



## Проблемы разработки графовых баз данных

*Г.Е. Засядко, А.В. Карпов*

*Московский политехнический университет*

**Аннотация:** В статье рассматриваются проблемы разработки графовых баз данных в условиях быстро растущего объема генерируемых и обрабатываемых данных. Целью статьи является анализ особенностей графовых моделей данных, демонстрация их преимуществ перед остальными, а так же постановка задачи о разработке нового способа представления графов в памяти компьютера.

**Ключевые слова:** база данных, структура данных, граф, модель данных, SQL, NoSQL, индексация, анализ данных, схема данных, представление графов.

В связи с последними тенденциями развития информационных технологий встает вопрос о необходимости увеличения гибкости и универсальности используемых структур данных. Повышается спрос в области анализа данных различного рода, как структурированных, так и не структурированных, зачастую требуется модифицировать не только данные, но и их структуру и взаимосвязи «на лету». Во многих существующих и новых информационных технологиях все больше преобладает «четвертая парадигма» [1]. Ставшие уже классическими для задач хранения и обработки информации SQL-базы данных не всегда отвечают предъявляемым требованиям. С каждым днем все большее применение находят различные NoSQL-решения. Причин этому несколько:

- резкий рост объемов обрабатываемых и хранимых данных [2];
- сложная и изменчивая структура данных;
- высокие требования к возможности распределенного хранения данных [3];
- нестандартные методы обработки данных.

Одним из самых перспективных направлений развития NoSQL-решений являются графовые базы данных [4]. Преимущества данных решений:

- разработка без задания схемы данных;
-



- наглядность;
- легкая масштабируемость;
- широкие возможности по описанию сложных данных;
- описание данных с большим количеством связей.

На сегодняшний день из всех существующих технологий и инструментов на рынке представлено очень малое количество тех, которые позволяли бы с большой гибкостью описывать объекты реального, окружающего нас мира, хранить полученные данные и выполнять их обработку, анализ и изменение самих данных «на лету», не выходя за рамки изначально используемой базы данных.

Таким образом, при решении определенного круга задач, например, таких как всевозможные научные исследования, описание и отслеживание бизнес-логики и бизнес-процессов, существует острая потребность в некотором хранилище, которое позволяло бы при помощи него осуществлять полный цикл сбора, хранения, обработки данных в единой адаптирующейся модели. Это могло бы дать значительный выигрыш не только в скорости доступа к данным и их обработке, но и в удобстве использования, упрощении и ускорении процесса разработки и интеграции, снижении избыточности данных и количестве применяемых инструментов, а соответственно и затрат на поддержку информационной системы в будущем.

Подобные многоцелевые решения пригодятся в первую очередь в сфере анализа данных, динамически изменяющихся в процессе работы с ними. Это те случаи, где можно столкнуться с ограничениями и неудобствами условно стандартных и распространенных средств хранения, таких как реляционные базы данных или многие NoSQL-решения, предоставляющие в большинстве случаев лишь широкие возможности по хранению и быстрому доступу к данным на физическом уровне.



Проблема заключается не только в скорости доступа к хранимым данным, но так же и в правильной их логической организации с минимизацией стоимости каждой операции, совершаемой над данными.

Одним из центральных методов представления данных об окружающем мире в информационных системах, является метод, основанный на выделении из доступного объема информации метаданных и непосредственно самих данных [5].

Суть данного подхода состоит в некотором анализе всего множества описываемых объектов, выделении на его основе некоторого количества различных классов и описании реальных объектов при помощи них. Каждый выделенный класс обладает определенным набором свойств, присущих ему.

В реальном же мире зачастую невозможно сразу определить класс объекта, либо объекты, отнесенные к одному классу, будут обладать различным набором свойств. Нет ничего идеального и поддающегося строгому описанию. Такой подход применим лишь для определенного круга задач, и способен отражать суть только статически описываемых объектов.

Перспективной моделью данных, подходящей под вышеперечисленные критерии и лишенной необходимости предварительной классификации описываемых объектов, является графовая модель данных, достаточно гибкая для описания данных любого рода и сложности, в силу своих структурных особенностей. Графы одинаково хорошо подходят для представления как слабо, так и для сложно структурированных объектов и систем [6, 7].

Для наглядной оценки в таблице №1 приведены результаты проведенного сравнения на соответствие основным критериям наиболее распространенных моделей баз данных.

Таблица №1. Сравнительные критерии для различных моделей баз данных



	Key-Value модель данных	Реляционная модель данных (РМД)	Объектно-ориентированная модель данных	Документ-ориентированная модель	RDF (Resource Description Framework) модель данных	Мультиграфовая модель (Property graph)
Разработка без задания схемы данных	-	-	-	+	+	+
Удобство описания объектов	-	+	+	+	-	+
Удобство оперирования большими объемами данных	+	-	-	+	+	+
Удобство оперирования большим количеством связей	-	-	-	-	+	+
Наглядность представления данных и их связей	-	-	-	-	+	+
Линейная масштабируемость	+	-	+	+	+	+
Возможность использования в децентрализованных системах	+	-	+	+	+	+

Помимо всего прочего графовую модель данных возможно представить как некоторый рекурсивный тип данных [8], что дает широкие возможности по применению различных способов его хранения и обработки. Здесь открывается целый спектр методов и алгоритмов для работы как с самой структурой графа и его логической организацией [9], так и непосредственно с данными, хранящимися в нем.

Области применения графовых моделей и методы работы с ними активно изучаются, ведется множество исследований, направленных на



унификацию языков запросов к базам данных, выбор канонических моделей и моделей ресурсов [10]. Такие исследования, безусловно, необходимы и играют одну из ключевых ролей.

Неотъемлемой частью успешной обработки данных на графовых моделях являются способы хранения и методы доступа к данным внутри используемых структур. Обычно способы хранения графов в памяти обладают высокой избыточностью и лишены возможности быстрого обхода и поиска, что приводит к сильному снижению скорости обработки. Таким образом, сейчас перед нами стоит задача по пересмотру и анализу существующих способов представления и хранения графовых моделей в памяти компьютера, выявление их особенностей, преимуществ и недостатков. Необходимо сконцентрировать силы на поиске нового, гибкого, легко адаптируемого под поставленные задачи способа представления графов. Искомая структура данных, должна будет способствовать оптимизации используемой памяти и логической организации, тем самым позволив снизить избыточность, ускорить процесс индексации и поиска хранимых данных, понизив алгоритмическую сложность обработки.

### Литература

1. Tony Hey, Stewart Tansley, Kristin Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond: Microsoft Research, 2009. p. 4.
2. Jeffrey Ullman, Database systems - the complete book. Pearson, 2009. p. 4.
3. Andrew S. Tanenbaum, Distributed Systems: Principles and Paradigms. Maarten van Steen, 2016. pp. 34 - 49.
4. Ian Robinson, Jim Webber, Graph Databases. O'Reilly, 2015. pp. 8 - 10.
5. Gavin Powell, Beginning Database Design. Wrox, 2006. p. 219.
6. Л.А. Гинис, Развитие инструментария когнитивного моделирования для исследования сложных систем // Инженерный вестник Дона, 2013, №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1806.



7. С.В. Астанин, Н.В. Драгныш, Н.К. Жуковская, Вложенные метаграфы как модели сложных объектов // Инженерный вестник Дона, 2012, №4. URL: [ivdon.ru/ru/magazine/archive/n4p2y2012/1434](http://ivdon.ru/ru/magazine/archive/n4p2y2012/1434).
8. Niklaus Wirth, Algorithms and Data Structures. Prentice-Hall, Inc, 1986. pp. 109-111.
9. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, Introduction to Algorithms, Third Edition. The MIT Press, 2010. pp 587 - 768.
10. С. А. Ступников, Отображение графовой модели данных в каноническую объектно-фреймовую информационную модель при создании систем интеграции неоднородных информационных ресурсов // Труды XV Всероссийской научной конференции RCDL, 2013. URL: [synthesis.ipi.ac.ru/synthesis/publications/13rcdl-graph/13rcdl-graph.pdf](http://synthesis.ipi.ac.ru/synthesis/publications/13rcdl-graph/13rcdl-graph.pdf).

### References

1. Tony Hey, Stewart Tansley, Kristin Tolle, The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond: Microsoft Research, 2009. p. 4.
  2. Jeffrey Ullman, Database systems - the complete book. Pearson, 2009. p. 4.
  3. Andrew S. Tanenbaum, Distributed Systems: Principles and Paradigms. Maarten van Steen, 2016. pp. 34 - 49.
  4. Ian Robinson, Jim Webber, Graph Databases. O'Reilly, 2015. pp. 8 - 10.
  5. Gavin Powell, Beginning Database Design. Wrox, 2006. p. 219.
  6. L.A. Ginis Inzhenernyj vestnik Dona (Rus), 2013, №3. URL: [ivdon.ru/ru/magazine/archive/n3y2013/1806](http://ivdon.ru/ru/magazine/archive/n3y2013/1806).
  7. S.V. Astanin, N.V. Dragnysh, N.K. Zhukovskaya. Inzhenernyj vestnik Dona (Rus), 2012, №4. URL: [ivdon.ru/ru/magazine/archive/n4p2y2012/1434](http://ivdon.ru/ru/magazine/archive/n4p2y2012/1434).
  8. Niklaus Wirth, Algorithms and Data Structures. Prentice-Hall, Inc, 1986. pp 109-111.
  9. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, Introduction to Algorithms, Third Edition. The MIT Press, 2010. pp. 587 - 768.
-



- 10.S. A. Stupnikov, Otobrazhenie grafovoy modeli dannykh v kanonicheskuyuob"ektno-freymovuyu informatsionnuyu model' prisozdanii sistem integratsii neodnorodnykhinformatsionnykh resursov. Trudy XV Vserossiyskoy nauchnoy konferentsii RCDL, 2013. URL: [synthesis.ipi.ac.ru/synthesis/publications/13rcdl-graph/13rcdl-graph.pdf](http://synthesis.ipi.ac.ru/synthesis/publications/13rcdl-graph/13rcdl-graph.pdf).