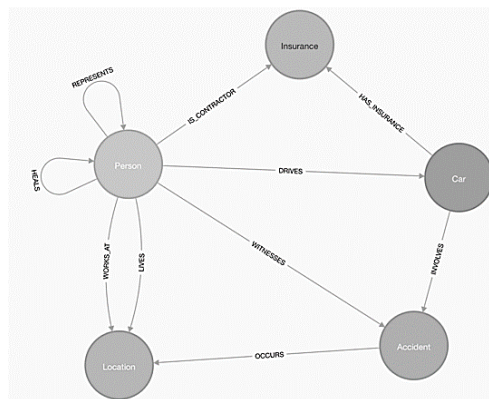


Рис. 1. Модель аварії у вигляді графа

Власний механізм обробки графів Neo4j підтримує високопродуктивні запити графіків на великих наборах даних, що дозволяє виявляти шахрайство в режимі реального часу.

Завдяки вбудованим функціям високої доступності Neo4j завжди доступні програми критичного виявлення шахрайства. Злочинець може вкрати інформацію про особу від десяти різних людей. Потім злочинець змішує і співпадає з номерами соціального страхування, адресами, номерами телефонів і адресами електронної пошти, щоб створити «нові» синтетичні ідентичності, які потім використовуються для відкриття банківських і кредитних карт, а також персональних кредитних ліній.

Стандартні технології та технології, включаючи реляційні бази даних, більше не справляються. Шаблони шахрайства стають все більш складними, і все що стосується аналізу і виявлення моделей. У Neo4j транзакції зберігаються як

графі, в якому пов'язані фрагменти даних, що дозволяє легко візуалізує ці відносини в реальному часі і швидко знаходити шахрайські моделі.

Саме в цьому розділі було проаналізовано декілька прикладів шахрайства на основі запитів до баз даних та візуалізація отриманих результатів у формі графів за допомогою бази даних Neo4j. Дані приклади показують різницю у роботі з даними в реляційній базі та графовій базі даних, швидкодії та продуктивності.

Шахраї часто створюють і керують сценаріями «заміни» учасників для того, щоб використати схему на максимум. Таким чином, одна аварія може мати конкретну людину, яка грає роль водія. При іншому нещасному випадку одна й та сама особа може бути пасажиром або пішоходом, а в іншому свідком. Розумне використання ролей може генерувати велику кількість дорогих фальшивих аварій, навіть з невеликою кількістю учасників.

Розумне використання ролей може генерувати велику кількість дорогих фальшивих аварій, навіть з невеликою кількістю учасників.

У сценарії, зображеному на рисунку 2, шість осіб у змові призводять до трьох помилкових аварій. Кожна людина грає роль «водія» одного разу і «пасажира» двічі.

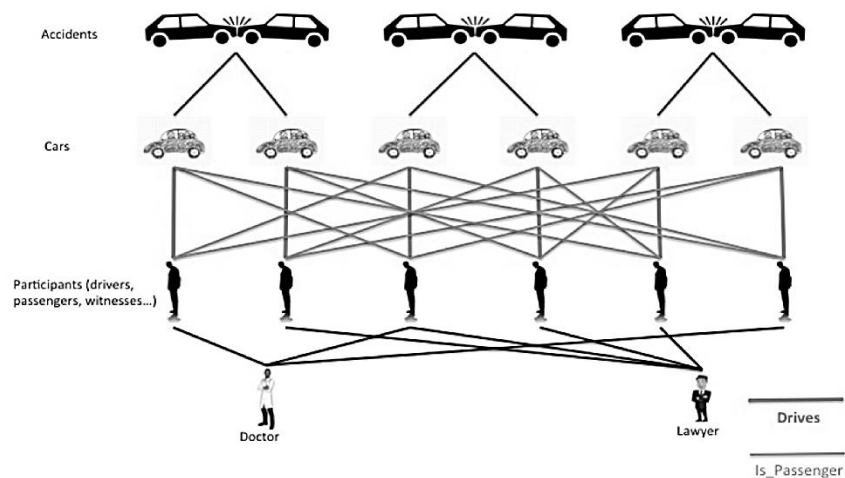


Рис. 2. Сценарій шахраїв, за участю декількох ролей

Графові бази даних, надають більш широкі можливості при аналізі великих обсягів даних. Доцільно їх використати для попередження (виявлення) шахрайських дій в банківській системі, якщо необхідно проаналізувати не об'єкти, а сутність зв'язків між ними.

Саме тому для подальшої роботи з виявленням шахрайських дій у страхуванні, обрано графову базу даних Neo4j, на основі якої, будуть розглянуті приклади вирішення шахрайських схем.

Четвертий 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.

У п'ятому розділі було розроблено практичні роботи для ознайомлення з роботою з базою даних Neo4j. Ознайомлення з веб-інтерфейсом, термінологією та CRUD операціями. Розглянуто, як Neo4j представляє дані у вигляді графа і як цей граф обходити. Описані методи та запити, які розкривають переваги роботи з оболонкою та мову Gremlin.

ЗАГАЛЬНІ ВИСНОВКИ

Магістерська наукова робота присвячена ознайомленню з графовою СУБД Neo4j, її властивостями та особливостями. Для розкриття всього потенціалу бази даних, було приведено декілька прикладів застосування її на практиці. Використовуючи реалізовану схему автошахрайства (рис. 2.), було проведено ряд запитів до бази даних в Neo4j, для візуалізації її роботи.

Підсумовуючи результати отримані в ході дослідницько-наукової роботи, можна зробити висновок, що для створення систем розкриття та упередження злочинних намірів в страховій сфері, доцільно використовувати Neo4j, для аналізу зв'язків між об'єктами.

АНОТАЦІЯ

Павлович Д.С. Аналіз страхових випадків з використанням **Neo4j**. – На правах рукопису.

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

Мета роботи – вивчити основні принципи функціонування графових баз даних, способи роботи з ними і продемонструвати можливості графової бази даних Neo4j, на прикладі аналізу шахрайських дій в сфері страхування.

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

Предмет дослідження – функції і можливості графової бази даних Neo4j, особливості її використання для аналізу шахрайських дій в сфері страхування.

Завдання 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 виявлення шахраїв; проаналізувати можливості і принципи роботи системи управління базами даних на прикладі сфери страхування; продемонструвати можливості графової бази даних Neo4j, на прикладі аналізу шахрайських дій в сфері страхування.

Методи дослідження. У процесі виконання поставлених у роботі завдань були використані методи аналізу літературних джерел, аналіз існуючих методів та технологій виявлення шахраїв, 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 мовою запитів Gremlin.

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: автостраховання, шахраї в страхуванні, шахрайські дії в автострахованні, шахрайські дії з банківськими картками, системи виявлення шахраїв в страхових.

ABSTRACT

to the master's scientific work on

“Analysis insurance fraud with Neo4j Database”

Student: Pavlovyh Diana Serhiivna

Scientific supervisor: Ph.D., Kulakovska Inessa Vasylivna

The given Master's Paper is aimed at research the basic principles of the functioning of graph databases, how to work with them and to demonstrate the capabilities of the Neo4j graph database, as an example of the analysis of fraudulent actions in the field of insurance.

The object of research - control systems and methods of using graph databases.

The subject of the study is the functions and capabilities of the Neo4j graph database, the peculiarities of its use for the analysis of fraudulent actions in the field of insurance.

Objectives of the study:

- to find out the concept of the graph database, the purpose and classification of control systems in graph databases;
- define the principles of designing graph database;
- features of the use of Neo4j;
- analyze existing methods and technologies for detecting fraudsters;
- analyze the possibilities and principles of the database management system on the example of the insurance industry;
- demonstrate the capabilities of the Neo4j graph database, based on an analysis of fraudulent activities in the field of insurance.

Research methods. In the process of fulfilling the tasks set forth in the work, methods of analysis of literary sources, analysis of existing methods and technologies for detecting fraudsters, and practical work with graph databases for testing and assimilating the acquired knowledge and skills were used.

The methodical part of the graduation project includes the developed practical works for acquaintance with the Neo4j graph database shell and the Gremlin query language.

In the special part “Labor Protection”, the working conditions at the workplaces of the personnel of a computer company were considered, and an instruction was developed for personnel actions in case of an emergency.

In general, the work consists of 89 pages, 6 tables, 21 figures, 1 annex.

Key words: *car insurance, fraudsters in insurance, fraudulent actions in auto insurance, fraudulent actions with bank cards, systems of detecting fraudsters in insurance.*

40 000 знаків з пробілами або 24 сторінки через півтора інтервали 14-м кеглем
 $0,4 \cdot 40000 = 16000$ знаків,