

# Концепция проекта по Big Data. Распределенный анализ данных Ethereum Blockchain

Автор:  
Алексеев Андрей Евгеньевич

Институт математики и информационных технологий

12 марта 2025 г.

- Экспоненциальный рост экосистемы Ethereum
  - 2+ млн транзакций ежедневно
  - 4000+ DApps и 200K+ смарт-контрактов
- Практические потребности
  - Обнаружение финансовых мошенничеств
  - Анализ транзакций Ethereum для выявления подозрительных активностей.
  - Отслеживание потоков средств

# Типы данных в блокчейне Ethereum

## Структурные данные

- Блоки (хеш, газ, временные метки)
- Транзакции (from/to, value)

## Семантические данные

- События смарт-контрактов
- Трассировки вызовов (traces)
- Состояние сети (storage)

- ❶ Статистический анализ
  - Распределение газовых цен
  - Активность майнеров
- ❷ Временные паттерны
  - Суточная активность DApps
  - Корреляция с курсом ETH
- ❸ Аномалии
  - Выбросы в размере транзакций
  - Подозрительные контракты

## Преимущества подхода

- Выявление сложных связей
- Обнаружение паттернов поведения
- Визуализация потоков средств

Применение: AML, расследование взломов, анализ DEX

## Стек технологий

- Ethereum ETL: Экспорт данных
- Apache Spark: Обработка данных
- GraphFrames: Графовый анализ
- Hadoop: Хранение данных