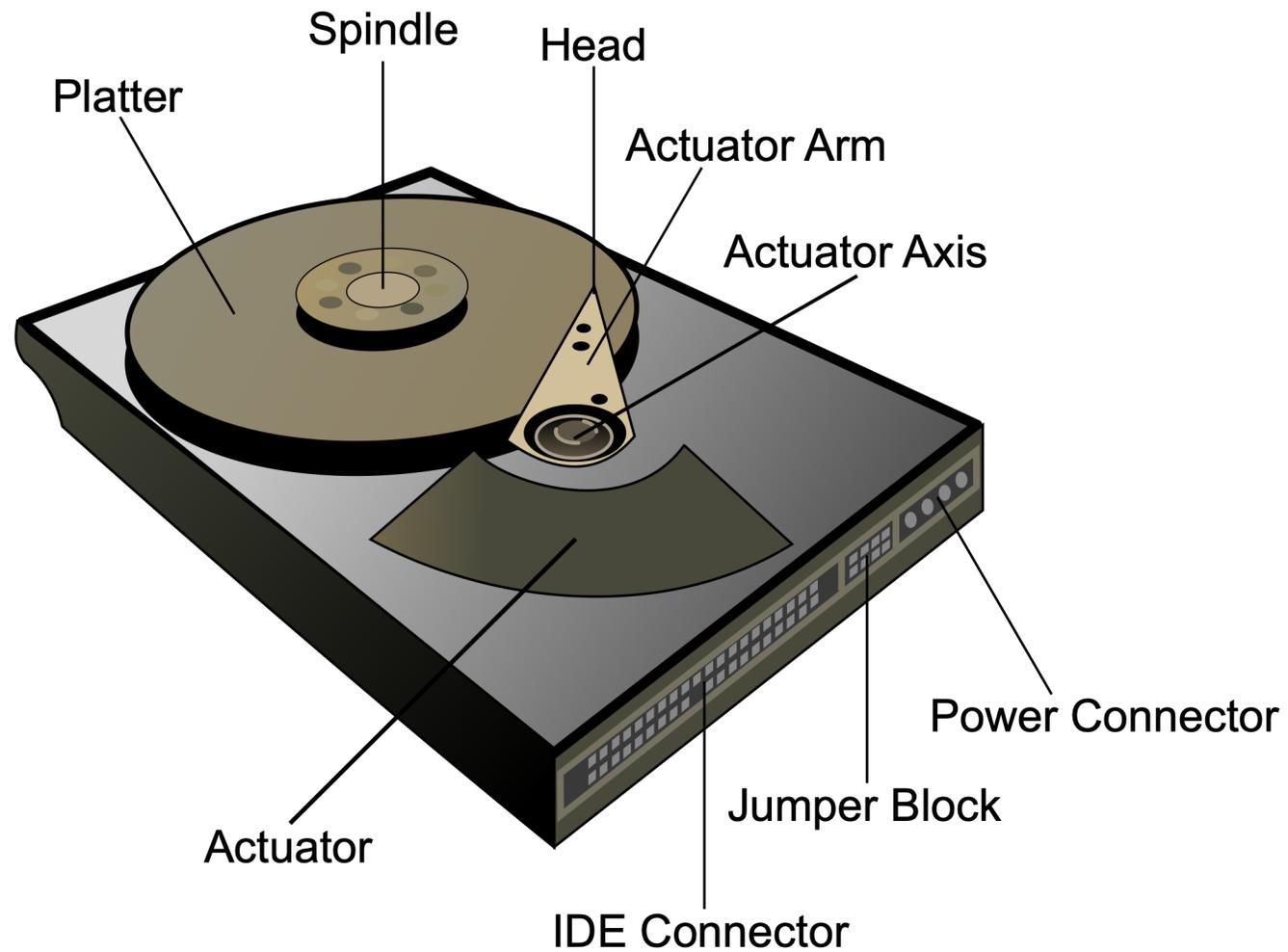


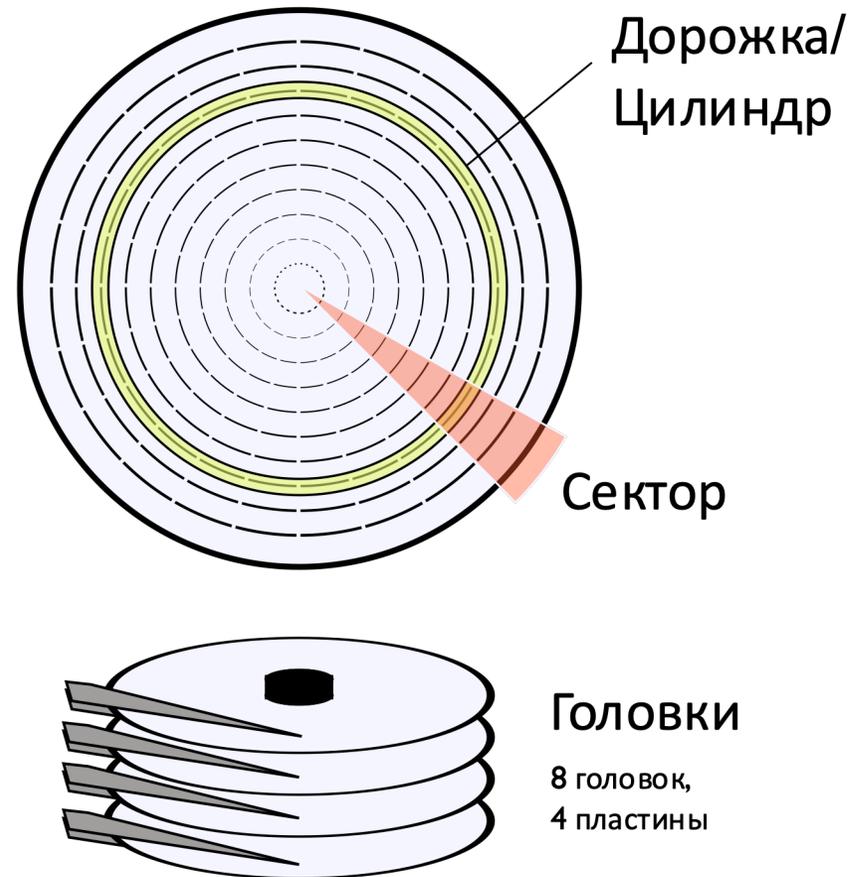
Жесткие диски и SSD

Устройство жесткого диска



Геометрия диска

- Цилиндр – все дорожки, соответствующие одному положению блока головок
- Сектор – минимальная единица записи (обычно 512 байт)



Секторы

- Преамбула обычно содержит
 - синхросигнал (маркер начала записи),
 - номер дорожки (для обнаружения ошибок позиционирования головки)
 - номер сектора



Производительность жесткого диска

- Скорость передачи данных (напр. если подключить диск SATA 3 к контроллеру SATA 2, скорость будет как у SATA 2)
- Скорость чтения сектора (определяется плотностью записи и скоростью вращения)
- Ротационная задержка (время подлета нужного сектора к головке)
- Задержка подачи головки
 - Современные диски сначала приблизительно кидают головку в нужную область, затем малыми движениями ищут нужную дорожку

Другие проблемы жестких дисков

- Чувствительность к ударам и вибрациям («посадка головок»)
- Шум
- Нагрев
- Механический износ
- Большое энергопотребление

Solid State Drive

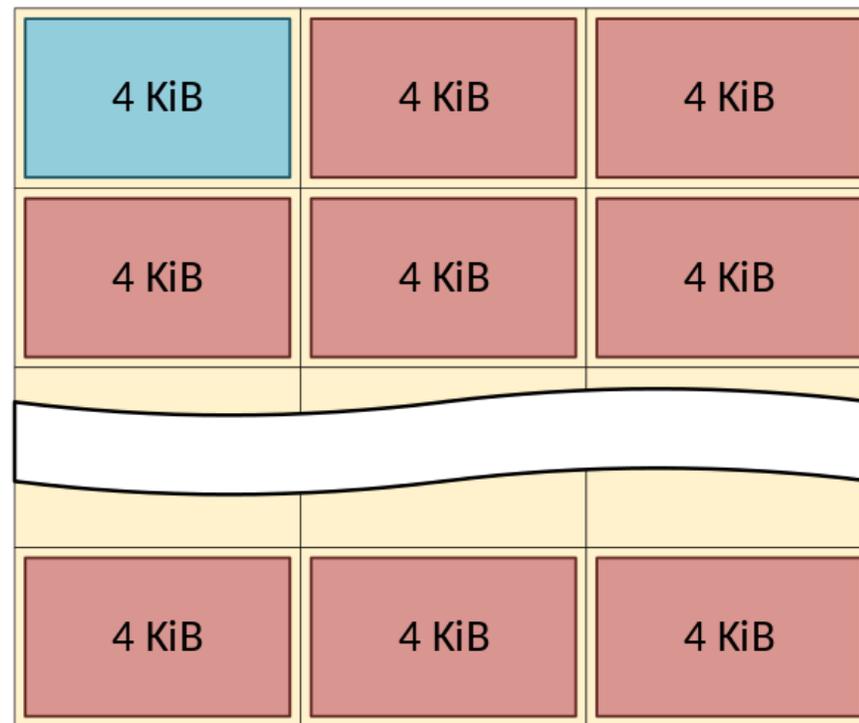
- Флэш-память – это электрически стираемое ПЗУ с блочным стиранием
- Бесшумные, быстрые, нет движущихся деталей
- Все виды ЭСПЗУ страдают ограничением на число циклов записи
- Решение: wear leveling

Wear leveling – пример структуры SSD

Data written
in 4KiB pages

Data erased
in 256KiB
blocks

64 writable pages
in 1 erasable block



Typical NAND flash pages and blocks

Wear leveling – сборка мусора

Block X	A	B	C
	D	free	free
	free	free	free
	free	free	free

Block Y	free	free	free
	free	free	free
	free	free	free
	free	free	free

1. Four pages (A-D) are written to a block (X). Individual pages can be written at any time if they are currently free (erased).

Block X	A	B	C
	D	E	F
	G	H	A'
	B'	C'	D'

Block Y	free	free	free
	free	free	free
	free	free	free
	free	free	free

2. Four new pages (E-H) and four replacement pages (A'-D') are written to the block (X). The original A-D pages are now invalid (stale) data, but cannot be overwritten until the whole block is erased.

Block X	free	free	free
	free	free	free
	free	free	free
	free	free	free

Block Y	free	free	free
	free	E	F
	G	H	A'
	B'	C'	D'

3. In order to write to the pages with stale data (A-D) all good pages (E-H & A'-D') are read and written to a new block (Y) then the old block (X) is erased. This last step is *garbage collection*.

Дисковые массивы

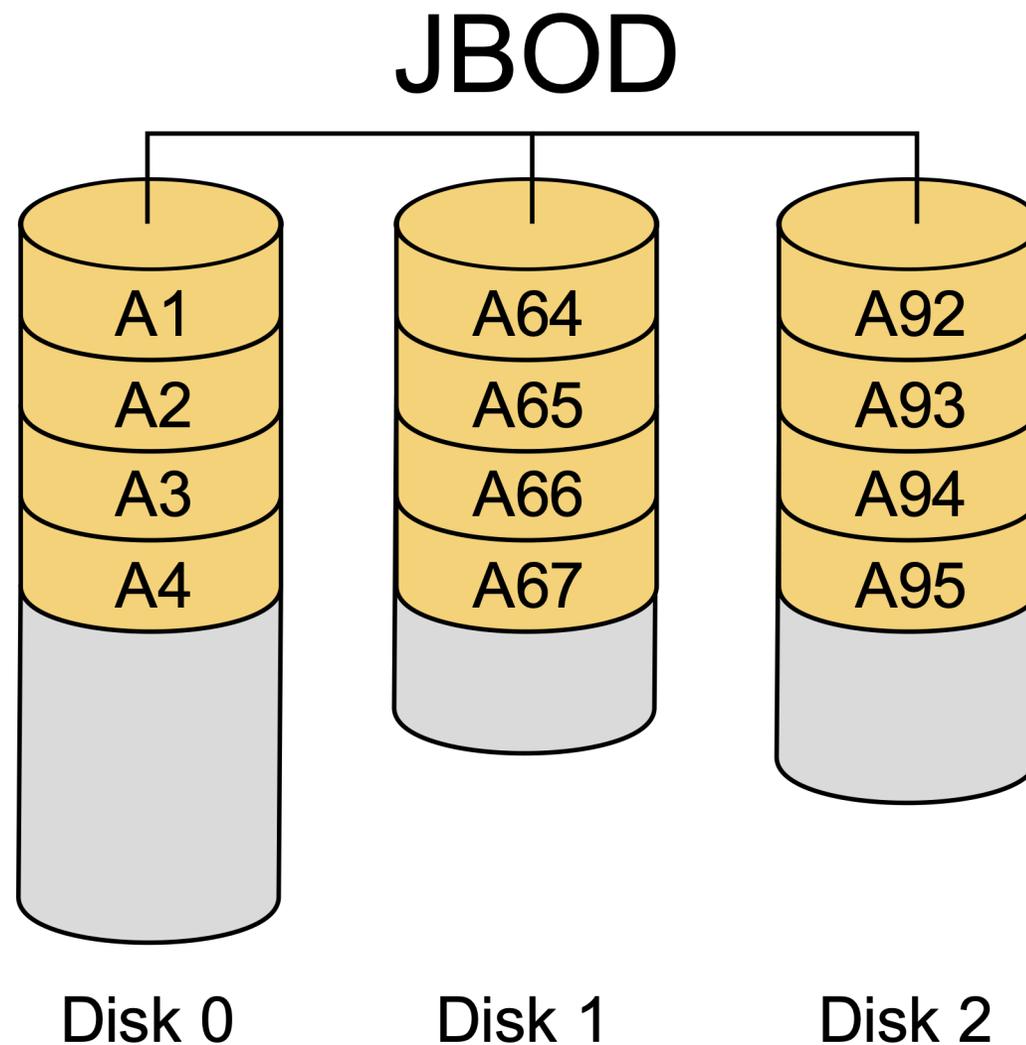
- RAID – Redundant Array of Inexpensive Disks
- Может быть реализован
 - Программно (на уровне драйверов диска)
 - Аппаратно (плата контроллера)
 - В отдельном корпусе (обычно используется для доступа iSCSI/NFS)

RAID 0

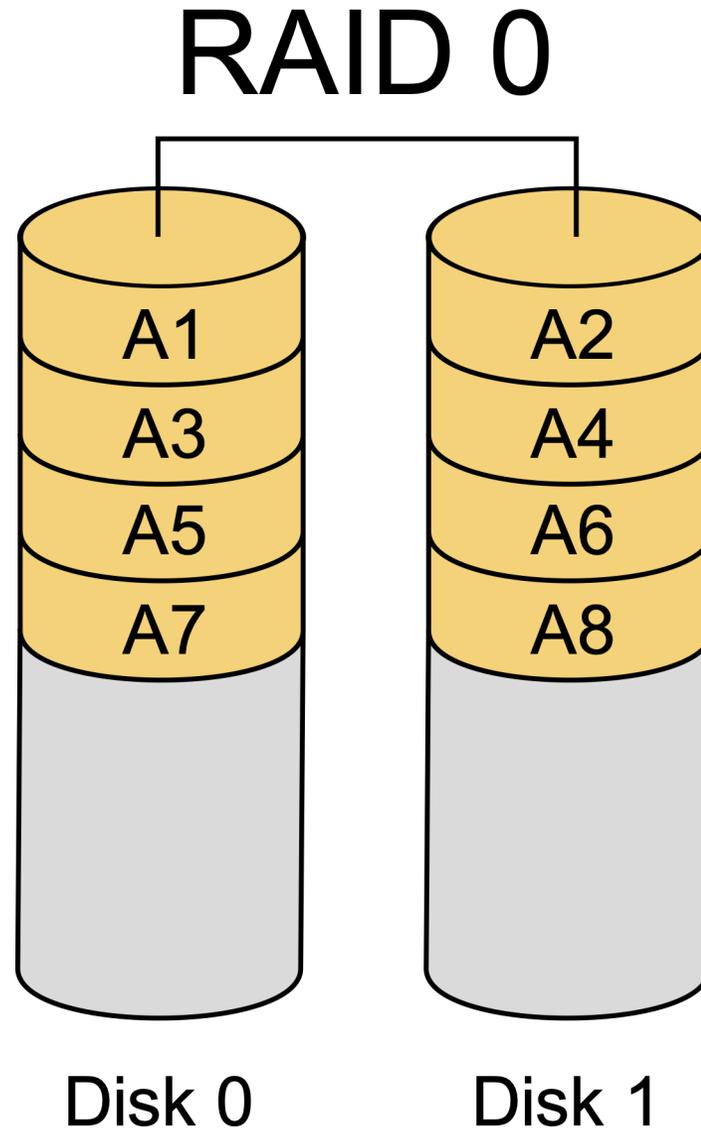
- логическое объединение двух или более дисков
 - Строго говоря, не RAID, так как не избыточно
 - Повышает производительность
 - Увеличивает единичный объем
 - Вероятность отказа всего массива $\approx N * P$,
где N – количество дисков,
 P – вероятность отказа одного

JBOD

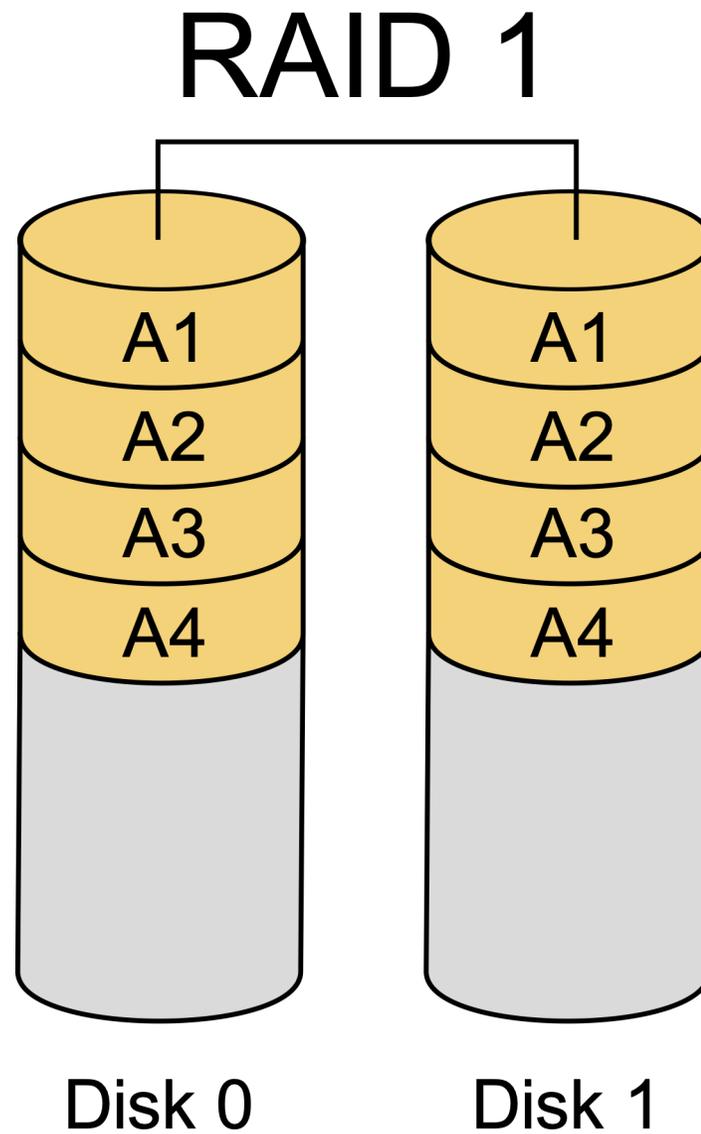
Just Bunch Of Disks,
также известно как
конкатенация



RAID 0
(striping)

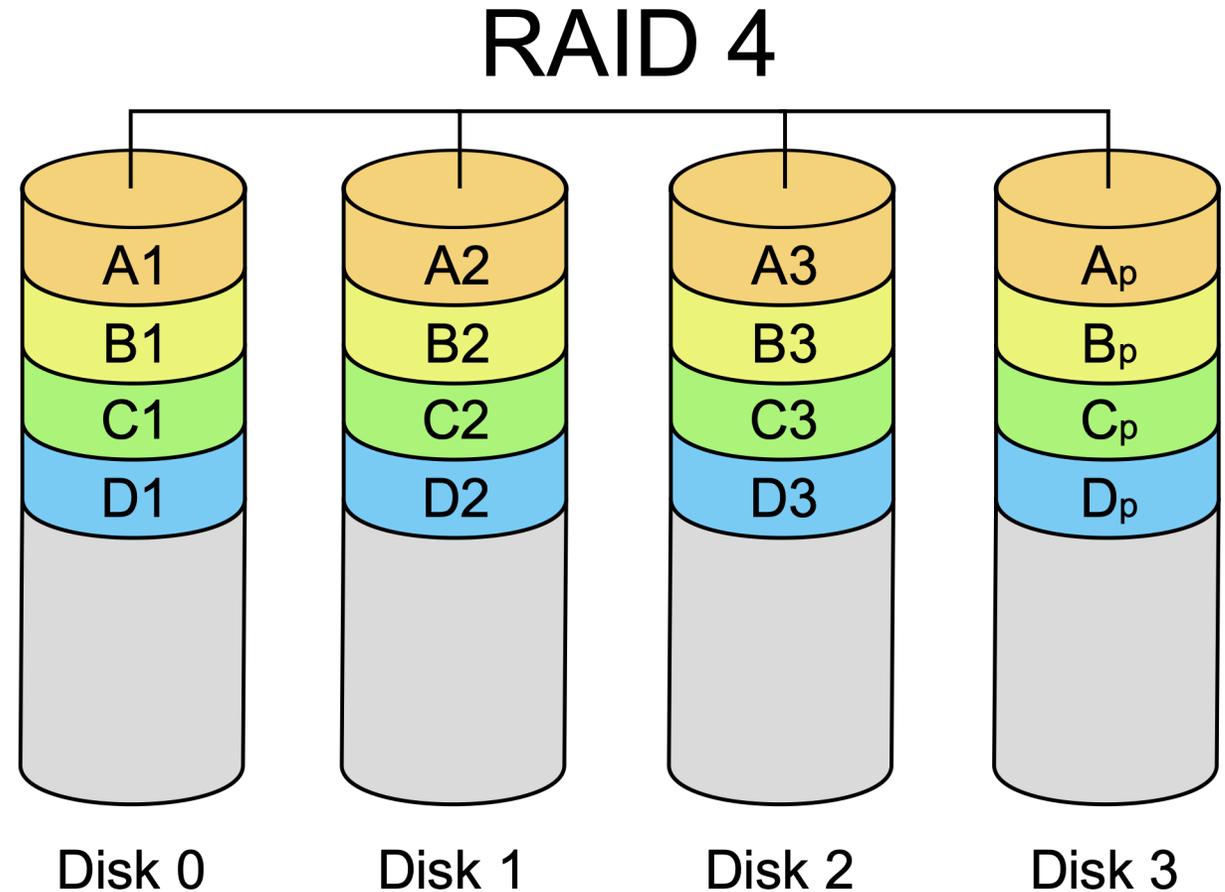


RAID 1 (зеркало)



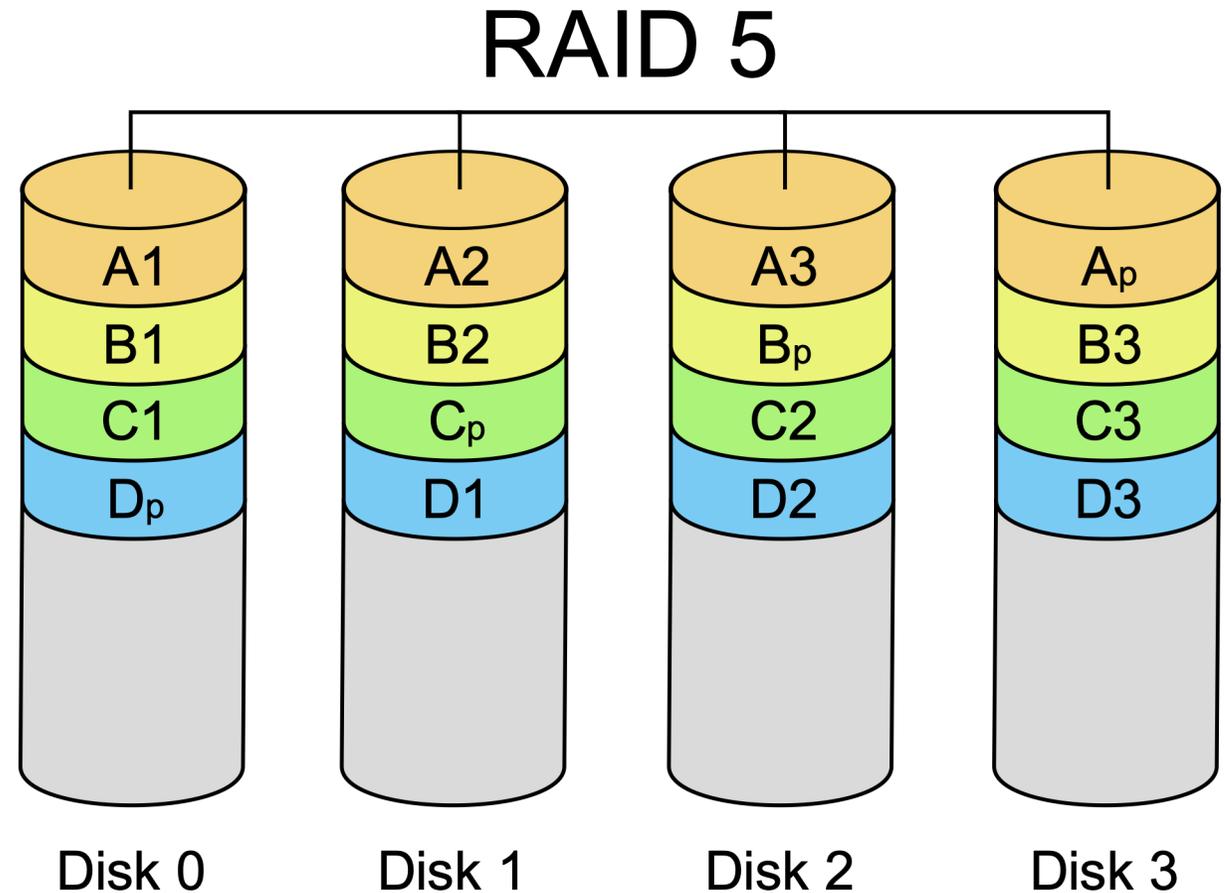
RAID 4

- Минимум 3 диска
- На одном диске размещена четность XOR A1..A3
- Запись в любой из секторов требует записи на диск с четностью



RAID 5

- Четность распределена по всем дискам



Многоуровневые RAID

- RAID 50 – JBOD или stripe set из двух массивов RAID5
- RAID 10 – зеркало из двух stripe set