

Разработка системы управления сбором, хранения и контроля Big Data в задачах геоакустики

Алена В. Хрыпченко¹

¹ *Институт вычислительной математики и математической геофизики СО РАН, проспект Академика Лаврентьева 6, г. Новосибирск, 630090, Россия*

Аннотация

Разработка системы управления базой геоакустических данных направлена на обеспечение эффективного хранения, обработки и доступа к этим данным. С учетом потребностей пользователей в сферах геофизики и геологии, система создается с целью обеспечить оперативный и точный доступ к информации для проведения научных исследований. В данной статье рассмотрим разработку системы для автоматизации управления большими объемами данных в задачах геоакустики с использованием фреймворка Spring, библиотеки визуализации Highcharts и базы данных MySQL. Результаты включают в себя разработанный инструмент для быстрого поиска, анализа и визуализации геоакустических данных, что позволит улучшить качество научных исследований и принимать обоснованные решения. Практическая значимость данной системы проявляется в повышении производительности и качества работы специалистов в области геофизики и геологии, сокращении времени на обработку данных и увеличении эффективности деятельности в области геоакустических исследований.

Ключевые слова

СУБД, геоакустические данные, система, MySQL, Spring, реляционная база данных, нереляционная база данных, Highcharts.

1. Введение

В настоящее время, управление большим объемом данных стало одним из важных аспектов в различных областях науки и технологий. С постоянным увеличением количества информации, ее разнообразием и сложностью, возникает необходимость в эффективной системе управления базами данных (СУБД). Эффективное управление базой данных позволяет хранить, организовывать, обрабатывать и анализировать данные с высокой скоростью и точностью.

В работе будут рассмотрены основные принципы работы СУБД в задачах геоакустики, а также методы и подходы к проектированию и разработке системы управления базой данных. Будет рассмотрена роль языка структурированных запросов SQL в создании и управлении базы данных, а также принципы работы распределенных баз.

Также в работе обсуждается вопрос о выборе наиболее подходящей СУБД для геоакустических задач. Рассматриваются основные характеристики, такие как производительность, масштабируемость, надежность и безопасность, и их роль в выборе подходящей системы.

Технологические аспекты разработки и внедрения базы данных также будут обсуждаться в данной работе. Описываются основные этапы процесса разработки СУБД, включая проектирование базы данных.

6th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2024), July 01–05, 2024, Irkutsk, Russia
EMAIL: ahrypchenko@yandex.ru (A. 1)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2024 Workshop Proceedings
DOI: 10.47350/ICCS-DE.2024.24

Практические аспекты внедрения базы данных в реальные системы также являются важной темой статьи. Будут рассмотрены проблемы и решения, связанные с миграцией данных, управлением изменениями, резервным копированием и восстановлением данных, а также безопасностью базы данных.

2. Постановка задачи

Геоакустические данные — это информация, получаемая при помощи геоакустических методов и технологий. Геоакустические данные имеют огромное значение для науки и промышленности по ряду причин.

В научных исследованиях они позволяют узнать больше о структуре земной коры и геологических процессах. Эти данные помогают геологам и геологоразведчикам определить глубину и состав слоев коры, местоположение и характеристики нефтяных и газовых месторождений, а также обнаружить и изучить подземные водные ресурсы и пустоши. Они также помогают исследователям в изучении динамики землетрясений и вулканической активности.

В промышленности геоакустические данные используются для разведки и добычи полезных ископаемых, таких как нефть и газ. Они позволяют определить вероятность наличия месторождения и оптимизировать процессы добычи, уменьшая затраты и увеличивая эффективность. Кроме того, геоакустические данные могут быть использованы при планировании строительства промышленных объектов, таких как нефтяные или газопроводы, позволяя избежать потенциальных проблем, связанных с геологическими условиями.

При работе с большими объемами данных в геоакустике возникают следующие трудности:

1. Ограниченные вычислительные ресурсы: обработка и анализ больших объемов данных требуют значительных вычислительных мощностей. Неправильная организация вычислительного процесса может привести к длительным задержкам и низкой производительности.

2. Затруднения в хранении данных: большие объемы данных требуют высокой емкости хранилища и эффективного способа их организации. Неправильное хранение данных может привести к потере информации или затруднить доступ к необходимым данным.

3. Высокая стоимость обработки данных: обработка больших объемов данных в геоакустике требует специализированного программного обеспечения и оборудования. Это может потребовать значительных финансовых затрат.

4. Сложность анализа данных: большие объемы данных могут содержать различные типы информации, что требует сложного анализа и обработки. Это может вызвать сложности в выделении и интерпретации нужной информации.

5. Проблемы с качеством данных: большие объемы данных могут быть подвержены различным искажениям и шумам. Это может привести к неточным результатам анализа или прогнозированию.

6. Угрозы безопасности данных: большие объемы данных требуют обеспечения их конфиденциальности и защиты от неавторизованного доступа. Недостаточная безопасность данных может привести к утечке информации или нарушению права на конфиденциальность [5].

Для оптимизации геоакустических данных ИВМиМГ было принято решение разработать систему по автоматизации, сбору и хранению Big Data.

3. База данных разрабатываемой системы

Для реализации системы по автоматизации и исследованию геоакустических данных была разработана база данных, состоящей из 23 таблиц. Более детально рассмотреть основные компоненты разработанной системы можно на рисунках ниже.

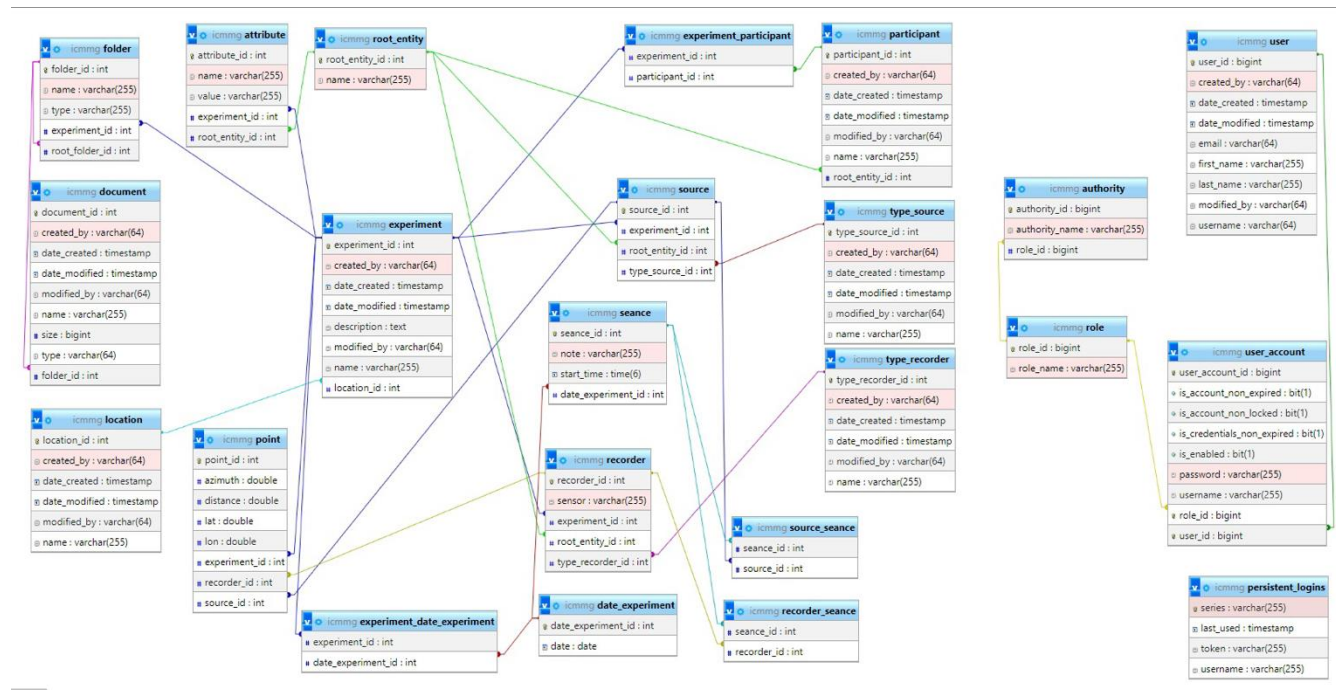


Рисунок 1: База данных для задач геоакустики

Разрабатываемая система построена с использованием Spring Framework на микросервисной архитектуре, представленной на рисунке 2, в которую входят: сервер авторизации. Он представляет из себя единую точку входа для клиентов системы использующий технологию OAuth2.0 для авторизации и Openid Connect для аутентификации. Gateway - для перенаправления запросов клиента к нужному сервису. Цепочка сервисов: реализующие сервисный слой и изменяющие состояние системы. Они связываются с хранилищами данных или выполняющие иную функцию, например, как Email Service рассылки (используется для отправки логов об ошибках на почту администратора). А также Web client для пользователей в браузере.

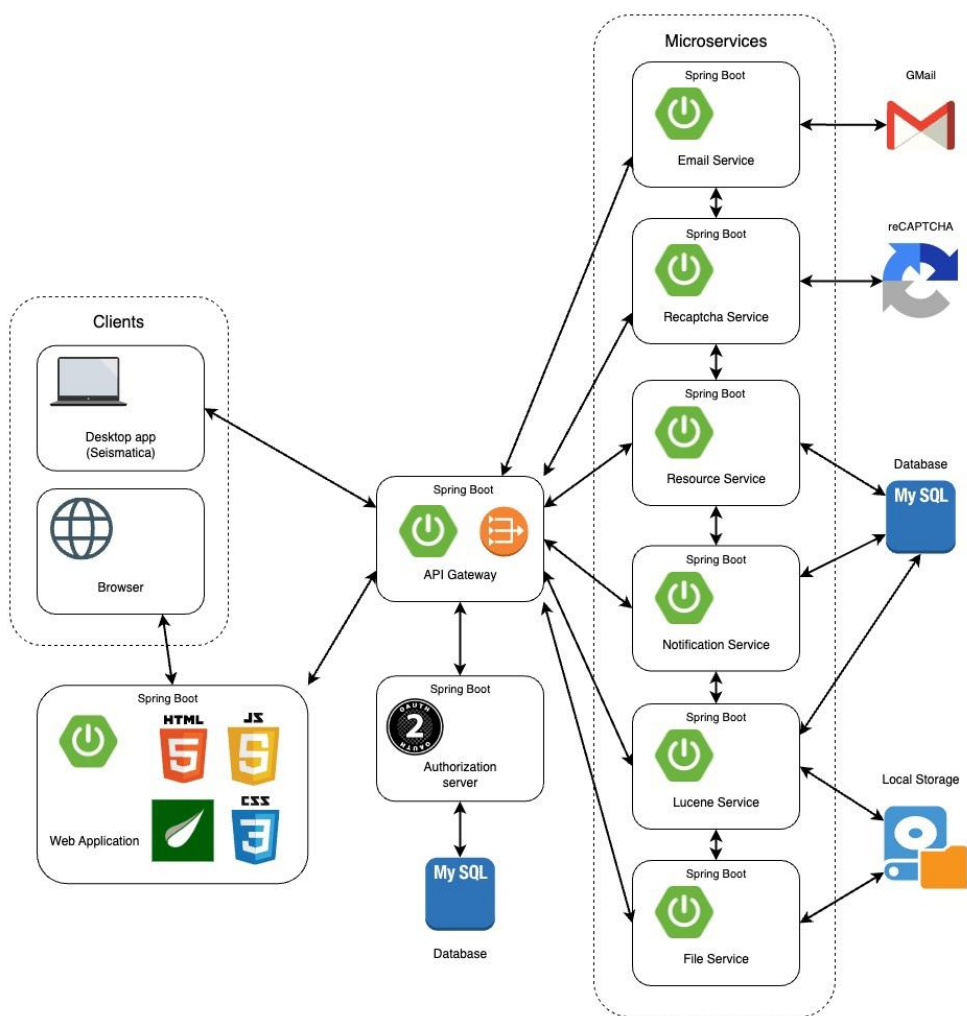


Рисунок 2: Архитектура системы

4. Краткая характеристика результатов

Мной была разработана система по управлению базой данных в задачах геокустики. После добавления системным администратором пользователя в систему, зарегистрированный может войти в систему (рисунок 3). Для безопасности утечки информации была сделана двухфакторная аутентификация.

Username

Password

☐ Remember me

Login

Рисунок 3: Форма входа в систему

Для регистрации нового эксперимента пользователю нужно заполнить форму, где нужно ввести название эксперимента, участников, датчики, местоположение, место проведение, даты, краткое описание эксперимента (рисунок 4).

EXPERIMENT

разрез Ерунак - Крест

профиль полигон Быстровка – г. Анжеро – Судженск, г. Белово, г. Прокопьевск

Add Location +

File Edit View Insert Format Tools Table Help

Paragraph B I

Тематика работ: В июле 2002 года был поставлен эксперимент по развертыванию в полевых условиях полномасштабной малой сейсмической группы с апертурой 2км x 2км на базе мобильных вибросейсмических комплексов для регистрации волновых полей вибрационных и взрывных источников сейсмических волн. Целью эксперимента было исследование характеристик малой сейсмической группы на основе четырех регистрирующих комплексов ВИРС-К, ВИРС-М, РОСА-1 и РОСА-2, которая может быть развернута в любом районе в соответствии с задачами исследований.

Методика работ: В экспериментах 2002 года малая группа была развернута на запасном полевом аэродроме вблизи г. Бердска с взаимно перпендикулярным расположением линейных сейсмических антенн регистрирующих комплексов в виде креста. Схема расположения креста и его пространственная ориентация представлены на рис. 1.

Группа располагалась в 30 км от вибросейсмического полигона п. Быстровка, где установлены мощные вибраторы,

p > strong Press °C0 for help 271 words

PARTICIPANT

#1 ИВМиМГ СО РАН

Add Participant +

Add Field +

SOURCE

#1 Взрыв

Add Source +

Point

#1 54.1 87.4 Azimuth Distance

Add Point +

Add Field +

#1 100 тонн серии по 10 тонн

Field Value

#2 Взрыв

Add Source +

Point

#1 54.2 86.42 Azimuth Distance

Add Point +

Add Field +

#1 119 тонн серии по 3,9 тонны

Field Value

RECORDER

#1 РОСА Д

Add Recorder +

Sensor

Point

#1 54.39 83.05 Azimuth Distance

Add Point +

Add Field +

#2 РОСА

Add Recorder +

Sensor

Point

#1 54.39 83.04 Azimuth Distance

Add Point +

Add Field +

#3 ВИРС-М

Add Recorder +

Sensor

Point

#1 54.39 83.04 Azimuth Distance

Add Point +

Add Field +

DATE

#1 03.07.2002

Seance

#1 13:32:28

☐ #1 Взрыв
 ☒ #1 РОСА Д

☒ #2 Взрыв
 ☒ #2 РОСА

☒ #3 ВИРС-М

Note

#2 04.07.2002

Seance

#1 15:44:39

☒ #1 Взрыв
 ☒ #1 РОСА Д

☐ #2 Взрыв
 ☒ #2 РОСА

☒ #3 ВИРС-М

Note

Save

Cancel

File Manager

Рисунок 4: Формы регистрации нового эксперимента

После чего эксперимент можно просмотреть, редактировать и удалить, если на то есть права у пользователя (рисунок 5).

PARTICIPANT	
# 1	ИВММГ СО РАН

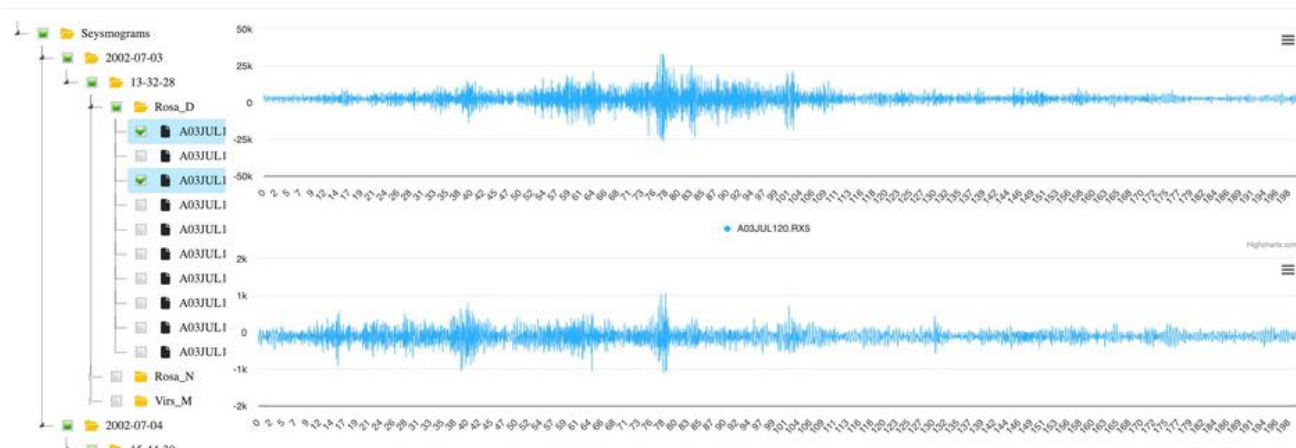
LOCATION	
профиль полигон Быстровка – г. Анжеро – Судженск, г. Белово, г. Прокопьевск	

RECORDER					
#	Name	Sensor	Point		
1	РОСА Д		Lat	Lon	Azimuth
			54.39	83.05	Distance
2	РОСА		Lat	Lon	Azimuth
			54.39	83.04	Distance
3	ВИРС-М		Lat	Lon	Azimuth
			54.39	83.04	Distance

SEANCE

#	Date	Seance			
1	03-07-2002	Start time	Source №	Recorder №	Note
		13:32:28	Взрыв	РОСА Д, РОСА, ВИРС-М	
2	04-07-2002	Start time	Source №	Recorder №	Note
		15:44:39	Взрыв	РОСА Д, РОСА, ВИРС-М	

SEISMOGRAM



FILES

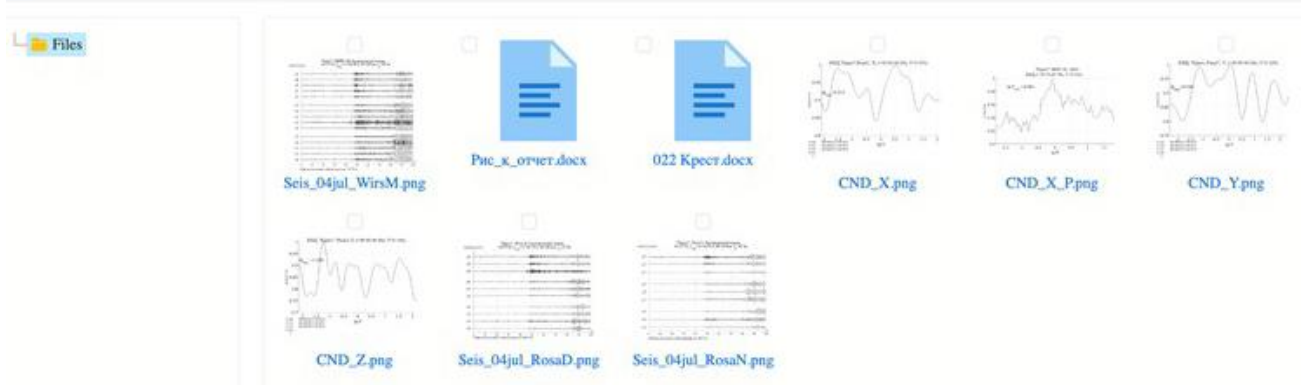


Рисунок 5: Формы просмотра и редактирования эксперимента

5. Заключение

Развитие технологий управления базами данных продолжает оказывать значительное влияние по обработке больших объемов информации. Важно не только следить за новшествами в этой области, но также уделять должное внимание практическим аспектам интеграции полученных решений. Именно проработанный подход к разработке и вдумчивое внедрение СУБД определяют успех её функционирования в долгосрочной перспективе.

Основной целью было предложить комплексный подход к разработке и внедрению систем управления базами данных, который позволял бы эффективно решать задачи хранения, управления и обработки данных. Для достижения этой цели были рассмотрены различные технологии по проектированию базы данных, их преимущества и недостатки. Была разработана система по управлению базой данных в задачах геоакстики. Подводя итоги хочется отметить,

что выбор технологии следует осуществлять с учётом конкретных задач и требований заказчиков.

Необходимо анализировать потребности в обработке и хранении данных, а также оценивать доступные ресурсы и ограничения, чтобы выбрать наиболее подходящую технологию.

В целом, статья представляет собой полезное исследование, которое дает читателям широкий обзор различных технологий и практических аспектов при разработке и внедрении систем управления базами данных. Результаты и выводы статьи могут быть полезны для специалистов, занимающихся разработкой и управлением баз данных, а также для научных исследователей в области информационных технологий.

6. Список источников

- [1] Kholodkov K.I., Aleshin I.M., Exact calculation of a posteriori probability distribution with distributed computing systems // Computer Research and Modeling, 2015, с. 539–542, 2020.
- [2] Malyshkin V. Active Knowledge, LuNA and Literacy for Oncoming Centuries // Programming Languages with Applications to Biology and Security: Essays Dedicated to Pierpaolo Degano on the Occasion of His 65th Birthday. Lecture Notes in Computer Science / Bodei C., Ferrari G., Priami C. (ред.). — Springer Cham, 2015. — Т. 9465. — 375 С. — DOI 10.1007/978-3-319-25527-9_19.
- [3] The sound amplifying forest with emphasis on sounds from wind turbines // Elis Johansson // Department of Civil and Environmental Engineering. Division of Applied Acoustics. Chalmers University of Technology. Sweden. – 2010. – 97 P.
- [4] L.George. Getting Started with Big Data Architecture. <http://blog.cloudera.com/blog/2014/09/getting-started-with-big-data-architecture/>
- [5] D. Laney 3D Data Management: Controlling Data Volume, Velocity, and Variety // META Group Inc. File 949. 2001.
- [6] Hu H., Wen Y., Chua T.-S., Li X. Towards scalable systems for big data analytics: a technology tutorial // IEEE Xplore Digital Library. 2014. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6842585>