

Datová úložiště CESNET

David Antoř

Přehled

- ▶ cíle datových úložišť
- ▶ možnosti využití
- ▶ výhledy budoucího rozvoje
- ▶ podpora projektů FR prakticky
- ▶ výhled plánů datových úložišť

Cíl: dlouhodobé uchování dat

- ▶ cíl: dlouhodobé uchovávání (vědeckých) dat
- ▶ uchování zajišťujeme na úrovni zachování binárních dat (bitstream)
 - ▶ správce úložiště nerozumí odborné povaze dat
 - ▶ ani nemůže rozhodovat o smysluplnosti jejich archivace
 - ▶ proto (zatím) neděláme LTP (Long-Term Preservation)
 - ▶ bitstream je nicméně předpokladem pro uchování dat
 - ▶ další kopie, geografické repliky, redundance v systému

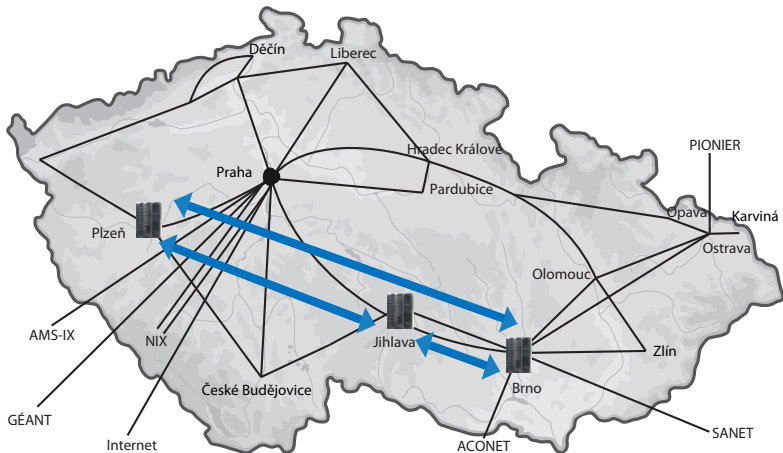
Možnosti využití datových úložišť

- ▶ zálohy
 - ▶ uživatelé mají primární data u sebe
 - ▶ na úložiště odkládají zálohu pro případ havárie
- ▶ archivace
 - ▶ uživatelé na úložiště odkládají cenná primární data
 - ▶ data nejsou často využívána
 - ▶ uživatelé nemají prostředky pro jejich uchování
- ▶ sdílení dat
 - ▶ distribuovaný tým potřebuje společně pracovat nad většími objemy dat, případně je zveřejňovat
- ▶ „něco jiného“
 - ▶ distribuce obsahu, jiné speciální aplikace

Možnosti využití datových úložišť

- ▶ a naopak: na co se vzdálené úložiště příliš nehodí
 - ▶ interaktivní práce zejména s velkým množstvím malých souborů
 - ▶ ukládání dat s potřebou přístupu v reálném čase
 - ▶ prioritou je spolehlivost uložení
 - ▶ přímý provoz systémů s nutností 24/7 přístupu
 - ▶ „toto není úložiště pod váš SAP“
 - ▶ na to je obecně vzdálené úložiště nevhodné
- ▶ vhodné použití pro podobné scénáře
 - ▶ primární 24/7 data u uživatele, na CESNET zálohy/repliky
 - ▶ rozdělení životně důležitých dat vs. „nice to have“ dat

Infrastruktura



Přenos dat na úložiště

- ▶ souborové
 - ▶ NFSv4 (známé uživatelům MetaCentra)
 - ▶ rsync, scp, FTPS
 - ▶ obvykle autentizované systémem Kerberos
 - ▶ existují GUI klienti i nástroje pro příkazovou řádku pro hlavní platformy
 - ▶ Globus Online – vysokorychlostní přenosy
 - ▶ výhledově CIFS (známý „síťový disk“ z Windows)
- ▶ speciální aplikace
 - ▶ gridové úložiště v systému dCache
 - ▶ FileSender, ownCloud, Globus
 - ▶ je možno i jiné – nutno řešit individuálně
- ▶ bloková zařízení (není preferováno)

Výhled

- ▶ cílem je udržet úložiště s otevřeným přístupem pro vědeckou komunitu
- ▶ ideové pilíře rozvoje
 1. technologie a její provoz
 2. organizace dat (Hadoop, databáze)
 3. aplikace – analýza velkých dat – včetně schopnosti fungovat jako testovací prostředí pro takové aplikace
 4. dlouhodobé uchování dat – poskytnutí nástrojů pro práci „aplikačního knihovníka“
- ▶ body 2 a 3 v úzké spolupráci s národním gridem MetaCentrum

Podpora projektu prakticky

- ▶ napište nám na `du-support@cesnet.cz`
- ▶ domluvíme se, co potřebujete
 - ▶ opravdu to potřebujeme pochopit
 - ▶ kolik dat máte, jak s nimi pracujete, ...
- ▶ připravíme konfiguraci úložiště
- ▶ typicky založíme virtuální organizaci
- ▶ dál si uživatele spravujete sami, vytváříte skupiny, ...
- ▶ ukládáte data přes FTP/rsync/NFSv4/...
 - ▶ my se staráme, aby na se vytvářely jejich repliky, ...

Úložiště a Fond rozvoje

- ▶ Fond rozvoje
 - ▶ vyhlašuje výzvu
 - ▶ posuzuje projekty samostatně a nezávisle
 - ▶ k posouzení si vyžádá stanovisko o proveditelnosti
- ▶ datová úložiště mohou podpořit pouze projekty
 - ▶ ve kterých je to technicky možné
 - ▶ které nejdou proti dlouhodobé strategii úložišť
- ▶ důsledky a doporučení
 - ▶ projekty na standardní použití úložišť (zálohy, archivaci) jsou triviálně podpořitelné
 - ▶ cokoli jiného nebo při jakýchkoli pochybnostech konzultujte předem

Kontakty

- ▶ <http://du.cesnet.cz>
- ▶ uživatelská podpora: du-support@cesnet.cz