

Власов Егор Александрович

студент IV курса Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (Северо-Западный институт управления РАНХиГС), Санкт-Петербург, Российская Федерация

ПРОЕКТИРОВАНИЕ АНАЛИТИЧЕСКИХ ВИТРИН И ИНТЕРФЕЙСОВ БИЗНЕС-АНАЛИТИКИ ДЛЯ МОНИТОРИНГА ГОСУДАРСТВЕННЫХ УСЛУГ

Аннотация. В статье рассматривается методология построения прикладных систем бизнес-аналитики для мониторинга электронного правительства. Основное внимание уделяется процессу проектирования аналитических витрин данных на базе гибридных хранилищ (PostgreSQL и ClickHouse), реализации интерактивных дашбордов в Apache Superset, а также внедрению современных практик UI/UX-прототипирования. Доказывается, что предложенный архитектурный подход позволяет кратно снизить время отклика системы и перевести процесс мониторинга из реактивной модели в проактивную, минимизируя риски нарушения регламентных сроков.

Ключевые слова: Business Intelligence, Apache Superset, аналитические витрины данных, ClickHouse, Data Marts, визуализация данных, мониторинг государственных услуг, SLA

Annotation. The article examines a methodology for developing application-oriented business intelligence systems for e-government monitoring. Particular attention is given to the design of analytical data marts based on hybrid data warehouse architectures combining PostgreSQL and ClickHouse, the implementation of interactive dashboards in Apache Superset, and the adoption of modern UI/UX prototyping practices. The study demonstrates that the proposed architectural approach can substantially reduce system response times and shift the monitoring process from a reactive to a proactive model, thereby minimizing the risks of non-compliance with established service deadlines.

Keywords: business intelligence, Apache Superset, analytical data marts, ClickHouse, data marts, data visualization, government service monitoring, SLA

Введение. Активная цифровизация государственных услуг и внедрение Системы межведомственного электронного взаимодействия (СМЭВ) приводят к экспоненциальному росту объемов логов и транзакционных данных. Для эффективного управления качеством предоставления услуг (как синхронных, так и длительных асинхронных процессов) необходимы аналитические инструменты, позволяющие трансформировать «сырые» данные в оперативные дашборды. Целью данного исследования является проектирование аналитического слоя и интерфейсов визуализации, обеспечивающих прозрачность межведомственных процессов и возможность предиктивной аналитики инцидентов и нарушений SLA (Service Level Agreement).

Гибридная архитектура хранилища и витрины данных (Data Marts). Для обеспечения высокой производительности BI-системы при работе с большими массивами транзакционных данных была спроектирована гибридная архитектура хранилища данных (DWH). Концепция базируется на жестком разделении транзакционной нагрузки и аналитических вычислений с использованием стека СУБД PostgreSQL и ClickHouse.

Реляционный сегмент в PostgreSQL обеспечивает надежное хранение нормативно-справочной информации (справочники ведомств, услуг, типов ошибок) и поддерживает ссылочную целостность [7]. Аналитический слой развернут в колоночной СУБД ClickHouse, где центральная таблица фактов использует специализированный движок MergeTree [5]. Секционирование данных настроено по месяцам, а разреженный индекс обеспечивает высокую скорость агрегации данных при минимальном потреблении оперативной памяти.

Однако прямая адресация тяжелых аналитических запросов со стороны BI-интерфейсов к сырой таблице фактов способна вызвать деградацию

аппаратных ресурсов (приводя к полному сканированию таблиц — Full Table Scan). Для решения этой проблемы был спроектирован выделенный слой агрегированных витрин данных (Data Marts) с использованием механизма материализованных представлений (Materialized Views).

В момент пакетной загрузки данных СУБД осуществляет преагрегацию инкремента «на лету» и фиксирует результаты в таблицах на базе высокопроизводительного движка SummingMergeTree [5].

Были разработаны две ключевые витрины:

- Часовая витрина для мониторинга технологической стабильности СМЭВ и фиксации отклонений времени отклика.
- Ежедневная витрина для аккумуляирования «тепловой карты» рисков нарушения нормативных сроков обработки заявок территориальными ведомствами.

Автоматизация обработки и математическое обогащение данных. Оркестрация процессов извлечения, трансформации и загрузки данных (ETL) реализована на базе фреймворка Apache Airflow [1, 4]. Логика трансформации, написанная на языке Python с применением библиотеки Pandas, гарантирует идемпотентность конвейера [6].

На этапе трансформации сырые массивы обогащаются расчетными метриками в рамках двух алгоритмических блоков:

1) Выявление технических аномалий: Расчет стандартизированного отклонения (Z-индекса) времени отклика шлюзов от скользящего среднего. Превышение порогового значения в 3.0 (по правилу трех сигм) автоматически маркируется как инфраструктурный сбой [2].

2) Предиктивный мониторинг SLA: Вычисление индекса риска R_j для длительных госуслуг (например, выдача паспорта). На основе отношения текущего срока к нормативному, взвешенного на коэффициент исторической эффективности ведомства, заявки распределяются по зонам риска («Зеленая», «Желтая», «Красная»).

Все математические вычисления векторизованы в оперативной памяти worker-узлов Airflow, после чего финализированный датасет отправляется пакетной вставкой (batch insert) в ClickHouse, что минимизирует фрагментацию дискового пространства [3, 5].

Завершающим этапом проектирования системы стала разработка дашбордов в Apache Superset. Архитектурное решение связи заключалось в делегировании вычислений: BI-система транслирует действия пользователя в SQL-запросы, обрабатываемые исключительно на уровне материализованных витрин, без выгрузки сырых массивов в оперативную память сервера визуализации.

Для ускорения цикла разработки аналитических панелей на этапе UI/UX-проектирования была применена методология ИИ-прототипирования. Использование генеративных языковых моделей (LLM) позволило оперативно декомпозировать бизнес-требования в иерархию компонентов дашборда и подобрать оптимальные цветовые паттерны. Это значительно сократило время подготовки макетов интерфейсов.

На базе предрассчитанных витрин были реализованы две интерактивные панели:

- 1) Оперативная панель мониторинга технических аномалий. Сфокусирована на выявлении деградации производительности шлюзов СМЭВ. Центральным элементом выступает двухосевой график временных рядов (Time-series chart), сопоставляющий среднее время отклика с расчетным Z-индексом (Рисунок 1). Дополнительно интегрирована тепловая карта (Heatmap), локализуемая узлы концентрации технических сбоев в матрице «Сервис – Время».

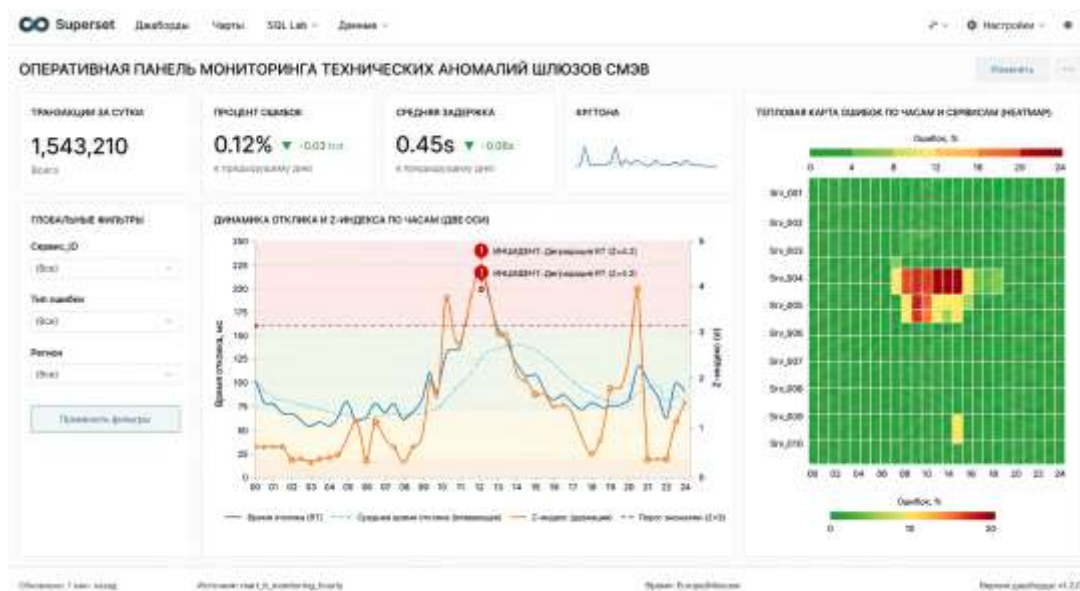


Рисунок 1. Оперативная панель мониторинга технических аномалий шлюзов СМЭВ

2) Аналитическая панель «SLA и зоны риска». Ориентирована на функциональное руководство. Основана на принципе рейтингования ведомств по предиктивному индексу R_j (Рисунок 2). Горизонтальная столбчатая диаграмма визуализирует распределение заявок по зонам риска, а система кросс-фильтров позволяет декомпозировать статистику от уровня макрорегиона до конкретного подразделения МФЦ.

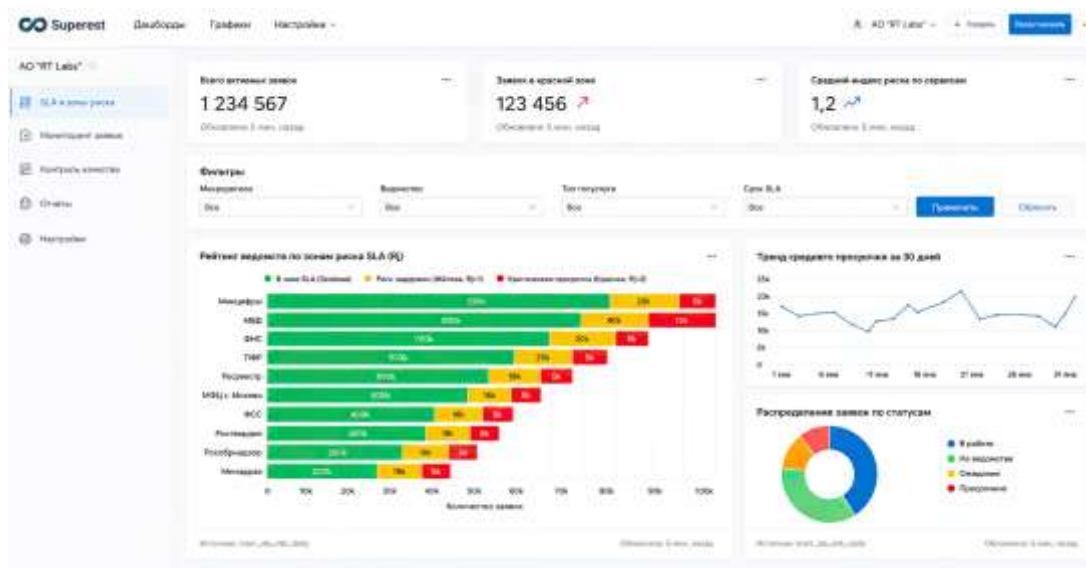


Рисунок 2. Интерфейс аналитической панели «SLA и зоны риска»

Оценка эффективности внедрения. Эмпирическая валидация спроектированной BI-системы, проведенная на основе репрезентативного массива данных, подтвердила высокую технологическую и управленческую эффективность предложенной архитектуры:

– Переход к колоночному сжатию ClickHouse позволил сократить объем хранения сырых данных в 11 раз (с 500 ГБ до 45 ГБ), что снижает капитальные затраты на аппаратное обеспечение.

– Использование слоя Data Marts ускорило рендеринг дашбордов и время отклика аналитических запросов с 120–150 секунд до эталонных значений менее 0,5 секунды.

– Интеграция математического аппарата выявления аномалий непосредственно в ETL-конвейер позволила снизить метрику MTTD (Mean Time To Detect) с 24 часов до 5 минут, обеспечив переход функционального мониторинга от реактивной модели к проактивной.

Заключение. Предложенная методология объединяет возможности гибридных хранилищ данных с современными инструментами визуализации и UI/UX-проектирования. Изоляция операционной нагрузки через агрегированные витрины данных и применение алгоритмов предиктивной оценки рисков (индекс R_j) позволяют создать масштабируемую аналитическую систему. Практическое применение данного комплекса обеспечивает прозрачность межведомственного электронного взаимодействия и предоставляет надежный инструментарий для непрерывного повышения качества государственных услуг.

Список литературы

1. Intorf, D. Apache Airflow Best Practices / D. Intorf, D. Storey, K. van Doorn. – Birmingham : Packt Publishing, 2024. – 188 p. – ISBN 978-1-80512-375-0.
2. Montgomery, D. C. Introduction to Statistical Quality Control / D. C. Montgomery. – 8th ed. – Hoboken : John Wiley & Sons, 2019. – 776 p.

3. Ruiter, J. de. Data Pipelines with Apache Airflow. 2nd Edition. Orchestration for Data and AI / J. de Ruiter, I. Cabral, K. Geusebroek, D. van der Ende, B. Harensiak. – Shelter Island : Manning Publications, 2026. – 600 p. – ISBN 978-1-633436-37-4.
4. Apache Airflow Documentation [Электронный ресурс] // Apache Software Foundation. – URL: <https://airflow.apache.org/docs/> (дата обращения: 22.05.2026).
5. ClickHouse Documentation [Электронный ресурс] // ClickHouse Inc. – URL: <https://clickhouse.com/docs/> (дата обращения: 14.05.2026).
6. pandas Documentation [Электронный ресурс] // The pandas Development Team. – URL: <https://pandas.pydata.org/docs/> (дата обращения: 20.05.2026).
7. PostgreSQL Documentation [Электронный ресурс] // The PostgreSQL Global Development Group. – URL: <https://www.postgresql.org/docs/> (дата обращения: 17.05.2026).

References

1. Intorf, D. Apache Airflow Best Practices / D. Intorf, D. Storey, K. van Doorn. – Birmingham : Packt Publishing, 2024. – 188 p. – ISBN 978-1-80512-375-0.
2. Montgomery D. C. Introduction to Statistical Quality Control. 8th ed. Hoboken: John Wiley & Sons, 2019. 776 p.
3. de Ruiter J., Cabral I., Geusebroek K., van der Ende D., Harensiak B. Data Pipelines with Apache Airflow. 2nd Edition: Orchestration for Data and AI. Shelter Island, NY: Manning Publications, 2026. ISBN 978-1-633436-37-4.
4. Apache Software Foundation. Apache Airflow Documentation. Available at: <https://airflow.apache.org/docs/> (accessed May 22, 2026).
5. ClickHouse Inc. ClickHouse Documentation. Available at: <https://clickhouse.com/docs/> (accessed May 14, 2026).
6. The pandas Development Team. pandas Documentation. Available at: <https://pandas.pydata.org/docs/> (accessed May 20, 2026).

7. The PostgreSQL Global Development Group. PostgreSQL Documentation.
Available at: <https://www.postgresql.org/docs/> (accessed May 17, 2026).