Disková pole (RAID)

Architektury RAID

- Základní myšlenka: snaha o zpracování dat paralelně.
- Pozice diskové paměti v klasickém personálním počítači – vyhovuje pro aplikace s jedním uživatelem.
- Řešení: data jsou distribuována na více disků, datová operace je realizována paralelně.
- Co to nabízí: kromě distribuování dat na více disků možnost zvýšení spolehlivosti – využití redundance (zdvojování disků nebo vygenerování a záznam informace, která umožní opravu).
- Paralelní přístup a detekci poruchy diskové paměti a příp. opravu.

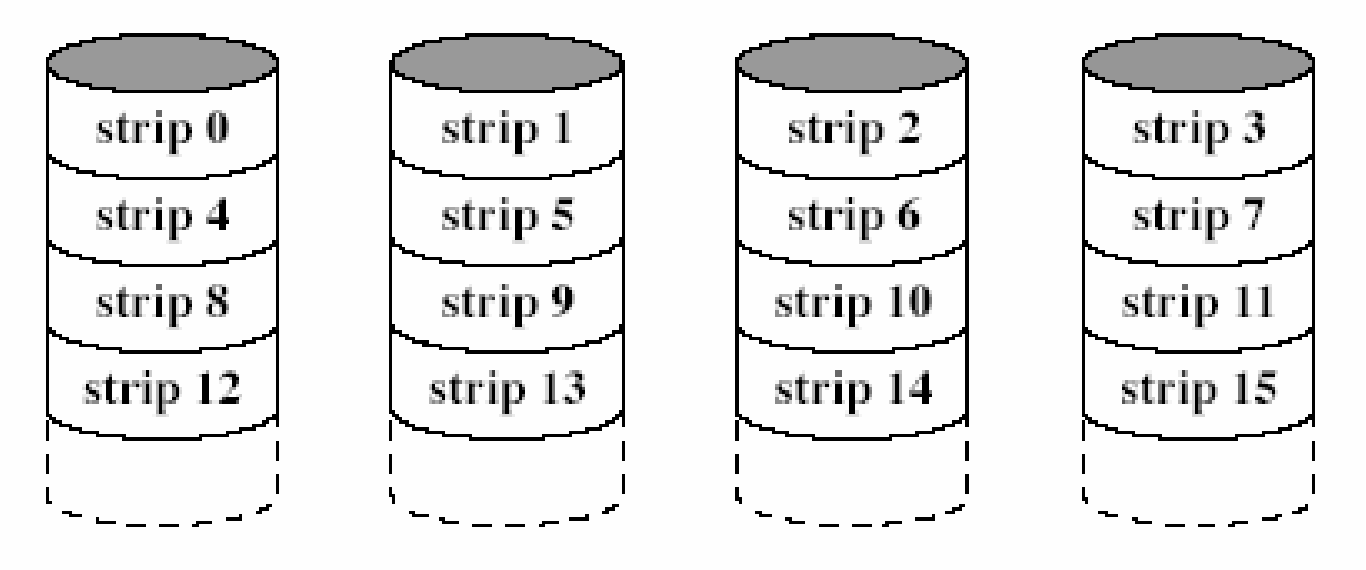
Architektury RAID

- Význam zkratky: **R**edundant **A**rray of **I**ndependent **D**isks.
- Jiné vysvětlení: **R**edundant **A**rray of **I**nexpensive **D**isks.
- Důvod zavedení RAID: reakce na zvyšující se rychlost procesoru.
- 7 úrovní.
- Neexistují hierarchické úrovně.
- Sada fyzických disků, operační systém je vidí jako jeden disk.
- Data jsou distribuována do všech fyzických disků.
- Možnost využití dodatečné kapacity pro uložení informace o paritě.

RAID 0

- Žádná redundance (všechny disku jsou využity pro uložení dat).
- Data rozdělena na všechny disky.
- Zvýšení rychlosti z těchto důvodů:
  - Požadovaná data jsou rozdělena na více disků.
  - Operace „vystavení“ (seek) je prováděna paralelně (současně na všech discích).
  - Všechny operace jsou prováděny paralelně.

RAID 0



Důležité:

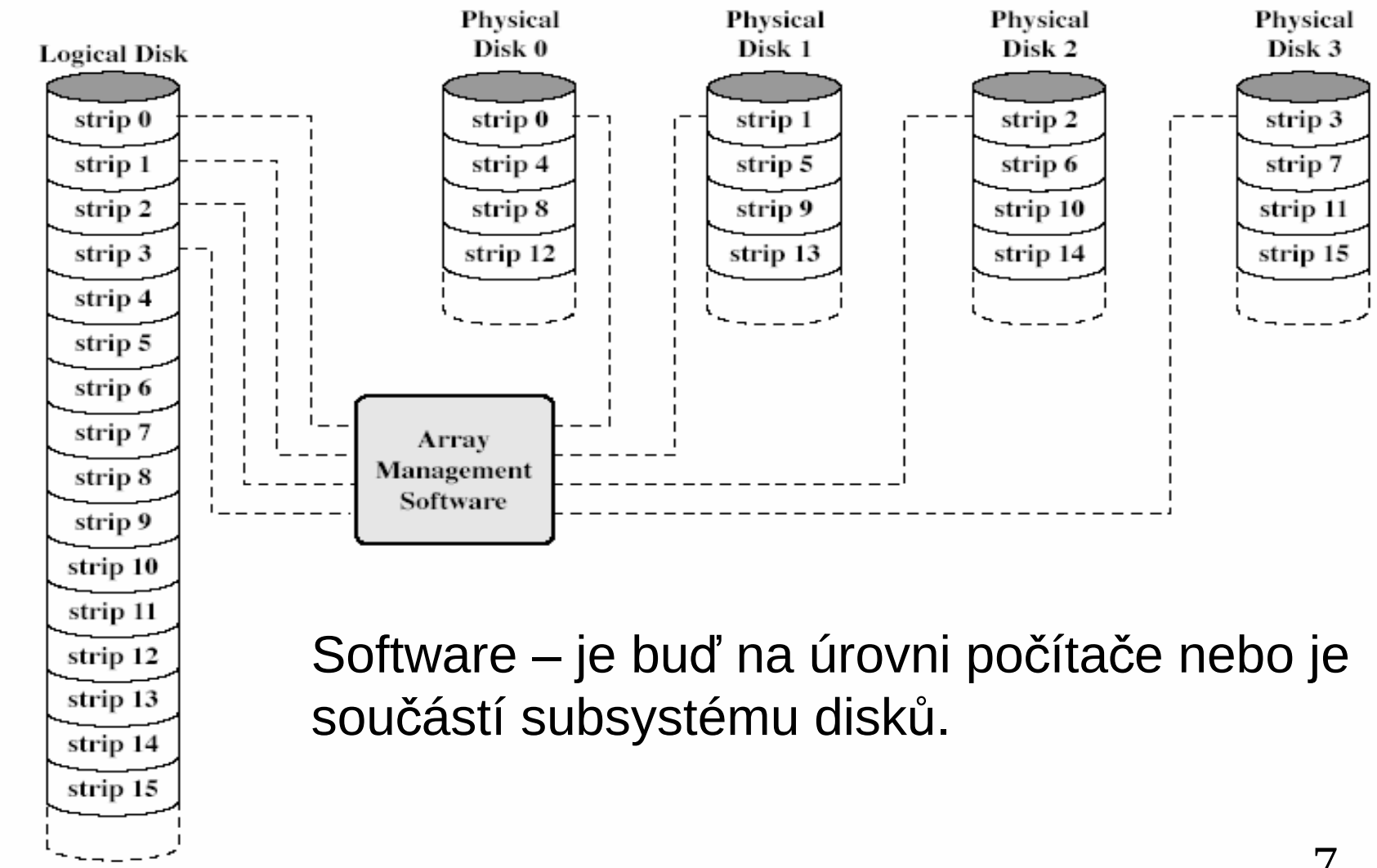
Dva různé V/V požadavky – data jsou na různých discích.

RAID 0

- Data jsou rozdělena na více disků (striped).
- Všechna uživatelská i systémová data jsou z hlediska OS uložena na jednom systémovém disku.
- Každý disk je rozdělen na **strips.**
- Příklad: pole n disků

  první strip na všech discích tvoří první **stripe**.
- Výhoda: současně je možné zpracovávat n prvků typu strip.

RAID 0



Software – je buď na úrovni počítače nebo je součástí subsystému disků.

RAID 0 – podpora vysoké rychlosti přenosu

- Paměť musí být dostatečně rychlá, tzn. na vysoké technické úrovni.
- Kvalitní spoj z diskové paměti (tvoří jeden celek s řadičem) do počítače.
- Rychlý řadič diskové paměti – každý disk v diskovém poli má svůj řadič, který autonomně řídí periferní operace.
- Rychlá systémová sběrnice.
- Rychlý procesor.
- **Toto vše platí obecně pro architektury RAID.**

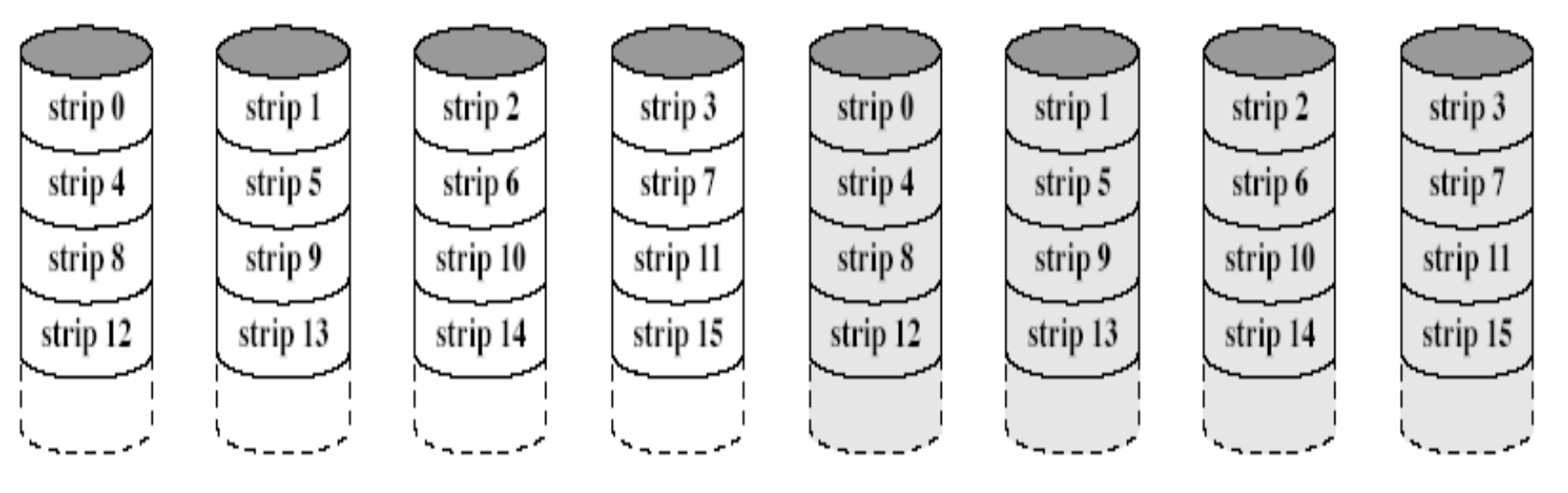
RAID 0 – charakter V/V požadavků

- V/V požadavky – požadavky na data.
- Ideální organizace dat – data, která spolu souvisejí (např. soubor) by měla být uložena v sousedních strip (v jednom stripe) – pak přenos paralelně.
- Výsledek – výrazně efektivnější operace související s diskem – hlavně rychlost přenosu (možnost přenosu jednoho souboru paralelně – jeho jednotlivé části).

RAID 0 – využití v systému zaměřeném na řešení transakcí

- Podpora vysokého objemu V/V požadavků.
- Řešení těchto požadavků – diskové pole umožňuje tyto požadavky vyváženě rozložit na více fyzických disků.
- Dvě situace – větší počet nezávislých transakcí nebo samostatné transakce, které je možné rozdělit na jistý počet asynchronních činností (termín „počet“ souvisí s počtem strip).
- Souvislost s velikostí strip – měl by být takový, aby řešení transakce nevyžadovalo více přístupů na disk.
- Představa: diskové pole v serverové stanici.

RAID 1



Architektura RAID 1

RAID 1

- Zrcadlené disky – snaha o zvýšení spolehlivosti.
- Každý strip je mapován na dva fyzické disky.
- Poznámka: v dalších typech řešeno způsobem, který nepředstavuje řešení typu „hrubá síla“, počítá se k datům parita (nevýhodné – při každé změně dat se musí znova počítat – režie).
- Výrazná redundance, nejsou však časové nároky na její realizaci – při změně se nepočítá informace, kterou jsou data zajištěna.
- Jednoduché zotavení při chybě.
- Nákladné.

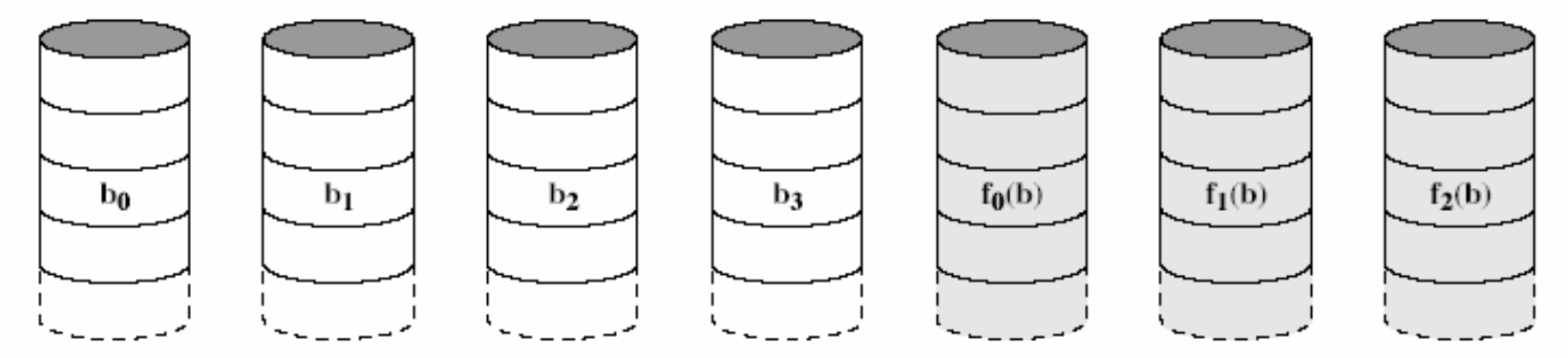
RAID 1

- **Čtení dat** – pouze z jednoho disku (z toho, kde to bude rychlejší – doba vystavení + rotační zpoždění).
- **Zápis dat** – provádí se paralelně na oba disky → výkon při zápise bude ovlivněn tím diskem, na němž to bude trvat déle (delší doba vystavení + rotační zpoždění).
- Zotavení po poruše – data se čtou z disku, který je funkční.
- Nevýhoda – vysoké náklady.

RAID 1 - využití

- Z hlediska výkonu tam, kde podstatnou část transakcí tvoří transakce „čtení“ (např. systémové disky – systémové programové vybavení a data - zálohování).
- Transakčně orientovaná aplikace – výhoda, pokud výrazným počtem transakcí jsou transakce typu čtení, horší stav v případě transakcí typu zápis.

RAID 2

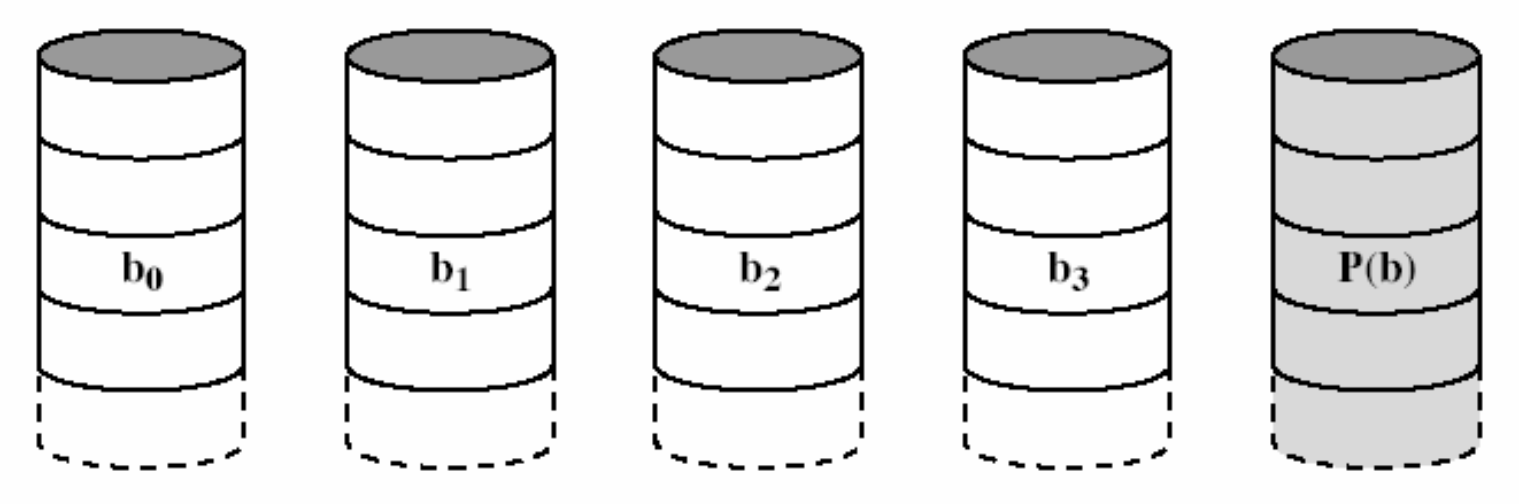


Čtyři disky – uložení dat, tři disky – informace pro opravu chyb

RAID 2

- Disky jsou synchronizovány, takže na všech discích jsou hlavy ve stejné pozici – z hlediska otáčení disku a vystavení.
- Paralelní přístup – na vyřízení každého V/V požadavku se podílejí všechny paralelně pracující disky.
- Velmi malé stripes - slabika/slovo.
- Informace potřebná pro opravu chyb se počítá z odpovídajících bitů na discích.
- Na paritní disky se ukládají bity vygenerované jako Hammingův kód z odpovídajících datových bitů.
- Velká redundance.
  - Nákladná technika.
  - Nepoužívá se (?).

RAID 3



Čtyři disky – uložení dat, jeden disk – informace pro opravu chyb.

RAID 3

- Organizováno podobně jako RAID 2.
- Pouze jeden redundantní disk bez ohledu na to, jak je rozsáhlé diskové pole.
- Jeden paritní bit pro každou sadu odpovídajících bitů.
- Data na discích, které mají poruchu, mohou být rekonstruována z existujících dat a parity.
- Vysoké rychlosti přenosu.

RAID 3 - využití redundance

- Pokud má disk poruchu, tak se přečte paritní bit a data se zrekonstruují ze zbývajících bitů, pro rekonstrukci se použije i bit paritní.
- Po výměně vadného disku je možné data ze zbývajících bitů zrekonstruovat.

- Rekonstrukce dat:
  Uvažujme diskové pole sestávající z pěti disků.
  X0 – X3 obsahují data
  X5 – paritní disk

- Schéma tvorby paritního bitu:
  X4(i) = X3(i) xor X2(i) xor X1(i) xor X0(i)

RAID 3 – rekonstrukce dat

- Předpokládejme, že disk X1 přestal fungovat, tzn. data z něj nejsou k dispozici.
- Výchozí vztah (předcházející stránka):

  X4(i) = X3(i) xor X2(i) xor X1(i) xor X0(i)
- K oběma stranám rovnice přičteme X4(i) xor X1(i)
- Pak dostaneme:

  X1(i) = X4(i) xor X3(i) xor X2(i) xor X0(i)

  Hodnota bitu z disku, který má poruchu, se vypočte ze zbývajících bitů.

- Závěr: bit z vadného disku se počítá ze zbývajících bitů, tento princip se používá pro RAID3 až RAID6 (tzv. redukovaný režim).

RAID 3

- Zápis v situaci, kdy je disk vadný:

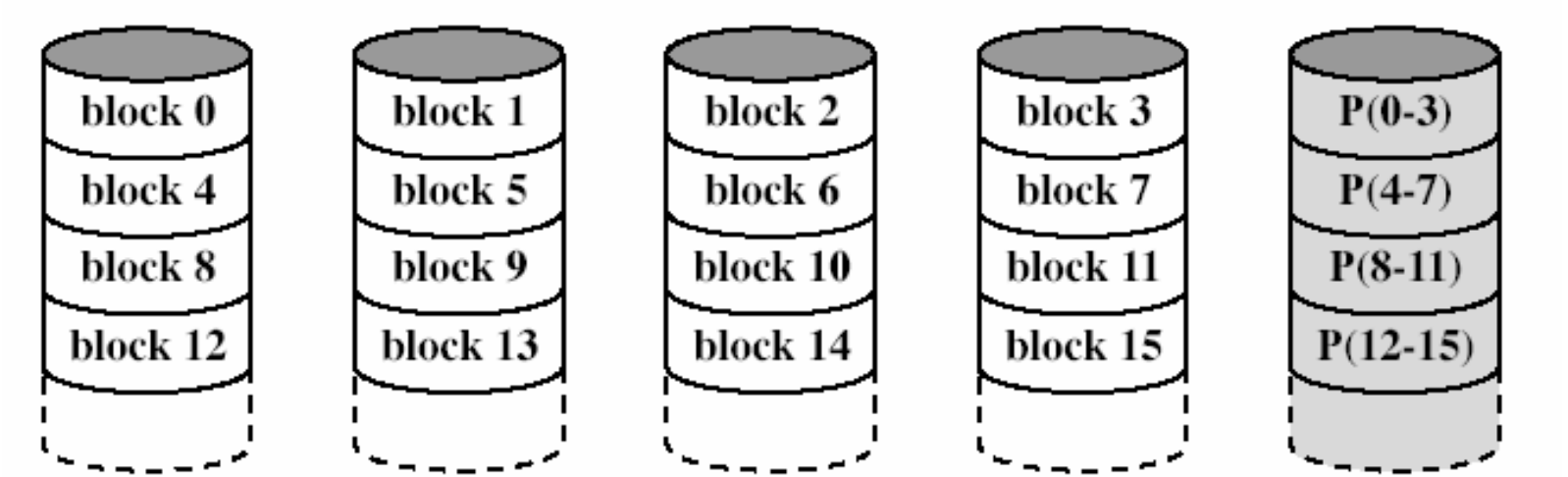
  Ze všech bitů se vytvoří paritní bit, vše se zaznamená (bez zápisu do vadného bitu – disk je vadný), po výměně disku se patřičné bity zrekonstruují.

- Výkon:

  Je vysoký, protože se čte z více disků současně, současně se však může vyřizovat pouze jeden požadavek.

  Nevýhoda: obtížně použitelné tam, kde je systém orientován na vyřizování transakcí.

RAID 4



Parita typu „block-level“.

RAID 4

- Každý disk je nezávislý.
- Vhodný pro aplikace s vysokým objemem V/V požadavků.
- Velké stripes.
- Parita se počítá bit po bitu přes celé stripes na každém disku.
- Parity se uloží na paritní disk.

RAID 4 - využití redundance

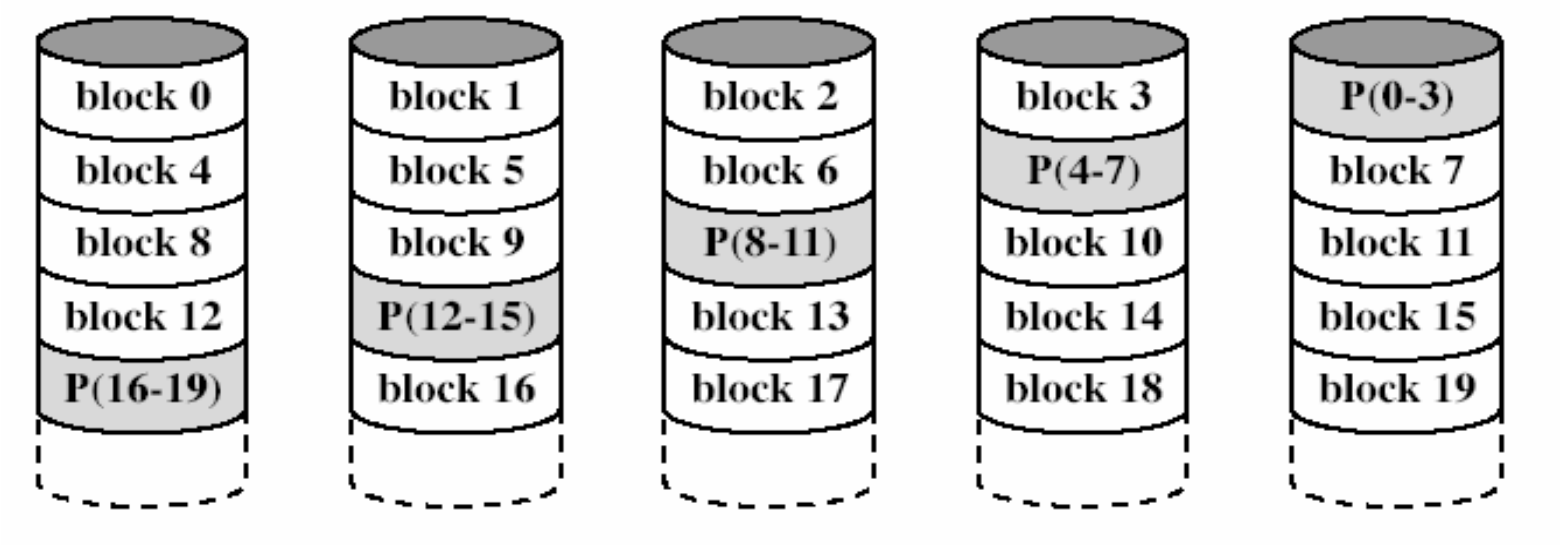
- Situace: potřebujeme zapisovat pouze na jeden disk – jak vypočteme paritu?
- Pole sestává z 5 disků: X0 – X3 – data, X4 – parita.
- X4(i) = X3(i) xor X2(i) xor X1(i) xor X0(i)

  = X3(i) xor X2(i) xor X1´(i) xor X0(i) xor X1(i) xor X1(i)

  = X3(i) xor X2(i) xor X1(i) xor X0(i) xor X1(i) xor X1´(i)

  = X4(i) xor X1(i) xor X1´(i)
- Má-li se vypočítat nová parita, tak se pro výpočet použije stará parita, původní hodnota bitu a nová hodnota bitu.
- Zápis: zapisuje se jak datový bit, tak i parita.

RAID 5

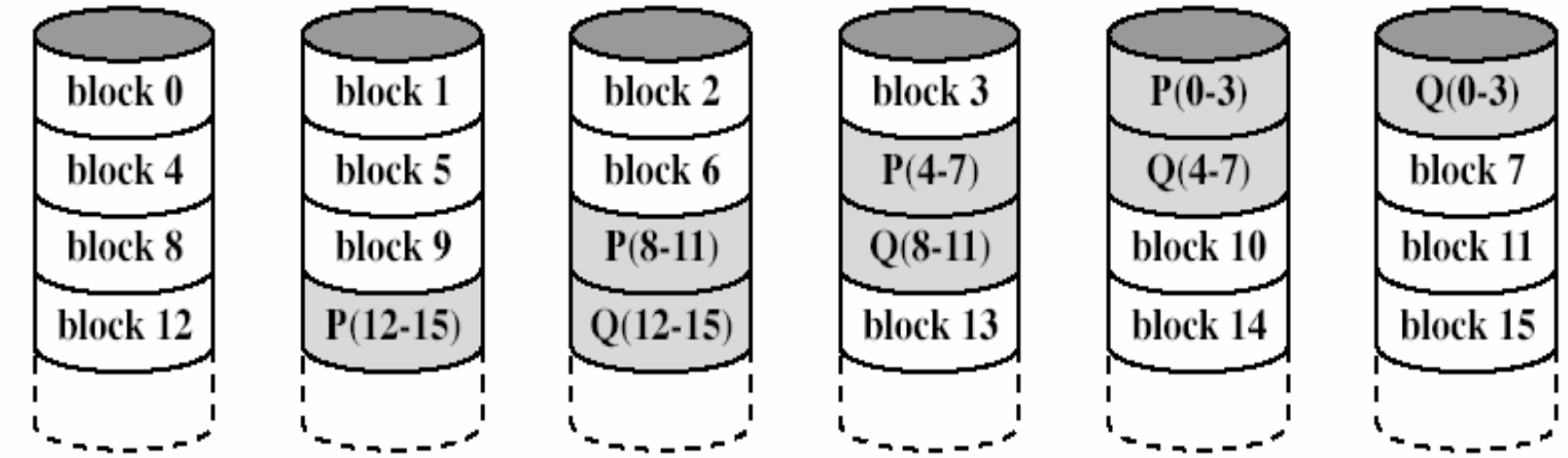


Parita typu „block-level distributed“

RAID 5

- Podobné s RAID 4
- Parita je uložena na všech discích.
- Všechny dosavadní mechanismy byly schopny napravovat problém, pokud nastal na jednom disku.
- Stav, kdy poruchu mělo více disků – neřešitelný.

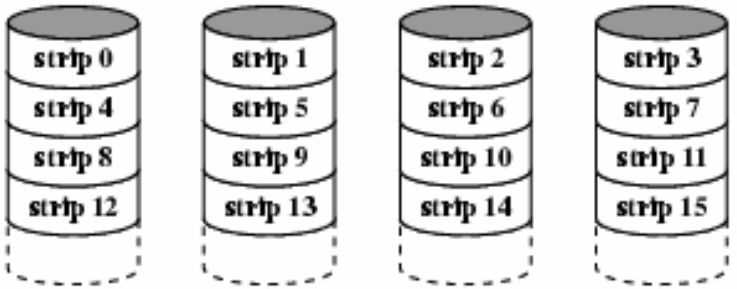
RAID 6


block 0
block 4
block 8
block 12
block 1
block 5
block 9
P(12-15)
block 2
block 6
P(8-11)
Q(12-15)
block 3
P(4-7)
Q(8-11)
block 13
P(0-3)
Q(4-7)
block 10
block 14
Q(0-3)
block 7
block 11
block 15

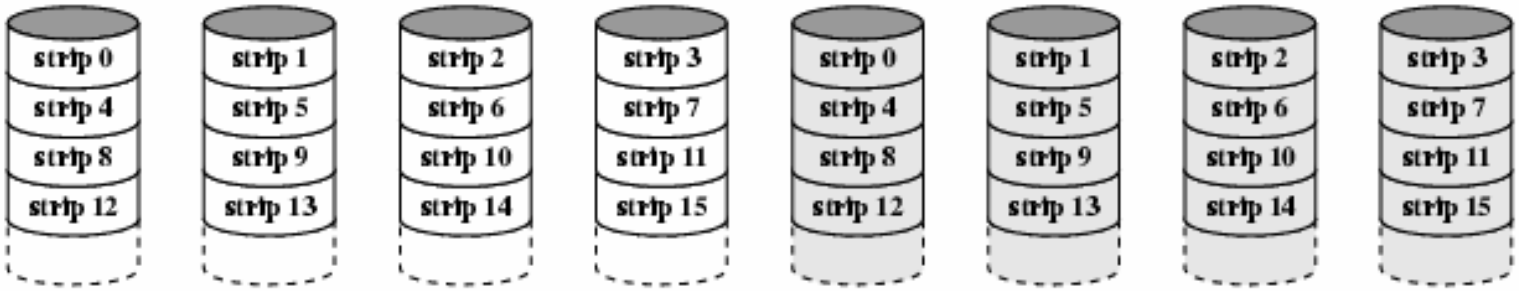
RAID 6

- Počítají se dvě parity.
- Parita se ukládá do samostatných bloků na různých discích.
- Je potřeba další dva disky navíc.
- Porucha dvou disků – je možná náprava dat.
- Porucha tří disků – neřešitelné.

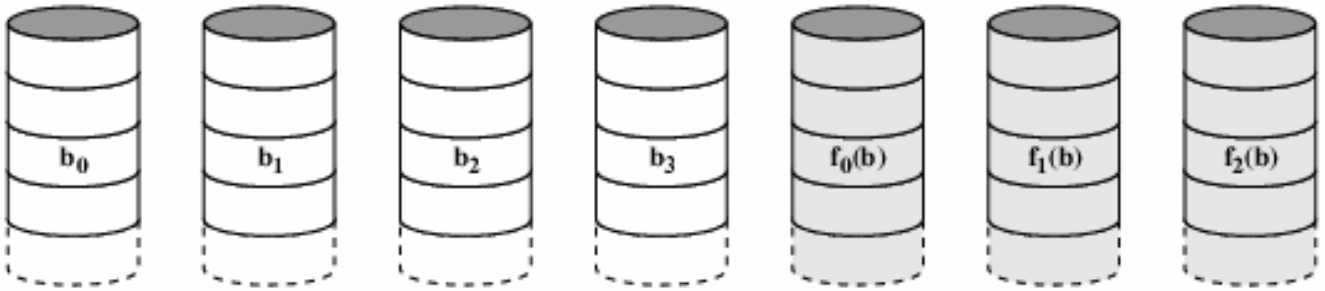
RAID 0, 1, 2



(a) RAID 0 (non-redundant)

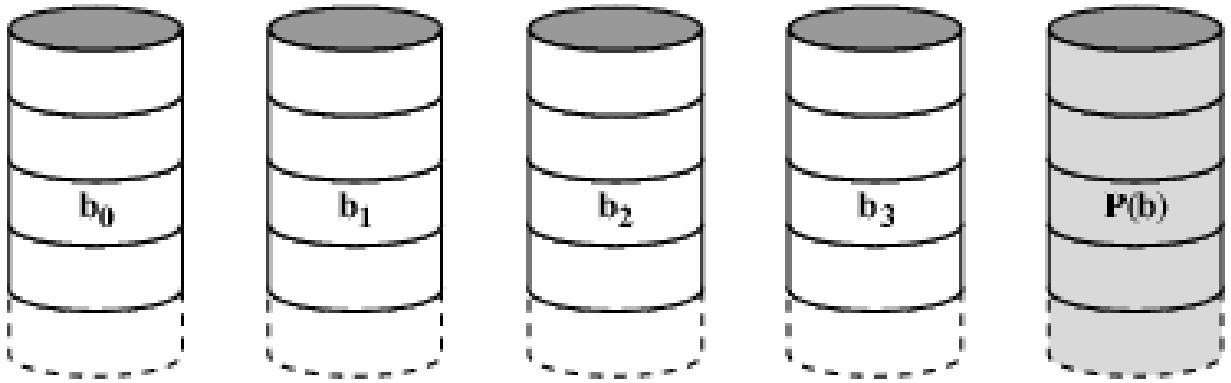


(b) RAID 1 (mirrored)

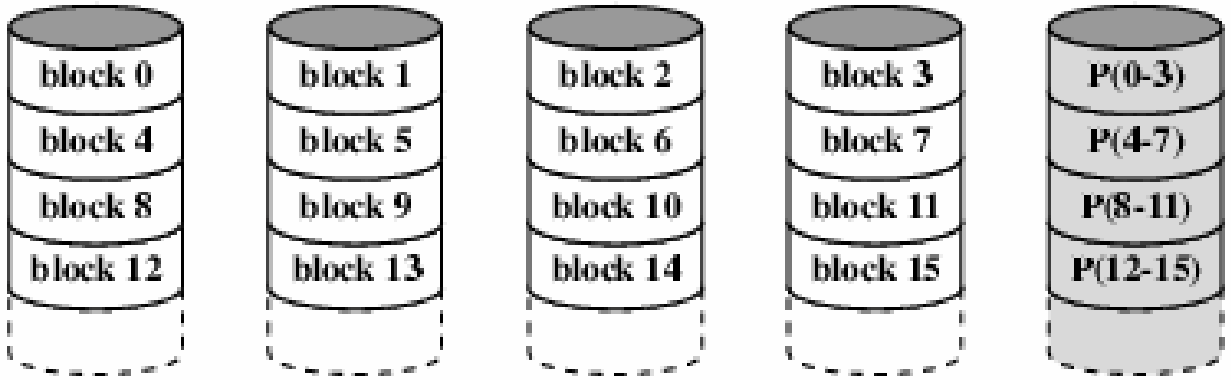


(c) RAID 2 (redundancy through Hamming code)

RAID 3 & 4

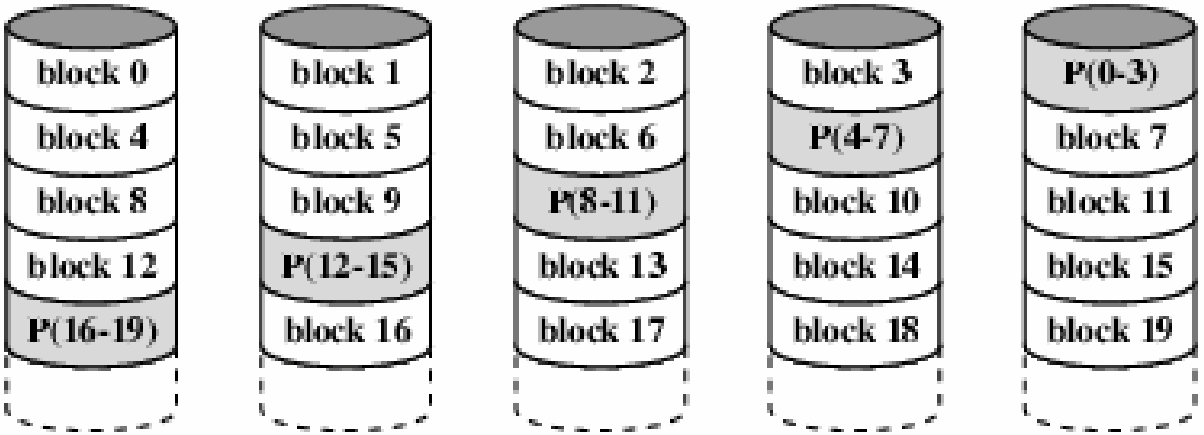


(d) RAID 3 (bit-interleaved parity)

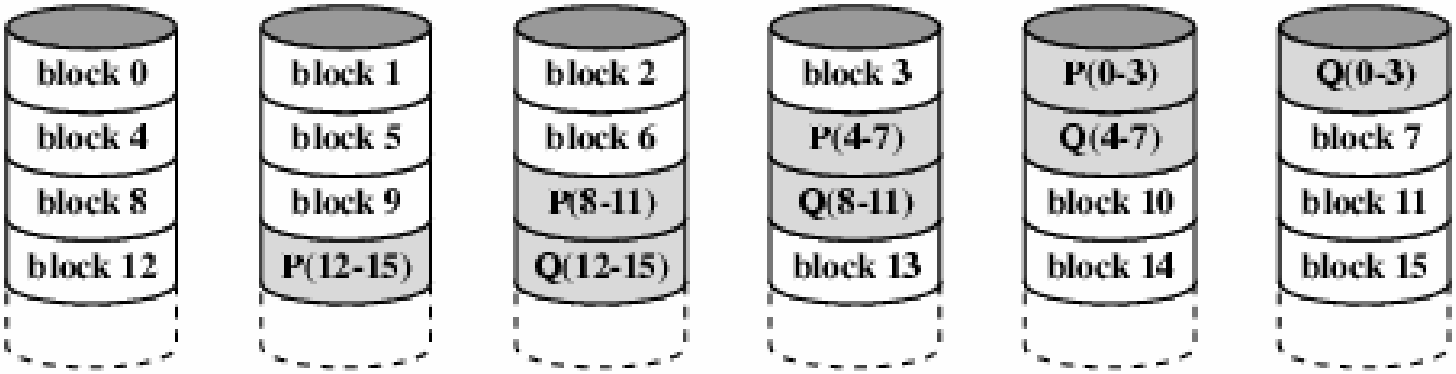


(e) RAID 4 (block-level parity)

RAID 5 & 6



(f) RAID 5 (block-level distributed parity)



(g) RAID 6 (dual redundancy)