

10.07.2025

Современные корпоративные хранилища данных

Концепции (warehouses, lakes),
решения (базы, платформы), процессы

techart.ru | promo.techart.ru

Регистраций: 73 (компаний – 59), в т.ч.:
- 46.6 % новых слушателей
- 31 топ-менеджер
- 21 постоянный (4+) слушатель программы
- 14 с вопросами и обратной связью



Ключевые компетенции «Текарт»

Консалтинг

- стратегия и стратегирование
- диагностика, бизнес-девелопмент
- трансформация
- цифровизация
- маркетинг
- организационное развитие
- бизнес-математика

Бизнес-аналитика

Интегрированный маркетинг

Дизайн

IT-решения и веб-разработка

Оргразвитие

Создание контента

25

лет на рынке бизнес-аналитики, маркетинга и digital

20 000

выполненных проектов

39 000+

клиентов и партнеров в экосистеме «Текарт»

4 место

рейтинга «Маркетинговый консалтинг» RAEX

160

сотрудников в штате

ТОП 100

в рейтинге работодателей Headhunter*

* среди компаний до 250 сотрудников



Программа непрерывного образования для руководителей по маркетингу

Основные треки программы:

Маркетинг (комплексное продвижение, SMM, SEO и т. д.), реклама

Martech, аналитика (технологии, данные, автоматизация, цифровые сервисы)

Дизайн, веб-разработка, контент, фото, видео

Маркетинговый анализ и стратегия

Организационное развитие

2

года работы

7 000+

участников из 1 900 компаний

31

спикер

550

пользователей [платформы](#)

План вебинара

1. Актуальные вызовы российских компаний: «зоопарк систем», ручные отчеты, недоверие к данным.
2. Этапы зрелости: от «электронных таблиц» к единому корпоративному хранилищу.
3. Ключевые концепции построения хранилищ данных.
4. Data Lakes vs DWH — не конкуренция, а синергия.
5. Российские решения и платформы для создания корпоративного хранилища данных.

«Идеальный мир»

04



«Идеальный мир»

Детализация по сотрудникам

| Сотрудник | План, ч. | Факт, ч. | % выполнения | Абсолютное выполнение плана, ч. | Планируемое распределение нагрузки, ч. | Загруженность, % | Отклонение от загрузки, ч. | % фактического выполнения задач |
|-------------------------|----------|----------|--------------|---------------------------------|--|------------------|----------------------------|---------------------------------|
| Итоги | 715,6 | 509,0 | 71% | -207 | 800 | | | 56% |
| + Сотрудник не назначен | 8,0 | 0 | 0 | ▼ -8 | 0 | 0% | 0 ● | 0% |
| + Сотрудник 94 | 32,0 | 54,0 | 169% | ▲ 22 | 80 | 100% | -48 ● | 75% |
| + Сотрудник 81 | 52,5 | 34,4 | 66% | ▼ -18 | 80 | 100% | -28 ● | 66% |
| + Сотрудник 80 | 85,0 | 56,0 | 66% | ▼ -29 | 80 | 100% | 5 ● | 63% |
| + Сотрудник 8 | 2,0 | 2,0 | 100% | ✓ 0 | 0 | 0% | 0 ● | 100% |
| + Сотрудник 67 | 1,5 | 1,0 | 67% | ▼ -1 | 40 | 50% | -39 ● | 100% |
| + Сотрудник 66 | 12,0 | 4,5 | 38% | ▼ -8 | 0 | 0% | 0 ● | 40% |
| + Сотрудник 62 | 22,5 | 22,5 | 100% | ✓ 0 | 0 | 0% | 0 ● | 85% |
| + Сотрудник 6 | 61,0 | 43,0 | 70% | ▼ -18 | 80 | 100% | -19 ● | 72% |
| + Сотрудник 48 | 81,1 | 58,6 | 72% | ▼ -23 | 80 | 100% | 1 ● | 95% |
| + Сотрудник 43 | 80,0 | 37,0 | 46% | ▼ -43 | 80 | 100% | 0 ● | 90% |

2024/Спринт 25

09.12.24 - 22.12.24

Обновлено 17.12.20:38

Последние данные 17.12.19:57

Утилизация

0%

0.0% ▼ к п.м.

Продуктивность

72%

0.0% ▼ к п.м.

Выполнение плана, часов

509,716

71% / 100%



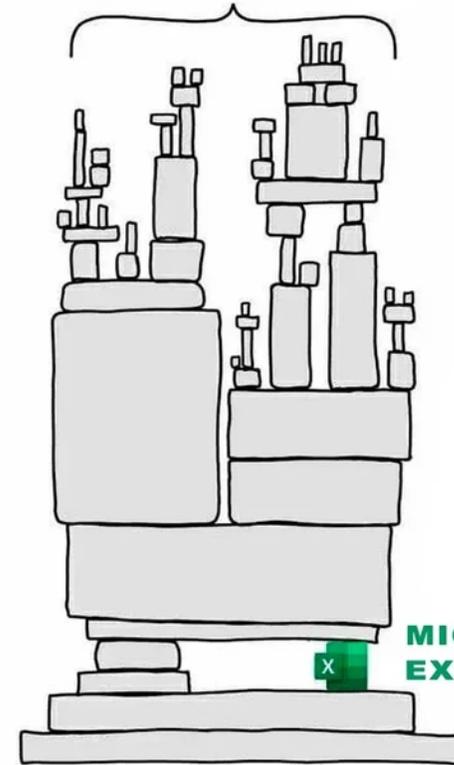
Детализация продуктивности, часов



Суровая реальность

- Отчеты собираются вручную из десятка систем.
- Данные маркетинга и данные по продажам изолированы.
- У разных сотрудников/отделов — разная отчетность.
- Недостаток информации для принятия решений.
- Недоверие к данным.
- Различия в НСИ.
- Неэффективные маркетинговые бюджеты, медленная реакция на рынок, упущенные возможности...

ВСЯ МИРОВАЯ ФИНАНСОВАЯ СИСТЕМА



От хаоса к упорядоченности

1. «Время героев».

Каждый сам формирует отчеты.

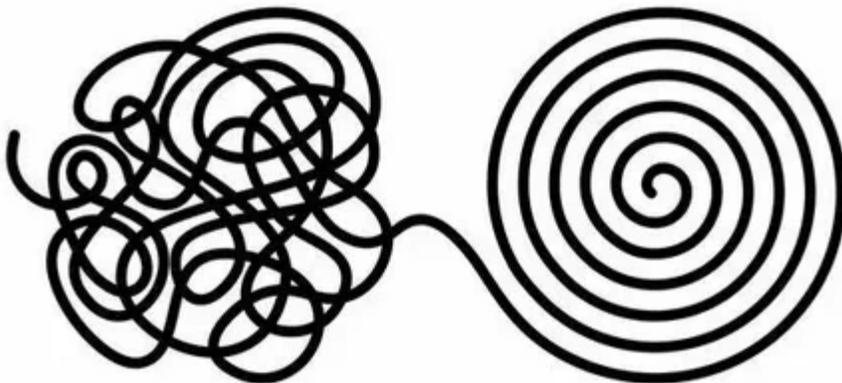
2. «Информационные острова».

Стихийно создаются автоматизированные сценарии формирования отчетов и хранения/обработки данных.

3. «Первая версия КХД от IT».

Наработки из п.2 объединяются в одну систему, без пересмотра самих сценариев.

4. «Полноценное хранилище».



Что такое хранилище данных?

Корпоративное хранилище данных (КХД, DWH) — централизованная система хранения данных для построения бизнес-аналитики и аналитических отчётов.



Предпосылки к внедрению хранилищ

- Большое количество источников данных.
- Не устраивает своевременность получения информации.
- Доработка имеющихся отчетов долгая/дорогая/сложная.
- Несовпадающие данные в разных отчетах.
- Ошибочные данные в отчетах.
- Путаница в показателях.
- Отчеты тормозят.
- Сложности с безопасностью и доступом.
- Неожиданно все может перестать работать.
- Не получается добиться нужной логики работы отчета.

Ключевые функции КХД

- Сбор, упорядочивание и анализ данных из разных систем в одном месте для их централизованного использования.
- Обеспечение единого источника аналитической информации для создания отчетов, проведения бизнес-анализа и принятия обоснованных управленческих решений на основе целостных данных.
- Обработка больших объемов разнообразных данных: расчет показателей, агрегация и фильтрация.
- Быстрое создание отчетов и информационных панелей.
- Оперативный мониторинг процессов в различных подразделениях, филиалах и центрах финансовой ответственности компании через единую точку доступа к данным.

Хранилище и база данных — есть ли разница?

КХД — это **не** большая база данных.

В отличие от традиционной базы данных, хранилище данных (DWH) обычно имеет трехуровневую архитектуру:

1 уровень — одна или несколько баз с данными, извлеченными из различных источников (например, транзакционные БД).

2 уровень — содержит сервер, который преобразует данные в структуру, подходящую для последующего анализа и выполнения сложных запросов.

3 уровень — приложения для анализа, визуализации и создания отчетов.

Данные перемещаются с уровня на уровень, трансформируются и предстают в виде согласованной информации, пригодной для анализа.

Хранилище и база данных — есть ли разница?

| | База данных | Корпоративное хранилище данных |
|--------------------|--|---|
| Цель использования | Используется для хранения текущих данных, которые часто обновляются и запрашиваются | Используется для хранения исторических данных и выполнения сложных аналитических запросов |
| Источники данных | Обычно используется для функционирования определенной системы (сайт, CRM и пр.) | Включает данные из разных источников |
| Объем данных | Обычно хранит меньший объем данных, так как фокусируется на текущих операциях | Хранит большие объемы исторических данных за длительные периоды времени |
| Производительность | Оптимизирована для высокой производительности при выполнении транзакционных операций | Оптимизировано для выполнения сложных запросов на больших наборах данных |
| Пользователи | Операционные сотрудники, разработчики, администраторы баз данных | Аналитики, дата-сайентисты, руководители |

Архитектура хранилища

- **Источники данных.**

Уровень, на котором собираются первичные данные: CRM- и ERP-системы, веб-сайт, биллинговая система и другие источники.

- **Хранилище.**

Вся поступившая в DWH разрозненная информация структурируется и приводится к нужному виду.

- **Витрины.**

На этом уровне выполняется преобразование массива данных в удобную для дальнейшего анализа структуру.

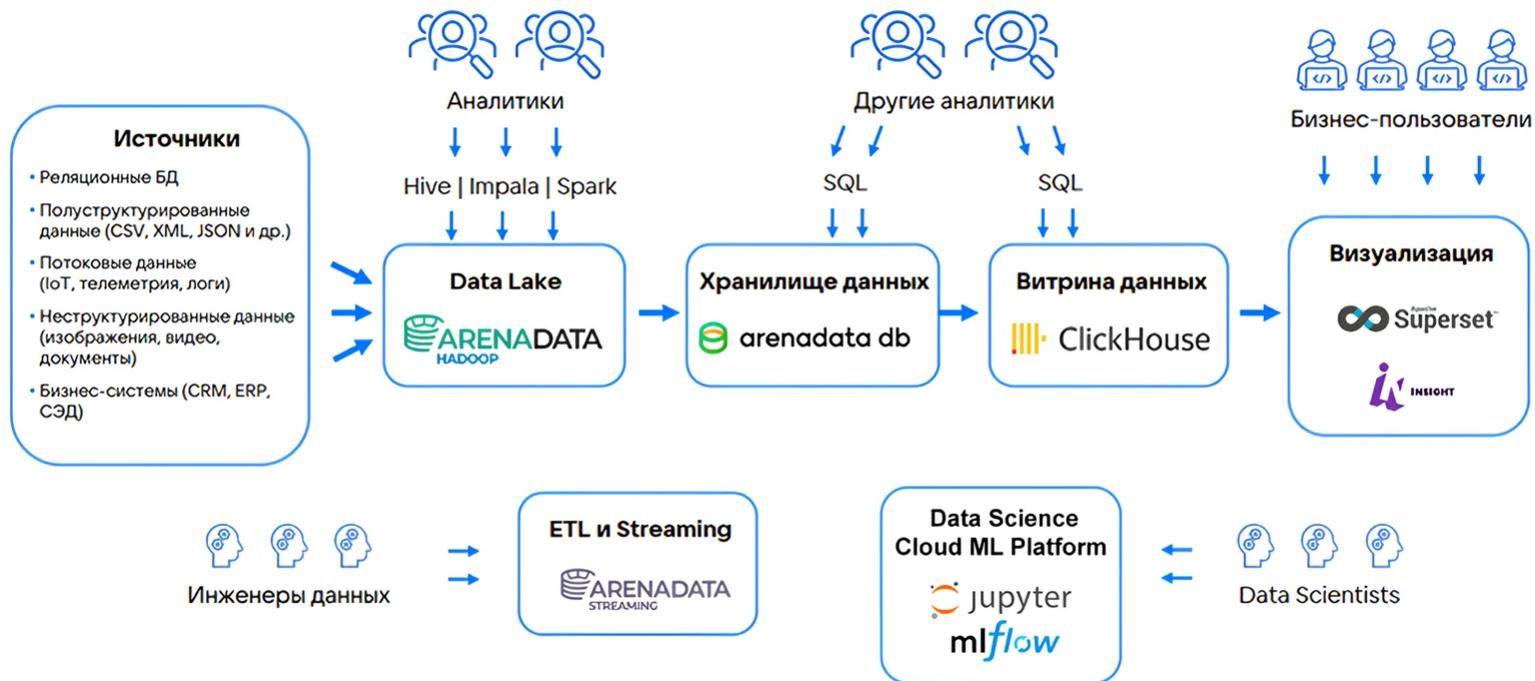
- **Сервисный уровень.**

Необходим для управления тремя предыдущими уровнями.

- **Интерфейс для бизнес-пользователей.**

Этот уровень агрегирует данные из витрин и хранилищ, приводя их к более простому и доступному виду. Пользователи выстраивают аналитику и работают с дашбордами и графиками, доступ к которым различается в зависимости от прав.

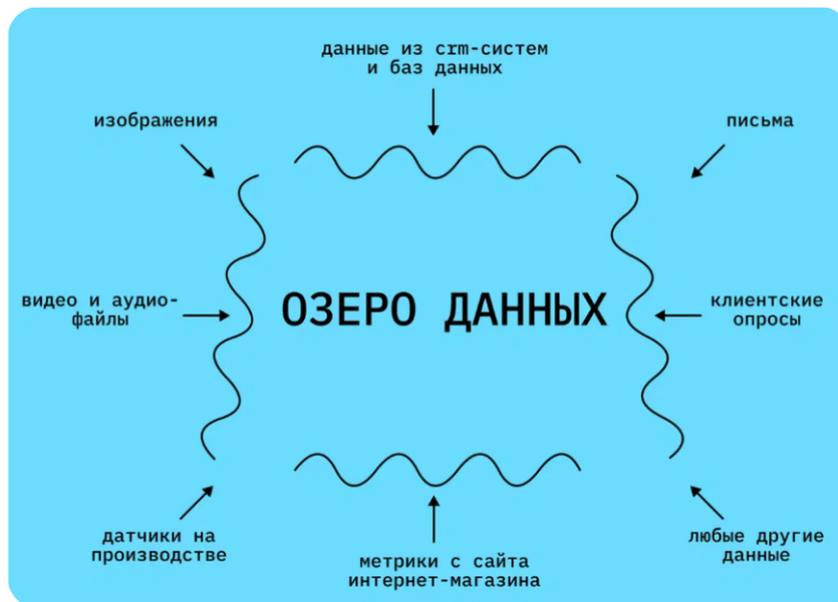
Архитектура хранилища



Озера данных (Data Lakes)

В Data Lake данные поступают и хранятся **в необработанном виде** без какой-либо упорядоченной структуры.

Подобного рода хранилища полезны, когда надо обрабатывать и анализировать данные из различных внешних источников, которые проблематично обработать и положить в таблицы.



Озера данных (Data Lakes)



Озеро данных можно рассматривать как шаблон проектирования — формализованные рекомендации, которые можно использовать для решения распространенных проблем при разработке инфраструктуры.

Ключевые особенности Data Lake

- Данные необработаны либо слабообработаны.
- Большой срок хранения данных.
- Есть возможность преобразования данных.
- Поддерживаются разные схемы чтения данных.

Кому неактуально

- Небольшой бизнес с минимальной IT-инфраструктурой и небольшим объемом поступающих данных.
- Тем, кто собирает структурированные данные, например, из баз или систем сбора метрик. В таком случае их сразу можно помещать в хранилища и использовать для аналитики.

Озера данных (Data Lakes)



Удобны для сбора, хранения и обработки больших потоков информации, которая поступает непрерывно. Если грамотно их использовать, они станут надежным инструментом для следующих отраслей:

- **Телекоммуникации.**

Озера данных часто применяются для хранения и анализа данных о клиентах, трафике, сетевых устройствах и других факторах, которые влияют на бизнес телекоммуникационных компаний.

- **Нефтегазовая промышленность.**

Data Lake собирают терабайты данных и используют их в прогнозных моделях для разведки месторождений, управления цепочками поставок и техническим обслуживанием.

- **Медицина.**

База данных о пациентах, их диагнозах и способах лечения может быть использована для автоматизации диагностики.

- **Розничная торговля.**

Data lake позволяет хранить и анализировать данные о продажах, клиентах, инвентаре и других факторах, которые важны для розничных компаний.

Озера и хранилища: сравнение

| Область сравнения | Data Lake | Хранилище данных |
|-----------------------|--|--|
| Сбор данных | Данные любой структуры и из любых источников | Данные приведены к единому виду (таблицы в БД) |
| Обработка данных | После сбора | Перед сбором |
| Основные пользователи | Специалисты по глубокому анализу данных | Оперативные пользователи |
| Стоимость хранения | Ниже | Выше |

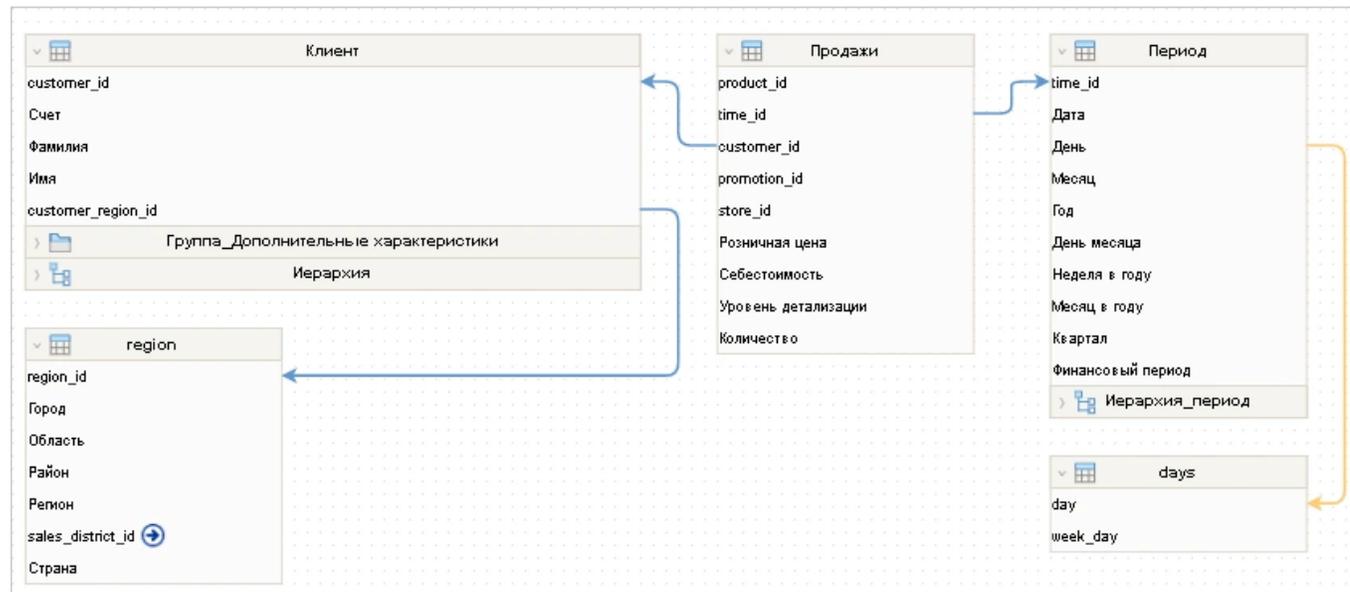
Структурированные данные

| product | date | cross_device_last_significantUTMCampaign | cross_device_last_significantUTMSource | visits | pageviews | bounces | avgVisitDurationSeconds |
|---------|------------|--|--|--------|-----------|---------|-------------------------|
| | 2024-01-01 | | yandexdirect | 4 | 5 | 2 | 35.5 |
| | 2024-01-01 | | yandexdirect | 1 | 1 | 0 | 15.0 |
| | 2024-01-01 | | | 1 | 1 | 0 | 23.0 |
| | 2024-01-01 | | yandexdirect | 13 | 16 | 0 | 20.53846154 |
| | 2024-01-01 | | yandexdirect | 1 | 1 | 1 | 0.0 |
| | 2024-01-01 | | yandexdirect | 1 | 1 | 0 | 15.0 |
| | 2024-01-01 | | yandexdirect | 2 | 2 | 1 | 8.0 |
| | 2024-01-02 | | yandexdirect | 4 | 5 | 0 | 24.0 |
| | 2024-01-02 | | yandexdirect | 1 | 1 | 0 | 15.0 |
| | 2024-01-02 | | yandexdirect | 1 | 1 | 0 | 26.0 |
| | 2024-01-02 | | | 1 | 1 | 0 | 15.0 |
| | 2024-01-02 | | yandexdirect | 15 | 15 | 6 | 9.13333333 |
| | 2024-01-02 | | yandexdirect | 1 | 1 | 1 | 0.0 |
| | 2024-01-02 | | yandexdirect | 2 | 2 | 1 | 8.0 |
| | 2024-01-03 | | yandexdirect | 5 | 5 | 0 | 15.4 |
| | 2024-01-03 | | | 1 | 1 | 0 | 16.0 |
| | 2024-01-03 | | yandexdirect | 17 | 18 | 2 | 109.47058824 |
| | 2024-01-03 | | yandexdirect | 1 | 1 | 1 | 0.0 |
| | 2024-01-03 | | yandexdirect | 1 | 1 | 0 | 16.0 |

Данные из Яндекс Метрики в таблице MS SQL Server

Методологии проектирования DWH

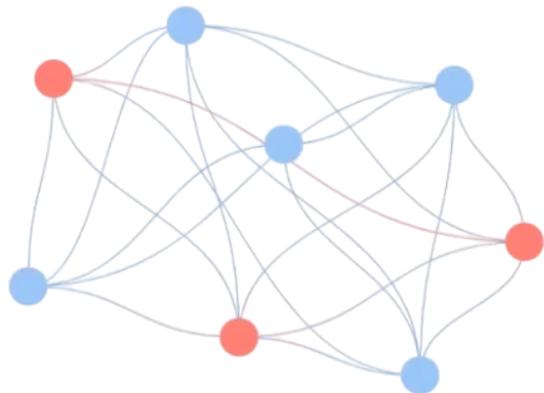
Схема «Снежинка»



Так же используются методологии Data Vault 2.0, Anchor Modeling и пр.

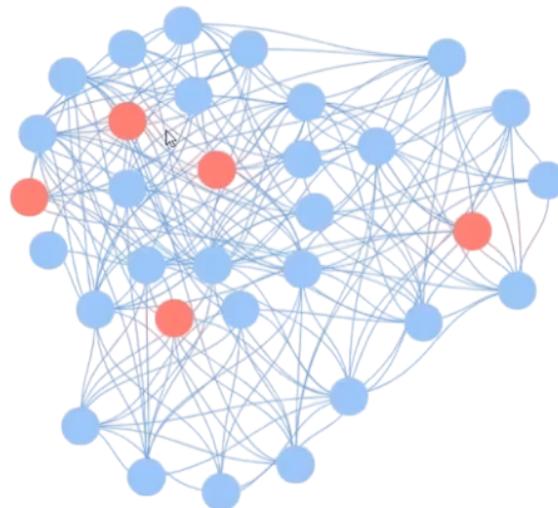
Методологии проектирования DWH

8 таблиц



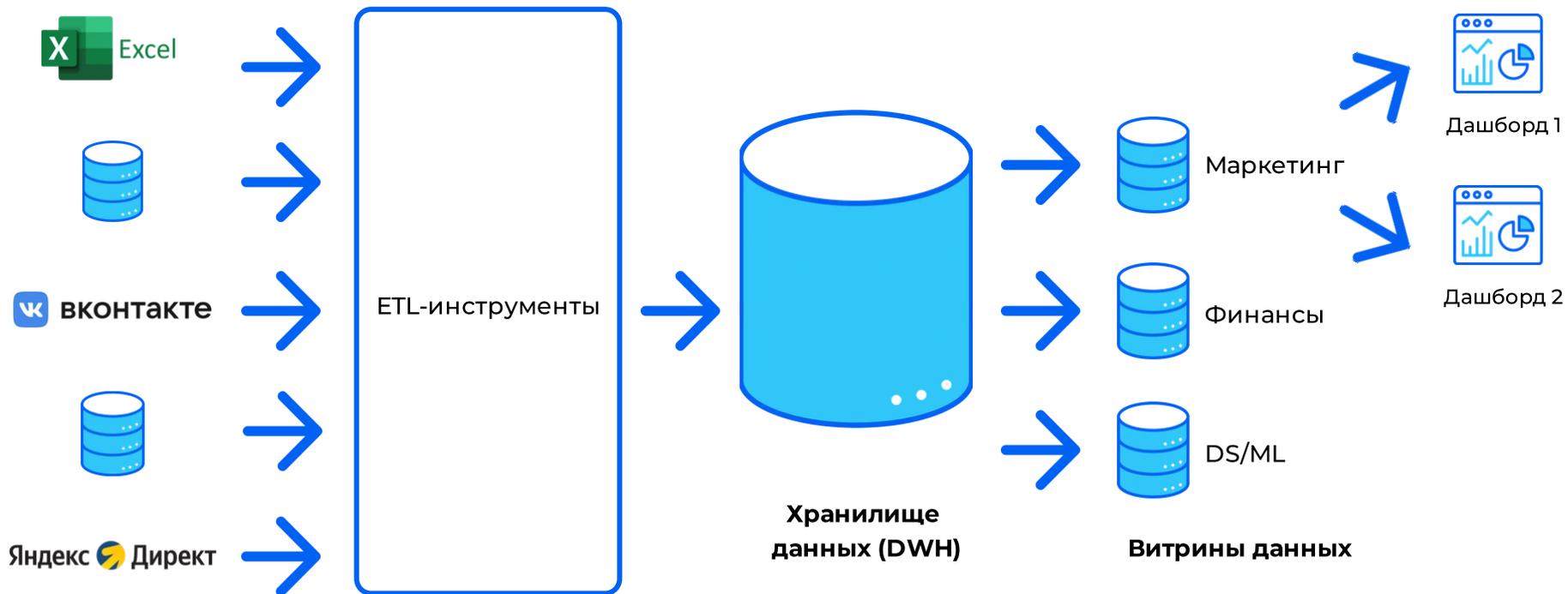
Через год
→

34 таблицы



Витрины данных

Data Mart или **витрина данных** — база данных, предназначенная для решения специализированной задачи или набора задач из одной предметной области. Фокус на конкретные бизнес-процессы: маркетинг, продажи, финансы, HR и т.д.



Витрины данных

Каталог витрин КХД

редактировал(а) sys: Управление данными (автоматическое обновление портала КХД) вчера в 08:07 PM

ВНИМАНИЕ!

Каталог содержит информацию о 2700+ витрин и загружается полностью за 1-1,5 минуты. Просьба не уходить со страницы на время загрузки.

? Узнать больше об инструменте

⚙️ Ролевая модель

Предметная область = Роль доступа = Схема = Витрина = Описание витрины = Описани

| № | Схема | Витрина | Описание витрины | Предметная область | Роль доступа | Поля | Описание полей |
|---|-----------|---|-------------------------|-----------------------------|----------------------|--|--|
| 1 | PRD_VD_DM | V_DAYS TOP-10 in Reporting | Справочник дней | Финансы (учет, PL, бюджеты) | base_role | DAY_ID , DAY_MONTH_STR , DAY_MONTH_NUMBER , DAY_YEAR_STR , DAY_YEAR_NUMBER , DAY_DATE_STR , DAY_OF_WEEK , DAY_OF_WEEK_SHORT_NAME , DAY_OF_WEEK_LONG_NAME , DAY_FIRST_DTM , DAY_LAST_DTM , YEAR_ID , QUARTER_ID , QUARTER_NUMBER , MONTH_ID , MONTH_NUMBER , MN , MONTH_DAYS_COUNT , WEEK_ID , WEEK_ID_2 , WEEK_ID_3 , WEEK_ID_4 , WEEK_YEAR_NUMBER , WEEK_MONTH_NUMBER , WEEK_ISO_YEAR , WEEK_ISO_NUMBER , IS_HOLID , PREV_YEAR_DAY , PREV_YEAR , PREV_MONTH_DAY , PREV_MONTH , PREV_WEEK_DAY , PREV_WEEK_ID , PREV_WEEK_ID_2 , LAST_YEAR_WEEK_ID_2 , PREV_WEEK_ID_3 , PREV_DAY , IS_PREV_HOLID , IS_EASTER | Дата // Порядковый номер в полной текстовой номер квартала) // t-1 месяце // ID Недели Идентификатор недели, а пятница, 1 Номер недели в год Дата минус месяц, как в WEEK_ID2) // И день // Признак явл |
| 2 | PRD_VD_DM | PPMS_ACTION_MASTER | Мастер акций | Маркетинг (Промо) | base_role | PPMS_ACTION_ID , SRC_PPMS_ACTION_ID , ACTION_NUMBER , GUID , PPMS_MARKET_ACTION_ID , ART_ID , FRMT_ID , CNTR , PREPACK_ART_ID , PPMS_HOLIDAY , SUPPLIER_REGULAR_KM_PRICE , FORECAST_CALC , SUPPLIER , PREPAID , BONUS , CREATE , MARKET_ACTION_PROJECT_ART_ID , PPMS_APPLYING_DISCOUNT_ID , BEGIN_SHIPMENT_DT , END_SHIPMENT_DT , BEGIN_DT , END_DT , IS_MANDATORY , FN_ROLE_NAME , PPMS_COMP_MECH_ID , PPMS_ACTIVITY_TYPE_ID , PPMS_ , LMASM_FRMST , ORDER_FORECAST_VOL , IS_ASTER_FORECAST , CATALOG_REGULAR_PRICE , IS_DEVIATION_EXCEEDED , NEXT_PROMO_DT , PPMS_LOGISTIC_APVL_ID , IS_SEND_TO_DMP , IS_PART_OF_DMP , SALE_PRICE_ACTN_LVL1 , DISC_ON_SHELF_ACTN_LVL1 , SALE_PRICE_ACTN_LVL2 | ID акции // ID акции контрагента склада , Промо-линейка // IS акции // Признак Нс поставки // ID Вид: действия цены // От пользователя, внеси отклонения // 1) // Скидка на полк (Прогрессивная - Количество товара ; |
| 3 | PRD_VD_DM | LOG_ACTIVITY_MASTER | Мастер логов | Операции | base_role | LOG_ACTIVITY_ID , DAY_ID , WMS_ID , ART_ID , CNTR_ID , CALC_TIME | |

Витрины данных

24

Страницы / ... / PRD_VD_DM 620 views

PRD_VD_DM ▸ V_DAYS

Создатель: sys: Управление данными (автоматическое обновление портала КХД), отредактировано 29 окт 2022

Справочник дней

🏠 Финансы (учет, PL, бюджеты) 📄

🏆 TOP-10 in Reporting 📄

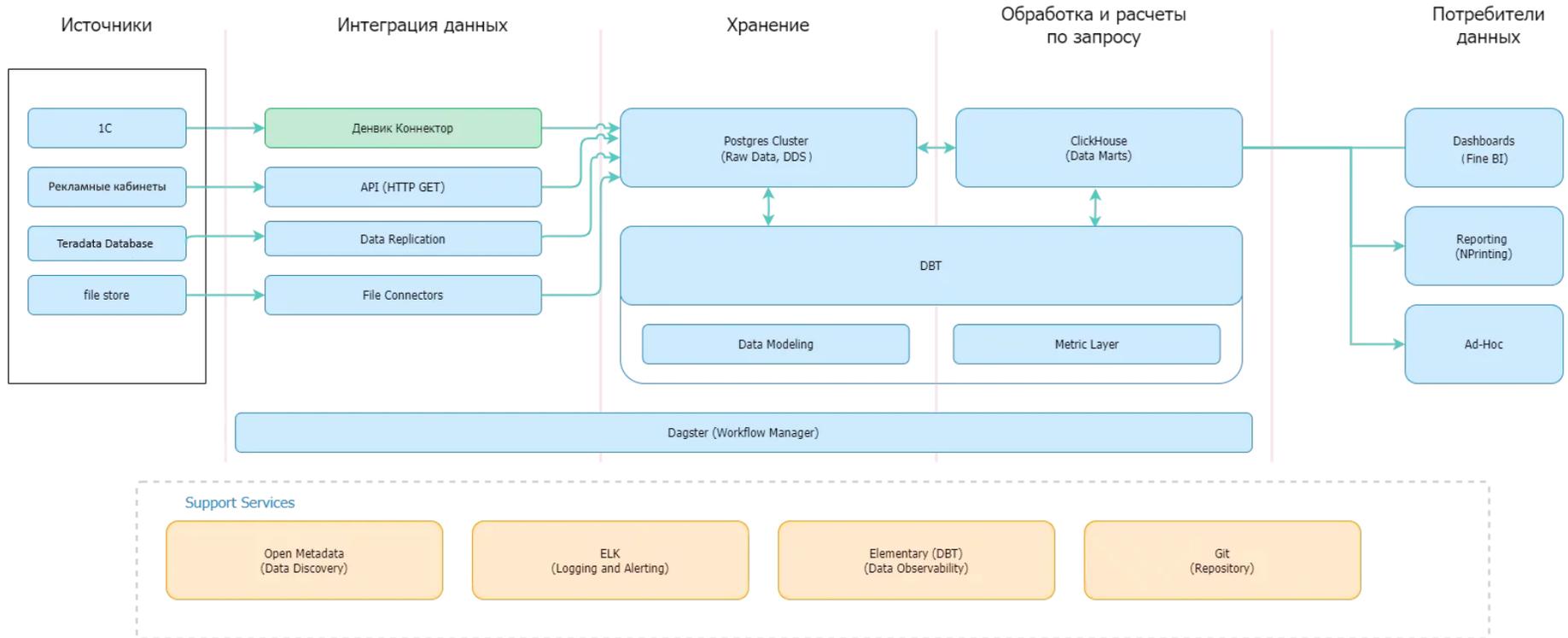
Поля Происхождение данных Использование

Если необходимо найти все витрины, которые содержат заинтересовавшее вас поле, воспользуйтесь [Каталогом витрин КХД](#) или чат-ботом MagnitData на этой странице.

| № | Поле | Описание поля | Тип данных |
|----|---------------------------|--|------------|
| 1 | DAY_ID <small>URI</small> | Дата | DATE |
| 2 | DAY_MONTH_STR | Порядковый номер дня в месяце (строковое поле) | VARCHAR |
| 3 | DAY_MONTH_NUMBER | Порядковый номер дня в месяце (числовое поле) | INTEGER |
| 4 | DAY_YEAR_STR | Порядковый номер дня в году (строковое поле) | VARCHAR |
| 5 | DAY_YEAR_NUMBER | Порядковый номер дня в году (числовое поле) | INTEGER |
| 6 | DAY_DATE_STR | Дата (строковое поле) | VARCHAR |
| 7 | DAY_OF_WEEK | День недели (число) | INTEGER |
| 8 | DAY_OF_WEEK_SHORT_NAME | День недели, в короткой текстовой записи (пн, вт, etc.) | VARCHAR |
| 9 | DAY_OF_WEEK_LONG_NAME | День недели в полной текстовой записи (понедельник, вторник, etc.) | VARCHAR |
| 10 | DAY_FIRST_DTM | Дата / время начала суток | TIMESTAMP |
| 11 | DAY_LAST_DTM | Дата / время конца суток | TIMESTAMP |

Архитектура корпоративного хранилища данных

25



ПО для построения хранилищ

- СУБД



- Оркестраторы



- Брокеры



- Коннекторы



- Инструменты трансформации



- BI



- Вспомогательные инструменты



Какие специалисты нужны для обслуживания КХД?

**Директор по данным
(Chief Data Officer ,CDO)**

Разработка стратегии управления данными компании. Координация работы всех специалистов, связанных с данными, и обеспечение соответствия деятельности стратегическим целям компании.

Архитектор (Chief Architect)

Разработка архитектуры хранилища данных, включая ETL-процессы (Extract, Transform, Load).

Владелец продукта (Product Owner/Data Product Manager)

Определение требований к продуктам на основе данных и управление их разработкой. Установление приоритетов в разработке функций и улучшений на основе потребностей пользователей.

Инженер данных (Data Engineer)

Объединение данных из различных источников в единое хранилище. Настройка и оптимизация баз данных для обеспечения высокой производительности.

**Администратор баз данных
(Database Administrator)**

Управление базами данных, обеспечение их надежности и доступности. Настройка механизмов защиты данных и управление доступом пользователей.

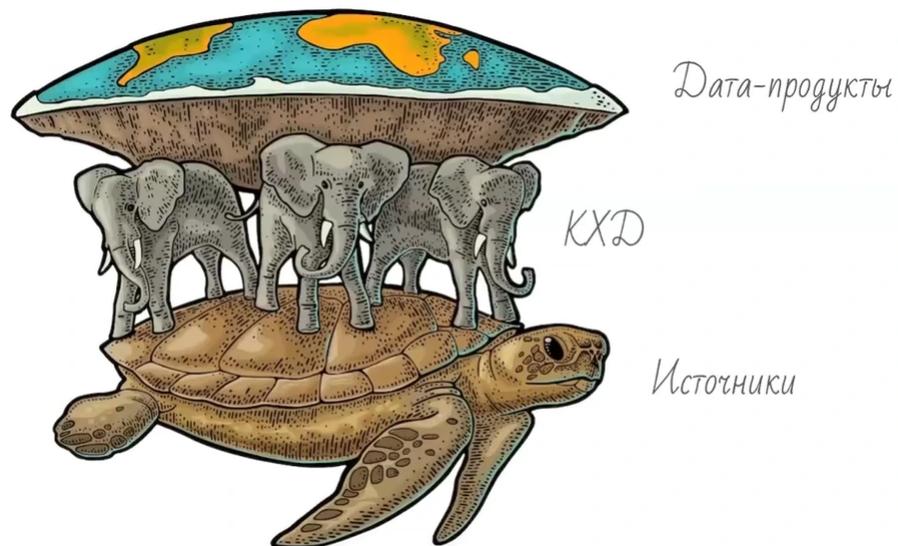
Конечные пользователи:

- Аналитики данных (Data Analyst)
- Бизнес-аналитики (Business Analyst)
- Разработчик BI (Business Intelligence Developer)
- Data Scientists
- Маркетологи и т.д.

Заключение

Грамотное внедрение корпоративного хранилища данных способно сократить затраты на процессы обработки данных:

- Интегрирует процессы работы с данными в одно приложение
- Предоставляет ключевые метрики для бизнеса
- Помогает ускорить обработку сложных запросов
- Обеспечивает анализ данных в режиме реального времени
- Совершенствует процессы принятия решений



Заключение

- **Начните с малого.**

Постройте одну ценную витрину данных (например, «Единый отчет по маркетинговым активностям»).

- **Фокус на бизнес-задачу.**

Решайте конкретную боль (например, «Сократить время сборки ежемесячного маркетингового отчета с 5 дней до 1 часа»).

- **Команда.**

Обязательно привлекайте бизнес-пользователей (маркетинг, аналитика)!
Технические специалисты (ИТ/аналитики данных) – исполнители, а не заказчики.

- **Данные важнее технологий.**

Сначала поймите, какие данные нужны для решения задачи, и где они живут.
Потом выбирайте инструмент.

- **Ключевой совет для маркетологов/аналитиков.**

Четко формулируйте свои потребности в данных и отчетности для ИТ/архитекторов.

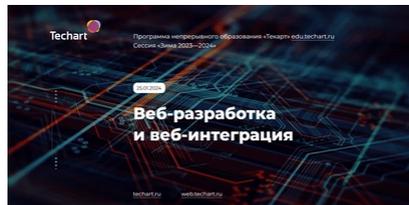
Наши материалы по теме

30



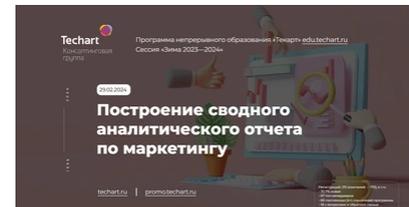
11.07.2027

[Архитектура корпоративных хранилищ данных \(КХД\)](#)



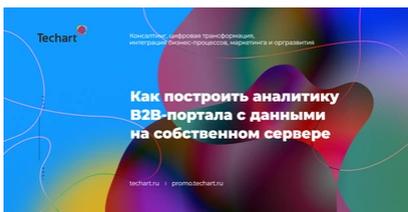
25.01.2024

[Веб-разработка и веб-интеграция](#)



29.02.2024

[Построение сводного аналитического отчета по маркетингу](#)



09.02.2023

[Как построить аналитику B2B-портала с данными на собственном сервере](#)



25.04.2024

[Yandex DataLens: обзор системы визуализации, построение дашбордов](#)

Персональное консультационное мероприятие по теме этого вебинара

Особенности:

- ✓ Будем разбирать именно ваши кейсы
- ✓ Ответим на любые вопросы
- ✓ Подсветим специфику ниши
- ✓ Максимально адаптируем контент под потребности компании
- ✓ Дадим конкретные рекомендации и инструкции
- ✓ Подготовим пошаговое руководство для решения проблемы

Программа может быть реализована в различных вариантах:

- вебинар;
- семинар / лекция;
- мастер-класс;
- стратсессия;
- круглый стол, мозговой штурм;
- диагностическая / консультационная сессия;
- деловая игра.

При необходимости мы можем собрать комбинацию из нескольких удобных вам вариантов обучения.

Форматы:

онлайн, оффлайн, гибрид.

Длительность:

от 1 часа до 2 дней.

Интересуюсь

**Андрей Титенко**

Заместитель директора по маркетингу,
руководитель практики технологического
маркетинга

titenko@techart.ru

Релевантные услуги «Текарт»:

- [Внедрение BI-систем \(business intelligence\)](#)
- [Аудит маркетинговых данных](#)
- [Визуализация данных](#)
- [Интеграция маркетинговых сервисов](#)
- [Веб и мобайл аналитика](#)

+7 495 790 75 91

Маркетинговая группа «Текарт»

techart.ru

Интегрированный маркетинг и PR

promo.techart.ru

IT-решения и веб-разработка

web.techart.ru

Дизайн-бюро

design.techart.ru

Креативное агентство

creative.techart.ru

Фотоагентство

photo.techart.ru

Аналитика и бизнес-планирование

research.techart.ru

Работа в «Текарт»

hr.techart.ru

Образовательная программа

edu.techart.ru**[TechartPulse](#)**

Рабочий пульс: проекты,
кейсы, публикации

**[Techart_ru](#)**

Системное развитие
бизнеса

**[Edu_Teachart](#)**

Непрерывное образование
для директоров по маркетингу