

Data gram

červen 2012

zpravodaj sdružení CESNET

číslo 28

Datová úložiště

Budovaná Velká infrastruktura CESNET, která představuje prostředí informačních technologií pro vědu a výzkum v ČR, obsahuje čtveřici důležitých složek pro práci vědeckých a výzkumných týmů:

- komunikační infrastrukturu,
- národní grid (výpočetní prostředí),
- infrastrukturu pro ukládání vědeckých dat a
- prostředí pro vzdálenou spolupráci.

Zatímco samotná komunikační infrastruktura spadá do prapůvodního zaměření sdružení a aktivity související s provozem a budováním výpočetního prostředí i nástrojů vzdálené spolupráce mají na CESNETu dlouhou tradici, služby úložné infrastruktury se teprve rodí.

Výhody externího úložiště

Proč vlastně budovat infrastrukturu pro ukládání dat? Uživatelé obvykle raději drží data na vlastním zařízení, pro dlouhodobé uchovávání nicméně takový

přístup není nejvhodnější. V řadě vědeckých týmů se o IT infrastrukturu stará jeden z jeho členů, přičemž tato činnost není primární oblastí jeho zájmu. Pro archivní data se používají nevhodná média a systematická péče o ně není nijak zajištěna.

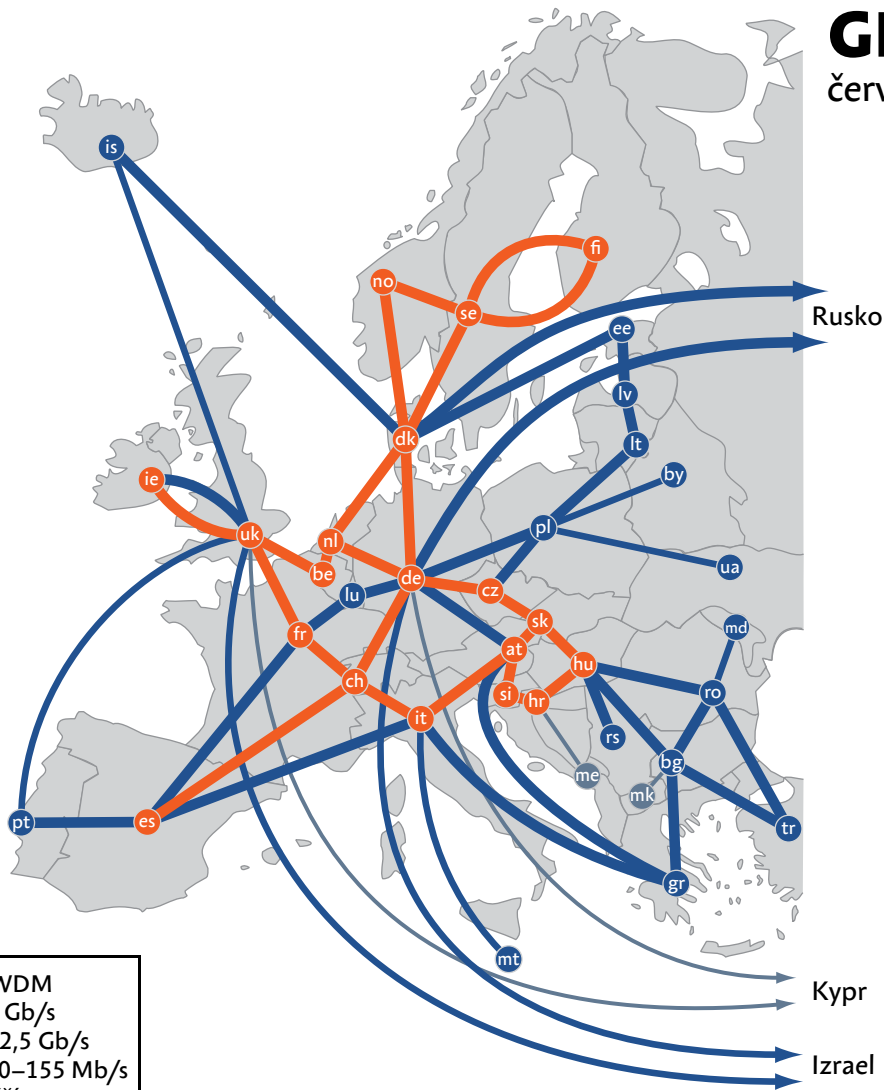
K dlouhodobému, spolehlivému, a přitom ekonomicky výhodnému ukládání dat je vhodné zdroje konsolidovat. Velké archivní úložiště je výhodnější než řada malých, dovoluje pořídit např. velkou páskovou knihovnu, zatímco malé objemy dat na jednotlivých pracovištích jsou drženy spíše na běžících discích. K velkému úložišti má smysl budovat tým profesionálních správců, kteří jsou schopni efektivně minimalizovat rizika ztráty dat. Úložiště na národní úrovni také může poskytnout nástroje pro řízené sdílení dat mezi členy jednotlivých týmů, aniž by si takové řešení museli vytvářet sami.

Potřeba dlouhodobě ukládat vědecká data přitom jasně existuje. Pomalu se stává standardem, že



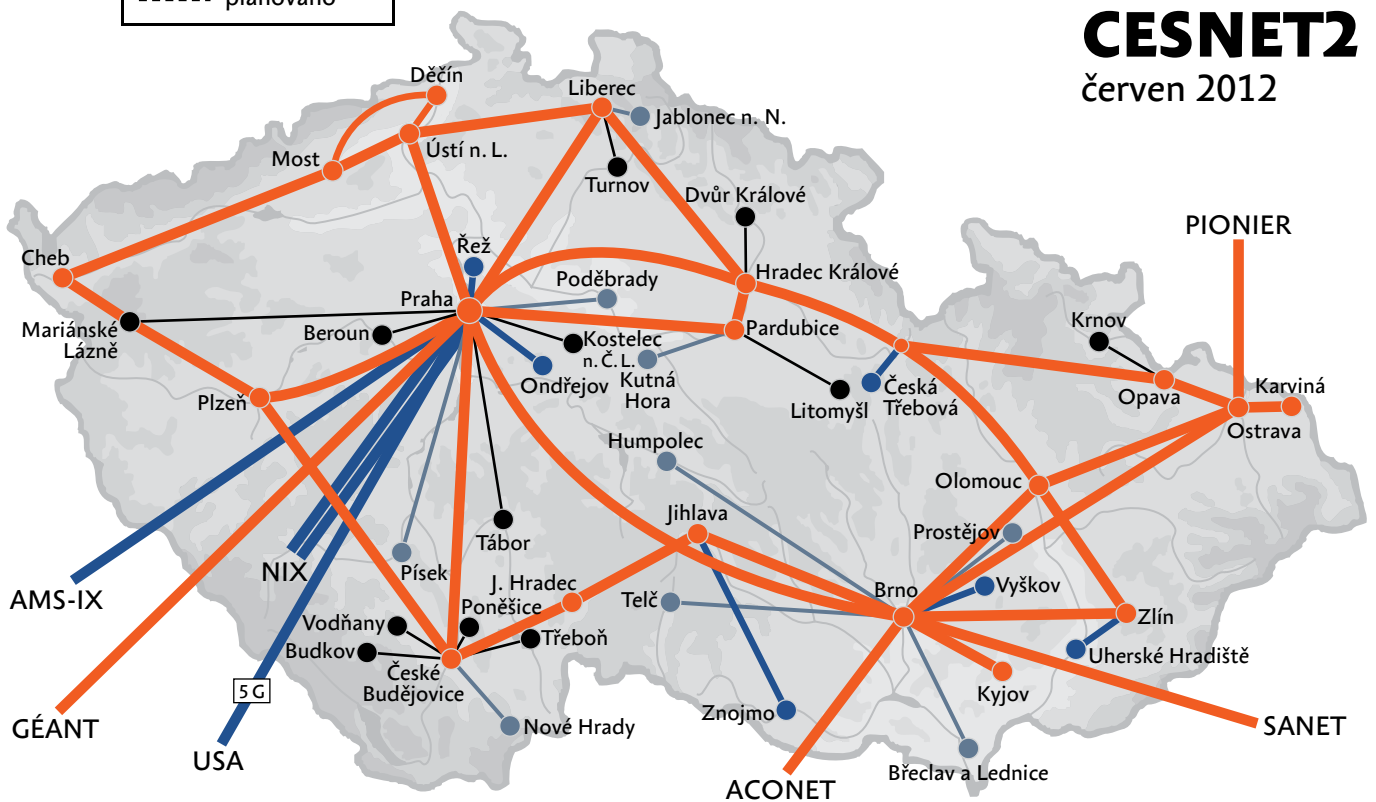
Topologie sítě GÉANT a CESNET2

GÉANT
červen 2012



- DWDM
- 10 Gb/s
- 1–2,5 Gb/s
- 100–155 Mb/s
- nižší
- - - plánováno

CESNET2
červen 2012



primární data z experimentů nebo výpočtů, která jsou nezbytná k vytvoření publikace, je vhodné uchovávat, a to jak s ohledem na potenciál jejich dalšího využití, tak i návratu a revizi publikovaných výsledků.

Možnosti využití

Budované úložiště CESNETu bude sloužit pro řadu nejrůznějších scénářů použití. Často to budou *zálohy*, kdy si uživatelé primární data drží u sebe a na úložiště CESNETu ukládají zálohy pro případ jejich poškození nebo havárie svého primárního úložiště.

Druhým typickým scénářem je *archivace*, kdy uživatelé na úložiště přesouvají data, která jsou sice cenná, ale se kterými nepracují příliš často, přitom nemají vlastní prostředky držet taková data u sebe.

Další možností je *sdílení dat*, kdy distribuovaný tým potřebuje společně zpracovávat větší objemy dat a řídit přístup k těmto datům mezi svými členy, případně je i vystavit někomu dalšímu.

Modely užití nejsou zdaleka omezeny na tyto případy, jsme připraveni s uživateli pracovat na jakémkoli způsobu využití, který pomůže jejich vědecké a výzkumné práci.

Budovaná úložiště

V rámci projektu OP VaVpI *eIGeR* budujeme postupně trojici úložišť ve třech městech: Plzni, Jihlavě a Brně. Začátkem letošního roku byla dokončena dodávka prvního z nich, které je umístěno v areálu Západočeské univerzity. V současnosti toto úložiště zpřístupňujeme prvním uživatelům. Jeho kapacita je cca 500 TB diskového prostoru a 3300 TB v páskové knihovně. Při současných poměrech ceny a výkonu lze odhadovat, že celková kapacita úložišť, kterou v projektu pořídíme, se bude pohybovat kolem 15 PB.

Všechna úložiště jsou koncipována jako *hierarchická*. To znamená, že se skládají z několika vrstev médií a řídicího systému, který automaticky přesouvá delší dobu nepoužívaná data do pomalejších vrstev, jako jsou páskové knihovny nebo vypínatelné disky. Takové řešení umožňuje udržet v rozumných mezích náklady na provoz systému, uživatele omezuje jen minimálně – přístup k datům, která již dlouho nikdo nepoužil, je pomalejší.

Přístup pro uživatele

Zásadním aspektem používání úložišť je organizace přístupu uživatelů. V prostředí e-infrastruktury CESNET zavádíme *jednotnou správu uživatelských účtů*, kde je identita uživatele podchycena na jednom místě a na tuto identitu jsou navázány jednotlivé mechanismy pro přístup k různým systémům, úložištím, výpočetním zdrojům MetaCentra i dalším poskytovaným službám a nástrojům. Uživatelé mají k dispozici jednotné webové rozhraní pro správu svého účtu.

Aby bylo možno rozsáhlý systém s mnoha uživateli spravovat, uživatelé se formují do tzv. *virtuálních organizací*, tedy skupin uživatelů, kteří mají společný zájem a vystupují jako celek. Virtuální organizace má správce, který za ni jedná s poskytovateli zdrojů, v našem případě datového prostoru. Správce dále rozhoduje

o podmínkách členství jednotlivých uživatelů. Prostřednictvím tohoto konceptu je možno například podchytit i distribuovaný spolupracující tým a poskytnout mu úložné kapacity dle jeho potřeb.

Součástí dohody o poskytnutí zdrojů mezi provozovatelem úložiště a virtuální organizací je také nastavení mnoha technických parametrů počínaje velikostí poskytovaného prostoru, přes dohodu, jakými protokoly budou uživatelé k úložišti přistupovat, a konče detailní politikou ukládání dat. Ta mimo jiné zahrnuje i možnosti zvýšit spolehlivost uložení pořízením více kopií dat nebo umístěním dat do několika fyzických úložišť. Jedna virtuální organizace může mít dohody o využívání více typů zdrojů, například prostoru datového úložiště a výpočetní kapacity v národním gridu.

Koncept virtuálních organizací směřuje zejména na organizované týmy se zvláštními potřebami. Jsme ale připraveni podporovat i *individuální uživatele*, kterým vyhovuje standardně nabízená sada protokolů pro přístup k úložišti. Pro takové budeme udržovat virtuální organizaci *Storage*.

Podmínkou využití zdrojů je příslušnost uživatele k akademické organizaci (což platí pro většinu služeb e-infrastruktury) a v případě virtuální organizace *Storage* po uživatelích požadujeme, aby každoročně uvedli v kratičké zprávě, k čemu úložiště využívají. Na úložišti je zaveden systém kvót, které slouží zejména k ochraně před chybou a přetížením úložiště takovou chybou způsobené. Uživatelé virtuální organizace *Storage* mohou standardně uložit až 500 GB dat, na základě stručně zdůvodněné žádosti kvóty dle potřeby navýšíme.

Pro *přístup k úložišti* postupně uvádíme do provozu řadu protokolů. Uživatelům *MetaCentra* je dobře znám síťový souborový systém NFSv4, který umožňuje připojit vzdálené úložiště, jako by to byl lokální disk. Uživatelé Linuxu mají podporu tohoto protokolu prakticky ve všech distribucích. Výhledově zavedeme i protokol CIFS, který je nativně podporován v prostředí Windows, kde opět připojí klasický „síťový disk“.

Mezi metody pro přenos souborů patří *rsync*, *scp*, *SFTP* (tedy FTP přes SSH) a *FTPS* (FTP nad SSL). Ve většině případů je autentizačním mechanismem Kerberos. Pro tyto protokoly existují jak uživatelsky přívětiví klienti pro Windows i Linux, tak i nástroje pro příkazovou řádku. Specifickou možností využití úložiště je gridový storage element realizovaný pomocí nástroje *dCache*. Na úložišti v omezené míře podporujeme také přímý přístup na bloková zařízení, což využívají některé typy zálohovacího software.

Aktuální stav

V současnosti uvádíme první úložiště v Plzni do zkušebního provozu. Přicházejí na něj první uživatelé a postupně jednáme se skupinami, které projeví zájem o používání infrastruktury. Je třeba vždy pochopit jejich způsob práce, abychom dokázali najít režim využití úložiště, který jim bude vyhovovat, a to jak po stránce technické, tak i organizační. S uživatelskými skupinami jsme otevření i vývojové spolupráci nad problémy ukládání dat.

Více o datových úložištích se dočtete na adrese

20 let Internetu v České republice

13. února uplynulo přesně dvacet let od oficiálního připojení tehdejšího Československa k síti Internet. Klíčovou událost domácí informatiky jsme si připomněli na slavnostním setkání, které jsme zorganizovali v den výročí na půdě pražského ČVUT. Právě výpočetní centrum ČVUT se stalo koncovým bodem historického prvního připojení s dnes úsměvnou přenosovou rychlostí 19,2 kb/s, z níž ovšem Internet využíval jen polovinu, druhá sloužila síti EARN.

Jednotlivé příspěvky, které na setkání zazněly, většinou mapovaly historický vývoj počítačových sítí a souvisejících technologií u nás od onoho připojení k Internetu až po současnost. Účastníci si například mohli připomenout první generaci domácí akademické sítě a její překotný rozvoj, kdy během čtyř let zvýšila

přenosové rychlosti páteřních tras z 19,2 kb/s v roce 1993 na 34 Mb/s, tedy bezmála 2000krát.

Během setkání vystoupilo několik významných domácích i zahraničních hostů: rektor ČVUT Václav Havlíček, ředitel CESNETu Jan Gruntorád, Jaroslav Bobovský ze společnosti BESET, publicista Jiří Peterka, spisovatel Ondřej Neff, výkonný ředitel CZ.NIC Ondřej Filip a ředitel ICANN Stephen D. Crocker. Virtuálně prostřednictvím videozáznamu přednesl svůj příspěvek i jeden ze zakladatelů Internetu, Vinton G. Cerf.

Prezentace i videozáznamy všech vystoupení jsou volně k dispozici. Máte-li zájem o malý výlet do historie síťování v České republice, navštivte

<http://www.cesnet.cz/akce/2012/20let-internetu/>



10 let roamingové infrastruktury eduroam

Kulaté výročí letos oslavila i služba, kterou mnozí z vás nepochybně využívají: roamingová infrastruktura *eduroam*, která usnadňuje připojení k Internetu uživatelům během návštěv mimo svou domovskou instituci.

Autorem koncepce, z níž vznikl *eduroam*, je nizozemec Klaas Wierenga, který 30. května 2002 zaslal svou ideu představitelům evropských sítí pro vědu, výzkum a vzdělávání. K zahájení provozu došlo v roce 2003 a o rok později *eduroam* dorazil do České republiky. V současné době již dosahuje do více než 50 zemí na pěti kontinentech.

Z původních šesti zapojených domácích organizací je dnes 38, které sdílejí připojení v bezmála 500 lokalitách. Svým pokrytím patří Česká republika mezi absolutní světovou špičku – z 5000 evropských lokalit se bezmála desetina nachází u nás.

Na přiloženém grafu vidíte růst počtu uživatelů v České republice za posledních pět let – v současnosti dosahuje až 6000 přístupů denně. Pokud *eduroam* dosud nevyužíváte, tady se dozvíte, jak na to:

<http://www.eduroam.cz/>

