



## SNAPSHOTY

cesta jak posunout zálohování a obnovu  
dále

**Michal Strnad, Jaroslav Kalus**

CESNET, z. s. p. o.

---

3. 3. 2019

- Datová úložiště a motivace pro workshop
- Popis prostředí pro hands-on
- Teorie a laborky

- Zajišťuje provoz a rozvoj národní infrastruktury pro ukládání dat pro výzkum a vývoj
- Aktuálně provozujeme čtyři úložiště založené na HSM
- Další HSM a nové servery pro object storage se právě instalují
- HSM úložiště jsou přístupná přes NFSv4, FTP, rsync, SCP, Globus ...
- Objektové úložiště pak přes S3/Swift, CephFS, RBD
- Další služby jako FileSender a ownCloud

- Čím dál větší množství dat
  - Riziko v podobě RAID technologie
  - Kontrolér v diskovém poli
  - Pomalý recall z pásek
  - Režie a pomalost souborových protokolů
  - Co když se během přenosu změní část dat?
- Tyto problémy vyřešíme kombinací technologií  
Ceph RBD a Btrfs (preview)

- Virtuál s CentOS 7
- Přihlašovací jméno, heslo a IP adresu máte na papíru
- Privilegovaný přístup na stroj přes sudo
- Přístup na úložiště CESNET (ostravské du4) přes servisní účet
- V případě problémů s jednotlivými kroky Vám pomůžeme
- Slidy, návody a další materiály  
<https://du.cesnet.cz/cs/workshop>
- Raději manuálové stránky než copy-and-paste ze slidů

- Jednorázová záloha či replika
  - Nástroj rsync
- Potřebujeme mít snapshoty celé datové oblasti nebo systému
  - Využijeme Btrfs snapshoty a send/recv
  - Jako cíl použijeme vzdálené RBD z Ceph clusteru
  - Primárně pro správce

## Jednorázová záloha či replika

- Prakticky na všech distribucích je možné rsync binárku nainstalovat z repa
- Pod sebou má SSH
- Delta přenos
- Používá ho velká část jiných aplikací (např. rsnapshot)
- Možnost skriptování
- Existuje grafická nadstavba Grsync

## Plný návod

<https://du.cesnet.cz/cs/navody/rsync/start>

- Základní použití nástroje
  - `rsync -av --progress ./directory`  
`labX@ssh.du4.cesnet.cz:~/VO_du_test-disk_only/`
- Umožňuje zachovávat práva a informace o vlastníkovi souboru (UID a GID) při přenosu na vzdálené úložiště v rozšířených attributech souborů
  - `rsync -av --numeric-ids`  
`--rsync-path="rsync --fake-super" ./directory`  
`labX@ssh.du4.cesnet.cz:~/VO_du_test-disk_only/`



- Pozor na lomenu na konci zdrojové cesty
  - `rsync -av ./directory/  
labX@ssh.du4.cesnet.cz:~/VO_du_test-disk_only/`
- Zvláštní obezřetnost je na místě při použití přepínače `--delete`
  - `rsync -av --delete ./directory  
labX@ssh.du4.cesnet.cz:~/VO_du_test-disk_only/`

Struktura pro připojení labX@ssh.du4.cesnet.cz

```
/home/labX:  
|-- .bash_history  
|-- .ssh  
|  \-- authorized_keys  
|-- VO_du_test-disk_only  
|-- VO_du_test-tape_tape  
\-- VO_du_test-tape_tape-shared
```

Budeme operovat jen v politice VO\_du\_test-disk\_only

1. Z webu <https://du.cesnet.cz/cs/workshop> stáhněte testovací data do VM a rozbalte archiv
2. Z VM se připojte za pomoci SSH na `ssh.du4.cesnet.cz`
3. Vytvořte adresář backup v politice `VO_du_test-disk_only`
4. Nahrajte do vytvořeného adresáře backup přes rsync rozbalený archiv
5. Upravte dle svého uvážení soubor `"include/keys/user-type.h"` na svém VM
6. Smažte soubor `"include/memory/jedec_ddr.h"` na svém VM
7. Zavolejte rsync tak, aby provedl změny i na straně serveru (CESNET úložiště)

## Řešení:

1. `yum install rsync`
2. `wget`  
`https://du.cesnet.cz/\_media/cs/kernel.tar.gz`
3. `tar -xf kernel.tar.gz`
4. `ssh labX@ssh.du4.cesnet.cz "mkdir`  
`./VO_du_test-disk_only/backup"`
5. `rsync -av --progress ~/include`  
`labX@ssh.du4.cesnet.cz:~/VO_du_test-disk_only/backup/`

Řešení:

1. `vim ~/include/keys/user-type.h`
2. `rm ~/include/memory/jedec_ddr.h`
3. `rsync -av --progress --delete ~/include  
labX@ssh.du4.cesnet.cz:~/VO_du_test-disk_only/backup/`

- Na velký počet souborů a velká disková pole již tradiční způsoby nestačí
- Řešením jsou snapshoty celé datové oblasti nebo systému
- Využijeme Btrfs snapshoty a send/recv
- Jako cíl použijeme vzdálené RBD z Ceph clusteru
- Instatní obnova
- Preview - ukázka nového přístupu k zálohování

- Copy-on-write file-systém
- Subvolumes (oddíly), snapshoty
- Checksumy dat i metadat
- RAID 0, 1, (pozor na 5 a 6), 10
- Defragmentace
- Scrubing
- Kompresce
- Přidání/odebrání disku

- send/recvie stabilní od kernelu 3.6
- btrfs-send - stream instrukcí, diff mezi dvěma subvolumes
- Full a inkremental mod
- Stream instrukcí umí zpracovat btrfs-recvie (user-space)



- Vylistování subvolumů a snapshotů
  - `btrfs subvolume show /`
- Zjištění výchozího subvolume
  - `btrfs subvolume get-default /`
- Zjištění obsazeného místa, ale ...
  - `btrfs filesystem df /`

- Vytvoření snapshotu
  - `btrfs subvolume snapshot [-r] <source> <dest>`
- Smazání snapshotu
  - `btrfs subvolume delete /path`
- Připojení / subvolume do cesty
  - `sudo mount /dev/sdaX /mnt/btrfs_test -o rw,relatime,space_cache,subvolid=5`

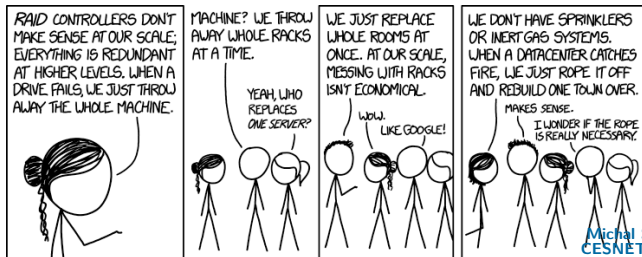
- Copy-on-write na souboru
  - `cp --reflink=auto`
- Snapshot adresáře
  - `cp -ax --reflink==always src/. dest`
- Automatické vytváření snapshotů před instalací balíků
  - `yum-plugin-fs-snapshot.noarch`

1. Namontujte si / subvolume do cesty /mnt/btrfs
2. Vytvořte subvolume /mnt/btrfs/@snapshots
3. Vytvořte pro /home snapshot připojený do cesty /mnt/btrfs/@snapshots
4. Smažte soubor "`~/include/xen/xen.h`" a následně ho zkuste obnovit ze snapshotu
5. Po úspěšné obnově smažte snapshot

## Řešení:

1. `sudo mkdir /mnt/btrfs`
2. `sudo mount /dev/vda3 /mnt/btrfs -o  
rw,relatime,space_cache,subvolid=5`
3. `sudo btrfs subvolume create  
/mnt/btrfs/@snapshots`
4. `sudo btrfs subvolume snapshot /home  
/mnt/btrfs/@snapshots`
5. `sudo rm ~/include/xen/xen.h`
6. `sudo cp  
/mnt/btrfs/@snapshots/home/labX/include/xen/xen.h  
~/include/xen/xen.h`
7. `sudo btrfs subvolume delete  
/mnt/btrfs/@snapshots/home/`

- Objektově orientované uložení
- Cluster sám udržuje minimální počet nastavených replik
- Protokoly S3/Swift, CephFS a RBD
- OSD, MON ...
- Často jako storage backend pro OpenStack



- Aktuálně pilotní provoz, testy s OpenStack
- Probíhá instalace nových serverů do Jihlav (HSM a Ceph)

- Ceph RADOS Block Device (RBD)
- Striping a replikace napříč clusterem
- Read-only snapshoty, revertování snapshotů
- Možnost připojit do Linuxu nebo QEMU KVM klientů
- RBD mirroring
- 10 Gbps jeden klient proti šesti serverům



- Nyní spojíme zmíněné technologie (RBD a Btrfs)
- Dostaneme možnost vytvořit zálohu subvolume/snapshotu (send/recv)
- Cílem však bude vzdálené blokové zařízení
- Zápis a případné čtení (obnova) dosahují rychlostí dostupné linky
- Provést to můžeme ručně nebo za pomoci skriptů (btrbk, snapper)

- Nainstalujeme základní Ceph balík
  - `sudo yum install ceph-common`
- Stáhněte `~/ceph.conf` z [https://du.cesnet.cz/\\_media/cs/ceph.conf](https://du.cesnet.cz/_media/cs/ceph.conf) do `/etc/ceph/`
- Zkopírujte `~/ceph.client.labX.keyring` do cesty `/etc/ceph/`

- Vytvoříme si image
  - `rd -n client.labX create DU-workshop/labX -s $((100*1024)) --image-format 2 --image-feature layering`
- Zkontrolujeme zdárné vytvoření
  - `rd -n client.labX list DU-workshop`
- Namapujeme RBD do systému
  - `sudo rd -n client.labX map DU-workshop/labX`
- Zkontrolujeme připojení ve zprávách od kernelu
  - `dmesg`

- Blokové zařízení nejdříve zašifrujeme přes dm-crypt/LUKS
  - `sudo yum install cryptsetup-luks`
  - `sudo cryptsetup -s 512 luksFormat --type luks2 /dev/rbdX`
- Kontrola nastavení
  - `sudo cryptsetup luksDump /dev/rbdX`

- **Blokové zařízení dešifrujeme**
  - `sudo cryptsetup luksOpen /dev/rbd0 luks_rbd`
- **Vytvoříme Btrfs na připojeném zařízení**
  - `sudo parted /dev/mapper/luks_rbd`
  - `mklabel gpt`
  - `mkpart primary btrfs 1MiB 100%`
  - `sudo mkfs.btrfs /dev/mapper/luks_rbdX`
- **Připojíme**
  - `sudo mkdir /mnt/rbd`
  - `sudo mount /dev/mapper/luks_rbdX /mnt/rbd`

- Odpojíme a uzamkneme
  - `sudo umount /mnt/rbd/`
  - `sudo cryptsetup luksClose /dev/mapper/luks_rbd`
  - `sudo rbd -n client.labX unmap DU-workshop/labX`

1. Vytvořte RBD o velikosti 100GB a připojte ho do svého VM
2. Blokové zařízení zašifrujte za pomoci dm-crypt/LUKS
3. Vytvořte na zařízení Btrfs file-systém a připojte ho do cesty /mnt/rbd

## Řešení:

- Nainstalujeme základní Ceph balík
  - `sudo yum install ceph-common`
- Stáhněte `~/ceph.conf` z [https://du.cesnet.cz/\\_media/cs/ceph.conf](https://du.cesnet.cz/_media/cs/ceph.conf) do `/etc/ceph/`
- Zkopírujte `~/ceph.client.labX.keyring` do cesty `/etc/ceph/`



- Vytvoříme si image
  - `rd -n client.labX create DU-workshop/labX -s $((100*1024)) --image-format 2 --image-feature layering`
- Zkontrolujeme zdárné vytvoření
  - `rd -n client.labX list DU-workshop`
- Namapujeme RBD do systému
  - `sudo rd -n client.labX map DU-workshop/labX`
- Zkontrolujeme připojení ve zprávách od kernelu
  - `dmesg`

- Blokové zařízení nejdříve zašifrujeme přes dm-crypt/LUKS
  - `sudo yum install cryptsetup-luks`
  - `sudo cryptsetup -s 512 luksFormat --type luks2 /dev/rbd0`
- Kontrola nastavení
  - `sudo cryptsetup luksDump /dev/rbd0`

- **Blokové zařízení dešifrujeme**
  - `sudo cryptsetup luksOpen /dev/rbdX luks_rbd`
- **Vytvoříme Btrfs na připojeném zařízení**
  - `sudo parted /dev/mapper/luks_rbd`
  - `mklabel gpt`
  - `mkpart primary btrfs 1MiB 100%`
  - `sudo mkfs.btrfs /dev/mapper/luks_rbdX`
- **Připojíme**
  - `sudo mkdir /mnt/rbd`
  - `sudo mount /dev/mapper/luks_rbdX /mnt/rbd`

- Nástroj pro práci se subvolumes
- Vytváří inkrementální snapshoty na zadané cestě
- Možnost definovat retention policy
- Přenos na více cílů i skrze SSH
- V repu nebo na <https://github.com/digint/btrbk>

- Ukázka konfiguračního souboru

Předpokládáme splněný lab 3.

1. Nainstalujte btrbk a projděte jeho nastavení
2. Opět namapujte, dešifrujte a připojte RBD (viz lab 3)
3. Upravte konfigurační soubor btrbk, aby
  - target\_preserve\_min byl na all
  - zálohoval se subvolume home
  - datovou retenci zvolte dle svého uvážení
4. Zavolejte btrbk s vaším konfiguračním soubor a volbou dryrun
5. Pokud je vše v pořádku, aplikujte ho
6. Zkuste smazat souboru z /home a následně ho obnovit ze snapshotu

- `sudo yum install btrbk`

```
target_preserve_min    all

snapshot_dir    @snapshots
volume /mnt/btrfs
  subvolume    home
    snapshot_create    ondemand
    target send-receive    /mnt/rbd/
```

- Ostatní data retention pravidla upravte dle svého uvážení nebo ponechte ve výchozím stavu

- `btrbk -c /etc/btrbk/btrbk.conf dryrun`
- `btrbk -c /etc/btrbk/btrbk.conf run`



The logo for cesnet, featuring the word "cesnet" in a white, lowercase, sans-serif font. Below the text is a horizontal row of seven small white squares, with the second and sixth squares from the left being slightly larger than the others.

cesnet  
"....."

Děkujeme za účast na workshopu!

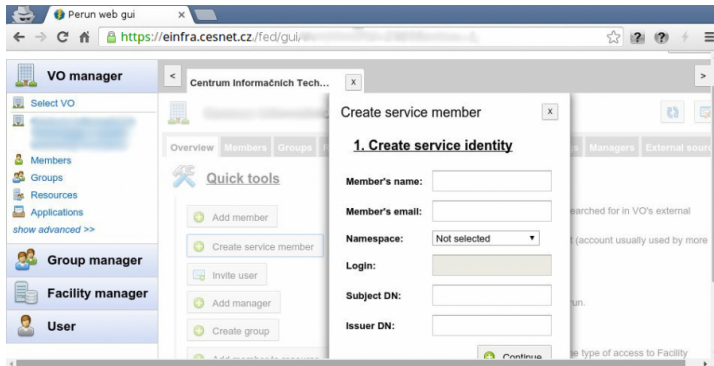
[www.du.cesnet.cz](http://www.du.cesnet.cz)

[du-support@cesnet.cz](mailto:du-support@cesnet.cz)

**cesnet**  
"...."

**Záložní slidy**





The screenshot shows a web browser window with the address bar displaying `https://einfra.cesnet.cz/fed/gul/`. The page title is "Perun web gui". The main content area is titled "Create service member" and contains the following elements:

- A sidebar on the left with a "VO manager" section containing links for "Select VO", "Members", "Groups", "Resources", and "Applications" (with a "show advanced >>" link). Below this are "Group manager", "Facility manager", and "User" sections.
- A main heading "2. Associate real users".
- A search input field, a "Search" button, an "Add" button, and a "Continue" button.
- A table with the following structure:

Name	User type	Count: 0	No items found.
Type in user's First name, Last name (or both) or Login or Email and press Search button.			
- A red warning message: "To be associated:".