

ПРИМЕНЕНИЕ ДОКУМЕНТО-ОРИЕНТИРОВАННЫХ СУБД В СИСТЕМАХ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Н.А. Хмурович¹, О.А. Мацкевич²

¹ *Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь, dev.nkh@gmail.com*

² *Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь, omatskevich92@gmail.com*

Abstract. The article contains the main idea of centralized distance learning systems and their objectives, the classical solution for data storage, the declaration of distributed data bases and the principals of information storage in NoSQL data bases in the distance learning systems.

Введение. Централизованные системы дистанционного обучения имеют большие перспективы на рынке программных продуктов. Такой метод предполагает самостоятельное изучение материала на расстоянии, периодический контроль прогресса и результатов освоения.

В современных веб-приложениях мы имеем дело с множеством запросов и действий пользователей. Такого рода информация имеет ценность с точки зрения поведенческого анализа субъекта. На основании действий пользователя могут определяться параметры: количество времени, проведенного пользователем за изучением теоретического материала по дисциплине, количество и распределение потраченного времени на выполнение различного рода заданий, своевременность начала выполнения учебных заданий. Возможен анализ вовлечения обучаемого в образовательный процесс по отдельным дисциплинам. Важной функциональностью является проверка уникальности произведенной письменной работы, сравнение с существующими аналогами.

При росте количества пользователей важной задачей становится хранение и быстрая агрегация большого объема данных. Серьезными требованиями к современным веб-приложениям являются быстрый отклик и максимальная производительность системы при работе с большим объемом информации.

Классический подход к хранению данных: использование реляционных СУБД. Реляционные СУБД стали стандартом при проектировании слоя хранения данных в современных приложениях.

Современные реляционные СУБД соответствуют требованиям ACID: Atomicity — Атомарность, Consistency — Согласованность, Isolation — Изолированность, Durability — Надежность.

В основу реляционной модели заложены жесткие схемы, строго регламентирующие структуру и связи сущностей.

Реляционные базы данных обеспечивают надёжное хранение данных, атомарность крупных операций и постоянную согласованность.

Однако для достижения подобного поведения используются различные механизмы, которые накладывают штрафы на производительность и ограничивают возможности масштабирования.

Использование NoSQL решений для хранения данных. С приходом огромных массивов информации и распределенных систем стало ясно, что обеспечить для них одновременно транзакционность набора операций и получить высокую доступность и быстрый отклик — невозможно.

Эта идея была была поолжена в основу CAP теоремы, которая гласит о том, что реализация распределенных вычислений не может достичь одновременно трёх свойств:

Consistency (Согласованность), Availability (Доступность), Partition Tolerance (Терпимость к разделению)

На практике существуют системы типа

CA: +Availability +Consistency -Partition tolerance

CP: +Consistency +Partition tolerance -Availability

AP: +Availability +Partition tolerance -Consistency

Понятие распределенных баз данных. С лавинообразным ростом пользователей и количества информации увеличились требования к производительности хранилищ данных. Необходимость обрабатывать большое количество информации за разумное время столкнулась с проблемой вертикальной масштабируемости баз данных.

Выходом из ситуации является горизонтальное масштабирование, когда несколько независимых серверов соединяются между собой, и каждый владеет своей репликой или частью данных, и обрабатывает только часть запросов. В такой архитектуре для повышения мощности хранилища достаточной мерой является добавление нового сервера в кластер. Процедурами шардинга, репликации, обеспечением отказоустойчивости, перераспределения данных в случае добавления дополнительного сервера в таких системах занимается СУБД.

Общие характеристики NoSQL баз данных.

1. Представление данных в виде агрегатов (aggregates).

Документоориентированные хранилища пропагандируют агрегирование и встраивание данных в документ, чтобы оперировать с этими сущностями как с целостными объектами.

2. В большинстве своём это open source решения, они бесплатны.

3. Возможность автоматически распределять данные между серверами.

4. Использование памяти, прозрачное кэширование - содержимое коллекций активно кэшируется для выборки.

Вопрос применения документоориентированных баз данных для построения систем дистанционного обучения зависит от цели и конечной аудитории пользователей. Если необходимо создать локальную систему для нужд одного университета, то острой необходимости задумываться о масштабировании нет. Однако, если мы ставим целью создать централизованную систему регионального уровня, то масштабирование является важной возможностью. В этом случае современные NoSQL СУБД могут стать очень эффективным инструментом для решения этой проблемы.

Литература

1. BASE: An ACID alternative - Dan Pritchett, <http://queue.acm.org/detail.cfm?id=1394128>
2. Why NoSQL? - Couchbase, <http://www.couchbase.com/why-nosql/nosql-database>
3. <http://ru.wikipedia.org/wiki/ACID>