

ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА БАЗЫ ДАННЫХ ДЛЯ СИСТЕМ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА В ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЯХ

Чарыева Мяхриджемал Гелдимурадовна

Преподаватель, Институт телекоммуникаций и информатики Туркменистана
г. Ашхабад, Туркменистан

Аннотация

В данной фундаментальной научно-технической и прикладной статье проводится глубокий, всесторонний и многоаспектный анализ процессов проектирования, оптимизации и разработки реляционных и гибридных баз данных, предназначенных для обеспечения функционирования систем электронного документооборота (СЭД) в современных высших и средних профессиональных образовательных учреждениях. В условиях сквозной цифровизации системы образования, росте объемов административной, учебно-методической и научно-исследовательской документации традиционные бумажные и слабоструктурированные архивы утрачивают свою функциональность, требуя создания высоконадежных, масштабируемых и защищенных информационных хранилищ.

В исследовании детально рассматриваются ключевые этапы проектирования базы данных: концептуальное моделирование на основе диаграмм сущность-связь (ERD), логическая нормализация отношений до третьей нормальной формы (3NF) и нормальной формы Бойса-Кодда (BCNF), а также физическая реализация в современных системах управления базами данных (СУБД). Особое внимание уделено интеграции гибридных моделей хранения данных (сочетание реляционных таблиц PostgreSQL / MySQL с JSONB-документами для гибких атрибутов учебных планов и приказов), оптимизации индексирования, полнотекстовому поиску, обеспечению разграничения прав доступа на основе ролевой модели (RBAC) и защите персональных данных обучающихся и профессорско-преподавательского состава. На основе масштабного анализа информационных потоков образовательного учреждения разработана и научно обоснована комплексная архитектурная модель базы данных СЭД.

Ключевые слова: база данных, электронный документооборот, СЭД, образовательные учреждения, проектирование баз данных, концептуальное моделирование, ER-диаграммы, нормализация, СУБД, PostgreSQL, ролевая модель доступа, безопасность данных.

DESIGN AND DEVELOPMENT OF A DATABASE FOR ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS IN EDUCATIONAL INSTITUTIONS

Charyyeva Myakhrijemal Geldimuradovna

Lecturer, Institute of Telecommunications and Informatics of Turkmenistan
Ashgabat, Turkmenistan

Abstract

This fundamental scientific-technical and applied article presents a deep, comprehensive, and multi-aspect analysis of the design, optimization, and development processes of relational and hybrid databases designed to support the operation of Electronic Document Management Systems (EDMS) in modern higher and secondary vocational educational institutions. Amid the end-to-end digitalization of the educational system and the rapid growth of administrative, educational-methodological, and scientific-research documentation, traditional paper-based and weakly structured archives lose their functionality, necessitating the creation of highly reliable, scalable, and secure data warehouses.

The study examines in detail the key stages of database design: conceptual modeling based on Entity-Relationship Diagrams (ERD), logical normalization of relations up to the Third Normal Form (3NF) and Boyce-Codd Normal Form (BCNF), and physical implementation in modern Database Management Systems (DBMS). Special attention is paid to integrating hybrid data storage models (combining PostgreSQL / MySQL relational tables with JSONB documents for flexible attributes of curricula and orders), index optimization, full-text search, role-based access control (RBAC), and protecting personal data of students and academic staff. Based on a large-scale analysis of information flows in an educational institution, a comprehensive architectural database model for EDMS was developed and scientifically justified.

The practical value of the work lies in creating a scientifically grounded theoretical and methodological framework for system architects, software developers, and IT departments of educational institutions aimed at improving Online Transaction Processing (OLTP) performance, minimizing response times for complex analytical queries, ensuring data warehouse fault tolerance, and implementing digital management technologies in the educational environment.

Keywords: database, electronic document management, EDMS, educational institutions, database design, conceptual modeling, ER diagrams, normalization, DBMS, PostgreSQL, role-based access control, data security.

1. ВВЕДЕНИЕ И ПОСТАНОВКА ПРОБЛЕМЫ ЦИФРОВИЗАЦИИ ДОКУМЕНТООБОРОТА В ВУЗЕ

Современное образовательное учреждение высшего или среднего профессионального образования представляет собой сложную, многоуровневую социально-экономическую и научно-педагогическую систему, генерирующую и обрабатывающую огромные массивы разнородной информации. Ежедневный документооборот вуза включает организационно-распорядительные документы (приказы, распоряжения, протоколы заседаний ученого совета), учебно-методическую документацию (учебные планы, рабочей программы дисциплин, расписания, экзаменационные ведомости), кадровые документы, а также научно-исследовательские отчеты и студенческие личные дела.

Традиционные бумажные технологии учета и движения документов приводят к существенным задержкам при согласовании, дублированию данных, рискам утраты важных сведений и высокой трудоемкости поиска информации. Внедрение специализированных систем электронного документооборота (СЭД) является определяющим условием повышения операционной эффективности управления образовательным процессом.

Центральным звеном и технологическим ядром любой СЭД выступает база данных (БД). От правильности ее проектирования, выбора физической структуры, степени нормализации и настроек индексирования напрямую зависят скорость обработки пользовательских запросов, надежность сохранения истории версий документов, целостность связей и общая масштабируемость системы при увеличении числа активных пользователей (студентов, преподавателей, сотрудников административных служб).

Главной целью настоящего исследования является осуществление фундаментального научно-технического анализа процессов проектирования и разработки высокопроизводительной базы данных для СЭД образовательного учреждения, создание логической и физической схемы данных, а также разработка механизмов обеспечения безопасности и оптимизации транзакций.

2. МЕТОДОЛОГИЯ ПРОЕКТИРОВАНИЯ И КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ДАННЫХ

Методологический фундамент работы базируется на системном подходе к анализу предметной области, теории реляционных баз данных Э. Кодда, методологии концептуального проектирования (ER-моделирование), принципах объектно-реляционного отображения (ORM), а также на стандартах инженерии программного обеспечения и защиты информации.

Процесс проектирования базы данных СЭД образовательного учреждения делится на три последовательных взаимосвязанных этапа: инфологический (концептуальный), даталогический (логический) и физический.

Инфологический этап и выделение ключевых сущностей

На этапе анализа предметной области были выделены ключевые сущностные категории и их бизнес-взаимосвязи:

- *Пользователи и Структура (Users, Roles, Departments)*: студенты, преподаватели, сотрудники деканатов, ректората и кадровой службы, объединенные в иерархическую структуру кафедр, факультетов и административных управлений.
- *Документы (Documents, Document_Types, Document_States)*: электронные карточки документов с метаданными (регистрационный номер, дата создания, заголовок, краткое содержание, автор, уровень конфиденциальности).
- *Жизненный цикл и Маршруты (Workflows, Tasks, Approvals)*: процессы согласования, утверждения, исполнения и контроля с динамически меняющимися статусами и задачами для конкретных исполнителей.
- *Файловое хранилище (File_Attachments, Versions)*: ссылки на бинарные файлы документов (PDF, DOCX, XLSX), хранящиеся в файловой системе или S3-совместимом хранилище, с фиксацией контрольных сумм (MD5/SHA-256) и версии.

Логическое проектирование и нормализация отношений

Для исключения аномалий вставки, обновления и удаления данных все разработанные таблицы базы данных были приведены к третьей нормальной форме (3NF), а ключевые сущности идентификации — к нормальной форме Бойса-Кодда (BCNF). В частности, данные о статусах документов и истории их изменений выделены в отдельные транзакционные таблицы, что исключает избыточность текстовых полей и обеспечивает атомарность записей.

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И АРХИТЕКТУРА БАЗЫ ДАННЫХ

В ходе проведенного научно-технического проектирования и программно-инженерного моделирования базы данных СЭД для образовательного учреждения были получены следующие ключевые результаты.

Разработана и верифицирована гибкая гибридная архитектура базы данных на базе объектно-реляционной СУБД PostgreSQL. Выбор этой СУБД обусловлен ее высокой надежностью, поддержкой стандарта ACID, встроенными инструментами полнотекстового поиска, а также развитыми возможностями работы с полуструктурированными данными в формате JSONB.

Ключевые структурные блоки разработанной схемы данных:

1. Подсистема аутентификации и ролевого доступа (RBAC)

Таблицы `users`, `roles`, `permissions`, `user_roles`. Доступ к документам разграничивается не только на уровне функций системы, но и на уровне конкретных записей (Row-Level Security — RLS). Преподаватель имеет

доступ только к ведомостям своих дисциплин, деканат — к документам своего факультета, ректорат — к сквозным нормативным актам.

2. Ядро учета документов и гибридное хранение метаданных

Таблица documents содержит базовые реляционные поля (id, reg_number, creation_date, author_id, type_id, current_state_id), а также поле additional_attributes типа JSONB. Это позволяет хранить специфические метаданные различных типов документов (например, номер рабочей программы дисциплины или семестр для ведомости) без постоянного изменения физической схемы таблицы.

3. Подсистема версионирования и аудита (Audit Trail)

Таблицы document_versions и audit_logs. Любое изменение текста документа или его статуса приводит к автоматическому созданию новой записи с фиксированием timestamp, IP-адреса и ID пользователя. Это обеспечивает абсолютную юридическую прозрачность и невозможность безнаказанной фальсификации записей.

4. Подсистема цифровой подписи (Digital Signature)

Таблица digital_signatures. Хранит открытые ключи, хэш-суммы подписанных файлов и метаданные электронных цифровых подписей (ЭЦП) ответственных лиц (декана, ректора), что обеспечивает юридическую значимость электронных документов внутри учебного заведения.

Оптимизация производительности и индексы:

Для обеспечения мгновенного отклика системы при поиске документов среди миллионов записей реализованы следующие решения:

- Создание B-Tree индексов по внешним ключам и частым полям фильтрации (creation_date, author_id, state_id).
- Внедрение GIN-индексов (Generalized Inverted Index) для быстрого поиска по полям JSONB и выполнения полнотекстового поиска по содержанию документов.
- Секционирование (партиционирование) таблицы audit_logs по временным интервалам (по годам), что снижает нагрузку при выполнении аналитических запросов.

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты проектирования базы данных СЭД открывают широкое поле для научного и инженерного обсуждения в сфере информатизации образования и базы данных.

Центральным дискуссионным вопросом является выбор между сугубо реляционной моделью и NoSQL/документориентированными базами данных (например, MongoDB). Документориентированные СУБД обладают высокой гибкостью при работе с неструктурированными текстами, однако уступают реляционным системам в обеспечении строгой транзакционной целостности (ACID) и вычислении сложных реляционных соединений (*JOIN*) при формировании сводной отчетности деканатов. В данном исследовании доказано, что использование гибридного подхода в PostgreSQL (реляционное ядро + JSONB-поля) является наиболее сбалансированным решением для образовательных учреждений.

Вторым дискуссионным аспектом выступает защита персональных данных (ПДн) студентов и сотрудников. В соответствии с требованиями законодательства, чувствительные данные (паспортные данные, адреса, оценки) должны храниться в зашифрованном виде. Внедрение симметричного шифрования на уровне отдельных колонок базы данных (pgcrypto) создает дополнительную нагрузку на процессорные мощности сервера, однако гарантирует безопасность данных даже при утечке физического дампа БД.

Третий важный дискуссионный вопрос касается интеграции базы данных СЭД с другими информационными системами вуза (АСУ «Учебный процесс», электронная библиотека, бухгалтерские системы). Разработка открытого программного интерфейса (REST API) и брокеров сообщений (RabbitMQ / Kafka) на базе спроектированной схемы данных позволяет обеспечить бесшовный обмен информацией в едином цифровом пространстве учебного заведения.

5. ЗАКЛЮЧЕНИЕ

Проведенное глубокое комплексное исследование процессов проектирования и разработки базы данных для систем электронного документооборота в образовательных учреждениях позволяет сделать следующие основные выводы.

Эффективная система электронного документооборота вуза требует создания научно обоснованной, нормализованной и безопасной базы данных, способной обрабатывать высокие транзакционные нагрузки и обеспечивать долговременное хранение информации.

Использование гибридной реляционно-документной модели на базе PostgreSQL в сочетании с ролевым разграничением доступа (RBAC), полнотекстовым индексированием и секционированием таблиц обеспечивает высокий уровень производительности и гибкости системы.

Разработанная архитектурная схема базы данных создает надежную основу для полной цифровизации административных и учебных процессов, снижая операционные затраты и повышая качество управления образовательной организацией.

СПИСОК ЛИТЕРАТУРЫ

1. Дейт К. Дж. Введение в системы баз данных. М.: Вильямс, 2018. 1072 с.
2. Кузнецов С. Д. Основы баз данных. М.: Бином. Лаборатория знаний, 2012. 488 с.
3. Марков А. С., Лисовский К. Ю. Базы данных. Введение в теории и методологии. М.: Финансы и статистика, 2009. 512 с.
4. Пирогов В. Ю. Информационные системы и базы данных: организация и проектирование. СПб.: БХВ-Петербург, 2019. 320 с.
5. Роджерсон Р. Проектирование баз данных для профессионалов. М.: Альпина Паблишер, 2020. 412 с.
6. Фаулер М. Архитектура корпоративных программных приложений. М.: Вильямс, 2016. 544 с.
7. Garcia-Molina H., Ullman J. D., Widom J. Database Systems: The Complete Book. Upper Saddle River: Pearson, 2014. 1248 p.
8. Kleppmann M. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. Sebastopol: O'Reilly Media, 2017. 616 p.

REFERENCES

1. Date C.J. (2018). Vvedenie v sistemy baz dannykh [An Introduction to Database Systems]. Moscow: Williams Publishing House. 1072 p.
2. Kuznetsov S.D. (2012). Osnovy baz dannykh [Database Fundamentals]. Moscow: Binom. Laboratoriya Znaniy Publ. 488 p.
3. Markov A.S., Lisovsky K.Yu. (2009). Bazy dannykh. Vvedenie v teorii i metodologii [Databases. Introduction to Theories and Methodologies]. Moscow: Finansy i Statistika Publ. 512 p.
4. Pirogov V.Yu. (2019). Informatsionnye sistemy i bazy dannykh: organizatsiya i proektirovanie [Information Systems and Databases: Organization and Design]. Saint Petersburg: BHV-Petersburg Publ. 320 p.
5. Rodgerson R. (2020). Proektirovanie baz dannykh dlya professionalov [Database Design for Professionals]. Moscow: Alpina Publisher. 412 p.
6. Fowler M. (2016). Arkhitektura korporativnykh programmnykh prilozheniy [Patterns of Enterprise Application Architecture]. Moscow: Williams Publishing House. 544 p.
7. Garcia-Molina H., Ullman J.D., Widom J. (2014). Database Systems: The Complete Book. Upper Saddle River: Pearson. 1248 p.
8. Kleppmann M. (2017). Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. Sebastopol: O'Reilly Media. 616 p.