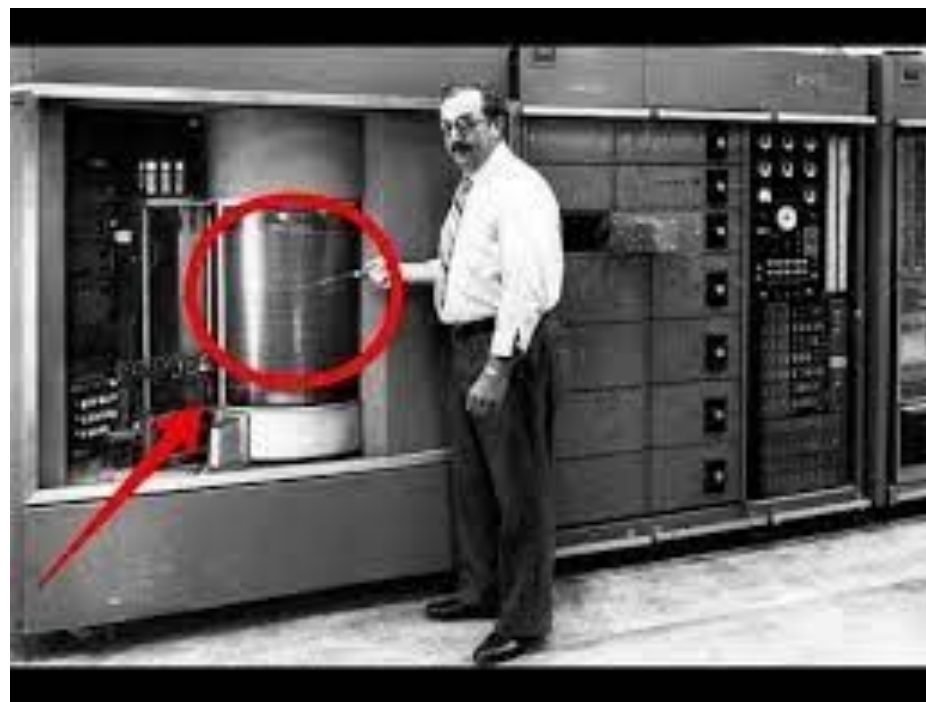


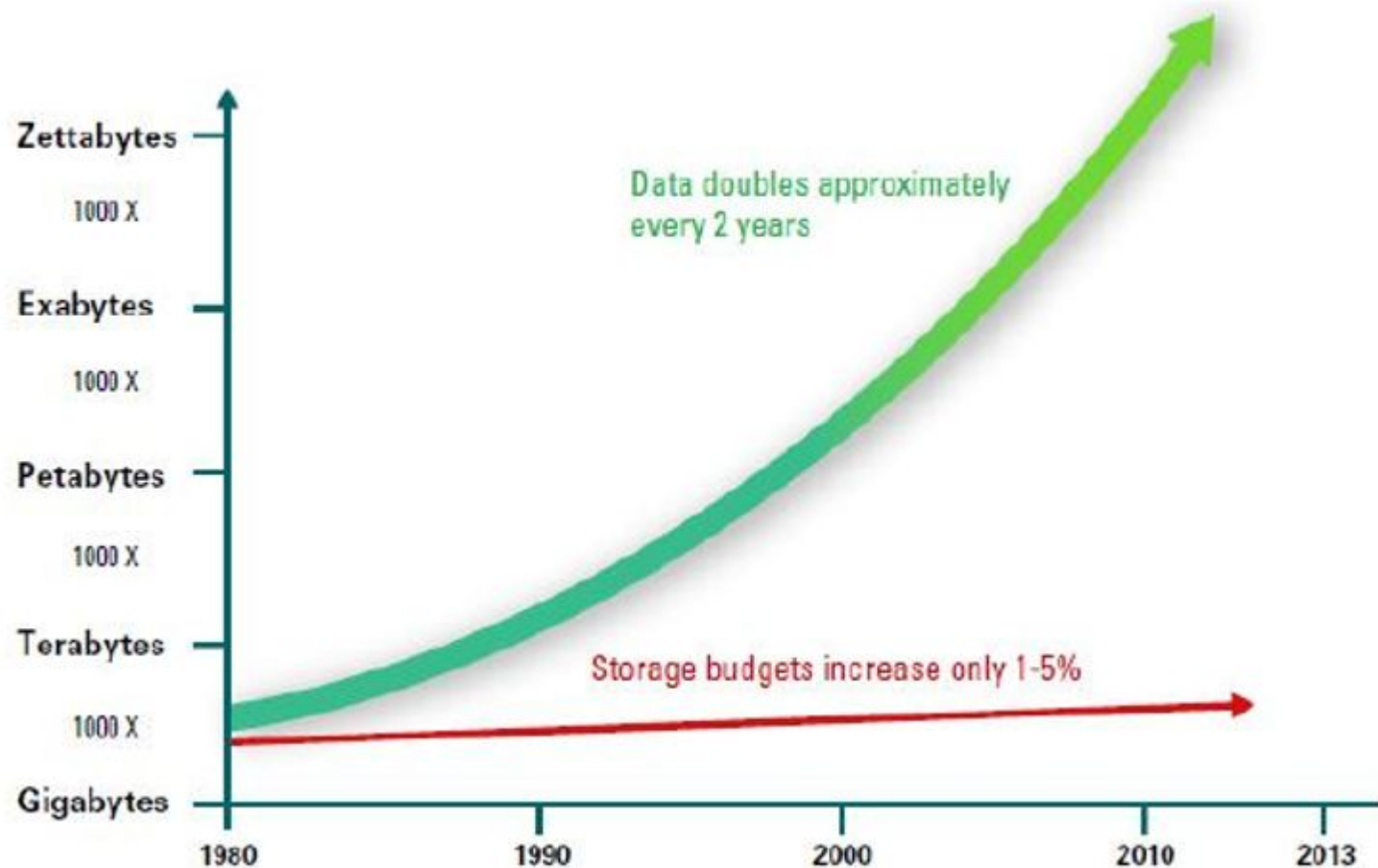
# Экосистема больших данных

# Предистория



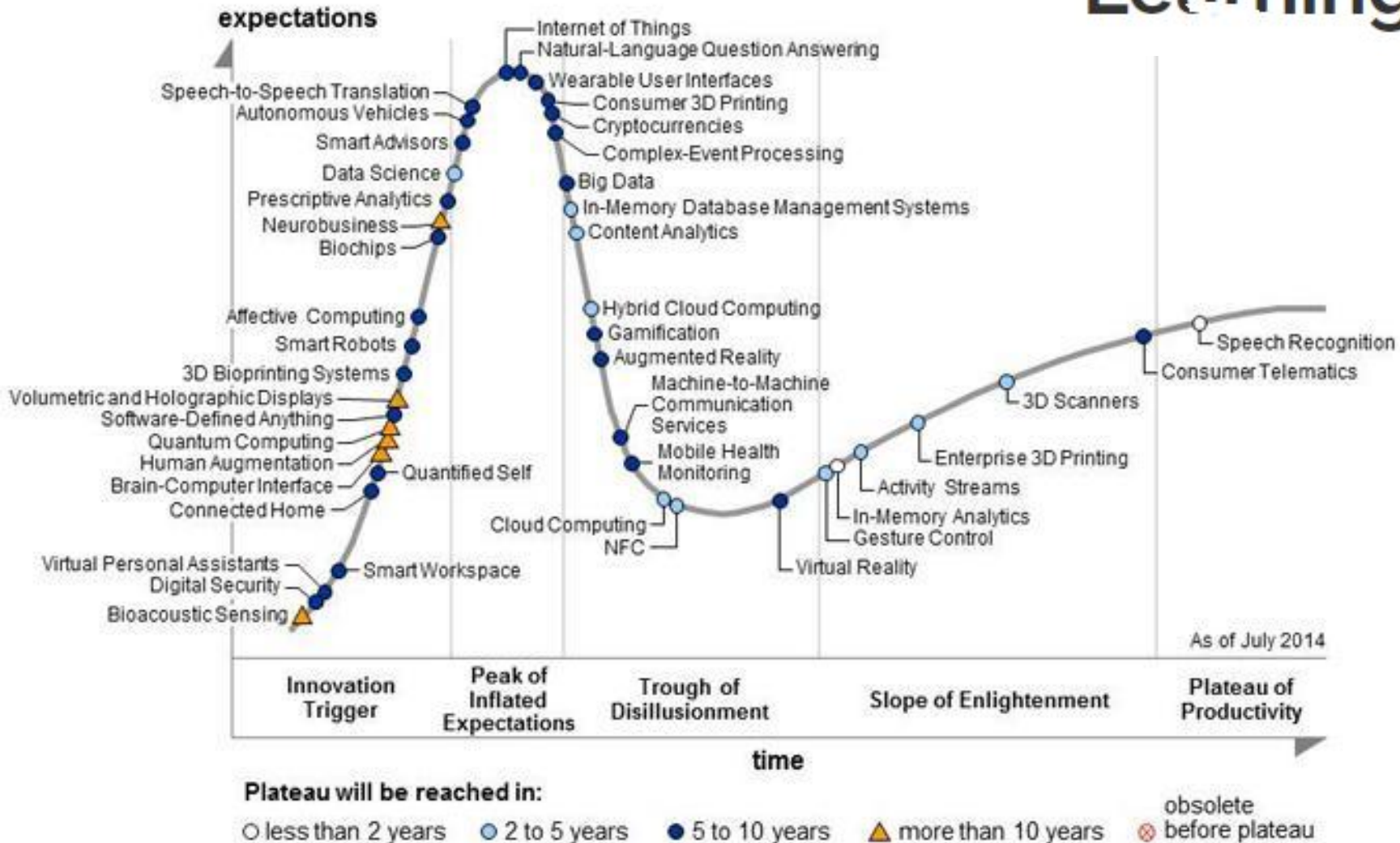
# Эволюция технологий хранения данных

- ▶ Все будет флэш



# Gartner 2015 Hype Cycle: Big Data is Out, Machine Learning is in

Реальность



► График 2014 года

# VVV

## Критерии (признаки) больших данных

- ▶ Volume
- ▶ Velocity
- ▶ Variety
- ▶ Veracity
- ▶ Viability
- ▶ Value
- ▶ Variability
- ▶ Visualization

# Виды больших данных

- ▶ Структурированные
  - ▶ Базы данных
- ▶ Слабо структурированные
  - ▶ XML- страницы
  - ▶ Email сообщения
- ▶ Неструктурированные
  - ▶ Текстовые данные
  - ▶ Изображения
  - ▶ Звук, видео
  - ▶ Мультимедиа

# Эволюция систем управления данными

## Database Management system

1980s

Flat File Database

Relational Database

Hierarchical Database

RDBMS

NoSQL

Key-value

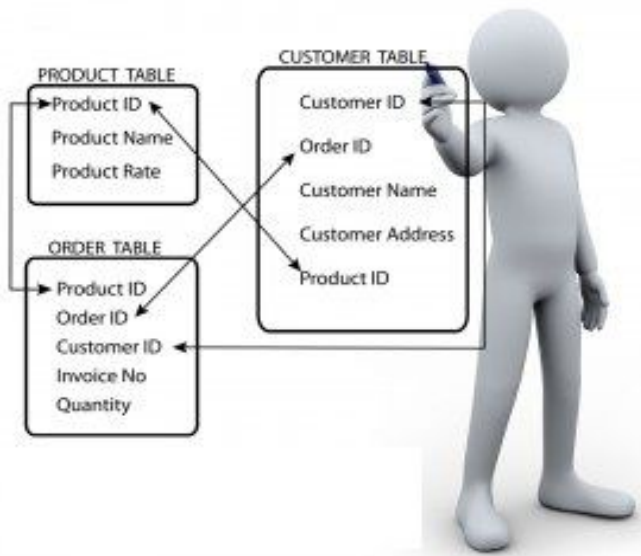
Column Oriented

Document Oriented

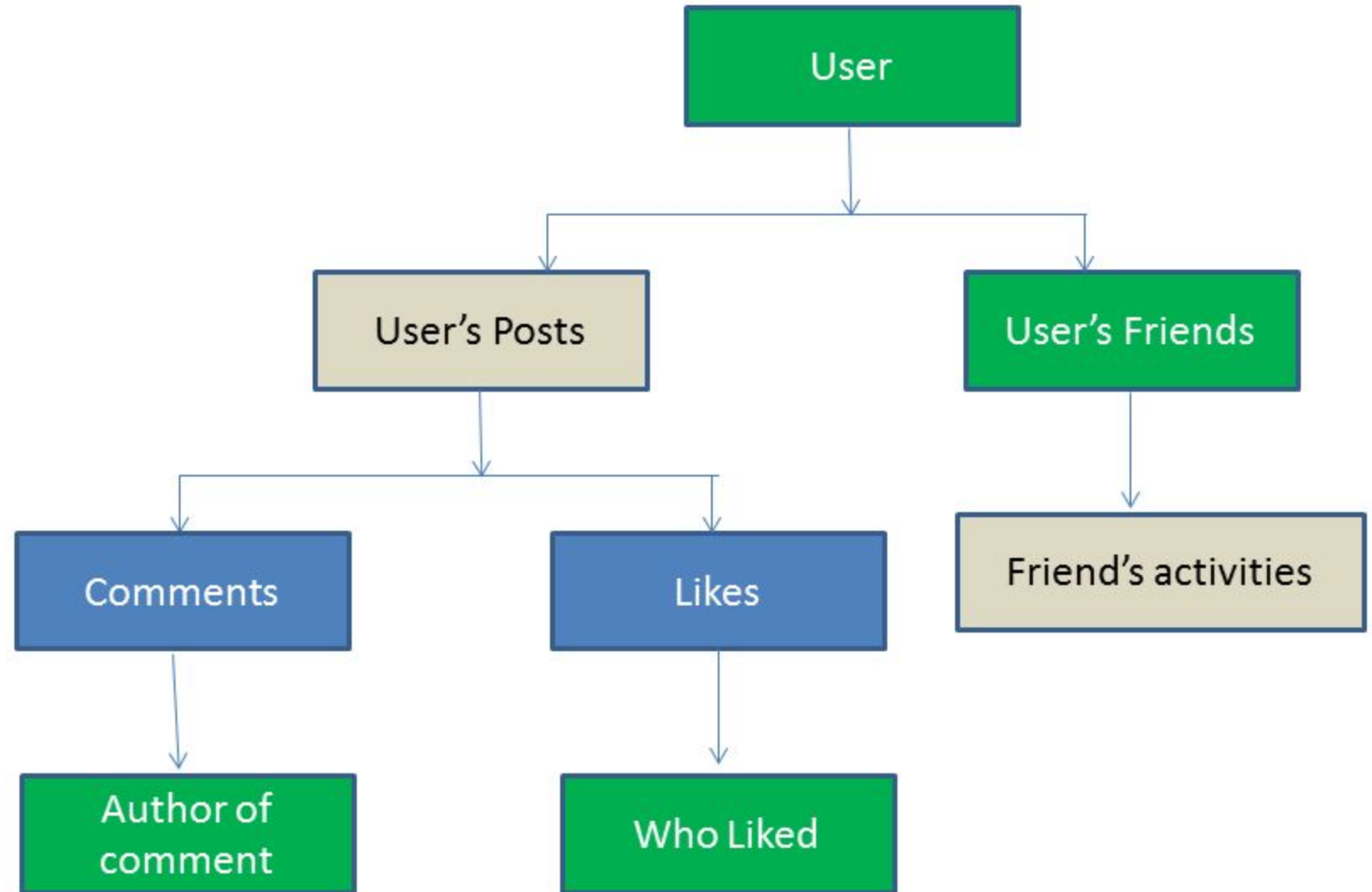
Graph DB

2014

Time of Popularity



# NoSQL базы данных



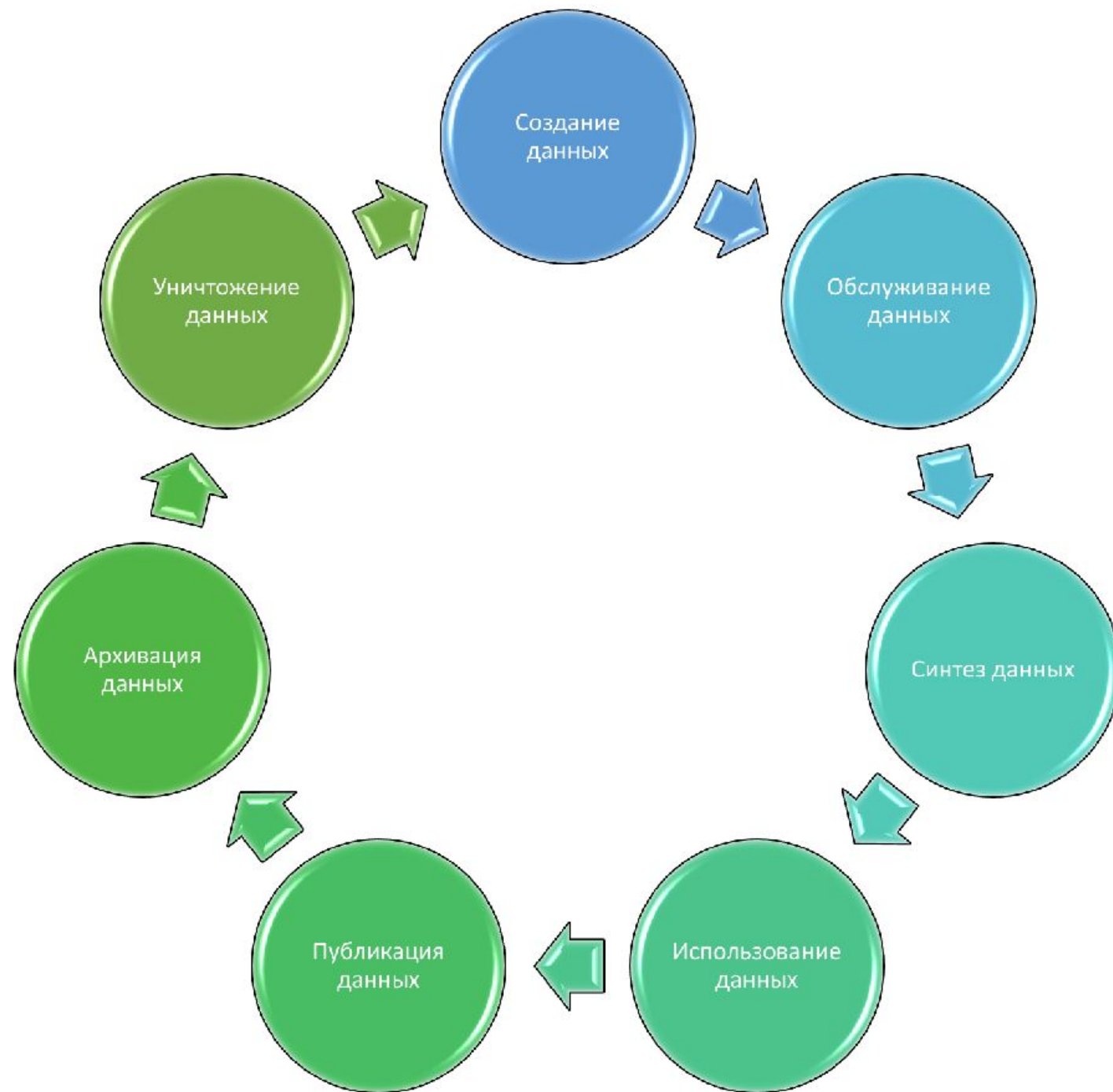


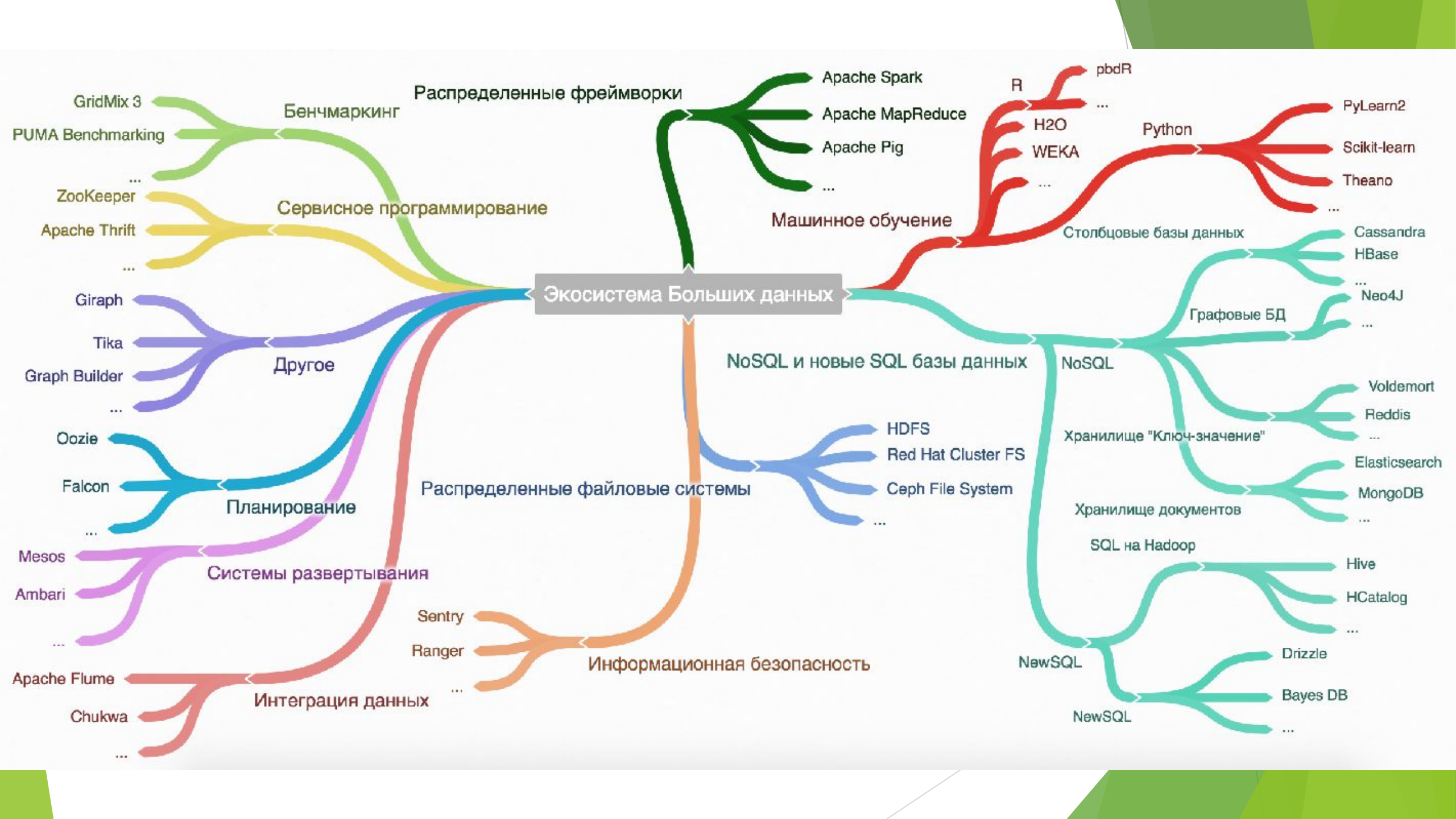
# Жизненный цикл данных

- ▶ - это последовательность этапов, которую набор (последовательность) данных проходит от начального этапа создания или получения до момента архивации или удаления.

- Создание
- Хранение
- Применение
- Удаление

ЖЦД





# Data analysis vice versa data mining

- ▶ Анализ данных
  - ▶ data analysis
- ▶ процесс исследования, фильтрации, преобразования и моделирования данных с целью извлечения полезной информации и принятия решений
- ▶ Интеллектуальный анализ данных
  - ▶ Data Mining
- ▶ метод анализа данных, который фокусируется на моделировании и открытии данных, а не на их описании

# Плюсы и минусы больших данных

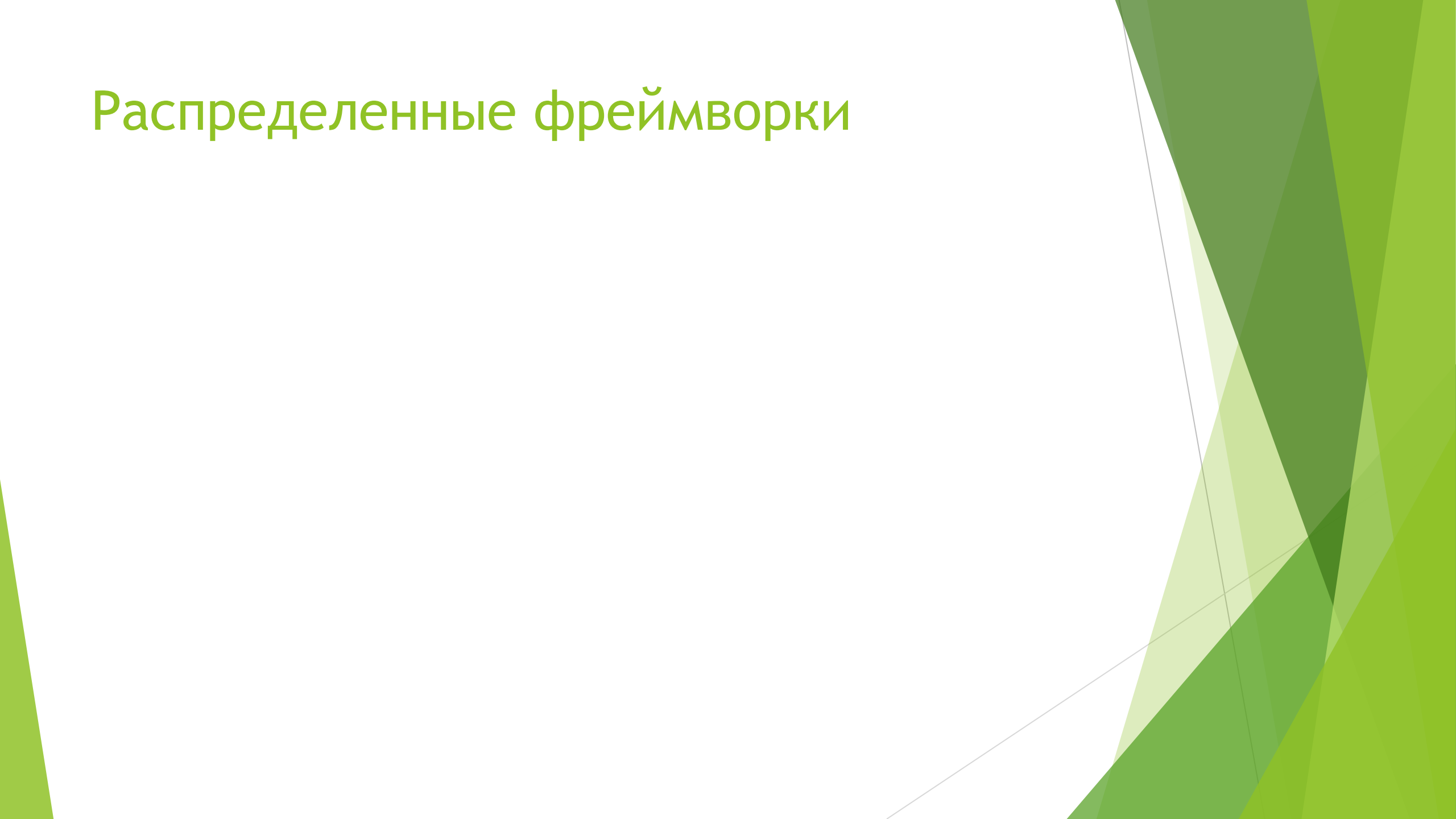
- ▶ Хранение и анализ данных - это расходная часть бюджета
- ▶ Результаты анализа больших данных не всегда могут принести реальную пользу
- ▶ В условиях цифровой экономики информация становится ценным активом,
- ▶ Базы больших данных имеют реальную коммерческую стоимость

# Области применения для анализа больших данных

# Распределенные файловые системы

- ▶ Распределенные файловые системы

# Распределенные фреймворки





# Тенденции применения больших данных

