

УДК 621.391

**ПРОЕКТ ЕДИНОЙ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ (ЕАП) ДЛЯ
ЦИФРОВОЙ ТРАНСФОРМАЦИИ ИНФОКОММУНИКАЦИОННОЙ
КОМПАНИИ: ПЕРЕХОД К УПРАВЛЕНИЮ, ОСНОВАННОМУ НА ДАННЫХ,
В РУП «БЕЛТЕЛЕКОМ»**

Д.М. ГОРБУКОВ

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь**Поступила в редакцию 17 апреля 2026*

Аннотация. В статье обоснована необходимость замены ручной табличной обработки данных на сквозную аналитическую платформу для РУП «Белтелеком». Описаны текущие проблемы: отсутствие единой витрины данных, высокая трудоёмкость отчётности, невозможность глубокого анализа, риски несогласованности данных. Предложена архитектура решения на базе Open Source компонентов (Kafka, ClickHouse, Airflow, Superset, Kubernetes, OpenMetadata, Keycloak). Показаны ожидаемые выгоды: сокращение времени подготовки отчётов до 80%, снижение ошибок на 90%, ускорение принятия решений, возможность прогнозной аналитики оттока и LTV.

Ключевые слова: цифровая трансформация, управление данными, телеком-аналитика, data platform, Open Source, ClickHouse, Apache Airflow, бизнес-интеллект, качество данных.

Введение

Ключевая проблема: система управления и анализа данных, построенная на ручной обработке с использованием табличных методов, не соответствует масштабам и динамике бизнеса компании. Это создает критические риски для операционного управления, финансовых результатов и конкурентной позиции РУП «Белтелеком».

Невозможность качественного drill-down-анализа: детализация (например, динамика ARPU с ежемесячной детализацией по каждой услуге и с детализацией по филиалам и ЗУЭС, УЭС) в рамках ручных методов приводит к созданию чрезвычайно сложных и хрупких моделей. Любая попытка «копнуть глубже» требует создания нового ручного отчета.

Слепые зоны в клиентском опыте. При абонентской базе в 2,5 млн пользователей интернета byfly и 2,53 млн зрителей TV ZALA, ручной анализ оттока и причин недовольства неэффективен. Это ведет к потере лояльных клиентов. Что особенно актуально в жесткой конкурентной среде, в которой работает компания.

Ручной труд и высокая трудоемкость. Большинство отчетов предположительно для управления коммерческой деятельностью и руководства формируются вручную с использованием формул (ВПР, расчетные столбцы и др. – распространённая методика подготовки таких отчетов), что требует значительных временных затрат и подвержено ошибкам.

Разрозненность данных. Данные извлекаются из разрозненных источников, что усложняет консолидацию и верификацию информации.

Низкая оперативность. Ежедневные отчеты готовятся с задержками из-за необходимости ручного сбора данных.

Сложность адаптации к изменениям. При изменении структуры филиалов (например, объединение финрегионов в Гомельском филиале) или критериев отчетности (например, новые тарифные планы) требуются ручные доработки формул и макросов. Постоянные доработки отчетов и ручные корректировки требуют вовлечения сотрудников.

Риски несогласованности данных. Отчеты из разных источников могут содержать противоречивые данные из-за различий в методологиях расчета.

Риски упустить выгоду от кросс-продаж. «Белтелеком» предлагает десятки услуг и пакетов («ЯСНА», «Бизнес-кейс», byfly, ZALA, «Видеоконтроль»), но без единой витрины данных сложно выявлять абонентов, готовых к апгрейду тарифа или подключению дополнительных услуг (например, к пакету «Видеоконтроль»).

Вопросы оперативного управления скидками при ценообразовании: компания активно использует региональные и стимулирующие скидки (до 60% на 18 месяцев). Без оперативного анализа их влияния на LTV (Lifetime Value) и отток клиентов невозможно оценить их истинную рентабельность и оптимизировать структуру.

В настоящей статье предложена архитектура сквозной аналитической платформы на Open Source стеке, позволяющая перейти от реактивного управления к проактивному, основанному на данных.

Текущее состояние, потенциал

«Белтелеком» обладает огромным массивом данных, но испытывает дефицит управляемой аналитики.

Масштаб данных: миллионы абонентов. Белтелеком – крупный игрок с 2,5 млн. интернет-абонентов и 2,53 млн. подписчиков ТВ. При этом у компании сотни тарифных комбинаций и множество сервисов (Evika, «Мой город» и др.), сложная тарифная структура с пакетами «ЯСНА», региональными различиями и программами лояльности.

Технические системы по предоставлению услуг (оборудование сети передачи данных, коммуникационные системы, сервисные платформы, билингвая система и др.) генерируют огромный объем операционной информации.

Ограниченная аналитика. Отсутствуют данные об внутренних системах аналитики. При этом в каждом сегменте услуг Белтелекома или региона страны присутствуют активные конкуренты. Принятие решений часто основано на агрегированных или устаревших отчетах, что не позволяет проводить глубокий анализ в разрезе отдельных сегментов.

Подтвержденная потребность. Успешный запуск платформы «Мой город» и активное развитие экосистемы услуг (от зарядных станций до видеонаблюдения) свидетельствует о том, что компания готова к внедрению сложных IT-решений для улучшения клиентского сервиса и операционной эффективности.

Проблемы:

- отсутствие единой платформы для управления данными;
- отсутствие автоматизированной верификации данных.

Заключение: текущая модель ручной обработки данных не является устойчивой. Она блокирует рост, маскирует реальные проблемы и препятствует переходу к управлению, основанному на данных, что ставит под угрозу конкурентные преимущества компании.

Предложения по решению и ключевые технологии

Для автоматизации отчетности предлагается внедрить сквозную платформу анализа данных, охватывающую весь жизненный цикл – от интеграции источников до визуализации.

Интеграция и обработка данных. Автоматизировать сбор данных из всех источников.

ETL/ELT-процессы для автоматизированного сбора данных из всех источников в реальном времени и их очисткой. Например, оборудование сети передачи данных, коммуникационные системы, сервисные платформы, билингвая система, системы учета абонентов и услуг, системы расчетов, системы технического учета и паспортизации, бухгалтерской системы и др. Реализовать ETL/ELT-процессы предлагается на базе Apache Airflow [7, 8, 9] для оркестрации загрузки данных из всех источников.

Apache Kafka и Apache Spark Streaming для потоковой обработки [1, 2, 3]:

- новые подключения услуг в реальном времени;
- оперативные данные по продажам;
- мгновенное обновление дашбордов.

Data Lake для консолидации сырых данных из всех систем. Единое хранилище для сырых данных с возможностью очистки и стандартизации. Пример: данные о населении регионов, блокировках абонентов, доходах.

Data Lake (например, на базе MinIO или HDFS) для консолидации сырых данных из всех систем.

Great Expectations [16, 17] и dbt [21] для автоматических проверок качества данных (полнота, уникальность, отсутствие аномалий).

Выгода:

- исключение ручного переноса данных между системами;
- сокращение времени на подготовку отчетов на 70–80%;
- автоматическое обнаружение проблем с данными.

Хранение и обработка. Гибридное хранилище данных: создать единое надежное хранилище для исторических и оперативных данных.

Использовать аналитическое Data Warehouse [4, 5, 6] как основной слой для агрегатов и витрин, а ClickHouse, как колоночная СУБД, оптимизированная для телеком-аналитики, имеющая высокую скорость агрегаций и низкую стоимость хранения.

Реляционное хранилище для метаданных, справочников и управленческой отчетности построить на PostgreSQL и Patroni:

- стандартизированные бизнес-метрики;
- автоматический пересчет градаций районов;
- единая методология расчетов (например, ARPU, приросты и т.д.).

Kubernetes [13, 14] для оркестрации контейнеров (Docker) с компонентами платформы, что обеспечит масштабируемость и отказоустойчивость.

Выгода:

- накопление единой версии данных для всех филиалов;
- мгновенный доступ к истории изменений (например, миграция регионов между группами);
- избежать эффекта «хрустального сервера»;
- высокая производительность даже при многолетней истории и миллионах абонентов.

Управление метаданными и качеством данных.

Data Catalog: OpenMetadata [15] или DataHub – для поиска данных, управления метаданными, lineage (происхождения данных).

Data Quality: Great Expectations [16, 17] (автоматические проверки), dbt tests [21], мониторинг аномалий.

Data Observability: метрики здоровья пайплайнов, оповещения о сбоях.

Выгода:

- прозрачность происхождения данных;
- снижение времени на поиск нужных данных;
- доверие к данным со стороны бизнеса.

Визуализация и аналитика. Обеспечить руководство интуитивными отчетами и дашбордами для ежедневного контроля.

Внедрение платформы Apache Superset [10, 11, 12] для создания интерактивных дашбордов и отчетов:

- Дашборды для тактического управления (управления коммерческой деятельностью): использование интерактивных фильтров и возможностей drill-down для самостоятельного анализа динамики продаж, выполнения планов и прогнозных моделей в реальном времени.
- Стратегические отчеты (руководство компании): создание единой панели KPI с использованием богатой библиотеки визуализаций Superset для отображения трендов и паттернов.

Настроить автоматические уведомления о рисках (снижение ARPU, рост блокировок).

Настроить API для интеграции с другими системами компании.

Выгода:

- сокращение времени формирования отчета с часов до минут;
- возможность drill-down до уровня филиала; точность данных за счет исключения ручных ошибок.

Безопасность.

Keycloak [18, 19] или Apache Ranger для централизованной аутентификации и авторизации.

Шифрование данных в покое (Data Lake, DWH) и при передаче (TLS).

Аудит доступа к данным и отчётам.

Маскирование чувствительных данных (персональные данные абонентов).

Интеграция с корпоративным LDAP/Active Directory.

Выгода:

- соответствие законодательству РФ о персональных данных;
- защита от утечек;
- прозрачность доступа.

Оркестрация и DevOps.

Apache Airflow [7, 8, 9] – оркестрация ETL/ELT-пайплайнов.

GitLab CI/CD или Jenkins для автоматического развертывания.

Terraform или Ansible (Infrastructure as Code).

Prometheus и Grafana для мониторинга платформы.

Выгода:

- воспроизводимость инфраструктуры;
- быстрое развёртывание новых окружений;
- высокая надёжность.

Прогнозная аналитика и Reverse ETL.

ML-платформа (например, на базе JupyterHub и MLflow) для:

- прогноза оттока (churn prediction),
- прогноза ARPU и LTV,
- рекомендательных систем для кросс-продаж.

Reverse ETL (инструменты типа Hightouch, Census или самописные конвейеры) для выгрузки сегментов и инсайтов обратно в CRM и биллинг.

Выгода:

- проактивное удержание клиентов;
- автоматические персонализированные предложения;
- рост LTV – пожизненной ценности клиента, упрощённая формула для телекома:

$LTV = (\text{Средний платёж в месяц} \times \text{Средняя продолжительность жизни клиента в месяцах}) - \text{Расходы на привлечение и обслуживание.}$

Ожидаемые выгоды от реализации проекта

Сокращение времени подготовки отчетов:

- До 80% за счет автоматизации (сейчас до 4-6 часов в день на ручные операции).
- Формирование отчетов с 2 дней до 15 минут. Доступ к данным в реальном времени.
- Сокращение цикла планирования на 40%.
- Мгновенное добавление новых услуг (1-2 дня вместо 2-3 недель).

Повышение точности данных:

- Единая версия данных для всех отчетов.
- Исключение ручных ошибок через единые алгоритмы расчета (например: ARPU по всем видам услуг; проникновение услуг с учетом специфики регионов таких как количество и плотность населения, доходы населения, наличия конкурентов в разных сегментах услуг и др.).
- Снижение ошибок данных на 90%.

Ускорение принятия решений:

- Формирование отчета в реальном времени (вместо ежедневных усилий).
- Ускорение принятия решений на 50%.
- Доступ к актуальным данным в реальном времени для анализа эффективности филиалов.

Операционная эффективность:

- Масштабируемость для бизнес-роста.
- Возможность быстрого добавления новых метрик (например, метрики для услуги «Окко» или «Литрес»).
- Возможность прогнозной аналитики.

- Повышение прозрачности процессов на 90%.
- Ускорение принятия стратегических решений.
- Снижение операционных рисков:
- Централизованный контроль данных и автоматическое логирование изменений.

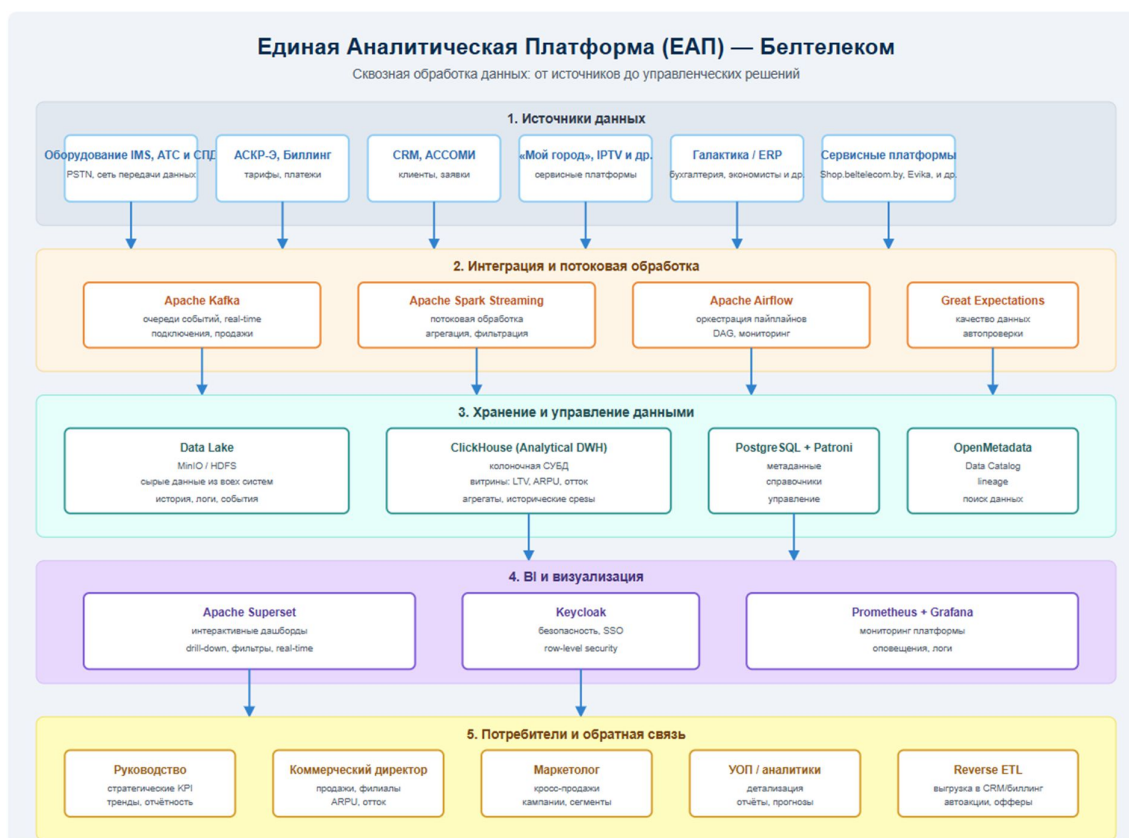


Рис. 1. Единая Аналитическая Платформа – Open Source стек для цифровой трансформации РУП «Белтелеком»

Заключение

Внедрение предложенной платформы позволит РУП «Белтелеком» совершить переход от фрагментированной ручной отчётности к единой управляемой экосистеме данных. Ключевые измеримые результаты:

- сокращение времени подготовки отчётов с 2 дней до 15 минут (*экономия до 80% трудозатрат*);
- снижение ошибок данных на 90% *за счёт автоматической верификации*;
- рост LTV и снижение оттока благодаря прогнозной аналитике и персонализированным предложениям.

Платформа построена на современном Open Source стеке, что обеспечивает низкую совокупную стоимость владения, масштабируемость и независимость от вендоров. Пилотный проект минимизирует риски, а поэтапное масштабирование позволяет контролировать затраты на каждом шаге.

Реализация данного решения создаст основу для цифровой трансформации управления данными и перехода к data-driven культуре, что напрямую повысит конкурентоспособность компании на рынке телекоммуникационных услуг.

UNIFIED ANALYTICS PLATFORM (UAP) PROJECT FOR THE DIGITAL TRANSFORMATION OF AN INFOCOMMUNICATION COMPANY: TRANSITION TO DATA-DRIVEN MANAGEMENT AT RUP «BELTELECOM»

D.M. HARBUKOU

Abstract. The document substantiates the need to replace manual spreadsheet-based data processing with an end-to-end analytics platform for RUP «Beltelecom». Current problems are described: lack of a single data mart, high reporting labor intensity, inability to perform deep analysis, and risks of data inconsistency. A solution architecture based on Open Source components (Kafka, ClickHouse, Airflow, Superset, Kubernetes, OpenMetadata, Keycloak) is proposed. Expected benefits include up to 80% reduction in reporting time, 90% reduction in data errors, faster decision-making, and predictive analytics for churn and LTV.

Keywords: digital transformation, data management, telecom analytics, data platform, Open Source, ClickHouse, Apache Airflow, business intelligence, data quality.

Список литературы

1. Apache Kafka [Электронный ресурс] // Официальный сайт Apache Kafka. – Режим доступа: <https://kafka.apache.org/>. – Дата доступа: 17.04.2026.
2. Kafka in Telecom: реальные кейсы использования [Электронный ресурс] // Confluent. – Режим доступа: <https://www.confluent.io/resources/customer-stories/telecom/>. – Дата доступа: 17.04.2026.
3. Архитектура Apache Kafka (диаграммы) [Электронный ресурс] // Apache Kafka Documentation. – Режим доступа: <https://kafka.apache.org/24/images/kafka-apis.png>. – Дата доступа: 17.04.2026.
4. ClickHouse [Электронный ресурс] // Официальный сайт ClickHouse. – Режим доступа: <https://clickhouse.com/>. – Дата доступа: 17.04.2026.
5. ClickHouse для телеком-решений: примеры использования [Электронный ресурс] // ClickHouse. – Режим доступа: <https://clickhouse.com/use-cases/telecom>. – Дата доступа: 17.04.2026.
6. Архитектура ClickHouse: колоночное хранение, партиционирование, сжатие [Электронный ресурс] // ClickHouse Documentation. – Режим доступа: <https://clickhouse.com/docs/en/development/architecture>. – Дата доступа: 17.04.2026.
7. Apache Airflow [Электронный ресурс] // Официальный сайт Apache Airflow. – Режим доступа: <https://airflow.apache.org/>. – Дата доступа: 17.04.2026.
8. Архитектура Apache Airflow (схема компонентов) [Электронный ресурс] // Apache Airflow Documentation. – Режим доступа: https://airflow.apache.org/docs/apache-airflow/stable/_images/airflow_architecture.png. – Дата доступа: 17.04.2026.
9. Astronomer. Data Engineering with Airflow: практические руководства [Электронный ресурс] // Astronomer. – Режим доступа: <https://www.astronomer.io/guides/>. – Дата доступа: 17.04.2026.
10. Apache Superset [Электронный ресурс] // Официальный сайт Apache Superset. – Режим доступа: <https://superset.apache.org/>. – Дата доступа: 17.04.2026.
11. Галерея дашбордов Apache Superset [Электронный ресурс] // Apache Superset Gallery. – Режим доступа: <https://superset.apache.org/gallery/>. – Дата доступа: 17.04.2026.
12. Архитектура Apache Superset [Электронный ресурс] // Apache Superset Documentation. – Режим доступа: <https://superset.apache.org/docs/contributing/architecture>. – Дата доступа: 17.04.2026.
13. Kubernetes. Компоненты кластера [Электронный ресурс] // Kubernetes Documentation. – Режим доступа: <https://kubernetes.io/docs/concepts/overview/components/>. – Дата доступа: 17.04.2026.
14. Kubernetes в телеком-индустрии: примеры внедрения [Электронный ресурс] // Cloud Native Computing Foundation (CNCF). – Режим доступа: <https://www.cncf.io/case-studies/telecom/>. – Дата доступа: 17.04.2026.
15. OpenMetadata. Архитектура платформы [Электронный ресурс] // OpenMetadata. – Режим доступа: <https://open-metadata.org/architecture>. – Дата доступа: 17.04.2026.
16. Great Expectations [Электронный ресурс] // Официальный сайт Great Expectations. – Режим доступа: <https://greatexpectations.io/>. – Дата доступа: 17.04.2026.
17. Great Expectations Gallery: примеры отчётов о качестве данных [Электронный ресурс] // Great Expectations. – Режим доступа: <https://greatexpectations.io/gallery>. – Дата доступа: 17.04.2026.
18. Keycloak. Архитектура и основные концепции [Электронный ресурс] // Keycloak Documentation. – Режим доступа: https://www.keycloak.org/docs/latest/server_admin/#core-concepts-and-terms. – Дата доступа: 17.04.2026.

19. Keycloak. Интеграция с BI-инструментами [Электронный ресурс] // Keycloak Guides. – Режим доступа: <https://www.keycloak.org/guides>. – Дата доступа: 17.04.2026.
20. AWS for Telecom. Референтная архитектура платформы данных [Электронный ресурс] // Amazon Web Services. – Режим доступа: <https://aws.amazon.com/telecom/data-platform/>. – Дата доступа: 17.04.2026.
21. dbt Labs. Современный стек данных в картинках [Электронный ресурс] // dbt Blog. – Режим доступа: <https://www.getdbt.com/blog/modern-data-stack-in-pictures/>. – Дата доступа: 17.04.2026.
22. Databricks. Архитектура Data Lakehouse для телекома [Электронный ресурс] // Databricks. – Режим доступа: <https://www.databricks.com/solutions/industries/telecommunications>. – Дата доступа: 17.04.2026.