Мультимедийные информационные системы (МИС)

Направление: 09.03.02 «Информационные системы и технологии»

17 декабря 2018 г.

Компьютерные системы хранения данных

shalaginov.com/2014/10/21/ликбез-№1-системы-хранения-данных

Системы хранения данных

- По-английски они называются одним словом – storage. Но
- На русский это слово переводится как «хранилище».
- Часто на слэнге «ИТ-шников» используют слово «сторадж» в русской транскрипции, или слово «хранилка»
- Поэтому будем использовать термин «системы хранения данных», сокращенно СХД, или просто «системы хранения».

- Любые устройства для записи данных: т.н. «флешки», компакт-диски (CD, DVD, ZIP), ленточные накопители (Tape), жёсткие диски (Hard disk) и пр.
- Жёсткие диски используются не только внутри компьютеров, но и как внешние USB-устройства записи информации
- Одна из первых моделей iPod’а – это небольшой жёсткий диск диаметром 1,8 дюйма, с выходом на наушники и встроенным экраном.

SSD – Solid State Drive/Disk

- В последнее время все большую популярность набирают т.н. **«твердотельные» системы хранения SSD** (Solid State Disk, или Solid State Drive), которые по принципу действия схожи с «флешкой» для фотоаппарата или смартфона, только имеют контроллер и больший объем хранимых данных.
- В отличие от HDD, SSD-диск не имеет механически движущихся частей.

HDD

SSD

Классификация систем хранения

По частоте использования хранимых данных, СХД можно подразделить на

- системы краткосрочного хранения (online storage),
- хранения средней продолжительности (near-line storage) и
- системы долговременного хранения (offline storage).

- К первым можно отнести жесткий диск (или SSD) любого персонального компьютера.
- Ко вторым и третьим – внешние системы хранения **DAS (Direct Attached Storage),** которые могут представлять собой массив внешних дисков (Disk Array).
- Их также можно подразделить на
  - JBOD (Just a Bunch Of Disks) «просто массив дисков»
  - iDAS (intelligent disk array storage), массив с управляющим контроллером

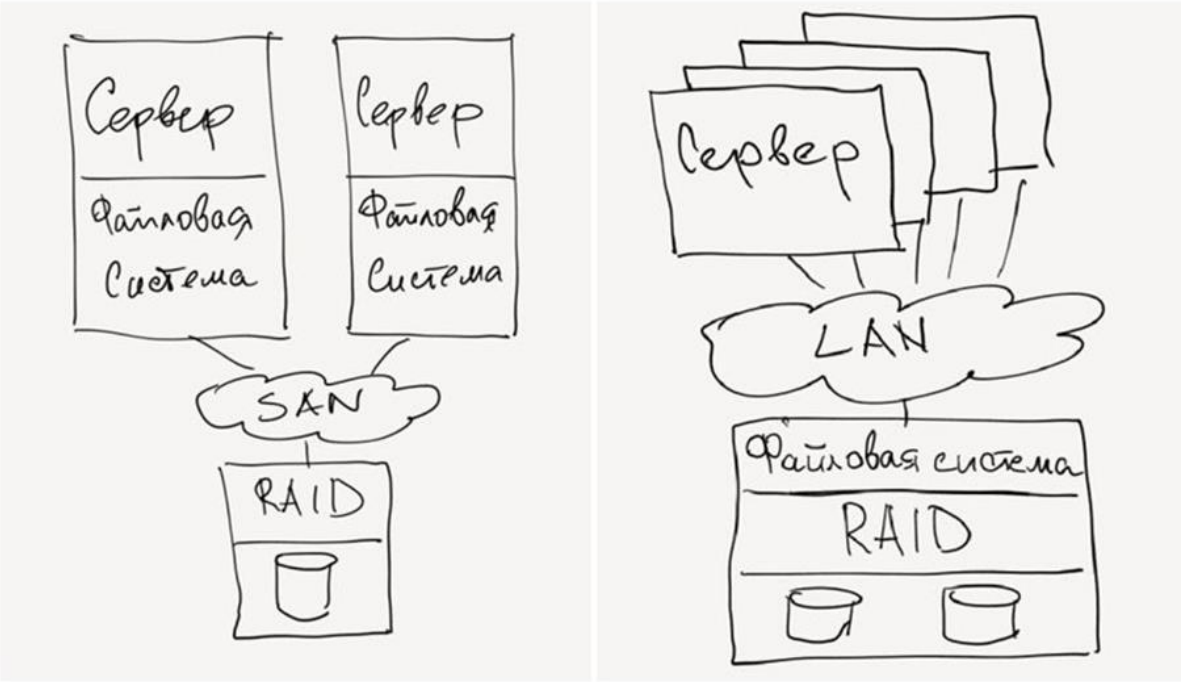
Внешние системы хранения DAS (Direct Attached Storage)

Внешние системы хранения бывают трех типов

- DAS (Direct Attached Storage),
- SAN (Storage Area Network) и
- NAS (Network attached Storage).

Даже опытные ИТ-шники часто не могут объяснить разницу между SAN и NAS

- Разница между SAN и NAS есть, и существенная (см. рис. справа).



SAN: с системой хранения связаны сами серверы через сеть области хранения данных SAN.
NAS: серверы связаны через локальную сеть LAN с общей файловой системой.

Основные протоколы СХД (1)

- **Протокол SCSI (Small Computer System Interface),** произносится как «ска́зи».
- Протокол, разработанный в середине 80-х годов для подключения внешних устройств к мини-компьютерам.
  - Преимущества: независимость от используемого сервера, возможность параллельной работы нескольких устройств, высокая скорость передачи данных.
  - Недостатки: ограниченность числа подключённых устройств, дальность соединения сильно ограничена.

FC (Fiber Channel): внутренний протокол между сервером и совместно используемой СХД, контроллером, дисками. Это широко используемый протокол последовательной связи, работающий на скоростях 4 или 8 Гигабит в секунду (Gbps).

- FC работает через оптоволокно (fiber), но и по меди тоже может работать.
- Fiber Channel – основной протокол для систем хранения FC SAN.

Основные протоколы СХД (2)

iSCSI (Internet Small Computer System Interface) стандартный протокол для передачи блоков данных поверх протокола TCP/IP («SCSI over IP»).

- iSCSI может рассматриваться как высокоскоростное недорогое решение для систем хранения, подключаемых через Интернет.
- iSCSI инкапсулирует команды SCSI в пакеты TCP/IP для передачи их по IP-сети.

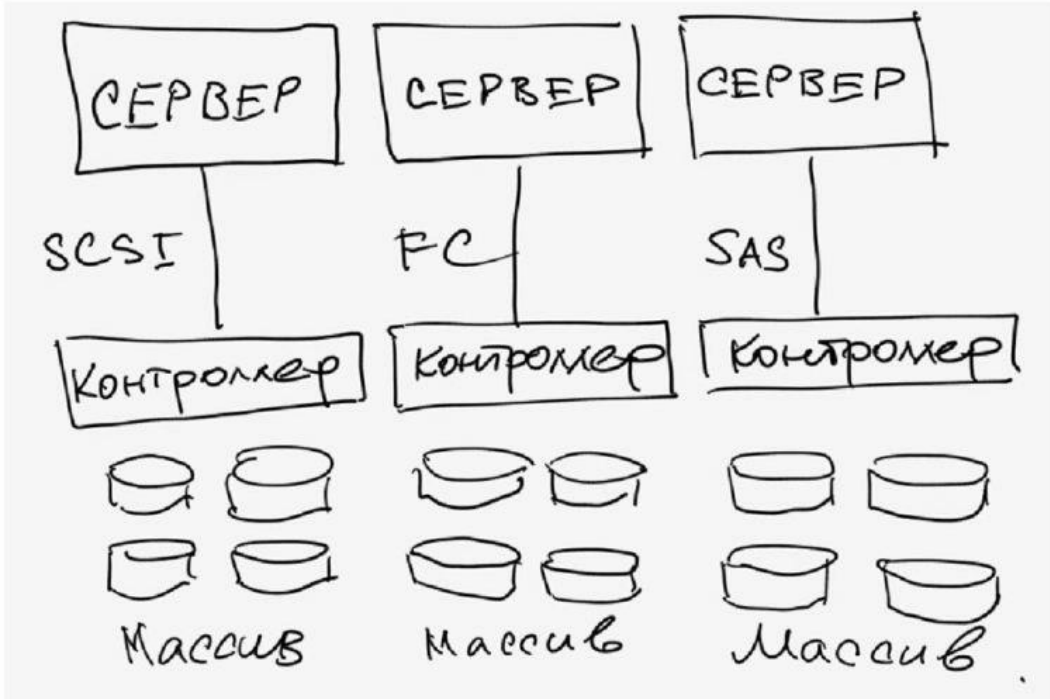
SAS (Serial Attached SCSI).

- SAS использует последовательную передачу данных и совместим с жёсткими дисками SATA.
- SAS может передавать данные со скоростью 3 Гбит/с или 6 Гбит/с, и поддерживает режим полного дуплекса, т.е. может передавать данные в обе стороны с одинаковой скоростью.

Типы систем хранения.

Три основных типа систем хранения:

- **DAS (Direct Attached Storage)**
- NAS (Network attached Storage)
- SAN (Storage Area Network)



СХД с непосредственном подключением дисков DAS были разработаны в конце 70-х годов, вследствие взрывного увеличения пользовательских данных, которые уже просто физически не помещались во внутренней долговременной памяти компьютеров (здесь речь идёт не о персоналках, их тогда ещё не было, а больших компьютерах, т.н. мейнфреймах).

Скорость передачи данных в DAS была невысокой, от 20 до 80 Мбит/с, но для тогдашних нужд ее вполне хватало.

Типы систем хранения.

Три основных типа систем хранения:

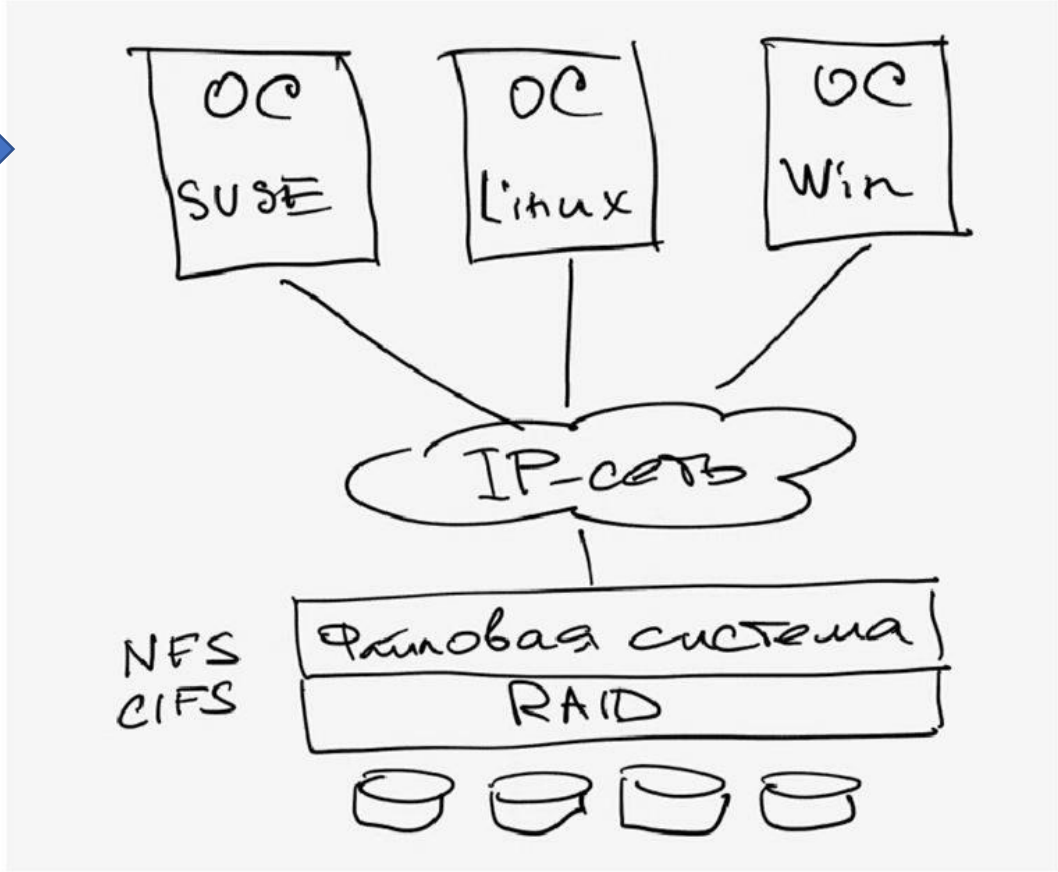
- DAS (Direct Attached Storage)
- **NAS (Network attached Storage)**
- SAN (Storage Area Network)

СХД с сетевым подключением NAS появились в начале 90-х годов.

Причиной стало быстрое развитие сетей и критические требования к совместному использованию больших массивов данных в пределах предприятия или сети оператора.

В NAS использовалась специальная сетевая файловая система CIFS (Windows) или NFS (Linux), поэтому разные серверы разных пользователей могли считывать один и тот же файл из NAS одновременно.

Скорость передачи: 1 - 10 Гбит/с.



Типы систем хранения.

Три основных типа систем хранения:

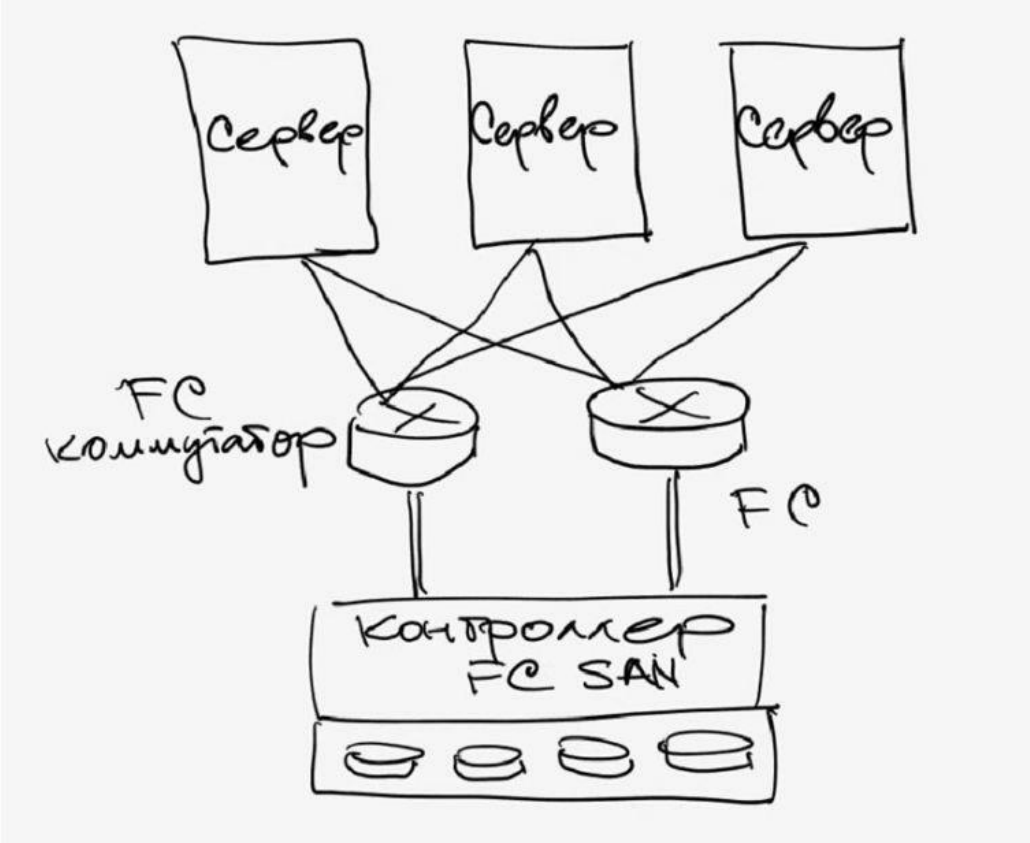
- DAS (Direct Attached Storage)
- NAS (Network attached Storage)
- **FC-SAN (FC Storage Area Network)**

В середине 90-х появились FC SAN. Их разработка была вызвана необходимостью организации разбросанных по сети данных.

Одно устройство хранения в SAN может быть разбито на несколько небольших узлов, называемых LUN (Logical Unit Number), каждый из которых принадлежит одному серверу.

Скорость передачи данных 2-8 Гбит/с.

FC SAN могли обеспечивать технологии защиты данных от потерь (snapshot, backup).



Типы систем хранения.

Три основных типа систем хранения:

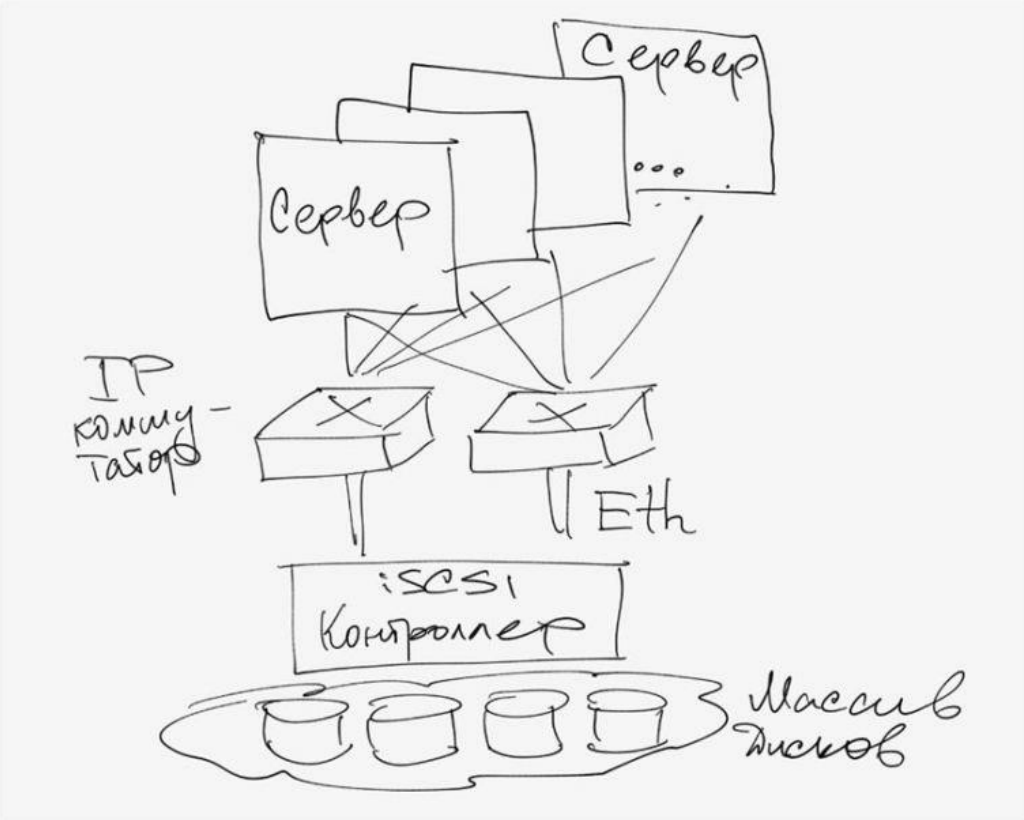
- DAS (Direct Attached Storage)
- NAS (Network attached Storage)
- **IP SAN (IP Storage Area Network)**

Разработаны в начале 2000-х годов.

FC SAN были дороги, сложны в управлении, а сети протокола IP находились на пике развития, поэтому и появился этот стандарт.

IP SAN подключались к серверам при помощи iSCSI-контроллера через IP-коммутаторы

Скорость передачи данных 1 - 10 Гбит/с.



Одно устройство хранения в SAN может быть разбито на несколько небольших узлов, называемых **LUN (Logical Unit Number),** каждый из которых принадлежит одному серверу.

Типы систем хранения - резюме

- SAN предназначены для передачи **блоков** данных в СХД
- NAS обеспечивают доступ к данным на уровне **файлов**
- Комбинацией SAN + NAS можно получить высокую степень интеграции данных, высокопроизводительный доступ и совместный доступ к файлам.
- Такие системы получили название unified storage – «унифицированные системы хранения».

- Унифицированные системы хранения: архитектура сетевых СХД, которая поддерживает как файлово-ориентированную систему NAS, так и блоко-ориентированную систему SAN.
- Эта СХД поддерживает практически все протоколы: FC, iSCSI, FCoE, NFS, CIFS.

Жёсткие диски (HDD и SSD)

Жёсткие диски можно подразделить на два основных типа:

- HDD (Hard Disk Drive), что и переводится как «жесткий диск»
- SSD (Solid State Drive, - т.н. «твердотельный диск»).

- В прошлом были и «**мягкие диски**», назывались «флоппи-диски» (из-за характерного "хлопающего" звука в дисководе при работе).
- Приводы для них ещё можно увидеть в системных блоках старых компьютеров, которые сохранились в некоторых госучреждениях.
- Однако, такие магнитные диски их вряд ли можно отнести к СИСТЕМАМ хранения.

«мягкие» диски (дискеты)

Характеристики жёстких дисков

Ёмкость

- В современных жестких дисках емкость измеряется в гигабайтах или терабайтах. Для HDD эта величина кратна ёмкости одного магнитного диска внутри коробки, умноженной на число магнитных, которых обычно бывает несколько.

Скорость вращения (только для HDD)

- Скорость вращения дисков внутри привода, измеряется в оборотах в минуту RPM (Rotation Per Minute), обычно составляет 5400 RPM или 7200 RPM.
- HDD с интерфейсами SCSI/SAS имеют скорость вращения 10000—15000 RPM.

Среднее время доступа

- Среднее время доступа = Среднее время поиска (Mean seek time) + Среднее время ожидания (Mean wait time), т.е. время извлечения информации с диска.

Скорость передачи данных

- Это скорости считывания и записи данных на жестком диске, измеряемая в мегабайтах в секунду (MB/S).

IOPS (Input/Output Per Second)

- Число операций ввода-вывода (или чтения-записи) в секунду (Input/Output Operations Per Second).
- IOPS – важный показатель, именно от него зависит быстродействие бизнес-приложений.

RAID

RAID (Redundant Array of Independent Disks) – массив независимых дисков с избыточностью хранения данных.

- Избыточность означает то, что все байты данных при записи на один диск дублируются на другом диске, и могут быть использованы в том случае, если первый диск откажет. Кроме того, эта технология помогает увеличить IOPS.

Основные понятия RAID –

- stripping (т.н. «располосование» или разделение)
- mirroring (т.н. «зеркалирование», или дублирование) данных.
- Их сочетания определяют различные виды RAID-массивов жёстких дисков.

Уровни RAID-массивов

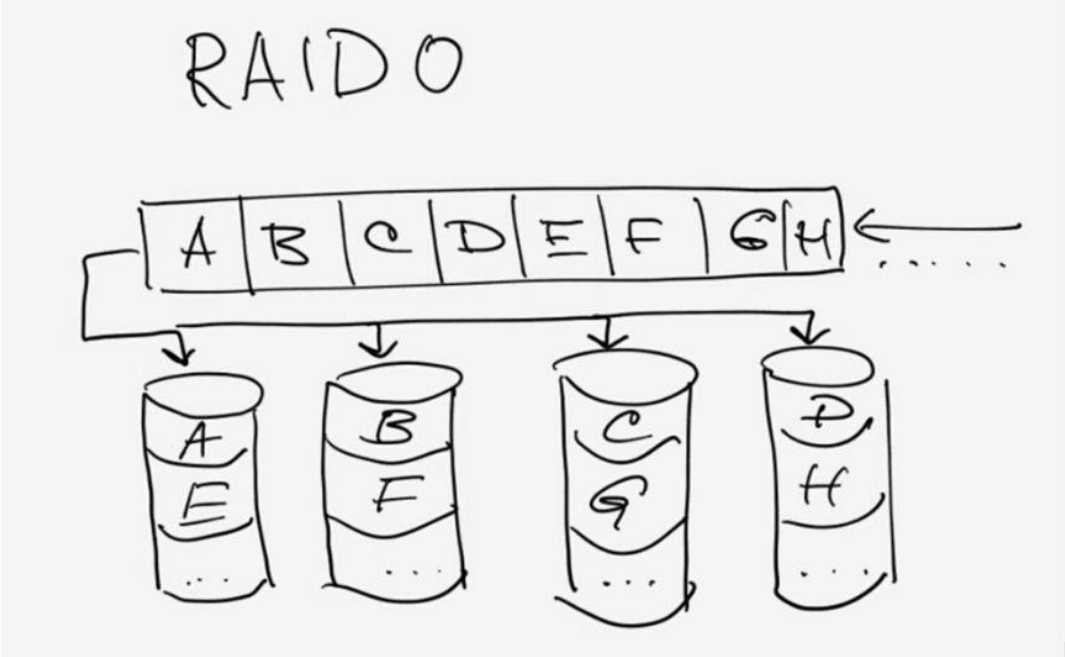
RAID 0	**Разделение данных без проверки на паритет.**
RAID 1	Дублирование данных без проверки на паритет.
RAID 3	Разделение данных с проверкой на паритет, с хранением данных паритета в выделенной области одного диска массива.
RAID 5	Разделение данных с проверкой на паритет, с хранением этих данных в разных областях всех дисков массива.
RAID 6	Разделение данных с проверкой на паритет, с хранением этих данных в разных областях всех дисков массива, а также с сохранением на отдельном диске для дополнительного резервирования.

Комбинации этих видов порождают ещё несколько видов RAID:

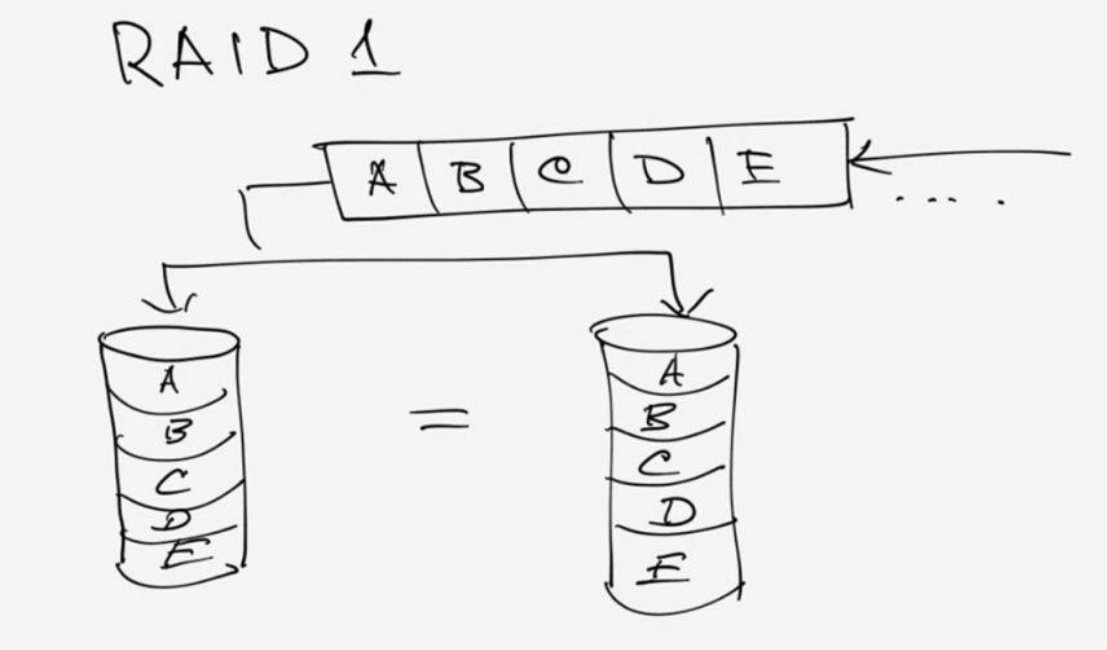
RAID 0+1	**Выполнение операций RAID 0 и RAID 1, т.е. сначала разделение, затем дупликация данных.**
RAID 10	Аналогично RAID 0+1 в обратном порядке: сначала выполняется RAID 1 (дублирование) затем RAID 0 (разделение).
RAID 50	Выполнение операций RAID 5, затем RAID 0, с целью увеличения производительности RAID 5

RAID

- RAID 0 (разделение):



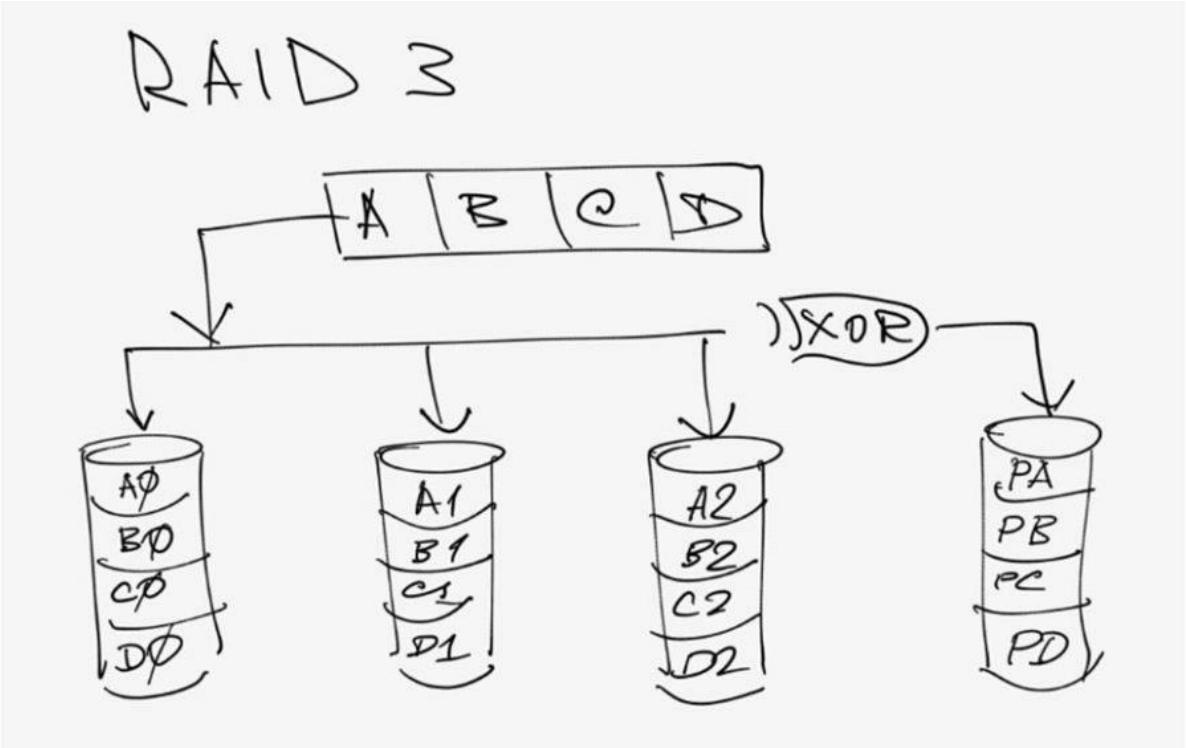
- RAID 1 (дублирование)



RAID

RAID 3.

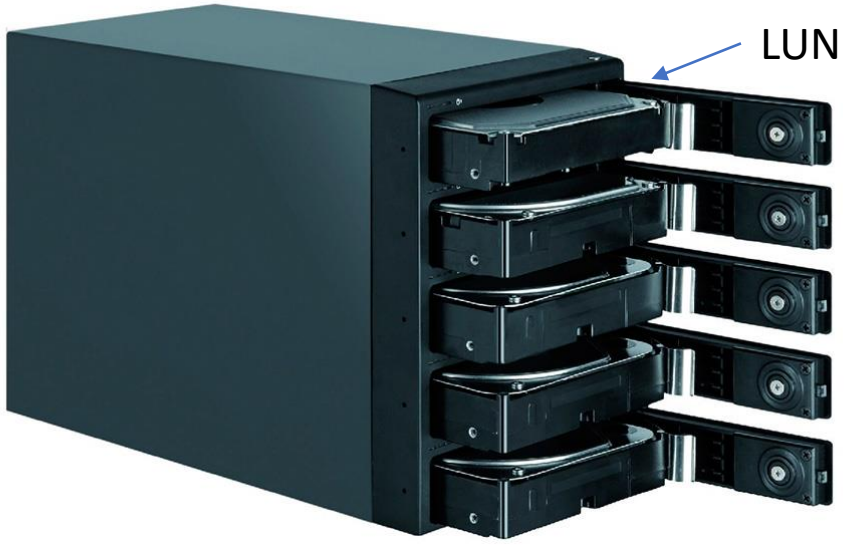
- XOR – логическая функция «исключающее ИЛИ» (eXclusive OR). При помощи неё вычисляется значение паритета для блоков данных A, B, C, D… , который записывается на отдельный диск.



Основные характеристики видов RAID

Уровень RAID	RAID0	RAID1	RAID5	RAID6	RAID10
Устойчивость к ошибкам	Низкая	Высокая	Средняя	Наивысшая	Высокая
Избыточность	Никакая	Дубли-рование	Паритет	Паритет	Дубли-рование
Доступная емкость	100%	50%	(N-1)/N	(N-2)/N	50%
Производительность	Наивысшая	Низкая	Средняя	Средняя	Высокая

N – Количество физических HDD или LUN



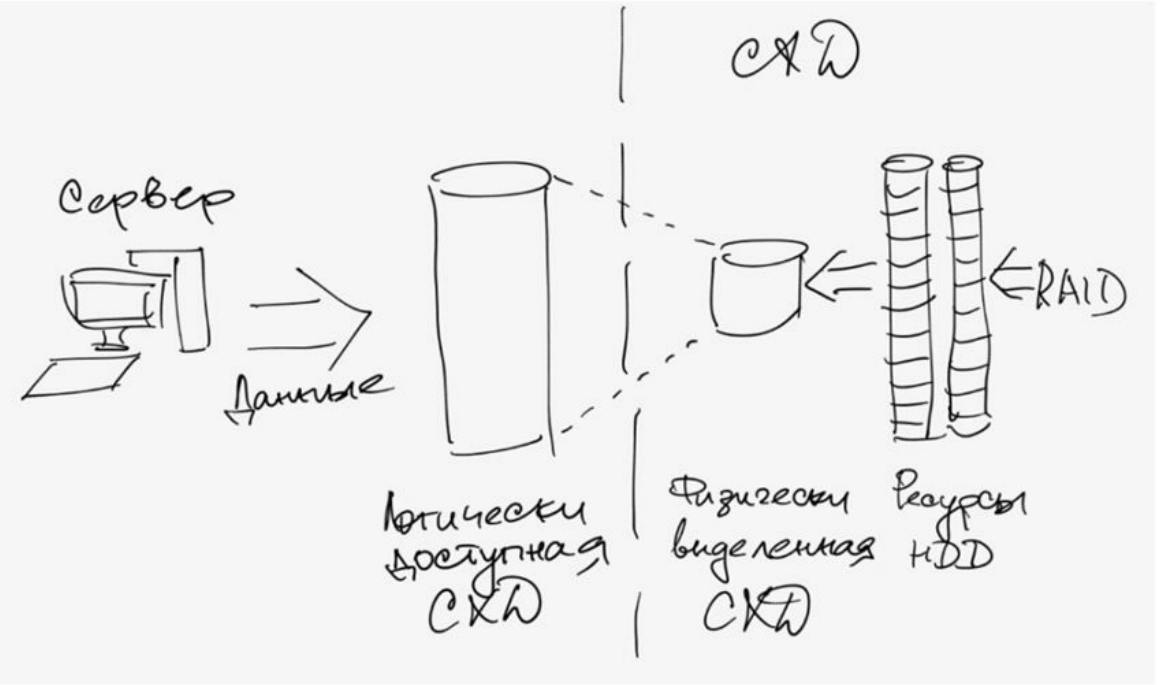
Программное обеспечение систем хранения

- **Управление и администрирование (Management):** управление и задание параметров инфраструктуры: вентиляции, охлаждения, режимы работы дисков и пр., управление по времени суток и пр.
- **Защита данных:** Snapshot («моментальный снимок» состояния диска), копирование содержимого LUN, множественное дублирование (split mirror), удалённое дублирование данных (Remote Replication), непрерывная защита данных CDP (Continuous Data Protection) и др.
- **Повышение надёжности**: различное ПО для множественного копирования и резервирования маршрутов передачи данных внутри ЦОД и между ними.
- **Повышение эффективности**:
  - Технология тонкого резервирования (Thin Provisioning),
  - автоматическое разделение системы хранения на уровни (tiered storage),
  - устранение повторений данных (deduplication),
  - управление качеством сервиса, предварительное извлечение из кэш-памяти (cache prefetch), разделение данных (partitioning),
  - автоматическая миграция данных, снижение скорости вращения диска (disk spin down)

Технология «тонкого резервирования» (Thin Provisioning)

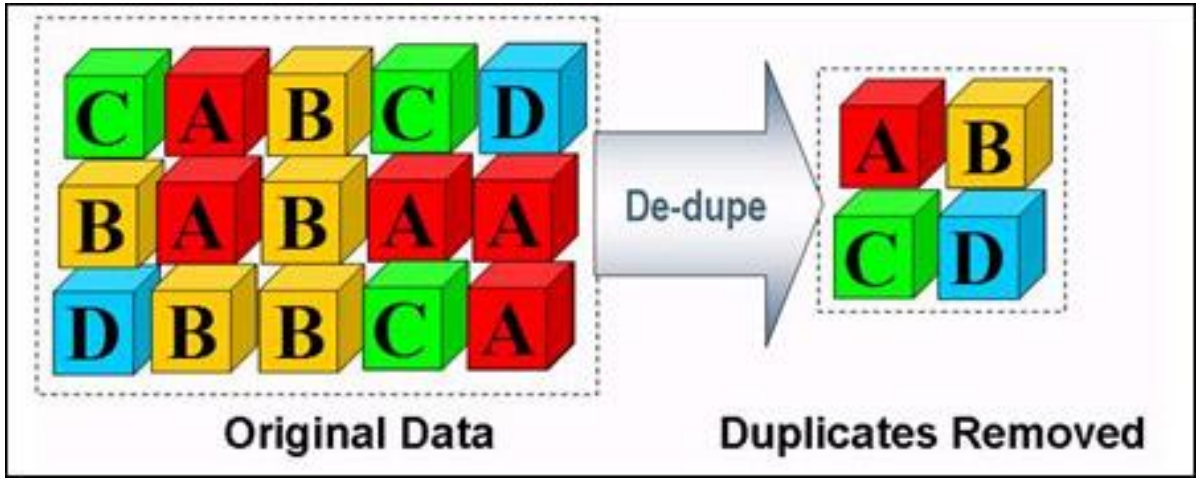
Для иллюстрации принципа «thin provisioning», можно привести банковский кредит.

- Когда банк выпускает десять тысяч кредитных карт с лимитом в 500 тысяч, ему не нужно иметь на счету 5 миллиардов, чтобы этот объём кредитов обслуживать.
- Пользователи кредитных карт обычно не тратят весь кредит сразу, и используют лишь его малую часть.
- Тем не менее, каждый пользователь в отдельности может воспользоваться всей или почти всей суммой кредита, если общий объем средств банка не исчерпан.



DEDUP

- Дедупликация (устранение повторений) данных (deduplication, DEDUP).
- Дедупликация устраняет повторы данных на пространстве диска, обычно используемого для резервирования данных.
- За счёт этого становится возможным значительно сократить требования к ёмкости системы хранения.



Disk spin-down

- Снижение скорости вращения диска (Disk spin-down) – «гибернация» (засыпание диска (hibernation)
- Если данные на каком-то диске не используются долгое время, то Disk spin-down переводит его в режим гибернации, чтобы снизить потребление энергии на бесполезное вращение диска на обычной скорости.
- При этом также повышается срок службы диска и увеличивается надежность системы в целом.
- При поступлении нового запроса к данным на этом диске, он «просыпается» и скорость его вращения увеличивается до обычной. Платой за экономию энергии и повышение надёжности является некоторая задержка при первом обращении к данным на диске

Snapshot и CDP (Continuous data protection)

- Snapshot – «Моментальный снимок» состояния диска, полностью пригодная к использованию копия набора данных на диске на момент съёма копии.
- Snapshot используется для частичного восстановления состояния системы на момент копирования. При этом непрерывность работы системы совершенно не затрагивается, и быстродействие не ухудшается.
- CDP (Continuous data protection) - Непрерывная защита данных, также известная как continuous backup или real-time backup, представляет собой создание резервной копии автоматически при каждом изменении данных.
- При этом становится возможным восстановление данных при любых авариях в любой момент времени, причем при этом доступны актуальная копия данных, а не тех, что были несколько минут или часов назад.

DR - Disaster Recovery

- Технология Disaster Recovery предполагает, что центр резервирования, используемый для восстановления данных при стихийных бедствиях, располагается на значительном удалении от места основного ЦОД, и взаимодействует с ним по сети передачи данных, наложенной на транспортную сеть, чаще всего оптическую.
- Использовать при таком расположении основного и резервного ЦОД, например, технологию CDP будет просто невозможно технически.

- **BW (Backup Window)** – «окно резевирования», время, необходимое для системы резевирования для того, чтобы скопировать принятый объем данных рабочей системы.
- **RPO (Recovery Point Objective)** – «Допустимая точка восстановления», максимальный период времени и соответствующий объем данных, который допустимо потерять для пользователя СХД.
- **RTO (Recovery Time Objective)** – «допустимое время недоступности», максимальное время, в течение которого СХД может быть недоступной, без критического воздействия на основной бизнес.

Три основополагающих понятия DR

Спасибо!

Алексей Шалагинов
ashalaginov@gmail.com
+7 925 0081486
Shalaginov.com