

ГРАФОВЫЕ БАЗЫ ДАННЫХ И ИХ СВОЙСТВА

Зорко П.А., Кулевич А.О., Хрол К.А.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Пилецкий И.И. – канд. физ.-мат. наук, доцент

Большая часть мировых данных хранится в базах данных (БД).

База данных – это информационная модель, позволяющая в упорядоченном виде хранить данные о группе объектов, обладающих подобным набором свойств [1]. Такой способ организации данных позволяет эффективно извлекать их. Данные можно просматривать, вставлять, обновлять и удалять по мере необходимости.

В настоящее время большой интерес приобретают графовые БД, т.к. они обеспечивают представление, более близкое к структуре реальных данных. Базы данных такого типа относятся к NoSQL базам данных.

Графовые базы данных обладают рядом преимуществ и достоинств по сравнению с другими БД: они, как и RDBS, обладают свойствами OLTP (Online Transaction Processing) & OLAP (Online Analytical Processing), поддерживают транзакции ACID (atomic, consistent, isolated и durable), чего не обеспечивает ни одна NoSQL БД.

Графовые технологии являются основой для построения интеллектуальных приложений, для применения алгоритмов искусственного интеллекта [2]. Еще одним отличием графовых баз данных является то, что они явно описывают зависимости между узлами данных, в то время как другие базы данных связывают данные неявными связями.

Главным отличием графовых баз данных от сетевых является то, что в сетевых базах данных нет возможности связывать произвольные записи друг с другом, что затрудняет работу с графо-ориентированными наборами данных, в то время как графовые базы данных предоставляют такую возможность.

Кроме того, сетевые базы данных используют записи с предопределенным числом полей, а графовые базы данных благодаря их гибкой структуре допускают произвольные пары ключ/значение как в узлах, так и в отношениях [3].

Графовые базы данных могут использоваться в самых разнообразных приложениях. Некоторые популярные области применения графовых баз данных включают:

- Социальные сети.
- Рекомендации по продукту в реальном времени.
- Сетевые схемы разнообразных физических объектов.
- Обнаружение мошенничества (в банках и сетях).
- Управление доступом.
- Поиск цифровых активов на основе графиков.
- Управление основными данными.
- Различные области искусственного интеллекта.

Использование графовых баз данных очень перспективно из-за простоты, наглядности и интуитивности графовой модели. Они часто не имеют схемы, что обеспечивает гибкость. Благодаря этому графовым моделям не требуется предварительно моделировать задачу в мельчайших подробностях, что очень удобно из-за быстро меняющихся бизнес-требований.

Присущая графам возможность расширения и модификации означает, что можно добавлять новые виды связей, новые узлы, новые метки и новые подграфы в существующую структуру, не нарушив при этом существующих запросов и функционала приложения.

Одной из веских причин выбора графовых баз данных является большой прирост производительности при работе со взаимосвязанными данными по сравнению с реляционными базами данных и NoSQL-хранилищами. Они могут более естественно масштабироваться до больших

наборов данных, поскольку обычно им не нужны операции объединения, которые часто могут быть дорогостоящими.

В отличие от других моделей, графовая модель данных может обеспечивать визуализацию с использованием обходов графа. Таким образом, предоставляется возможность быстро исследовать и просматривать данные графа, чтобы получить ценную информацию.

Графовые базы данных создаются и управляются с помощью графовых систем управления базами данных (СУБД). В настоящее время разработаны различные графовые СУБД.

Первой и одной из популярнейших на данный момент платформ является платформа Neo4j Graph. Она включает компоненты, которые позволяют разрабатывать приложение с поддержкой графов.

Графическая платформа Neo4j оптимизирована для хранения, картографирования, анализа и обхода сетей связанных данных [4].

С помощью интуитивно понятного отображения данных и связей между ними Neo4j обеспечивает возможность использования в реальном времени таких технологий, как: искусственный интеллект, машинное обучение; управление основными данными; управление идентификацией и доступом. Механизм хранения Neo4j использует массивы фиксированного размера для хранения данных графа и может обеспечить доступ к данным за время $O(1)$ [5].

В Neo4j используется SQL-подобный язык запросов Cypher. Cypher – это декларативный язык запросов для взаимодействия с графовой БД. Это мощный, оптимизированный для графов язык, который понимает и использует преимущества хранимых соединений.

Он состоит из предложений, ключевых слов и выражений, таких как предикаты и функции (например: MATCH, WHERE, LIMIT, AND, p.firstName = "Vsevolod").

```
MATCH (p:Professor)
WHERE p.lastName = "Petrov"
RETURN p
```

На основе полученных знаний авторами была разработана собственная графовая база данных, которая описывает наборы связей между группами специальности «Информатика и технологии программирования», преподавателей, которые ведут занятия у данных групп, и соответствующие им кафедры.

Графовое представление данной БД приведено на рисунке 1.

В данном графе узлы – это объекты, а дуги – отношения между объектами. В разработанной базе данных используются узлы трех типов: преподаватель (Professor); группа (Group); кафедра (Department).

Для каждого узла-преподавателя указаны его ФИО, ученая степень, электронная почта, рабочий номер телефона и рабочая аудитория.



Рисунок 1 – Графовая база данных и ее представление

57-я Научная Конференция Аспирантов, Магистрантов и Студентов БГУИР, Минск, 2021

Для каждого узла-группы дополнительно указаны курс, ФИО старосты, количество человек в группе, а для групп 1–2 курсов – ФИО куратора группы.

Для каждого узла-кафедры дополнительно указаны ФИО заведующего кафедрой, а также полное название кафедры.

Результатом научной работы стала разработанная база данных, которую можно легко интегрировать в различные приложения. Так, например, она может быть использована в приложении, отслеживающем расписание групп. Можно оперативно получать данные о сотрудниках и студентах.

Список использованных источников:

1. База данных – Википедия [Электронный ресурс]. – 2021. – Режим доступа : https://ru.wikipedia.org/wiki/База_данных.
2. Интеллектуальная система комплексного анализа данных интернет-источников / М. П. Батура [и др.] // *BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов VI Междунар. науч.-практ. конф., Минск, 20–21 мая 2020 г. : в 3 ч. Ч. 1 / редкол. : В. А. Богуш [и др.]*. – Минск : Бестпринт, 2020. – С. 220–241.
3. Робинсон, Я. Графовые базы данных: новые возможности для работы со связанными данными / Я. Робинсон, Дж. Вебер, Э. Эцифрем ; пер. с англ. Р. Н. Рагимова ; науч. ред. А. Н. Кисилев. – 2-е изд. – М. : ДМК Пресс, 2016. – 256 с.
4. Graph Database Platform | Graph Database Management System | Neo4j [Электронный ресурс]. – 2021. – Режим доступа : <https://neo4j.com>.
5. *History of Databases and Graph Database – Bitnine Global Inc.* [Электронный ресурс]. – 2020. – Режим доступа : <https://bitnine.net/blog-graph-database/history-of-databases-and-graph-database/>.