

# Лекция – 2

## Системы хранения данных и Центр обработки данных

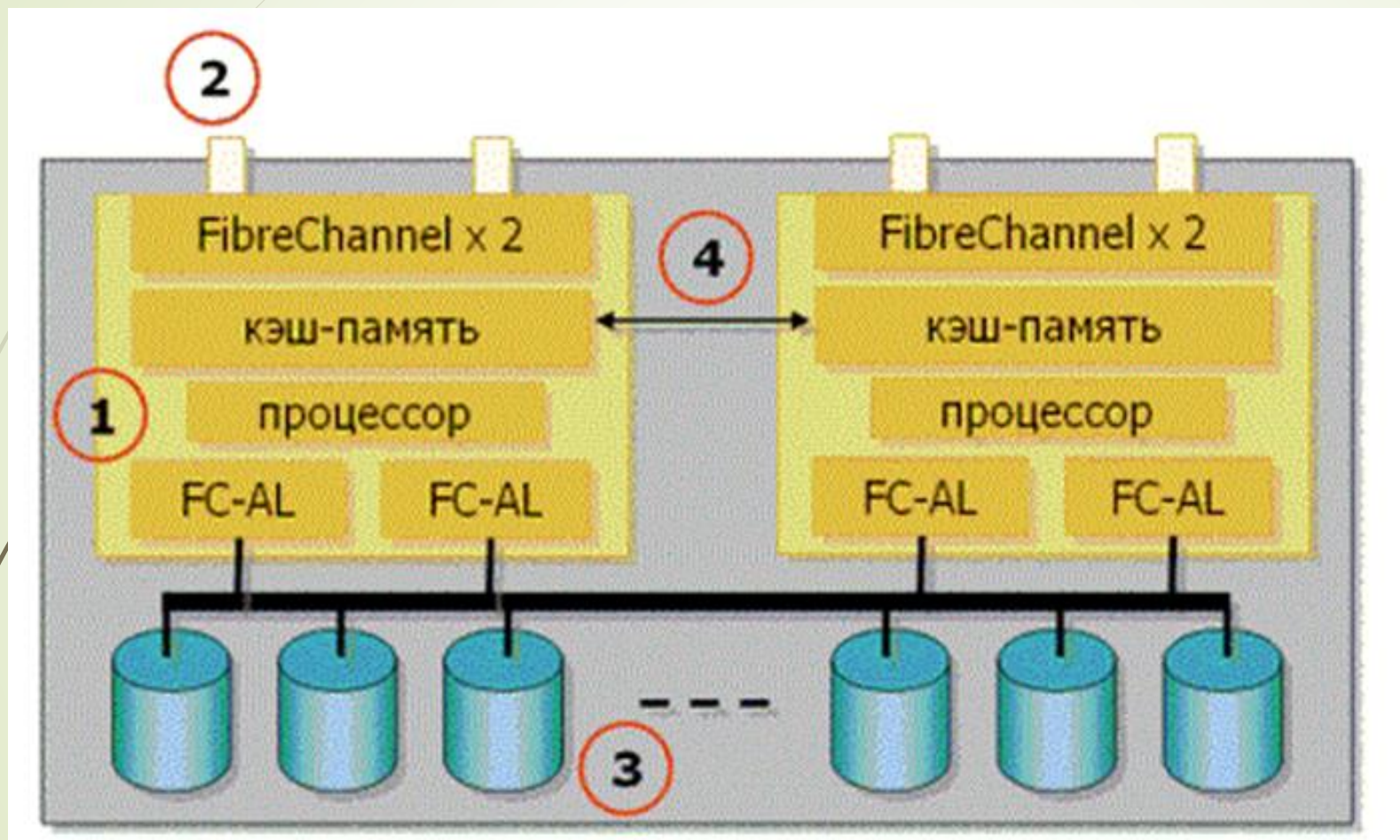
## Что такое система хранения данных?

- **Система хранения данных** (или просто **СХД**) – это так называемый «конгломерат» специализированного оборудования и программного обеспечения, который предназначен для хранения и передачи больших массивов информации. Система хранения данных позволяет организовать хранение информации на дисковых площадках с оптимальным распределением ресурсов.
- Как правило, СХД содержит следующие подсистемы и компоненты:
  - устройства хранения (дисковые массивы, ленточные библиотеки)
  - инфраструктуру доступа к устройствам хранения
  - подсистему резервного копирования и архивирования данных
  - программное обеспечение управления хранением
  - систему управления и мониторинга.

# Система хранения данных (СХД)

- Это особый ресурс информационного центра
  - Данные направляются из приложения к устройству хранения посредством различных компонентов, которые все вместе называются средой хранения данных.
  - Выделяют три основных компонента этой среды
    - хост, система связи (система передачи данных),
    - система связи и
    - устройство хранения данных (хранилище).

## Пример структурной схемы СХД:



- 1. Контроллер
- 2. Внешний интерфейс
- 3. Жёсткие диски
- 4. Кэш-память



## Классификация СХД по архитектурному признаку

- Единственно верным критерием для классификации СХД является программная архитектура, используемая в том или ином решении, поскольку от этого зависят все основные характеристики системы. Остальные компоненты СХД зависят от того, какая именно программная архитектура была выбрана разработчиками.
- Стоит понимать, что есть лучшие решения для каких-то конкретных ситуаций или видов нагрузки, но не существует универсального идеального решения.

## Выделяют 4 типа архитектур СХД

- Кластеризованная архитектура;
- Слабо связанные архитектуры;
- Сильно связанные архитектуры;
- Распределённая архитектура без общих элементов.

# 1. Кластеризованная архитектура

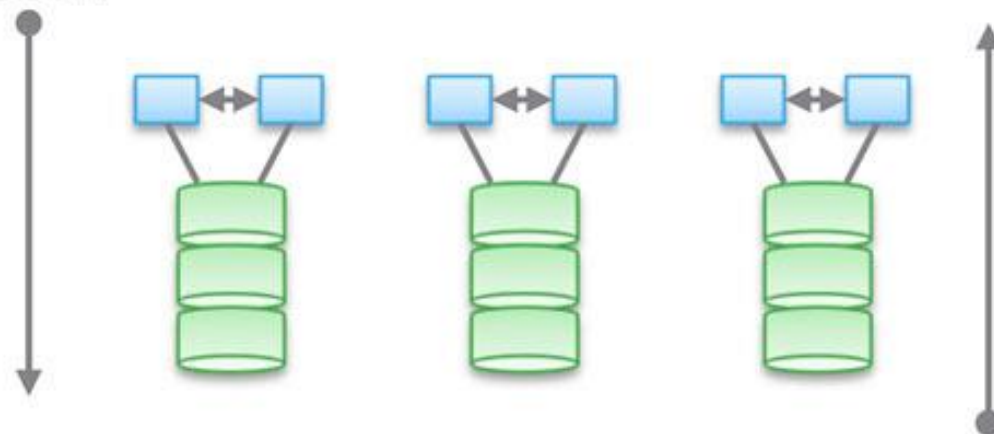
(вертикально масштабируемые решения)

СХД может быть:

- A/A (оба контроллера активны);
- A/P (один контроллер пассивен);
- A/S (каждый контроллер обрабатывает свой объём данных).

Низкие задержки  
при записи  
(нераспределённый  
ввод-вывод)

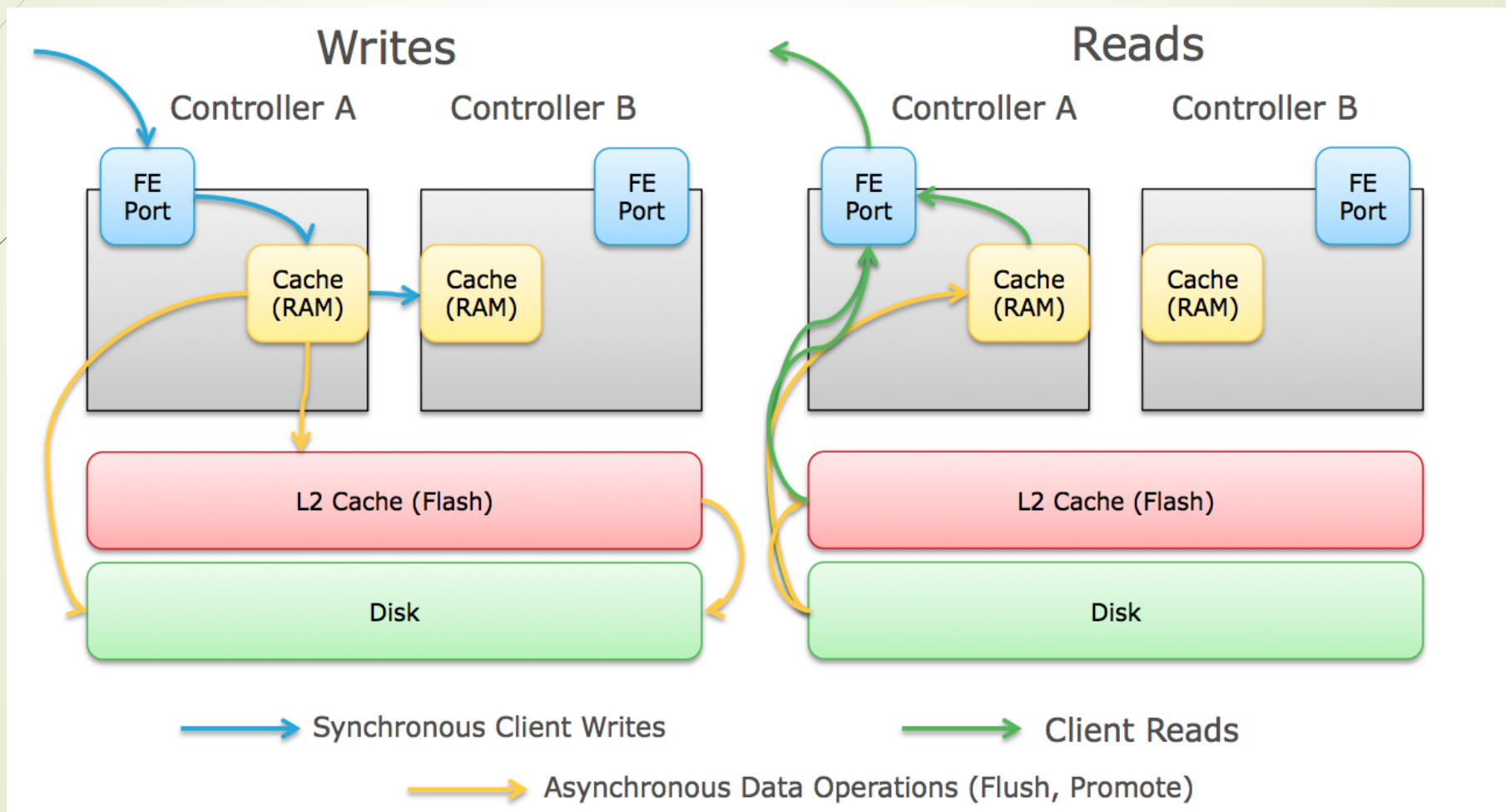
«Балансировка» с помощью  
мобильности виртуального  
сервера и миграции



Данные всегда находятся  
на одной управляющей машине,  
распараллеливаемость невысока

# 1. Кластеризованная архитектура

(вертикально масштабируемые решения)





# 1. Кластеризованная архитектура

(вертикально масштабируемые решения)

## **Плюсы кластеризованной архитектуры:**

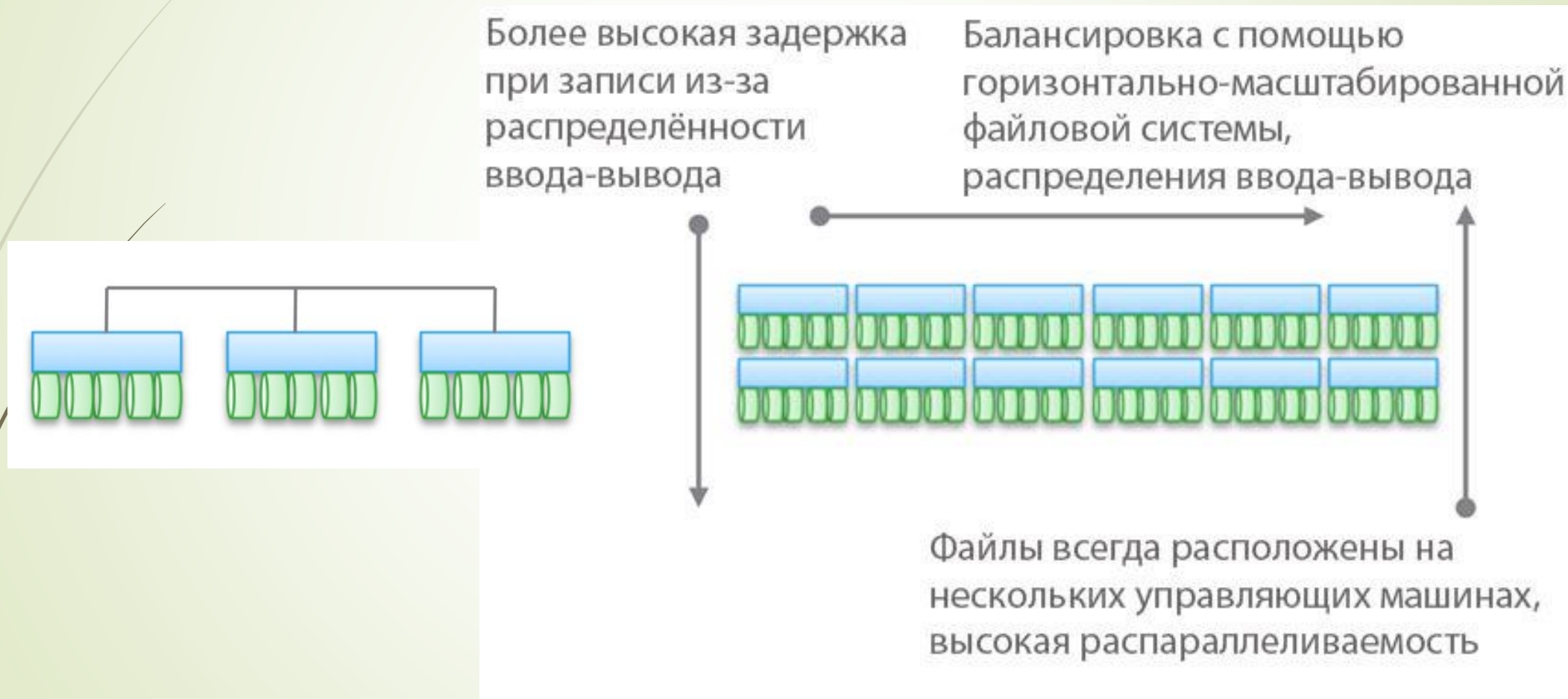
- Задержка между контроллерами практически нулевая;
- Скорость работы зависит от производительности дисков;
- Так как задержки в таких решениях минимальны, то такие СХД идеальны для транзакционной модели доступа к данным;
- Архитектуру легко поддерживать и настраивать.

## **Минусы кластеризованной архитектуры:**

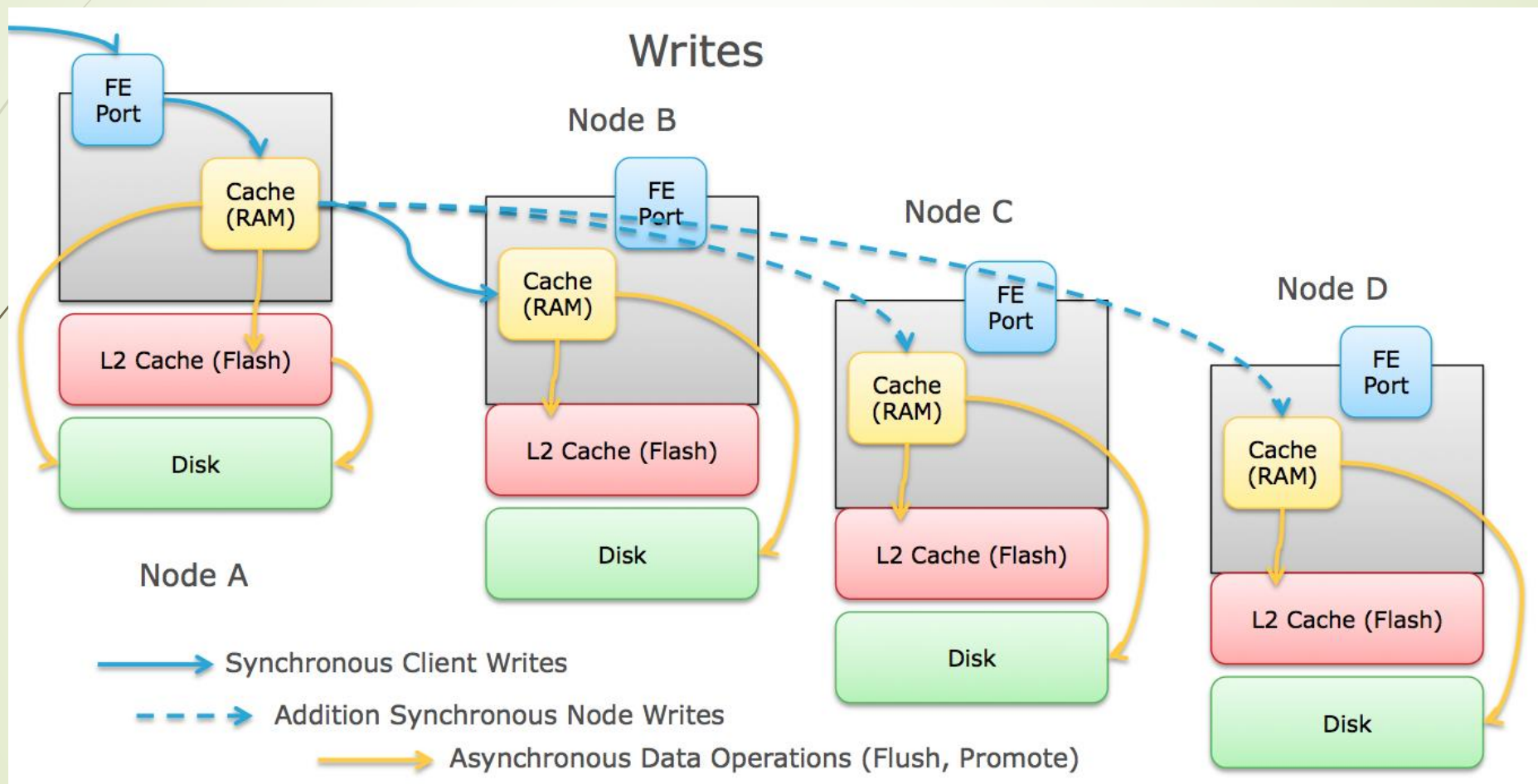
- Масштабируемость ограничена мощностью контроллера;
- Доступность данных гарантируется переключением между контроллерами, что может привести к понижению производительности при выходе одного контроллера из строя.

**Примеры реализации:** EMC VNX, NetApp FAS, Pure Storage, Tintri, Nimble и Nexenta.

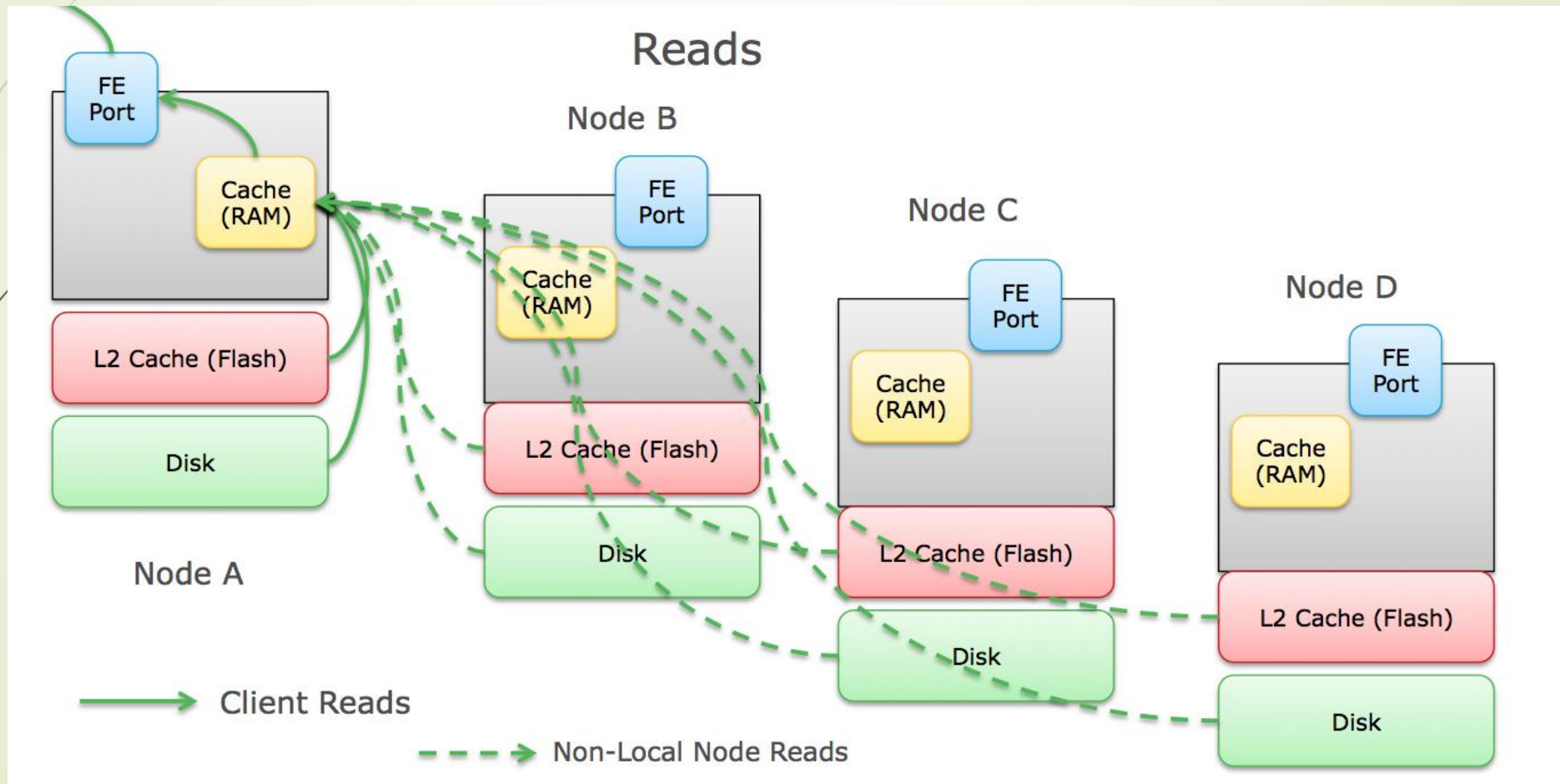
## 2. Слабо связанные архитектуры (горизонтально масштабируемые решения)



## 2. Слабо связанные архитектуры (горизонтально масштабируемые решения)



## 2. Слабо связанные архитектуры (горизонтально масштабируемые решения)





## 2. Слабо связанные архитектуры

(горизонтально масштабируемые решения)

### **Плюсы слабо связанной архитектуры:**

- Задержки на запись практически нулевые от клиента к массиву – основные задержки вызывает копирование между нодами;
- Быстрый локальный кэш для минимизации задержек синхронизации;
- Отличная производительность при последовательном чтении;
- Практически не используются специализированные аппаратные решения.

### **Минусы слабо связанной архитектуры:**

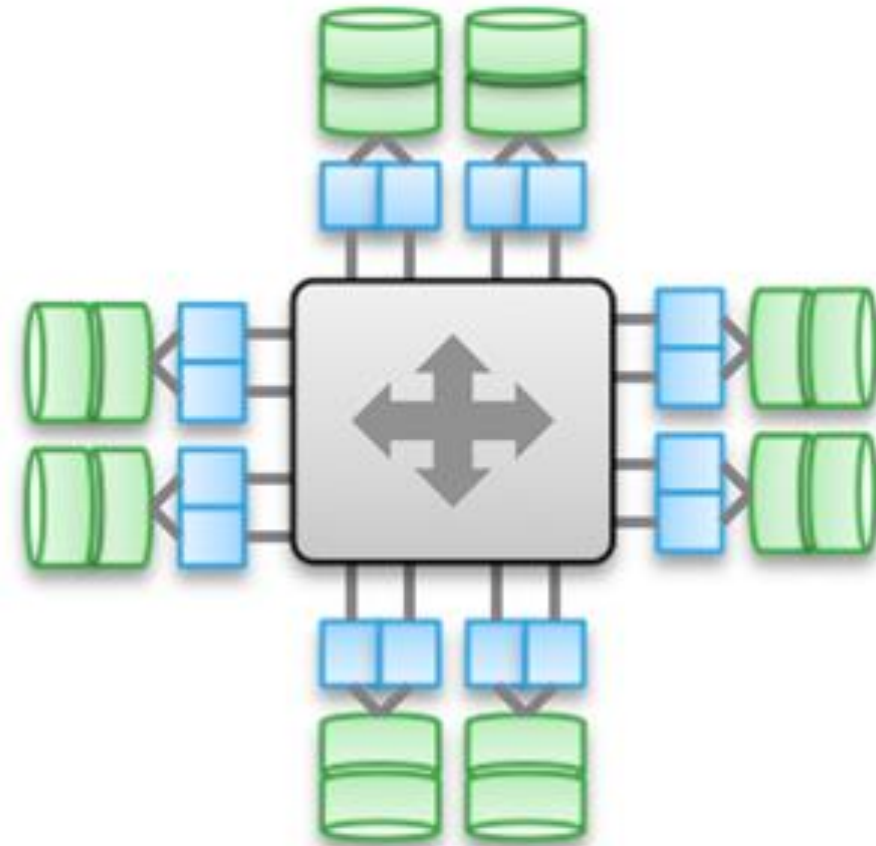
- Чем больше нодов, тем больше ресурсов требуется на координацию;
- Создание дискового пула является сложной фоновой задачей;
- Некоторые приложения не могут использовать все преимущества такой архитектуры, что приводит к высокой локальной нагрузке.

**Примеры реализации:** EMC Isilon, Dell Equallogic, VMware VSAN, IBM SONAS.

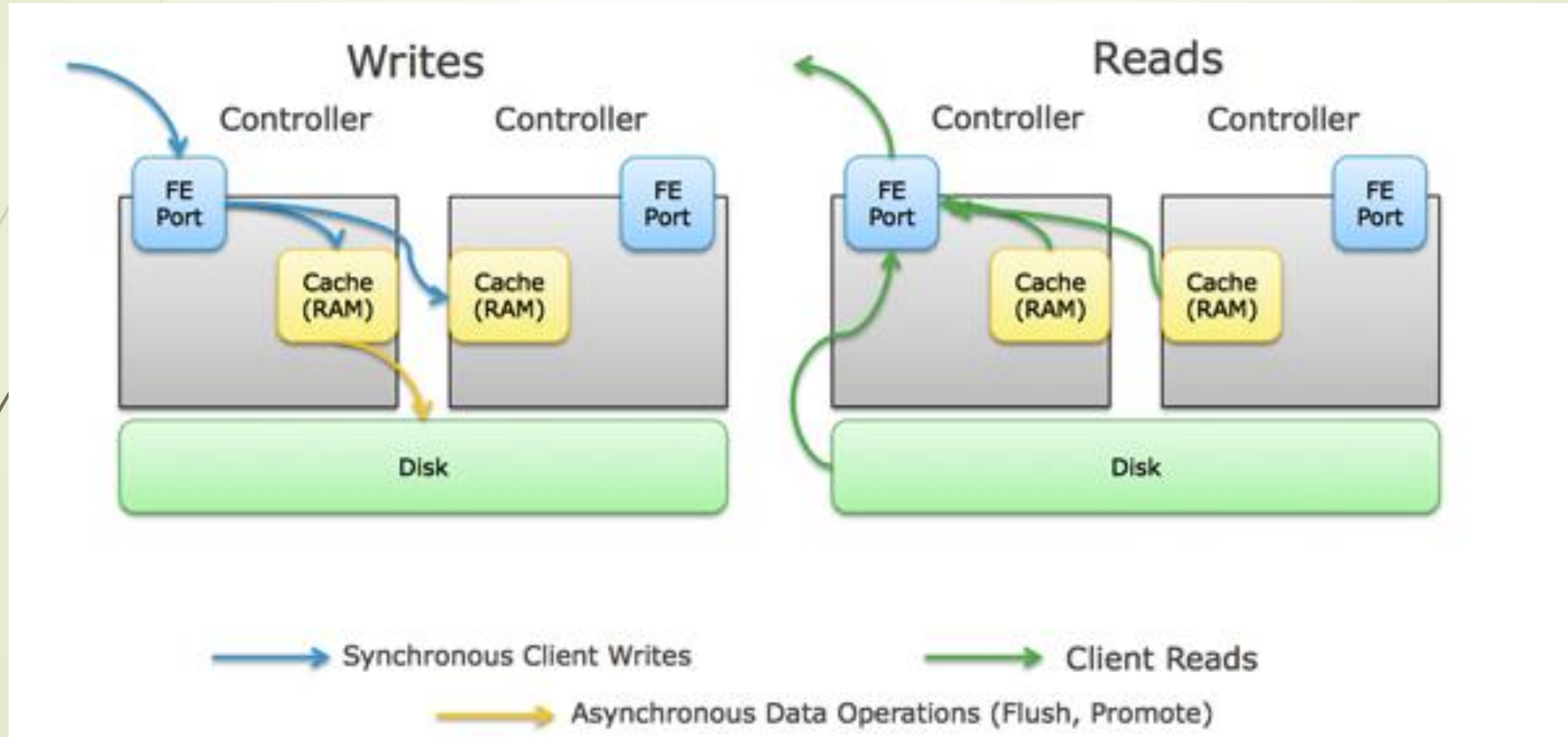


### 3. Сильно связанные архитектуры (горизонтально масштабируемые решения)

- ▶ В сильно связанной архитектуре СХД используется общая память между контроллерами (кэш или метаданные), а сами данные разнесены по разным компонентам.
- ▶ Архитектура разрабатывалась так, чтобы при выходе из строя любого элемента (запланированном или нет) операции ввода-вывода оставались относительно сбалансированными.



### 3. Сильно связанные архитектуры (горизонтально масштабируемые решения)



### 3. Сильно связанные архитектуры

(горизонтально масштабируемые решения)

#### **Плюсы сильно связанной архитектуры:**

- Задержки между контроллерами практически нулевые;
- Задержки записи данных зависят в основном от физических дисков;
- Идеально подходит для высоконагруженных систем;
- При ограничении масштабируемости гарантируется производительность.

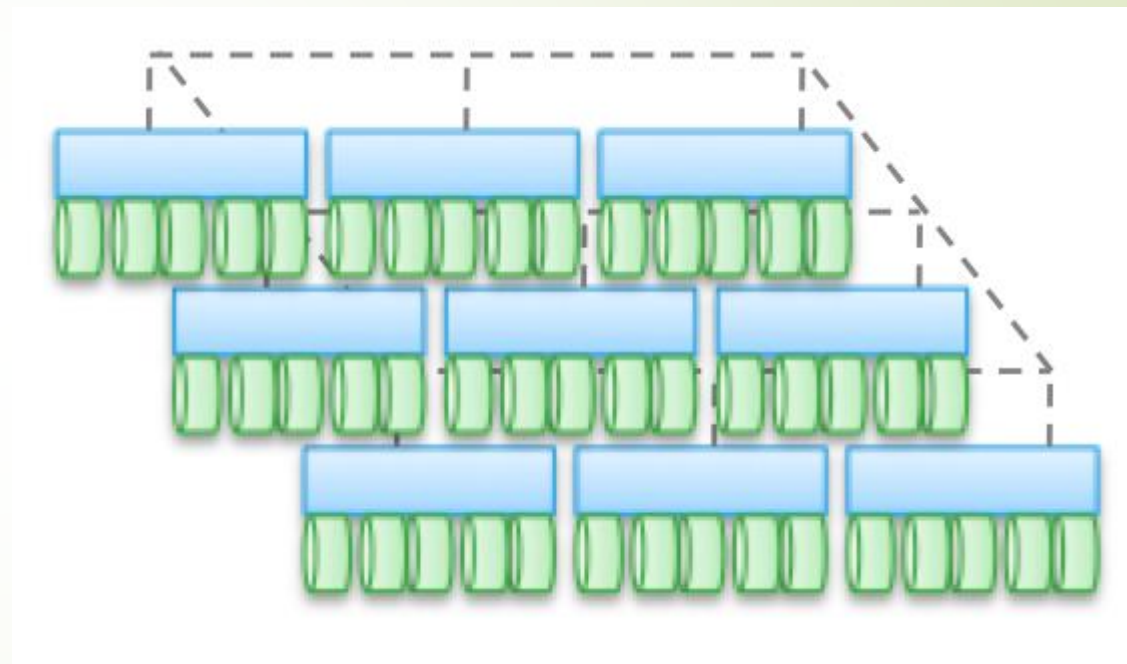
#### **Минусы сильно связанной архитектуры:**

- Масштабируемость ограничена максимальным количеством контроллеров;
- Для архитектуры характерна высокая сложность программного кода, которая влияет на количество добавляемых служб обработки данных, поскольку это представляет собой трудную вычислительную задачу.

**Примеры реализации:** EMC Symmetrix VMAX, HDS VSP, HP 3Par.

## 4. Распределённая архитектура (без совместного использования ресурсов)

- В распределённых системах не используется общая память, данные разнесены по разным компонентам.
- Данные записываются на один компонент, и с определённой периодичностью копируются на другие компоненты для обеспечения защищённости.
- СХД с распределённой архитектурой не транзакционны.

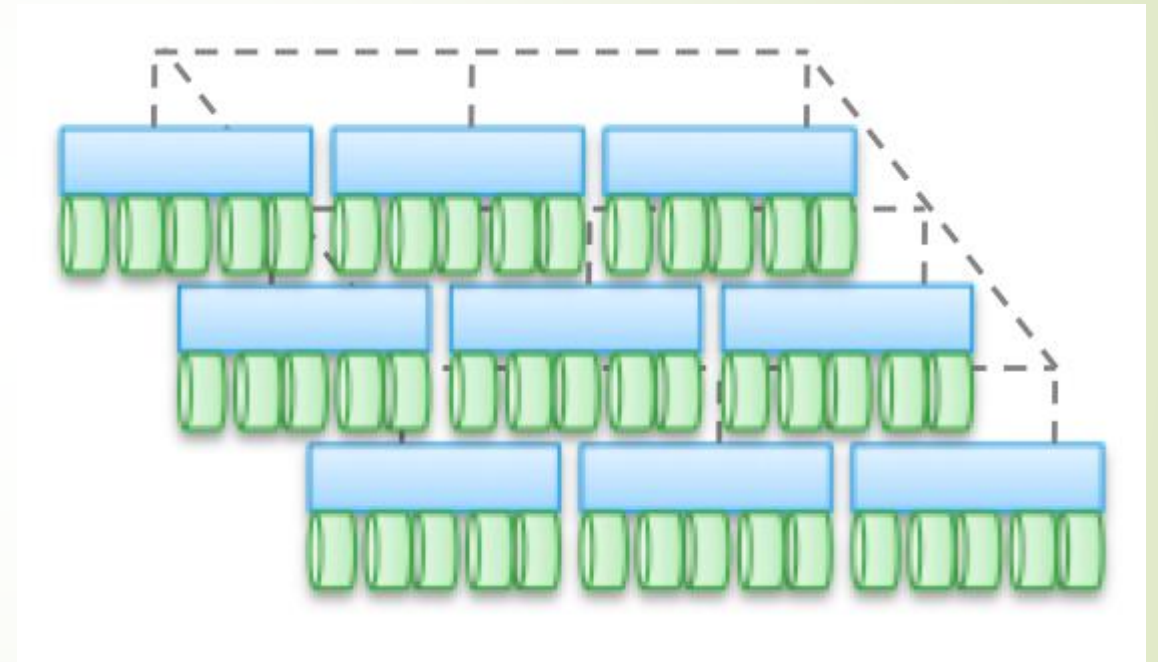




## 4. Распределённая архитектура

(без совместного использования ресурсов)

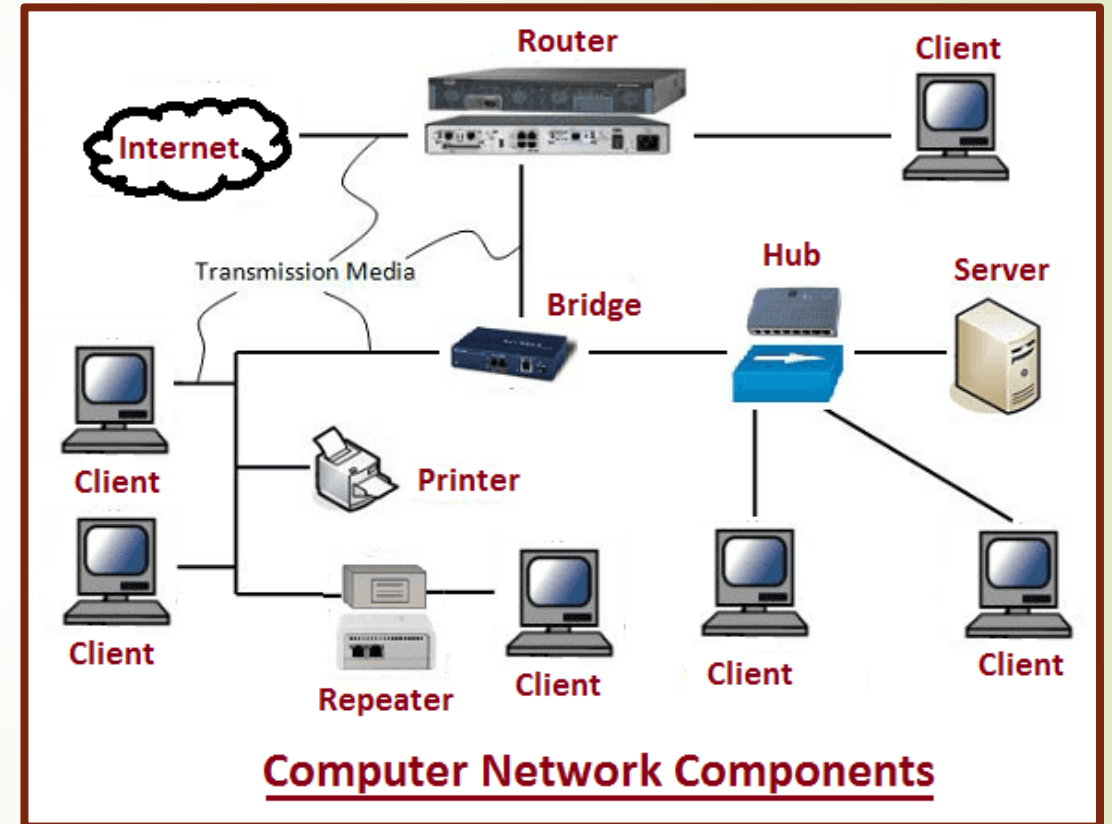
- Связь между компонентами практически всегда осуществляется посредством Ethernet (это просто, универсально и дёшево), при этом система не является отказоустойчивой.
- Архитектура является самой масштабируемой из всех перечисленных.
- Архитектура не зависит от оборудования и может быть развёрнута на любом «железе».
- Примеры реализации: Haystack, Atmos, ViPR, Ceph, Swift, HDFS.





# Основные сетевые компоненты

1. Абонентские системы
2. Сетевое оборудование
3. Коммуникационные каналы



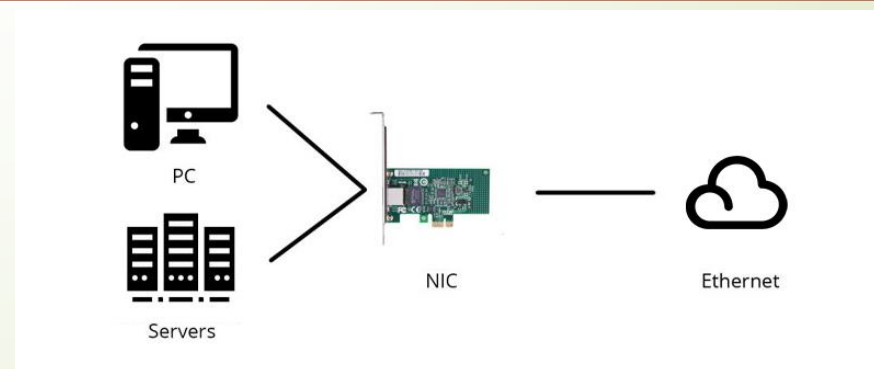
# HBA & NIC

20

**Хост-адаптер шины** (HBA, host bus adapter) - вид компьютерных комплектующих: плата адаптера, устанавливаемая в компьютер и служащая для подключения накопителей (устройств хранения информации) или сети, имеющих в качестве интерфейса шинную организацию, отличную от имеющихся в компьютере изначально.

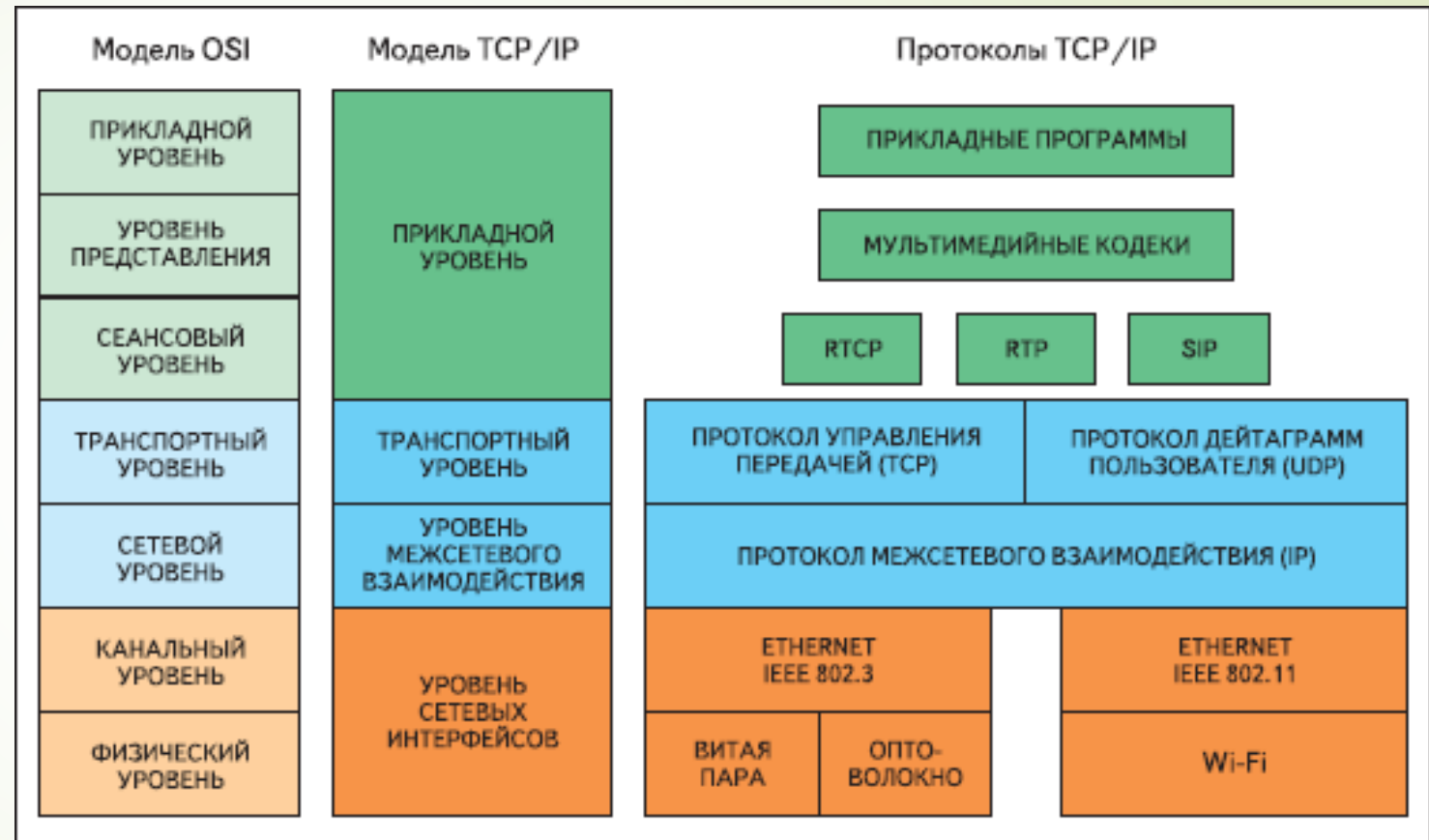


**Сетевая плата** (network interface controller), также известная как сетевая карта, сетевой адаптер (в терминологии компании Intel, Ethernet-адаптер — по названию технологии — дополнительное устройство, позволяющее компьютеру взаимодействовать с другими устройствами сети. В настоящее время в персональных компьютерах и ноутбуках контроллер и компоненты, выполняющие функции сетевой платы, довольно часто интегрированы в материнские платы для удобства, в том числе унификации драйвера и удешевления всего компьютера в целом.



# OSI vs. TCP/IP

- Модель OSI устанавливает глобальный стандарт, определяющий состав функциональных уровней при открытом взаимодействии между компьютерами
- Стек протоколов TCP/IP — набор сетевых протоколов, на которых базируется интернет



# Виды протоколов

22

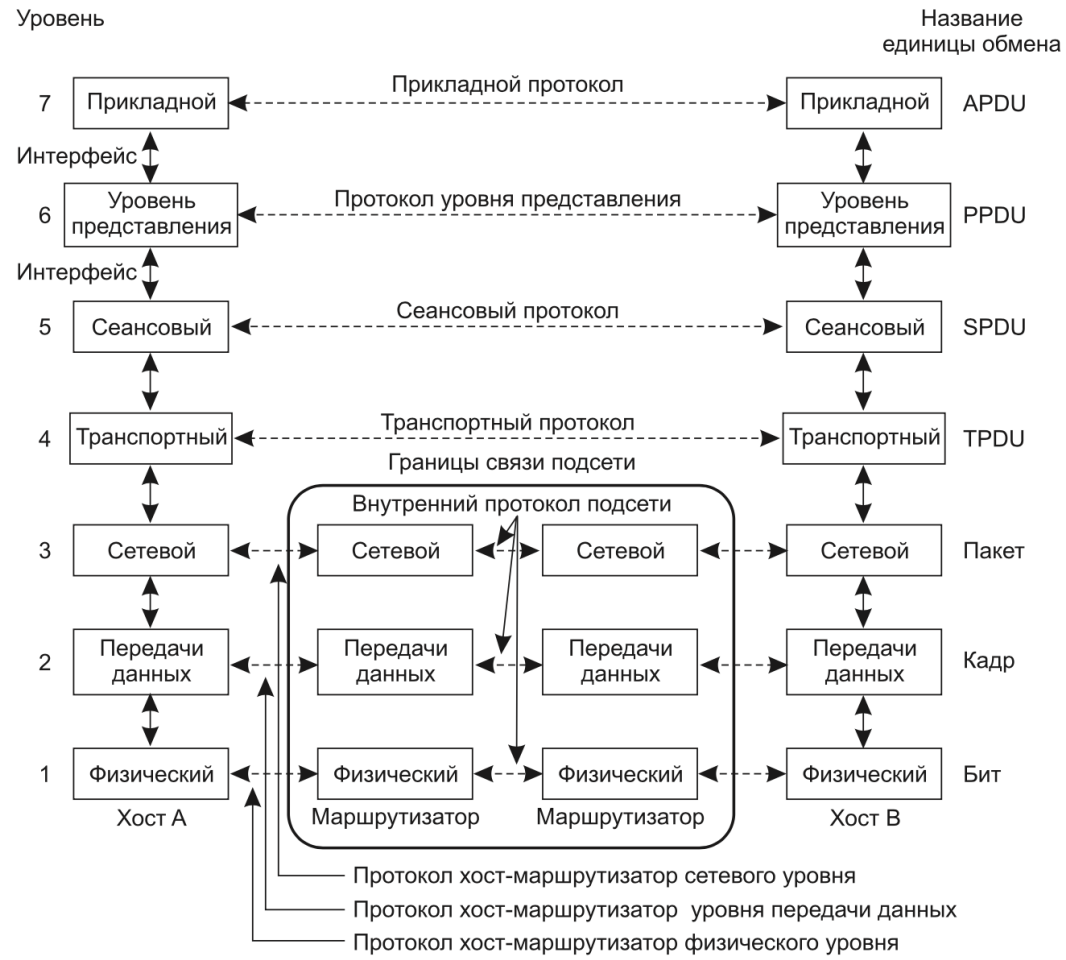


Рис. 1.17. Эталонная модель OSI

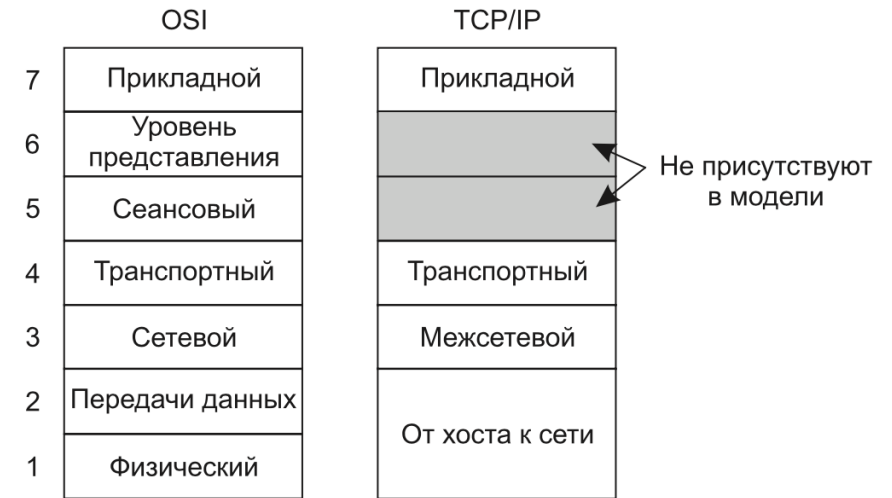


Рис. 1.18. Эталонная модель TCP/IP

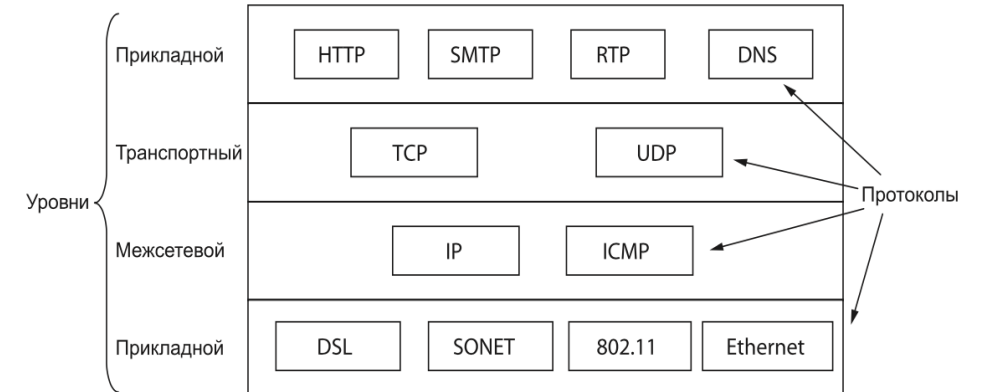


Рис. 1.19. Протоколы и сети в модели TCP/IP



# Сетевые операционные системы

- Комплекс программ, обеспечивающих в сети обработку, хранение и передачу данных. Для организации сети кроме аппаратных средств, необходима также сетевая операционная система. Операционные системы сами по себе не могут поддерживать сеть. Для дополнения какой-нибудь ОС сетевыми средствами необходима процедура инсталляции сети.

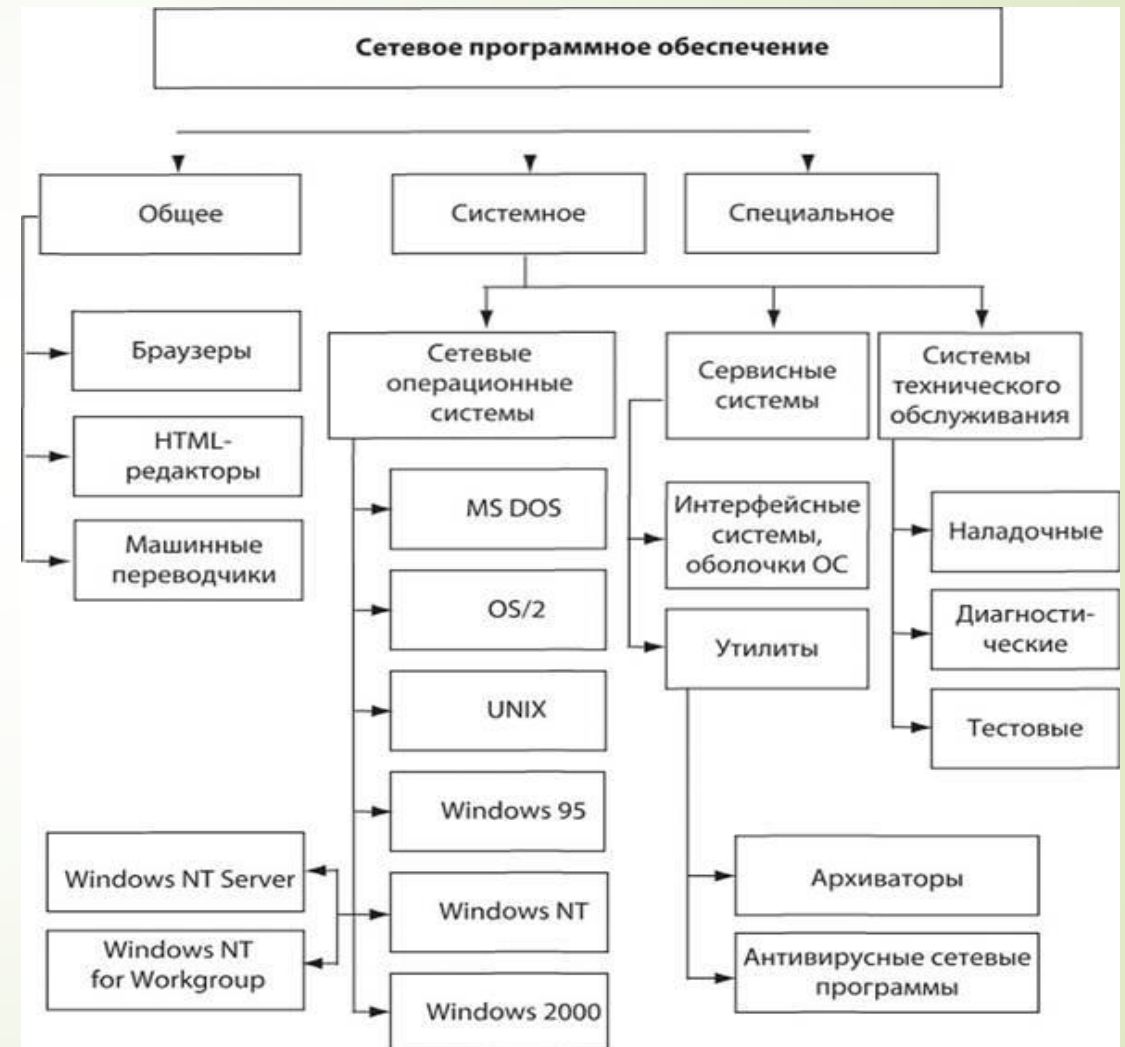
Windows NT;  
Windows for  
Workgroups;  
LANtastic;  
NetWare;  
Unix;  
Linux и т.д.





# Сетевое программное обеспечение (Сетевые службы)

- клиент сети
- сетевая карта
- протокол
- служба удаленного доступа



## Центр обработки данных (ЦОД)

Это- Совокупность оборудования, аппаратуры и сооружений, которая содержит в себе ресурсы хранения, обработки, и передачи данных по сети, а также другие IT-ресурсы для централизованного предоставления мощностей по обработке данных

- Ключевые элементы ЦОД
  - Приложение
  - Система управления базами данных (СУБД)
  - Хост или Сервер
  - Сеть
  - Система хранения данных
- Эти элементы работают вместе для соответствия требованиям по обработке данных

# Приложение

26

- Программное обеспечение, которое предоставляет логику для вычислительных операций
- Приложения, развертываемые в ЦОД:
  - Бизнес-приложения – email, системы электронного документооборота, системы принятия решений
  - Приложения для управления – управление ресурсами, настройка производительности, виртуализация
  - Приложения для защиты данных – бэкап, репликация
  - Приложения безопасности – аутентификация, антивирус
- Ключевые характеристики ввода-вывода приложения (I/O)
  - Интенсивное чтение или интенсивная запись (Read intensive vs. write intensive)
  - Последовательное или случайное (Sequential vs. Random)
  - Размер операции ввода-вывода (I/O size)

# Система управления базой данных (СУБД)

- База данных – структурированный подход к хранению данных в логически организованных, взаимосвязанных таблицах
  - Помогает оптимизировать хранение и получение данных
- СУБД контролирует создание, поддержку и использование баз данных
  - Обрабатывает запросы приложений на данные
  - Инструктирует ОС для получения подходящих данных от системы хранения
- Популярными СУБД являются: MySQL, Oracle RDBMS, SQL Server, etc.

**Хост** – Ресурс, который запускает приложения при помощи имеющихся вычислительных компонент

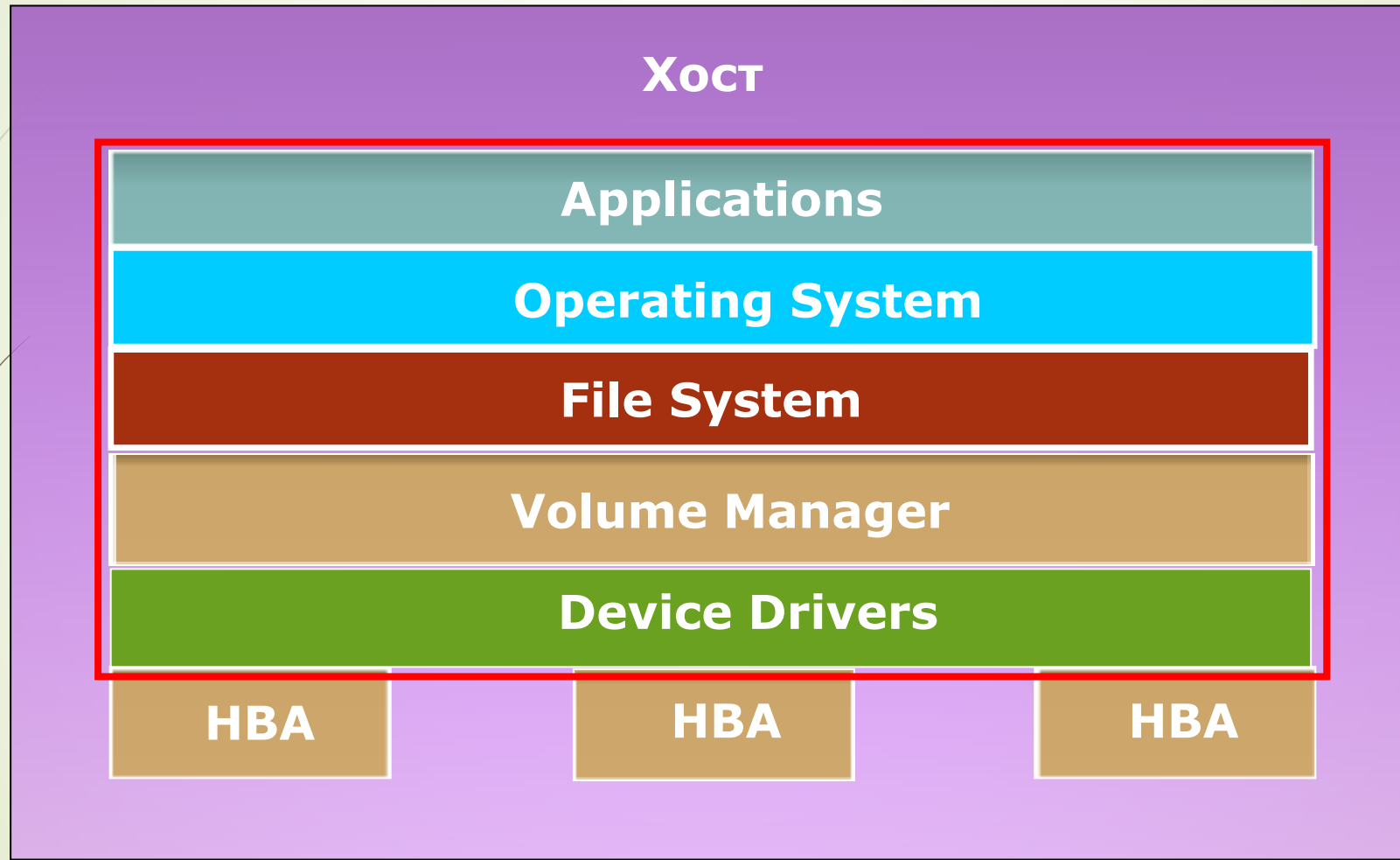
- Примеры: Мэйнфрейм, сервер, ноутбук, персональный компьютер, планшет, etc.
- Состоит из аппаратных (hardware) и программных (software) компонент
- Аппаратные (hardware) компоненты
  - CPU, память, устройства ввода-вывода
- Программные (software) компоненты
  - Операционная система, драйверы устройств, файловая система, менеджер логических томов, etc



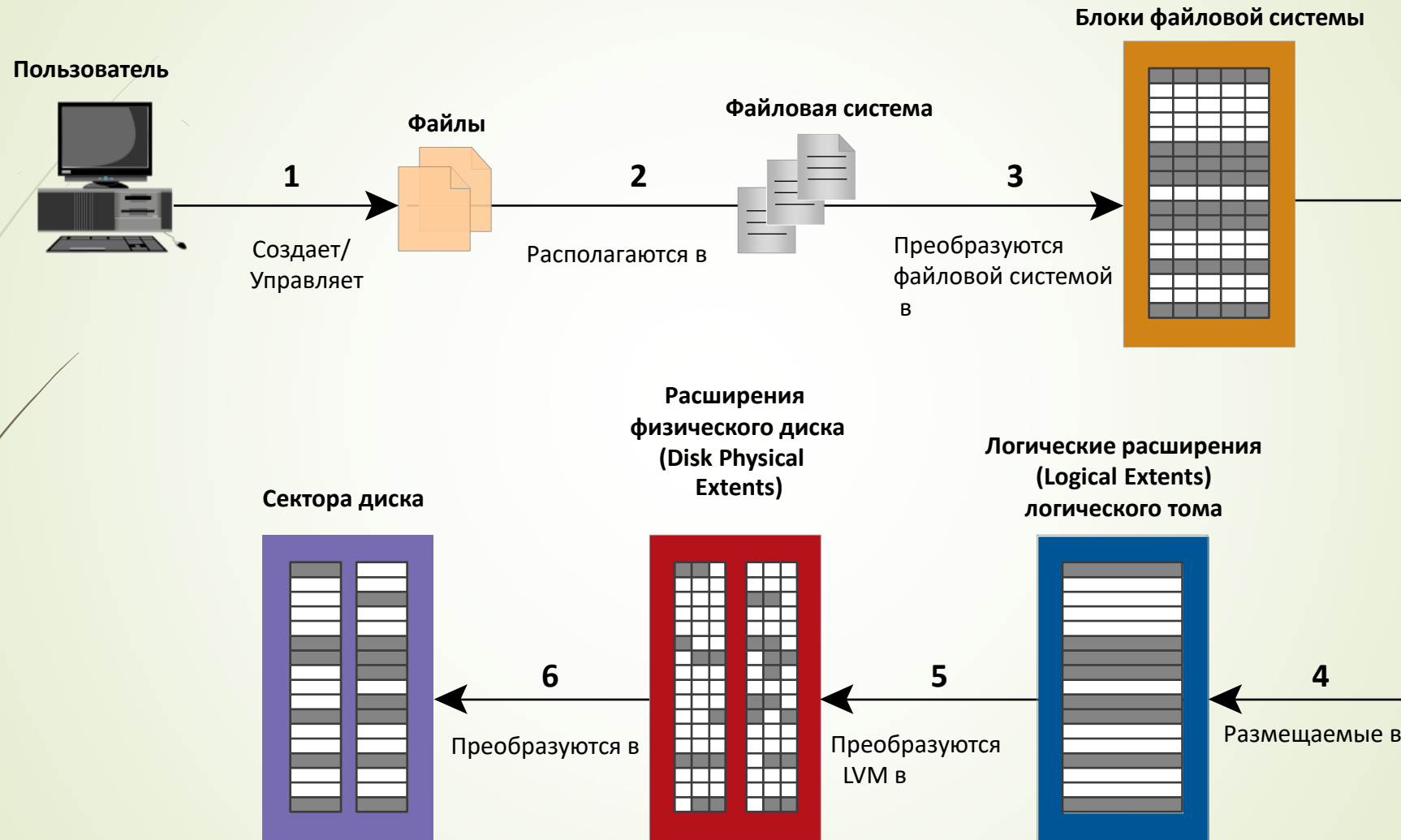


# Логические компоненты хоста

29

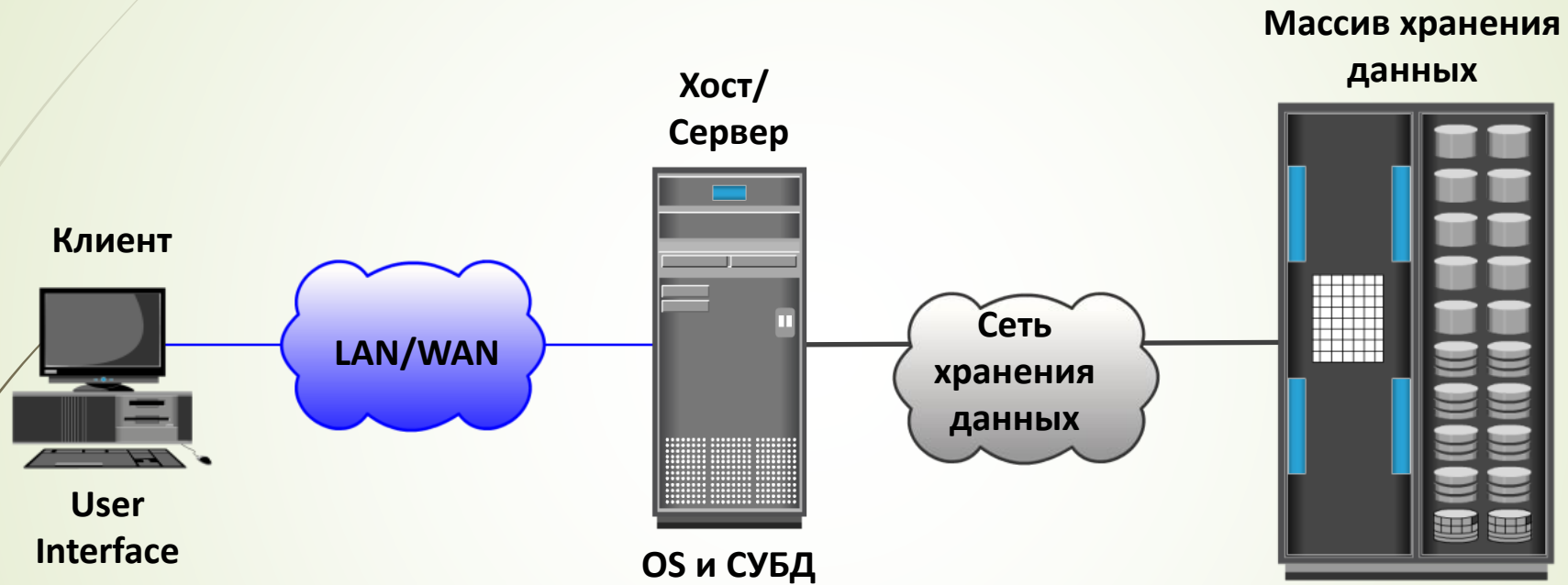


# Файловая система



# Структура ЦОД

Пример системы обработки онлайн-заказов



# Классический ЦОД (CDC)

## Ключевые элементы ЦОД

Приложение

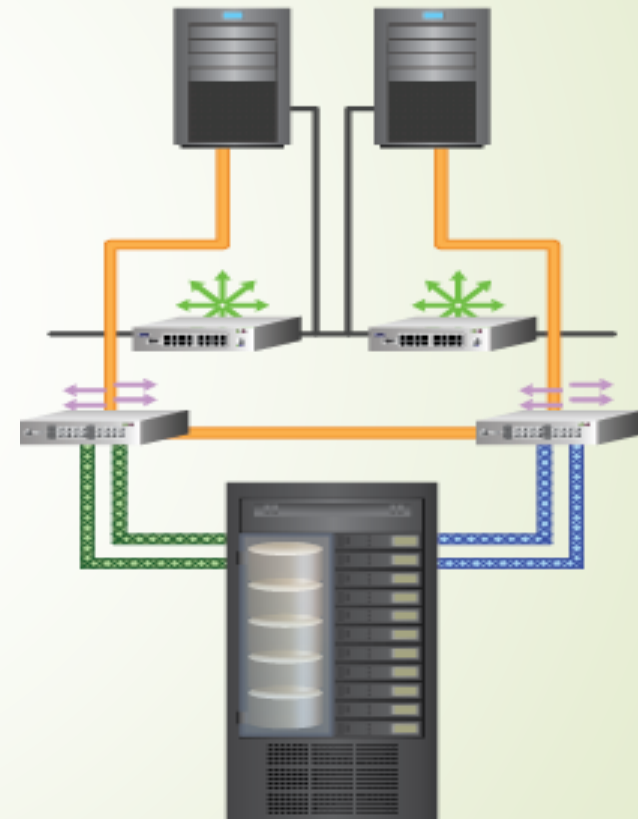
Система управления базами данных (СУБД)

Хост или Сервер

Система хранения данных

Сеть

Классический ЦОД





# Ключевые требования к ЦОД

33

- Доступность – учитывается время и удобство доступа для клиента
- Безопасность – как от несанкционированного доступа, так и от случайного доступа к ресурсам, не предназначенным данному клиенту
- Масштабируемость – операции ЦОД должны включать возможность находить дополнительные ресурсы в случае необходимости
- Производительность и объем – должна быть оптимальной для высокоскоростной обработки запросов.
- Целостность данных – наличие средств, гарантирующих правильное хранение и передачу данных.

УПРАВЛЯЕМОСТЬ может достигаться путем автоматизации и снижения роли человека при выполнении стандартных заданий

## ИТОГИ

Основные понятия, раскрытые в данной лекции:

- Система хранения данных (СХД)
- Архитектуры СХД
- Центр обработки данных (ЦОД)
- Классический ЦОД