

Pevné disky a SSD disky

Základní kostra:

- Co je pevný disk, základní popis konstrukce
- Co je SSD disk
- Porovnat mezi sebou z hlediska použití, výhod, nevýhod
- Parametry pevných a ssd disků
- Připojovací rozhraní
- Diskové konfigurace do diskových polí
- Nasazení v profesionálním světě

Co je pevný disk:

Elektromagnetická paměť typu RAM-RWM určená pro permanentní ukládání. Založeno na elektromagnetickém snímání a ovlivňování magneto-citlivé plochy.

Základní komponenty:

- Elektronika (obsahuje cache paměť, řídicí čipy, interface a případně část řadiče – jen u SAS)
- osa pro plotny
- motor pro plotny
- plotna (plotny)
- rameno
- osa pro rameno
- motor pro rameno
- čtecí a zápisové hlavy

Co je SSD disk?

Elektronická paměť, podobná flash principu, upravená pro použití jako trvalé úložiště. Typu RAM-RWM, non-volatilní.

Obsahuje: Řídicí elektroniku (tj. cache paměti, řídicí čipy, interface) a samotné paměťové čipy.

Porovnat mezi sebou z hlediska využití, výhod a nevýhod:

Pevný disk	SSD disk
Vysoké kapacity	Extrémní rychlost (podle modelu)
Variabilita provedení	Kompaktní rozměry
Neomezené zápisové cykly	Nižší spotřeba
Nízká cena za GB	Bez pohyblivých částí
Všechna provedení interface	Vysoká odolnost proti otřesům
Nízká odolnost proti otřesům	Vysoká cena za 1GB
Vyšší spotřeba napájení	Vyšší citlivost na ESD
Mechanické opotřebení	Omezený počet zápisových cyklů
	Neexistují některá provedení
	Může být problém s teplotami

	Nízké dostupné kapacity

Parametry pevných disků

Kapacita – jednotka TB, u SSD pak GB. Maximální kapacity dnes až 3TB, ale pouze u SATA (resp. MDL SAS) interface, u SAS a FC je maximální kapacita 900GB u LFF formátu a 600GB u SFF formátu.

Tovární provedení (velikost) – 2,5“, 3,5“ (nebo také SFF a LFF = Small form factor, Large Form Factor)

Výrobní technologie (u SSD) – MLC (levnější, pomalejší) a SLC (dražší, rychlejší)

Velikost vyrovnávací paměti – v MB (obvykle 32-64MB)

Rychlost otáček (pevné disky) – 4500 (netbooky, notebooky), 5400 (netbooky, notebooky), 7200 (notebooky, běžná PC, některé servery, pracovní stanice), 10000 (servery, pracovní stanice, komerční disková pole), 15000 (servery, komerční disková pole)

Určení – Málo sledovaná vlastnost, přesto podstatná. Každý disk je pro něco určen, tj. má určitou cílovou skupinu využití. A rozdíly jsou dost zásadní.

- Notebook disky (výlučně SATA) – disky určené pro provoz v zařízeních s akumulátory, tj. důraz na nízkou spotřebu energie, provozní režim 8/24, designovány na velmi častá vypnutí. Malé form factory (SFF).
- Energy save disky – speciální disky určené jako sklady pro velké kapacity bez požadavků na výkon ale s požadavkem na úspornost. Standardně v 3,5“ provedení (LFF). Úspory se docílují pomocí snížení rychlosti otáček.
- Desktop disky (výlučně SATA) – disky určené pro běžné použití v režimu 8/24 hodin, tedy provoz zhruba 8-12 hodin denně, jsou optimalizované na časté zapínání a vypínání
- Workstation disky (SATA) – výkonné disky určené pro náročné uživatele pro použití v polích i samostatně s režimem 16/24, čili cca 16-18 hodin denně, jsou designovány na pravidelné vypínání
- Server disky – SATA/MDL SAS – výkonnější disky určené do malých serverů, výlučně pro práci v diskových polích s provozním režimem 24x5 až 24x7, designovány na občasné vypínání pro servisní účely.
- Enterprise disky – SAS/FC – nejvýkonnější disky určené od středních serverů po velké BCS stroje (Business critical). Designovány na provozní režim 24x7x365 a na trvalý chod bez vypnutí (vypnutí se zde předpokládá řádově na desítky minut ročně)

Další funkce – U disků můžeme také vyhledávat různé typy funkcí. Především by disk měl podporovat **S.M.A.R.T.**, je to diagnostická funkce sledující chování disku podle určitých kritérií na jejichž základě vyhodnocuje, zda je disk ještě zdravý, nebo je potřeba ho vyměnit a to ještě dřív, než k závadě dojde. S.M.A.R.T. musí ale podporovat i základní deska resp. čipová sada.

NCQ – speciální technologie firmy Seagate určená pro inteligentní řazení dat pro zpracování

Hotplug/hotswap – je technologie umožňující výměnu pevného disku v diskovém poli za chodu, jedná se o elektronickou ochranu proti poškození disku při vyjmutí za chodu.

Připojovací rozhraní

V současnosti jsou všechna rozhraní sériová. (v minulosti byla paralelní, u PATA 2 disky na kabel, u SCSI až 7 disků na kabel, sdílela se rychlost kabelu a rychlost se přizpůsobovala nejpomalejšímu zařízení).

SATA – nejčastější rozhraní, velmi jednoduché, navrženo na snadnou instalaci, má i externí provedení, umožňuje hotplug/hotswap ve spojení s řadiči SAS. Připojuje se červenými plochými kabely (6 vodičů). Diskový řadič obvykle přímo v čipsetu, ale může existovat i v externích provedeních (karty, čipy na základní desce). Současný standard SATA III, nebo také SATA6G s maximální propustností per port 6Gbit/s.

SAS – Enterprise, resp. podnikové řešení interface. Založeno na SCSI architektuře v sériovém provedení. SCSI (Small Computer Storage Interface) je speciální komunikační protokol mezi pevnými disky a příslušnými řadiči. Je založen na optimalizaci a garanci datových toků, umožňuje prioritizovat jednotlivé datové toky, umožňuje řídit poměr rychlosti zápisu a čtení, umožňuje používat vyrovnávací paměť řadiče pro zrychlení datových toků. SAS rozhraní je drahé na výrobu, je složité a vyžaduje instalaci části „řadiče“ na pevný disk. Řadiče nejsou přímo v čipsetu, ale buď v podobě externích čipů, případně externích karet. SAS umožňuje v základu Hotplug/hotswap a podporuje vylepšený monitoring stavu pevného disku. Připojuje se na straně pevného disku podobným konektorem jako SATA, obsahuje navíc čtyři piny pro SCSI komunikaci. Napájení mají oba typy interface shodné.

Díky specifickým vlastnostem se SAS používá především pro profesionální sféru serverů, diskových polí, najít ho ale lze i v pracovních stanicích.

SATA i SAS jsou vzájemně částečně kompatibilní. SATA disky je možné připojit k SAS řadičům, ale nikdy ne SAS disky k SATA řadičům (chyběla by podpora pro SCSI). SATA to SAS se používá tam, kde je potřeba velká kapacita s nižším výkonem a spolehlivostí (například různé formy záloh, nebo velkokapacitní úložiště pro nekritická data).

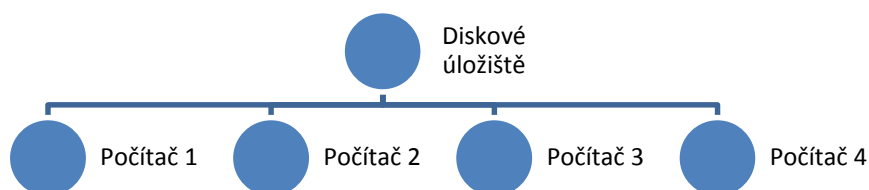
Základní rozdíly SAS/SATA

SATA	SAS
Jednoduché	Složitější
Levné	Nepoměrně dražší
Maximální možné capacity	Omezená kapacita (max. 900GB)
Pomalejší	Rychlé
Méně spolehlivé	Vysoce spolehlivé
Chybí interní inteligence	S interní inteligencí a monitoringem
Možné připojit k SATA i SAS řadičům	Možno připojit jen k SAS řadiči

SATA i SAS jsou interface, které se používají přímo proti pevným diskům a v rámci jednoho počítače.

Existují však i rozhraní, která mohou být sdílena několika počítači v jednom okamžiku, tj. umožňují zapojení do tzv. SAN (Storage area network), tedy do sítě postavené z diskových úložišť.

Čili v zásadě nabídneme nějaké diskové úložiště (tzv. Storage) více počítačům zároveň. Viz schéma.



Využívá se to především pro tzv. clustering, čili vytváření virtuálního systému z několika systému fyzických (navenek se celé řešení tváří jako jeden počítač).

Typy SAN interface:

Shared SAS – tato varianta využívá výše uvedený interface, který jenom umí nabídnout více počítačům zároveň, je to extrémně levné, ale má to značná technická omezení. Rychlost připojení úložišť je v násobcích 6Gbit/s

iSCSI – nejpodobnější klasické síti, využívají se standardní síťové prvky (ethernet standard), v počítačích iSCSI potřebuje pouze síťovou kartu. Lze ho provozovat na současných rychlostních standardech Gigabit ethernet, 10Gigabit a 40Gbit. Optimální variantou je 10Gbit. Dražší než SharedSAS a podstatně složitější na návrh.

FC (Fibre channel) – Je nejvyšším stupněm interface. Je postaven na čistě optickém řešení kabeláže, má nejvyšší inteligenci a umožňuje připojit přes sebe i pevný disk (na rozdíl od iSCSI). V počítači vyžaduje tzv. HBA (Host Bus adaptér). Nominální rychlost jednoho připojení je 8Gbit/s. (existují i 4Gbit/s a 2Gbit/s). Je to velmi drahé, ale absolutně spolehlivé s možností připojovat pevné disky na desítky kilometrů. Současný standard pomalu začíná adaptovat i novou rychlost 16Gbit/s. Základní konfigurace FC SAN obsahuje:

- a. HBA – karta v počítači
- b. SAN switch – speciální přepínač pro protokol Fibre channel
- c. Diskový kontroler (řadič)
- d. Diskové police s jednotlivými disky

Pro ilustraci se nejnižší cena FC zařízení pohybuje v řádu set tisíc Kč. (řadiče kolem 100k, switche cca 80k, HBA ca 20k, optické vysílače cca 6 za kus. Cena jednoho přípojného bodu do počítače se pohybuje kolem 40k. (20k HBA, 2x6k za optické vysílače).

Budoucnost SAN se bude pravděpodobně opírat o kombinace výše uvedených technologií. Současnost například postupně vyřazuje FC jako připojovací interface směrem k diskům a nahrazuje jej levnější, ale stejně výkonnou technologií SAS. Hlavní vývoj však lze očekávat na straně SAN jako takové. Již v současnosti existují technologie jako FCoE (Fibre channel over Ethernet), které využívají

standardu klasických sítí (včetně nastupujících standardů 40Gbit Ethernet a 100Gbit Ethernet) a které skrz ně propouštějí komunikační standard FC. V počítači pak místo dedikovaného (vyhrazeného) host bus adaptéru najdeme tzv. CNA (Converged Network Adapter), který po jednom kabelu dokáže transferovat jak diskovou komunikaci, tak standardní síť. Stále se však jedná spíše o hudbu budoucnosti. (především kvůli obrovským nákladům, které jsou ještě vyšší, než u klasického Fibre channelu).

Diskové konfigurace do polí

Diskové pole – termín označuje logické seskupení více pevných disků/SSD disků do jednoho zařízení. Tedy máme-li několik disků shodné kapacity (např. 6x100GB), navenek se v konfiguraci diskového pole budou tvářit jako disk jeden s různými kapacitami (600GB, 300GB, nebo 400GB, nebo 500GB podle konfigurace).

Diskové pole má buď hardware, nebo software podobu. Jedná se tedy jen o jakýsi algoritmus schopný spojit více diskových zařízení dohromady.

RAID – Redundant Array of Independent disk – označuje právě onen algoritmus, který se data na disky sdružené do diskového pole, ukládají.

RAID konfigurace se používá tam, kde je potřeba:

- a. Vyšší výkon (využívá se potenciál všech disků dohromady)
- b. Vysoká spolehlivost a dostupnost (tam, kde si nemůžeme dovolit vadný disk)

Důležité RAID režimy:

- a. **RAID0** – je režim kdy spojujeme do diskového pole dvojici pevných disků, výsledkem je sečtená kapacita (u 100GB disků to pak je 200GB). Je velmi rychlé (výkony se sčítají), ale také velmi nespolehlivé (pád jednoho z dvojice disků = ztracená data. Hodí se tam, kde nejsou důležitá data, ale výkon. (např. caching pro video atd.)
- b. **RAID1** – je režim, kdy spojujeme do diskového pole dvojici pevných disků, výsledkem je kapacita jednoho z dvojice disků (U 100GB disků to je 100GB). Je to velmi rychlé (výkony se sčítají), ale také velmi spolehlivé, toleruje se výpadek jednoho disku, pokud vypadne, data pořád zůstávají kompletní. Tento RAID se používá především pro instalace operačních systémů.
- c. **RAID5** – nepoužívanější typ RAID režimu, minimálně musí být tři disky. Data se rozkládají v prouzcích (stripe) a každý uložený blok dat má vždy na dvou discích data a na jednom disku tzv. paritu, pomocí které je možné provést dopočet při výpadku jednoho z disků. (čili je to sada informací z obou provozních disků, ze kterých umíme sestavit výsledný obsah při výpadku jednoho z disků. Je velmi rychlý na čtení, trochu pomalejší na zápis, nicméně lze do RAID5 konfigurace použít celkem vysoký počet disků (běžně až 16), a tím se pak razantně násobí i výkon. RAID5 toleruje vždy výpadek jen jednoho disku.
- d. **RAID6** – Nejvyšší standardní konfigurace, Opět se ukládají data a tzv. parita, u RAID6 dokonce dvojí. Tolerován je tedy výpadek až dvou disků. Minimální počet disků jsou čtyři. Tento typ pole je celkem rychlý na čtení, ale velmi pomalý na zápis (důvodem je nutnost zápisu dvojí parity). Proto na RAID6 konfiguraci je vhodných min. 8 disků, kde výkonnost na zápisu už celkem vyhovuje. Standardní maximum je opět 16 disků.

Existují i další RAID režimy, například 10,50,60 a 100. Jedná se však vždy jen o analogie 0,1,5,6. Např. RAID10 jsou dvě a více RAID0 pole spojená ještě jednou dohromady do RAID1 (RAID0+RAID0 spojená jako RAID1), minimem jsou pro tento typ pole čtyři disky (2 a 2 disky). RAID50,60 jsou obdobné, ale vycházejí z RAID5 a RAID6.

Rovněž se občas používají tzv. Spare disky. Lze si to představit tak, že vyrobíme standardní diskové pole RAID5 se sedmi disky (budeme tedy tolerovat výpadek jednoho disku a přijdeme o 1/7 kapacity) a k němu přiřadíme ještě tzv. spare. Tento disk nebude součástí diskového pole, ale bude v pohotovosti čekat. V okamžiku, kdy dojde k poruše jednoho z disků v RAID5 poli, Spare se automaticky aktivuje a systém sám dopočítá na tento spare disk veškerá data. Vadný disk se vymění a stane se z něj další spare.

Využití ve firemním prostředí

Ve firemním prostředí dominují klasické pevné disky nad disky SSD. Především z důvodů extrémně vysoké ceny, v porovnání se SSD disky v běžné prodejní síti, stojí SSD disky pro servery až desetinásobek ceny při shodné kapacitě.

Ve firemním prostředí se využívá konfigurace diskového pole takřka bez výjimky. Jde o oba parametry zmíněné výše tj. rychlost i dostupnost (spolehlivost).

Více se využívají disky typu SAS.

Tam, kde je potřeba výkon, dochází ke slučování spíše menších kapacit pevných disků, ale v o to větším počtu. Není výjimkou pole s 16ti disky x 146GB.

Využívají se hardwarové typy diskových polí se sdílenými interface, tj. Shared SAS, iSCSI, Fibre channel. Tato zařízení umožňují škálovat i na řády stovek, či tisíců pevných disků v jednom logickém systému.

Firmy často používají SAN infrastrukturu pro geografické rozložení dat (tj. například diskové pole rozložené mezi centrálou a pobočkou atd.)

V současnosti existují disky SAS-MDL, což jsou interně disky typu SATA, které ale mají více elektroniky a hlavně jsou na vstupu typu SAS. Jsou o trochu výkonnější než SATA disky. Hlavním důvodem je dostupnost všech SAS funkcí i na úrovni disku typu SATA.

Firmy používají RAID řešení, používají se všechny typy kromě RAID0, která se využívá jen pro specifické účely.

Na vyšší RAID režimy je potřeba výkonný řadič obvykle typu SAS (ale i FC), řadič vyžaduje cache paměť zálohovanou buď z baterie, nebo z flash paměti (kvůli možnému výpadku napájení). Velikost takové cache dnes dosahuje až 2GB DDRAM.

Pevné disky ve firmách se mění v daleko kratších cyklech.

Výrobci:

Western Digital, Seagate, Hitachi, Samsung

SSD disky: Corsair, Intel, OCZ, Patriot, Kingston atd.

Praktické ukázky:

Ukázka diagnostiky pevného disku, výpis SMART, ukázka měření teplot atd.

Ukázka konfigurace disků do diskového pole

Ukázka připojovacích rozhraní SAS a SATA (porovnání)

Ukázka technologie Hotswap.

Ukázka SAN protokolu iSCSI

Ukázka konfigurace SAN diskového pole