

# Распределенные системы хранения и обработки данных

Владислав Белогрудов, EMC

[vlad.belogrudov@gmail.com](mailto:vlad.belogrudov@gmail.com)

# Содержание курса

- Системы хранения данных
- Сети хранения данных
- Обеспечение непрерывности бизнеса
- Безопасность данных
- Облачные вычисления

# Для чего нужен этот курс?

- Понимание устройства сложных промышленных информационных систем
- Умение управлять данными
- Повышение квалификации
- Сертификация
- Работа
- ..

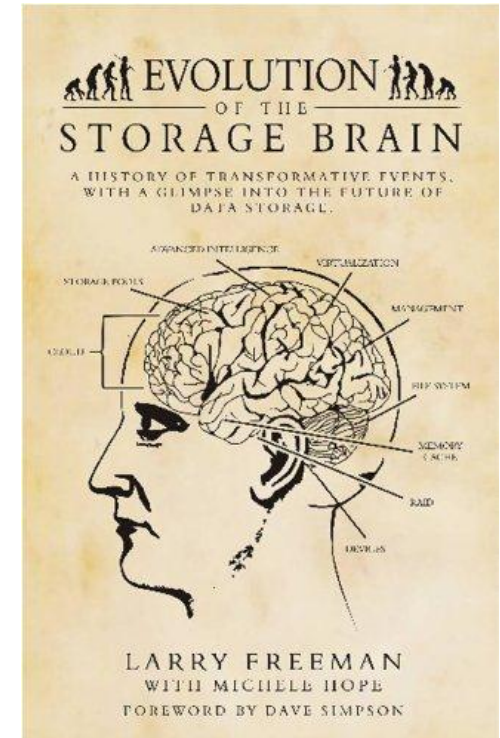
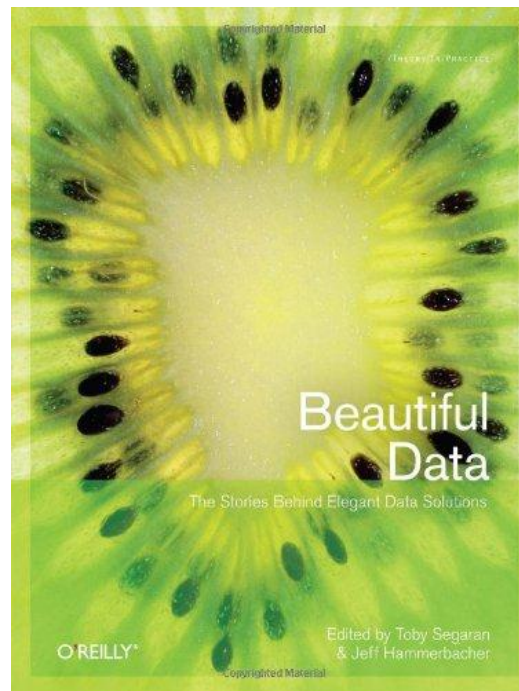


# Задания, работы, оценки

- 100 баллов за весь семестр:
  - 20% контрольные и семинары
  - 10% «домашки»
  - 70% курсовая работа
- Итоги:
  - 40..59                   => 3
  - 60..79                   => 4
  - 80..100               => 5



# Рекомендуемая литература



# Лекция 1

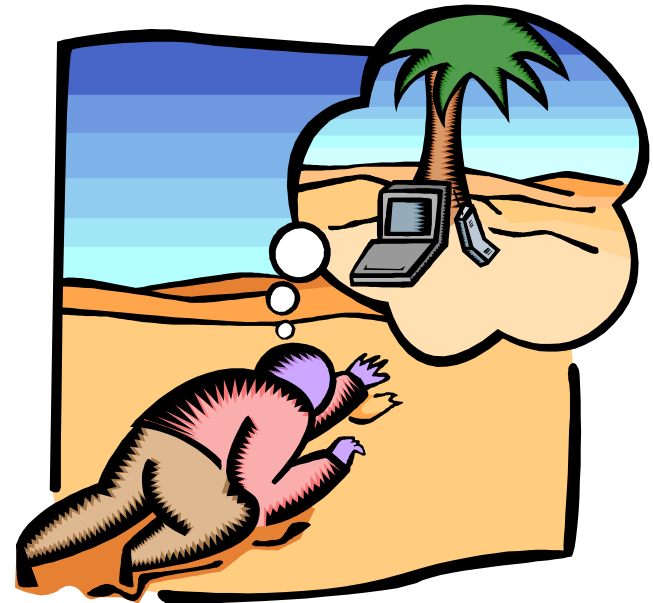
## Введение в системы хранения данных (СХД)

# Содержание лекции

- Данные и информация
- Типы данных
- Большие данные
- Эволюция СХД
- Архитектура ЦОД
- Характеристики ЦОД
- Жизненный цикл информации
- Иерархическое управление носителями

# Роль информации в повседневной жизни

- Информационно-зависимое общество
- Значимость и объем экспоненциально растут





# Требования к технологиям хранения информации

- Доступность
- Защита
- Масштабируемость
- Производительность
- Сохранность
- Вместимость
- Управляемость



# Данные и информация

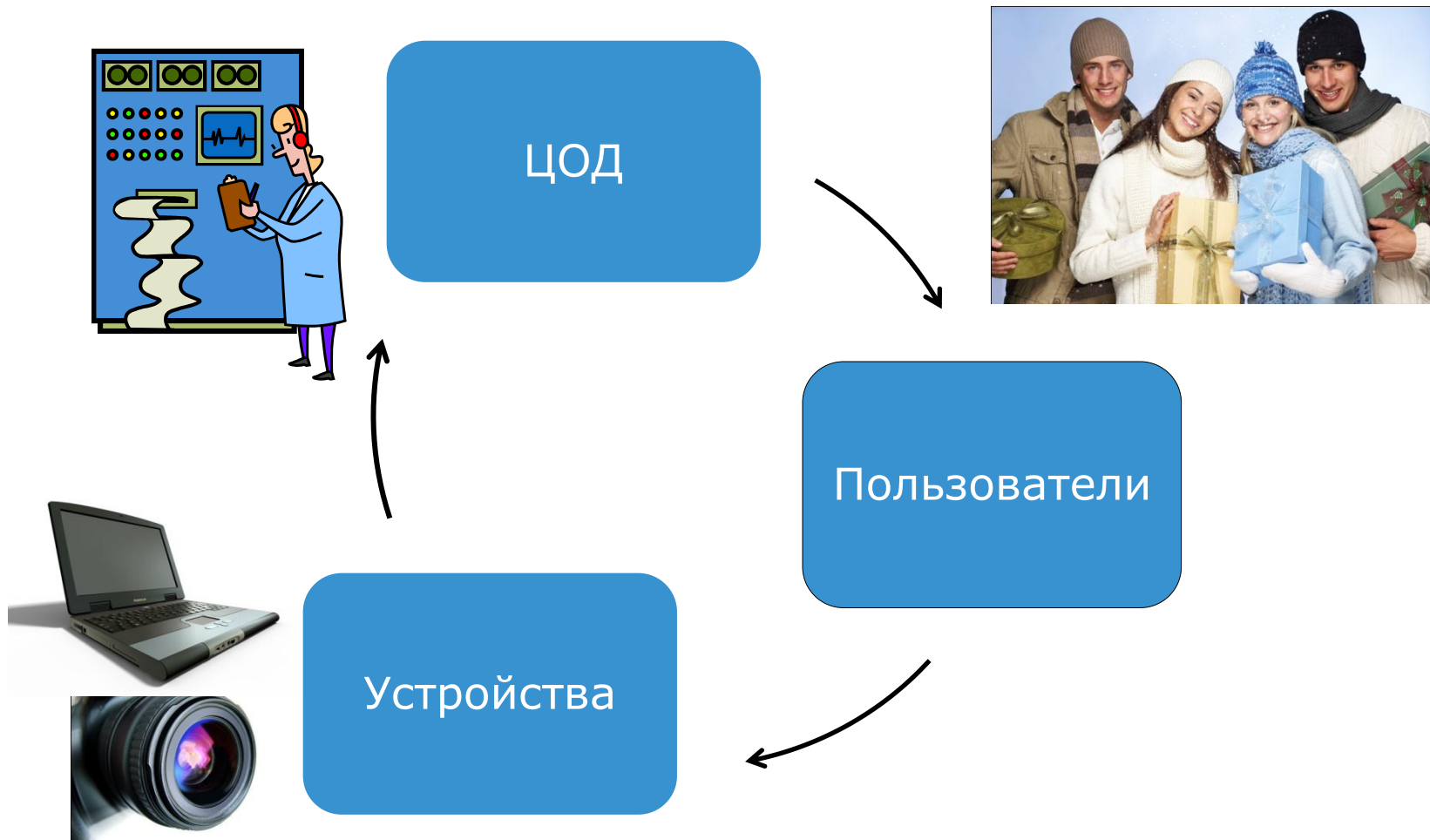
- Данные – совокупность фактов из которых можно сделать выводы.
- Информация – сведения и знания, извлекаемые из данных.
- Данные генерируются все большим числом устройств и людей.



# Рост цифровых данных

- Улучшение производительности обработки цифровых данных
- Снижение цен на цифровые носители
- Появление высокоскоростных технологий передачи

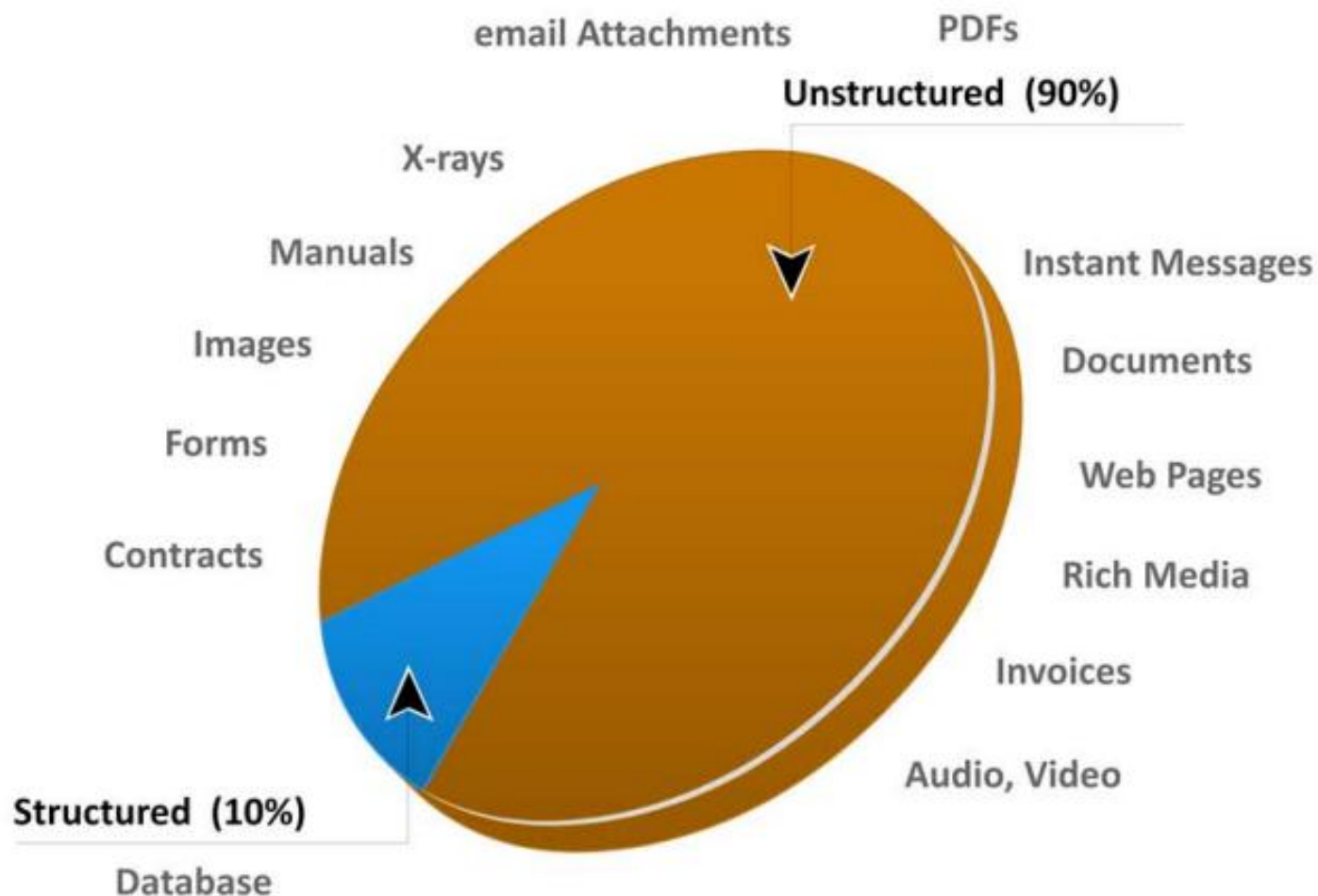
# «Производители» и «хранители» данных



# Типы данных в зависимости от способа управления и хранения

- Структурированные – организуются в ряды, колонки, таблицы.. удобны для обработки – базы данных
- Неструктурированные данные – фотографии, письма, сообщения, документы

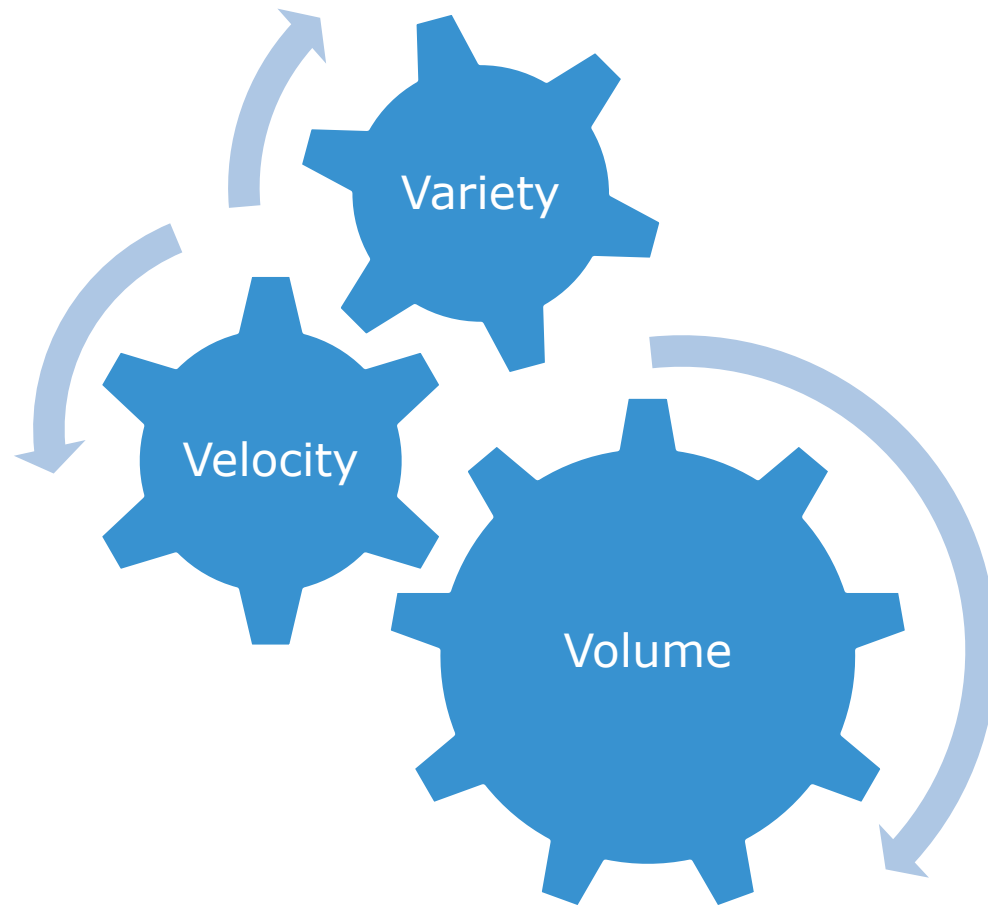
# Структурированные и неструктурированные данные



# Большие данные

- **Большие данные** - серия подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов.
- Объем больших данных настолько огромен, что невозможно его обработать (сохранить/прочитать) без значительных временных затрат обычными средствами.

# 3-V больших данных

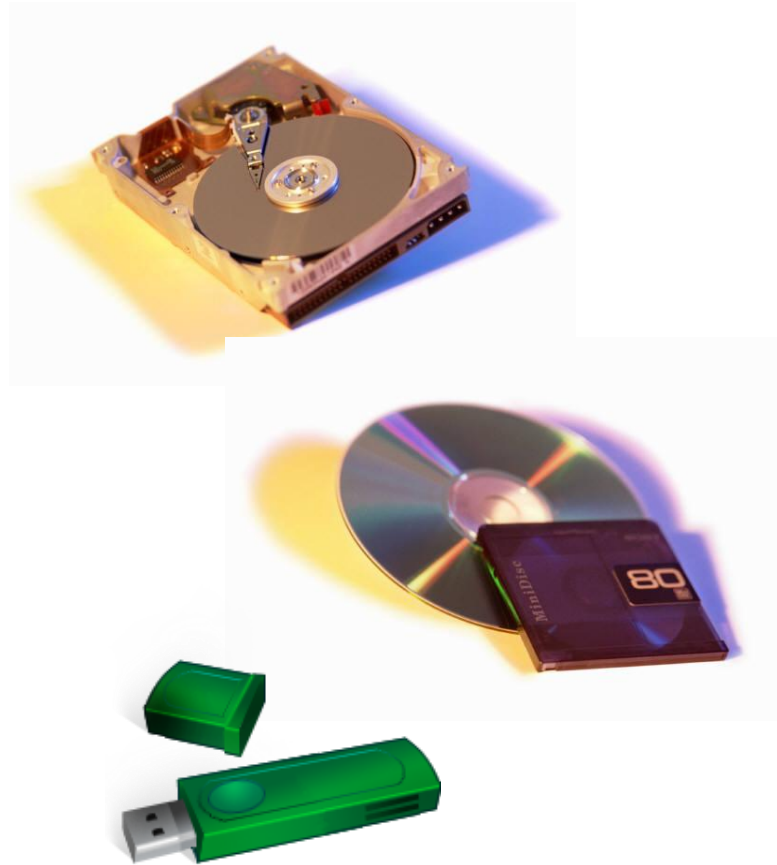




# Методы анализа больших данных

- Data Mining
- Кластеризация
- Регрессионный анализ
- Краудсорсинг
- A/B тестирование
- Анализ временных рядов
- Анализ тональности текста
- ...

# Устройства для хранения данных



# Как это начиналось



5 мегабайт из 1956-го

# Сколько данных помещается в голове?



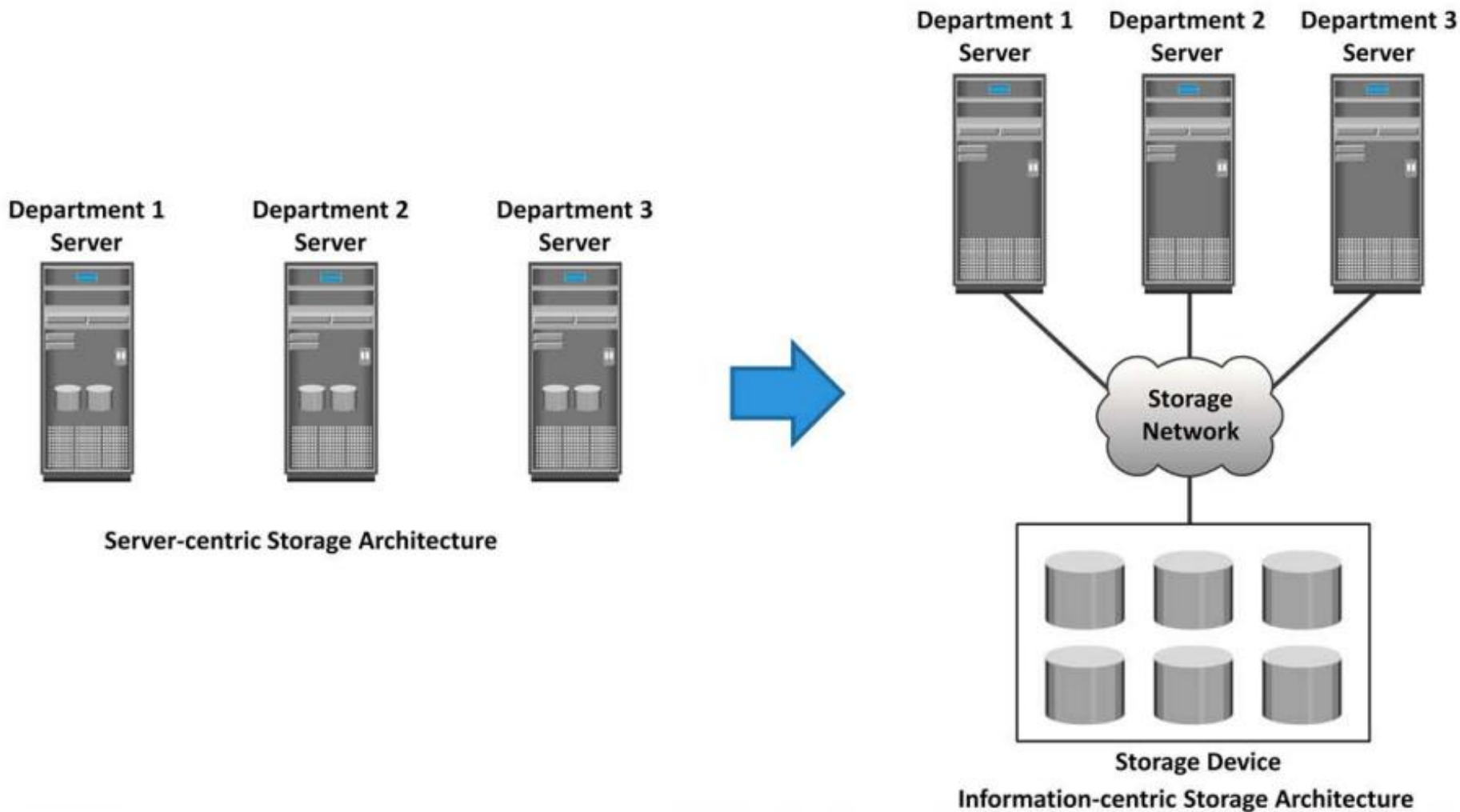
Сколько данных помещается в голове?

1..1000 Терабайт



<http://www.moah.org/exhibits/archives/brains/technology.html>

# Эволюция архитектуры СХД

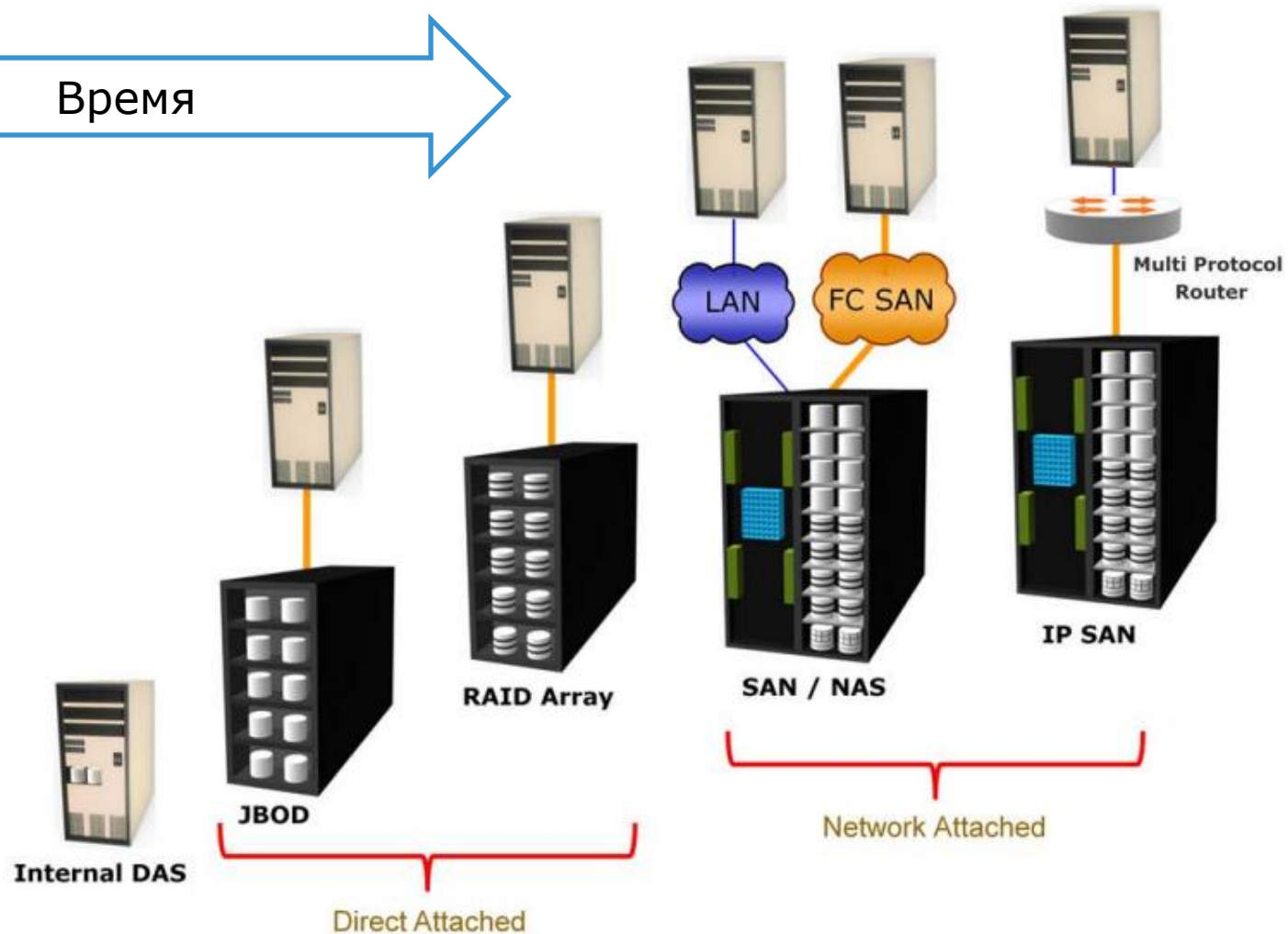


# Решаемые архитектурные проблемы

- Рост количества серверов
- Незащищенность
- Неуправляемость
- Увеличение расходов



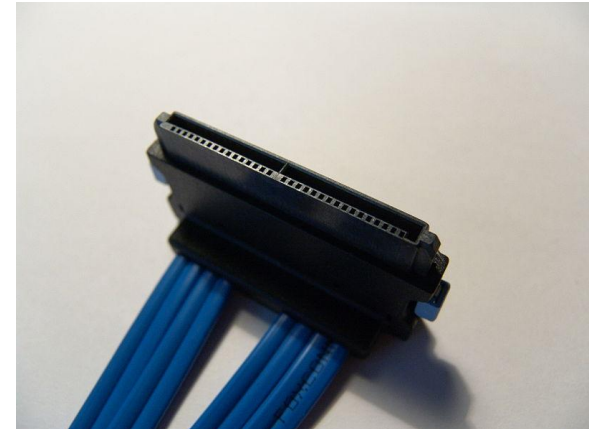
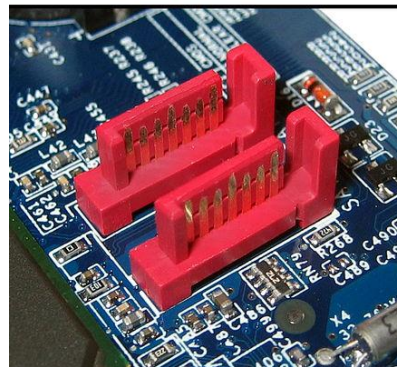
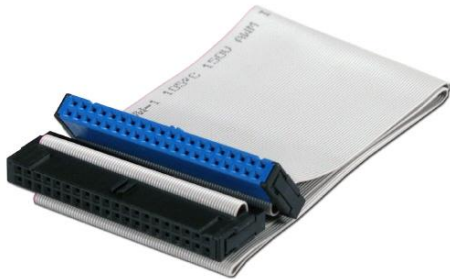
# Эволюция технологий СХД





# DAS – Directly Attached Storage

- Система хранения прямого подключения
- Встроенная или внешняя
- Протоколы - IDE/ATA, SATA, SCSI, SAS, FC



# RAID – Redundant Array of Independent Disks

- Избыточный массив (недорогих :) независимых дисков
- Повышение скорости доступа или надежности хранения
- Уровни – 0, 1, 3, 4, 5, 6, 1+0, 0+1 ..



# NAS – Network Attached Storage

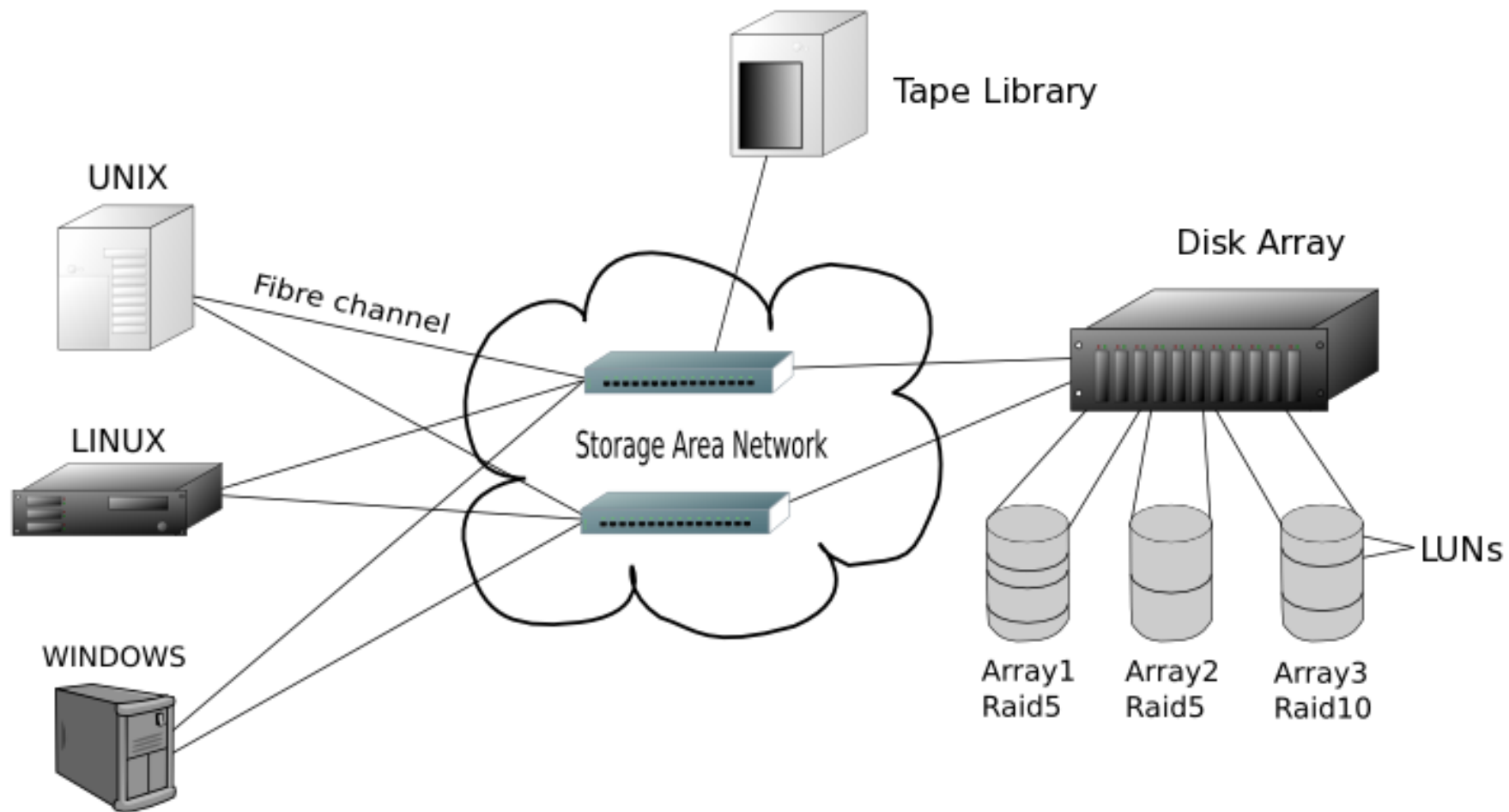
- Сетевое устройство хранения данных
- Доступ по локальной сети (TCP/IP)
- Протоколы передачи (файлов) - CIFS/NFS..
- Отличия от обычных файловых серверов:
  - Специализация
  - Производительность
  - Расширяемость
  - Надежность



# SAN – Storage Area Network

- Сеть хранения данных
- Специальное высокопроизводительное оборудование (СХД, кабели, карты, коммутаторы)
- FC и FCoE – физическая среда
- SCSI – протокол передачи данных (блоков)

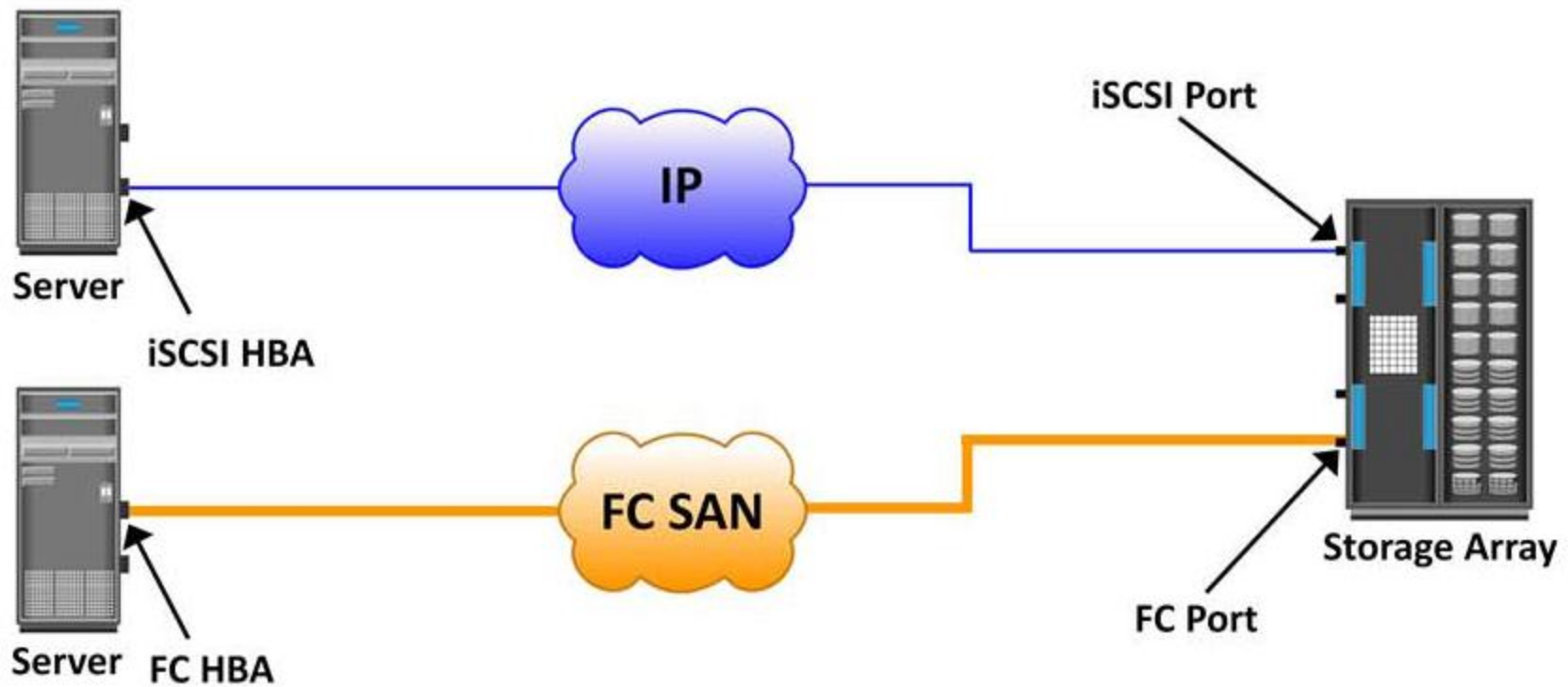
# SAN наглядно



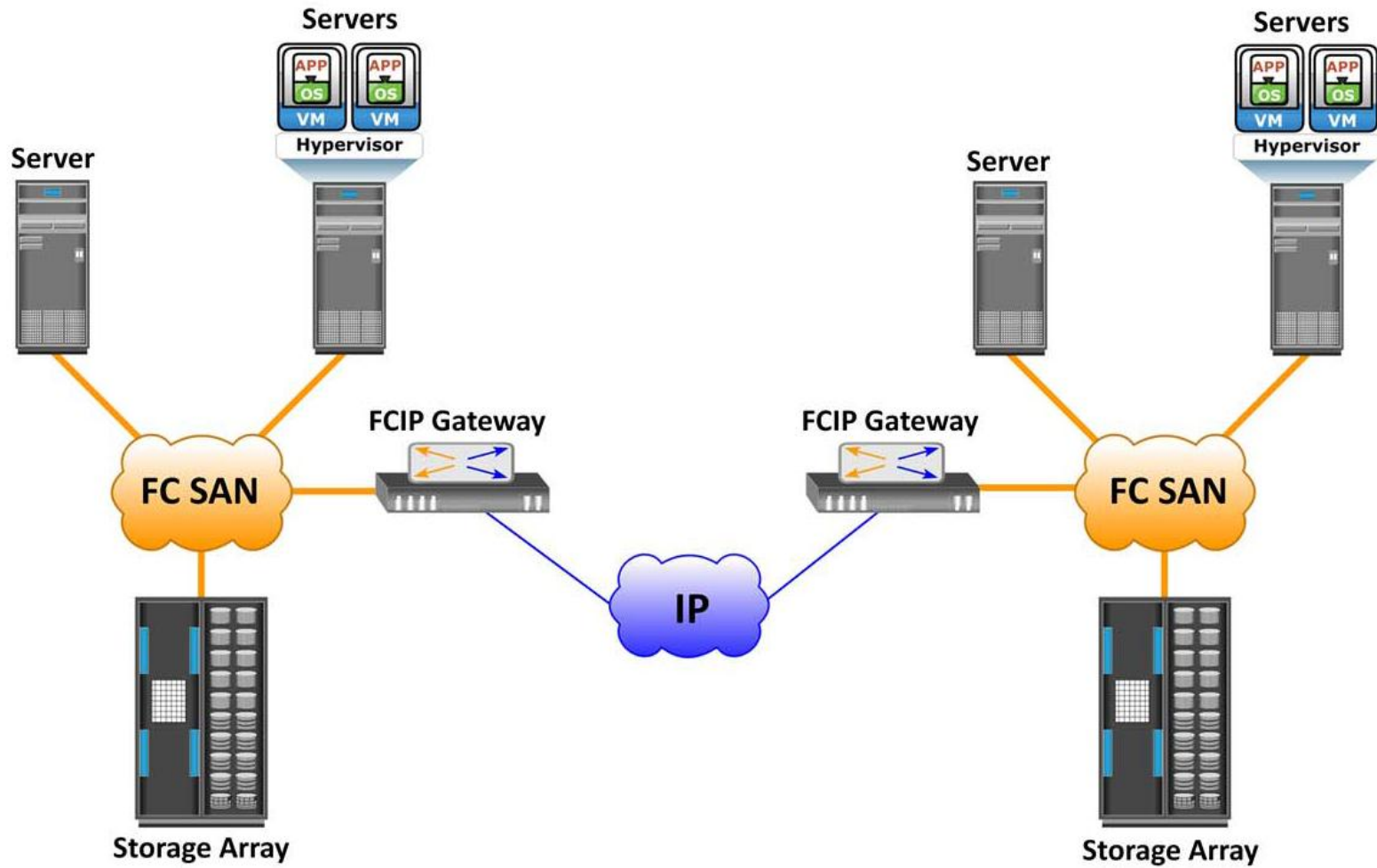
# IP SAN

- SAN через IP сети
- iSCSI, FCIP
- Большие расстояния чем в дорогих SAN
- Дешевле оборудование

# iSCSI



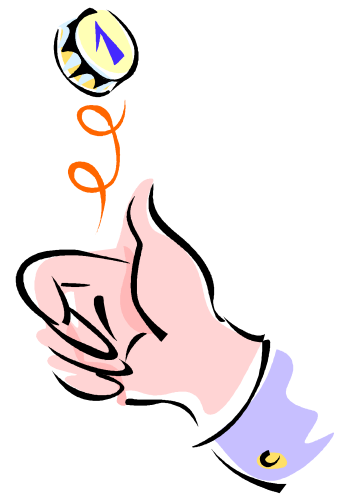
# FCIP





# FCoE

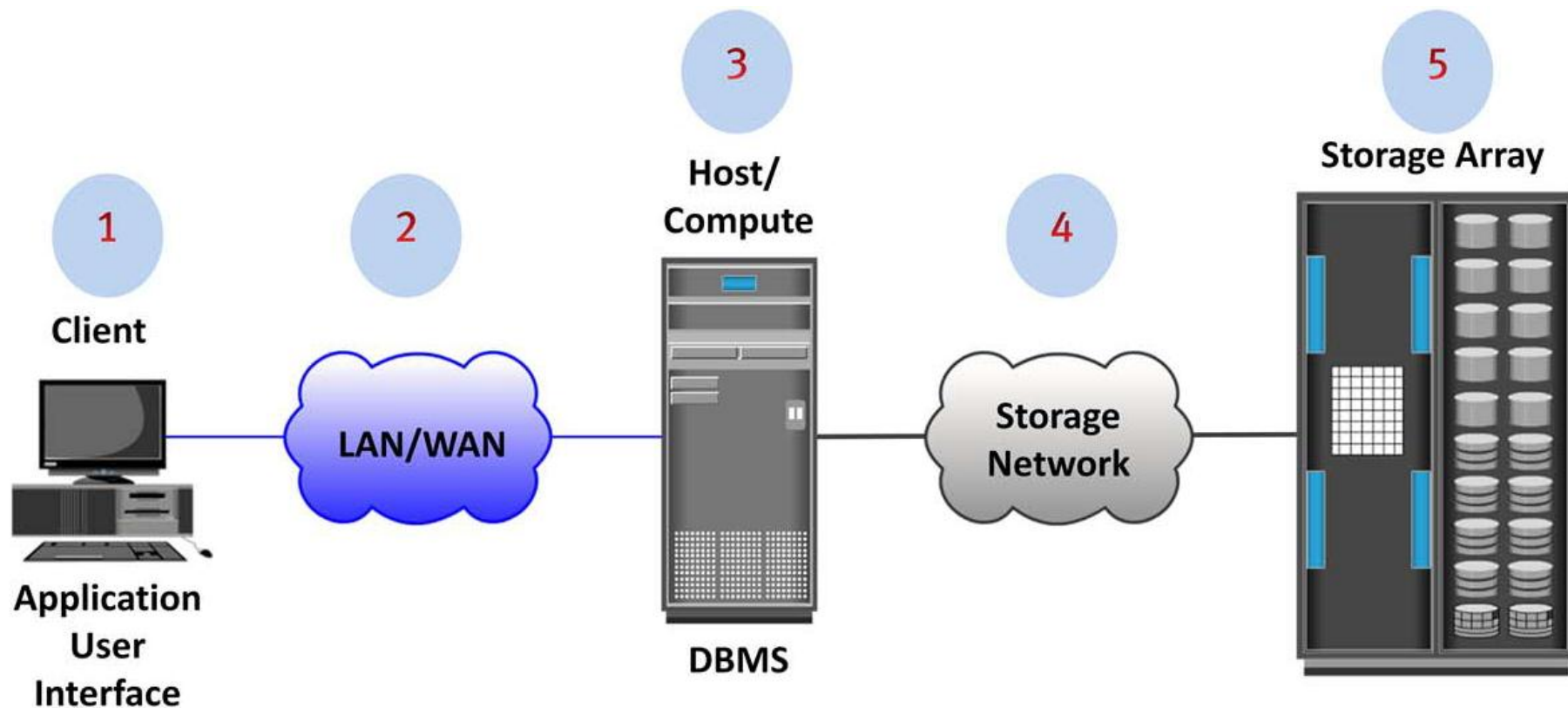
- Одна сеть для всего – мин. в 2 раза меньше соединений и проводов
- Lossless Ethernet
- Серверы видят два адаптера (две сети) – интернет и хранения данных
- Отлично подходит для «облачных» ЦОД



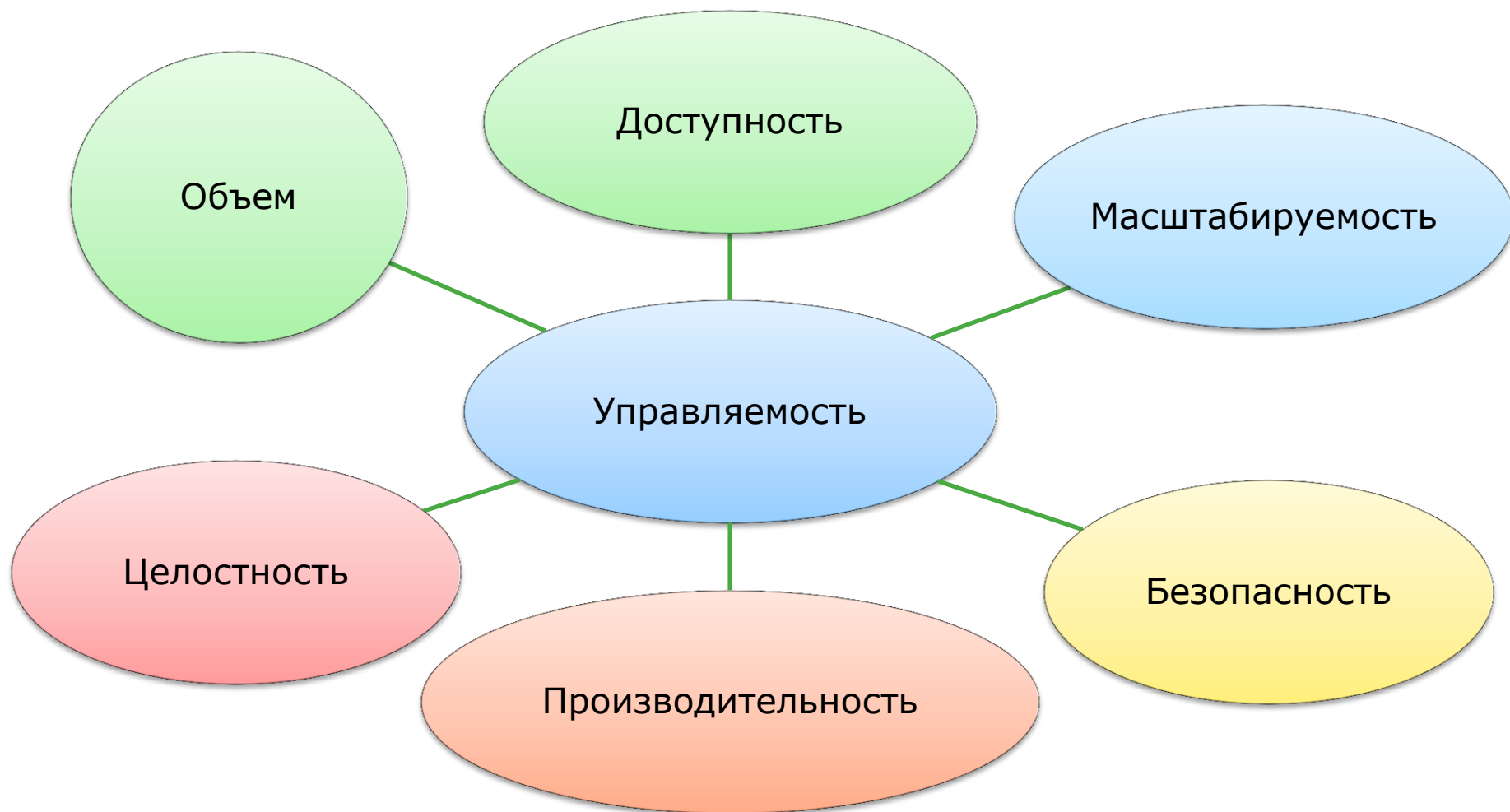
# Архитектура ЦОД

- ЦОД = центр обработки данных, информационный центр, data center
- Специализированное место (зал, здание, комплекс), объединяющее в себе серверы, СХД, сети, коммутаторы и другое оборудование для централизованного решения задач по хранению и обработке данных

# ЦОД - элементы



# Характеристики ЦОД



# Ключевые проблемы управления информацией

- Стремительный рост объемов
- Усиление информационной зависимости
- Изменение ценности во времени

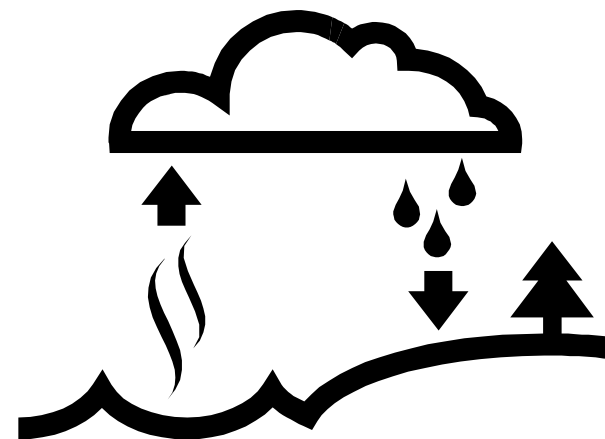


# Жизненный цикл информации

- Жизненный цикл информации – изменение ценности и других свойств с течением времени
  - Прогноз погоды
  - Гарантийные обязательства
  - Фотографии
  - ..
- Information Lifecycle Management (ILM) – управление жизненным циклом информации

# ILM

- Стратегия управления данными
- Ориентация на бизнес
- Централизация
- Целостность
- Гетерогенность
- Оптимизация ресурсов



# Реализация ILM

- Классификация данных и приложений
- Выполнение стратегии (от создания до удаления)
- Управление процессами
- Организация уровневого хранения





# Этапы внедрения ILM

1. Установка СХД
2. Классификация данных (уровни, циклы, политики)
3. Перемещение «вручную»
4. Автоматизация основных процессов
5. Полное внедрение для всех приложений

# Многоуровневое хранение

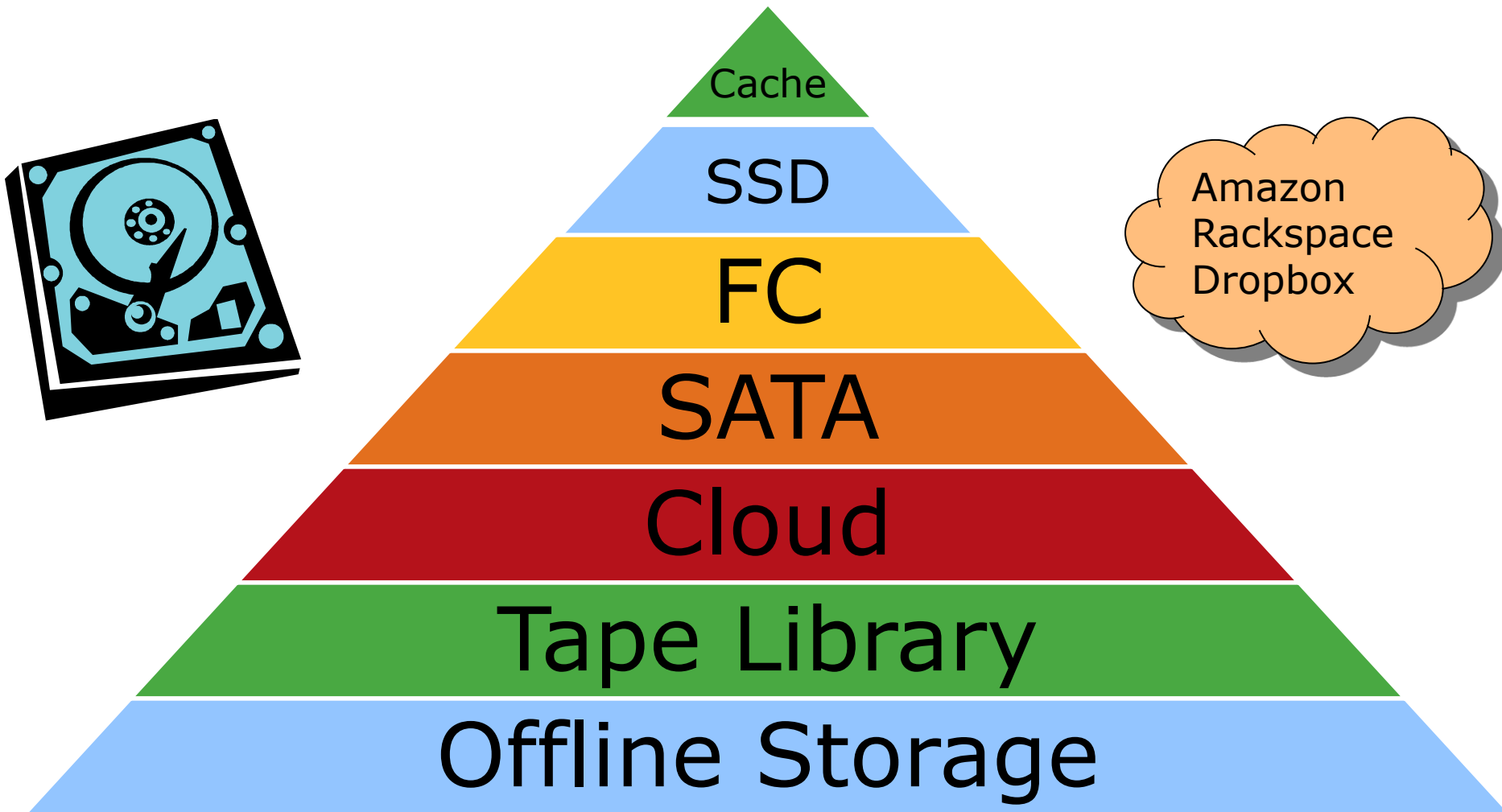
- Иерархическое хранение информации
- FAST – Fully Automated Storage Tiering

СХД

данные



# Иерархии уровней хранения



# Спасибо!

EMC<sup>2</sup>®