

ПРИМЕНЕНИЕ КОРПОРАТИВНОГО ХРАНИЛИЩА ДАННЫХ ДЛЯ ПОВЫШЕНИЯ ОБОСНОВАННОСТИ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ В КРУПНЫХ ПРОМЫШЛЕННЫХ КОМПАНИЯХ

Дробкова О.С.

Московский государственный технический университет имени Н.Э. Баумана,
Москва, Россия

drobkova.os@bmstu.ru

Аннотация. В условиях постоянно увеличивающихся объемов данных промышленным компаниям необходимо создание надежного и гибкого корпоративного хранилища, которое станет основой для повышения эффективности принятия управленческих решений. В статье приведен обзор патентной активности, технических и экономических свойств технологии проектирования корпоративных хранилищ данных.

Ключевые слова: корпоративное хранилище данных, Data Vault 2.0, управленческие решения, крупные промышленные компании.

Введение

В настоящее время цифровая трансформация играет ключевую роль в развитии бизнеса, проблема эффективного управления данными и принятия на их основе управленческих решений становится все более актуальной. С каждым годом объемы данных растут, а скорость их обработки увеличивается, что предъявляет новые требования к системам хранения и обработки информации. Таким образом, корпоративные хранилища данных становятся неотъемлемой частью успешного бизнеса. В условиях постоянно растущих объемов и сложности данных крупным промышленным компаниям необходимы современные методы проектирования хранилищ, которые обеспечивают масштабируемость, адаптивность и высокое качество данных.

Основными причинами необходимости разработки и внедрения корпоративного хранилища данных являются: во-первых, большое количество источников данных (как внешних, так и внутренних) и пользователей данных; во-вторых, накопление исторических данных, которые необходимо анализировать и на их основе строить прогнозы; в-третьих, необходимость соблюдения конфиденциальности и защиты [1,2].

Для успешного внедрения корпоративного хранилища данных необходимо проанализировать цели и задачи, которые необходимо решить бизнесу, и, в зависимости от полученных результатов, выбрать наиболее подходящую методологию проектирования для повышения эффективности принятия управленческих решений.

1. Методология

В статье использовался метод анализа публикаций для анализа и систематизации передового зарубежного опыта и современной отечественной практики в области применения хранилищ данных применительно к специфике деятельности крупных промышленных компаний.

Разработка и внедрение, а также преимущества и недостатки корпоративных хранилищ данных исследовались в работах зарубежных и отечественных ученых [1,3-6].

Кроме того, в статье применен патентный анализ в области проектирования корпоративных (аналитических) хранилищ данных.

2. Результаты

Корпоративное хранилище данных (Data Warehouse – DWH) – это единое и структурированное представление данных, предназначенное для сбора и аналитической обработки данных для поддержки принятия управленческих решений [7].

Поскольку патентный анализ позволяет оценить текущий уровень и тренды технологического развития, в работе проведен анализ патентной активности по странам за 2013-2023 гг.

2.1. Патентный анализ технологии корпоративного хранилища данных

Для анализа сформирован запрос и проведен патентный поиск в системе «Orbit Intelligence» за 2013-2023 гг., в результате которого было найдено 2853 патента. Анализ данных проведен в приложении Orange Data mining.

На рисунке 1 представлена диаграмма со странами публикации патентов в области технологии хранилище данных на оси ординат и с годами выпуска исследуемого объекта на оси абсцисс.

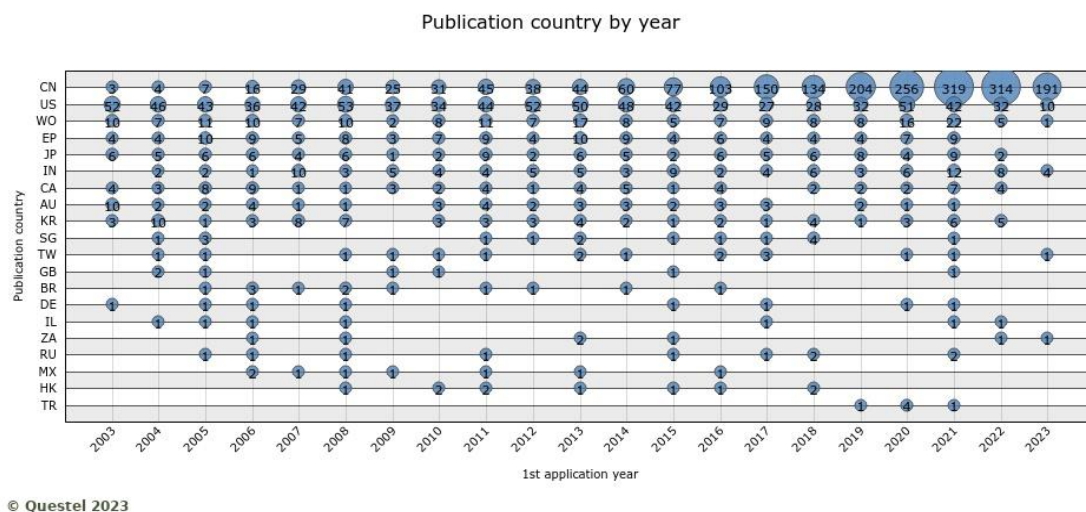


Рис. 1. Патентная активность с учетом стран выдачи патентов

На основе полученной диаграммы можно сделать вывод о том, что активное развитие технологии наблюдается с 2018 года. Для объяснения такой тенденции существуют несколько причин: во-первых, рост объемов данных, требующих масштабируемых и эффективных систем хранения; во-вторых, в последние годы требования к данным стали более динамичными, что предполагает гибкость и скорость обработки; в-третьих, облачные технологии стали более доступными и популярными, что побудило разработчиков и инженеров сфокусироваться на создании более совершенных и эффективных облачных хранилищ данных. Также, за последние годы появились новые технологии и инструменты, позволяющие эффективно хранить, обрабатывать и анализировать данные. Пиковый показатель патентной активности был достигнут в 2021 году, незначительное снижение наблюдалось в следующем году. В настоящий момент показатель соизмерим с аналогичным показателем 2019 года.

На рисунке видно, что Китай является лидером по патентованию в сфере технологии хранилищ данных, в США данная сфера развивается чуть менее активно, но тоже на достаточно высоком уровне. Для более точного анализа приведена столбчатая гистограмма с проиллюстрированным количеством патентов в каждой стране (рис. 2).

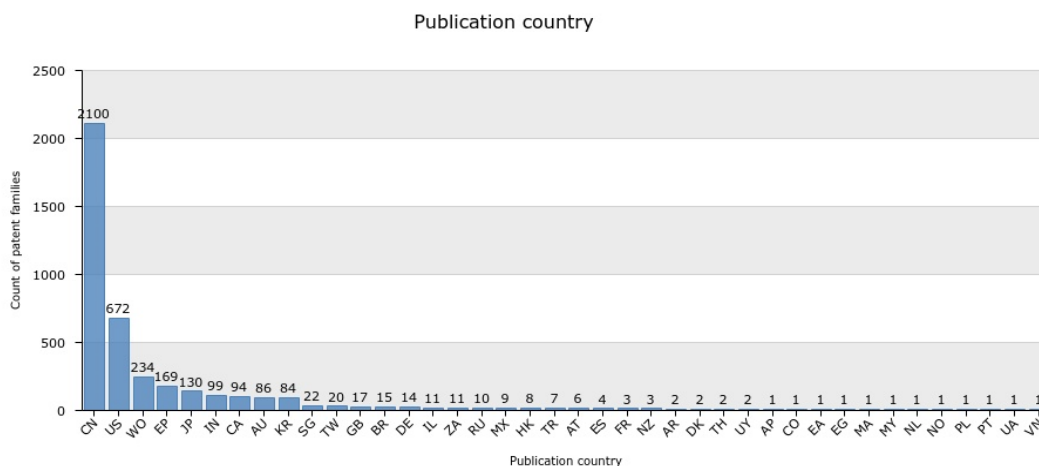


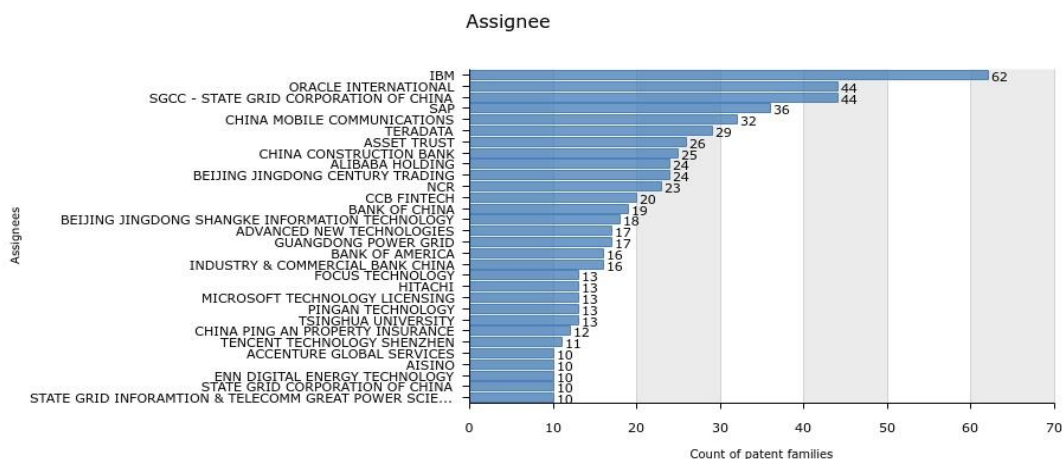
Рис. 2. Гистограмма географической структуры патентования технологии хранилище данных

Китай является лидером в развитии технологии хранилищ данных по нескольким причинам: во-первых, Китай имеет самое большое население в мире, что создает огромный спрос на хранение и обработку данных. Благодаря своему большому рынку и высокому уровню потребительской активности, китайские компании могут инвестировать крупные суммы в разработку и инновации в

области хранилищ данных. Во-вторых, политика правительства активно поддерживает развитие сектора информационных технологий и хранилищ данных, включая создание специальных экономических зон, налоговые льготы, стимулирование инвестиций и другие меры. В-третьих, Китай имеет большое число квалифицированных специалистов в области информационных технологий. Активная система образования и исследовательских программ, а также высокие инвестиции в области науки и технологий позволяют Китаю привлекать и развивать талантливых специалистов в этой сфере.

Несмотря на текущую ситуацию, российские компании активно ищут замену иностранным решениям в области хранения данных, рассматривая отечественные разработки, которые обеспечивают необходимый уровень функциональности и надежности.

Далее была проведена выборка по заявителям (рис. 3).



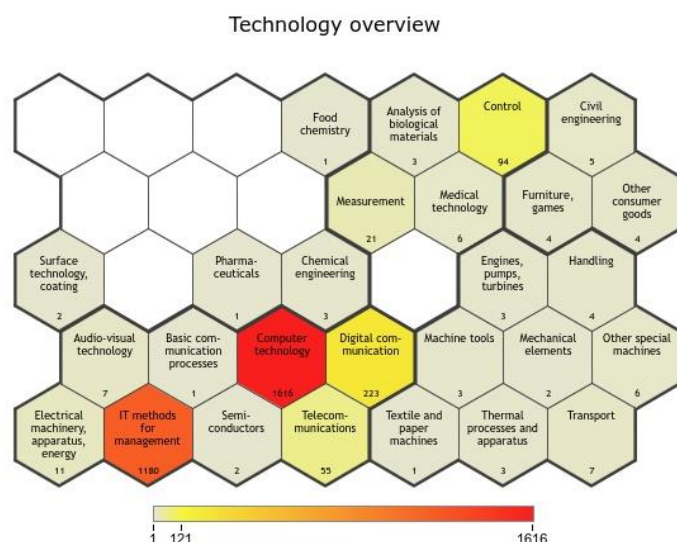
© Questel 2023

Рис. 3. Гистограмма структуры выборки по заявителям

Из полученных данных видно, что наибольшую активность демонстрируют крупнейшие ИТ-компании США и Китая. Компании IBM, Oracle, SAP, Teradata являются крупными мировыми поставщиками программного обеспечения и решений в области информационных технологий. Они имеют широкий спектр продуктов и услуг, которые включают в себя системы управления базами данных, аналитические инструменты, облачные решения и т.д. Также, для компаний характерно глобальное присутствие и сотрудничество с крупными корпорациями и организациями по всему миру.

State Grid Corporation of China (Государственная электросетевая корпорация Китая) является одним из лидеров по количеству патентов как в Китае, так и в мире. Компания стала интеллектуальным монополистом, в значительной степени опирающимся на национальную инновационную систему Китая – в частности, на государственные исследовательские организации и государственное финансирование, а также на инновационную и энергетическую политику [8].

Ниже представлена структура выборки по технологическим направлениям (рис. 4). Каждый шестиугольник представляет технологическое направление с указанным количеством патентов в сфере технологии хранилищ данных.

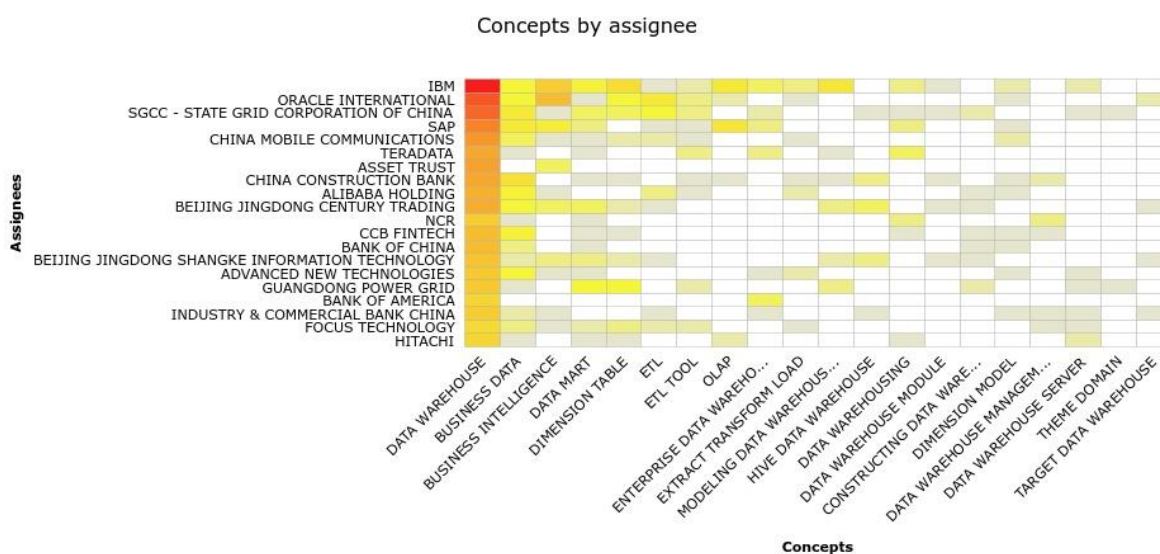


© Questel 2023

Рис. 4. Структура выборки по технологическим направлениям

Наибольшее количество патентов представлено в сфере компьютерных технологий, IT-инструментов для управления (менеджмента), цифровых коммуникаций. Существуют патенты для различных отраслей, например, транспорта, медицины, фармацевтики, аудиовизуальных технологий.

На рисунке 5 показана диаграмма ключевых слов по патентообладателям.



© Questel 2023

Рис. 5. Диаграмма ключевых слов по патентообладателям

Исходя из представленных данных можно сделать вывод о том, что патенты компаний в исследуемой области были связаны с сферой хранилищ данных, бизнес-данных, BI-систем, в меньшей степени – витрины данных, таблиц измерений.

Таким образом, анализ тенденций патентования в сфере технологии хранилище данных позволяет сделать вывод о значимости данной области, интенсивном развитии и конкуренции между ведущими игроками.

2.2. Обзор технических свойств технологии проектирования корпоративного хранилища данных по методологии Data Vault 2.0

В рамках данной работы рассматривается методология проектирования корпоративного хранилища данных Data Vault 2.0, которая предлагает целостный подход к проектированию, разработке и развертыванию хранилищ данных [9]. Подход Data Vault был изобретен Дэном Линстедтом для Министерства Обороны США в 2000 г. В 2013 году данная технология была доработана с учетом

новых вызовов и требований и представлена версия Data Vault 2.0 [10]. Модель Data Vault 2.0 применяется в IT-сфере, телекоммуникациях, финансовом и нефтегазовом секторе. В условиях цифровой трансформации особое значение приобретает внедрение данной модели в крупных промышленных компаниях.

Выбор методологии проектирования влияет на структуру, организацию и доступность данных в хранилище данных, а также на усилия и время, необходимые для его разработки и внедрения. В конечном итоге, выбор методологии влияет на итоговый результат – качество аналитических отчетов, позволяющих обоснованно принимать управленческие решения и реализовать концепцию Data Driven Decision Making («решения на основе данных») [11]. Существует три наиболее популярных модели проектирования корпоративного хранилища данных: Снежинка, Data Vault 2.0, Anchor Modeling [12]. Каждая из моделей имеет свои преимущества и недостатки. На сегодняшний день технология проектирования по методологии Data Vault 2.0 превосходит другие модели за счет следующих особенностей:

- в условиях ограниченности ресурсов затраты на разработку и доработку сравнительно низкие;
- высокая скорость доступа к данным;
- гибкость и простота моделирования;
- простой доступ для анализа данных;
- прослеживаемость во времени изменения хранимой информации за счет временных меток;
- масштабирование бизнеса без перепроектирования IT-решения для соответствия новым требованиям.

Структура Data Vault 2.0 предусматривает длительное хранение больших объемов разнородной информации в виде трех сущностей: хаб, связь и сателлит [13].

Внедрение новой архитектуры для реализации хранилища затрагивает изменения структуры данных, включая их организацию, связи и форматы хранения. Одним из основных объектов изменений является процесс объединения данных в хранилище, настройка интеграций с различными системами, например, ERP или CRM и прочих источников, обеспечение их согласованности, определение прав доступа и разграничения видимости данных.

Архитектура корпоративного хранилища данных состоит из нескольких уровней:

- область сбора первичных данных, где объединяются данные из разных отделов компании, разных источников;
- ядро, в котором весь поступивший в область сбора объем разрозненных данных приводится к нужным структурам и ключам, именно этот ключевой компонент хранилища обеспечивает целостность и полноту данных;
- первичные (помогают решать относительно простые задачи) и вторичные (для сложных аналитических отчетов и нетиповых задач) аналитические витрины, построенные на основе структурированных данных, преобразованных в массивы, которые удобно использовать для решения конкретных задач;
- сервисный слой, отвечающий за управление тремя предыдущими уровнями и обеспечивающий мониторинг процессов в хранилище данных, оперативное выявление и устранение технических ошибок.

Следует отметить, что существует тесная связь между хранилищами данных и BI-системами (Business intelligence), которые используют данные для анализа, отчетности и принятия решений [14].

Таким образом, корпоративное хранилище данных хранит агрегированную информацию, относящуюся ко всей компании, ко всем бизнес-единицам. Это центральное хранилище данных в информационной системе предприятия, которое служит основой для принятия управленческих решений.

2.3. Обзор экономических свойств технологии проектирования корпоративного хранилища данных по методологии Data Vault 2.0

В условиях внедрения новых технологий, стремительного нарастания объемов данных и возникновения потребности в быстрой обработке и длительного хранения внедрение корпоративного хранилища данных позволяет достичь этих целей и принести значительные экономические выгоды для компании. Грамотное применение методологии Data Vault 2.0 при проектировании корпоративного хранилища данных приводит к снижению затрат на обработку и анализ данных благодаря более эффективной структуре хранилища. Например, компания, использующая Data Vault 2.0, может сократить затраты на обработку данных примерно на 20% за счет высвобождения временного ресурса [2].

Если рассматривать экономические выгоды от внедрения корпоративного хранилища данных на трех уровнях управления, то можно привести следующие результаты. На оперативном уровне снижаются затраты на обработку данных, расходы на персонал, наблюдается значительная экономия времени. На тактическом уровне снижаются затраты на перепроектирование бизнес-процессов, что позволяет изменить подходы к управлению компанией и, как следствие, повысить прибыль компании. Благодаря улучшенной аналитике данных компания может увеличить выручку за счет точного прогнозирования спроса и адаптации к рыночным трендам. На стратегическом уровне экономические выгоды заключаются в увеличении рентабельности инвестиций и выручки от новых сегментов рынка, повышении стоимости крупной промышленной компании и др. К качественным характеристикам экономических выгод также можно отнести изменение стратегии компании с учетом аналитических отчетов и новые подходы к ведению бизнеса. В качестве примера можно привести компанию «Аэрофлот», в которой внедрение хранилища данных позволило не только прогнозировать спрос на определенную продукцию, но и осуществить мониторинг программы инновационного развития, как это указано на сайте группы «Аэрофлот» [15]

Крупные промышленные компании, такие как ПАО «СИБУР Холдинг» и ПАО «Газпром нефть», используют корпоративные хранилища данных по методологии Data Vault 2.0 и обращают внимание, что данная технология позволяет качественно изменить процесс принятия управленческих решений [16,17].

Как заявлено на сайте компании КРОК внедрение корпоративного хранилища данных позволяет освободить ИТ-персонал от рутинных задач и сократить затраты на обслуживание. При этом затраты на обслуживание сокращаются до 3 раз [18]. По опыту реализованных проектов финансовая выгода от оптимизации бизнес-процессов на основе анализа больших данных составляет до 60% [18].

Эффективное управление данными и аналитика, обеспечиваемые корпоративным хранилищем, способствуют принятию более обоснованных и качественных решений. Анализ данных в хранилище по методологии Data Vault 2.0 позволяет выявить потенциальные риски и угрозы заранее, что помогает компании принимать меры по их снижению или предотвращению. Возможность применения self-service инструментов за счет высокой структурированности данных позволяет повысить оперативность обработки гипотез и сократить процесс, который обычно занимает несколько дней, до нескольких минут. Более достоверная и своевременная информация, которую обеспечивает Data Vault 2.0, улучшает процессы принятия решений на всех уровнях промышленной компании.

Основные экономические цели, которые могут быть достигнуты посредством внедрения корпоративного хранилища данных, представлены в таблице 1.

Таблица 1. Экономические цели, которые достигаются посредством внедрения корпоративного хранилища данных по методологии Data Vault 2.0

Экономические цели	Основание для реализации цели
Повышение обоснованности принятия управленческих решений	Глубокая бизнес-аналитика
	Построение отчетов и аналитических дашбордов
	Алгоритмы машинного обучения
	Интеграция с системами бизнес-аналитики и управления данными
	Установление закономерностей и тенденций экономических процессов
	Выявление сильных и слабых сторон компании
	Построение сложных многофакторных моделей
Оптимизация процессов загрузки и обработки данных	Аудит
	Поддержка параллельной обработки и запросов
	Предварительная обработка данных и индексация
	Автоматизация процессов очистки и обновления данных
Повышение производительности бизнеса	Увеличение доступности и скорости доступа к данным
	Оптимизация процессов обработки и анализа данных
	Снижение времени реакции на запросы и улучшение отзывчивости системы
	Обнаружение «узких» мест в бизнес-процессах
Повышение конкурентоспособности компании	Предоставление актуальной и достоверной информации для принятия стратегических решений
	Выявление новых возможностей и угроз на рынке
	Прогнозирование спроса
	Оценка эффективности бизнес-стратегий и их корректировка
Снижение затрат на хранение данных	Архивирование и сжатие данных
	Управление жизненным циклом данных

Экономические цели	Основание для реализации цели
Увеличение прибыльности	Дедупликация данных
	Компрессия и оптимизация хранения
	Оптимизация ценообразования и управление затратами
	Выявление наиболее прибыльных продуктов, сегментов рынка, клиентов и каналов сбыта
	Прогнозирование финансовых показателей и планирование бюджета

Источник: разработано автором

Применение корпоративного хранилища данных позволит достичь поставленных экономических целей и повысить обоснованность принятия управленческих решений.

3. Заключение

Таким образом, основной целью применения технологии проектирования корпоративного хранилища данных по методологии Data Vault 2.0 является повышение обоснованности принятия управленческих решений в крупных промышленных компаниях. Ключевыми преимуществами Data Vault 2.0 являются гибкость, масштабируемость и эффективность. Применение корпоративного хранилища данных позволяет получить существенные экономические выгоды в условиях ограниченности ресурсов с одной стороны, и стремительного растущего объема данных с другой стороны. Модульная конструкция Data Vault 2.0 позволяет легко добавлять новые атрибуты или предметные области без изменения существующей модели. Использование Data Vault 2.0 позволяет сохранить все версии данных, что обеспечивает лучшую прослеживаемость, упрощает аудит и способствует снижению расходов на поддержание и развитие системы.

Применение методологии Data Vault 2.0 способствует повышению качества данных и экономии времени, что в свою очередь ведет к развитию крупных промышленных компаний.

Литература

1. *Сынкова А. О.* Разработка корпоративного хранилища данных: преимущества и подходы к проектированию // Актуальные проблемы прикладной математики, информатики и механики: сб. трудов Междунар. науч. конф – Воронеж: Научно-исследовательские публикации, 2023. – С. 916-918.
2. Что такое хранилище данных? [Электронный ресурс]. Официальный сайт онлайн платформы Astera.com. URL: <https://www.astera.com/ru/type/blog/data-warehouse-definition> (дата обращения: 17.05.2024).
3. *Салий В. В., Кухаренко Л. В., Ищенко О. В.* Цифровая трансформация экономики и внедрение хранилищ данных на основе больших данных в инфраструктуру компании // Вестник Академии знаний. 2021. №. 3 (44). С. 208-214.
4. *Некрасов А. А., Гаврилов С. О., Беленькая М. Н.* Средства создания хранилищ данных // Телекоммуникации и информационные технологии. 2021. Т. 8. №. 1. С. 75-80
5. *Turken G. et al.* Research on Creating a Data Warehouse Based on E-Commerce // 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST). IEEE, 2023. P. 16-20.
6. *Maheshwari N.* A Practical Guide to Data Warehousing in a Post-Big Data World. — Berkeley (CA): Apress, 2019. 352 p.
7. Что такое Data Warehouse и как оно помогает экономить бизнесу [Электронный ресурс]. Официальный сайт IT-компании ООО «Клевер Солюшнс». URL: <https://www.qlever.ru/services/dwh>. (дата обращения: 17.05.2024)
8. *Rikap C.* Becoming an intellectual monopoly by relying on the national innovation system: the State Grid Corporation of China's experience // Research Policy. 2022. Vol. 51. №. 4. P. 104472
9. *Колесникова П. Е.* Business intelligence и хранилище данных // Информационно-телекоммуникационные системы и технологии. 2020. С. 129-130.
10. Что такое Data Vault: моделирование КХД для архитектора Big Data [Электронный ресурс]. Официальный сайт Школы Больших Данных (ООО «Учебный центр «Коммерсант»). URL: <https://bigdataschool.ru/blog/what-is-data-vault.html> (дата обращения: 17.05.2024).
11. *Чернышева Ю.Г.* Бизнес-анализ - новые возможности аналитики для управления // Учет и статистика. 2022. №1 (65). С. 69-76. URL: <https://cyberleninka.ru/article/n/biznes-analiz-novye-vozmozhnosti-analitiki-dlya-upravleniya> (дата обращения: 20.05.2024).
12. Снежинка, Data Vault, Anchor Modeling. Какая методология проектирования DWH подойдет для вашего бизнеса? [Электронный ресурс]. Официальный сайт IT-сообщества Habr.com. URL: <https://habr.com/ru/articles/786822> (дата обращения: 17.05.2024).
13. Data Vault [Электронный ресурс]. Официальный сайт Публичной облачной платформы Yandex Cloud. URL: <https://yandex.cloud/ru/docs/glossary/datavault>. (дата обращения 22.05.2024).
14. *Гриненко Ю.К., Матов М.Б.* Особенности применения ВІ-технологий в условиях развития цифровой экономики // Московский экономический журнал. 2022. №. 6. С. 670-682.

15. Внедрение хранилищ данных и Data Lake [Электронный ресурс]. Официальный сайт Группы компаний «РАМАКС». URL: <https://www.ramax.ru/directions/bi-systems/data-lake> (дата обращения: 22.05.2024).
16. Особенности DATA-DRIVEN в нефтехимии [Электронный ресурс]. Официальный сайт ИТ-кластера в контуре нефтехимической группы СИБУР ДИДЖИТАЛ. URL: <https://sibur.digital/84-osobennosti-data-driven-v-neftekhimii> (дата обращения: 22.05.2024).
17. Управление данными в «Газпром нефть» [Электронный ресурс]. URL: <https://gsbhsefiles.ru/cfo/m6b/%D0%9C%D0%B0%D1%81%D0%BB%D0%BE%D0%B2.pdf> (дата обращения: 22.05.2024).
18. Корпоративное хранилище данных [Электронный ресурс]. Официальный сайт компании КРОК Облачные сервисы. URL: <https://cloud.croc.ru/cloud-services/database> (дата обращения: 22.05.2024).