Лекция ВШЭ · 23 мая 2026

Основы платформ данных

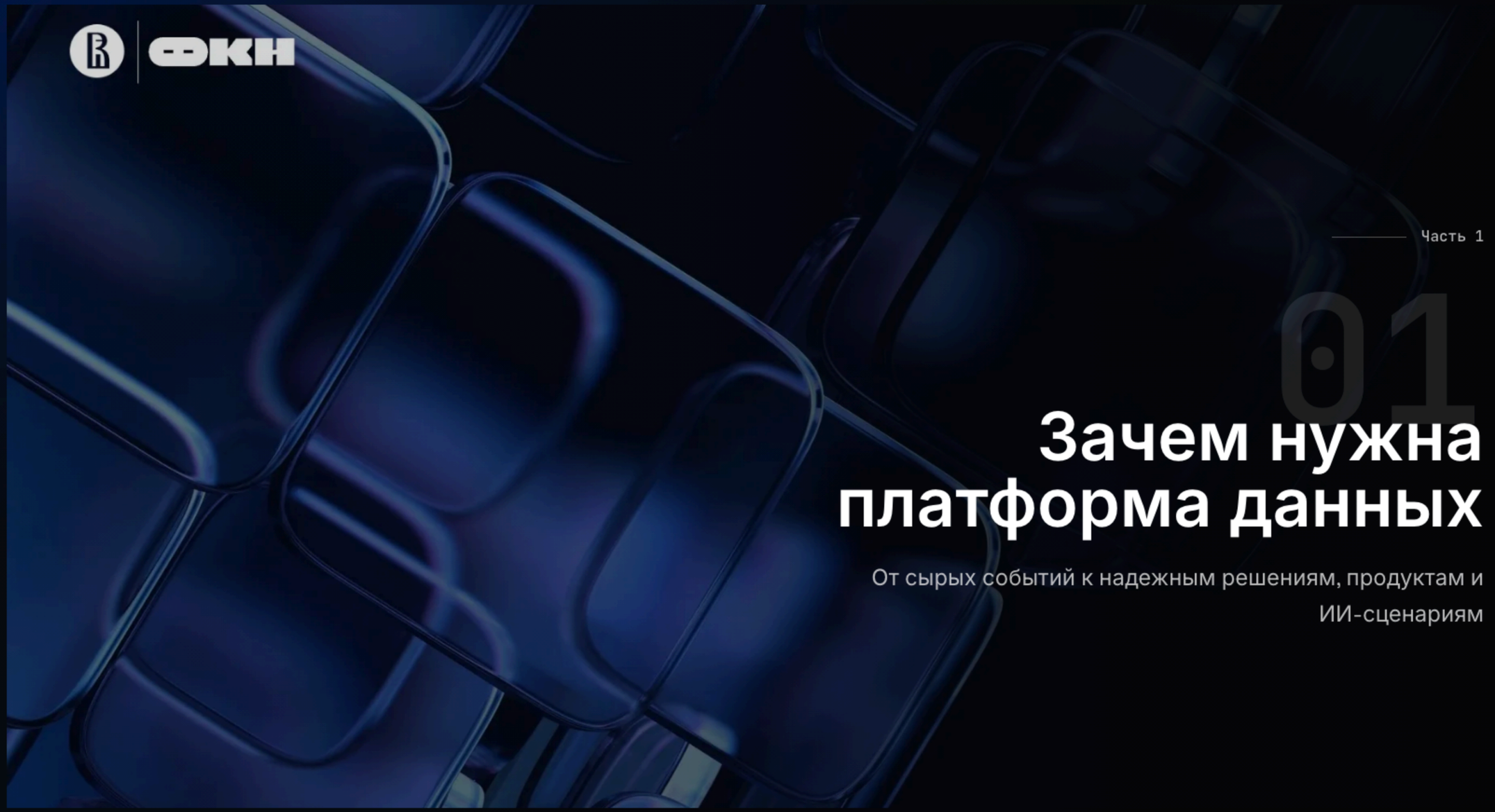
Базовое введение: зачем нужна платформа данных, как устроены ETL/ELT-контуры и как измерять зрелость

Александр Поломодов

Спикер Контекст

Technical Director & Fellow, T-Технологии

- Архитектура и RnD в инженерии.
- AI в SDLC на масштабе.
- Фокус: платформа данных как продукт.



Часть 1

01

Зачем нужна платформа данных

От сырых событий к надежным решениям, продуктам и ИИ-сценариям

Данные становятся производственным контуром

Скорость

От источника до витрины, API, модели или отчёта

Доверие

Владелец, качество, происхождение, доступ и свежесть

Масштаб

Один контур: аналитика, продукты, ML и ИИ

Скорость, доверие и масштаб — свойства всего контура

Проблема Без платформы

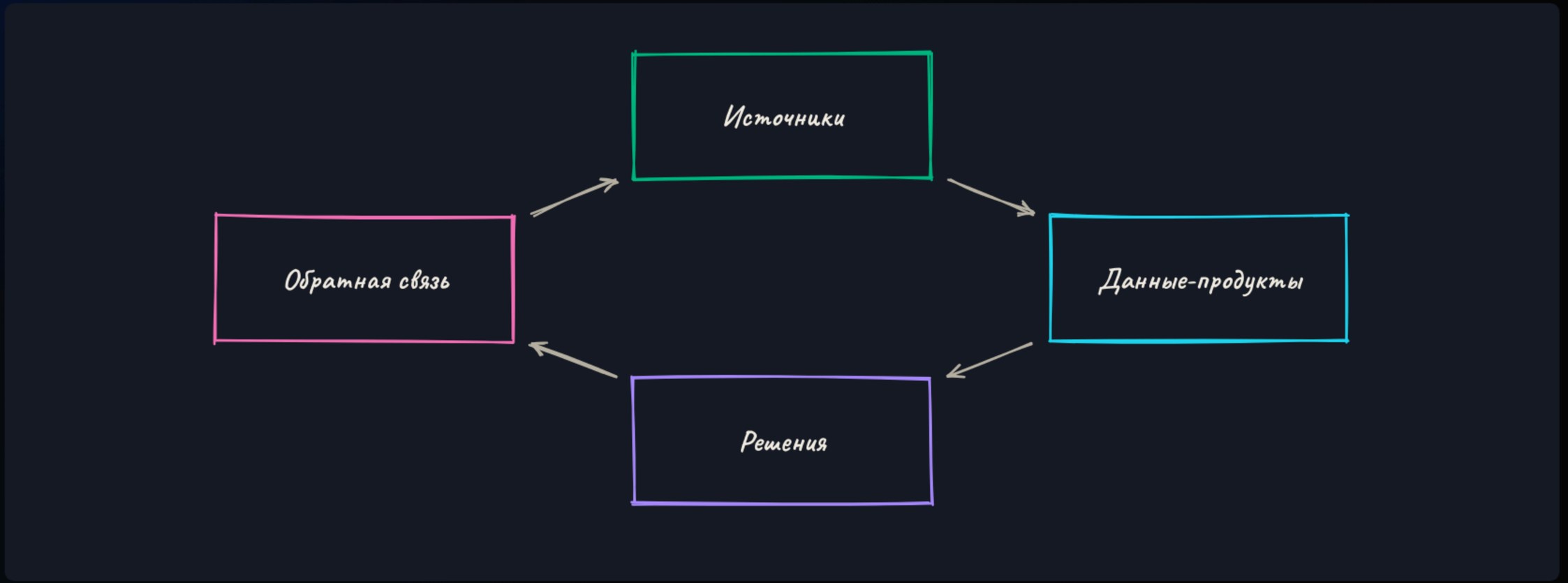
Хранилища и скрипты перестают работать

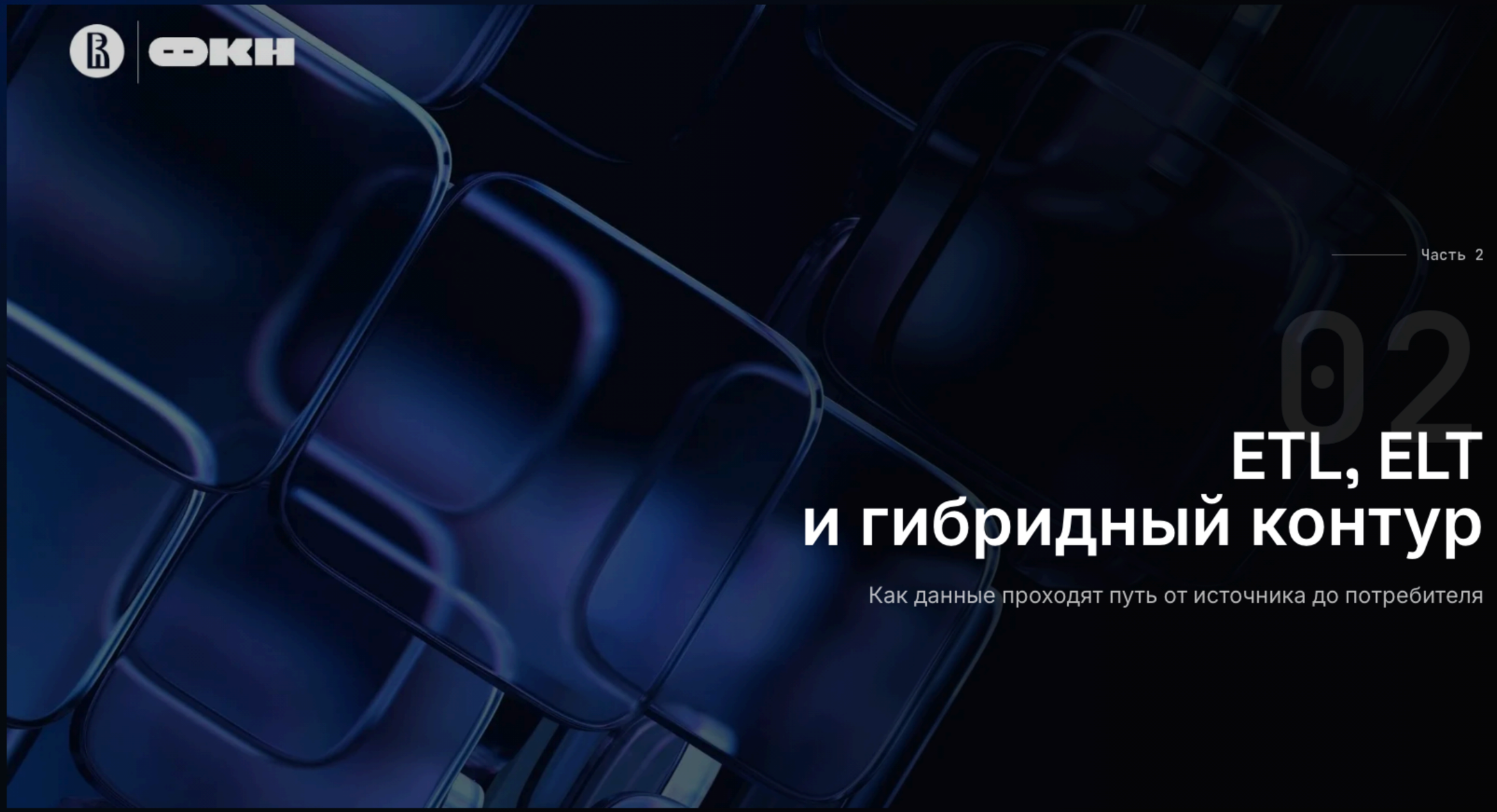
На масштабе ломается ответственность

- Команды заново ищут источники.
- Бизнес-логика прячется в SQL.
- Инциденты данных видны слишком поздно.
- Стоимость не связана с ценностью.

Платформа связывает данные с решениями

Контур | Контур ценности





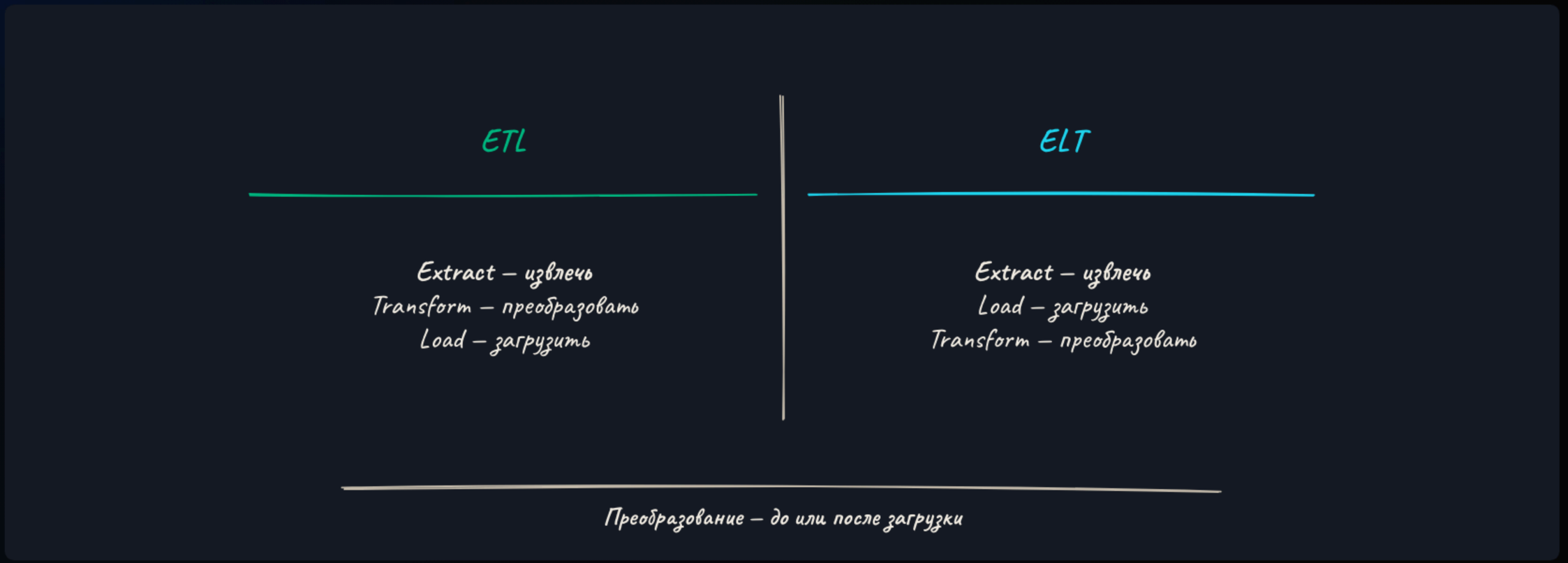
Часть 2

02

ETL, ELT и гибридный контур

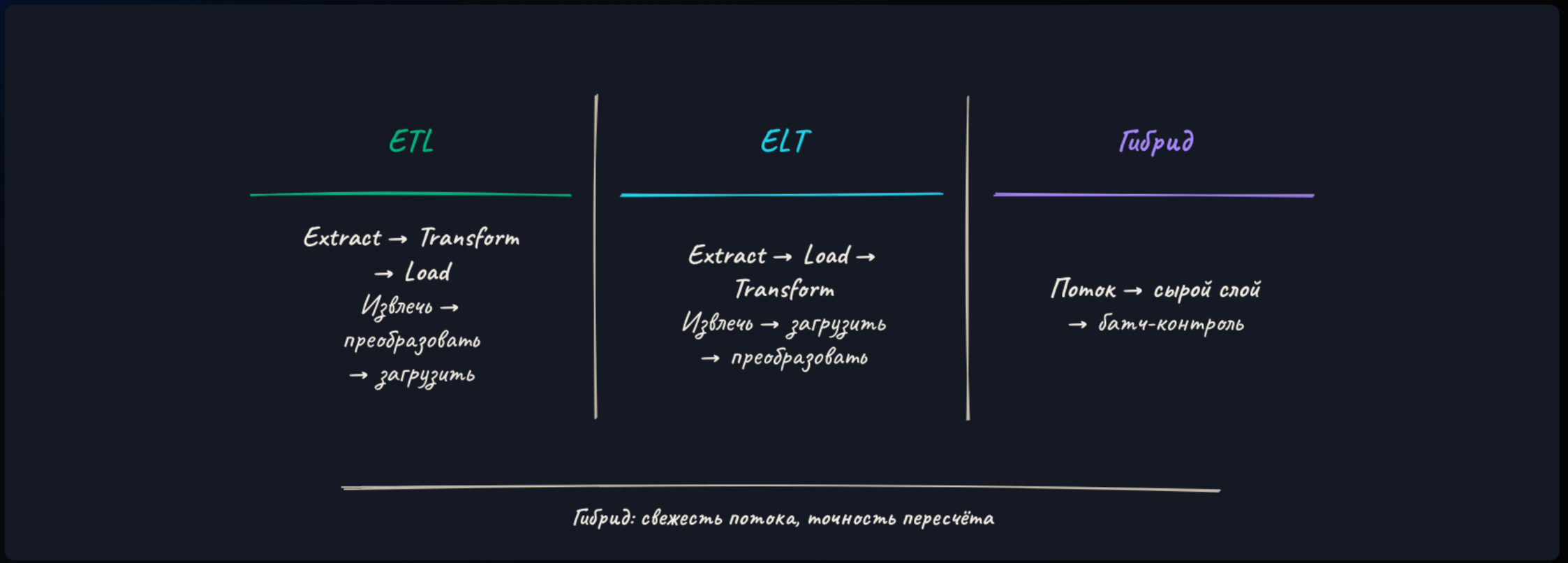
Как данные проходят путь от источника до потребителя

ETL и ELT по-разному строят преобразования



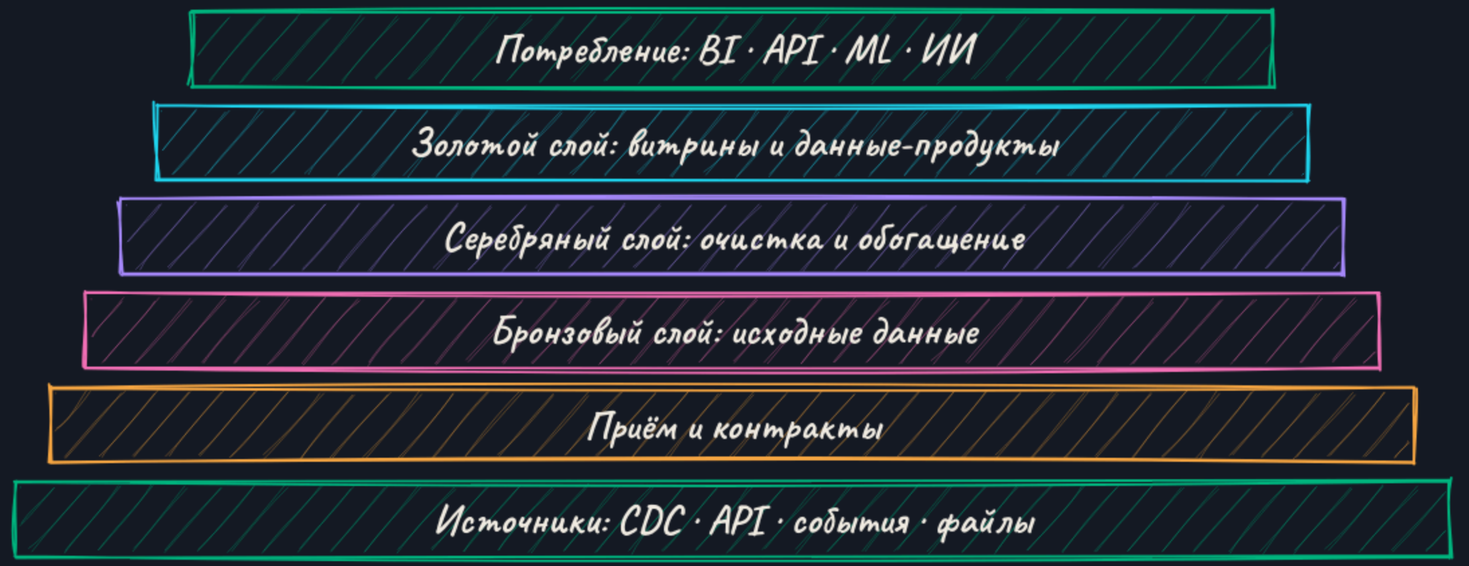
ETL, ELT и гибридный lakehouse-контур

Архитектура Гибрид

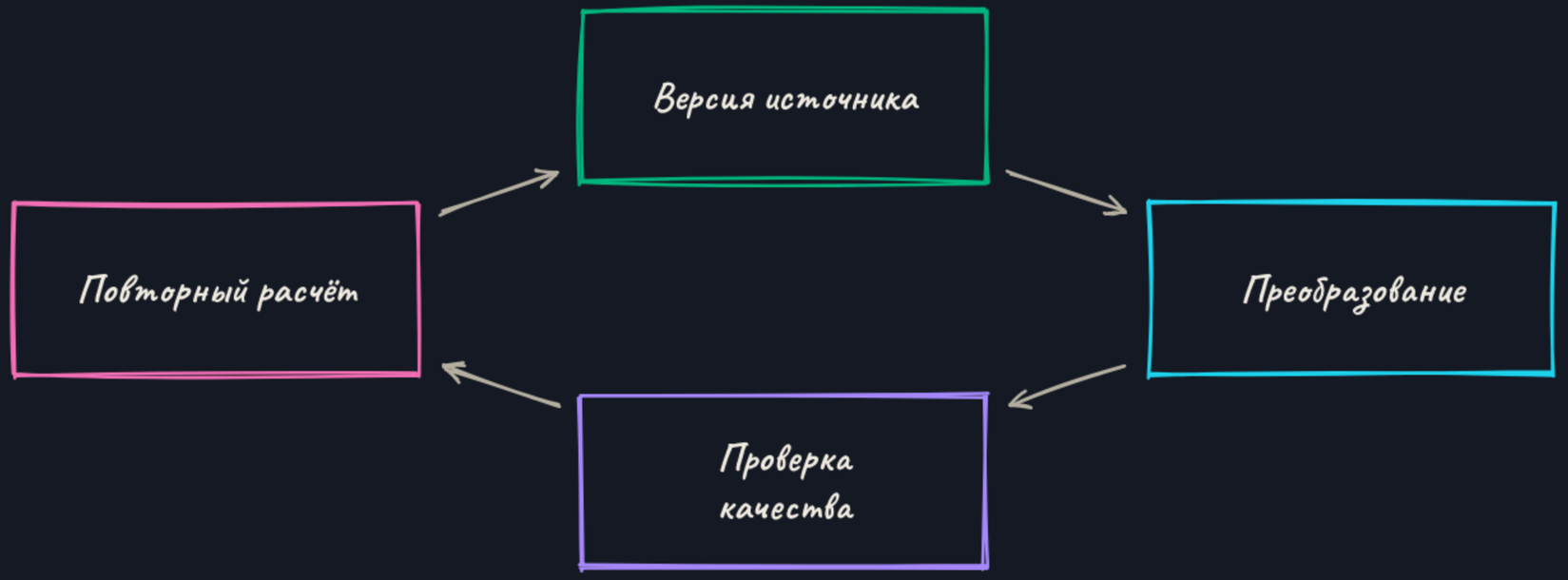


Опорная архитектура платформы данных

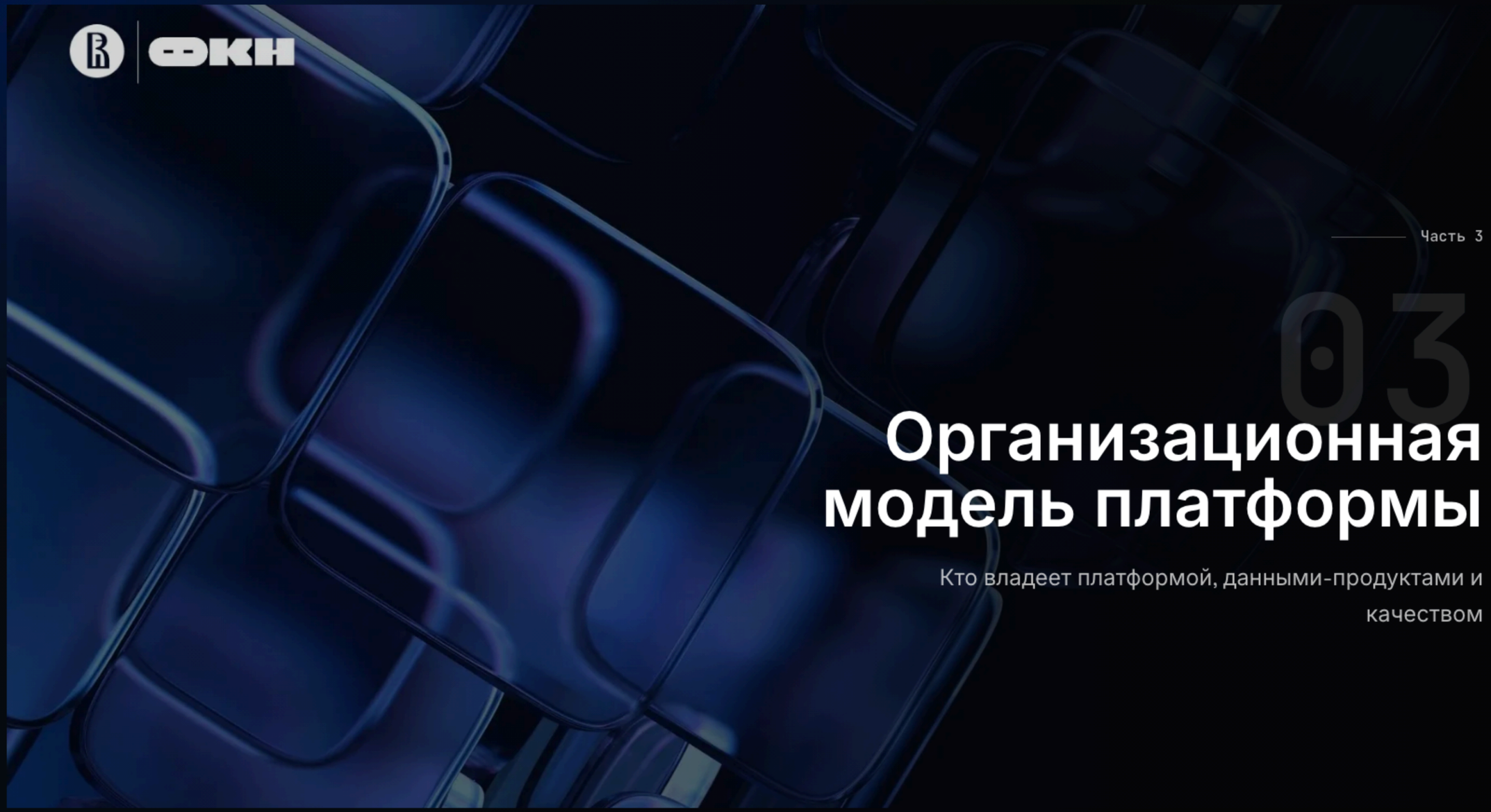
Слои | Опорная схема



Пайплайн: пересчитываемый и объяснимый



«Ровно один раз» редко реалистично; дозагрузка истории — отдельно



Часть 3

03

Организационная модель платформы

Кто владеет платформой, данными-продуктами и качеством

Модели Ответственность

Платформа данных — это еще и оргдизайн

Технология не взлетит без понятного распределения ответственности

Централизованная

Одна команда задает инструменты и правила.

Гибридная

Платформа владеет общим слоем, доменные команды владеют данными-продуктами и SLA.

Федеративная

Data Mesh: домены + общий стандарт.

Центр → гибрид → федеративная модель

Центр	Гибрид	Федерация
Одна команда	Общая платформа	Доменные владельцы
Риск очереди	Доменные продукты	Общие правила

Data Mesh Без мифа

Data Mesh без платформы — хаос

Федерации нужен общий слой

+ Что отдают доменам

- Смысл, бизнес-правила, приоритеты.
- Данные-продукты и потребители.
- Свежесть, полнота, совместимость.

+ Что оставляют платформе

- Самообслуживание: загрузка, хранение, вычисления, оркестрация.
- Каталог, lineage, доступ, качество.
- Финансовые защитные ограничения и инженерные стандарты.



Часть 4

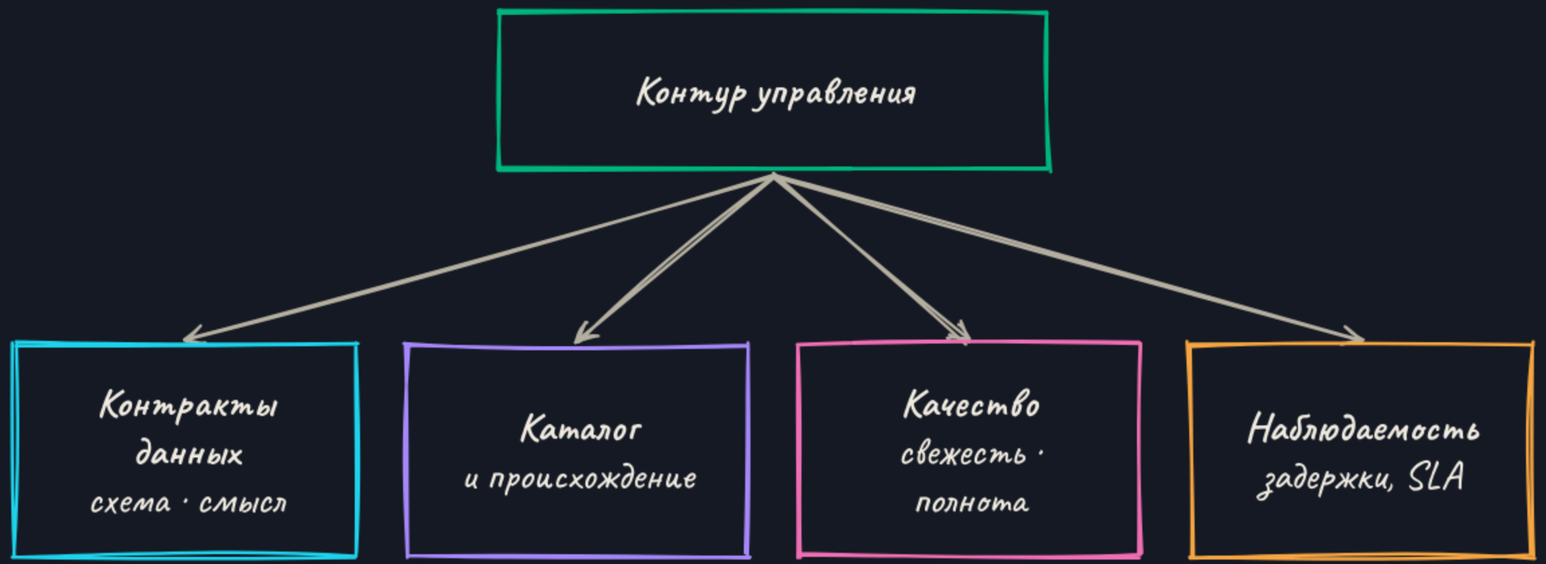
04

Управление и надежность

Контракты, lineage, качество, наблюдаемость и стоимость

Базовый контур управления

Управление Минимум



Антипаттерны Риски

Где платформы данных чаще всего ломаются

Ответственность ломается раньше инструментов

Один гигантский DAG

Локальное изменение становится системным риском.

Скрытая логика

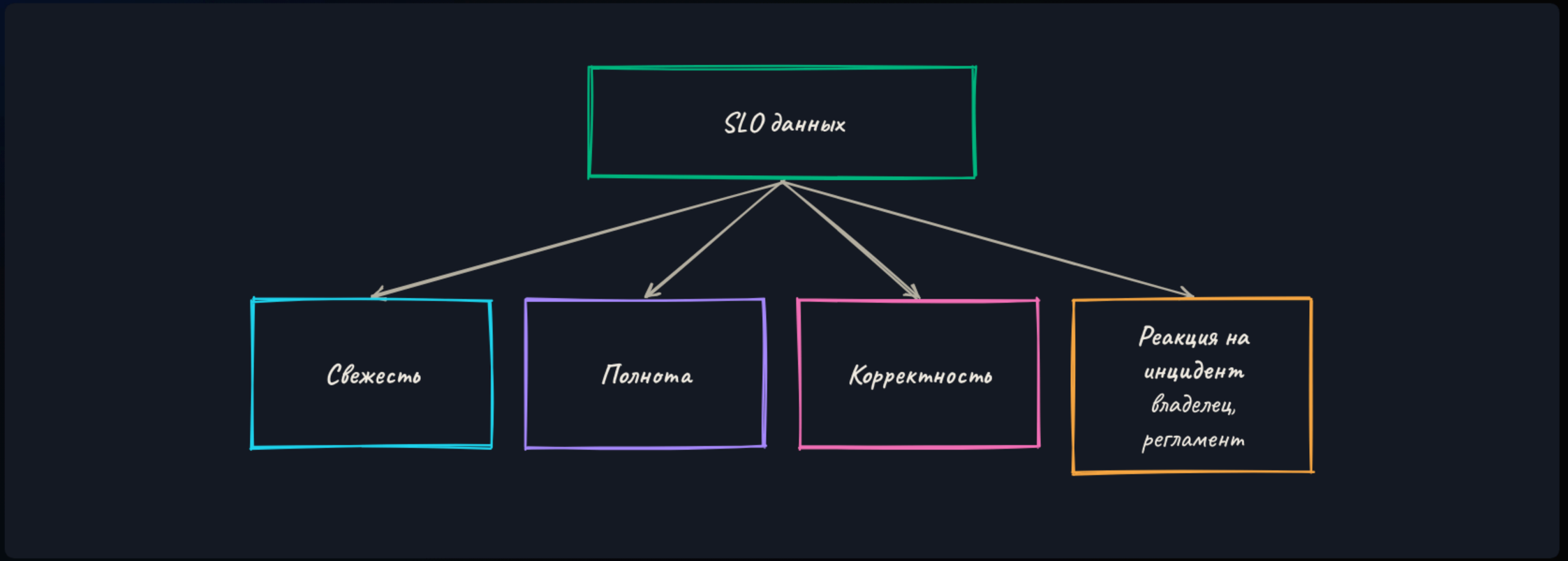
SQL-правила без тестов, версий и ревью.

Нет экономики

Не видно стоимость витрин и хранения.

SLO данных должен быть явным

Практика SLO





Часть 5

05

Метрики зрелости платформы

Как понять, что платформа стала продуктом, а не набором инструментов

Метрики зрелости платформы данных

Зрелость | Метрики

Скорость	Надёжность	Повторное использование	Экономика
Первый датасет < 1 недели	Пайплайны ≥99,9% SLO свежести ≥95% Инциденты схем ↓ за квартал	Доля переиспользования > 60%	Стоимость инсайта под контролем

Первые 120 дней зрелости

Первые 120 дней План

Сначала карта данных и владельцы

- 0-30: источники, потребители, владельцы, инциденты.
- 30-60: контракты, каталог, качество, свежесть.
- 60-120: шаблоны, регламенты, showback.
- Дальше: метрики → ценность для продукта.

Что забрать с собой

Платформа данных — это управляемый путь от источника до решения

- ETL, ELT и гибрид — разные компромиссы.
- Оргмодель важна как стек.
- Зрелость = скорость, качество, свежесть, стоимость.
- Дальше — DataOps/MLOps в продакшене.

- Дальше: DataOps / MLOps для ИИ-бизнеса

Ссылки и материалы

Источники | Материалы

Опорные источники

- Архитектура конвейеров ETL/ELT: ↗system-design.space
- Интервью с Николаем Головым (2025): ↗system-design.space
- Документация OpenLineage, Apache Airflow и dbt.
- Материалы о Data Mesh и операционных моделях.