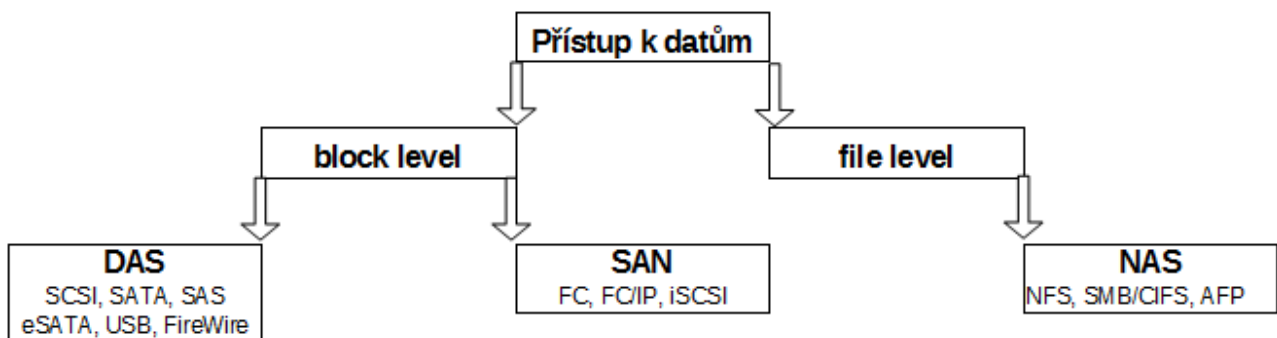


1 Dělení úložišť podle přístupu k datům

Základní rozdělení datových úložišť je podle způsobu, jakým přistupují k uloženým datům. Mohou získávat **soubory** (adresují soubory) pomocí souborově založených (file-based) protokolů jako NFS nebo SMB/CIFS. Nebo přistupovat přímo k médiu a číst **diskové bloky**, které převádí na soubory, což je běžné u lokálně připojených disků nebo SAN sítí.

Obr. 1 – dělení storage - Directly Attached Storage, Storage Area Network, Network attached storage

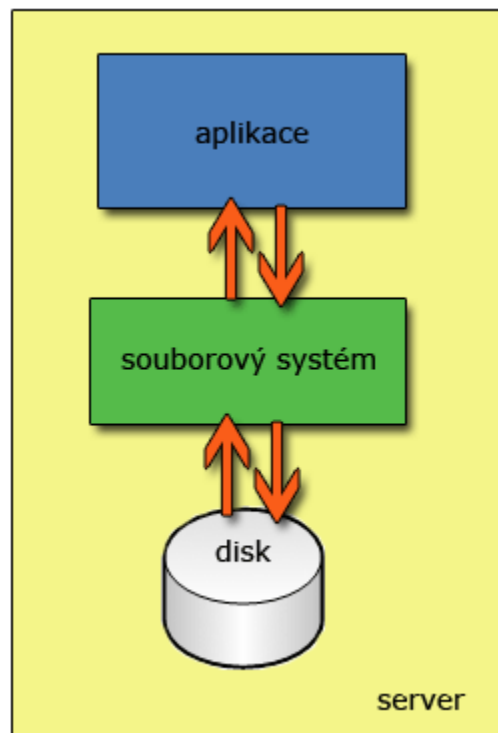


2 Architektury připojení storage

2.1 DAS - Directly Attached Storage

Úložné zařízení (disky) je připojené přímo k počítači. Jedná se o disky, které jsou připojeny přímo k internímu nebo externímu diskovému řadiči. Tedy buď disky v počítači, nebo jednoduché diskové pole, které je připojené k počítači spojem Point to Point.

Obr. 2 – schéma DAS



Příkladem DAS zařízení jsou klasické magnetické disky (HDD), SSD, optické mechaniky a úložiště na externích discích.

Hlavními rozhraními pro připojení DAS jsou protokoly **ATA, SATA, eSATA, NVMe, SCSI, SAS, USB, IEEE 1394** (Firewire), ale i **FC (Fibre Channel)** v režimu **point-to-point**.

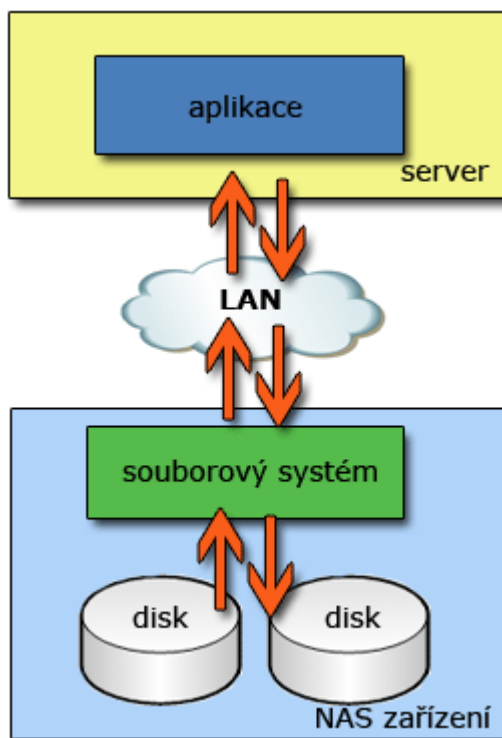
2.1.1 Výhody a nevýhody DAS

DAS je přístupné z počítače, ke kterému je připojené pomocí **host bus adaptéru (HBA)**. Mezi počítačem a DAS není potřeba síťové zařízení (hub, switch, router), mezi DAS a počítačem je vztah 1:1.

2.2 NAS - Network Attached Storage

NAS je označení pro datové úložiště připojené k místní síti LAN. Data toho úložiště mohou být poskytována různým uživatelům. NAS nemusí mít pouze funkci souborového serveru, ale může mít i jiné specializované funkce. Například klient P2P sítě, webový server, UPnP/DLNA server a další. Většinou obsahuje nějaký vestavěný počítač, který má za úkol sdílení dat a podporu různých protokolů.

Obr. 3 – schéma NAS



Tato zařízení si získala popularitu od roku 2010, když začala být používána pro sdílení dat mezi několika počítači. V porovnání mezi jinými síťovými úložišti jsou NASy rychlejší, mají lehčí administraci a snadnější nastavení.

NAS obsahuje jeden a více pevných disků, které se můžou slučovat do větších datových struktur nebo můžou vytvořit RAID pole. Levnější zařízení podporují hlavně RAID 0 a 1. Přístup k datům je většinou pomocí protokolů NFS, SMB/CIFS nebo AFP.

2.2.1 Historie NAS

V roce 1980 firma Newcastle Connection vedená Brianem Radnellem a jeho kolegy představila a demonstrovala v Newcastle Univerzitě vzdálený přístup k datům přes několik počítačů používající UNIX. Serverový operační systém Novell NetWare a NCP protokol byly vytvořeny až v roce 1983. Stejně s tím vytvořila firma Sun Microsystems v roce 1984 systém NFS, který dovozoval jejím síťovým klientům sdílet jejich data po síti. Firmy 3Com, IBM a Microsoft vytvořily protokol LAN Manager. 3Com 3Server a 3+Share software vytvořili první vestavěný server pro open system servery. Tento server obsahoval potřebný software, hardware a podporu několika disků.

Od této doby začala, díky dobré odezvě zákazníků, vytvářet celá řada firem svoje zařízení. Zatímco 3Com byla první firma, která vytvořila vestavěný NAS pro osobní počítače, tak Auspex System byli první, kdo vytvořil vestavěný NFS server pro použití v UNIX marketu. Skupina Auspex inženýrů se spojila, aby vytvořila v roce 1990 integrovaný NetApp filter, který měl podporovat jak Windowsovský CIFS a UNIXovský NFS protokol. Měl by být taky lehce nasaditelný a měl obsahovat jednoduchou administraci. Toto odstartovalo výrobu opravdových NAS, které známe dnes.

V roce 2009 začali výrobci NAS (hlavně CTERA Networks and NETGEAR) představovat možnosti online zálohy dat. Tato funkce byla přímo integrovaná v zařízeních a měla podporovat přímo obnovení zálohy online.

Obr. 4 – ukázky NAS zařízení



2.2.2 Použití NAS

NAS může být využíván i více způsoby, než jenom jako centrální jednotka, která poskytuje různým uživatelům data. Může být použit také jako levný a jednoduchý email nebo web server poskytující úložiště. Největší využití nacházejí NAS tam, kde uživatel potřebuje uložit velké množství multimediálních dat. V dnešní době je NAS levnější než rackový server, takže se menším firmám vyplatí používat NAS kvůli ceně, ale například i kvůli velikosti. Cena se liší hlavně v tom, kolik NAS podporuje připojených disků, jaké RAID podporuje a jestli obsahuje nějaké další připojení (například FireWire).

2.2.3 NAS a open source

Mezi NASové distribuce Linuxu a FreeBSD patří FreeNAS, CryptoNAS, NASLite, Gluster, Openfiler, OpenMediaVault a TurnKey souborový server postavený na základu Ubuntu. Všechny distribuce jsou vytvořeny tak, aby se pomocí nich dal NAS snadno nastavit, a většina z nich používá k nastavení webový prohlížeč.

Můžou být spuštěny z virtuálních zařízení, LiveCD, jako bootovatelný flashdisk nebo přímo z připojeného disku. Podporují protokol Samba (SMB démon), protokol NFS (NFS démon) a protokol FTP (FTP démon), které jsou v těchto distribucích volně dostupné.

Dražší řešení umí nabídnout funkce jako je zachycování obrazu disku (tzv. snapshot), zrcadlení, počítání kontrolních součtů atd.

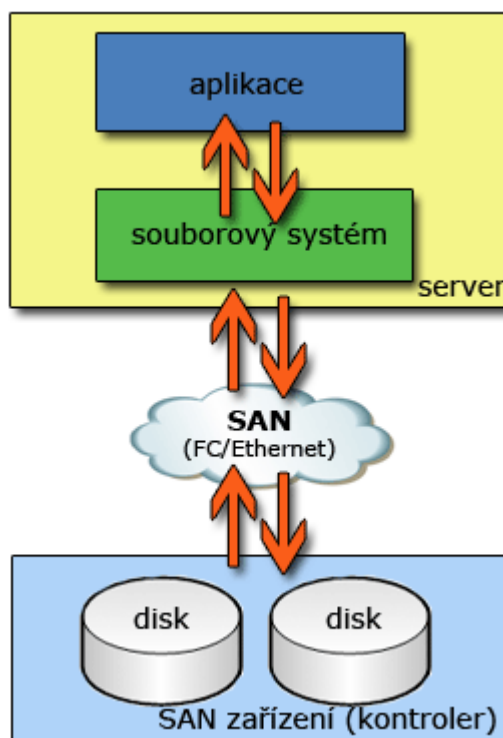
2.2.4 Protokoly používané u NAS

- CIFS (SMB)
- NFS
- FTP
- SFTP
- HTTP
- UPnP
- Apple Filing Protocol (AFP)
- rsync
- Andrew file system (AFS)

2.3 SAN - Storage Area Network

Storage area network (zkratka SAN) je dedikovaná (oddělená od LAN, WAN, atd) datová síť, která slouží pro připojení externích zařízení k serverům (disková pole, páskové knihovny a jiná zálohovací zařízení).

Obr. 5 – schéma SAN



SAN vznikla hlavně kvůli narůstajícím potřebám na zabezpečení a konsolidaci dat. Díky poměrně vysokým pořizovacím nákladům se SAN budují hlavně ve větších společnostech, které vyžadují vysokou dostupnost svých služeb, rychlé odezvy v transakčně orientovaných aplikacích a co největší škálovatelnost (rozšiřitelnost). Další oblastí, která zvyšuje použití SAN je virtualizace (migrace VM mezi fyzickými nody). V poslední době se však díky klesající ceně začínají malé SAN rozšiřovat i do stále menších společností. SAN zařízení tak postupně nahrazují tzv. DAS (Direct Attach Storage) technologii, kde součástí serveru byly všechny disky či zálohovací zařízení, nejčastěji připojené pomocí SAS (SCSI) řadiče či interního RAID řadiče.

2.3.1 Výhody a nevýhody SAN

DAS koncept má samozřejmě řadu nevýhod. Není možné fyzicky oddělit data a datové zdroje od samotného serveru a mnohdy se stává, že komplikovanější havárie serveru, či živelní pohroma (požár, povodeň a jiné) způsobí ztrátu dat. Dále mohou nastat obtíže při migraci dat na jiný server, servery mají omezenou diskovou kapacitu a není prakticky

možné mít tzv "no single point of failure" architekturu. Naproti tomu SAN oproti DAS nabízí:

- Fyzické oddělení dat a serverů. Jednotlivé prvky mohou být od sebe vzdáleny až desítky km
- Sdílení zdrojů (diskových zařízení, zálohovacích zařízení) mezi jednotlivými servery (velmi omezeně umí toto i DAS)
- Vyšší propustnost
- Definici redundantních cest ke zdrojům (velmi omezeně umí toto i DAS)
- Podporu pro clusterová řešení a architekturu "no single point of failure"
- Podpora pro tzv. Disaster Recovery sites

2.3.2 Druhy sítí

Běžné SAN sítě využívají Fibre Channel protokol (SCSI), kde jako nosné medium je použitý optický kabel (možno ale i metalika).

- **Fibre channel protocol (FCP)**, mapující SCSI přes optické nebo metalické medium (používá se hlavně optika, metalické spoje jsou definovány standardem). Aktuálně nejběžněji používaný. Ve variantách 1 Gbit/s, 2 Gbit/s, 4 Gbit/s, 8 Gbit/s, 10 Gbit/s, 16 Gbit/s, 32 Gbit/s a 128 Gbit/s.
- **FICON (Fibre Connection)** mapování přes Fibre Channel (užívaný mainframe servery).
- **iSCSI, iFC** , mapující SCSI nebo FC přes TCP/IP
- **HyperSCSI** , mapující SCSI přes Ethernet
- **ATA** přes Ethernet, mapování ATA přes Ethernet
- **SCSI a/nebo TCP/IP** mapování přes InfiniBand (IB)

2.3.3 Topologie FC

Fibre channel protokol definuje 3 různé topologie:

- Point-to-Point – nejjednodušší topologie, nelze rozšiřovat, 2 zařízení přímo propojená
- Arbitrated Loop – všechna zařízení zapojená do kruhu (obdoba token ringu), od roku 2010 téměř nepoužíváno
- Fabric – obdoba moderního ethernetu

Samotná SAN síť využívá hlavně Fabric topologii, s ostatními je možné se setkat například v koncových zařízeních pro propojení jednotlivých komponent. V praxi se třeba Arbitrated

Loops se používá v některých diskových polích pro propojení řadičů a jednotlivých expanzí s disky.

2.3.4 Základní stavební kameny fabric sítí

Host Bus Adapter (HBA) - Speciální karta, která připojuje jednotlivé servery do SAN. Analogií v DAS architektuře by byl SCSI (RAID) řadič. Každé HBA (a kterýkoliv jiný node či uzel v SAN) má unikátní 64-bitovou **WWN** (world wide name) adresu (analogie ethernetovské MAC adresy), která ji jednoznačně identifikuje v rámci SAN. Adresa se skládá z prefixu výrobce a samotného čísla karty.

Obr. 6 – HBA karta



Parametry:

8/4/2Gbps auto-negotiation

režimy point-to-point, switched fabric

2x LC konektor na MM vlákna

2Gbps na 500m, 8Gbps na 150m

PCI Express x8

Fabric Switch - zařízení, které propojuje jednotlivé prvky v rámci fabric SAN a zajišťuje jejich komunikaci (analogie ethernet switchu). Uzlem může být buď nějaký server (HBA), koncové zařízení nebo jiný FC switch. Každý switch obsahuje určitý počet portů pro připojení těchto zařízení. Vyrábějí se jak malé 8 nebo 16 portové switche tak i velké (tzv directory), které jich obsahují až 256. Počet zařízení v celé fabric je limitován na 65535 prvků.

Obr. 7 – FC switch – Qlogic SANbox – 20x port (LC) 8 Gbps, 4x 10Gb stack port, dual PSU



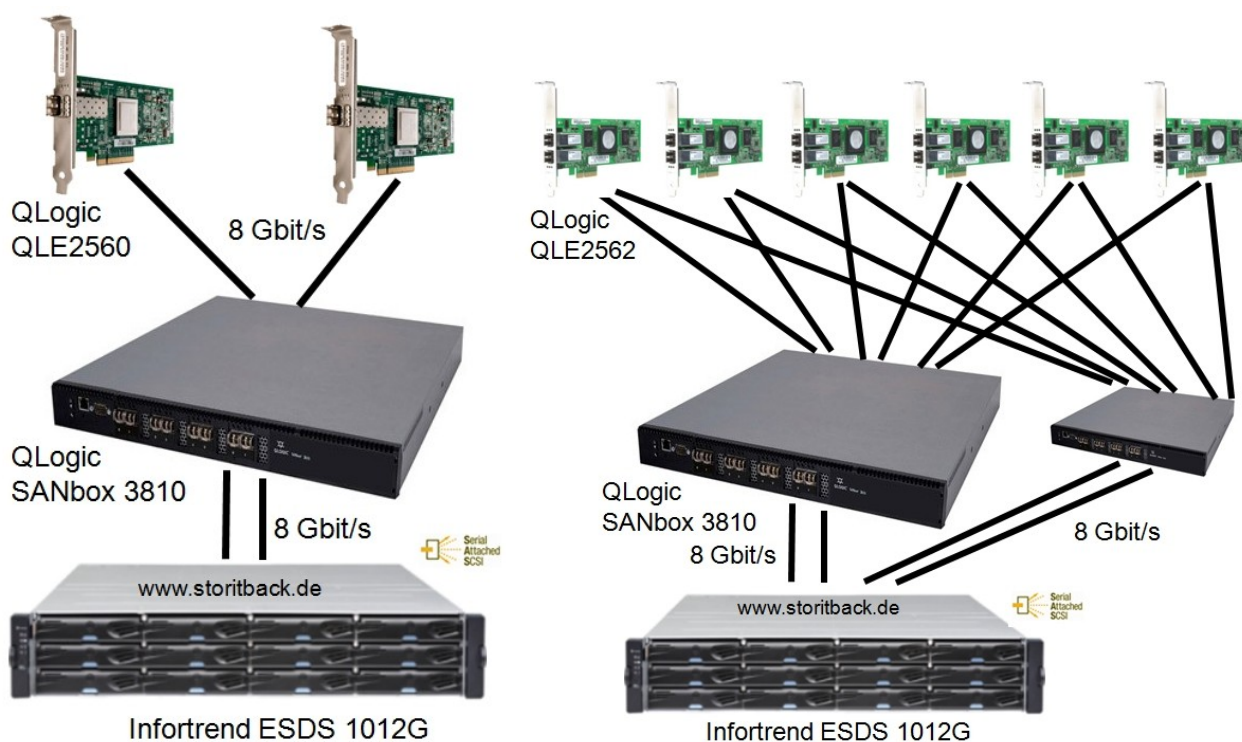
Koncová zařízení - většinou disková pole nebo jiné zařízení pro ukládání dat, archivaci či zálohování.

Obr. 8 – FC switch Infortrend - ESDS 3016GE – pomocí host modulu se volí cílová technologie k serverům

ESDS 3016GE Detailed front and back product views



Obr. 9 – schéma zapojení SAN - siglepath a multipath



Obr. 10 – SAN – celkové zapojení (CDN77 Rack Chicago 22. 9. 2016)



2.3.5 Komunikace v rámci fabric

Jeden z nejdůležitějších mechanismů je v rámci fabric je tzv. **zoning** (analogií v sítích Ethernet je protokol 802.1q – tzv. VLAN), pomocí nějž může správce SAN definovat pomocí zón zařízení, které spolu mohou komunikovat. Například pokud má naše infrastruktura diskové pole, páskovou knihovnu, 5 serverů a jeden zálohovací server, může být zónování navrženo tak, aby všechny servery byly schopny komunikace s diskovým polem, a pouze zálohovací server bude mít nadefinovanou zónu pro další komunikaci s páskovou knihovnou. Zoning je propagován v celé SAN a stačí tedy definovat příslušnou zónu na kterémkoliv fabric switchi. Používají se 2 způsoby:

- Definicí zón nad jednotlivými fyzickými porty switchů
- Definicí zón nad WWN adresami

V praxi je lepší používat zóny nad WWN, protože předejdeme potížím s komunikací při přepojení zařízení z jednoho portu do jiného. Port zoning se používal hlavně dříve, kdy wwn zoning ještě nebyl vždy podporován. Použití zoningu je ve většině případů holá nutnost. Jednak některá zařízení to přímo vyžadují, ale hlavně předejdeme nepředvídatelným okolnostem, které mohou způsobit některé nody přistupující na zařízení komunikující již s jiným nodem apod.

2.3.6 LUN masking

LUN masking zajišťuje další granularitu SAN. Jak jsme si popsali výše, zoning tedy zajišťuje rozdělení fabric na nody, které spolu mohou komunikovat a které ne. To však pro většinu zařízení není postačující. Pokud máme například diskové pole se 2 raid kontrolery, zoningem pouze definujeme, které servery s těmito kontrolery mohou komunikovat. Na diskovém poli však mohou být definovány desítky a desítky logických disků a i ty je potřeba přiřadit konkrétnímu serveru. Proto se na diskových polích definuje tzv mapování. Každý virtuální disk dostane přiřazen tzv Logical Unit Number (LUN) a ten se poté mapuje na konkrétní WWN adresy v SAN. Kombinace zoningu a mapování tedy zajistíme, že všechny servery vidí diskové pole, ale každý server vidí jen disky, které mu patří.

2.3.7 Sdílení médií

V rámci SAN je možné sdílet jednotlivá zařízení, ať jsou to třeba virtuální logické disky nebo páskové knihovny. To se hodí například při budování clusterů, filesystémů na kterých požadujeme paralelní operace a nebo LAN-less či Server-less zálohování, kdy veškerý objem dat zálohujeme přes SAN síť a významně tím šetříme provoz na LAN síti.

SAN v podstatě umožňuje spojování skladovacích „ostrovů“ použitím vysokorychlostní sítě. Operační systém vidí zařízení v SAN jako lokálně připojené zařízení. V případě logických disků vypublikovaných na diskovém poli vidí řadu lokálních disků, na kterých může nakonfigurován některý z filesystémů který daný OS podporuje, či některý z disků může být přímo použit například databázovým softwarem (viz Oracle DB na raw volumech či ASM). Jeden z problémů přístupu k datům v SAN je ten, že SAN řeší sdílení logických disků mezi jednotlivými nody, avšak OS většinou nikoliv, protože současný design většiny

filesystémů neumožňuje paralelní přístup. Pokud je tato funkcionalita požadována, je řešena pokročilými souborovými systémy jako např. ADvFS (True64), GPFS(Linux, AIX), VMFS (VMWare), které jsou většinou na komerční bázi. Navzdory těmto problémům SAN napomáhá k lepší konsolidaci dat a také k jejich efektivnějšímu a rychlejšímu přístupu než je tomu například při použití technologie NAS.

2.3.8 Kompatibilita FC prvků

Jeden z dřívějších problémů s Fibre Channel SAN byl, že karty a další hardware od různých výrobců nebyl zcela kompatibilní. Ačkoli základní skladovací protokoly FCP byly vždy jako standardy, některé z vyšších funkcí nepracovaly korektně. Mnoho řešení byla na trh vydána ještě před dokončením standardů.

Kombinované úsilí členů Storage Networking Industry Association (SNIA) situaci zlepšilo. Ale i dnes nejdražší zařízení mohou vykazovat problémy s hardwarem různých firem.

2.3.9 SAN přes Fibre Channel (FC)

Původně jediný typ SAN sítě. Využívá se optických vedení **Fibre Optics** (v současnosti o rychlosti 10 a 16 Gbps) a protokolu **FCP**. Jsou potřeba speciální FC switche (pokud nemáme pouze spojení Point-to-Point) a servery musí mít **Fibre Channel Host Bus Adapter** (HBA), který nahrazuje SCSI řadič. Každý HBA má svoji unikátní adresu **World Wide Name - WWN** (obdoba MAC adresy u síťové karty), HBA může mít jeden nebo více portů, každý port má svou **port WWN** (WWNP) adresu a celý HBA má *node WWN* (WWNN) adresu. HBA zajišťuje komunikaci po síti a zapouzdřuje SCSI do FCP (a opačně).

2.3.10 SAN přes TCP/IP - iSCSI

SAN přes FC je značně drahá záležitost, takže se v době, kdy se začal používat **gigabitový ethernet** (v té době se používalo FC s rychlostí 2Gbps) objevil standardizovaný protokol **Internet Small Computer System Protocol** (iSCSI). Ten zapouzdřuje SCSI do TCP/IP protokolu, takže může fungovat na běžné ethernetové síti (podmínkou je minimální rychlost 1 Gbps). Teoreticky může fungovat na již existující síti (switchích a síťových kartách), ale v praxi se rozhodně doporučuje vybudovat oddělenou síť pro SAN (pro datový provoz není problém vytížit 1Gbps síť).

TCP protokol sice není pro datový přenos, tak dobrý jako **FCP protokol**, který byl speciálně pro tento účel vyvinut, ale je dostatečný. V dnešní době jsou SAN sítě založené na iSCSI značně rozšířené. Je tu ovšem otázka síťové karty. iSCSI se může provozovat přes klasickou síťovou kartu, ale pak veškeré zpracování (zapouzdřování SCSI do iSCSI) provádí procesor počítače. Druhá možnost je zakoupit speciální **iSCSI HBA**, který provádí tyto úkony a počítači již předává normální SCSI.