

МАТЕМАТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ И МОДЕЛИ В ЭКОНОМИКЕ

DOI: 10.6060/ivecofin.2026682.784

УДК: 33:004.65

ИССЛЕДОВАНИЕ И АНАЛИЗ СОВРЕМЕННЫХ МЕТОДИК ПОСТРОЕНИЯ АРХИТЕКТУР ДАННЫХ

С.С. Гудзь, О.А. Цуканова

Савелий Сергеевич Гудзь (ORCID 0009-0000-0252-9264)

Санкт-Петербургский электротехнический университет «ЛЭТИ», ул. Профессора Попова, 5, Санкт-Петербург, 197022, Россия

E-mail: sawa.gudz@gmail.com

Ольга Александровна Цуканова (ORCID 0000-0002-9442-2268)

Санкт-Петербургский государственный университет, Университетская набережная, д. 7-9, Санкт-Петербург, 199034, Россия

E-mail: zoa1999@mail.ru

В статье рассматриваются современные подходы к построению архитектур данных для задач оценки и рационализации ИТ-инфраструктуры предприятия в условиях роста объема, скорости обновления и разнообразия источников информации. Актуальность исследования обусловлена усложнением ИТ-ландшафта организаций, необходимостью интеграции технических, операционных и финансовых данных, а также повышением требований к обоснованности управленческих решений в сфере ИТ. Целью исследования является уточнение и апробация методики построения аналитической платформы, обеспечивающей консолидацию разрозненных данных, сохранение полной истории изменений и поддержку расчёта ключевых показателей эффективности инфраструктуры. В работе проведён сравнительный анализ архитектурных подходов Inmon, Kimball и Data Vault 2.0 с точки зрения параметров гибкости, масштабируемости, трудоёмкости внедрения и способности поддерживать историчность данных. По результатам анализа обоснован выбор методологии Data Vault 2.0 как наиболее адаптивной модели для динамично изменяющейся ИТ-среды. Разработанная методика реализована с использованием технологического стека Apache Airflow, dbt и PostgreSQL в рамках парадигмы ETL/ELT, предусматривающей загрузку данных с последующей трансформацией внутри целевой СУБД. Апробация выполнена на примере создания витрины данных для автоматизации отчётности ИТ-интегратора на основе CRM Bitrix24. В результате внедрения достигнуто сокращение времени подготовки аналитической отчётности, повышение производительности выполнения запросов и выявление резервов оптимизации инфраструктурных ресурсов. Расчётный экономический эффект пилотного сегмента свидетельствует об актуальности внедрения и использования предлагаемой методики на типичном предприятии, и предполагается, что целесообразно ее масштабировать на другие направления управления ИТ-инфраструктурой.

Ключевые слова: Архитектура данных, Data Vault 2.0, ETL/ELT, ИТ-инфраструктура, рационализация затрат, аналитическая платформа, Apache Airflow, dbt, витрина данных.

RESEARCH AND ANALYSIS OF MODERN DATA ARCHITECTURE DESIGN METHODOLOGIES

S.S. Gudz, O.A. Tsukanova

Saveliy S. Gudz (ORCID 0009-0000-0252-9264)

Saint Petersburg Electrotechnical University "LETI", Professor Popov St., 5, Saint Petersburg, 197022, Russia

E-mail: sawa.gudz@gmail.com

Olga A. Tsukanova (ORCID 0000-0002-9442-2268)

Saint Petersburg State University, Universitetskaya Emb., 7–9, Saint Petersburg, 199034, Russia

E-mail: zoa1999@mail.ru

The article examines modern approaches to data architecture design for the assessment and rationalization of enterprise IT infrastructure under conditions of increasing data volume, velocity, and heterogeneity. The relevance of the study is due to the growing complexity of organizational IT environments, the need to integrate technical, operational, and financial data, and increasing requirements for evidence-based IT management decision-making. The purpose of the study is to develop and validate a methodology for building an analytical platform that ensures integration of heterogeneous data sources, preservation of full historical data, and support for calculating key IT infrastructure performance indicators. A comparative analysis of the Inmon, Kimball, and Data Vault 2.0 architectural approaches has been conducted in regards to flexibility, scalability, implementation complexity, and ability to support historical data tracking. Based on the results of the analysis, Data Vault 2.0 has been selected as the most suitable architectural model for dynamically changing IT environments due to its adaptability and structural extensibility. The proposed methodology has been implemented using a modern data engineering technology stack, including Apache Airflow for workflow orchestration, dbt (Data Build Tool) for data transformation, and PostgreSQL as the target database management system. The implementation follows the ELT paradigm, where data transformation is performed within the target database environment. The methodology is validated through the development of a data mart designed to automate analytical reporting for an IT integrator using data extracted from the Bitrix24 CRM system. The implementation results demonstrate a significant reduction in analytical report preparation time, improved query performance, and identification of IT infrastructure optimization opportunities. The calculated economic effect of the pilot segment demonstrates the relevance of implementing and using the proposed methodology at a typical enterprise, and it is assumed that it can be scaled to other areas of IT infrastructure management.

Keywords: Data architecture, Data Vault 2.0, ETL/ELT, IT infrastructure, cost rationalization, analytical platform, Apache Airflow, dbt, data mart.

Для цитирования:

Гудзь С.С., Цуканова О.А. Исследование и анализ современных методик построения архитектур данных. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2026. № 02(68). С. 99-109. DOI: 10.6060/ivecofin.2026682.784

For citation:

Gudz S.S., Tsukanova O.A. Research and analysis of modern data architecture design methodologies. *Ivecofin*. 2026. № 02(68). P. 99-109. DOI: 10.6060/ivecofin.2026682.784 (in Russian)

ВВЕДЕНИЕ

Современная ИТ-инфраструктура предприятия представляет собой совокупность взаимосвязанных аппаратных, программных и сервисных компонентов, функционирующих в условиях постоянных изменений. Рост числа информационных систем, использование облачных сервисов, расширение интеграционных контуров и увеличение объема телеметрических данных усложняют задачи мониторинга, анализа и управления. Рационализация ИТ-инфраструктуры – процесс, направленный на оптимизацию затрат, повышение производительности и надежности – является не разовой акцией, а непрерывным циклом управления, основанным на данных (data-driven) [1].

Ключевое затруднение связано с фрагментированностью информационного пространства. Сведения, необходимые для оценки состояния инфраструктуры, распределены по различным системам, отличаются по структуре, частоте обновления и смысловой интерпретации. Отсутствие согласованной архитектурного подхода приводит к высокой трудоемкости расчета показателей и снижению доверия к аналитическим результатам [2]. В

перспективе преодоление семантических разрывов между источниками может быть дополнено технологиями извлечения знаний на основе больших языковых моделей (LLM) и retrieval-augmented generation (RAG), что открывает возможности для построения экспертных систем предприятия [3].

Информационная база формируется за счет нескольких классов источников. Метрики функционирования инфраструктуры поступают из систем мониторинга, включая Zabbix, Prometheus и SolarWinds, где фиксируются показатели загрузки вычислительных ресурсов, операций ввода-вывода и доступности сервисов. Сведения о составе инфраструктуры и взаимосвязях ее элементов извлекаются из систем управления конфигурациями и сервис-десков, таких как ServiceNow и Atlassian Jira, где также аккумулируются данные об инцидентах и изменениях. Финансовый контур дополняет картину за счет учета затрат на оборудование, лицензии и облачные сервисы, тогда как журналы приложений и системного ПО позволяют анализировать поведение пользователей, ошибки и время отклика. Прямой анализ таких «сырых»

данных затруднен вследствие различий в форматах, периодичности обновления и семантической интерпретации, что приводит к усложнению и удлинению процесса расчета ключевых показателей эффективности (KPI), а также повышает вероятность возникновения ошибок [4].

Для объективной оценки состояния инфраструктуры необходимо рассчитать ряд ключевых показателей эффективности, которые включают в себя:

- доступность сервисов, представляющее отношение разницы между общим временем и временем простоя к общему времени, выраженное в процентах;
- среднее время восстановления (MTTR), рассчитываемое как суммарное время простоя по инцидентам, деленное на их количество;
- среднее время между отказами (MTBF), определяемое как суммарное время работы, деленное на число отказов;
- уровень утилизации ресурсов, представляющий собой отношение средней используемой мощности к общей мощности в процентах;
- показатель ИТ-затрат на единицу бизнес-результата (Cost per Unit).

В случае отклонения фактических значений от нормативных возникает необходимость в использовании специализированной аналитической платформы, обеспечивающей консолидацию данных из разнородных источников, их согласование и очистку, эффективное хранение с возможностью отслеживания исторических изменений, а также быстрый и удобный доступ к данным для последующего расчета метрик и формирования отчетности. Расчет данных метрик требует интеграции технических и финансовых данных, их очистки, согласования и историзации. Это обуславливает необходимость разработки специализированной аналитической платформы, основанной на устойчивой архитектуре данных [5].

МЕТОДЫ ИССЛЕДОВАНИЯ

Гипотеза исследования: если уточнить и доработать методику Data Vault 2.0 – добавить информацию о происхождении данных, обеспечить устойчивую загрузку при изменяющихся источниках, учитывать технико-экономические показатели при обработке данных – то можно создать аналитическую платформу, в которой данные будут более достоверными и прозрачными для удовлетворения бизнес-потребностей по анализу информации различных аспектов предпринимательской деятельности. При этом повысится надежность и согласованность показателей ИТ-инфраструктуры.

Методы исследования: сравнительный анализ архитектурных подходов к проектированию хранилищ данных; проектирование и реализацию пилотной системы (прототипирование); инструментальные средства оркестрации (Apache

Airflow) и трансформации данных (dbt); методы оценки экономической эффективности (расчет потенциальной экономии на основе анализа утилизации ресурсов и снижения трудозатрат).

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

При построении аналитической платформы для оценки состояния ИТ-инфраструктуры необходимо выбрать соответствующую архитектурную модель хранилища данных [6]. В практике управления данными и ИТ-аналитике доминируют три основных подхода.

Традиционная концепция корпоративного хранилища, предложенная Б. Инмоном, сфокусирована на построении централизованной нормализованной модели (EDW), которая выступает в качестве единого источника достоверной информации. На её основе формируются специализированные витрины данных для аналитических задач. Такой подход создает высокий уровень согласованности и управляемости данных, однако требует значительных временных и организационных затрат на внедрение и слабо адаптируется к быстро меняющимся требованиям.

Методология многомерного моделирования, разработанная Р. Кимбаллом, в рамках которой хранилище строится «снизу вверх» через последовательное создание витрин, ориентированных на конкретные бизнес-процессы, является альтернативным подходом [7]. Использование денормализованных схем, таких как «звезда» или «снежинка», существенно ускоряет выполнение аналитических запросов. Вместе с тем подобный подход нередко приводит к фрагментации данных и усложняет их последующую интеграцию при расширении числа источников.

В условиях высокой изменчивости ИТ-инфраструктуры особое внимание привлекает методология Data Vault 2.0, предложенная Д. Линстедом как развитие идей гибкого и масштабируемого хранилища [8]. В отличие от рассмотренных подходов она ориентирована на разделение ключевых бизнес-сущностей, связей между ними и описательных атрибутов, что изолирует изменения и минимизирует влияние на существующую модель. Такая организация данных облегчает подключение новых источников без переработки всей структуры хранилища, а также обеспечивает полную историзацию за счет фиксации всех изменений атрибутов во времени. Дополнительным преимуществом является возможность параллельной загрузки данных из независимых источников, что особенно важно в распределенной ИТ-среде [9].

Сравнительный анализ архитектурных моделей представлен в табл. 1.

Таблица 1. Сравнительный анализ подходов к проектированию хранилищ данных
Table 1. Comparative analysis of approaches to data warehouse design

Критерий	<i>Inmon (EDW)</i>	<i>Kimball (Dimensional)</i>	<i>Data Vault 2.0</i>
Целевое назначение	Единый источник информации	Быстрая аналитика для бизнес-процессов	Гибкость, масштабируемость, сохранение истории
Структура данных	Высоко нормализованная	Денормализованная (звезда/снежинка)	Гибридная (Хабы, Ссылки, Спутники)
Гибкость к изменениям	Низкая	Средняя	Высокая
Историчность	Ограниченная	Ограниченная (тип 2 SCD)	Полная (на уровне атрибутов)
Сложность внедрения	Высокая	Средняя	Средняя/Высокая (требует дисциплины)
Подходит для	Стабильные, комплексные корпоративные отчеты	Оперативная бизнес-аналитика, дашборды	Интеграция множества источников, изменение требований, аудит

Выбор архитектуры в рамках исследования был обусловлен особенностями предметной области. Анализируемая ИТ-инфраструктура характеризуется высокой динамикой изменений, множественностью источников и необходимостью ретроспективного анализа состояний. В этих условиях применение Data Vault 2.0 представляется обоснованным, поскольку именно этот подход позволяет одновременно обеспечить гибкость развития модели, полноту хранения истории и устойчивость к изменениям структуры входных данных [10, 11]. В отличие от классических моделей, его использование ориентировано не только на поддержку аналитики, но и на обеспечение трассируемости преобразований данных, что имеет принципиальное значение при оценке достоверности расчетных показателей.

Следует отметить, что предлагаемая архитектура не сводится к прямому воспроизведению канонической методологии Data Vault 2.0, описанной в работах Д. Линстеда. В рамках исследования она была адаптирована к специфике задач анализа ИТ-инфраструктуры, где ключевое значение имеют не только интеграция разнородных источников, но и согласование технических, операционных и финансовых данных в едином аналитическом контуре. В отличие от типовых реализаций, ориентированных преимущественно на хранение и историзацию данных, предложенный подход акцентирует внимание на обеспечении их последующей интерпретируемости для расчета комплексных метрик эффективности [12].

Таким образом, новизна подхода заключается в том, что наиболее эффективная интеграция разнородных источников может быть достигнута за счет уточнения структуры бизнес-слоя, интеграции показателей различной природы и организации

сквозной трассируемости преобразований от исходных событий до агрегированных показателей. При этом практическая реализация выходит за рамки демонстрационного прототипа и рассматривается как апробация адаптированной методики, учитывающей ограничения традиционных подходов при работе с динамичной ИТ-инфраструктурой.

Реализация архитектуры Data Vault 2.0 предполагает использование технологического стека, поддерживающего парадигму ELT (Extract, Load, Transform), в рамках которой загрузка данных в целевую систему предшествует их преобразованию, а основные операции трансформации выполняются средствами целевой СУБД [13, 14]. Такой подход позволяет повысить производительность обработки данных, обеспечить централизованное управление логикой преобразований и упростить масштабирование аналитической платформы [15].

Организация процессов загрузки и обработки данных была реализована с использованием платформы Apache Airflow, которая применяется для управления вычислительными workflow и описывает их в виде направленных ациклических графов (DAG) [16]. В разработанной архитектуре Airflow выполняет роль координирующего слоя: он отвечает за планирование задач, контроль их последовательности и мониторинг выполнения. Особое значение имеет возможность формализации зависимостей между этапами обработки, что позволяет обеспечить целостность конвейера данных при работе с множеством источников. Помимо этого, Airflow используется как единая точка запуска трансформационных процедур, реализованных средствами dbt, что повышает воспроизводимость и управляемость всего процесса [17].

Для реализации логики трансформации данных применялся инструмент dbt (Data Build

Tool), ориентированный на выполнение преобразований непосредственно в среде целевой базы данных с использованием SQL и механизма шаблонизации Jinja2. Использование dbt позволило реализовать принцип «трансформация как код», при котором все модели данных, включая таблицы и представления, описываются декларативно и учитываются в системе контроля версий [18]. Встроенные механизмы dbt обеспечивают автоматическое построение графа зависимостей между моделями, что позволяет корректно определять порядок их выполнения, а также реализуют функции тестирования и документирования данных. В рамках разработанной архитектуры dbt использовался для последовательного преобразования данных из слоя Raw Vault, содержащего исходные данные, в слой Business Vault, реализующий бизнес-логику, и далее в аналитические витрины данных, предназначенные для формирования отчетности и выполнения аналитических запросов [19].

В качестве целевой системы управления базами данных была выбрана PostgreSQL, обладающая необходимыми характеристиками надежности, производительности и функциональности для реализации архитектуры Data Vault 2.0. PostgreSQL поддерживает работу с реляционными и еще не полностью структурированными данными, включая формат JSON, и обеспечивает эффективное выполнение аналитических запросов. Применение PostgreSQL в сочетании с описанным стеком инструментов обеспечивает целостность хранения, поддержку историчности и достаточную производительность для решения поставленных задач [20].

В области аудита ИТ-инфраструктуры возникает специфический класс задач, где на первый план выходят прослеживаемость происхождения данных и устойчивость к неполноте информации, а не только скорость построения отчетов или простота модификации схемы. Существующие публикации по внедрению Data Vault 2.0 либо обходят эти аспекты стороной, либо решают их частными техническими подходами, не формулируя общих принципов. Восполнение этого пробела необходимо учесть в предлагаемой уточненной методике.

В традиционной модели Data Vault 2.0 происхождение данных не формализовано на уровне структуры хранилища — предполагается, что сведения о надежности источника и контексте получения информации хранятся в метаданных ETL-процессов или вовсе не фиксируются. Однако для объективной оценки состояния ИТ-инфраструктуры принципиально важно различать данные, поступившие из систем автоматического мониторинга (с высокой частотой и потенциальными шумами), из конфигурационных баз (относительно

статичные, но устаревающие) и из финансовых систем (точные, но агрегированные за периоды). Без явного указания источника любые попытки сопоставить, например, загрузку сервера с затратами на его обслуживание теряют доказательную силу: невозможно определить, вызвано ли расхождение ошибкой измерений или разной периодичностью обновления. Включение атрибута происхождения в состав спутников Data Vault 2.0 превращает хранилище не просто в коллекцию исторических фактов, а в *информационную среду с встроенной системой аудита достоверности*. С научной точки зрения, это позволяет в дальнейшем строить модели оценки надежности агрегированных показателей в зависимости от композиции источников, что выходит за рамки сугубо инженерной задачи построения хранилища.

С технической точки зрения — первым этапом проектирования будет реализация слоя Raw Vault. На данном этапе выполняется идентификация ключевых сущностей предметной области, таких как серверы, сервисы и инциденты, а также определение связей между ними и выделение изменяемых атрибутов. В соответствии с архитектурными принципами Data Vault 2.0 сущности представляются в виде хабов, связи — в виде линков, а атрибутивные характеристики — в виде спутников. Использование данной модели обеспечивает независимость структурных компонентов, возможность расширения модели без модификации существующих структур и сохранение полной истории изменений атрибутов. Это позволяет реализовать ретроспективный анализ состояния инфраструктуры и снизить трудоемкость интеграции новых источников данных.

Вторым шагом будет решение вопроса к организации процессов загрузки данных в условиях *нестационарной доступности источников*. Классические рекомендации по внедрению Data Vault 2.0 исходят из предположения о регулярном и предсказуемом поступлении данных; сбои рассматриваются как исключительные ситуации, обрабатываемые на уровне мониторинга. Между тем, анализ практики работы с инфраструктурными данными показывает, что нестабильность API систем мониторинга и сервис-десков является не исключением, а нормой, обусловленной как техническими ограничениями (лимиты на число запросов, таймауты), так и организационными факторами (плановые остановки сервисов). Поэтому необходимо включить разработку и реализацию автоматизированных конвейеров загрузки данных с использованием системы оркестрации Apache Airflow. В рамках этого технического решения создаются направленные ацикли-

ческие графы выполнения задач, обеспечивающие регулярное извлечение данных из исходных систем, включая системы мониторинга, конфигурационные базы данных и сервис-дески, с последующей загрузкой в слой Raw Vault. Автоматизация процессов загрузки обеспечивает актуальность данных, снижает вероятность ошибок, связанных с ручной обработкой информации, и создает основу для регулярного обновления аналитической информации.

В большинстве работ по хранилищам данных финансовые показатели либо рассматриваются как один из рядовых атрибутов, либо вовсе выносятся в отдельные витрины, проектируемые независимо от технической части. Однако при анализе ИТ-инфраструктуры технические и экономические показатели неразрывно связаны через аспекты по загрузке сервера, где имеет смысл только в сопоставлении с затратами на его содержание, а стоимость простоя сервиса – с вероятностью инцидента. Методика предусматривает, что правила расчёта амортизации, классификации утилизации и оценки эффективности закладываются в слой бизнес-логики не как внешние отчёты, а как *неотъемлемая часть трансформационной модели*. Именно в этой части реализация бизнес-логики преобразования данных и формирования аналитических витрин с использованием инструмента dbt выполняют задачи по очистке, стандартизации и интеграции данных из различных источников, а также применение бизнес-правил, необходимых для расчета аналитических показателей. В результате формируются специализированные структуры данных, оптимизированные для выполнения аналитических запросов и формирования отчетности. Использование подхода декларативного описания трансформаций обеспечивает прозрачность логики обработки данных, воспроизводимость результатов и возможность централизованного управления преобразованиями.

Заключительным этапом уточненной методики является верификация и анализ полученных результатов, включающие проведение тестирования качества данных, расчет ключевых показателей эффективности и оценку состояния инфраструктурных ресурсов. На основе полученных аналитических данных формируются выводы о текущем состоянии ИТ-инфраструктуры, выявляются неэффективно используемые ресурсы и определяются направления оптимизации. Это создает информационную основу для принятия управленческих решений, направленных на повышение эффективности использования инфраструктуры и снижение эксплуатационных затрат [21].

Для апробации методики был реализован проект по созданию витрины данных для одного

из заказчиков ИТ-интегратора. Целью было автоматизировать формирование отчёта об актуальной клиентской базе для проектного офиса. В рамках реализации вышеизложенных аспектов методологии была детально проработана архитектура витрины. Общая логика потока данных для данного кейса представлена на рис. 1.

Ключевые компоненты:

- 1) Источники: данные извлекались через стандартное и кастомное API Bitrix24 из сущностей CRM и высокоуровневых справочников [22].
- 2) Сырой слой (Raw Vault): данные загружались в исторические таблицы (*_history) с сохранением всех технических метаданных (дата загрузки, источник), что соответствует принципам Data Vault 2.0.
- 3) Слой трансформации: с помощью dbt были созданы промежуточные модели (stg_*) для очистки и базовой стандартизации, а затем бизнес-модели (mart_*), в которые были интегрированы справочники и применена бизнес-логика.
- 4) Конечная витрина: Результирующее представление агрегирует данные из трех основных источников в единую денормализованную структуру, готовую для анализа.

Для проектируемой витрины был разработан и настроен DAG (Directed Acyclic Graph) в Apache Airflow, который выступал в роли центрального оркестратора всего конвейера обработки данных. Его основная задача заключалась в координации и автоматическом выполнении последовательности операций. Ежедневно, по расписанию, DAG инициировал процесс инкрементальной загрузки данных, обращаясь к API системы Bitrix24. Загрузка выполнялась с использованием параметра LAST_ACTIVITY_TIME, что позволяло получать только измененные с момента последней синхронизации записи из ключевых сущностей. Полученные сырые данные с сохранением полной истории и метаданных загружались в сырой слой Raw Vault целевой базы данных PostgreSQL.

Сразу после успешной загрузки данных Airflow автоматически запускал выполнение проекта dbt (Data Build Tool), который отвечал за весь цикл трансформации. Структура dbt-проекта была четко разделена по стадиям обработки. На начальном этапе, в рамках моделей staging, происходила базовая очистка и стандартизация данных: выполнялся маппинг исходных технических названий полей на понятные бизнес-атрибуты (в соответствии с таблицами сырого слоя осуществлялось сопоставление с новыми полями для будущей витрины данных), а также приведение типов данных, например, преобразование строк в формат даты или целых чисел.

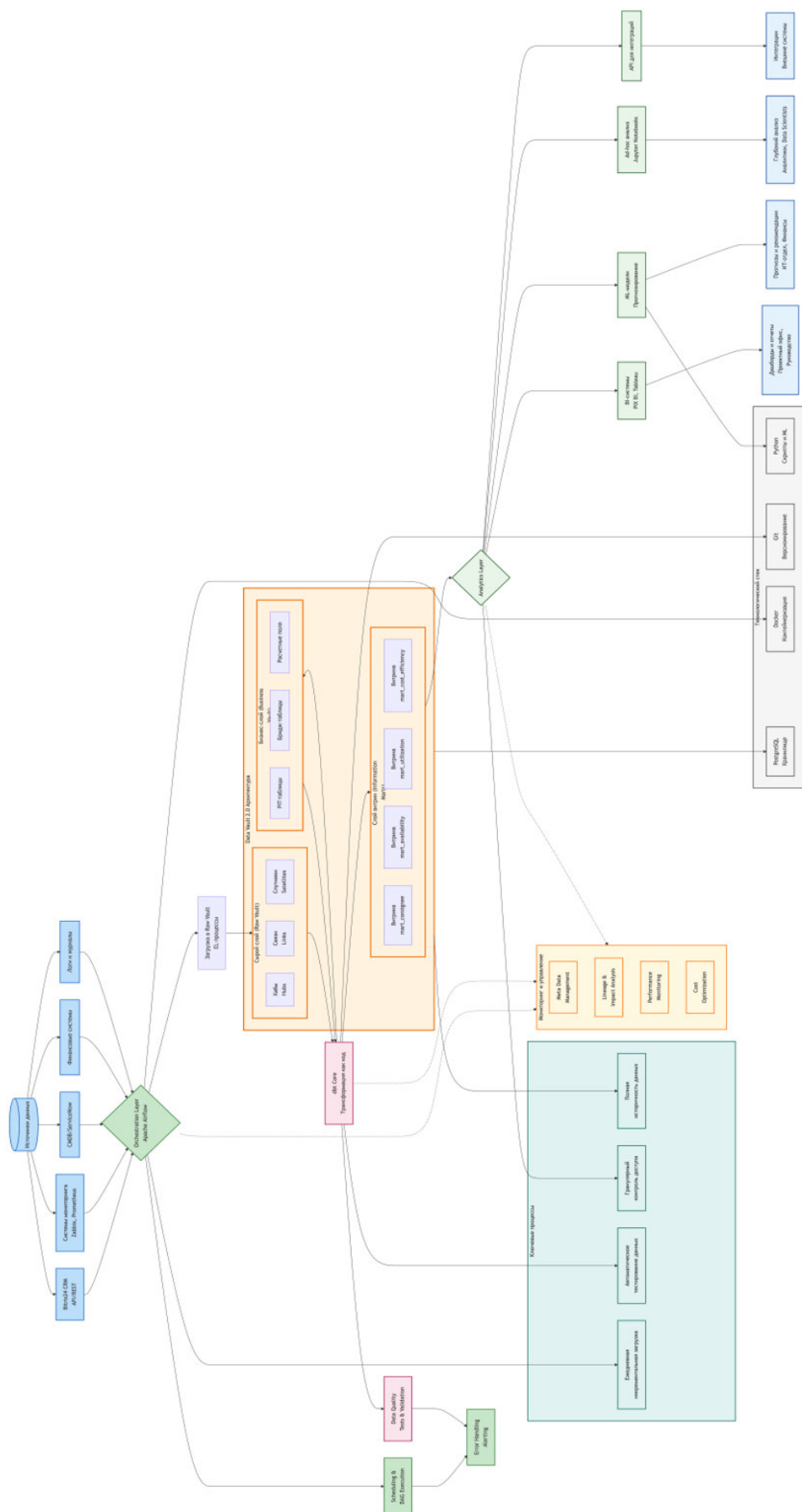


Рисунок 1. Общий поток данных проекта по ИТ-инфраструктуре
Figure 1. General IT infrastructure project data flow

На следующем этапе, в бизнес-моделях *marts*, реализовывалась основная логика преобразований. Здесь данные из разных источников обогащались и соединялись. Ключевой операцией стало объединение (JOIN) данных о грузополучателе с информацией об ответственном лице и связанными логистическими расстояниями. Также применялись специфические бизнес-правила, такие как агрегация значений из полей-массивов в читаемые строки с помощью функции `array_to_string`. Результатом этого этапа стали готовые таблицы.

Финальным шагом являлось создание потребительского представления. Оно реализует финальное соединение подготовленных моделей в единую денормализованную структуру, полностью соответствующую архитектуре, описанной в документации. Для обеспечения надежности и качества данных в dbt-проект были интегрированы автоматические тесты, проверяющие целостность внешних ключей, допустимость диапазонов дат и другие критерии, формализованные в критериях проверки качества данных на этапе документации. Таким образом, весь процесс – от загрузки сырых данных до формирования готовой к анализу витрины – был полностью автоматизирован, воспроизводим и контролируем. Общая схема построения витрины показана на рис. 2.

ЗАКЛЮЧЕНИЕ

Пилотная апробация предложенной методики на ограниченном сегменте данных CRM Bitrix24 и сопутствующих источников позволила получить предварительные результаты, требующие осторожной интерпретации с учётом масштаба и условий проведённого эксперимента. На данном этапе речь идёт не о статистически значимом подтверждении универсальной эффективности методики, а о демонстрации её работоспособности в контролируемой среде и выявлении факторов, которые должны быть учтены при планировании более масштабных внедрений.

Верификация исходной гипотезы, предполагавшей, что адаптация архитектуры Data Vault 2.0 путём введения атрибута источника записи, механизма адаптивной загрузки и встраивания экономической логики в слой трансформаций обеспечит измеримые операционные и экономические преимущества, была выполнена частично. Подтверждено, что в условиях пилотного контура время выполнения типового аналитического запроса сократилось с 15–20 минут до 2 секунд, а трудозатраты на подготовку ежемесячной отчётности снизились с 16 до 2 человеко-часов. Однако эти значения получены без параллельного тестирования альтернативных архитектурных решений (например, построения витрины по методологии Кимбалла или использования бессхемного подхода на основе JSON-хранилищ) на том же наборе

данных. Следовательно, зафиксированное улучшение нельзя однозначно атрибутировать именно предложенным концептуальным расширениям, а не самому факту перехода от разрозненных операционных источников к централизованному хранилищу. Полная верификация гипотезы потребовала бы сравнительного эксперимента с несколькими архитектурными вариантами и статистической обработки результатов на репрезентативной выборке инфраструктурных объектов — такая работа выходит за рамки настоящей публикации и составляет предмет дальнейших исследований.

Важным результатом стало повышение качества и достоверности данных. Использование встроенных механизмов тестирования инструмента dbt позволило реализовать автоматический контроль целостности и корректности данных, включая проверку уникальности ключевых атрибутов, согласованности связей и допустимости значений. Применение принципа декларативного описания трансформаций с их версионированием в системе контроля версий обеспечило воспроизводимость результатов и прозрачность логики обработки данных, что повышает доверие к аналитическим выводам и снижает риски принятия решений на основе некорректной информации [23].

Экономический эффект от внедрения методики выразился в выявлении резервов оптимизации инфраструктурных ресурсов и снижении трудозатрат на подготовку отчетности. В ходе анализа были выявлены три физических сервера с устойчиво низким уровнем утилизации (менее 15%), а также 20 избыточных лицензий программного обеспечения, не используемых в полном объеме. Потенциальная экономия от оптимизации использования данных ресурсов оценивается в 240 тыс. руб. в год. Дополнительный экономический эффект достигнут за счет сокращения трудозатрат аналитического персонала на подготовку отчетности с 16 до 2 человеко-часов в месяц, что соответствует 336 тыс. руб. в год. Совокупный расчетный экономический эффект для пилотного сегмента составил 576 тыс. руб. в год. Предложенный способ оценки экономического эффекта опирается на измеримые параметры (утилизация оборудования, число активных пользователей, трудозатраты) и типовые статьи расходов (амортизация, энергопотребление, стоимость лицензий), что делает его воспроизводимым за пределами пилотного контура. Методика расчёта не содержит специфических допущений, привязанных исключительно к CRM Bitrix24 или конкретной конфигурации серверного парка, и может быть адаптирована для анализа других компонентов ИТ-инфраструктуры путём подстановки актуальных для организации значений стоимости ресурсов и нормативов загрузки.

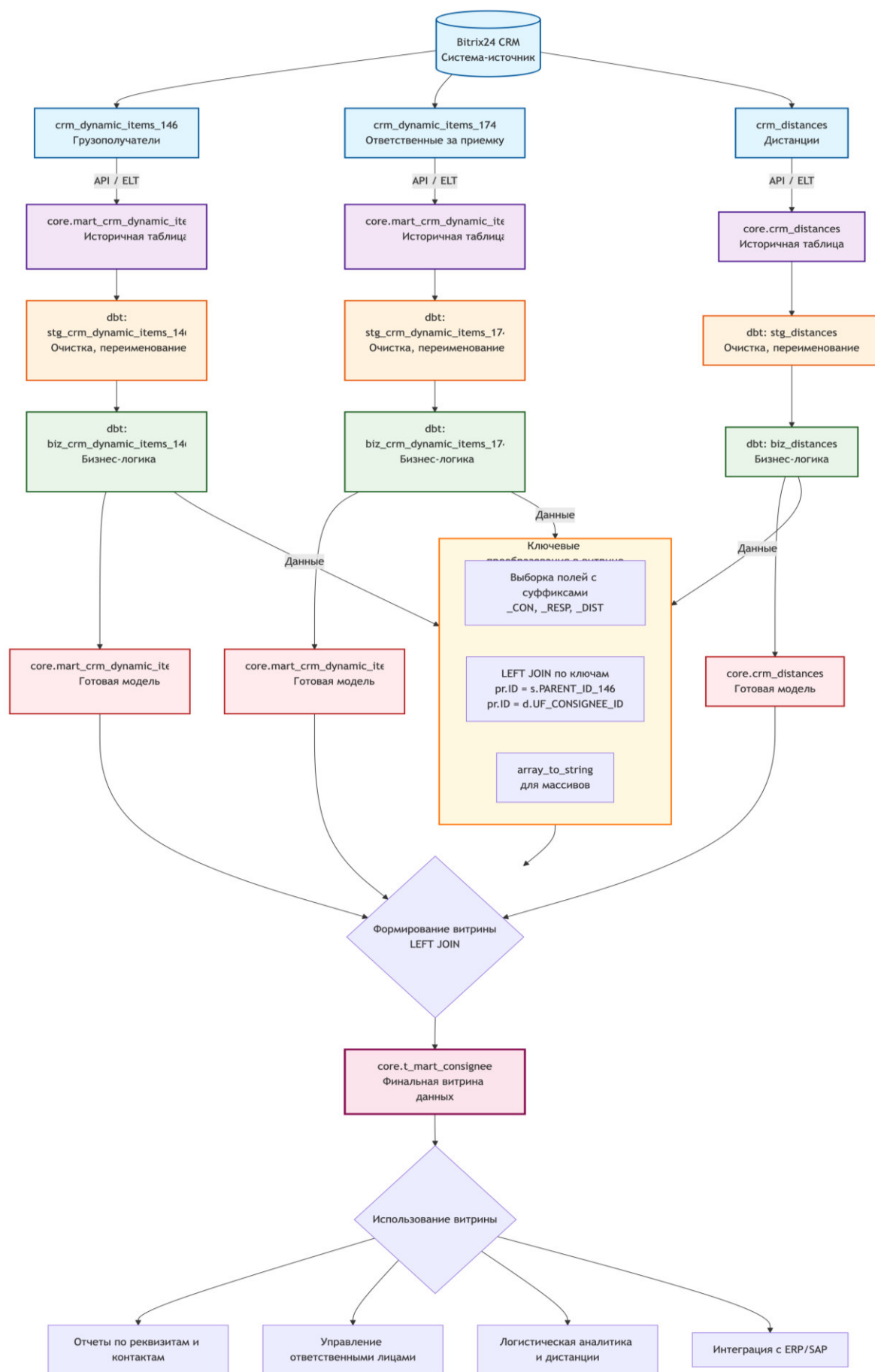


Рисунок 2. Общее представление архитектуры витрины данных
Figure 2. Overview of the data mart architecture

Помимо прямого экономического эффекта, внедрение аналитической платформы обеспечило формирование единого информационного пространства для анализа инфраструктурных и операционных данных. Это позволило обеспечить согласованность аналитических показателей, исключить расхождения в отчетных данных между подразделениями и создать основу для последующего расширения аналитической платформы на другие направления управления ИТ-инфраструктурой, включая мониторинг техниче-

ских ресурсов, анализ инцидентов и оценку эксплуатационных затрат [24]. Таким образом, результаты апробации подтверждают, что разработанная методика обеспечивает повышение эффективности обработки и анализа данных, снижение трудоемкости подготовки отчетности и создание информационной основы для принятия обоснованных управленческих решений в области управления ИТ-инфраструктурой.

*Авторы заявляют об отсутствии
конфликта интересов.*

The authors declare no conflict of interest.

ЛИТЕРАТУРА

1. Гудзь С.С., Цуканова О.А. BI-системы как инструмент data-driven управления предприятием. *Петербургский экономический журнал*. 2024. № 2. С. 124–134.
2. Лаверычев М.А., Саркисова И.О., Соколов А.А. Семантическая интеграция и адаптация данных в гетерогенных корпоративных информационных системах. <https://cyberleninka.ru/article/n/semanticheskaya-integratsiya-i-adaptatsiya-dannyh-v-geterogennyh-korporativnyh-informatsionnyh-sistemah/viewer>.
3. Полехин А.С., Галиаскаров Э.Г., Космина С.В. Исследование возможностей LLM и RAG для построения экспертной системы предприятия. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2025. № 3(65). С. 111–118. DOI: 10.6060/ivecofin.2025653.738.
4. Воробьева Г.Р. Методологические основы обработки неоднородной пространственно-временной информации в системах поддержки принятия решений. *Системная инженерия и информационные технологии*. 2023. Т. 5. № 2. С. 40–50.
5. Lennerholt C., van Laere J., Söderström E. Challenges in data warehouse development: a systematic literature review. *Information Systems Frontiers*. 2020. Vol. 22. P. 635–659.
6. Sawadogo P., Darmont J. Data warehouse architectures: from data marts to data lakes. *Journal of Data and Information Quality*. 2021. Vol. 13. N 2. P. 1–20.
7. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Indianapolis: John Wiley & Sons. 2013. 600 p.
8. Linstedt D., Olschimke M. Building a Scalable Data Warehouse with Data Vault 2.0. Morgan Kaufmann. 2015. 684 p.
9. Линстедт Д. Data Vault 2.0: Справочник по моделированию гибкого хранилища данных для предприятия. М.: Альпина Паблишер. 2021. 478 с.
10. Konakov P.O., Smolentseva T.E. Development of enterprise data warehouse architecture using Data Vault 2.0 methodology. *Stolyipin Bulletin*. 2022. N 8. P. 45–53.
11. Злобина А.В., Кочергин К.С. Обзор методологии и архитектуры Data Vault. *Современные информационные технологии и ИТ-образование*. 2021. Т. 17. № 4. С. 897–906.
12. Космина С.В., Галиаскаров Э.Г. Искусственный интеллект как инструмент подготовки профессионалов для обслуживания технических систем предприятий. *Известия высших учебных заведений. Серия «Экономика, финансы и управление производством» [Ивэкофин]*. 2024. № 3(61). С. 104–112. DOI: 10.6060/ivecofin.2024613.695.
13. Espinoza F., Maryska M., Doucek P. Modern data architectures: evaluation framework for selecting suitable data platforms. *Quality Innovation Prosperity*. 2025. Vol. 29. N 2. P. 45–62.
14. Machado I.A., Costa C., Santos M.Y. Data mesh: concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science*. 2021. Vol. 196. P. 263–271.

REFERENCES

1. Gudz S.S., Tsukanova O.A. BI systems as a tool for data-driven enterprise management. *Peterburg economic journal*. 2024. N 2. P. 124–134. (in Russian).
2. Laverychev M.A., Sarkisova I.O., Sokolov A.A. Semantic integration and adaptation of data in heterogeneous corporate information systems. <https://cyberleninka.ru/article/n/semanticheskaya-integratsiya-i-adaptatsiya-dannyh-v-geterogennyh-korporativnyh-informatsionnyh-sistemah/viewer>. (in Russian).
3. Polekhin A.S., Galiaskarov E.G., Kosmina S.V. Exploring the capabilities of LLM and RAG for building an enterprise expert system. *Ivecofin*. 2025. N 3(65). P. 111–118. DOI: 10.6060/ivecofin.2025653.738. (in Russian).
4. Vorobeva G.R. Methodological foundations for processing heterogeneous spatial and temporal information in decision support systems. *Systemic engineering and information technologies*. 2023. Vol. 5. N 2. P. 40–50. (in Russian).
5. Lennerholt C., van Laere J., Söderström E. Challenges in data warehouse development: a systematic literature review. *Information Systems Frontiers*. 2020. Vol. 22. P. 635–659.
6. Sawadogo P., Darmont J. Data warehouse architectures: from data marts to data lakes. *Journal of Data and Information Quality*. 2021. Vol. 13. N 2. P. 1–20.
7. Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Indianapolis: John Wiley & Sons. 2013. 600 p.
8. Linstedt D., Olschimke M. Building a Scalable Data Warehouse with Data Vault 2.0. Morgan Kaufmann. 2015. 684 p.
9. Linstedt D. Data Vault 2.0: A Guide to Agile Data Warehouse Modeling. Moscow: Alpina Publisher. 2021. 478 p. (in Russian).
10. Konakov P.O., Smolentseva T.E. Development of enterprise data warehouse architecture using Data Vault 2.0 methodology. *Stolyipin Bulletin*. 2022. N 8. P. 45–53.
11. Zlobina A.V., Kochergin K.S. Review of Data Vault methodology and architecture. *Modern information technologies and IT education*. 2021. Vol. 17. N 4. P. 897–906. (in Russian).
12. Kosmina S.V., Galiaskarov E.G. Artificial intelligence as a tool for training professionals to service technical systems of enterprises. *Ivecofin*. 2024. N 3(61). P. 104–112. DOI: 10.6060/ivecofin.2024613.695. (in Russian).
13. Espinoza F., Maryska M., Doucek P. Modern data architectures: evaluation framework for selecting suitable data platforms. *Quality Innovation Prosperity*. 2025. Vol. 29. N 2. P. 45–62.
14. Machado I.A., Costa C., Santos M.Y. Data mesh: concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science*. 2021. Vol. 196. P. 263–271.
15. Priebe T., Neumaier S., Markus S. Finding your way through the jungle of big data architectures. *IEEE International Conference on Big Data*. 2021. P. 5994–5996.

15. **Priebe T., Neumaier S., Markus S.** Finding your way through the jungle of big data architectures. *IEEE International Conference on Big Data*. 2021. P. 5994–5996.
16. **Михайлов А.Н., Аксенова М.В.** Использование Apache Airflow для оркестрации процессов обработки данных. *Прикаспийский журнал: управление и высокие технологии*. 2023. № 2 (62). С. 60–71.
17. Apache Software Foundation. Apache Airflow Documentation. <https://airflow.apache.org/docs/>.
18. **Khaire S., Ware R., More R.** Transforming data engineering with DBT: modeling and automation techniques. *International Journal of Advanced Research in Science, Engineering and Management*. 2024. Vol. 11. N 6. P. 112–118.
19. dbt Labs. dbt Documentation. <https://docs.getdbt.com/>.
20. PostgreSQL Global Development Group. PostgreSQL Documentation. <https://www.postgresql.org/docs/>.
21. **Астраханцева И.А., Астраханцев Р.Г., Бобков С.П.** Агентные интеллектуальные системы с точки зрения системного анализа. *Современные наукоемкие технологии. Региональное приложение*. 2025. № 4(84). С. 101–109. DOI: 10.6060/snt.20258404.00014.
22. 1С-Битрикс. Документация REST API Bitrix24. https://dev.1c-bitrix.ru/rest_help/.
23. **Хасанова А.М., Абдуллаев Э.И.** Интеллектуальный анализ процессов по данным журналов событий информационных систем. *Вестник компьютерных и информационных технологий*. 2022. N 5 (215). С. 30–38.
24. **Зимнуров М.Ф., Астраханцева И.А.** Методология создания многосвязных структур данных с применением LLM в рабочих проектах. *Современные наукоемкие технологии. Региональное приложение*. 2025. № 1(81). С. 76–83. DOI: 10.6060/snt.20258101.0009.
16. **Mikhailov A.N., Aksenova M.V.** Using Apache Airflow for data processing orchestration. *Prikaspiyskiy journal: management and high tech*. 2023. N 2 (62). P. 60–71. (in Russian).
17. Apache Software Foundation. Apache Airflow Documentation. <https://airflow.apache.org/docs/>.
18. **Khaire S., Ware R., More R.** Transforming data engineering with DBT: modeling and automation techniques. *International Journal of Advanced Research in Science, Engineering and Management*. 2024. Vol. 11. N 6. P. 112–118.
19. dbt Labs. dbt Documentation. <https://docs.getdbt.com/>.
20. PostgreSQL Global Development Group. PostgreSQL Documentation. <https://www.postgresql.org/docs/>.
21. **Astrakhantseva I.A., Astrakhantsev R.G., Bobkov S.P.** Agent-based intelligent systems from the point of view of system analysis. *Modern High Technologies. Regional Application*. 2025. N 4(84). P. 101–109. DOI: 10.6060/snt.20258404.00014. (in Russian).
22. 1С-Bitrix. Bitrix24 REST API Documentation. https://dev.1c-bitrix.ru/rest_help/. (in Russian).
23. **Khasanova A.M., Abdullaev E.I.** Process mining based on event logs of information systems. *International Journal of Open Information Technologies*. 2022. N 5 (215). P. 30–38. (in Russian).
24. **Zimnurov M.F., Astrakhantseva I.A.** Usage of large language models for building multi-vectored and multi-linked data model of work process. *Modern High Technologies. Regional Application*. 2025. N 1(81). P. 76–83. DOI: 10.6060/snt.20258101.0009. (in Russian).

Поступила в редакцию 25.02.2026
Принята к опубликованию 10.03.2026

Received 25.02.2026
Accepted 10.03.2026