

PROTOCOL MTP

ЭВОЛЮЦИЯ ПРЕДСТАВЛЕНИЯ И ПЕРЕДАЧИ ДАННЫХ

[#ProfitCloudDay](#)

Алматы, 2026

Области исследований

Фундаментальные основы

Научно-исследовательский центр, ориентированный на создание новых моделей вычислений и структур данных.

Теория вычислимости

P NP PSPACE EXPTIME EXPSPACE



Прикладная математика



Вычислительная физика



Теория инф. структур



Архитектура распр. систем



Ключевой результат

Разработана **модель вычислительных представлений данных**. На её основе сформирован собственный shortcut метод, положенный в основу протокола **МТР**.

Глобальные тенденции (2026 → 2030)

Создание данных (YB)

1.0 ↑x5

Прогноз к 2030 году (vs 150-180 ZB в 2026)

East-West Трафик (Прогноз)

180 → 900 ZB ↑x4

Внутри ЦОД трафик растет быстрее интернета

ИНСАЙТ

Генеративный ИИ создаёт синтетические данные, где машинная генерация допускает достижение **1 YB/год** к 2030 г. Инфраструктура не успевает за темпами генерации.

Динамика объёмов и разрыв

Показатель	~2026	2030 (прогноз)	Рост
Создание данных	150–180 ZB	400–1000 ZB	x2–x5
Инфраструктурное хранение	20–23 ZB	60–80 ZB	x3
East-West трафик (ЦОД)	~180 ZB	600–900 ZB	x3–x4
Интернет-трафик	10–12 ZB	25–35 ZB	x2–x3

Инфраструктура хранения, сети и передача

Более \$1 трлн

Ежегодные глобальные расходы (2026)

Энергопотребление ЦОД

800 ТВт·ч ⚡ High

ИИ-чипы - главный потребитель энергии

⚠ ЭНЕРГЕТИЧЕСКИЙ БАРЬЕР

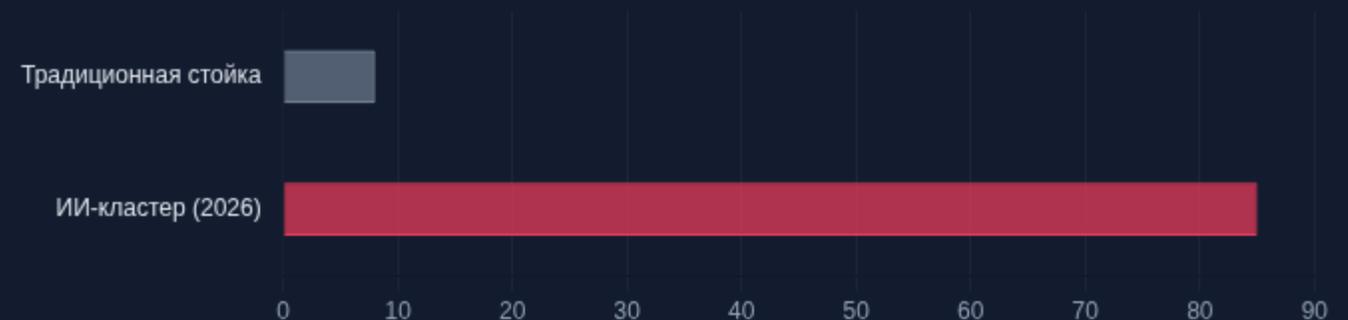
Расходы на питание и охлаждение начинают конкурировать с CAPEX. Стоимость эксплуатации всё чаще определяется энергией, а не дисками. Объём данных растёт быстрее, чем создаётся экономическая ценность.

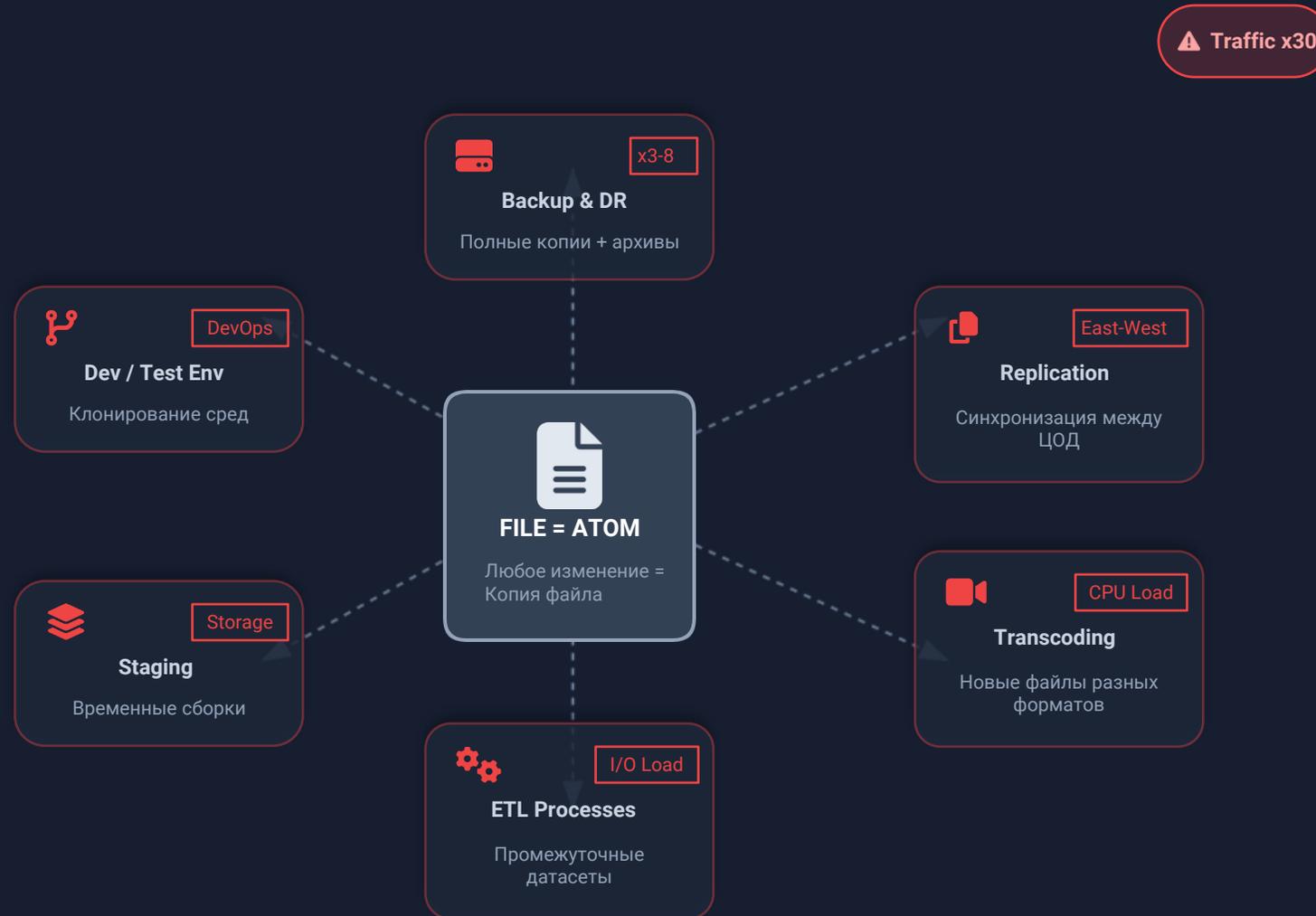
Структура расходов и нагрузки

Метрики 2026

Параметр	Значение (2026)	Динамика
Инфраструктура хранения	~\$530 млрд	Медленное снижение \$/GB
Сети и передача	\$550–650 млрд	Высокий рост
Энергопотребление ЦОД	~500–800 ТВт·ч	Критический рост
Плотность стоек	50–100 кВт	Рост ×10 (vs 5-10 кВт)

Рост плотности мощности (кВт)





Узкие места архитектуры



Файлоцентричность

Файл является неделимым атомом. Любое изменение требует создания новой копии файла, увеличивая нагрузку на storage.



East-West Трафик x25-30

Внутренний обмен внутри кластеров и между репликами генерирует колоссальную нагрузку, превышающую внешний трафик.



Энергоемкость и I/O

Постоянная перезапись и перемещение полных массивов данных приводят к конкуренции расходов на питание с CAPEX.

"Современные решения оптимизируют инфраструктуру, но не уделяют достаточно внимания эффективности структур данных."

Три фундаментальные проблемы



1. Линейный рост данных

Файл воспринимается как неделимый атом. Любое, даже минимальное отличие создаёт новый файл целиком.



2. Загруженность сетей

Передача данных равна физическому перемещению. Каждый трансфер требует перемещения полного массива байтов по каналу.



3. Принципы безопасности

Периметральная защита вокруг файла. Необходимость хранить, копировать и шифровать каждый объект целиком, создавая накладные расходы.

КРИТИЧЕСКАЯ ТОЧКА ОТРАСЛИ

Инфраструктура оптимизируется, но принцип передачи и хранения не менялся десятилетиями.

⚡ Энергетический барьер достигнут

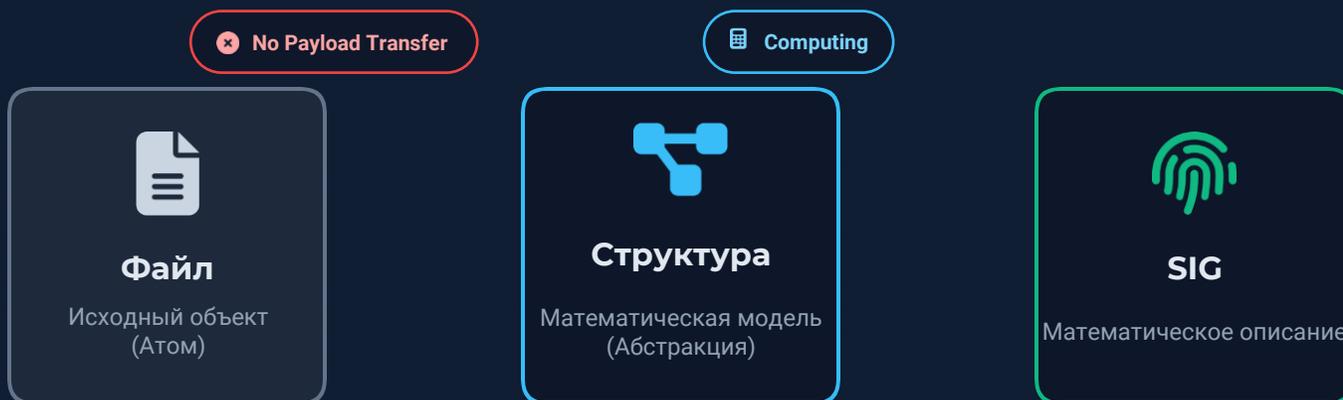
🔌 Инфраструктурный парадокс

Рост инфраструктуры компенсирует архитектурную неэффективность, но не устраняет её причину. Мы строим более мощные дата-центры, чтобы обслуживать устаревшие принципы работы с данными.

🌡️ Экономика энергии

Расходы на питание и охлаждение начинают конкурировать с CAPEX, а стоимость эксплуатации всё чаще определяется уже энергией, а не стоимостью оборудования.

Современный подход к управлению данными приблизился к своему пределу масштабируемости.



Файл как таковой перестает быть единицей передачи и хранения.
Работа ведётся с математической моделью данных.

Новая парадигма

-  **Математическое описание**
SIG - это полное математическое представление файла, а не его физическая копия.
-  **Это НЕ является:**
 - × Компрессией или архивом
 - × Классической дедупликацией
 - × Переносом полезной нагрузки
-  **Bit-to-Bit Восстановление**
Файл становится производной структурой. Гарантируется 100% точность восстановления оригинала.

 **SIG-Structure Data**

Файл перестаёт быть **атомарной единицей** инфраструктуры.

В новой архитектуре файл становится лишь производной функцией от математической структуры.

**ТРАНСФОРМАЦИЯ ФУНКЦИЙ ФАЙЛА****Не единица передачи**

Больше не перемещаем полные массивы

**Не единица резервирования**

Устранение избыточных физических копий

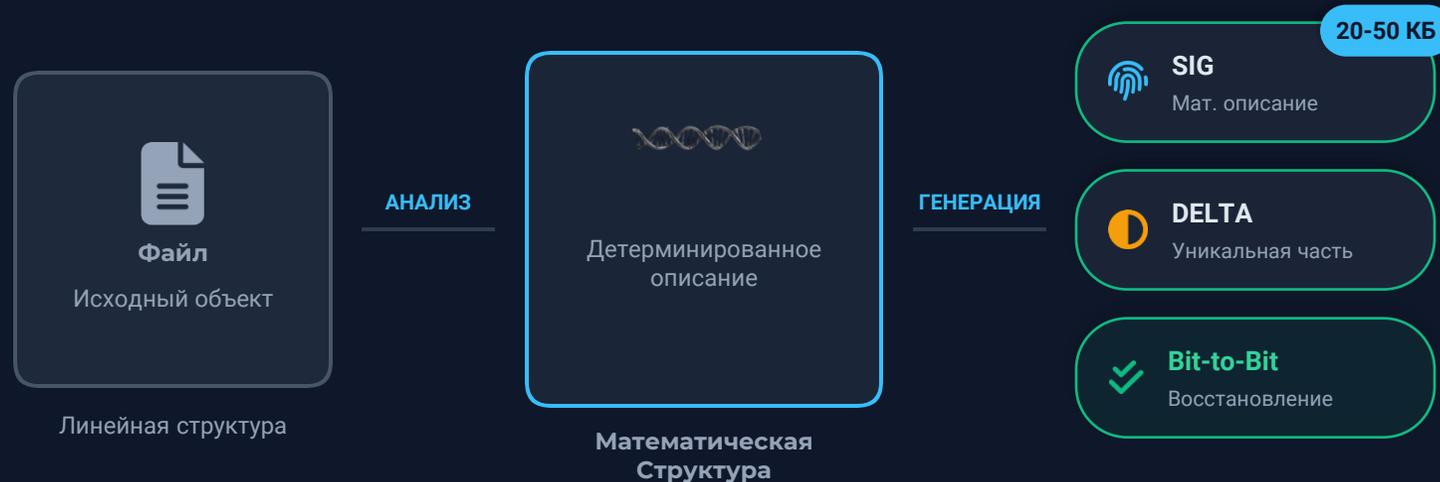
**Не единица синхронизации**

Синхронизация только изменений структуры

**Вся работа инфраструктуры переводится на SIG.**

(SIG - полное математическое представление данных)

Базовый принцип



Ключевые компоненты

-  **Математическая структура**
Файл трансформируется в детерминированное математическое значение. Это не хеш, а вычисляемое описание содержимого.
-  **SIG (Signature)**
Компактный дескриптор (20-50 КБ). Не содержит полезной нагрузки. Не является архивом или компрессией.
-  **Оптимизированная передача**
Передается только SIG и минимальная Delta к базе MTP. Экономия трафика до 20x.
-  **Bit-to-Bit Restore**
Гарантированное побитовое восстановление файла. Файл становится производной функцией от структуры.

Свойства Базы данных МТР

Роль протокола в системе

Протокол формирует собственную структурную базу, которая заменяет множественное хранение файлов требующих многократные вычислительные процессы

Архитектурные теги

Canonical Base

Pre-trained

Binary Optimized

Non-Linear



Детерминированная

Строго определенная математическая структура, гарантирующая предсказуемость результата.



Индивидуализированная

Собственная бинарная структура, уникальная для окружения, что повышает безопасность.



Насыщаемая

База данных МТР достигает плато, после которого новые данные требуют минимум места.



Сублинейный рост

Объём базы не растёт линейно с объёмом входящих данных ($O(1)$ или $O(\log n)$ в пределе).

✓ Результат внедрения

База данных МТР становится **единственным набором данных**, к которому обращаются все операции восстановления и передачи.

Сравнение архитектурных подходов хранения

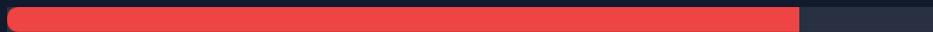
Comparison

AS-IS: Сейчас



Физический объём

3-8x



Множественные копии

Реплики и транскоды хранятся как полноценные отдельные файлы.

Линейный Backup

Резервное копирование = полная физическая копия файла.

Сервисные дубликаты

Промежуточные копии ETL и staging занимают место.

VS

МТР: Новая парадигма (уровень 1)



Физический объём

1.0 - 0.6x



Структурное представление

Транскоды и реплики - это виртуальные проекции, а не файлы.

База данных МТР

База данных МТР вместо множества избыточных реплик.

Умный Backup

Backup = SIG (метаданные) + дельта к Базе данных МТР.
Мгновенное восстановление.



Входные данные: 1 ПБ
Хранение: 3-8 ПБ



Входные данные: 1 ПБ
Хранение: 1 ПБ

Сравнение архитектурных подходов ВЫЧИСЛИТЕЛЬНЫЕ ПРОЦЕССЫ

Optimization & Efficiency
MTP Protocol Layer 1



🕒 Сейчас (AS-IS)

⌘ Повторная компрессия данных

🌀 Повторное хеширование

🛡️ Повторная проверка целостности

👁️ Rebuild архивов и Staging-сборки

🔄 Частично повторные ETL процессы



🚀 С внедрением МТР (уровень 1)

✓ Устранение повторной компрессии

✓ Исключение проверки копий

⊘ Отсутствие Rebuild backup-файлов

✍️ Сокращение Write Amplification

Общее снижение

Повторных вычислений

↓ **50%**

Сравнение архитектурных подходов передача данных

Экономия Трафика

2-7x

✓ Efficiency

Многokратное снижение объемов передаваемых данных

Сетевая Нагрузка

-75%

↓ Load

Снижение нагрузки на внутренние каналы ЦОД



АРХИТЕКТУРА ПЕРЕДАЧИ

Вместо полного файла передается **SIG (20–50 КБ)** и дельта. Внутренний обмен (East-West) сокращается радикально.

Сравнительный анализ передачи 1 ГБ (уровень 1)

Параметр	Сейчас (AS-IS)	С МТР (To-Be)	Эффект
Объем трафика (на 1 ГБ)	3.0 – 8.0 ГБ	SIG + Дельта	↓ 2–7x
Структура пакета	Полный файл	20–50 КБ + Дельта	Оптим.
Влияние на сеть	Линейный рост	Сублинейное	-75% Load



Максимальная эффективность

Потенциал протокола MTP

Передача

Размер SIG

20–50 КБ

- ✓ Дельта в **5–20 раз меньше** 1 ГБ входного файла
- ✓ При насыщенной БД трафик составляет всего **1–3%** от исходного объема (**от 10 ПБ**)
- Кратное снижение нагрузки на каналы связи (East-West)

Хранение

Физический объём

1.0x на первом уровне

от логического (вместо 3–8x)
до 0.6x после насыщения

- ✓ Устранение избыточных копий
- ✓ Замена множества реплик на **1 Базу данных MTP**
- Существенная экономия дискового пространства

Вычисления

Повторные операции

~0

Устранение дублирующих процессов

- ✓ Устранение повторной **компрессии**
- ✓ Устранение повторного **хеширования**
- ✓ Исключение проверки целостности копий



От линейного роста данных к сублинейному

Рост базы данных навсегда перестает быть линейным по отношению к объему входных данных. С увеличением объёма данных эффективность хранения и передачи возрастает экспоненциально.



Мини-ИИ в ядре протокола

В ядре протокола встроен **адаптивный алгоритмический модуль**, который обучается на паттернах ваших данных.



Это не внешняя ML-модель, а встроенный оптимизационный механизм ядра, работающий в реальном времени.



Адаптивная оптимизация

Автоматическая настройка параметров обработки под специфику потока данных.



Admission-политика

Интеллектуальное управление приемом данных в базу MTP для предотвращения избыточности.



Стабилизация базы

Контроль роста размера базы данных MTP при масштабировании объемов.

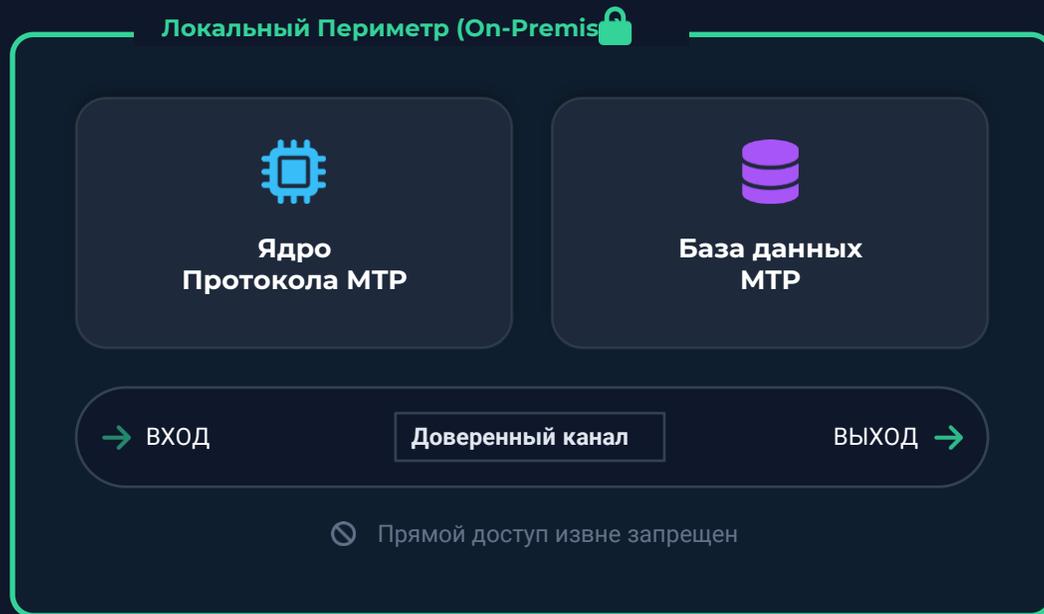


Обучение

Модель обучается на ваших данных и работает исключительно локально.

MTP Security Protocol & Cryptography

Контур изоляции



Принципы защиты



SIG — математическое описание

SIG не содержит сжатых исходных данных. Это детерминированное значение, которое бесполезно без доступа к локальной базе и алгоритмам.



Приватность дельты

Передаваемая дельта не раскрывает структуру исходного файла и не позволяет восстановить данные.



Индивидуальная структура БД

Уникальная бинарная структура базы исключает возможность воспроизведения данных вне конкретного доверенного окружения.

Производительность и оптимизация

Метрики скорости и устранение избыточности



Скорость обработки

Генерация SIG

1 ГБ ≈ 1 сек

Восстановление

0.2–0.5 сек

на 1 ГБ данных



Узкие места

Лимитирующий фактор

I/O Дисков

Производительность упирается в диск, а не в процессор

 **CPU не перегружен** - алгоритм оптимизирован

 Скорость зависит от физической пропускной способности SSD/NVMe



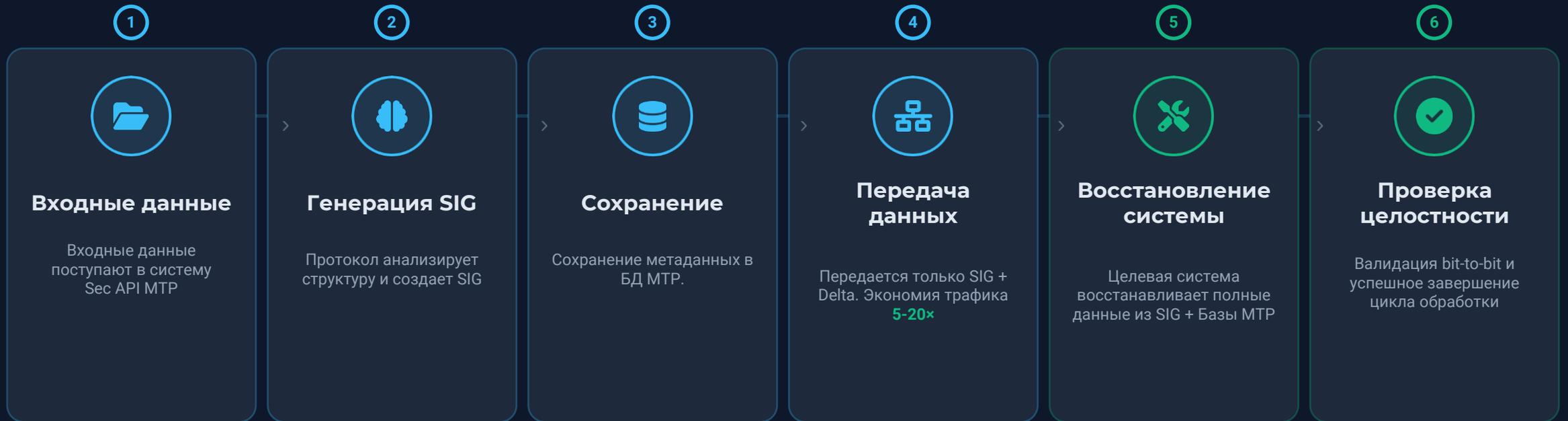
Что устраняется

Эффективность МТР

**Лишнее
хранение
и процессы**

- × Повторная передача полного файла
- × Репликация неизменённых блоков
- × Ваксип как полная копия файла
- × Повторная компрессия и проверки копий
- × Любые избыточные вычислительные процессы

Жизненный цикл обработки данных



Непрерывный цикл оптимизации

Каждый проход обогащает базу данных МТР, делая последующие операции обработки и передачи еще более эффективными (насыщение базы).

Экономическая эффективность

Комплексный анализ затрат и выгоды



 Storage

- 50-75%

Снижение физического объёма хранения данных

 Network

2-7x

Кратная экономия на трафике East-West

 Compute

до -50%

Снижение повторных вычислений и операций I/O

 Backup

- 50-95%

Уменьшение объёма резервных копий

 Restore

×10-30

Ускорение восстановления данных

 Энергия

Пропорционально

Снижение потребления от сокращения стоек



Ежегодно для крупных ДЦ

> сотни \$ млрд

Для отрасли



Критерии внедрения

Протокол демонстрирует максимальную эффективность в инфраструктурах, где критичны **стоимость хранения и скорость передачи**.

Условия эффективности:

 Большие массивы данных

 Повторяемая структура

 Высокая стоимость хранения

 Дорогая передача данных

Сферы применения



Дата-центры

Оптимизация стоек



Облака

Cloud Storage & IaaS



Backup / DR

Мгновенное восстановление



Big Data

Аналитические озера



Медиа-платформы

Стриминг и архивы



Государственные и изолированные системы

СМЕНА ПАРАДИГМЫ ДАННЫХ

МТР — это не оптимизация хранения
это эволюция фундаментальных принципов
цифрового мира.



Protocol MTP Ready 2026



Файл перестаёт быть атомом

Мы отказываемся от устаревшей концепции файла как неделимой единицы. Любое отличие больше не порождает новую сущность.



Структура — первична

Первичной сущностью становится математическая структура. Данные не копируются физически, а вычисляются алгоритмически через SIG и дельту.

Protocol MTP - это переход от эпохи избыточного копирования к эре чистого математического смысла.



Эволюция представления и передачи данных