# Backup Milan Brchel

Obsah

[Backup 1](#_Toc445058151)

[Zálohování dat](#_Toc445058152) 2

[Archivace dat](#_Toc445058153) 2

[Operace s daty](#_Toc445058154) 3

[Problematika zálohování v podnikové sféře](#_Toc445058155) 4

[Historická média pro uložení dat](#_Toc445058156) 4

[Současná média pro uložení dat](#_Toc445058157) 5

[Ukládání dat](#_Toc445058158) 6

[Podle topologie](#_Toc445058159) 6

[Podle vzdálenosti](#_Toc445058160) 7

[Typy záloh](#_Toc445058161) 7

[Modely zálohování dat](#_Toc445058162) 9

[Grandfather-Father-Son](#_Toc445058163) 9

[Round-robin](#_Toc445058164) 9

[Hanojská věž](#_Toc445058165) 10

[Disaster Recovery](#_Toc445058166) 11

[Zdroje](#_Toc445058166) 11

## Zálohování dat

*„Zálohování je mechanismus, při kterém jsou vybraná data (nemusí to být tedy všechna) ukládána na jiné médium. V případě zničení původního média jsou data obnovena ze zálohy.“*

Záložní kopie dat se vytváří dle nastavené firemní politiky a pro každou organizaci tak existují jiná pravidla. Většina zálohovaných dat je především určena jen např. pro obnovení operačního systému, pro obnovu databáze nebo pro obnovu nedávno vytvořených dokumentů apod., není tedy určena pro soubory, které nemají nic společného s chodem firmy (filmy, osobní dokumenty atd.). Zálohování dat se provádí z důvodu jejich ochrany před ztrátou – za účelem jejich obnovy po neočekávaně nastalé havárii.

Pokud se médium, na kterém jsou data primárně uložena, jakýmkoliv způsobem poškodí nebo dokonce zničí (dojde k havárii), budou data na něm uložená nenávratně ztracena. Díky záloze na jiné médium lze tomuto negativnímu jevu předejít. Záloha dat se provádí na dobu několika dní, maximálně několika týdnů, aby byla zajištěna aktuálnost těchto dat. K zálohování je třeba mít na rozdíl od archivace výkonné záznamové zařízení, které musí mít rychlou odezvu a zajištěnu stálou konektivitu. Pro uložení dat na časový úsek, jako jsou měsíce nebo roky, se využívá archivace dat.

## Archivace dat

*„Archivace představuje trvalé odstranění dat ze systému, které již nemusí být dostupné on-line.“*

Smyslem archivace je uchování a zajištění dlouhodobé dostupnosti digitálních dat řádově od několika měsíců až po desítky let. Tato data se ukládají na nosiče s dlouhou životností a dobrým bezztrátovým uchováním informace. Takovýmito nosiči bývají nejčastěji magnetické pásky a vybrané pevné disky.

Archivace dat je znatelně levnější než zálohování, protože u zálohovacích médií jsou vývojáři nuceni držet krok s technologickým vývojem v oblasti rychlosti přístupu k datům a velikosti úložného prostoru, zatímco pro účely archivace postačí vlastnit starší a pomalejší úložná média.

# Operace s daty

S digitálními daty se mohou provádět různé úpravy, vedoucí k zamezení nebo v některých případech dokonce podpoření jejich ″kopírování / mazání″. Jednotlivé datové operace, které jsou nejčastěji prováděny, jsem popsal níže.

* **Duplikace dat** - jejím principem je vytvoření dvou záloh na dvou různých mediích a na různých místech. Zvyšuje se tím zabezpečení dat proti jejich ztrátě a rychlost při jejich obnovení po katastrofě. Duplikace ale také sebou nese nevýhodu v potřebě vlastnit mnoho úložného prostoru (je to drahé). Pro odstranění takovýchto nadbytečných dat se využívá metoda zvaná deduplikace.
* **„Deduplikace dat** zajišťuje v celém ukládaném datovém objemu vyhledávání a odstraňování duplicitních údajů. Díky tomu je možné uložit více datových souborů na menší úložný prostor.“ Deduplikace dat je metoda snížení požadavků na úložiště odstraněním redundantních dat, takže určitý datový obsah zabírá prostor na disku pouze jednou. Pokud jsou nalezena duplicitní data, nahradí je ukazatel na již uložený soubor údajů, a vlastní data jsou zahozena, neboli "de-duplikována". Deduplikace dat se provádí na úrovni bloků, nikoli na úrovni souborů. Deduplikace zmenšuje objem ukládaných dat až o 95%. Jedná se o poměrně novou technologii, která v budoucnu bude velmi potřeba, protože objem ukládaných dat ročně roste téměř exponenciálně.
* **Komprese dat** je taková operace s daty, při níž dochází za pomoci kompresního algoritmu ke zmenšení celkového objemu dat. Komprese dat je zvlášť výhodná při archivaci dat. Kompresi je možné rozdělit do dvou základních skupin na bezztrátovou a ztrátovou.
* **Replikace dat** – „Replikace je automatický proces kopírování a aktualizace dat na více počítačů v síti – obvykle do geograficky rozdílného území, pro zabezpečení dat před ztrátou při nastalé katastrofě v jedné budově.“ [5] Změny aplikované na jednu část jsou zachyceny a uloženy lokálně před posláním a aplikováním na každou vzdálenou lokaci. Replikace poskytuje uživateli rychlý lokální přístup ke sdíleným datům a chrání dostupnost dat z aplikací, protože existují alternativní možnosti přístupu k datům. Replikace je zaměňována také s pojmem duplikace – jedná se téměř o ten samý jev.
* **Redundance dat** je označení pro nadbytečný výskyt stejných hodnot v celém objemu dat. Výhodou redundance je, že může sloužit jako prostředek ke zvyšování spolehlivosti a odolnosti proti chybám, zároveň však její nevýhodou je právě ona nadbytečnost sama, která zvyšuje potřebu velké úložné kapacity.
* **Redundant Array of Inexpensive/Independent Disks (RAID)** – RAID se dá do češtiny přeložit jako ″vícenásobné diskové pole laciných/nezávislých disků″. Jak již bylo řečeno, je to metoda zabezpečení dat proti selhání pevného disku. Toto zabezpečení dat se provádí za pomoci ukládání dat na více na sobě nezávislých pevných disků, několika různými způsoby (základních je 7, ostatní způsoby jsou kombinací předchozích, uvedu pouze ty nejpoužívanější).
  + **RAID 0** – několik disků je pouze spojeno do jednoho logického celku (při poruše jednoho členu ztratíme data).
  + **RAID 1** – provádí zrcadlení (kopírování stejných) dat na dva nezávislé disky, při poruše jednoho o data nepřijdeme.
  + **RAID 5** – data jsou rozprostřena na více disků (je rychlejší čtení), na každém disku jsou uloženy samo opravné kódy, ze kterých se při havárii jednoho pole dokáží ztracená data dopočítat

# Problematika zálohování v podnikové sféře

V podnikovém prostředí je zálohování a archivace dat mnohem důležitější, než jak je tomu u domácích uživatelů. Jednotlivé organizace potřebují svá data mít aktuální a neustále k dispozici, ale zároveň je potřebují ochránit před ztrátou nebo odcizením.

Za kritická data můžeme označit taková data, která mají pro organizaci užitnou hodnotu (účetní informace, údaje o zákaznících, objednávkách, firemní dokumenty, pracovní výsledky, roční uzávěrky, emailová korespondence, data generovaná specifickými aplikacemi apod.).

# Historická média pro uložení dat

**Děrný štítek** – vyráběl se z tenkého kartonu, informace na něm byla reprezentována dírkou na určité pozici. Místa pro otvory byla uspořádána do matice. Nejběžněji používanou kapacitou bylo 80 znaků na jeden děrný štítek.

**Děrná páska** – vyráběla se z papíru a byla nejvíce využívána především v 60. letech 20. století, kdy sloužila jako primární uložiště dat. Byla však pomalá a měla velmi malou kapacitu pro uložení dat.

**Magnetická páska** – mohla uložit např. i 10 000 děrných štítků. Její výhoda byla hlavně ve škálovatelnosti, opakovaném použití a vyšších rychlostech záznamu a čtení dat. V současnosti se opět používá zejména pro archivaci dat.

**Disketa** – bylo to malé, magnetické datové médium, které se vyznačovalo velkým rozšířením u uživatelů, malou výrobní cenou, ale také se vyznačovalo malou kapacitou a nevelkou životností.

**Optický disk CD/DVD** – optická média si díky své poměrně velké kapacitě, ceně a všestrannému použití velmi rychle rozšířila a téměř na dalších 20 let se stala nejvyužívanějším zálohovacím médiem. Kapacita tohoto nosiče dat byla v minulosti poměrně dostačující, v současnosti jsou na ústupu.

# Současná média pro uložení dat

**USB Flash disk** – je to paměťové zařízení používané převážně jako náhrada za diskety, CD i DVD. Postupným vývojem se jeho kapacita zvětšila natolik, že může konkurovat dnešním HDD i SSD diskům (na trhu můžeme najít Flash disk s kapacitou 1TB+). Další výhodou jsou jeho malé rozměry a díky USB portu ho lze použít na velmi rozmanité škále zařízení.

**Optický Blu-ray disk** – je nástupcem dnes již kapacitně nedostačujícího CD a DVD. Nabízí mnohonásobně větší kapacitu (řádově desítky GB dat) při zachování stejných rozměrů a vlastností běžných CD a DVD. Stojí však za zmínku říci, že mechanika určená pro zápis a čtení Blu-ray disků není zatím tolik rozšířena, jako je tomu u mechaniky pro CD a DVD.

**Pevný disk (HDD)** – hlavním důvodem velkého rozšíření HDD se stal poměr velké kapacity disku a přívětivé pořizovací ceny. Je to první médium, na které se dají bezpečně zálohovat data. Jedinou jeho nevýhodou je jeho mechanická konstrukce – může dojít ke ztrátě dat. Nicméně se HDD stalo na velmi dlouhou dobu levným a populárním uložištěm pro různá data, které se hojně využívá i dnes.

**SSD disk** – tento typ diskového média na rozdíl od HDD neobsahuje pohyblivé mechanické části a má mnohem menší spotřebu energie. To jsou jeho hlavní přednosti. Další hlavní výhodou je vysoká rychlost čtení i zápisu. Pro uložení dat je použita nevolatilní flash paměť. SSD mají poměrně malou životnost – jsou limitovány maximálním počtem zápisů do stejného místa, což je jejich hlavní nevýhoda. SSD disky se začínají postupně používat i v datových centrech k odstranění tzv. úzkých. Velkou výhodou může být použití hybridního řešení v podobě využití SSD a HDD disků dohromady v jednom zařízení. Operační systém, často používané aplikace a databáze mohou být uloženy na velmi rychlém SDD disku, zatímco data, ke kterým není potřeba přistupovat tak často, mohou být uložena na běžném HDD. Menší potřeba zápisu na SDD disk šetří jeho opotřebení.

**Magnetická páska** – přestože se jedná o sekvenční médium (Data se na něj musí nahrávat s konstantním datovým tokem (nevyužívá se náhodného přístupu jako je tomu u disků), a proto je třeba mít i přídavné zařízení, které konstantní datový tok zajistí) využívané převážně v minulosti pro ukládání dat, dnes po řadě let vývoje patří páska neodmyslitelně k moderním archivačním médiím hlavně kvůli nízké ceně média a kvůli dobrému uchování informace po několik desítek let.

Výrazně vpřed se posunula nejen kapacita jednotlivých pásek, která může být v současnosti až 6,25 TB (specifikace LTO-6 v komprimované podobě), ale také se významně zrychlily zapisovací a čtecí vlastnosti tohoto média. V současnosti pásky pracují s technologií zvanou LTO (Linear Tape-Open). Tato technologie je stále ve vývoji a je rozdělena na několik generací, které jsou označeny čísly 1–8. Aktuálně jsou na trhu k dispozici pásky s označením LTO 6, které mohou pojmout až 6,25 TB dat a pyšní se rychlostí čtení dat až 400 MB/s. Během několika následujících let je ohlášen nástup generace s označením 7 a 8. LTO 8 umožní uložit na jediný magnetický pásek až 32 TB dat a využívat rychlost čtení až 1180 MB/s. Doba skladovatelnosti dat na pásce je uváděna až 30 let.

# Ukládání dat

Pokud jakákoliv organizace zvažuje, jaké výhody jim zálohování v budoucnu může přinést, je třeba si nejprve položit několik otázek, které se zálohováním úzce souvisí: Jak velká úložná kapacita bude potřeba? Bude možné v případě potřeby přidat další úložný prostor? Jaká bude dostupnost dat (postačí v budoucnu jen noční zálohování, nebo budou potřeba redundantní disková pole)? A v neposlední řadě nastane také otázka, jak moc drahé bude řešení pro ukládání dat? Níže popsané informace by měly pomoci při výběru správného řešení pro ukládání dat.

## Podle topologie

**DAS (Directly Attached Storage)** je základní architektura navržená pro maximální výkon. Diskové pole je zde externě napojeno přímo pomocí kabelu k serveru (PC stanici). Každý takovýto server vlastní pouze své disky - u tohoto návrhu nebylo počítáno se sdílením dalších disků. Serverové systémy s architekturou DAS jsou většinou postaveny na technologiích SCSI, SAS, SATA2 nebo Fiber Channel. DAS přistupuje k datům blokově. DAS je implementován pouze v síti LAN, kde lze využít switch pro připojení k více serverům.

**SAN (Storage Area Networks)** je dedikovaná datová síť (oddělená od LAN, WAN atd.) navržená pouze k datovým přenosům – slouží pro zpracování velkého množství dat (přistupuje k datům blokově). Architektura SAN vyžaduje infrastrukturu sestavenou z tzv. FC switchů (Fiber Chanel switch – optický switch), na rozdíl od DAS využívá SAN externí RAID řadiče a využívá optické kabely. SAN se proto hodí pro střední až velké společnosti, které požadují vysokou dostupnost svých služeb, rychlé odezvy a škálovatelnost. Pořizovací náklady SAN jsou poměrně vysoké, protože je nutné vybudovat specifickou infrastrukturu. Hlavním přínosem a zároveň výhodou SAN je možnost sdílet datová pole na více serverech (mohou to být i virtuální logické jednotky nebo páskové knihovny). Operační systém vidí zařízení v SAN jako lokálně připojené zařízení. Zatímco u DAS byla disková pole vázána přímo na konkrétní server (kabely mohly být dlouhé maximálně pár metrů), u SAN lze provést fyzické oddělení dat a serverů od sebe i na desítky kilometrů. SAN je využita v datových centrech.

**NAS (Network Attached Storage)** je datová síť postavená na využití síťového protokolu TCP/IP v místní síti LAN. Architektura NAS je určena k efektivnímu sdílení souborů v síti na úrovni souborového systému (může mít i jiné specializované funkce jako je webový server, klient P2P apod.) – je využívána speciálně pro zálohování dat. K NAS lze přistupovat jak v rámci sítě LAN, tak i prostřednictvím vzdáleného přístupu přes WAN. Pokud není používán server jakožto prostředník dat (tento server má za úkol sdílení a podporu různých protokolů), může být použit také jako levný a jednoduchý server pro loadbalancing (rozvažování zátěže sítě), což je jedna z jeho výhod. Další velkou výhodou NAS řešení je využití již stávající infrastruktury (TCP/IP síť) – s tím je spojena i jednoduchost jeho implementace a správy, ale také nízká cena celého řešení (v dnešní době proto je toto řešení hojně využíváno).

## Podle vzdálenosti

**Lokální** – pod tímto pojmem si můžeme představit taková uložiště, která používáme i v běžné domácnosti. Ukládat nebo rovnou zálohovat naše data můžeme na pevný disk našeho počítače, který je připojen pomocí SATA kabelu, na externí disk připojený pomocí USB kabelu nebo FireWare (nahrazuje klasické SCSI). Dnes lze také, co se týče paměti a rychlosti, využít velmi dobré, nové Flash disky podporující USB 3.0

**Vzdálené** – pod tímto pojmem si nejlépe můžeme představit již výše zmíněné uložiště NAS nebo DAS.

# Typy záloh

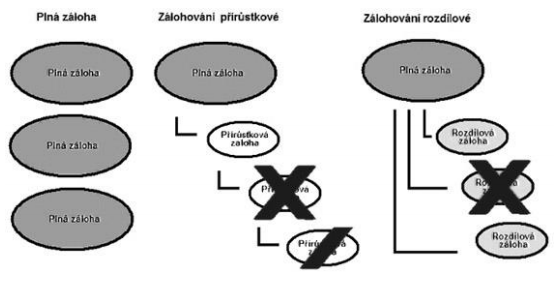
**Nestrukturovaná** - Nestrukturovaným úložištěm může být větší množství disket, [CD](https://cs.wikipedia.org/wiki/Kompaktn%C3%AD_disk), [DVD](https://cs.wikipedia.org/wiki/DVD) medií s minimem informací o záloze. Tento způsob je nejjednodušší, ale není příliš oblíben u větších firem.

**Plná záloha** – je výchozím bodem pro všechny ostatní zálohy a obsahuje všechna data (soubory a složky), jež jsou vybrána pro zálohování. Plná záloha by byla ideálním typem pro všechna zálohování, protože je nejvíce komplexní a soběstačná. Nicméně takovéto zálohování zabere mnoho času, proto se příliš nevyužívá. Úplná záloha je často omezena na týdenní nebo měsíční periodu a je spuštěna převážně jen přes noc. Jedině plná záloha nám ale poskytuje možnost zcela a rychle obnovit všechny zálohované soubory, protože obsahuje kompletní data. Plná záloha ovšem sebou nese řadu nevýhod, např. zálohování je pomalejší v porovnání s jinými druhy záloh, požadavky na úložný prostor jsou ze všech typů u plné zálohy nejvyšší a také je tu bezpečnostní riziko (odcizení disku).

**Inkrementální záloha (přírůstková)** – je to záloha, kde se vytvoří úplná záloha pouze jednou, a to hned na začátku, poté další záloha uloží změny oproti ní (bude uložen pouze přírůstek takových dat, která se jakýmkoliv způsobem změnila oproti původní úplné záloze). V pořadí třetí přírůstková záloha opět uloží pouze ty změny, které nastaly oproti druhé záloze. Přírůstkové zálohování poskytuje rychlejší způsob zálohování dat, než opakování plných záloh (čas potřebný k obnovení zálohy může být zlomkový oproti času potřebného k provedení obnovy z úplné zálohy). Další výhodou je úspora úložného prostoru, potřebného pro tento typ zálohování. Ovšem velkou nevýhodou je, že pokud dojde ke ztrátě nebo poškození jakékoliv části dat i v jednom z přírůstků, nelze již obnovit data ani z dalších přírůstků, které za tímto poškozeným následují. Proto je velmi žádoucí provádět po určitém čase znovu plnou zálohu.

**Diferenciální záloha (rozdílová)** – je to taková záloha, která obsahuje všechny soubory, které se změnily od poslední úplné zálohy. Výhodou diferenciální zálohy je, že se zkrátí čas obnovení v porovnání s plnou zálohou.

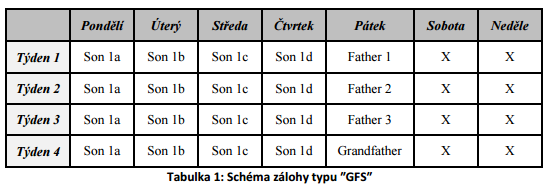
*Rozdíl mezi přírůstkovou a rozdílovou zálohou* je takový, že zatímco přírůstková záloha ukládá pouze změny z hlavní zálohy a poté už jen změny z přírůstků, rozdílová záloha ukládá všechny soubory, které se změnily od poslední úplné zálohy. Pokud dojde k poškození některé z diferenciálních záloh, nemá to vliv na žádnou jinou diferenciální zálohu (zálohy na sobě nejsou závislé). Obnovení rozdílovou zálohou je rychlejší než obnovení přírůstkovou, protože jsou potřeba pouze dva soubory – poslední plná záloha a poslední rozdíl. Další výhodou je i rychlejší proces zálohování, než je tomu u plné zálohy. Nevýhodami tohoto typu zálohování jsou: pomalejší obnovení, než jak je tomu u obnovy z plné zálohy, samotný proces zálohování je pomalejší, než je tomu u přírůstkové zálohy, a také jsou zde vyšší požadavky na úložný prostor, než je tomu u přírůstkové zálohy.



# Modely zálohování dat

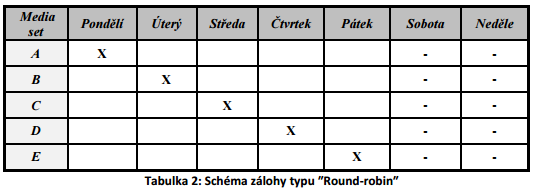
## Grandfather-Father-Son

Typ Grandfather-Father-Son (GFS) – tento typ rotace je určený především pro zálohování na pásku, ale funguje stejně dobře i u ostatních typů zálohování. V GFS se definují tři sady záloh – denní, týdenní a měsíční. Denní záloha je prováděna v určité (přednastavené) dny, např. od pondělí do čtvrtka – každé samostatné denní záloze se říká, že je to tzv. "media set" (každý se provádí na samostatné pásky nebo disky) a ve schématu je označen jako Son (syn). Na těchto media setech probíhají inkrementální zálohy, které jsou každý týden pravidelně přepisovány. Po každém týdnu se dělá nová skupina záloh označená jako Father (otec). Záloha typu Father je plná záloha. Z každé takovéto zálohy se další týden provádí opět denní záloha typu Son (inkrementální). Na konci každého měsíce se provádí znovu plná záloha označovaná jako Grandfather (dědeček). Tento typ pásky se na rozdíl od ostatních z knihovny vyjme a uloží se na jiné místo jakožto bezpečnostní záloha. Na místo této pásky se dá páska nová, která se opět po dalším měsíci vyjme a uloží na jiné bezpečné místo. Pásky typu Son a Father jsou v průběhu měsíce neustále cyklicky přepisovány. Tyto pásky je doporučeno vyměnit každé dva roky, aby byla zajištěna jejich bezchybná funkčnost.



## Round-robin

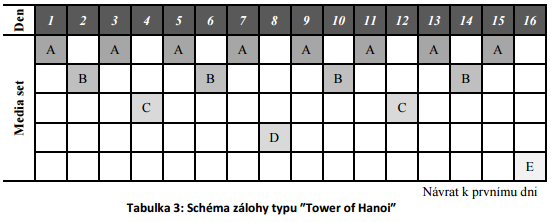
Typ Round-robin (schéma s jednou páskou pro každý den) – je to jeden z nejstarších typů pro rotaci záloh. Jedno z nejjednodušších schémat může být takové, že pro každý pracovní den v týdnu, kdy provádíme úplnou denní zálohu dat, máme vyhrazenu jednu samostatnou pásku. Rotace dle tohoto schématu nám umožňuje obnovu dat maximálně o jeden týden zpět. Po tomto jednom týdnu se pásky začnou opět znovu jednotlivě přepisovat v pořadí od pondělí do pátku. Schéma Round-robin je vhodné především pro malé podniky, které mohou využít interní nebo externí páskovou mechaniku. Je vhodné také pro organizace, kterým postačuje maximálně týdenní záloha dat. Také je možné použít někdy i dnes levnější zařízení NAS.



## Hanojská věž

Typ Hanojská věž (Tower of Hanoi) – tento typ rotující zálohy nese stejný název jako jméno jedné logické hry. Nachází se zde pět media setů – media set A (použit pro každý druhý den), media set B (použit pro každý čtvrtý den), media set C (použit pro každý osmý den) a media sety D a E (použity střídavě každý šestnáctý den).

Zálohování začíná s media setem A, a pak se s ním pokračuje každý druhý den. Další záloha probíhá na media set B – na ten se začínají zálohovat data jen ty dny, na které nepřipadá zálohování media setu A (každý čtvrtý den). Záloha na media set C začíná v den, kdy se neprovádí záloha na media set A, a ani na media set B (opakuje se každou osmou zálohu). Pro media sety D a E je zálohování nastaveno tak, že se neprovádí v den, kdy probíhá záloha A, B nebo C a opakuje se každou šestnáctou zálohu. Schéma je poměrně náročné na manuální správu. Je doporučeno použít zálohovací software přímo k tomu učený. Tento typ rotující zálohy umožňuje periodické vyjmutí media setu za účelem archivace.



# Disaster Recovery

Disaster Recovery (běžně označováno jako D/R) se do českého jazyka dá přeložit jako ″Obnova po katastrofě″ nebo ″Obnova po havárii″ – jsou to předem připravené zásady a postupy, které vedou k ochraně dat před výpadkem a zároveň slouží jako manuál pro případné obnovení IT infrastruktury v důsledku nastalé katastrofy/havárie. Katastrofou/havárií rozumíme buď nastalou živelnou pohromu nebo lidskou chybu, ať už úmyslnou či neúmyslnou, která zapříčiní nefunkčnost systému. Můžeme si tak představit např. poškození hardwaru, pád operačního systému, zastavení chodu kritické firemní aplikace, problémy s hardwarem – tyto a mnoho dalších příčin mohou způsobit nemalé problémy v jakékoliv společnosti. Pokud se těmto 30 problémům nepředchází a nesnižují se rizika přerušení fungování vnitřních procesů, je velmi pravděpodobné, že jednoho dne tento katastrofický scénář nastane.

# Zdroje

<http://www.diplomovaprace.cz/2013/9/Bakalarska_prace_Pavel_Junek.pdf>

<https://cs.wikipedia.org/wiki/Zálohování_dat>