

А.В. МАКАРОВ

**Вопросы применения NoSQL СУБД
в информационно-
образовательных системах**

УДК 004.04:004.822

Муромский институт
(филиал) ФГБОУ ВПО
«Владимирский
государственный
университет имени
А.Г. и Н.Г. Столетовых»

В статье анализируются вопросы применения NoSQL СУБД в информационно-образовательных системах. Формулируются требования к СУБД на основе предлагаемой модели учебно-образовательного контента. Приводятся результаты сравнения производительности NoSQL и реляционных СУБД в задачах хранения и обеспечения доступа к учебно-образовательной информации. Обосновывается выбор средств программной реализации информационно-образовательной системы на уровне хранения данных.

In article the issues of using NoSQL DBMS in information-educational systems are analyzed. Requirements to DBMS on the basis of offered model of an educational content are formulated. Results of comparison of productivity NoSQL and relational DBMS in problems of storage and providing of access to the educational information are resulted. The choice of means of program realization of information-educational system at data storage level is proved.

Сегодня рост востребованности многопользовательских информационно-образовательных систем, построенных на основе веб-технологий, актуализирует целый круг вопросов, связанных с технологическим обеспечением их разработки. От правильности выбора соответствующих средств напрямую зависят дидактические возможности и эффективность функционирования образовательной системы в целом. Сегодня одним из ключевых вопросов при разра-

ботке информационно-образовательной системы является вопрос выбора оптимальных средств хранения и обеспечения доступа к учебно-образовательному контенту. Учебно-образовательный контент, в своей основе представляемый дидактическим материалом и результатами образовательной деятельности учащихся, обладает рядом специфических характеристик, которые определяют выбор соответствующей СУБД:

– Слабая структурируемость. Учебный материал, а в особенности результаты образовательной деятельности учащихся, крайне сложно целостно представить посредством единовременно спроектированных структур, таких как таблицы реляционных СУБД. Невозможность изначально предусмотреть все нюансы предметной области наряду с ее постоянно изменчивостью ведет к дополнительным сложностям перепроектировки структуры базы данных в случае реляционных СУБД [1].

– Семантическая составляющая. При решении задачи сохранения исходной целостности формализуемого дидактического материала необходимо учитывать одну из его важнейших составляющих — смысловые связи между единицами учебного материала.

– Сетевая структура. Очевидным является тот факт, что множественные семантико-логические взаимосвязи учебно-образовательного контента представляются в виде сетевой структуры. Таким образом, при выборе типа СУБД следует учитывать, что учебно-образовательный контент по своей сути является «знаниями», но не «данными» в общем смысле этого слова.

В ходе исследований вопросов формализации учебного материала и результатов образовательной деятельности в рамках информационно-образовательных систем с учетом обозначенных выше специфических требований предметной области, авторами была предложена следующая информационная модель учебно-образовательного контента [2], базовая единица которой представлена на рисунке 1.

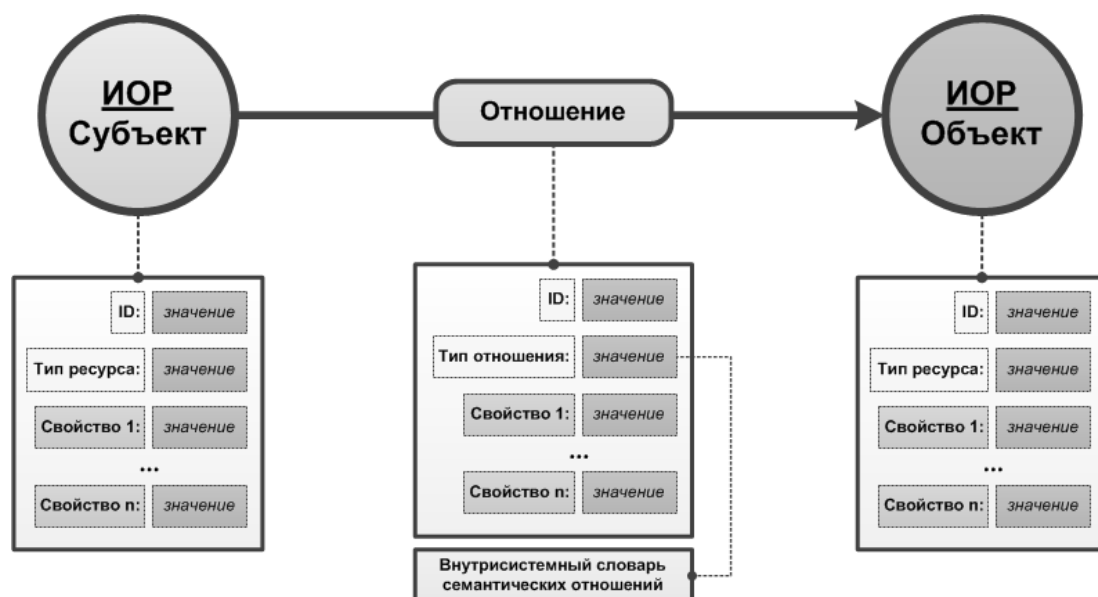


Рис. 1 — Базовая единица информационной модели учебно-образовательного контента

В качестве базовой единицы информационной модели учебно-образовательного контента предлагается использовать тройку «субъект»–«предикат/отношение»–«объект», как атомарную составляющую информационной модели, представляющей собой взвешенный орграф [3]. При этом «субъект/объект» и «предикат/отношение» обладают набором некоторых свойств (списком пар «ключ–значение»), определяющих их содержание.

Исходя из приведенных спецификаций предметной области, а также предлагаемой её информационной модели, становится очевидно, что использование в этих условиях реляционных баз становится нерациональным. Причина этого заключается главным образом в том, что реляционные СУБД эффективны только лишь в том случае, пока данные не содержат множество взаимосвязей, требующих при их выборке ресурсоемкое объединение таблиц [1]. Оптимальным выбором в рассматриваемом случае, когда предметная область определяется информационным полем образовательного процесса, являются графовые NoSQL СУБД. Определим основные критерии (на базе предлагаемой модели) оценки графовых СУБД на предмет соответствия их решаемым задачам в рамках информационно-образовательных систем.

Таблица 1

Сравнение графовых NoSQL СУБД

| Критерии | Neo4J | HyperGraphDB | DEX | InfoGrid |
|-------------------------|-------|--------------|-----|----------|
| Свойства вершин | + | + | + | + |
| Свойства ребер | + | +/- | + | - |
| Направленность ребер | + | - | + | + |
| Язык запросов | + | - | - | - |
| Обход графа | + | + | + | + |
| Поддержка RDF-хранилища | + | - | - | - |
| Поддержка OWL-онтологий | +/- | - | - | - |
| ACID | + | +/- | - | +/- |

Помимо качественных критериев оценки СУБД, не менее важными критериями оценки эффективности ее работы является количественные показатели скорости выполнения запросов и физического размера БД при равном количестве хранимой информации. Обратимся к данным, полученным в ходе исследования производительности СУБД Neo4J в сравнении с широко распространенной СУБД MySQL [4]. Наиболее показательными является результаты сравнения скорости выполнения полнотекстовых поисковых запросов к символьным данным, составляющих основу учебной информации, СУБД MySQL и Neo4J, где L — длина поисковой строки (Таблица 2).

Таблица 2

Результаты выполнения запросов к символьным БД, мс

| | MySQL | Neo4J | MySQL | Neo4J | MySQL | Neo4J |
|---------------|--------|-------|--------|-------|--------|-------|
| | L=5 | | L=7 | | L=8 | |
| 1000char8k | 15,0 | 41,6 | 11,1 | 41,6 | 15,6 | 36,3 |
| 5000char8k | 131,6 | 41,8 | 126,0 | 33,0 | 91,9 | 41,6 |
| 10000char8k | 269,0 | 41,5 | 263,1 | 42,6 | 249,9 | 41,5 |
| 100000char8k | 3224,1 | 41,5 | 3077,4 | 41,8 | 2834,4 | 36,4 |
| 1000char32k | 41,6 | 42,6 | 31,9 | 41,4 | 31,9 | 35,4 |
| 5000char32k | 242,9 | 41,5 | 188,5 | 38,5 | 152,0 | 41,5 |
| 10000char32k | 468,8 | 41,6 | 382,1 | 41,5 | 298,8 | 36,3 |
| 100000char32k | 4859,1 | 33,3 | 4703,3 | 41,5 | 2769,6 | 41,5 |

Результаты, приведенные в таблице 2, наглядно показывают, насколько в рассматриваемых задачах полнотекстового поиска в

условиях множественных взаимосвязей данных СУБД MySQL проигрывает в производительности СУБД Neo4J. Учитывая, что операции поиска информации в информационно-образовательных системах является одной из наиболее важных, время их исполнения является критичным параметром в выборе соответствующей СУБД.

На сегодняшний день СУБД Neo4J наиболее полно удовлетворяет сформулированным требованиям и является оптимальным выбором средств хранения и обеспечения доступа к учебно-образовательному контенту в рамках информационно-образовательных систем, построенных на базе веб-технологий. Поддержка RDF-хранилищ и OWL-онтологий обеспечивает возможность тесной интеграции с уже существующими базами знаний и различными семантическими веб-сервисами на базе спецификаций, предложенных консорциумом W3C. Наряду с рассмотренными в таблице 1 характеристиками СУБД Neo4J стоит отметить кроссплатформенность СУБД благодаря реализации на языке Java, а также несколько типов развертывания СУБД и доступа к ней из клиентского приложения:

- Прямое внедрение в приложение Java;
- Интеграция с приложением по технологии Object-mapping'a;
- Neo4J Standalone Server;
- Neo4J REST API;

В настоящее время автором ведется разработка информационно-образовательной системы на основе веб-фреймворка Ruby on Rails и СУБД Neo4J, что позволяет сделать ряд выводов на основе практического опыта использования рассматриваемых решений.

Сегодня использование СУБД Neo4J значительно упрощает разработку информационно-образовательных систем, сокращает время на проектирование логической модели БД, гибкая информационная модель позволяет сохранять целостность исходной семантико-логической структуры учебно-образовательного контента, а также обеспечивает большую скорость выполнения запросов к БД по сравнению с реляционными СУБД. Встроенные алгоритмы обхода графа, а также возможность легкой разработки новых, значительно упрощает решение задач построения и корректировки индивидуальных образовательных траекторий, выполнения структурного

анализа дидактического материала и операций автоматизированной диагностики уровня усвоения учебного материала.

Литература

1. *M. Kleppmann* Should you go beyond relational databases? – 2009 – URL: <http://carsonified.com/blog/dev/should-you-go-beyond-relational-databases/>. (дата обращения: 05.02.2012).

2. *Макаров А.В., Ан А.Ф.* Концептуальные основы проектирования информационно-образовательной среды непрерывного физического образования // Современные проблемы науки и образования. – 2011. – № 6; URL: www.science-education.ru/100-5211 (дата обращения: 06.02.2012).

3. *Макаров А.В.* Метод автоматизированной диагностики уровня усвоения учебного материала на основе семантико-логических взаимосвязей контрольных заданий // Алгоритмы, методы и системы обработки данных. — 2011, №18. URL: http://amisod.ru/index.php/index.php?option=com_content&view=article&id=61:amisod-2011-3-18-makarov&catid=15:amisod-3-18-2011&Itemid=111 (дата обращения: 07.02.2012).

4. *C. Vicknair, M. Macias, Z. Zhao, X. Nan, Y. Chen, D. Wilkins* A Comparison of a graph database and relation database. – 2010 – URL: <http://cs.olemiss.edu/~ychen/publications/conference/> (дата обращения: 07.02.2012).