

**РАЗРАБОТКА ГРАФИЧЕСКОЙ НОТАЦИИ МОДЕЛИРОВАНИЯ
МУЛЬТИМОДЕЛЬНЫХ БАЗ ДАННЫХ**

Клемешева А. С. (Университет ИТМО)

**Научный руководитель – доцент, кандидат физико-математических наук,
Ромакина О. М.** (Университет ИТМО)

Введение. Мульти модельные базы данных, обеспечивающие поддержку нескольких моделей данных в рамках одной СУБД, возникли в результате потребности разработчиков в хранении и обработке большого объема данных разных типов. Такие базы данных предоставляют возможность выбора наиболее подходящей модели данных для каждого конкретного случая без переноса данных в другую систему.

NoSQL технологии, в том числе мульти модельные базы данных, предполагают отсутствие жесткой схемы данных, тем не менее, для работы с данными знание их структуры необходимо. Как правило, в нереляционных СУБД схема данных формируется не на этапе проектирования, а извлекается позже, проводится «обратный процесс проектирования» [1]. Такой подход имеет свои недостатки: при отсутствии логической модели, созданной при проектировании и демонстрирующей структуру базы данных в том виде, в котором она планировалась разработчиком, может быть утеряна часть информации о том, почему были приняты те или иные решения о разделении, например, на разные сущности. Тогда позже в процессе реструктуризации и оптимизации базы данных могут быть не учтены эти детали.

Стремительное развитие NoSQL технологий в целом и мульти модельных систем в частности привело к тому, что у разработчиков нет единого стандарта моделирования таких баз данных. Существуют определенные рекомендации по проектированию NoSQL баз данных, зависящие от конкретной модели, но даже касательно одной модели данных нет идеального решения и единого мнения относительно верного подхода к ее моделированию [2].

Следует учитывать, что у мульти модельных баз данных структура сложнее, чем у одно модельных, поскольку для корректной совместной работы нескольких моделей используются меж модельные ссылки или встраивание одной модели в другую [3]. Сложность архитектуры таких систем требует тщательного анализа и моделирования при проектировании. Разработчикам следует фиксировать принятые решения относительно структуры мульти модельных баз данных на логической модели и в документации проекта.

Основная часть. Мульти модельные базы данных чаще всего поддерживают следующие распространенные модели данных [4]: реляционную, графовую, документоориентированную, колоночную, ключ-значение.

Многие мульти модельные базы данных основаны на одной главной модели, чаще всего реляционной. Это связано с тем, что реляционные СУБД изначально были самым популярным типом баз данных, и когда они перестали в полной мере справляться с растущим объемом неструктурированных или слабоструктурированных данных, реляционные СУБД нередко расширяли для поддержки также и других моделей, более эффективных в работе с такими данными.

В работе [6] была предложена графическая нотация для моделирования основных видов NoSQL моделей данных: документной, графовой, колоночной и ключ-значение. Эта нотация также подходит для моделирования реляционных баз данных, поскольку основана на известной нотации Мартина («Вороньи лапки»).

В публикации рассматривалась возможность моделирования одно модельных нереляционных баз данных, и в ней не были предусмотрены никакие обозначения, позволяющие моделировать мульти модельные базы данных. Однако, поскольку мульти модельные базы данных сочетают в себе несколько моделей, все обозначения, подходящие этим моделям, актуальны и для мульти модельной схемы.

Но помимо этого требуется специальное обозначение, которое будет указывать, что логическая модель относится именно к мультимодельной базе данных, а не одномодельной. Кроме того, может возникнуть такая ситуация, когда при разработке мультимодельной базы данных разработчику необходимо показать, что определенные данные могут использоваться только в одной модели из тех, которые поддерживает эта база данных, а в других эти данные не могут быть представлены – например, из соображений производительности.

Поэтому помимо обозначения мультимодельной базы данных необходимо сформировать символы отдельных моделей данных.

В итоге были выбраны обозначения, продолжающие идею применения в нотации простых символов. Они представлены далее:

- 1) [[[M]]] – Multi-model
- 2) [R] – Relational
- 3) [D] – Document
- 4) [G] – Graph
- 5) [K] – Key-value
- 6) [C] – Column
- 7) [GEO] – Geographic

Таким образом, одна модель – это пара квадратных скобок с заключенной между ними первой (первыми) буквами названия модели, а мультимодельная база данных – три пары таких же скобок и буква «M» в середине. Были рассмотрены обозначения для основных и наиболее распространенных моделей, модель «Geographic» приведена как пример того, что эти обозначения при необходимости могут быть легко расширены на новые модели данных.

Располагать эти символы на схемах предполагается следующим образом: логическая модель делится на несколько полей, каждое поле содержит только те сущности, которые относятся к модели, обозначение которой стоит в левом верхнем углу поля.

Значок модели отображается в левом верхнем углу по двум причинам. Во-первых, при проектировании схемы данных необходимо первым делом определить, какая модель требуется для хранения и обработки этих данных, во-вторых, схема может расширяться, обычно это происходит в направлении «вправо и вниз», следовательно, из соображений удобства чтения и во избежание необходимости сдвигать значок при добавлении новых сущностей, он должен располагаться в условном начале схемы.

Заключение. Задача логического моделирования систем, поддерживающих одновременно несколько моделей данных, является актуальной, ее решение позволит облегчить разработку и поддержку таких СУБД. Разработанная нотация, расширенная для моделирования мультимодельных баз данных, может стать одним из вариантов решения. Предложенный подход был апробирован на предметной области, предполагающей хранение большого объема данных с высокой вариативностью их видов.

Список использованных источников

1. Abdelhedi F. и др. Reverse Engineering Approach for NoSQL Databases. // Big Data Analytics and Knowledge Discovery, 22nd International Conference, DaWaK 2020, Bratislava, Slovakia, September 14–17. 2020. С. 60–69.
2. Vera H. и др. Data Modeling and NoSQL Databases - A Systematic Mapping Review // ACM Computing Surveys. 2021. Т. 54. С. 1–26.
3. Koupil P., Hricko S., Holubová I. A universal approach for multi-model schema inference // Journal of Big Data. 2022. Т. 9. № 1. С. 97.
4. DB-Engines Ranking [Электронный ресурс]. URL: <https://db-engines.com/en/ranking> (дата обращения: 25.01.2025).
5. Ромакина О. М. и др. Разработка графической нотации моделирования NoSQL баз данных // XXI Век: итоги прошлого и проблемы настоящего плюс. 2024. Т. 13. № 2 (66). С. 32–42.