

УДК 004.658

Ключевые проблемы внедрения ETL/ELT-процессов и методы их преодоления

- Битченков Дмитрий Игоревич** — магистрант 2-го курса направления 09.04.02 «Информационные системы и технологии». Научные интересы: интеграция данных, ETL/ELT-процессы, большие данные. E-mail: dmanbgu@gmail.com
- Злобина Маргарита Сергеевна** — студент бакалавриата 3-го курса направления 09.03.01 «Информатика и вычислительная техника». Научные интересы: анализ данных, ETL/ELT-процессы, большие данные. E-mail: zzzlobina.ms@gmail.com
- Ермаков Сергей Геннадьевич** — д-р техн. наук, профессор, заведующий кафедрой «Информационные и вычислительные системы». Научные интересы: обработка данных, корпоративные хранилища данных, безопасность данных. E-mail: ermakov@pgups.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Битченков Д. И., Злобина М. С., Ермаков С. Г. Ключевые проблемы внедрения ETL/ELT-процессов и методы их преодоления // Интеллектуальные технологии на транспорте. 2025. № 3 (43). С. 41–48. DOI: 10.20295/2413-2527-2025-343-41-48

Аннотация. Рассматривается проблематика использования процессов ETL и ELT в системах обработки данных. Эти процессы играют ключевую роль в интеграции и анализе данных, однако их применение сопряжено с рядом проблем, требующих научно-теоретического осмысления. **Цель:** изучение проблем, возникающих при построении систем обработки данных, включая высокие временные затраты на настройку и оптимизацию, вопросы качества данных, масштабируемости и безопасности. **Результаты:** предложены рекомендации по выбору между ETL и ELT, а также решения для оптимизации процессов, обеспечения качества и безопасности данных. **Практическая значимость:** заключается в разработке новых подходов повышения эффективности работы с корпоративными хранилищами данных. **Обсуждение:** подчеркивается необходимость оптимизации процессов ETL и ELT, повышения качества и безопасности данных, обоснованного выбора архитектуры хранилищ данных для эффективной работы с информацией.

Ключевые слова: ETL, ELT, интеграция данных, проблемы качества данных, масштабируемость, безопасность данных, выбор архитектуры, оптимизация процессов, хранилища данных

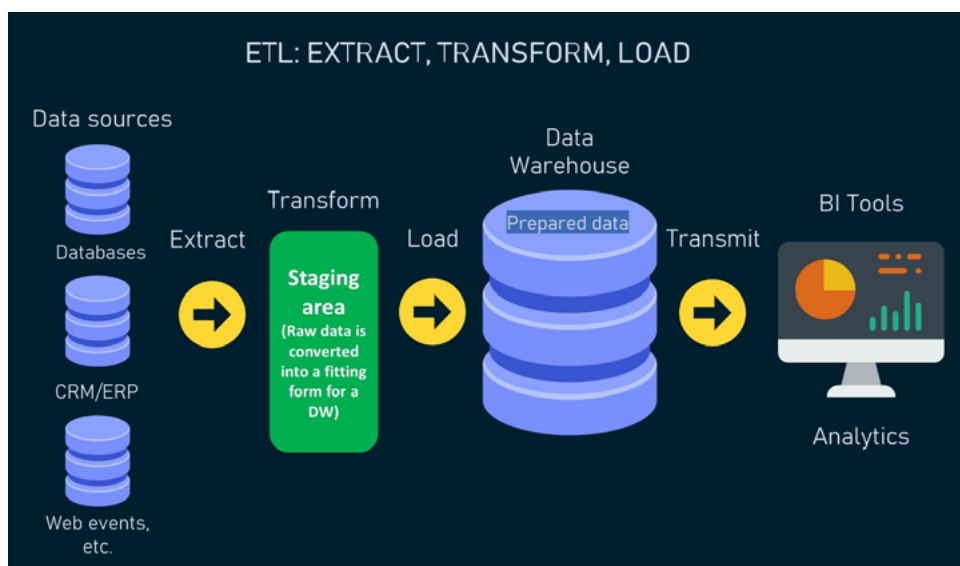
2.3.5 — математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей; 2.3.6 — методы и системы защиты информации, информационная безопасность (технические науки)

Введение

В современных системах обработки данных ETL- и ELT-процессы занимают важное место, выступая, по мнению исследователей, важной частью архитектуры [1]. Благодаря этим процессам компании могут собирать данные из различных источников, преобразовывать их для последующего анализа и загружать в корпоративные хранилища. Тем не менее реализация процессов обработки данных

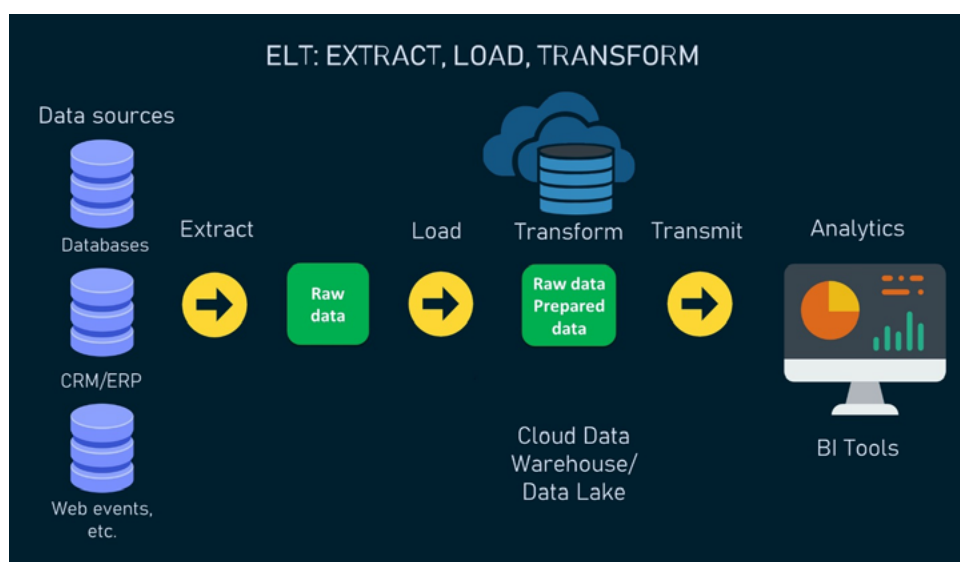
сопряжена с рядом сложностей, требующих комплексного анализа.

ETL (Extraction, Transformation, Loading) — процесс извлечения неподготовленных данных из различных источников, обработки в промежуточном хранилище и передача в единую целевую систему. Схема реализации ETL-процесса представлена на рис. 1.



Источник: <http://www.altexsoft.com/blog/etl-vs-elt>

Рис. 1. Рабочий процесс ETL



Источник: <http://www.altexsoft.com/blog/etl-vs-elt>

Рис. 2. Рабочий процесс ELT

В управлении информационными системами ETL-процессы рассматриваются как основной способ организации импорта данных [2]. В частности, инструменты ETL предназначены для интеграции данных с целью удовлетворения требований систем управления реляционными базами данных и/или традиционных хранилищ данных, поддерживающих OLAP (online analytical processing), то есть аналитическую обработку данных онлайн. Для корректной работы OLAP и запросов на языке SQL необходимо, чтобы массивы данных были стандартизованы с

помощью ряда преобразований, которые должны быть выполнены перед передачей данных в хранилище. Более того, исследователи подчеркивают и важность оптимизации самих ETL-процессов для корректного и бесперебойного импорта данных [3].

ELT (Extraction, Loading, Transformation), в свою очередь, представляет собой процесс, в котором, в отличие от ETL, иной порядок процессов: после сбора данных из источников они передаются непосредственно в центральный репозиторий, где проходят последующую обработку (рис. 2).

В целом ETL-процессы реализуются благодаря развитию аналитических платформ, способных эффективно выполнять сложные вычисления прямо на стороне хранилища данных [4].

Ключевые проблемы внедрения ETL/ELT-процессов и методы их преодоления

Среди ключевых проблем, характерных для работы с корпоративными хранилищами данных, можно выделить трудоемкость процессов извлечения и преобразования данных, требующих значительных усилий для настройки и оптимизации. Указанная проблема обостряется, если данные поступают из множества разнородных источников и имеют разные форматы — часто это приводит к задержкам получения актуальной информации и, как следствие, невозможности оперативно принимать обоснованные бизнес-решения, особенно на стороне работы аналитиков [5].

Среди современных ETL- и ELT-инструментов (Talend, AWS Glue, Informatica, Apache NiFi) есть решения, позволяющие компаниям извлекать и обрабатывать данные и в результате сокращать затраты на ресурсы благодаря автоматизации процессов. Дополнительное снижение времени развертывания обеспечивается внедрением стандартизированных шаблонов преобразования данных, а производительность увеличивается за счет использования облачных технологий с готовыми коннекторами для распространенных источников данных.

Другой сложностью выступает обеспечение качества данных. Исходные наборы данных часто содержат противоречивые записи, пропущенные или ошибочные значения, которые при недостаточном контроле приводят к передаче некорректной информации в хранилище данных. Построение аналитических выводов на основе таких данных делает их недостоверными, что снижает эффективность бизнес-процессов. Например, отсутствие проверки и очистки данных на этапе передачи при реализации проекта для крупного ретейлера повлекло за собой замедление процесса анализа из-за накопления 15 % дублирующихся записей в целевой системе [6].

Эффективным способом обеспечения качества данных являются системы Data Quality Management, которые представляют собой набор практик по контролю, повышению и поддержанию качества данных в организации [7]. Такие системы автоматизируют процессы выявления и исправления ошибок, минимизируют риски появления ошибочных данных за счет правил проверки и очистки данных на этапе преобразования, обеспечивают регулярный аудит и мониторинг качества данных. Например, использование Apache NiFi позволяет автоматизировать проверку данных с помощью встроенных процессоров, таких как ValidateRecord и DetectDuplicate. В одном из проектов для финансовой компании внедрение таких инструментов сократило количество ошибок на 30 %. Исследователи подчеркивают, что Apache NiFi является одним из наиболее эффективных решений на рынке на сегодняшний день [8].

Также в качестве проблемы следует выделить масштабируемость ETL и ELT-процессов, обусловленную ростом объема данных и численности их источников. Игнорирование этой проблемы приводит к увеличению потребности организаций в вычислительных и человеческих ресурсах, увеличению затрат на поддержание и расширение информационных систем. Более того, преобразование данных на локальной инфраструктуре организаций приводит к снижению эффективности в связи с ограниченностью ресурсов. Так, обработка 1 ТБ данных на локальных серверах телекоммуникационной компании при работе над проектом для этой организации могла занимать более 12 часов. Для нивелирования описанной проблемы целесообразно использовать облачные сервисы: AWS Glue, Azure, Google Cloud. Кейс-анализ, проведенный в ходе настоящего исследования, свидетельствует о том, что в корпоративных информационных системах время обработки 1 ТБ данных после перехода на AWS Glue может быть сокращено до 2 часов благодаря использованию распределенных вычислений, а стоимость поддержки информационной инфраструктуры снижается за счет масштабирования ресурсов в зависимости от уровня нагрузки на облачных платформах.

Таблица 1

Сравнение преимуществ и недостатков подходов ETL и ELT

Критерии	ETL	ELT
Скорость обработки	Медленнее для больших объемов данных	Быстрее благодаря использованию мощностей целевой системы
Гибкость	Ограничена этапом трансформации	Высокая, так как трансформация выполняется после загрузки
Сложность реализации	Высокая, требует мощных серверов	Низкая, использует ресурсы облачных платформ
Стоимость	Высокая в связи с затратами на инфраструктуру	Сравнительно более низкая благодаря использованию облачных ресурсов

Также стоит обратить внимание на безопасность данных. ETL- и ELT-процессы неотрывно связаны с трансфером чувствительной информации между различными системами, что приводит к рискам утечек данных и несанкционированного доступа. В результате организация может столкнуться с серьезными финансовыми последствиями. Например, утечка данных в одной из крупных компаний в 2021 году повлекла за собой штраф в размере \$ 5 млн за нарушение общего регламента по защите данных (General Data Protection Regulation, GDPR).

Для снижения вероятности возникновения описанных рисков компаниям требуется реализация надежных механизмов защиты данных на всех этапах работы с ними, включая шифрование при передаче и хранении данных, регулярные аудиты безопасности систем и тестирование на наличие уязвимостей, а также внедрение систем авторизации и аутентификации (Identity and Access Management, IAM). В качестве примера можно привести облачную систему хранения Amazon S3, которая обеспечивает шифрование данных как на уровне сервера, так и на уровне клиента благодаря использованию AES-256. Дополнительно необходимо упомянуть обязательное соответствие стандартам (GDPR, HIPAA, ISO 27001) единых корпоративных хранилищ данных. Исследователи подчеркивают необходимость внимательного отношения к рискам при выборе программных продуктов для данных целей, особенно в условиях ограниченного использования зарубежного ПО и тенденции к его импортозамещению [9].

Наконец, проблемным может стать выбор между внедрением ETL и ELT, так как требует содержательного анализа требований компаний,

архитектуры данных и уже используемых технологий. Оба решения имеют свои преимущества и недостатки, вследствие чего неправильный выбор может привести к снижению производительности бизнес-процессов, увеличению количества затрачиваемых ресурсов и, как следствие, ухудшению качества аналитики.

Для выбора оптимальной методологии организациям целесообразно провести многоаспектную оценку потребностей с учетом специфики корпоративной информационной системы и фокусируемых в ней данных [10]. ETL-процесс подойдет в случае необходимости сложных преобразований на небольших объемах данных. ELT следует внедрять в случае работы с большими объемами данных, при наличии цели обеспечения гибкости информационной системы за счет разделения этапов загрузки данных и их последующей трансформации. Также можно использовать комплексный подход, при котором часть информации проходит через ETL-процесс, в то время как другая часть — через ELT. Основные преимущества и недостатки каждого решения приведены в табл. 1.

Так, для небольших объемов данных (до 100 ГБ) ETL может быть более подходящим, поскольку позволяет выполнять сложные преобразования до загрузки. В свою очередь, для больших объемов данных (свыше 1 ТБ) ELT предпочтительнее, так как использует вычислительные мощности целевой системы, такие как Snowflake, BigQuery или Redshift.

При наличии потребности быстрой обработки данных ELT может быть более эффективным, так как загрузка данных происходит без предварительной трансформации. ETL же подходит для сце-

нариев, при которых критически важна точность данных перед загрузкой.

Также для сложных преобразований, требующих значительных вычислительных ресурсов, ETL может быть более подходящим в силу возможности выполнять их на этапе трансформации. ELT же лучше подходит для простых или инкрементальных преобразований, которые можно выполнить после загрузки данных.

В целом ELT является более экономичным решением, так как использует облачные ресурсы и не требует мощных локальных серверов, что актуально для организаций в условиях нестабильности экономической среды и явной потребности в оптимизации финансовых издержек: кейсы показывают, что ELT может существенно снизить затраты на инфраструктуру (до 25 %). ETL как альтернатива требует значительных инвестиций в инфраструктуру, особенно для обработки больших объемов данных: здесь важен системный предварительный анализ экономической целесообразности интеграции.

Практические рекомендации при реализации ETL- и ELT-процессов

Опираясь на результаты проведенного исследования, можно заключить, что для автоматизации и мониторинга ETL- и ELT-процессов приоритетно использование таких инструментов, как Apache Airflow, Talend или Informatica. В частности, Airflow позволяет создавать сложные рабочие процессы (DAGs), отслеживать их выполнение в реальном времени.

Для регулярного аудита данных рекомендуется внедрение системы контроля качества данных, приоритетные варианты — Great Expectations, Deequ или Dataiku. Указанные инструменты помогают выявлять аномалии, дубликаты и некорректные данные на ранних этапах.

Также целесообразным видится обучение сотрудников современным технологиям и стандартам безопасности в рамках курсов по облачным платформам (AWS, Azure, GCP) и инструментам ETL и ELT (Talend, Informatica, dbt).

Следует отметить, что регулярное тестирование процессов и их оптимизация способствует повышению производительности: использование инкрементальной загрузки данных вместо полной загрузки может значительно сократить время обработки, поэтому важно обеспечить непрерывность процесса мониторинга. В ходе исследования было установлено, что при реализации проекта построения хранилища данных крупной телекоммуникационной компании обработка 2 ТБ данных занимала более 24 часов на локальной инфраструктуре, тогда как после перехода на ELT с использованием Google BigQuery время обработки сократилось до 3 часов, а затраты на инфраструктуру снизились на 40 %. Также при реализации схожего проекта для одной из крупных финансовых компаний наблюдалась проблема низкого качества данных из-за отсутствия проверки на этапе загрузки. Решить эту проблему позволило внедрение Apache NiFi для автоматической валидации данных: численность ошибок сократилась на 50 %.

Заключение

Процессы ETL и ELT стали неотъемлемой частью инфраструктуры работы с данными, однако их реализация сопряжена с проблемами, систематизированными в рамках данного исследования. Для повышения эффективности работы с информацией организациям важно не только оптимизировать эти процессы, но и обеспечить должный уровень качества и безопасности данных, обоснованно выбирать решения в области построения архитектуры хранилищ данных.

Рациональный выбор между подходами ETL и ELT, наряду с их корректной реализацией, способен существенно улучшить управление данными, повысить точность аналитики и сократить издержки. В условиях стремительного роста объемов информации особое значение приобретает гибкость систем и готовность к их адаптации, что требует от специалистов постоянного обновления знаний и профессиональных навыков.

СПИСОК ИСТОЧНИКОВ

1. Остудина К. А., Киреев В. С. Архитектура хранения и обработки текстовых открытых данных // Интеллектуальные технологии в науке и образовании: материалы Международной научно-практической конференции (Новочеркасск, Россия, 24–25 ноября 2023 г.). Новочеркасск: Лик, 2023. С. 102–108.
2. Соколов В. А., Пономарева К. А. ETL-процессы как способ организации импорта данных для обеспечения функций управления / В. А. Соколов, К. А. Пономарева; науч. рук. Р. И. Кузьмич // Проспект Свободный — 2023: материалы XIX Международной научной конференции студентов, аспирантов и молодых ученых (Красноярск, Россия, 24–29 апреля 2023 г.). Красноярск: Сибирский федеральный ун-т, 2023. С. 3350–3353.
3. Упаева П. В. Оптимизация ETL-процессов: обзор отечественного рынка // Вестник науки. 2024. № 6 (75), Т. 4. С. 1218–1224.
4. Пьянзин А. М., Лайша А. К., Аносов М. С. Обзор современных решений для хранения и структурирования данных в аддитивных технологиях // Технологии аддитивного производства. 2025. Т. 3, № 1. С. 43–48.
5. Блохин В. В., Лосев В. С. Автоматизация процесса актуализации данных в витринах: как упростить работу аналитиков // Финансовая экономика. 2024. № 10. С. 100–103.
6. Тавторкин Н. О., Куляшова Н. М. ETL-процессы в работе с данными // Будущее науки — 2024: сборник научных статей 11-й Международной молодежной научной конференции (Курск, Россия, 18–19 апреля 2024 г.): в 5 т. Т. 4. Курск: Университетская книга, 2024. С. 175–178.
7. Макаров В. В., Волчик О. В. Управление данными в системах менеджмента качества // Экономика и качество систем связи. 2024. № 4 (34). С. 4–13.
8. Лабинский А. Ю. Программные средства обработки больших объемов данных // Природные и техногенные риски (физико-математические и прикладные аспекты). 2024. № 4. С. 45–52. DOI: 10.61260/2307-7476-2024-2023-4-45-52.
9. Мандрика О. С., Ермаков С. Г. Модели построения корпоративного хранилища данных // Информатизация и связь. 2024. № 2. С. 22–26. DOI: 10.34219/2078-8320-2024-15-2-22-26.
10. Яковенко Е. С. Основные понятия и примеры программных инструментов в сфере обработки больших данных // Научные исследования в современном мире: опыт, проблемы и перспективы развития: сборник научных статей по материалам XIV Международной научно-практической конференции (Уфа, Россия, 19 апреля 2024 г.). Уфа: НИЦ Вестник науки, 2024. С. 563–567.

Дата поступления: 25.08.2025

Решение о публикации: 02.09.2025

Key Challenges of Implementing ETL/ELT Processes and Methods to Overcome Them

- Dmitry I. Bitchenkov** — 2nd year Master's Degree Student in 09.04.02 Information Systems and Technologies. Research interests: data integration, ETL/ELT processes, big data. E-mail: dmanbgu@gmail.com
- Margarita S. Zlobina** — 3rd year Bachelor's Degree in 09.03.01 Informatics and Computer Technology. Research interests: data mining, ETL/ELT processes, big data. E-mail: zzzlobina.ms@gmail.com
- Sergey G. Ermakov** — D. Sci. in Engineering, Professor, Head of the Information and Computing Systems Department. Research interests: data processing, corporate data warehouses, data security. E-mail: ermakov@pgups.ru

Emperor Alexander I St. Petersburg State Transport University, 9, Moskovsky ave., Saint Petersburg, 190031, Russia

For citation: Bitchenkov D. I., Zlobina M. S., Ermakov S. G. Key Challenges of Implementing ETL/ELT Processes and Methods to Overcome Them. *Intellectual Technologies on Transport*, 2025, No. 3 (43), Pp. 41–48. DOI: 10.20295/2413-2527-2025-343-41-48. (In Russian)

Abstract. *This paper examines the challenges associated with the use of ETL and ELT processes within data processing systems. These processes are of pivotal significance in the realms of data integration and analysis. However, a number of challenges that necessitate a robust foundation in scientific and theoretical understanding accompanies the implementation of these processes. **Purpose:** to examine the challenges associated with the development of data processing systems, with a particular focus on the significant time input associated with configuration and optimization, and issues concerning data quality, scalability, and security. **Results:** a number of recommendations and solutions have been proposed for choosing between ETL and ELT, for the optimization of processes, and for insured data quality and security. **Practical significance:** new approaches have been developed to improve the efficiency of work with corporate data storage. **Discussion:** the necessity to optimize ETL and ELT processes, to improve data quality and security, and to make a justified choice of data storage architecture for effective information management is emphasized.*

Keywords: ETL, ELT, data integration, data quality issues, scalability, data security, architecture selection, process optimization, data storage

REFERENCES

1. Ostudina K. A., Kireev V. S. Arkhitektura khraneniya i obrabotki tekstovykh otkrytykh dannykh [Architecture of Storage and Processing of Textual Open Data], *Intellektualnye tekhnologii v nauke i obrazovanii: materialy Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [Intelligent Technologies in Science and Education: Materials of the International Scientific and Practical Conference], Novocherkassk, Russia, November 24–25, 2023. Novocherkassk, Lik Publishing House, 2023, Pp. 102–108. (In Russian)
2. Sokolov V. A., Ponomareva K. A. ETL-protsessy kak sposob organizatsii importa dannykh dlya obespecheniya funktsiy upravleniya [ETL processes as a way of organizing data import to ensure management functions], *Prospekt Svobodnyy — 2023: materialy XIX Mezhdunarodnoy nauchnoy konferentsii studentov, aspirantov i molodykh uchenykh* [Prospect Svobodnyy — 2023: Proceedings of the XIX International Scientific Conference for Undergraduate, Postgraduate, PhD Students and Early Career Researchers], Krasnoyarsk, Russia, April 24–29, 2023. Krasnoyarsk, Siberian Federal University, 2023, Pp. 3350–3353. (In Russian)
3. Upaeva P. V. Optimizatsiya ETL-protsessov: obzor otechestvennogo rynka [Optimization of ETL Processes: An Review of the Domestic Market], *Vestnik Nauki*, 2024, No. 6 (75), Vol. 4. Pp. 1218–1224. (In Russian)
4. Pyanzin A. M., Laisha A. K., Anosov M. S. Obzor sovremennykh resheniy dlya khraneniya i strukturirovaniya dannykh v additivnykh tekhnologiyakh [Overview of Modern Solutions for Data Storage and Structuring in Additive Technologies], *Tekhnologii additivnogo proizvodstva* [Additive Fabrication Technologies], 2025, Vol. 3, No. 1, Pp. 43–48. (In Russian)
5. Blokhin V. V., Losev V. S. Avtomatizatsiya protsessa aktualizatsii dannykh v vitrinakh: kak uprostit rabotu analitikov [Automation of the Process of Data Acquisition in Storefronts: How to Simplify the Work of Analysts], *Finansovaya ekonomika* [Financial Economics], 2024, No. 10, Pp. 100–103. (In Russian)
6. Tavtorkin N. O., Kulyashova N. M. ETL-protsessy v rabote s dannymi [ETL processes in working with data], *Budushchee nauki — 2024: sbornik nauchnykh statey 11-y Mezhdunarodnoy molodezhnoy nauchnoy konferentsii* [The Future of Science — 2024: Collection of Scientific Articles of the 11th International Youth Scientific Conference], Kursk, Russia, April 18–19, 2024, Vol. 4. Kursk, Universitetskaya Kniga Publishing House, 2024, Pp. 175–178. (In Russian)

7. Makarov V. V., Volchik O. V. Upravlenie dannymi v sistemakh menedzhmenta kachestva [Data Management in Quality Management Systems], *Ekonomika i Kachestvo Sistem Svyazi*, 2024, No. 4 (34), Pp. 4–13. (In Russian)
8. Labinskiy A. Yu. Programmnye sredstva obrabotki bolshikh obemov dannykh [Processing Software Big Data], *Prirodnye i tekhnogennye riski (fiziko-matematicheskie i prikladnye aspekty) [Natural and Man-Made Risks (Physico-Mathematical and Applied Aspects)]*, 2024, No. 4, Pp. 45–52. DOI: 10.61260/2307-7476-2024-2023-4-45-52. (In Russian)
9. Mandrika O. S., Ermakov S. G. Modeli postroeniya korporativnogo khranilishcha dannykh [Models for Building an Enterprise Data Warehouse], *Informatizatsiya i svyaz [Informatization and Communication]*, 2024, No. 2, Pp. 22–26. DOI: 10.34219/2078-8320-2024-15-2-22-26. (In Russian)
10. Yakovenko E. S. Osnovnye ponyatiya i primery programmnykh instrumentov v sfere obrabotki bolshikh dannykh [Basic concepts and examples of software tools in the field of big data processing], *Nauchnye issledovaniya v sovremennoy mire: opyt, problemy i perspektivy razvitiya: sbornik nauchnykh statey po materialam XIV Mezhdunarodnoy nauchno-prakticheskoy konferentsii [Scientific Research in the Modern World: Experience, Problems and Development Prospects: Proceedings of the XIV International Scientific and Practical Conference]*, Ufa, Russia, April 19, 2024. Ufa, NIC Vestnik Nauki Publishing House, 2024, Pp. 563–567. (In Russian)

Received: 25.08.2025

Accepted: 02.09.2025