

Zálohování dat

Úvod

Jedna z jednodušších otázek, na kterou stačí selský rozum a aspoň si to jedno přechíst. Ještě že tam tyhle otázky jsou :c. Kdybych neměl zálohu tak bych si všechna tahle scripta musel dělat znova :x. Naštěstí jsem měl a jen se mi nepodařilo obnovit jednu otázku, kterou jsem vypracoval ještě jednou.

Definice

Ukládání dat z provozních médií na média zálohovací. Záložní data jsou vyvolána, pokud jsou aktuální data poškozeny nebo ztraceny. Může probíhat pravidelně i nepravidelně.

Důvod

Hlavním důvodem, proč vůbec zálohovat data je abychom předešli jejich ztrátě. Nejčastěji můžeme přijít o data důsledkem selháním hardwaru, lidským faktorem – například neodborným zásahem do zařízení, krádeží zařízení nebo jiným manuálním poškozením – například přírodními vlivy.

Druhým, pro normálního uživatele zvláštním důvodem, proč zálohovat je kvůli verzování. Je to proces, při kterém zálohujeme určité změny např. v textovém dokumentu a vidíme, kdy a kde byly změny provedeny a kdo je udělal. Toto je velmi žádané ve velkých IT firmách, kde si takto lidi mohou porovnávat zdrojový kód určitého programu. Tyto změny lze jednoduše zpětně vrátet. Nejznámější program toto umožňující je Git.

Způsoby zálohování dat

Úplná záloha

Každá záloha obsahuje veškerá uživatelská data tzn. že nezalohujeme aplikace a programy, které lze znovu nainstalovat. Nepotřebujeme dalšího k obnovení. Nutnost buď překopírovat celou složku anebo vytvořit jinam. Není to elegantní řešení.

Tento typ zálohy je nejrozměrnější a je 1:1.

Úplná záloha + inkrementální záloha

Obsahuje pouze změny, které proběhly od poslední inkrementální zálohy. Podobná verzování. Je nutno ponechat celý řetězec zálohovaných změn a úplnou zálohu. Pokud v řetězci chybí jediná záloha, nelze obnovit ty, které následovaly. Kladnou vlastností je, že zabírají méně místa než úplná záloha a že lze postupně oddělovat změny, pokud soubor byl již zálohován špatně.

Úplná záloha + rozdílová záloha

Diferenciální zálohy nezabírají příliš místa. Diferenciální záloha zaznamenává změny, které proběhly od plné zálohy. Pro obnovu je potřeba mít příslušnou Diferenciální a plnou zálohu. Diferenciální zálohy nejsou na sobě závislé. Pokud dojde k poškození některé z diferenciálních záloh, nemá to vliv na žádnou jinou diferenciální zálohu.

Obraz disku

Jedná se o bitovou kopii disku, obsahuje opravdu všechna data. Což znamená že zkopíruje i neviditelné soubory i prázdná místa na disku atd.

Zásady zálohování dat

Pravidlo 3-2-1 (Petr Krogh)

3. Alespoň tři kopie dat
2. Alespoň dva typy médií
1. Jedna záloha mimo své pracoviště

Pravidelné zálohování – mít systém, kdy byla poslední záloha a kdy bude další

Přehlednost záloh – organizaci složek a dat

Zálohujeme pouze důležitá data – kvůli snižování místa

Kam zálohovat

Data můžeme zálohovat mnoho způsoby a na mnohá zařízení. Většinu z nich denně využíváme.

Pevný disk

Využívají se většinou externí disky. Zálohovat je možná manuálně, nebo lze případně využít některý z programů pro zálohování - například WD Backup. Pevné disky mají velmi dlouhou životnost a jsou dostupné za nízké ceny. Pevné disky se používají také pro RAID (viz RAID). Náchylné na otřesy – pohyb.

Optický disk

Nejčastěji se využívají CD-R, DVD-RW nebo DVR-RAM. Nevýhoda CD je jeho velikost, DVD je vhodnější, ale stále má malou kapacitu.

Největší nevýhodou optických disků je jejich životnost. Po několika desítkách let se médium stane nečitelným. Jedná se o nejlevnější variantu záloh.

Flash disk

Flash disky se využívají hlavně pro rychlý přenos dat a nejsou moc vhodné pro jejich zálohu.

Nevýhodou je náchylnost na poškození a malá velikost, díky které lze Flash disk lehce ztratit. Jsou ale odolné (stejně jako SSH) proti prachu a otřesům. Výhodou Flash disků je snadná dostupnost a nízká cena.

SSD

Na SSD disky se nezalohuje často z důvodu vysoké ceny. Jsou ovšem méně náchylné na zničení a více rychlé. Takže pokud máte hodně peněz a potřebujete přenášet zálohy a rychle je rozkopírovat tak je to správné médium pro vás.

Domácí server

Sever, který je vyroben z „domácích“ komponent. Je levnější než NAS a může být i stejně účinný.

Může mít ale i nižší výkon a životnost než NAS. Je zde specifický operační systém pro jeho provoz. (FreeNAS, OpenMediaVault, unRAID) Nutná administrace.

Může stačit i jednoduché Raspberry Pi. Pro tyto účely je i možné koupit specifické komponenty přímo dělané na tyto servery např speciální procesory, RAM, či 24/7 HDD, ale nejsou potřeba.

NAS Network Attached Storage

Jedná se o zařízení, které spojuje jeden nebo více pevných disků (SSD) a připojuje je do lokální sítě, kde jsou dostupné více lidem najednou. Na rozdíl od externích disků je konstruován pro provoz 24hodin denně (Nutnost pořídit specifické). Má za úkol sdílet data pomocí dostupných protokolů.

Může mít i vlastní webový server. Nevýhodou je vyšší pořizovací cena + cena disků. Podporuje i řešení RAID.

Cloudové úložiště

Naše soubory máme neustále k dispozici, ale zároveň jsou stále v bezpečí na serveru. Vždy hrozí, že se k našim datům dostane někdo další. Většina online úložišť nabízí prostor zdarma do určité kapacity. Nejčastěji se setkáme s kapacitou kolem 15 GB. Využívá se převážně pro zálohování mobilních zařízení, například pro zálohu fotek a videí z mobilních telefonů.

Nejznámějšími poskytovateli jsou Google – Google Disk, Microsoft – OneDrive, Apple – iCloud, Mega - Mega Cloud.

Verzované úložiště

Tato úložiště jsou většinou jako cloudová úložiště k dispozici na internetu za poplatek. Uchovává ale veškeré informace o dění s daty a využívá speciální programy pro jejich kontrolu, testování a sledování změn. Využívá se hlavně pro zdrojové kódy a IT projekty, protože tam se nejlépe uplatní všechny tyto funkce. Nejběžnějším zástupcem je GitHub, který je hojně používán mezi velkými korporacemi pro jeho jednoduchost a 100% funkčnost. Používá Git program, který spravuje všechny změny a zaznamenává je. Druhým populárním úložištěm je GitLab, který také používá Git program, avšak je zadarmo a má lépe nastavitelné funkce. Obě služby poskytují srovnatelný balíček služeb.

Umožňuje, aby na stejném projektu pracovalo více lidí současně a zároveň zajišťuje, že změny provedené kterýmkoli vývojářem jsou konzistentní a distribuovány ostatním. Pro větší projekty, nebo projekty, na nichž se podílí více lidí (často na geograficky vzdálených místech), je použití systému pro správu verzí prakticky nezbytností.

CVS

Jeden z nejstarších verzovacích systémů (1980s). CVS se typicky používá jako víceuživatelská klient/server aplikace. Samotný program CVS je dodáván v základní verzi pro použití v příkazovém řádku, existují však i nadstavby pro použití v grafickém režimu. Nemá podporu a není tak spolehlivý.

SVN

Náhrada za CVS. – snaží se vylepšit nedostatky (například nemožnost přesunu nebo kopírování adresářů). Velmi dobrá dokumentace. Hodně pluginů a možností přizpůsobení. Pořád ale obsahuje hodně chyb a je pomalejší.

Mercurial

Multiplatformní verzovací systém. Mercurial je určen primárně pro použití v příkazovém řádku, dostupné jsou ale i grafická uživatelská rozhraní. Lepší dokumentace. Vysoká výkonnost a škálovatelnost.

RAID

Je to metoda využívání více než jednoho disku pro zabezpečení nebo zrychlení počítače. Budeme se zde věnovat jen těm nejdůležitějším, protože RAID polí je opravdu mnoho.

RAID 0

Není to skutečný RAID, protože nemáme žádné nadbytečné informace.

JBOD *Just a bunch of disks*

Data se ukládají postupně na několik disků za sebou a při selhání jednoho disku mohou být některé soubory zachovány.

Prokládání

Disky musí mít stejné parametry. Je potřeba 2 a více disků. Dojde ke zrychlení čtení a zápisu. Při výpadku jednoho z disků přijdeme o data z ostatních disků. Nejčastější využití má při ukládání dočasných velkých souborů - například při střihu videa.

RAID 1

První RAID. Disky musí mít stejné parametry. Funguje na principu zrcadlení - data se ukládají současně právě na dva disky. Nedojde ke zvýšení výkonu, ale v případě výpadku jednoho z disků máme jeho kopii. Bohužel není možné využít obou kapacit pro nová data – technicky poloviční kapacita, protože se druhý disk plní nadbytečnými daty.

RAID 5

Disky musí mít stejné parametry. Je zapotřebí minimálně třech disků. Data se prokládají mezi dva disky, třetí disk obsahuje výsledek výpočtu XOR z dat těchto dvou disků, a tak poskytuje samo opravné kódy pro 2 disky. Rychlejší čtení, ale pomalejší zápis. Pokud jeden ze 2 disků selže, lze opravit pomocí samo opravných kódů.

RAID 6

Disky musí mít stejné parametry. Zvyšuje výkon a zaručuje dvojnásobnou ochranu. Zapotřebí jsou minimálně čtyři disky, z čehož dva vždy obsahují paritu. Můžeme přijít o dva z disků, aniž bychom ztratili uložená data. Má pouze hardwarové řadiče. Rychlost jeho zápisu je nižší než rychlost čtení disků.

RAID 01

Disky musí mít stejné parametry. Využívá minimálně 4 disky. Funguje na principu RAID 0, který dále zrcadlíme metodou RAID 1. Jelikož se veškeré soubory ukládají dvakrát, kapacita disků je o polovinu menší.

RAID 10 Stripování

Disky musí mít stejné parametry. Funguje podobně jako RAID 01, ale 2 disky nejdříve zrcadlíme pomocí RAID 1 a až poté se použije RAID 0. Opět nám zůstane 50 % z celkové kapacity.

Zdroje

1. https://cs.wikipedia.org/wiki/Zálohování_dat
2. <https://www.jaknainternet.cz/page/1180/zalohovani/>
3. https://en.wikipedia.org/wiki/Mainframe_computer#Differences_from_supercomputers
4. https://cs.wikipedia.org/wiki/Magnetick%C3%A1_p%C3%A1ska
5. <https://cs.wikipedia.org/wiki/Mainframe>
6. <https://cs.wikipedia.org/wiki/Verzov%C3%A1n%C3%AD>
7. https://cs.wikipedia.org/wiki/Z%C3%A1lohov%C3%A1n%C3%AD_dat
8. <https://www.acronis.cz/kb/diferencialni-zaloha/>
9. <https://en.wikipedia.org/wiki/RAID>
10. <https://cs.wikipedia.org/wiki/RAID>