

Zálohování pro začátečníky

Ondřej Caletka

o.caletka@sh.cvut.cz

<http://shell.sh.cvut.cz/~oskar>



<http://sut.sh.cvut.cz>

Osnova

- Definice pojmu
- Požadavky na zálohování
- Co zálohovat
- Kam zálohovat
 - vsuvka o šifrování
- Zálohovací produkty
- Záloha databází



Co je zálohování

- Ochrana před ztrátou dat v důsledku
 - selhání hardwaru
 - výrobní vada
 - konec životnosti
 - vnější vlivy (bouřka, přepětí,...)
 - **živelní pohromy**
 - **chyby obsluhy**
 - **odcizení HW**
 - **zapomenutí šifrovacího klíče**

Co není zálohování

- RAID
 - chrání pouze před selháním HW
 - nechrání před živly
 - nechrání před chybou obsluhy
- High Availability systémy
 - zvyšují dostupnost
 - chyba obsluhy okamžitě poškodí všechny nody

Požadavky na zálohování

- Musí být jednoduché.
 - Musí se dělat pravidelně.
 - Mělo by uchovávat historické kopie pro případ, že omyl neobjevíme včas
 - Mělo by zacházet hospodárně s místem na disku
 - Obnovení zálohy by mělo být snadné a rychlé.
 - V enterprise prostředí se obnovení pravidelně zkouší a jsou zpracovány postupy.
 - V domácím prostředí se obvykle spokojíme s ujištěním, že data „tam někde jsou.“
-
-

Co zálohovat?

Dva základní přístupy:

- Zálohujeme jen potřebné
 - Nezálohujeme např. operační systém
 - Obnovení zálohy je příležitost updatovat :)
 - Obnovení bude nějakou dobu trvat
 - Při obnovení nejspíše něco málo ztratíme
 - Zálohujeme toho co nejvíc
 - Nezálohujeme jen /tmp, /var/run, /proc, atd.
 - Down-time je pro nás kritický
 - Při incidentu nemáme čas reinstalovat OS
 - Při obnovení ztratíme jen práci od poslední zálohy
-
-

Kam zálohovat?

- Stejný disk
 - nouzovka, nechrání proti odcizení
 - Externí USB/eSATA disk
 - Ideální pro notebooky
 - Nenosit stále s sebou
 - Rychlost USB je omezující
 - Síťový disk/páska
 - Ideální pro desktopy/servery
 - Šifrovat, nebo nešifrovat?
 - Rozhodně šifrovat na cestě po síti.
 - Šifrovat na disku, který nosíme s sebou.
 - Nešifrovat na disku v trezoru
-
-

Vsuvka o šifrování

1. Šifrovaný overlay pomocí EncFS

- připojí šifrovaný adresář
- On-the-fly mapuje 1 šifrovaný soubor na jeden dešifrovaný
- Nevznikají úniky volného místa

2. LUKS kontejner

- Je nutno předem dobře odhadnout velikost

1. Vytvoříme (řídový) soubor obrazu:

```
dd if=/dev/zero of=bup.img bs=1 seek=100G count=0
```

2. Napamujeme na loopback

```
losetup /dev/loop0 bup.img
```

3. Naformátujeme LUKSem a namapujeme

```
cryptsetup luksFormat /dev/loop0
```

```
cryptsetup luksOpen /dev/loop0 cbackup
```

4. Vytvoříme souborový systém, atd... (na /dev/mapper/cbackup)

Zálohovací produkty

- BACULA
 - HP Data Protector
 - DropBox, Unison - ve skutečnosti nejde o zálohování, je ale vedlejším efektem
 - Windows Zálohování
 - Apple Time-Machine, FlyBack (OSS klon)
 - cp -a
 - tar
 - rsync
 - **rdiff-backup**
-
-

cp -a, tar

- Zálohování je v podstatě kopírování
- `cp -a` zachová atributy
- `tar` (Tape ARchiver) vytvoří jeden velký soubor
 - Bude složitější obnovit jen část
 - Pozor na kompresi, ne `gzip`!
 - 1 špatný bit zničí celý soubor
 - Umí i inkrementální zálohy – vhodné na pásky

rsync

- Názvem připomíná rcp, rsh, atd...
 - Ve skutečnosti je to inteligentní, síťově transparentní kopírovač.
 - Přenáší jen změny, jedním směrem
 - Jak ale pozná změny?
 - Podle data poslední modifikace
 - Spočítá částí souborů hashe na obou stranách
 - => Těžce snáší velké soubory
 - Historické kopie pomocí hardlinků
 - Google: rsync hardlink
http://www.mikerubel.org/computers/rsync_snapshots
-
-

rdiff-backup

- Síťově transparentní inkrementální záloha
- Používá rsync-like přístup
- Použití velmi jednoduché, např.:

```
rdiff-backup --remove-older-than 10B ${CIL}
rdiff-backup --include-globbing-filelist ${list} \
-v5 / ${CIL}
```

- První příkaz smaže desátou historickou verzi
 - Druhý příkaz zazálohuje vše kromě vyjmenovaných souborů
 - Cílový adresář nemusí nutně být na linux filesystemu, ale je to lepší.
-
-

rdiff-backup - obnovení

- Cílový adresář obsahuje plnou verzi aktuální zálohy, můžeme ji přímo použít
- Chceme-li historickou verzi, použijeme přepínač `-r <datum> ...`
- ...nebo si najdeme příslušný diff v adresáři `rdiff-backup-data`

<http://www.root.cz/clanky/zaloha-dat-pomoci-rdiff-backup/>

Pozn: Pro síťové spojení `rdiff-backup` používá `ssh`, ale na obou koncích tunelu musí existovat binárka `rdiff-backup`

Zálohování databází

- Dva druhy záloh: dump × záloha dbfile
 - Dbfile nikdy nezalohujeme při běžící databázi!
 - existuje skript pro použití LVM snapshotu
 - Google: mysql backup lvm snapshot
 - Dbfile se nehodí pro archivační účely.
 - Dbfile umožní velice rychlou obnovu.
 - máme-li navíc (binární) log aktivních transakcí, můžeme dojet až těsně k místu destrukce
 - Dump se hodí pro archiv.
 - dá se obnovit v podstatě kdekoli, kdykoli
 - v případě MySQL je to textový SQL soubor
-
-

Maminčiny exploity



© Randall Munroe. Překlad: Tomáš Znamenáček písmo: Martin Stiborský

Závěr

- Díky za pozornost!
 - Zálohujte, zálohujte, zálohujte!
 - Co nemáš zálohované, jak bys ani nikdy neměl, takže tě nemusí mrzet, když o to přijdeš.
 - Prostor pro dotazy.
 - Příští týden CouchDB.
 - EOF □
-
-