



Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.9

## ПРИМЕНЕНИЕ ТЕХНОЛОГИИ ФАБРИКИ ДАННЫХ

**Тикки Д.А., Никольский В.Е., Самошкин Н.С., Сафонова Т.В., <sup>1</sup>Мокряк А.В.**  
*ФГБОУ ВО "РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ" Санкт-Петербург, Россия (192007, город Санкт-Петербург, Воронежская ул., д. 79)*

<sup>1</sup>*ФГБОУ ВО "САНКТ-ПЕТЕРБУРГСКИЙ УНИВЕРСИТЕТ ГОСУДАРСТВЕННОЙ ПРОТИВОПОЖАРНОЙ СЛУЖБЫ МИНИСТЕРСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ ПО ДЕЛАМ ГРАЖДАНСКОЙ ОБОРОНЫ, ЧРЕЗВЫЧАЙНЫМ СИТУАЦИЯМ И ЛИКВИДАЦИИ ПОСЛЕДСТВИЙ СТИХИЙНЫХ БЕДСТВИЙ ИМЕНИ ГЕРОЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ГЕНЕРАЛА АРМИИ Е.Н.ЗИНИЧЕВА", Санкт-Петербург, Россия (196105, г. Санкт-Петербург, Московский проспект, д.149), e-mail: mokryakanna@mail.ru*

В данной работе исследуется применение технологии фабрики данных в современных информационных системах. Фабрика данных является эффективным инструментом для управления, хранения и обработки больших объемов данных. Рассматриваются основные принципы построения фабрики данных, ее основные компоненты и преимущества от ее использования. Также рассматриваются основные вызовы и проблемы, с которыми сталкиваются организации при внедрении фабрики данных.

В работе проводится анализ существующих методов и подходов к реализации фабрики данных, а также оцениваются их преимущества и недостатки. Рассматривается роль фабрики данных в повышении эффективности бизнес-процессов и принятии обоснованных решений на основе данных.

В заключении делается вывод о том, что применение технологии фабрики данных является актуальным и перспективным направлением развития информационных систем. Рекомендации по внедрению фабрики данных в организации и ее использованию для оптимизации бизнес-процессов и принятия управленческих решений

Ключевые слова: Фабрика данных, управление данными, аналитика, инновации, безопасность данных, эффективность, будущее данных.

## APPLICATION OF DATA FACTORY TECHNOLOGY

**Tikki D.A., Nikolsky V.E., Samoshkin N.S., Safonova T.V., <sup>1</sup>Mokryak A.V.**  
*RUSSIAN STATE HYDROMETEOROLOGICAL UNIVERSITY, St. Petersburg, Russia (192007, St. Petersburg, Voronezhskaya str., 79)*

<sup>1</sup>*ST. PETERSBURG UNIVERSITY OF THE STATE FIRE SERVICE OF THE MINISTRY OF THE RUSSIAN FEDERATION FOR CIVIL DEFENSE, EMERGENCIES AND ELIMINATION OF CONSEQUENCES OF NATURAL DISASTERS NAMED AFTER THE HERO OF THE RUSSIAN FEDERATION, GENERAL OF THE ARMY E.N. ZINICHEV, St. Petersburg, Russia (196105, St. Petersburg, Moskovsky prospekt, 149), e-mail: <sup>1</sup>mokryakanna@mail.ru*

This paper examines the application of data factory technology in modern information systems. The data factory is an effective tool for managing, storing and processing large amounts of data. The basic principles of building a data factory, its main components and the benefits of using it are considered. The main challenges and problems faced by organizations when implementing a data factory are also considered.

The paper analyzes the existing methods and approaches to the implementation of the data factory, as well as assesses their advantages and disadvantages. The role of the data factory in improving the efficiency of business processes and making informed decisions based on data is considered.

---

**In conclusion, it is concluded that the use of data factory technology is an urgent and promising direction for the development of information systems. Recommendations for implementing a data factory in an organization and using it to optimize business processes and make management decisions.**

---

Keywords: Data factory, data management, analytics, innovation, data security, efficiency, the future of data.

## **Введение**

В эпоху, где структурированные данные становятся неотъемлемым ресурсом для принятия управленческих решений, концепция фабрики данных вступает в силу, предлагая компаниям уникальную возможность преобразовывать объемы информации в ценные знания. Фабрика данных — это не просто технологическое решение, но и философия управления данными, которая разбивает стену между разнородными источниками данных, создавая единую платформу для их сбора, обработки и анализа [1].

Эта инновационная концепция дает возможность компаниям не только управлять своими информационными ресурсами результативно, но и применять их для прогнозирования трендов, принятия аргументированных решений и создания стратегических преимуществ.

В данной обзорной статье по фабрике данных исследуем ее назначение, основные элементы, преимущества и недостатки, проанализируем будущие перспективы использования технологии, где она сможет стать ключевой составляющей по управлению информацией [1, 2].

## **Концепция технологии фабрики данных**

Фабрика данных представляет собой централизованную систему, разработанную для сбора, обработки, хранения и анализа данных с целью обеспечения эффективного управления информацией. Это инновационное решение, предназначенное для преодоления вызовов, связанных с растущим объемом данных в современных компаниях.

Основные характеристики фабрики данных:

1. Сбор данных: фабрика данных обеспечивает многоканальный сбор данных из различных источников, включая внутренние и внешние источники, устройства интернета вещей, социальные сети и другие.
2. Хранение данных: системы хранения данных в фабрике предоставляют надежное и масштабируемое хранилище, позволяя эффективно управлять разнообразными типами данных.
3. Обработка данных: процессы ETL (Extract, Transform, Load) автоматизированы для обеспечения качественной обработки данных, включая их очистку, трансформацию и загрузку.
4. Анализ данных: фабрика данных предоставляет инструменты для проведения разнообразного анализа, включая статистический, предсказательный, исследовательский анализ данных, а также машинное обучение [3, 4].

Причины перехода на использование фабрики данных:

1. Обеспечение единой точки доступа к данным, для облегчения их поиска и использования.
2. Оптимизация процессов сбора, обработки и анализа данных сокращает временные затраты и повышает эффективность.
3. Аналитика данных в фабрике предоставляет фирмам обширные знания для принятия обоснованных стратегических и операционных решений.

В фабрике данных процесс сбора данных является первым и критически важным этапом. Различные источники данных, такие как внутренние системы, веб-сервисы, датчики интернета вещей, социальные сети и внешние базы данных, служат основой для формирования информационного базиса [5].

1. Внутренние источники: фабрика данных интегрирует данные из различных подразделений компании, таких как финансы, маркетинг, продажи, что обеспечивает комплексное понимание внутренних процессов.

2. Внешние источники: информация из внешних источников, таких как рыночные исследования, статистика отрасли, также включается для создания более полного образа внешней среды.

3. Устройства интернета вещей: данные от сенсоров и устройств интернета вещей обеспечивают реальное время и контекстную информацию, что существенно для оперативных решений [6, 7].

Хранение данных в фабрике представляет собой фундаментальный аспект, где обеспечивается сохранность и доступность информации.

1. Хранилища данных: реляционные и нереляционные базы данных, хранилища данных в облаке и внутрипредприятийные хранилища обеспечивают оптимальное распределение и структурирование данных.

2. Безопасность данных: реализация мер безопасности, таких как шифрование, управление доступом и регулярное аудирование, гарантирует сохранность конфиденциальной информации [8].

Процессы ETL (Extract, Transform, Load) в фабрике данных обеспечивают эффективную обработку данных перед их загрузкой в хранилище.

1. Очистка данных: устранение ошибок, дубликатов и несоответствий для повышения качества данных.

2. Трансформация данных: преобразование данных в необходимый формат для обеспечения их совместимости и удобства анализа.

3. Загрузка данных: автоматизированный процесс загрузки данных в хранилище для последующего использования в аналитике.

Анализ данных в фабрике предоставляет компаниям инсайты, необходимые для принятия обоснованных решений.

1. Статистический анализ: использование методов статистики для выявления тенденций и паттернов в данных.

2. Предсказательный анализ: прогнозирование будущих событий на основе исторических данных и трендов.

3. Машинное обучение: применение алгоритмов машинного обучения для автоматического извлечения знаний из данных.

Фабрика данных, интегрируя эти компоненты, создает основу для оперативного и стратегического управления информацией в организации. Далее рассмотрим преимущества использования технологии фабрики данных [7].

### **Преимущества и недостатки фабрики данных: баланс инноваций**

Введение в мир фабрики данных невозможно без более глубокого понимания ее плюсов и минусов. Понимание преимуществ и недостатков фабрики данных играет решающую роль

в принятии решения о внедрении данного инновационного подхода к управлению информационными данными.

На Рисунке 1 отмечены основные преимущества технологии фабрики данных.



Рисунок 1 – Преимущества использования технологии фабрики данных

К недостаткам технологии фабрик данных можно отнести:

- сложность интеграции;
- риски безопасности;
- требования к ресурсам;
- необходимость культурных изменений;
- поддержка и обновления.

Понимание этих аспектов поможет предприятиям взвесить плюсы и минусы фабрики данных, прежде чем погружаться в мир инновационных подходов к управлению данными [9].

### **Примеры успешной реализации фабрики данных**

В мире быстро развивающихся технологий и данных, несколько ведущих компаний демонстрируют впечатляющий успех во внедрении и использовании фабрик данных, что привело к значительным улучшениям в их операционной эффективности и стратегической проницательности.

Amazon, гигант электронной коммерции, применяет фабрику данных для обработки и анализа огромных объемов данных от миллионов транзакций ежедневно. Фабрика данных Amazon обеспечивает бесперебойный сбор, хранение и анализ данных, что дает возможность точно предсказывать

Netflix, лидер в области потокового видео, использует фабрику данных для улучшения рекомендаций контента. Алгоритмы анализа данных в режиме реального времени обрабатывают информацию о просмотренных фильмах и предоставляют персонализированные рекомендации, что способствует удержанию аудитории [5].

Spotify, музыкальный стриминговый сервис, внедрил фабрику данных для анализа музыкальных предпочтений своих пользователей. Это позволяет Spotify создавать кураторские плейлисты, предлагать новые треки и улучшать общий пользовательский опыт [3].

Uber использует фабрику данных для обработки миллиардов транзакций поездок и анализа данных местоположения. Это позволяет компании оптимизировать маршруты, прогнозировать спрос и обеспечивать высокий уровень обслуживания клиентов [4].

Эти примеры являются иллюстрациями того, как успешное внедрение фабрики данных приводит к более эффективной работе компаний в условиях современной цифровой экономики [8, 9].

### **Вызовы и решения внедрения фабрики данных**

Вопреки своему впечатляющему потенциалу, внедрение фабрики данных не лишено вызовов. Компании, стремящиеся использовать это инновационное решение, часто сталкиваются с несколькими сложностями, которые требуют внимательного рассмотрения и стратегического подхода.

Интеграция данных из различных источников может стать сложной задачей из-за разнообразия форматов и структур данных. Решение включает в себя использование мощных инструментов ETL и разработку стандартов для обеспечения единообразия данных.

С увеличением количества собираемых данных возрастают и риски нарушения безопасности. Эффективные меры безопасности, такие как шифрование, регулярный мониторинг и обучение сотрудников, необходимы для предотвращения угроз и утечек.

Обработка и анализ больших объемов данных требует значительных вычислительных ресурсов. Облачные решения и технологии, такие как Apache Hadoop и Spark, могут предоставить масштабируемость для эффективной обработки больших данных [5].

Внедрение фабрики данных часто требует изменений в культуре и управлении организации. Недостаточная готовность сотрудников к новым технологиям и процессам может стать преградой. Решение включает в себя обучение и активное вовлечение персонала.

Фабрика данных требует постоянной поддержки и обновлений, чтобы оставаться актуальной и эффективной. Компании должны разработать стратегии по управлению жизненным циклом фабрики данных, включая регулярные обновления и мониторинг производительности.

Преодолев эти вызовы, компании могут наслаждаться преимуществами фабрики данных и эффективно использовать информацию для принятия стратегических решений. Однако ключевым моментом остается систематический подход и готовность к непрерывному улучшению.

### **Будущее фабрик данных – это эволюция информационного потока данных**

В мире постоянных инноваций и быстро меняющихся технологий фабрика данных становится ключевым компонентом успешной цифровой трансформации. Взгляд в будущее подсказывает, что эта эволюция информационного потока данных только начинается, и фабрики данных будут играть все более важную роль в повседневной деятельности компаний.

Фабрика данных будет активно использовать возможности искусственного интеллекта и машинного обучения для более точного прогнозирования, автоматизации процессов анализа

и создания персонализированных решений. Это откроет новые горизонты для бизнес-аналитики и предсказательной аналитики.

Фабрики данных будут все чаще полагаться на облачные технологии для повышения гибкости, масштабируемости и доступности. Это позволит компаниям эффективно использовать ресурсы в облаке, обеспечивая высокую производительность и сокращая затраты на инфраструктуру.

С ростом числа устройств интернета вещей, фабрика данных будет активно вовлечена в обработку и анализ данных от сенсоров и устройств. Это откроет новые возможности для предсказательного обслуживания, мониторинга состояния оборудования и улучшения операционной эффективности.

DataOps, как эволюция DevOps в сфере управления данными, станет неотъемлемой частью фабрик данных. Это подразумевает ускорение цикла разработки данных, автоматизацию процессов и активное взаимодействие между различными отделами компании [3, 6].

Развитие технологий визуализации данных позволит более наглядно и эффективно представлять результаты анализа. Интерактивные дашборды и интуитивные отчеты станут стандартом, обеспечивая легкий доступ к ценной информации для решений на разных уровнях управления.

Будущее фабрик данных обещает бескрайние возможности для инноваций и улучшений в управлении данными. Компании, готовые к адаптации и внедрению новых технологий, смогут максимально воспользоваться преимуществами, которые фабрика данных предоставляет в стремительно развивающемся цифровом мире [10].

## **Выводы**

В завершение проведенного исследования мира фабрик данных, становится ясным, что эта концепция не просто инструмент, но кладезь потенциала для компаний, стремящихся выйти вперед в эпохе данных. Фабрика данных не только упрощает процессы управления информацией, но и становится стержнем для инноваций, определяя будущий характер работы компаний в цифровой эре.

Надежность в обработке данных, стремительный доступ к информации, защита конфиденциальности и постоянное развитие — вот критические составляющие, делающие фабрику данных ценным активом для предприятия.

Сегодня, когда бизнес-ландшафт стремительно эволюционирует, умение преобразовывать массы данных в ценную информацию становится ключевым преимуществом. Фабрика данных, в свою очередь, представляет собой инструмент этого преобразования, открывая перед компаниями бескрайние горизонты возможностей.

Внедрение фабрики данных — это не просто инвестиция в технологии, но и в стратегическое видение будущего, где данные становятся неотъемлемой частью принятия решений и инноваций. Это путеводитель к тому, как преобразовать данные в ценный ресурс, способствующий успеху и росту компании.

Таким образом, фабрика данных остается не просто технологическим инструментом, но вехой в эпохе данных, где компании обретают не только возможность управлять информацией, но и преобразовывать ее в источник вдохновения, инноваций и устойчивого развития.

## Список литературы

1. Рекомендательные системы: принцип работы и порядок разработки. Электронный ресурс – Режим доступа: <https://www.simbirsoft.com/blog/machine-learning-ne-tolko-neyronki/> (Дата обращения 10.04.2024)
2. Булгакова А.В., Сафонова Т.В., Кутикова В.С. Классификация нейронных сетей Информационные технологии и системы: управление, экономика, транспорт, право. 2023. №1 (45). С.11-18
3. Современная платформа данных: подробное руководство. Электронный ресурс – Режим доступа: <https://www.fivetran.com/learn/modern-data-architecture> (Дата обращения 11.04.2024)
4. Булгакова А.В., Сафонова Т.В., Диденко А.Ю. Этапы разработки и внедрения нейронной сети в проект Информационные технологии и системы: управление, экономика, транспорт, право. 2023. №1 (45). С.87-92.
5. DevOps. Электронный ресурс – Режим доступа: <https://www.linkedin.com/pulse/devops-dataops-mlops-aiops-key-elements-what-engineer-jhagc9tf#:~:text=DataOps%20is%20an%20approach%20to,quality%2C%20and%20data%20delivery%20processes.> (Дата обращения 12.04.2024)
6. AWS Migration Hub. Электронный ресурс – Режим доступа: <https://aws.amazon.com/ru/migration-hub/> (Дата обращения 13.04.2024)
7. Сафонова Т.В., Колбина О.Н., Истомина Е.П., Каламбет М.В., Яготинцева Н.В., Peculiarities of creating a database for the IoT system of urban forest management in the city of St. Petersburg IOP Conference Series: Earth and Environmental Science. 2021
8. Тикки Д.А., Никольский В.Е., Сафонова Т.В., Самошкин Н.С., Авакян Е.В. Использование облачных технологий для оптимизации бизнес-процессов Информационные технологии и системы: управление, экономика, транспорт, право. 2023. № 1 (45). С. 76-79.
9. Полтавцева Е.А., Сафонова Т.В. Облачные решения для развития производства Информационные технологии и системы: управление, экономика, транспорт, право. 2023. № 1 (45). С. 80-86.
10. Субботина В.В., Сафонова Т.В. Раскрытие ключевой роли Data-Engineering в современной работе с данными Информационные технологии и системы: управление, экономика, транспорт, право. 2023. № 2 (46). С. 83-88.

## References

1. Recommender systems: operating principle and development procedure. Electronic resource – Access mode: <https://www.simbirsoft.com/blog/machine-learning-ne-tolko-neyronki/> (Date of access: 04/10/2024)
2. Bulgakova A.V., Safonova T.V., Kutikova V.S. Classification of neural networks Information technologies and systems: management, economics, transport, law. 2023. No. 1 (45). P.11-18
3. The Modern Data Platform: A Comprehensive Guide. Electronic resource – Access mode: <https://www.fivetran.com/learn/modern-data-architecture> (Date of access: 04/11/2024)

4. Bulgakova A.V., Safonova T.V., Didenko A.Yu. Stages of development and implementation of a neural network in a project Information technologies and systems: management, economics, transport, law. 2023. No. 1 (45). P.87-92.
  5. DevOps. Electronic resource – Access mode: <https://www.linkedin.com/pulse/devops-dataops-mlops-aiops-key-elements-what-engineer-jhagc9tf#:~:text=DataOps%20is%20an%20approach%20to,quality%2C%20and%20data%20delivery%20processes>. (Date of access: 04/12/2024)
  6. AWS Migration Hub. Electronic resource – Access mode: <https://aws.amazon.com/ru/migration-hub/> (Date of access: 04/13/2024)
  7. Safonova T.V., Kolbina O.N., Istomin E.P., Kalambet M.V., Yagotintseva N.V., Peculiarities of creating a database for the IoT system of urban forest management in the city of St. . Petersburg IOP Conference Series: Earth and Environmental Science. 2021
  8. Tikki D.A., Nikolsky V.E., Safonova T.V., Samoshkin N.S., Avakyan E.V. Using cloud technologies to optimize business processes Information technologies and systems: management, economics, transport, law. 2023. No. 1 (45). pp. 76-79.
  9. Poltavtseva E.A., Safonova T.V. Cloud solutions for production development Information technologies and systems: management, economics, transport, law. 2023. No. 1 (45). pp. 80-86.
  10. Subbotina V.V., Safonova T.V. Disclosure of the key role of Data-Engineering in modern work with data Information technologies and systems: management, economics, transport, law. 2023. No. 2 (46). pp. 83-88.
-