



PÁSKOVÁ KNIHOVNA VS. OBJEKTOVÁ TECHNOLOGIE: BOJ O MÍSTO VE SKLADU ČASOPROSTORU

Michal Strnad

CESNET, z. s. p. o.

8. 10. 2023



- Datová úložiště CESNET
- Diskové pole
- Páskové knihovny
- Objektové úložiště

cesnet Datová úložiště jedním slidem

- Zajišťuje provoz a rozvoj národní infrastruktury pro ukládání dat pro výzkum a vývoj
- Aktuálně provozujeme čtyři úložiště založené na object storage, jedno HSM a jedno používá jen disková pole
- Nové servery pro další object storage se právě instalují, další soutěžíme
- HSM a úložiště obsahující jen disková pole jsou přístupná přes NFSv4, FTP, rsync, SCP, ...
- Objektové úložiště pak přes S3/Swift, CephFS, RBD
- Další služby jako FileSender, ownCloud, OpenAccess rep.

Diskové pole

- Riziko v podobě RAID technologie
- Kontrolér v diskovém poli
- Omezené možnosti zrcadlení
- Pozor na Massive Array of Idle Disk (MAID) pole

Pásková knihovna

- Vysoce kapacitní pásy (větší než disky)
- Pásy se snadno přemísťují (off-site záloha)
- Skladovatelnost od 15 do 30 let (archív)
- Náchylnost poškození magnetickým polem, velkou vlhkostí či teplotou
- Sekvenční přístup - pomalá rychlost čtení a pomalý seek time
- Omezený počet použití jedné pásky (dobré je zkusit rotovat)
- Omezený počet robotů a hlavně drivů



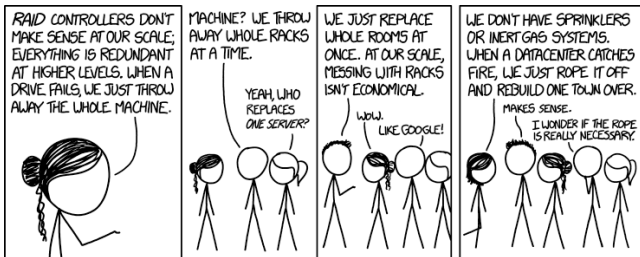
HSM - Hierarchical Storage Management

- Migrační politika (přesun mezi tiery)
- Méně často používaná data jsou migrována
- Zaplnění vyhrazeného prostoru
- i-nody a soubory do 64kB
- Pro uživatele transparentní
- Pomalý recall z pásek
- Režie a pomalost souborových protokolů
- Nutno mít frontendy/aplikační servery



■ Objektová uložště

- Hodně serverů, hodně disků, silné síťové spoje
- Cluster sám udržuje minimální počet nastavených replik/EC
- Větší odolnost proti výpadku serveru, racku, či dokonce celého datacentra



■ Co je ten objekt?

ID	Binary Data	Metadata
1234	0101010101010100110101010010 0101100001010100110101010010 0101100001010100110101010010	name1 value1 name2 value2 nameN valueN

Objektová uložení vs klasický souborový systém

- Flat struktura
- Rozšířená metadata pro vyhledávání, organizaci a správu obsahu
- Rozšiřitelnost a škálovatelnost - dynamicky rozšiřování bez omezení hierarchie adresářů
- Redundantní kopie (replikace) nebo EC
- Jemnější kontrolu přístupu k objektům a metadatům
- Ceph, GlusterFS, IBM Spectrum Scale, Scality ...

- Objektová uložení - Ceph
 - Protokoly S3/Swift, CephFS a RBD
 - CRUSH
 - MON, OSD, PG

- Ceph clustery v podání CESNET DU
 - každý cluster má desítky serverů a tisíc až dva tisíce disků (MTFB)
 - server - 4x10 Gbps / 2x25 Gbps, 300 - 750TB, min. 256 GB RAM
 - cluster připojen N x 100 Gbps

■ Objektová uložení - Ceph

- Obrovské čtecí a zapisovací rychlosti (hlavně větších datových bloků)
- Vhodné na D2D zálohy (rotace dat) a jejich rychlou obnovu
- Dobré i pro živější data
- Bulk storage
- Warm archives
- Stoupající rožšířenost S3 protokolu
- Blokovaná zařízení pro virtuály ale i jiné použití
- V případě potřeby file-systému máme CephFS

- Objektová uložení - Ceph
 - Hodně HW znamená
 - Velká úmrtnost HW (i jeden disk může udělat problém)
 - Je třeba detailní monitoring
 - Velká spotřeba ele.
 - Komplikovanější technologie
 - Nároky na síť (latence -> IOPS) a i znalosti
 - Dual-stack není ještě v plné formě

Pásková knihovna vs. objektová technologie

- Rychlost
- Škálovatelnost
- Odolnost proti výpadku disků, serverů ...
- Větší nároky na znalosti adminů

- Zvu vás na prezentaci a navazující workshop ohledně Cephu na OpenAltu
- We are hiring!

The logo for cesnet, featuring the word "cesnet" in a white, lowercase, sans-serif font. Below the text is a graphic element consisting of a series of white dots of varying sizes, arranged in a pattern that suggests a network or data flow.

cesnet
.....

Děkuji za pozornost

www.du.cesnet.cz

du-support@cesnet.cz

michal.strnad@cesnet.cz

