



Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE

Řešení pro správu velkých dat s využitím služby pro datová úložiště CESNET - MS 2015-05-14

14.5.2015 Malé sbírky



Program

- Představení projektu
- Portfolio služeb DÚ CESNET
- Stručné zhodnocení stávajícího využívání DÚ CESNET
- Případové studie uživatelů DÚ na PřF UK
- Naše zkušenosti s DÚ CESNET
- Diskuze



Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE

CESNET



Sdružení CESNET založily v roce 1996 veřejné vysoké školy a Akademie věd České republiky. CESNET buduje, provozuje a rozvíjí českou národní e-infrastrukturu určenou pro potřeby vědy, výzkumu, vývoje a vzdělávání.

Založeno: 1996

Představení projektu:



Záměr sdružení CESNET v oblasti Přístup do e-infrastruktury

- integrace datových úložišť CESNET do zálohovacích procesů instituce
- propojení úložných systémů instituce s datovými úložišti CESNET
- oblast náročných výpočtů a zpracování dat
- přenosy audiovizuálního obsahu s vysokým rozlišením
- využití dělených displejů při zobrazování a sdílení obrazu s vysokým rozlišením ve výzkumu, vývoji, výuce a další kreativní činnosti (Sdružení CESNET poskytne jako podporu vybranému týmu na období 6 měsíců s možností prodloužení vlastní mobilní dělený display s rozlišením 7860x2160 pixelu, který je řízen systémem SAGE).

Důležitým výsledkem budou zpracované případové studie a doporučení užitečná pro celou akademickou komunitu.



Představení projektu:

Cíl1 - připravit standardizovaná řešení pro ukládání dat v DÚ Cesnet

Cíl	Výstup	Komentář
C1 Zavedení CIT jako agreátora a správce virtuální organizace (VO_prf) pro přístup vědeckých pracovišť fakulty k datovým úložištím CESNET	V1-1 Analýza, vyhodnocení a publikování stavu VO_prf na semináři správců IT vědeckých pracovišť	Pokračování semináře z roku 2014, vyhodnocení zkušeností zavedených pracovišť
	V1-2 Zpracování checklistu pro zájemce o využití služeb DÚ CESNET	Aktualizace dotazníků a návodů na základě V1-1
	V1-3 Testování, analýza a vyhodnocení zálohovacích metod pro typizovaná řešení	Navrh "instatntních" řešení pro různé kategorie uživatelů, konzultace se zástupci OSSS UPa
	V1-4 Zavedení dalších žadatelů do VO_prf	Podle způsobu řešení a připravenosti infrastruktury, pokračování pilotního projektu z r 2014
	V1-5 Analýza, via V1-1 vč. konzultace s pracovištěm další VVŠ	Pokračování semináře se zástupci IT pracovišť, návrhy postupů a metrik, konzultace se zástupci OSSS UPa
	V1-6 Zpracování a nasazení jednotné metodiky pro typizovaná řešení	Zavedení dalších žadatelů podle metodik, vytvoření a publikace case study
	V1-7 Prezentace na pracovišti jiné VVŠ	Oponentura case study a test proveditelnosti na OSSS UPa přechod z pilotního na standardní provoz
	V1-8 Publikace výsledků, standardní nasazení v produkčním prostředí	Vyhodnocení, změny na základě připomínek a zkušeností z pilotního provozu, dokončení a publikace finální podoby metodik a postupů



Představení projektu:

Cíl2 - zlepšit interní síťovou a serverovou konektivitu

Cíl	Výstup	Komentář
C2 Zlepšení interní síťové a serverové konektivity	V2-1 Analýza provozu stávající VO_prf	Analýza rychlosti přenosu a objemů ukládaných dat
	V2-2 Revize infrastruktury	Analýza a kontrola úzkých hrdel přenosů dat v infrastruktuře, návrhy na vylepšení
	V2-3 Posílení výkonu a připojení virtuálního prostředí a vybraných částí síťové vrstvy	Příprava a implementace řešení

Cíl3 - vypracovat návody a postupy, poskytovat podporu a součinnost pro vědecká pracoviště při ukládání dat v DÚ Cesnet

Cíl	Výstup	Komentář
C3 Publikace výsledků a ověření řešení na pracovišti jiné VVŠ	V3-1 Prezentace stavu využití VO_prf na semináři IT správců	Prezentace a panelová diskuse se správci IT zařazenými do VO_prf a ostatními zájemci via V1-1
	V3-2 Návody a dotazníky pro uživatele a zájemce o služby v rámci VO_prf	Příprava a publikace dokumentů
	V3-3 Prezentace navržených postupů a aktuální dokumentace	Pracovní seminář za účasti zástupců OSSS UPa
	V3-4 Koordinace výsledků a zkušeností z pilotních provozů pracovišť jako základ postupů a metrik	Publikace materiálů ze semináře se zástupci IT pracovišť a OSSS UPa
	V3-5 Případová studie a test proveditelnosti	Oponentura případové studie a test proveditelnosti na OSSS UPa
	V3-6 Metodiky a postupy	Publikace metodiky a postupů

Cíl4 - podporovat využití dalších služeb e-infrastruktury CESNET (Filesender, ownCloud, videokonference, Metacentrum atd.)



Klasické možnosti využití úložišť

- zálohy
 - uživatelé mají primární data u sebe
 - na úložiště odkládají zálohu pro případ havárie
 - buď pro zálohování jednotlivých strojů
 - nebo i agregovaně – IT oddělení zálohuje celou katedru
- archivace
 - uživatelé na úložiště odkládají cenná primární data
 - data nejsou často využívána
 - uživatelé nemají prostředky pro jejich uchování
 - individuální přístup koncových uživatelů vs. „laboratorní archivár“



Klasické možnosti využití úložišť

- sdílení dat
 - distribuovaný tým potřebuje společně pracovat nad většími
 - objemy dat, případně je zveřejňovat
 - typicky koncoví uživatelé



IT infrastruktura na PŘF UK

- vhodné pro produkční (“živá”) data formou sdílení nebo jako souborové systémy virtuálních serverů
- nevhodné pro zálohování a archivaci dat
- v minulém roce došlo k rozšíření úložných kapacit, které se využívají mj. právě v kombinaci s DÚ CESNET
- správci dat mohou provádět zálohování a archivaci z centralizovaného úložiště PŘF UK do datového úložiště CESNET (metodická podpora ze strany CIT, v současnosti probíhá výběr a testování vhodných technologií, aby bylo možné uživatelům doporučit konkrétní postupy)



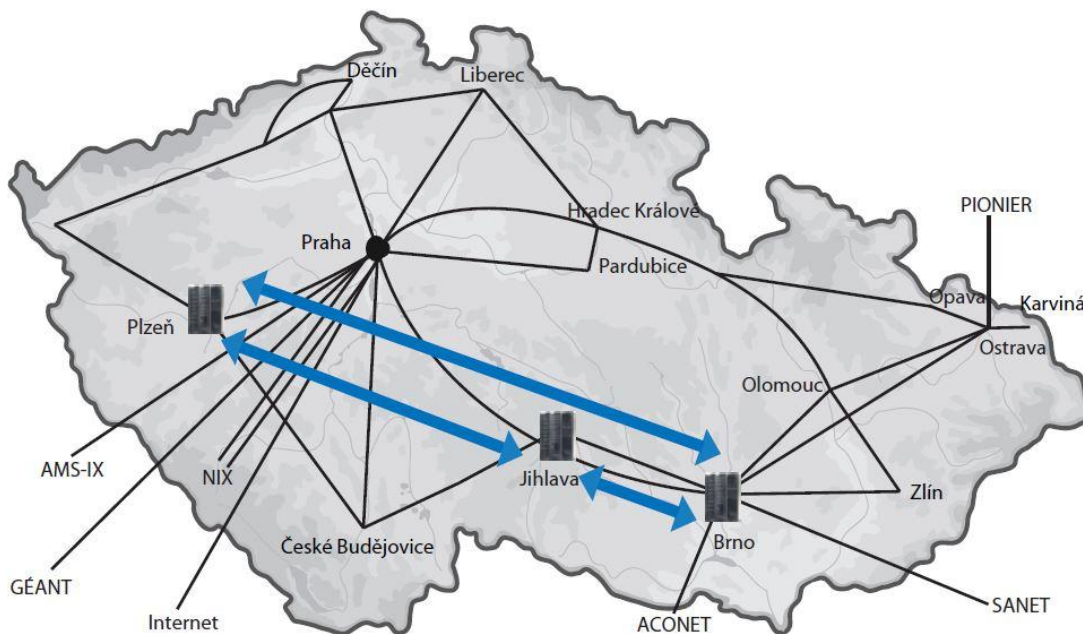
Lokální vs. centralizovaná a externí úložiště

- uživatelé raději drží data na vlastním zařízení
 - to ale nedává rozumné garance dlouhodobého uchování
 - používají se nevhodná média, chybí systematická péče
 - data je lépe uchovávat v živém systému
- konsolidace zdrojů – národní úložiště vědeckých dat (CESNET)
 - umožňuje pořídit zařízení dostatečné velikosti
 - spravované specializovaným týmem
 - s možností sdílení dat mezi skupinami uživatelů



Služba datových úložišť - infrastruktura

- 3 geograficky oddělená úložiště - Plzeň, Jihlava, Brno
- poblíž přístupových bodů páteřní sítě CESNET2





Služba datových úložišť - infrastruktura

- cca 21PiB, PŘF UK jako člen CESNET má zarezervovanou část této kapacity
- distribuovaný a hierarchický model (HSM Hierarchical Storage Management)
- princip “tierů” (vrstev) - méně často používaná data jsou odsouvána na levnější média s vyšší kapacitou
- přístup k dlouho nepoužitému souboru až řádově minuty



Služba datových úložišť - infrastruktura

- Plzeň - v areálu ZČU
 - Tier1
 - 500 TiB diskového prostoru FC disky
 - Tier 2
 - 3 584 TiB na vypínatelných SATA discích
 - Tier 3
 - 4 800 TiB na páskách (Spectra Logic T-Finity)





Služba datových úložišť - infrastruktura

Aktuálně jsou dostupné následující migrační politiky.

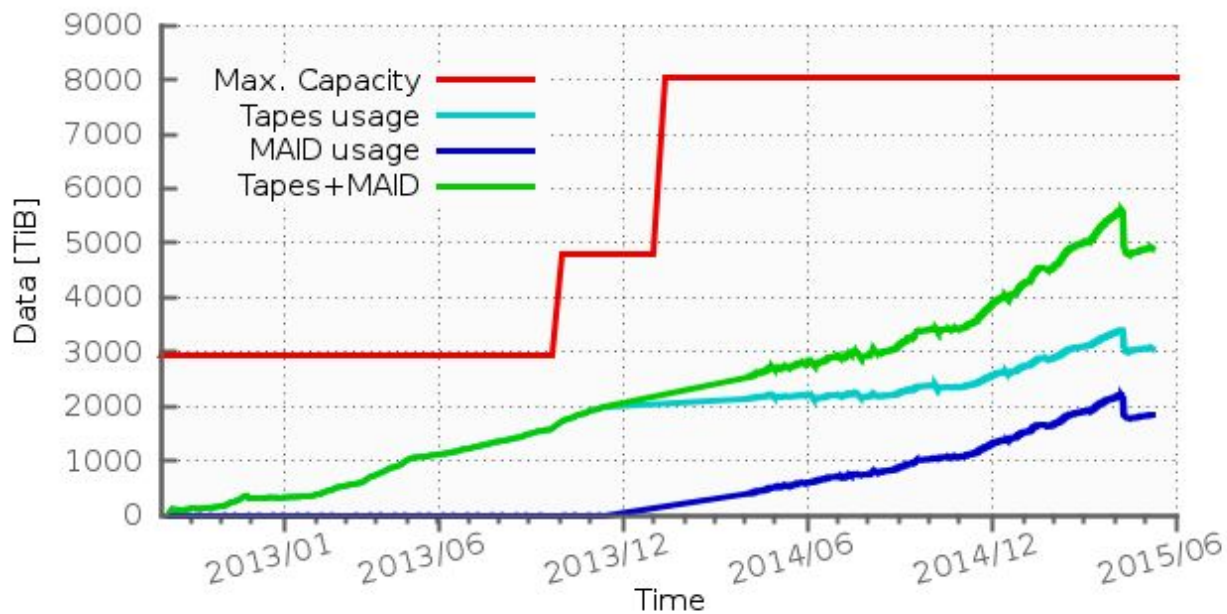
- **tape** - data jsou na SATA discích a migrována na jednu sadu pásek (vhodné pro druhou nebo třetí zálohu dat)
- **tape_tape** - pro DÚ Plzeň platí, že data jsou na SATA discích a kopie na dvou různých páskách (vhodné pro standardní data a první zálohy)
- **disk_only** - data jsou jen na discích (RAID) a nikdy se nepřesouvají ani nezálohují (vhodné pro rychloobrátková data, např. zálohy, které budou za jednotky dnů odstraněny)
- **cache_tape** - primární data jsou na rychlých Fibre Channel discích, jedna kopie na páskách a SATA discích. Pokud dojde místo na SATA oblasti, vytvoří se druhá kopie na páskách, tj. celkem dvě různé pásy (vhodné pro rychlejší přístup ke standardním datům).



Služba datových úložišť - statistiky DÚ Plzeň

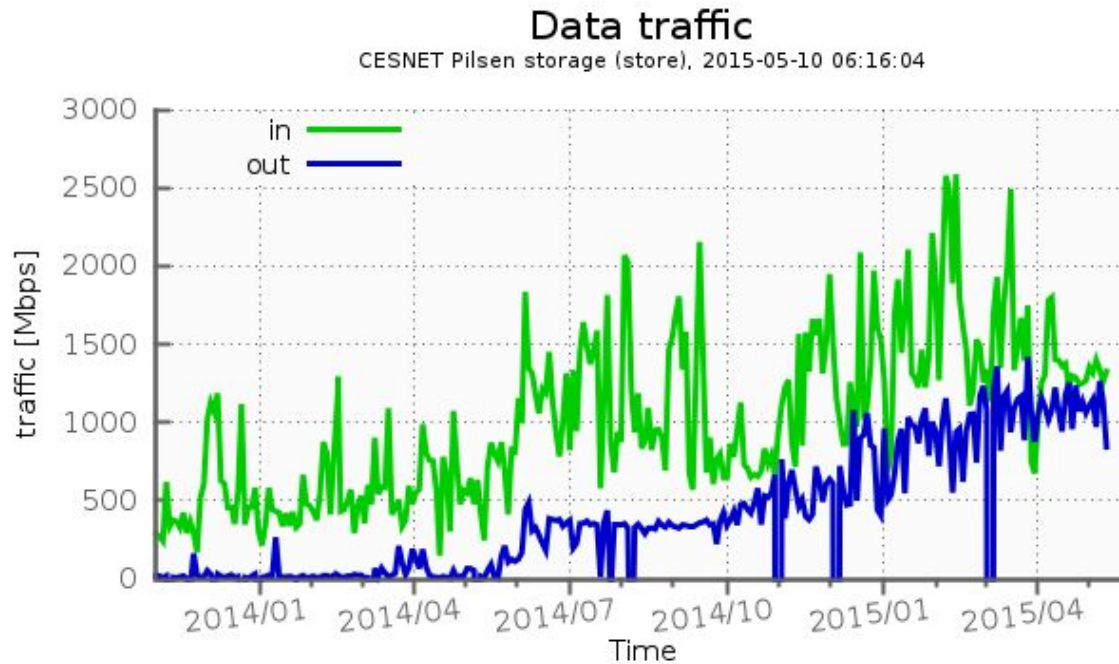
Data stored on tapes/MAID

CESNET Pilsen storage, 2015-05-10 06:16:01





Služba datových úložišť - statistiky DÚ Plzeň



Služba datových úložišť - využití na PŘF UK

Pracoviště	Vedoucí týmu	Typ dat	Objem dat (GB)	Poznámka
Katedra zoologie (1700)	RNDr. František Šťáhlavský, Ph.D.	data z mikroskopů	1000	archiv
Katedra aplikované geoinformatiky a kartografie (3700)	RNDr. Lucie Kupková, Ph.D.	data KRMAP	15000	archiv
Laboratoř konfokální a fluorescenční mikroskopie (1035)	Mgr. Ondřej Šebesta	data z konfokálního mikroskopu	10000	archiv
Katedra genetiky a mikrobiologie (1400)	Mgr. Martin Kuthan, Ph.D.	zálohy dat z mikroskopů pro pracovní skupinu	2340	záloha
Ústav pro životní prostředí (7500)	prof. RNDr. Evžen Stuchlík, CSc.	zálohy dat terénní stanice Velký Pálenec (Blatná)	350	záloha
Katedra botaniky (1200)	Mgr. Viera Mrázová	skeny položek z herbářových sbírek ve vysokém rozlišení	400	záloha
Mapová sbírka (3040)	Mgr. PhDr. Eva Novotná	záloha dat z lokálního úložiště	20000	záloha
Knihovna geografie (3020)	Mgr. PhDr. Eva Novotná	záloha fulltextů databáze GEOBIBLINE	30	záloha
		celkem objem dat	49120	



Služba datových úložišť - využití na PŘF UK 04/2015

Počáteční čas	Trvání	IP adresa	Bajty na vstupu
20.4.2015 14:59:00	1 t, 9 h, 32 m, 6.608 s	karto21.natur.cuni.cz	8 GB (38.8%)
1.4.2015 9:03:00	3 t, 5 d, 5 h, 30 m, 28.407 s	stahlavsky.natur.cuni.cz	6.09 GB (29.6%)
1.4.2015 12:36:00	3 t, 1 d, 3 h, 22 m, 46.370 s	cell-r.natur.cuni.cz	3.96 GB (19.2%)
2.4.2015 9:50:00	3 t, 23 h, 29 m, 55.635 s	provis.natur.cuni.cz	2.47 GB (12.0%)
14.4.2015 11:10:00	2 t, 1 h, 43 m, 29.167 s	pracovna.natur.cuni.cz	80.44 MB (0.4%)
16.4.2015 14:10:00	3 h, 29 m, 11.575 s	ts8.natur.cuni.cz	81.43 KB (0.0%)
21.4.2015 14:24:00	50.722 s	helix2.natur.cuni.cz	19.27 KB (0.0%)
27.4.2015 13:31:00	3 m, 18.181 s	eduroam170.natur.cuni.cz	11.48 KB (0.0%)
		ftp.du1.cesnet.cz	20.6 GB (100.0%)



Služba datových úložišť - využití na PŘF UK 04/2015

Počáteční čas	Trvání	IP adresa	Bajty na vstupu
1.4.2015 12:47:00	3 t, 6 d, 6 m, 51.560 s	botany.natur.cuni.cz	630.73 GB (74.1%)
9.4.2015 14:00:00	2 t, 20 h, 45 m, 56.540 s	analysis.natur.cuni.cz	219.87 GB (25.8%)
19.4.2015 18:23:00	1 h, 2 m, 47.098 s	nothofagus.natur.cuni.cz	129.66 MB (0.0%)
16.4.2015 14:05:00	1 t, 4 d, 22 h, 49 m, 15.309 s	ts8.natur.cuni.cz	1.1 MB (0.0%)
28.4.2015 9:25:00	2 h, 23 m, 37.280 s	ts5.natur.cuni.cz	83.42 KB (0.0%)
5.4.2015 16:03:00	2 t, 3 h, 39 m, 32.137 s	velesb2.natur.cuni.cz	37.21 KB (0.0%)
		ssh.du1.cesnet.cz	850.73 GB (100.0%)



Služba datových úložišť - využití na PŘF UK 04/2015

Počáteční čas	Trvání	IP adresa	Bajty na vstupu
31.3.2015 23:59:00	3 t, 6 d, 12 h, 54 m, 11.970 s	data.natur.cuni.cz	397.59 GB (100.0%)
17.4.2015 14:48:00	1 t, 3 d, 22 h, 2 m, 12.195 s	botany.natur.cuni.cz	2 KB (0.0%)
		nfs.du1.cesnet.cz	397.59 GB (100.0%)



Služba datových úložišť - popis

Datové úložiště je provozováno v „**best effort**“ režimu.

To znamená, že provozovatel DÚ **vyvíjí veškeré úsilí k zaručení dostupnosti a bezpečnosti datových služeb.**

Vysoká dostupnost a bezpečnost datových služeb je dána také redundancí hardwarových komponent, bezpečnostními politikami HSM systému, zálohováním požadovaných dat pro jednotlivé uživatele či VO a v neposlední řadě parametry technické a systémové podpory, kterou v režimu 7×24 smluvně zajišťuje dodavatel DÚ.



Služba datových úložišť - popis

Provozovatel ale **nedává žádné záruky** na persistenci uložených dat a zříká se tak právní zodpovědnosti za jejich uchování, a to zejména kvůli zásahu vyšší moci.

Data uživatelů zásadně patří uživatelům.

Uživatel bere na vědomí, že informace o uživateli a jeho činnosti na úložišti jsou ukládány pro účely administrativy, provozu, statistik, monitoringu a bezpečnosti.



Služba datových úložišť - pravidla

- pro využití služeb DÚ CESNET platí [provozní pravidla](#)
- od 1.2.2015 platí **nová verze** provozních pravidel
- nově ošetřují případ, kdy uživatel neprodlouží členství ve virtuální organizaci nebo ztratí přístup k federaci, ale předtím, než členství ztratí, si nezkopíroval svá data. V takovém případě se uživateli data jednorázově vydají



Služba datových úložišť - pravidla

výtah z pravidel:

- Uživatel smí DÚ využívat pouze pro přenos a ukládání dat, která souvisejí s činnostmi ve vědě, výzkumu, vývoji, šíření vzdělanosti, kultury a prosperity, detailně viz „Zásady přijatelného užití Velké infrastruktury CESNET (Acceptable Use Policy, AUP)“
- Uživatel nesmí užívat DÚ způsobem porušujícím platné zákony (zejména autorský zákon a zákon o ochraně osobních údajů), obcházet administrativní a bezpečnostní opatření pro přístup k DÚ, nebo se o takové obcházení pokoušet.



Služba datových úložišť - přehled protokolů DÚ

- NFSv4
- FTPS
- SCP/SFTP
- Rsync
- **nově CIFS**



Služba datových úložišť - přehled protokolů DÚ

NFSv4

- dostupný zdarma pro klienty využívající Linux
- pro uživatele Windows je dostupný komerční klient NFS Maestro
- na operačním systému Mac OS je problém s mapováním identit a tudíž je tento protokol veskrze nepoužitelný
- bez pomoci dalších nástrojů jako je tar, nelze pomocí tohoto protokolu zachovat původního vlastníka

Výhody

- umí připojit svazek tak, aby se tvářil jako lokální disk
- reconnect
- nemá problémy s hardlinky



Služba datových úložišť - přehled protokolů DÚ

FTPS

- je protokol FTP, který pro přihlášení, a volitelně i pro data, využívá TLS protokol zajišťující šifrování přenášených dat. FTP je vhodné především pro přenos velkých souborů. Nešifrované FTP není podporováno.

Nevýhody

- nelze zachovat vlastníka původních souborů
- standardně také ani čas modifikace souborů

Klienti

- řádkový klient lftp
- Total Commander
- WinSCP
- FileZilla
- řádkový klient CurlFtpFS



Služba datových úložišť - přehled protokolů DÚ

SCP/SFTP

- je protokol využívající SSL, vhodný pro přenos velkých souborů

Výhody

- nemá problém s přenosem dalších informací o souborech (oprávnění, časy modifikace a podobně)

Nevýhody

- má problémy s hardlinky

Klienti

- řádkový klient rsync
- GUI klient Grsync



Služba datových úložišť - přehled protokolů DÚ

RSYNC

- Rsync je jednoduchý nástroj vhodný pro situace, kdy není vyžadována historie záloh, jen přesná kopie.

Výhody

- Umožňuje zachovávat práva a informace o vlastníkově souboru při přenosu na vzdálené úložiště v rozšířených atributech souborů

Nevýhody

- rychlost přenášení nových souborů limitována na úroveň cca 40 souborů za vteřinu bez ohledu na jejich velikost (rsync dělá při kopírování nového souboru několik metadata operací)

Klienti

- řádkový klient pro scp
- řádkový klient pro sftp
- WinSCP
- FileZilla



Služba datových úložišť - způsob užití DÚ

- na fakultě založena virtuální organizace **VO_prfuk**
- VO je skupina uživatelů se společným zájmem, kteří vystupují jako celek
- VO má správce, který
 - jedná s poskytovatelem zdrojů
 - rozhoduje o podmínkách členství jednotlivých uživatelů
- dohoda o poskytování zdrojů mezi VO a správcem zdroje
 - popisuje nastavení technických parametrů
 - velikost poskytovaného prostoru
 - použité přístupové protokoly
 - politika ukládání dat, počty kopií/replik
- VO může využívat různé typy zdrojů



Služba datových úložišť - způsob použití DÚ

Uživatel resp. pracoviště kontaktuje CIT

Domluví se způsob přístupu, plánovaný objem dat atp., adresářová struktura a seznam lidí ve skupině pro sdílení dat












Uživatel provede registraci do naší virtuální organizace přes




- https://perun.metacentrum.cz/perun-registrar-fed/?vo=VO_prfuk

!Členství ve virtuální organizaci je třeba jedenkrát ročně prodloužit!



Přihlásit účtem

-  Univerzita Hradec Králové 
-  Univerzita Jana Evangelisty Purkyně v Ústí nad Labem 
- Univerzita Karlova v Praze** 
- Univerzita Palackého v Olomouci 
- Univerzita Pardubice 
- Univerzita Tomáše Bati ve Zlíně  Tomas Bata University in Zlin
- Ústav chemických procesů AV ČR 
- Ústav experimentální botaniky AV ČR 
- Ústav fotoniky a elektroniky, Akademie věd ČR 

  CESNET 



Přihláška do VO DataC PrFUK



Prfuk je virtuální organizace uživatelů DataC PrFUK, které jsou v rámci datových úložišť poskytovány domluvené objemy dat. Členství v prfuk je schvalováno administrátory DataC PrFUK, podmínkou členství je souhlas s pravidly.

Položky označené hvězdičkou jsou povinné.

Osobní údaje

Titul před
jménem

Jméno*

Titul za jménem

Preferovaný e-
mail*

Kontaktní
telefon*

Organizace*

Preferovaný
jazyk*

Zvolte jazyk, ve kterém chcete dostávat novinky a notifikace.



Přihlašovací údaje

Některé služby VI CESNET vyžadují přihlášení specifickým uživatelským jménem a heslem. Zvolte si, prosím, uživatelské jméno a heslo ke službám VI CESNET a dobře si je zapamatujte.

**Uživatelské
jméno***

Uživatelské jméno musí začínat malým písmenem a obsahovat pouze malá písmena, číslice a podtržení. Doporučujeme délku nanejvýš 8 znaků.

Heslo*

Heslo musí být 8 až 20 znaků dlouhé a obsahovat alespoň 3 písmena a alespoň 1 znak jiný než písmeno. Nesmí obsahovat jméno, příjmení ani login, a to ani pozpátku. Ve druhém políčku zadejte heslo znovu pro kontrolu.

Doplňující údaje pro členství ve virtuální organizaci prfuk

**Plánované
využití VO***

Uvedte stručně, jaká data hodláte ukládat. Uvedte také Vaše nadstandardní požadavky, požadavky, které nejsou pokryty položkami formuláře, případně jiné skutečnosti, které považujete za podstatné pro vyřízení přihlášky.

**Katedra/
ústav/výzkumná
skupina***

Zadejte název a adresu fakulty nebo ústavu.

Stiskem tlačítka 'Stát se členem virtuální organizace VO_prfuk' souhlasím s pravidly využití datových úložišť CESNET.

Stát se členem VO_prfuk



Služba datových úložišť - způsob použití DÚ

Wiki DÚ CESNET - návody a postupy

- <https://du.cesnet.cz/wiki/doku.php>

Aktuální adresy serverů (hostů) pro Plzeň:

- ftp.du1.cesnet.cz
- ssh.du2.cesnet.cz
- nfs.du3.cesnet.cz



Služba datových úložišť - způsob použití DÚ

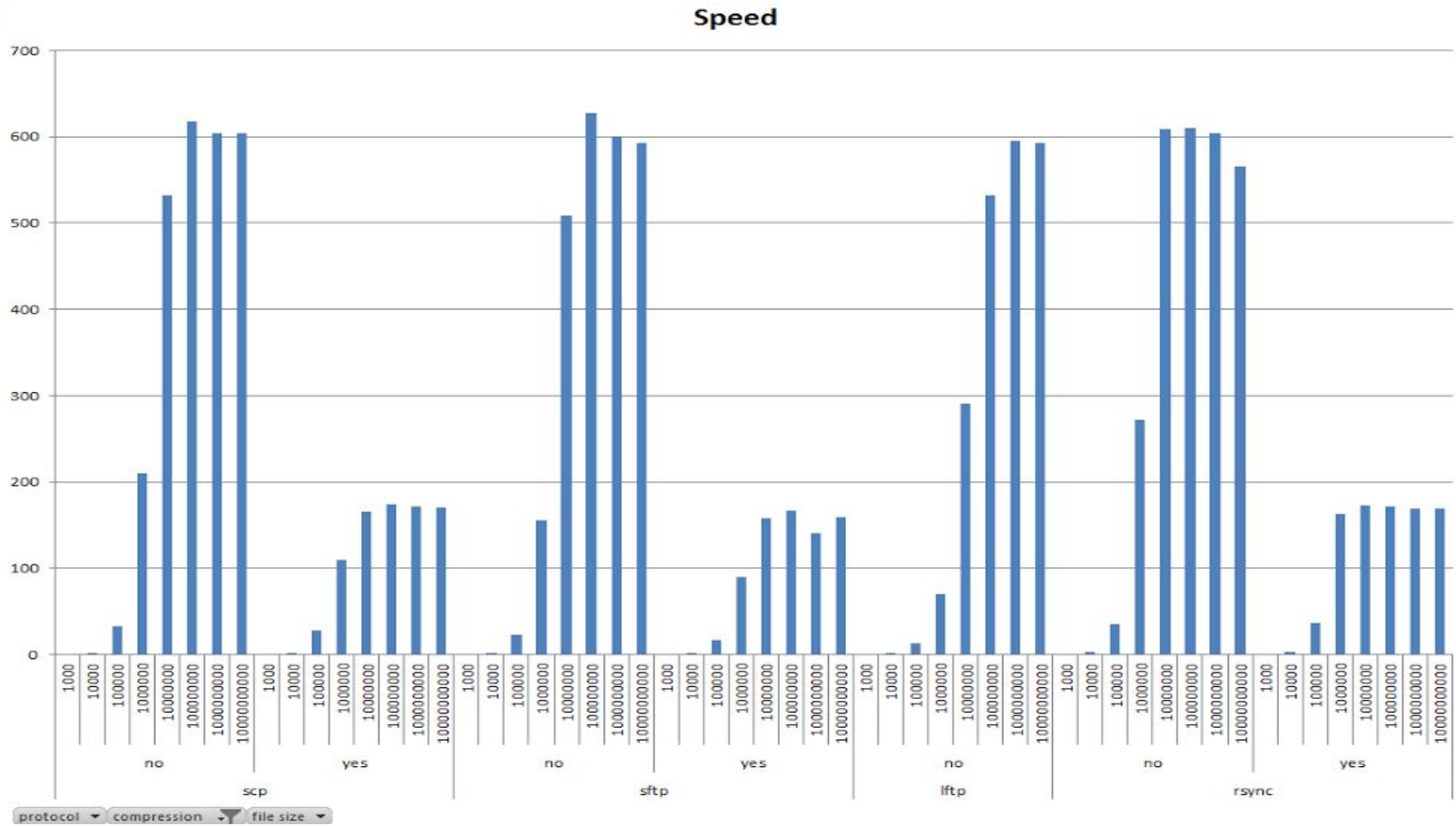
Je doporučeno zvolit takový zálohovací nástroj, který pro zjištění, zda se soubory na primárním úložišti změnily, nevyžaduje kontaktovat server, na který se zálohuje tzn. drží vlastní index odzálohovaných souborů

Pokud je limitujícím faktorem rychlosti přenosu množství zálohovaných souborů, či je průměrná velikost souboru menší než řádově desítky MB, je vhodné soubory seskupovat do větších buď pomocí jednoduchých nástrojů jako je tar, případně lépe pomocí zálohovacích softwarů, které umí zálohovat do souborů předem dané velikosti (virtuálních pásek), typicky několik GB. Příkladem takového softwaru může být open source produkt Bacula.

2014/2015 projekt “Gigabit do každého datového rozvaděče”, který umožní zájemcům o službu DÚ navýšení rychlosti připojení na 1Gb/s.



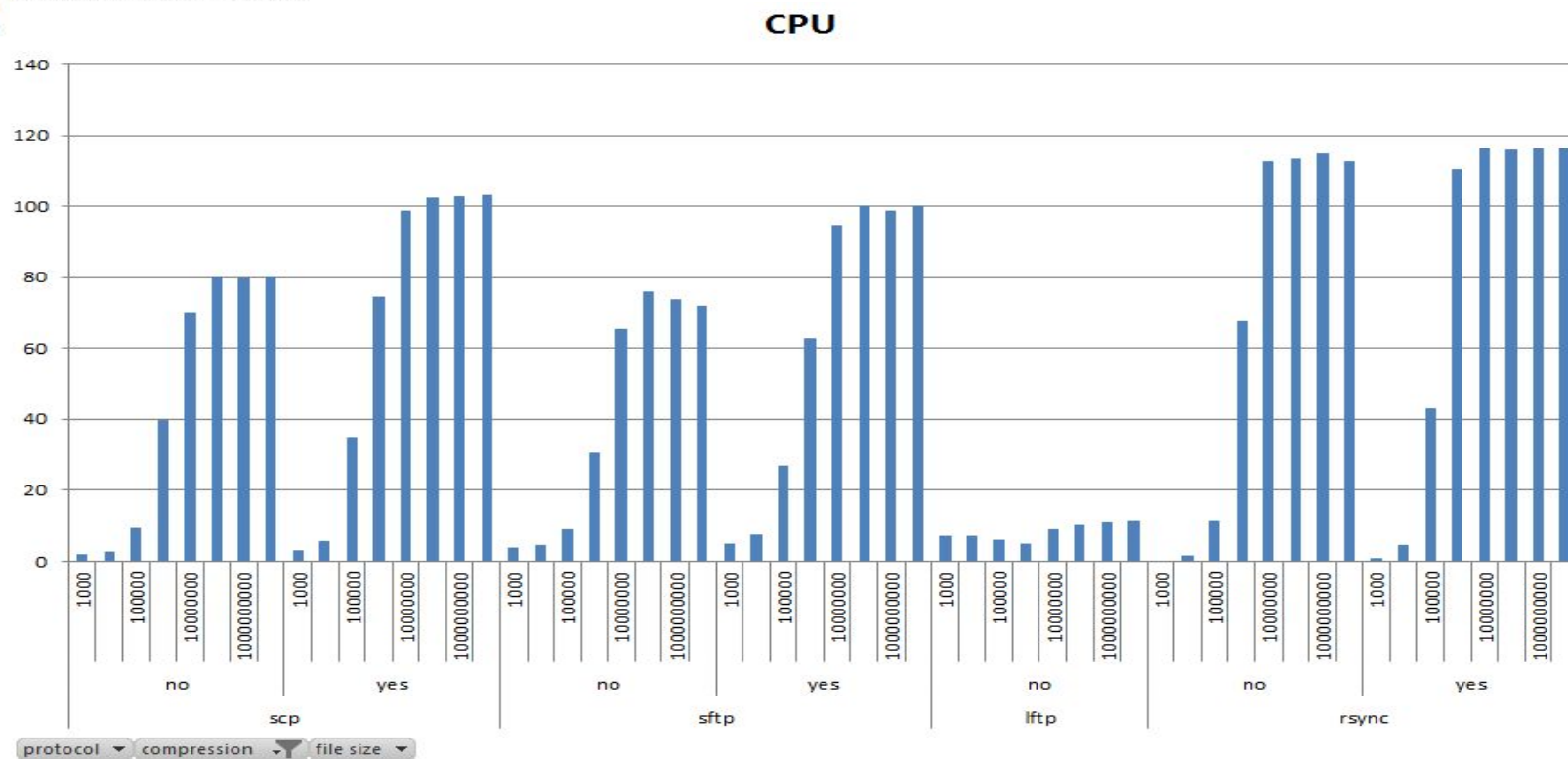
Služba datových úložišť - výkon DÚ



Přenosová rychlost je optimální při menším počtu velkých souborů, protože klesá režie cílového souborového systému. Doporučujeme vše před přenosem “zazipovat”.



Služba datových úložišť - výkon DÚ



- při přenosu velkého množství malých souborů je úzké hrdlo na straně úložiště a klient není příliš vytížen - je lepší přenášet menší počet velkých souborů než hodně malých
- z pohledu zátěže klienta je ideální použít “lftp”;



Služba datových úložišť - **Filesender**

- filesender.cesnet.cz
- sdílení velkých souborů (Adobe Flash 2GB, HTML5 až 500GB)
- max. 100 adresátů
- najednou lze poslat pouze jeden soubor (využít např. zip, rar; nastavit nulovou kompresi)
- příjemce nemusí být členem akademické obce
- doba uložení max. 1 měsíc
- přihlášení přes eduID loginem a heslem do CAS
- možnost využití “voucherů” pro externí spolupracovníky



Služba datových úložišť - **ownCloud**

- webové rozhraní owncloud.cesnet.cz
- v provozu od dubna 2014
- opensource SW pro ukládání, automatickou synchronizaci a sdílení souborů
- alternativa pro Dropbox, OneDrive, Google Drive, ...
- přihlášení přes eduID loginem a heslem do CAS
- prostor 100GB, nahrát lze najednou 16GB
- obsahuje aplikace pro kontakty, kalendář, úkoly, vytvoření textového ODT souboru, integrovaný prohlížeč obrázků, PDF, OpenDocument souborů, txt souborů, ...



Služba datových úložišť - **ownCloud**

- pro přístup k datům se používá WebDAV (Web-based Distributed Authoring and Versioning), ke kontaktům CardDAV, pro kalendář CalDAV
- hostel.eduid.cz/ - pro uživatele z institucí, které zatím nejsou zapojeny do federace identit eduID.cz
- zdarma synchronizační klient pro Linux/Windows/Mac OS + placené app (cca 20Kč) pro mobilní platformy Android, iOS, Windows Phone
- možnost připojit jako disk přes WebDAV např. ve Windows příkazem
net use r: https://owncloud.cesnet.cz/remote.php/webdav/ /user:cislo@cuni.cz

Poslední verze je 1.8.1. vs. problematická verze 1.8.0



Služba datových úložišť - **GlobusOnline**

- nutný účet na www.globusonline.org - lze navázat na Google účet
- služba slouží pro kopírování velkého objemu dat na velké vzdálenosti za pomoci **webového rozhraní**
- data jsou přenášena mezi tzv. endpointy. Jeden endpoint může být Váš home adresář vytvořený v datových úložištích CESNET a druhý buď jiné vzdálené úložiště podporující GlobusOnline nebo vaše pracovní stanice/notebook



Služba datových úložišť - **MetaCentrum**

- pokud potřebujete i výpočetní zdroje
- gridová infrastruktura
- možnost bezplatného využití výpočetní a úložné kapacity a řady aplikačních programů, za všechny např. Matlab, Maple, Gaussian,...
- další informace na <https://metavo.metacentrum.cz/>



Případové studie uživatelů DÚ na PŘF UK - shrnutí samostatných prezentací

Případová studie KRNAP Lucie Červená, 14.5.2015 Malé sbírky PŘF UK

Pracovníci katedry aplikované geoinformatiky a kartografie zpracovávají vědecká data (ortofota, hyperspektrální data AISA a APEX, digitální modely terénu a povrchu) z Krkonošského národní parku. Data měli uložena na fyzických nosičích (HDD) a jejich přenos a zpracování bylo velmi nekomfortní. V rámci DÚ CESNET byla vytvořena skupina a došlo ke zkopírování dat z HDD. Celkový objem je cca 15TB.

Zkušenosti s rychlostí přenosu dat

- Při stahování dat je rychlost zhruba 8 - 10 MB/s, většinou se zpracovává fronta souborů, tj. stahují se dva soubory najednou, každý rychlostí cca 4 - 5 MB/s.
- Při stahování dvou linií AISA dat (celkem cca 600 GB) běží stahování část dne a v noci, do rána vždy bez problémů doběhlo.

Přístup k datům

- Využíváme Filezillu, která vyhovuje.
- Jedinou nevýhodou je, že data jsou uložena na jednom účtu (konkrétně mém) a nelze je nasdílet konkrétním kolegům v síti, co mají taktéž zařízené účty (nebo jsme zatím nezjistili jak, podařilo se mi najít pouze možnost zpřístupnění všem uživatelům z PŘF, ale vybraným nikoliv).

Případová studie YCG Martin

Kuthan, 14.5.2015 Malé sbírky PŘF UK

Dobře zavedené odborné pracoviště pod záštitou Katedry genetiky a mikrobiologie PŘF UK a nově také jako součást projektu BIOCEV se dlouhodobě zabývá genetikou, morfologií, růstovou dynamikou a dalšími aspekty kvasinek. Produkuje významná data, například snímky z mikroskopů ve vysokém rozlišení a genomická data. DÚC využívá jako sdílený prostor pro správu dat, vzhledem ke specifickým požadavkům na řešení přístupových práv do úložiště přistupuje nepřímo s využitím fakultního virtuálního serveru.

Řešení zálohování

- vyhrazená složka v DÚ CESNET, připojená protokolem NFSv4 + Kerberos autentizace
- opensource zálohovací nástroj duply/duplicity jako prozatimní řešení vybrané na základě srovnání několika testovaných nástrojů
- plná záloha 1x měsíčně, přírůstky 1x denně, retence 3 měsíce



Naše zkušenosti s DÚ CESNET

- plně funkční řešení, multiprotokolový přístup, nadstavbové služby
- profesionální podpora - jsou nám hlášeny plánované výpadky, provozní problémy jsou aktivně řešeny, prioritou provozovatele je dostupnost a ochrana dat
- pro využívání DÚ CESNET je třeba, aby uživatel prošel procesem registrace - někdy se setkáváme s tím, že uživatel registraci nedokončí a zbytečně se tak odkládá využívání služeb (teprve po registraci je možné zadat nastavení přístupových práv z naší strany)
- CIT jakožto integrátor zprostředkovává nastavení složek a přístupových práv - webové rozhraní, ne vše funguje automaticky, na straně CESNETu jsou součástí procesu i ruční zásahy operátorů, v zájmu zrychlení je zde přislíbena postupně vyšší míra automatizace



Dotazník

Jaké HW a SW řešení pro ukládání dat v současnosti používáte?

Jaké typy dat spravujete a jakým způsobem je používáte?

(záloha/archiv/db/filesystem)

Jaký je celkový objem dat ve Vaší správě? Kolik % dat plánujete ukládat na DÚ CESNET?

Jste spokojeni se službami DÚ CESNET po cca roce používání?

Máte zájem dále využívat DÚ CESNET?

Které z variant řešení ukládání dat preferujete?

- interní úložiště PŘF UK
- DÚ CESNET
- owncloud
- vlastní - jaké



Dotazník k semináři „Integrace služeb DÚ CESNET do procesu zálohování a archivace dat vědeckých pracovišť PŘFUK“

Dotazník

Prosíme o vyplnění a
předání dotazníku na
závěr semináře

Děkujeme
CIT

Jméno a příjmení: číslo pracoviště:

1. Jaké HW a SW řešení pro ukládání dat v současnosti používáte?

.....

2. Jaké typy dat spravujete a jakým způsobem je používáte? (záloha, archiv, databáze, filesystem, atd.)

.....

.....

3. Jaký je celkový objem dat ve vaší správě? Kolik % dat plánujete ukládat na DÚ
CESNET?

do 1 TB 1-10 TB nad 10 TB %dat

4. Jste spokojeni se službami DÚ CESNET po cca roce používání?

ANO / NE / ZATÍM NEVYUŽÍVÁM

5. Máte zájem dále využívat DÚ CESNET?

ANO / NE

6. Které z variant řešení ukládání dat preferujete? (i více možností)

a) interní úložiště PŘF UK

b) DÚ CESNET

c) OwnCloud

d) vlastní – jaké:



Shrnutí

Seminář 14.5.2015

- představení projektu FR CESNET
- aktualizované informace k DÚ CESNET
- využívání a zkušenosti DÚ CESNET
- sběr dat (dotazník)

Dotazy a připomínky



Děkujeme za pozornost