

e-Infrastruktury – komplexní nástroje pro podporu vědy, výzkumu i výuky

Petr Holub, David Antoš, ÚVT MU

V rozvinutých zemích světa jsou již desítky let budovány různé infrastrukturní komponenty pro podporu vědy, výzkumu i výuky: od počítačových sítí, přes výpočetní systémy, až po různé multimediální služby. Tyto komponenty či služby však doposud často trpí tím, že byly vyvíjeny samostatně a nedostatečně zohledňovaly možnost svého dalšího využití. Kupříkladu sice máme síťové spoje s kapacitou 10 Gb/s i více, ale kolik služeb a aplikací dokáže tyto sítě využít? To sice neznamená, že by páteřní spoje s těmito kapacitami nebyly využívány v důsledku agregace provozu od mnoha uživatelů současně, avšak řada uživatelů by mohla potenciál těchto sítí zužít mnohem lépe, měla-li by k tomu nástroje. Obdobným příkladem je autentizace – doposud si řada institucí budovala své vlastní autentizační databáze vesměs založené na jménech a heslech, často i několik různých uvnitř jedné instituce. Důsledkem toho je, že uživatel pak má celou řadu identit a přihlašovacích údajů, které si v lepším případě pamatuje a v horším je má napsané na papírku. Až v poslední době se díky autentizačním federacím [1] začíná dařit tento přístup měnit.

V kontextu tohoto článku si e-infrastruktury na-
definujeme jako integrované služby pro vědu a výzkum založené na počítačových systémech a síťových infrastrukturách. Pod pojmem integrované budeme rozumět skutečnost, že jednotlivé komponenty jsou cíleně vyvíjeny tak, aby na sebe navazovaly, a počítají s tím, že budou skládány do větších celků poskytujících lepší funkcionality.

Oblast e-infrastruktur je bohužel semeništěm bezobsažných frází, které nemají praktického významu, což vede často potenciální uživatele k přesvědčení, že jim e-infrastruktura nenabídne žádný praktický užitek. Cílem tohoto článku je ukázat, jaké výhody mohou e-infrastruktury přinést uživatelům (je nezbytné, aby za vzletnými abstraktními formulacemi byla také technická

představa) a proč má pro vývojáře smysl investovat zvýšené úsilí do toho, aby jejich komponenty mohly být zapojeny do e-infrastruktur. A také chceme podtrhnout fakt, že bychom se o budování podobných infrastruktur měli aktivně snažit v rámci ČR.

Tento článek volně navazuje na článek o výpočetních a úložných infrastrukturách publikovaný v minulém čísle Zpravodaje [2], ovšem pokusíme se na problematiku podívat komplexněji a poněkud více vizionářsky, přičemž se budeme snažit neztratit ze zřetele technologickou rozumnost a realizovatelnost popisovaných systémů.

1 Současný stav – několik příkladů

Třebaže integrace různých komponent je běžným trendem pro osobní počítače až do takové úrovně, že uživatel jednotlivé složky systému nedokáže rozlišit, u větších a distribuovaných systémů se jedná o samostatné a často nedobře integrované komponenty. Jedna a tatáž funkcionality, která by mohla být implementována v samostatné a dobře odladěné vrstvě, bývá navíc duplikována v každé aplikaci zvlášť, pokaždé ovšem s jinými omezeními a chybami. Uvedme si několik příkladů situací, které se právě e-infrastruktury snaží vylepšit.

Přímo učebnicovou ukázkou mohou být datová úložiště ÚVT [2]. Většina uživatelů preferuje přístup protokolem CIFS (Samba), případně je nucena jej použít, protože k datům přistupuje např. z řídicího počítače mikroskopu, který prakticky nic jiného nezvládne. Implementace protokolu CIFS nejsou připraveny na integraci s externí správou identit (vyjma LDAPu, což ovšem pro nás nebylo použitelné), takže uživatelům je nutno vytvořit další uživatelské jméno a heslo. Do serveru sice lze podporu správy identit implementovat, bohužel by to vyžadovalo úpravy i na straně klienta, což ve většině případů znamená zásah do kódu OS Windows.

U většiny systémů vyvíjených na MU se už naštěstí integrace autentizačních mechanismů daří, takže používají například primární nebo sekundární heslo IS. Prakticky standardně je integrace dosažitelná u služeb přístupných přes webové rozhraní.

Mezi institucemi je situace daleko obtížnější. Pokud na Fakultě informatiky MU přednáší externista například z VUT, je třeba vydat mu zaměstnaneckou čipovou kartičku a zavést ho do našeho informačního systému, aby dokázal použít projekční a zvukovou techniku v posluchárnách dvorního traktu. Zatím se jedná o ojedinělé případy, ale při očekávané úrovni spolupráce v rámci projektu CEITEC¹ bude třeba systémy řízení přístupu MU a VUT buď spojit do federace, nebo vést všechny záznamy o přístupových bodech a oprávněných osobách v obou institucích.

Dalším příkladem může být síťový souborový systém NFSv4 exportovaný na výpočetní klastry MetaCentra². Občas se stane, že jeho výkon na chvíli prudce poklesne. Je mimořádně obtížné takovou situaci nasimulovat, takže téměř nelze odladit. Kdyby byl v době problému k dispozici rozumný monitoring, s rozhraním jak pro administrátory, tak i pro pokročilé uživatele, zachycující stav sítě a zároveň chování NFS serveru, dala by se alespoň izolovat komponenta, která problémem s výkonem způsobuje.

Konečně, nedávno jsme připravovali nástroje pro spolupráci na projektu, na kterém se podíleli lidé z MU, UK, ČVUT, ZČU, CESNETu a dalších institucí. Bylo potřeba připravit přístup do interní wiki projektu, CVS, virtuální místnosti ve dvou(!) videokonferenčních systémech, a všechny tyto položky, jakkoli jsou standardními nástroji pro naši práci, bylo nutno nastavit ručně, nemluvě o problémech s řízením přístupu, které jsme řešili pro každou komponentu zvlášť. Zejména jakékoli další změny, kupříkladu přidání dalšího uživatele do skupiny, pak znamená netriviální pravděpodobnost chyby ze strany administrátorů.

2 Základní služby

Zastavme se nyní u základních a všeobecně přijímaných služeb e-infrastruktur. Systémy pro ukládání a archivaci dat byly z pohledu služeb nabízených v současnosti zmíněny v minulém článku [2]. I k těmto komponentám se vrátíme z poněkud širšího pohledu.

¹<http://www.ceitec.cz/>

²<http://meta.cesnet.cz/>

2.1 Systémy pro ukládání, archivaci a sdílení dat

Ukládání dat pokrývá celou řadu služeb od ultrarychlých a kapacitně omezených služeb (např. disková pole vybudovaná z SSD disků), přes persistentní ukládání dat do on-line úložných systémů (např. velkokapacitní disková pole integrovaná do hierarchických úložných systémů), až po off-line archivaci (typicky páskové knihovny), která sice také může být automatizovaná, avšak vzhledem k mechanickým procesům (přesunům pásek, byť robotizovaným) řádově pomalejší v porovnání s předešlými.

Proč vlastně uvažovat o rozdělování úložných služeb do různých škatulek? Na rozdíl od výpočetních kapacit či sítí, které jsou svojí přirozeností služby transientní (tedy řekněme charakteru „udělej a zapomeň“), významná část dat vyžaduje dlouhodobé ukládání. Požadavky na kapacity datových úložišť mají tendenci neustále růst a jde o to, jak tyto požadavky naplnit finančně udržitelným způsobem. Aby si laskavý čtenář mohl učinit představu o potřebných kapacitách úložných systémů: již dnes máme na univerzitě skupiny, které produkují nebo ve velmi blízké budoucnosti plánují produkovat experimentální či simulační data v rádech půl petabyte za rok³ na jednu skupinu – přičemž s novými přístroji a zvyšováním dostupnosti výpočetních kapacit tyto objemy mohou ještě výrazným způsobem narůst. Rychlý přístup je ovšem třeba pouze k datům, jejichž zpracování probíhá; pro dlouhodobé skladování dat je samozřejmě nevhodné používat rychlá disková pole, která navíc svým provozem spotřebovávají značná množství energie. Páskové knihovny pro odkládání archivních dat jsou pro tyto účely mnohem lepším řešením – delší čas na získání souboru z archivu nevádí, páska je výrazně spolehlivější než disky nebo třeba DVD-R a s výjimkou občasné potřeby energeticky nenáročného převinutí nespotebovává elektřinu.

Úložné systémy samy o sobě představují pouze výchozí komponentu a úkolem e-infrastruktur je

³... což odpovídá více jak 110 000 DVD

jejich integrace s ostatními technologiemi takovým způsobem, aby byly pro uživatele co nejlépe použitelné. Z pohledu síťové infrastruktury se to týká nejen síťové kapacity, ale i protokolů pro přístup k datům. Protokoly běžně dostupné na koncových stanicích uživatelů jako např. Samba/CIFS jsou obvykle výkonnostně omezeny na nejvýše desítky megabitů za sekundu, a to ještě v lokálních sítích s minimální latencí. Na druhou stranu sice existují experimentální protokoly pro efektivní přenos dat na dnešních sítích s kapacitami v řádu jednotek a desítek gigabitů za sekundu, a to dokonce i v rozsáhlých sítích s vyššími latencemi, avšak tyto protokoly povětšinou neposkytují rozhraní systému souborů tak, aby s nimi běžné aplikace mohly přímo pracovat. Co víc, mohou vyžadovat nemalé zásahy na stanicích, na kterých jsou nasazeny, což může být v přímém rozporu s požadavky na tyto stanice, zejména pokud jejich primární rolí je řízení specializovaných přístrojů (mikroskopy, sekvenátory, NMR či RTG strukturní analýza) a spolehlivý sběr dat z nich.

2.2 Výpočetní prostředí

Požadavky na výpočetní prostředí se velmi liší mezi různými aplikacemi. Některé disciplíny si vystačí s uživatelskou vlastní výpočetní stanicí s případnou hardwarovou akcelerací (např. pomocí GPU⁴). Jiné vyžadují tisíce procesorů, ale problémy jsou efektivně paralelizovatelné nad distribuovanou pamětí (např. systémy PC klastrů). Z finančního pohledu bývají nejproblematičtější aplikace, které jsou paralelizovatelné pouze nad pamětí sdílenou a které vyžadují velké množství procesorů. Zde typicky nezbývá než nasadit na řešení velmi nákladné superpočítače.

Zatímco výstavba vlastních klastrových řešení nepředstavuje po hardwarové stránce zásadní problém, komplikace nastávají s vývojem, nasazením a udržováním softwarového prostředí a u rozsáhlejších klastrových systémů i s údržbou celého systému. Pro určité typy aplikačních klastrů (např. pro dávkové/neinteraktivní úlohy)

může být nejvýhodnější využít výsledků evropských gridových projektů jako je EGEE, na nichž se ostatně ČR významným způsobem podílela (např. participací CESNETu na vývoji middleware pro workload management). Flexibilnější podporu různým typům aplikací včetně aplikací interaktivních by měla poskytnout v blízké budoucnosti podpora virtualizovaná architektura vyvíjená v rámci projektu MetaCentrum [3].

V případě potřeby klasických superpočítačů bude potřeba pečlivá koordinace na národní úrovni, protože cena těchto systémů je obvykle několikanásobně vyšší v porovnání s klastrovými řešeními. Ve stadiu příprav je nyní v ČR projekt IT4Innovations⁵, jehož podání se plánuje mezi velkými projekty VaVpl.

Sdílené výpočetní systémy musí být interoperabilní se zbytkem infrastruktur – musí být možno používat data z úložných kapacit pro výpočty, systém musí být integrován s bezpečnostními mechanismy i monitoringem. Uživatelská podpora těchto systémů může využívat nástroje pro spolupráci (např. videokonference a sdílení pracovní plochy), aby mohla uživatelům pomoci diagnostikovat jejich problémy a nalézt řešení, aniž by za nimi musela cestovat.

Výpočetní prostředí však zdaleka nejsou pouze o hardwarových platformách, neméně významnou roli hraje výpočetní software používaný koncovými uživateli. Zatímco nákup softwarových licencí a starost o zprovoznění softwaru na koncových stanicích si obvykle zvládne vyřešit každý uživatel sám za sebe, bývá tatáž činnost v případě rozsáhlých výpočetních zdrojů podstatně komplikovanější a je vhodné ji řešit koordinovaným způsobem. U komerčních nástrojů je třeba řešit licence pro běh v paralelním prostředí, které lze navíc často mezi uživatelskými komunitami sdílet, protože ne vždy dokáže celá skupina využít maximální počet licencí a tudíž se licence mohou v čase sdílet⁶.

⁵<http://www.it4i.eu/>

⁶Zde je samozřejmě přirozenou obavou uživatelů, aby jim v době, kdy budou potřebovat maximální počet licencí, byly skutečně k dispozici. To lze efektivně řešit nákupem dostatečného množství licencí navíc a dále lze při koordinaci nákupů dosáhnout speciálních podmínek tak, aby se např. licence daly dokupovat či pronajímat zpětně v případě špičkového překročení potřebného počtu licencí, či

⁴General-Purpose computation on GPUs.
<http://www.gpgpu.org/>
<http://en.wikipedia.org/wiki/GPGPU>

2.3 Vizualizace a další zpracování výsledků

Výsledky získané při experimentech vyžadují obvykle další zpracování, aby se v nich vědecký pracovník zorientoval. V některých případech se jedná opět o využití výpočetních kapacit popsaných v předchozím případě, často se ale dá uvažovat o různých formách vizualizace (např. vizualizace velkých 2D dat pomocí SAGE [4] či paměťově náročný volumetrický rendering), případně dolování dat (data mining). Ačkoli samotné zpracování dat nemusí vyžadovat velké výpočetní kapacity, je třeba efektivně přistupovat k datovým zdrojům. Navíc počítače provádějící vizualizace musí být připojeny k samotným vizualizačním zařízením, a tudíž k nim musí být dovedena příslušná síťová kapacita – v případě zmíněných dělených displejů SAGE mohou datové toky dosahovat Gb/s.

2.4 Uživatelská rozhraní a prostředí pro spolupráci

Prostředí pro spolupráci nezahrnují zdaleka jen videokonferenční prostředí, tolikrát zmiňované i zde ve Zpravodaji. Dalšími komponentami jsou sdílení aplikací a pracovních ploch, sdílení dat a spolupráce nad nimi, či uživatelská rozhraní s podporou spolupráce – jednoduchými portály (wiki) počínaje, specifickými aplikacemi pro danou oblast konče (např. sdílené anotační nástroje pro popisování medicínského obrazového materiálu). Na to je ovšem třeba, aby tato prostředí byla napojena jak na bezpečnostní infrastrukturu, tak i na systémy sdílení dat. Samozřejmě je potřeba slušného síťového propojení, ovšem pro specifické aplikace se může jednat až o nároky dosti extrémní: při spolupráci více dělených displejů SAGE navíc mohou datové toky vzrůst až na desítky Gb/s, pro což je s dnešními technologiemi třeba opravdu specifické síťové podpory.

Další důležitou oblastí, která s těmito nástroji souvisí, je práce na dálku, tedy z domu a na cestách. Zde je potřeba poskytovat specifické nástroje, schopné pracovat nezávisle na fyzickém i síťovém umístění a přes síťové spoje s relativně nízkou propustností.

dosáhnout na neomezené multilicence.

Nástroje pro spolupráci a práci na dálku se často dají využít také pro vzdálené ovládání přístrojů a pro jejich sdílení. V některých případech je ovšem třeba implementovat do ovládacích programů lokální logiku, která je schopna se vyrovnat se skutečností, že došlo k přerušení práce s přístrojem v důsledku ztráty konektivity.

2.5 Integrace inteligentních budov

Poslední komponentou, kterou si v tomto článku představíme a kterou samu o sobě těžko zařadit přímo do služeb, jsou inteligentní budovy a jejich propojení s e-infrastrukturami. Řada moderních budov již pro své ovládání a monitorování využívá různých počítačových systémů [5, 6]. Integrace federalizovaných autentizačních mechanismů s ovládáním přístupových bodů (tj. vstupů do místností, laboratoří či provozů) by elegantně vyřešila problém s přístupem do prostor CEITECu jak pro pracovníky z MU, tak pro pracovníky z VUT.

Dalším přirozeným styčným bodem jsou požadavky na budovy ze strany e-infrastruktur: od stavební připravenosti pro různé komponenty (nejen výpočetní sály, ale také optimalizace místností vzhledem k využití audiovizuální techniky), přes rekuperaci tepla generovaného výpočetní a audiovizuální technikou, až po monitoring budov. Čidla monitorující teplotu, vlhkost a jiné parametry v různých místnostech, a to i v jejich různých částech, mohou být využity pro optimalizaci chlazení přístrojů, pro detekci anomálií, ale také pro protokolování laboratorních podmínek při provádění experimentů. Výsledky monitoringu je třeba alespoň po jistou dobu ukládat a poskytovat k nim zabezpečený přístup, čímž se opět vracíme k využití dalších e-infrastrukturních komponent.

3 Middleware

Jak již nejspíše pravidelný čtenář Zpravodaje znamenal (např. v [1]), termín middleware se používá pro označení komponent, které jsou logicky umístěny mezi infrastrukturou a koncovými aplikacemi. Z pohledu architektonického se jedná o vrstvu abstrakce podobnou různým abstrakcím v operačním systému (např. ovladače různých typů zařízení), která zajišťuje společné

funkce pro více aplikací, aby si je každá aplikace nemusela implementovat po svém.

3.1 AAA

Typickým zástupcem middleware jsou autentizační, autorizační a účtovací systémy (označované jako AAA - authentication, authorization, accounting). V rámci e-infrastruktur je ovšem potřeba nasazovat takové systémy, které jsou přirozeně distribuované a umožňují efektivní fungování napříč administrativně nezávislými institucemi a současně mohou být použity k autentizaci různých služeb. Zatímco například federativní bezpečnostní model [1] první část podmínky velmi dobře splňuje, k naplnění druhé části je potřeba ještě hodně vývojového úsilí - v současné době je pomocí federací relativně snadné řídit přístup k webovým aplikacím, ovšem například integrace s protokoly pro přenosy a sdílení dat je mnohem komplikovanější.

3.2 Monitorování

Dalším typickým zástupcem jsou monitorovací systémy. Jak již bylo naznačeno v kapitole o službách, prakticky žádná ze služeb se neobejde bez monitorování jak samotné služby, tak i všech ostatních komponent, na kterých závisí. Navíc monitorování rozsáhlých distribuovaných systémů - což e-infrastruktury bez pochyby jsou - přináší řadu stále se opakujících problémů s návrhem a implementací efektivního sběru, agregace, ukládání a presentace dat z monitoringu. Zde nacházíme také vysvětlení toho, proč by monitoring neměl být pouze ušit na míru každé aplikaci samostatně. Také presentace výsledků monitoringu musí podporovat všechny komponenty, a to jak na úrovni zobrazení výsledků pro administrátory (kteří na základě výsledků monitoringu mohou napravovat problémy, případně chování infrastruktur optimalizovat), tak i pro koncové uživatele (kteří jednak mohou sledovat výkon svých aplikací, ale také se v případě problémů efektivně rozhodovat, jak chybu napravit - zda tak mohou učinit sami, nebo zda je třeba kontaktovat někoho

dalšího a v tom případě také koho. Strukturování monitoringu s popisem závislostí komponent také zrychluje detekci příčin rozsáhlých výpadků (např. jeden síťový prvek, který způsobí nedostupnost celého segmentu sítě a všech služeb, které na daném segmentu běží).

4 Výhled do budoucna

Vize e-infrastruktur pro uživatele se skládá z celé palety služeb, jejichž nejběžnější zástupce jsme si ukázali v předchozím textu. Samozřejmě si ti, kteří se e-infrastruktury snaží budovat, těžko mohou dělat apriorní nárok na znalost toho, co dělají různí vědci a pedagogové ve svých oborech a co je pro jejich práci nejlepší. Podobně jako i u komerčních služeb, je třeba se snažit zanalyzovat jejich potřeby i dosavadní způsoby práce a se znalostí dostupných technologií navrhnout jejich co nejlepší podporu.

Přípravě na budování výše popsanych technologií se věnujeme v rámci přípravy středoevropského technologického institutu CEITEC a zkušenosti získané při jeho přípravě (a doufáme, že i realizaci) nabízíme celé univerzitě.

Čtenářům se však dle našeho názoru rozhodně vyplatí sledovat další vývoj e-infrastruktur v rámci ČR. Aktivity na jednotlivých institucích by měly poskytnout základní komponenty infrastruktur, které budou sloužit primárně výzkumným týmům na těchto institucích sídlících. Na národní úrovni bude současně třeba najít silné hráče, kteří budou schopni a ochotni věnovat se technologicky rozumné koordinaci mezi institucemi a budování centrálních služeb tak, aby byla do budoucna možná efektivní spolupráce mezi jednotlivými institucemi. Role sdružení CESNET v minulých letech v oblasti počítačových sítí (sít' národního výzkumu CESNET2), bezpečnostních otázek (federace identit, Eduroam, distribuované týmy pro řešení bezpečnostních incidentů), koordinace a integrace výpočetních zdrojů (projekt MetaCentrum) či v oblasti nástroje pro spolupráci (např. VoIP síť propojující telefonní ústředny [7]) ukazují, že tento model má potenciál fungovat.

Literatura

- [1] D. Kouřil, M. Kuba, M. Osovský, R. Peša, M. Procházka. Federace identit aneb spolčení totožností. *Zpravodaj ÚVT MU*. ISSN 1212-0901, 2007, roč. XVIII, č. 2, s. 1-7.
- [2] D. Antoš „Služby ÚVT pro vědu a výzkum.” *Zpravodaj ÚVT MU*. ISSN 1212-0901, 2008, roč. XIX, č. 2, s. 2-5.
- [3] L. Matyska, série článků *Virtualizace výpočetního prostředí*, *Zpravodaj ÚVT MU*. 2006 - XVII/2, 2007 - XVII/3, 2007 - XVII/5.
- [4] J. Matela, E. Hladká. „Obraz jako Brno.” *Zpravodaj ÚVT MU*. ISSN 1212-0901, 2009, roč. XIX, č. 3, s. 11-14.
- [5] F. Procházka, P. Glos. „Přístupová práva (nejen) v IS BAPS a jak na ně.” *Zpravodaj ÚVT MU*. ISSN 1212-0901, 2004, roč. XIV, č. 3, s. 11-14.
- [6] P. Glos. „Pasportizace budov a místností MU.” *Zpravodaj ÚVT MU*. ISSN 1212-0901, 2005, roč. XV, č. 4, s. 9-11.
- [7] E. Hladká, J. Růžička. “Telekonference v síti CESNET2.” *Zpravodaj ÚVT MU*. ISSN 1212-0901, 2008, roč. XIX, č. 1, s. 5-7. □