

# CNEWS FORUM КЕЙСЫ ОПЫТ ИТ-ЛИДЕРОВ

КАК ИЗМЕНИЛИСЬ  
ПОДХОДЫ К СБОРУ  
И ХРАНЕНИЮ ДАННЫХ

The logo for CNews Forum Кейсы features a stylized icon of a cluster of dots on the left, followed by the text 'cnews' in a lowercase sans-serif font, and 'FORUM КЕЙСЫ' in a smaller, uppercase sans-serif font below it.

cnews  
FORUM КЕЙСЫ

Никита Титов  
Эксперт по цифровой трансформации  
и управлению данными

# Подходы



## Никита Титов

Эксперт по цифровой трансформации  
и управлению данными, промышленный  
сектор

Telegram: @nikita\_ai

**Качество данных** – критический фактор успешной реализации ИИ и аналитических решений

**Нет управления данными** → нет бизнес-эффекта.  
Data Governance обязателен

**Источник новых данных** – это описание бизнес-процессов компании

# Содержание



- Текущее состояние
- Куда движется рынок?
- Кейс
- Выводы
- Приложения

—

— Текущее состояние

# Данные стали основным производственным ресурсом



## Раньше

- Автоматизация учета
- Поддержка отчетности
- Источник исторической информации

## Сегодня

- Основа для принятия решений
- «Топливо» для ИИ
- Источник конкурентных преимуществ
- База для цифровых продуктов

Компании больше не собирают данные ради отчетов — они собирают данные ради действий и автоматизации решений

# Подходы к хранению данных

|                 | DWH                               | Data Lake                                  | Data Mesh  |
|-----------------|-----------------------------------|--|--|
| Архитектура     | Централизованная                  | Централизованная                           | Децентрализованная                                       |
| Кто отвечает?   | Единая команда DWH / ETL-инженеры | Единая команда Data Lake / Инженеры данных | Профильные доменные команды (бизнес-юниты)               |
| Гибкость данных | Низкая (только по модели)         | Максимальная (сырые данные)                | Зависит от домена, но в целом высокая                    |
| Основная цель   | Бизнес-аналитика, BI-отчеты       | Машинное обучение, Big Data, архивы        | Масштабирование управления данными в крупных корпорациях |

# Как изменился сбор данных

---

## Раньше

Периодическая загрузка:

- Раз в сутки
- Раз в неделю
- Пакетная обработка

## Сейчас

Потоковая обработка:

- Онлайн
- Событийная модель
- Реальное время

# Роль Data Governance



## Раньше

- Контроль доступа

## Сейчас

- Политика данных
- Качество данных
- Владельцы данных
- Каталогизация
- Data Lineage
- Соответствие требованиям регуляторов

# Текущее состояние

---

- Уровни зрелости управления данными на рынке сильно различаются: от устоявшихся практик до первых шагов формирования работы с данными
- Акцент на real-time: потоковая обработка данных приобретает приоритет над периодической загрузкой
- Управление данными выделилось в самостоятельную функцию корпоративного управления и стало необходимым условием цифровой трансформации бизнеса

---

**Куда движется рынок?**

# От хранения данных к цифровому представлению организации

За последние 30 лет компании прошли путь:

**Базы данных → DWH → Data Lake → Data Mesh → Data Platform**

Следующий этап уже формируется сегодня:

**Data Platform → Digital Twin → AI-Native Enterprise**

Главное изменение заключается не в способах хранения данных. Оно заключается в том, что предприятия начинают формировать цифровое представление своих знаний, процессов, активов и взаимосвязей

Эта среда становится фундаментом для ИИ-трансформации, цифровых двойников компаний и появления интеллектуальных систем поддержки управленческих решений

# Кейс: Цифровой двойник департамента



Цель создать управляемую цифровую систему, где процессы, данные, показатели и знания сотрудников объединены в единый цифровой двойник, отражающий текущее состояние организации и помогающий прогнозировать последствия решений и усиливать работу сотрудников с помощью ИИ

## Знания

- Регламенты, политики, методологии и инструкции
- Экспертиза сотрудников и накопленный опыт
- База решений, практик и типовых сценариев

## Модель

- Структура департамента, роли и зоны ответственности
- Процессы, проекты, ресурсы и взаимосвязи между ними
- Цели, показатели эффективности и правила управления

# Задачи при разработке цифрового двойника департамента

Создать цифровой двойник не столько технологическая, сколько организационная задача

**Задача №1.** Знания находятся в головах сотрудников

- Критически важные знания часто не описаны в документах и системах
- Многие решения принимаются на основе опыта, неформальных правил и экспертных договоренностей
- Потеря эксперта приводит к потере части организационного знания

Для создания цифрового двойника необходимо извлекать, структурировать и формализовывать знания организации

**Задача №2.** Описываем предприятие как единую систему

- Процессы, проекты, сотрудники, данные, ресурсы и показатели существуют не изолированно
- Изменение одного элемента влияет на другие элементы организации
- Описание отдельных процессов не позволяет понять работу компании целиком

Цифровой двойник требует системного подхода к модели бизнеса, отражающего объекты организации и взаимосвязи между ними

Цифровой двойник организации начинается не с данных и не с ИИ. Он начинается с формализации знаний и построения целостной модели бизнеса как единой системы.

# Возвращение Knowledge Management в эпоху ИИ

## Почему тема снова стала актуальной

На протяжении многих лет компании инвестировали в:

- ERP
- BI
- Data Warehouse
- Data Lake

Но значительная часть корпоративной экспертизы продолжала оставаться:

- В документах
- В презентациях
- В переписке
- в головах сотрудников

В результате данные накапливались быстрее, чем знания становились доступными

# Источники и сбор данных

## Источник данных

Сотрудники



## ETL

Бизнес аналитики

Специалисты по организационному дизайну



Человек – абсолютный источник и главный инструмент извлечения данных. Эффективная архитектура данных строится на принципах передачи и фиксации знаний от человека к человеку

# Хранение данных

## Для цифрового двойника необходимы графовые и векторные БД

### Ограничения классических реляционных БД

Количество взаимосвязей в оргструктуре часто превышает количество самих объектов

В реляционной модели такие связи становятся сложными и плохо масштабируются

### Роль графовых БД

Графовые БД позволяют хранить:

#### Узлы

- сотрудники
- подразделения
- проекты
- программы
- продукты
- активы

#### Ребра

- подчиняется
- участвует
- финансирует
- зависит
- поставляет
- использует

### Графовые БД позволяют:

- строить цифровой граф предприятия

На практике граф становится основой цифрового двойника организации

### Роль векторных БД

Большая часть корпоративных знаний находится в неструктурированном виде:

- документы
- регламенты
- письма
- протоколы
- отчеты
- презентации

### Векторные БД позволяют:

- находить скрытые связи
- искать знания по смыслу
- связывать документы с процессами и проектами
- формировать контекст для ИИ-агентов

# Что изменил генеративный ИИ



Генеративный ИИ требует не только данные, но и контекст

**Для получения качественных результатов необходимо:**

- структурировать корпоративные знания
- каталогизировать документы
- связывать знания с процессами и объектами предприятия
- обеспечивать актуальность информации

# Выводы



## Про данные

- Корпоративные хранилища данных создаются для анализа прошлого
- Платформы данных создаются для управления настоящим
- Цифровые двойники компаний создаются для моделирования будущего

## Про процессы

- Смена парадигмы: цифровизация перестает быть исключительно технологической задачей и становится вопросом организационного дизайна
- Человекоцентричность: эксперт-сотрудник возвращается в роль главного источника данных и передачи знаний
- Новая цель: предприятия начинают формировать единое цифровое представление активов, процессов и взаимосвязей
- Ренессанс управления знаниями: Управленческий опыт и процессы получают приоритетное развитие как фундамент для работы ИИ



# Приложения

— Текущее состояние

# Что изменилось за 30 лет

| Было            | Стало               |
|-----------------|---------------------|
| SQL БД          | Data Mesh           |
| ETL             | ELT                 |
| Batch           | Streaming           |
| Отчеты          | Ai                  |
| Таблицы         | Документы + Векторы |
| Централизация   | Децентрализация     |
| DWH             | Data Platform       |
| Пользователь BI | Ai Agent            |

# DWH / Data Warehouse

## (Традиционные хранилища данных)

---

**Суть:** централизованная база данных, оптимизированная под аналитику и отчетность

**Подход:** «Схема при записи» (*Schema-on-write*). Данные перед загрузкой тщательно очищаются, трансформируются и структурируются.

**Тип данных:** строго структурированные (реляционные таблицы)

### Плюсы

Высокая скорость стандартных запросов, надежность, идеальная отчетность

### Минусы

Сложно и дорого масштабировать, не подходит для неструктурированных данных (видео, логи, аудио)

# Data Lake (Озеро данных)

---

**Суть:** огромный централизованный репозиторий, хранящий информацию в исходном виде

**Подход:** «Схема при чтении» (*Schema-on-read*). Данные сваливаются «как есть», а структуру им придают только в момент использования

**Тип данных:** любые — структурированные, полуструктурированные и неструктурированные

## Плюсы

Дешевое хранение Big Data,  
гибкость, идеальная среда для Data  
Science и машинного обучения

## Минусы

Без должного порядка может превратиться  
в «болото данных» (Data Swamp), сложный  
поиск нужных данных для рядовых бизнес-  
пользователей

# Data Mesh



**Суть:** организационный и архитектурный фреймворк, который полностью отказывается от централизации

**Подход:** децентрализация и доменное владение. Каждое бизнес-подразделение (например, логистика, маркетинг, финансы) само отвечает за свои данные и предоставляет их как готовый продукт (*Data as a Product*)

**Тип данных:** распределенные данные любой структуры, объединенные общими стандартами управления

## Плюсы

Масштабируемость для гигантских корпораций, устранение узких мест в виде единой команды инженеров, высокая скорость вывода фич

## Минусы

Высокий порог входа, риск дублирования усилий, сложность внедрения для небольших компаний

# Цифровые двойники и Ai Native

# Digital Twin of Organization (DTO) и архитектура нового поколения

**Digital Twin of an Organization (DTO)** – это динамическая цифровая модель организации, объединяющая данные о процессах, сотрудниках, проектах, ресурсах, финансах и клиентах для понимания работы компании, прогнозирования последствий решений и повышения эффективности управления

Возникает архитектура:

**Data Lake (Data Mash) → Graph DB → Vector DB → AI Agents → Digital Twin**

Где:

- Data Lake (Data Mash) хранит факты
- Graph DB хранит связи
- Vector DB хранит знания
- AI интерпретирует происходящее
- DTO становится цифровым отражением компании

# Ai Native



**Компания, в которой искусственный интеллект встроен в основу бизнеса, а не является отдельным инструментом**

## **AI участвует в каждом бизнес-процессе**

От принятия решений до выполнения операционных задач

## **Цифровой двойник компании**

Структура, процессы, данные, регламенты и знания представлены в машиночитаемом виде

## **Knowledge Management как стратегический актив**

Знания сотрудников извлекаются, структурируются и становятся доступны AI-системам

## **AI-агенты как цифровые сотрудники**

Выполняют аналитические, операционные и управленческие функции наряду с людьми

## **Управление на основе данных и прогнозов**

Решения принимаются не по отчетам прошлого периода, а на основе прогнозных моделей и симуляций

## **Непрерывная автоматизация**

AI самостоятельно выявляет узкие места, предлагает улучшения и запускает изменения

## **Единая корпоративная память**

Информация не теряется при смене сотрудников и доступна всей организации через AI-интерфейсы

## **Человек + AI как новая модель работы**

Каждый сотрудник использует персонального AI-помощника для повышения производительности и качества решений