

Různé cesty do MetaCentra

MetaCentrum propojuje zdroje a uživatele

MetaCentrum poskytuje nástroje pro náročné výpočty a zpracování dat různorodým uživatelům napříč obory od jednotlivců (typicky student nebo výzkumný pracovník) až po velké vědeckovýzkumné týmy. Zdroje MetaCentra (výpočetní a úložné kapacity, software, speciální hardware, síť) realizující potřeby uživatelů jsou také nehomogenní, disponující různými vlastnostmi, schopnostmi i omezeními. Variabilita zdrojů je dána několika faktory. Především záměrem pokrýt různé typy požadavků a vyvážit poměr dostupných technologií podle zájmu uživatelů. Dalším cílem je „posouvání horizontů“, tj. poskytnout přístup na nové platformy, které si uživatelé mohou vyzkoušet a tím hledat cesty k rozvoji svého oboru (GP-GPU karty, SGI UV2 s velkou sdílenou pamětí, nyní Hadoop platforma). Výpočetní zdroje jsou doplněny odpovídající úložnou kapacitou, která je strukturována podle účelu a rozprostřena geograficky napříč ČR. Její členění odpovídá poměru výkon/kapacita, začíná u vysokorychlostních úložišť pro běžící výpočty (tzv. scratch) a končí integrací se službami Oddělení datových úložišť, které poskytují vysoké kapacity a možnost dlouhodobého uchování dat.

Základní vizí MetaCentra je mapovat co nejlepším způsobem potřeby uživatelů a charakter jejich aplikací na dostupné zdroje a jejich vlastnosti. Naplněním této vize je umožnit uživatelům přístup ke zdrojům různými způsoby, které představují různé úrovně přizpůsobení konkrétním potřebám.

MetaCentrum jako grid

Grid je základní metodou přístupu ke zdrojům MetaCentra. Zpracovávané úlohy jsou zpravidla dávkové, délka trvání úlohy v rozsahu hodin až týdnů. Uživatelé specifikují nároky na zdroje (hardwarové požadavky, softwarové licence) na jejichž základě jsou úlohy centrálně plánovány plánovačem, který je zároveň odpovědný za odpovídající přidělování a optimální využití dostupných kapacit. Prostředí gridu je jednotné, všechny uzly mají centrálně spravovaný jednotný operační systém a centrálně spravované software. Sem patří rozsáhlé vývojové prostředí (kompilátory, debugery, paralelní knihovny) a aplikační vybavení z širokého spektra oborů. V gridu podporujeme paralelní úlohy ve formě OpenMP a MPI i využití GP-GPU karet.

Gridové prostředí je flexibilní, umožňuje dedikovat část pravomocí a zodpovědností na uživatele a jejich skupiny, zejména pravomocí z oblasti přidělování zdrojů a instalace a konfigurace aplikačního SW. Za tímto účelem jsou zřizovány tzv. virtuální organizace (VO). Ty tvoří samostatnou administrativní doménu sdružující uživatele se stejnými zájmy a potřebami. Stále však platí, že uživatelé i celé skupiny se musí přizpůsobit požadavkům nezbytným pro sdílené a centrálně spravované prostředí (například v oblasti bezpečnosti, verzí OS a knihoven, základního rámce pro řízení přístupů ke zdrojům, apod.).

MetaCentrum jako cloud

Cloudové rozhraní k MetaCentru představuje možnost spouštět namísto výpočetních úloh celé (virtuální) stroje – výpočetní uzly. Uživatel tím získá naprostou volnost v konfiguraci a řízení prostředí pro běh výpočtů, stává se totiž vlastníkem a administrátorem souboru dedikovaných virtuálních výpočetních uzlů. Cloudové rozhraní je určeno pro ty uživatele, kterým přinese vyřešení jejich specifických potřeb či odstranění překážky, kterou pro ně představuje přizpůsobení se sdílenému distribuovanému výpočetnímu prostředí. Příkladem mohou být požadavky na odlišný operační systém (včetně MS Windows), vlastní prostředí pro správu úloh nebo požadavek mezinárodního projektu na instalaci totožného prostředí na všech spolupracujících institucích.

Cloudové prostředí je vhodné i v případech, kdy uživatelská skupina potřebuje provozovat permanentní služby spojené s jejími výpočty. Často se jedná o specializované uživatelské rozhraní pro správu úloh, správu a zpřístupnění dat, zpřístupnění služby dalším uživatelům, apod. V tomto případě je možné po odladění služby požádat o přenesení do režimu trvalých virtuálních strojů, kde je garantována vyšší dostupnost, permanentní jméno stroje, apod.

Podstatnou vlastností cloudového přístupu je to, že zavádí jako nový zdroj virtuální síť. Připojením virtuálních (výpočetních) uzlů do uživatelské sítě vzniká virtuální cluster, který může dynamicky měnit svoji kapacitu (počet a typ uzlů). Síť může být privátní (chráněná), napojená do vlastního prostředí uživatele (virtuální expanze lokálního clusteru) nebo může dokonce díky integraci s pokročilými službami sítě CESNET propojovat distribuovaný tým napříč ČR.

MetaCentrum jako platforma

Pro zpracování rozsáhlých nestrukturovaných dat se prosazuje metoda MapReduce výpočtů, kde jsou data rozdělována přes uzly clusteru a zpracovávána paralelně na jednotlivých uzlech clusteru. MetaCentrum poskytuje platformu Hadoop, ve které je možné tento typ úloh popsat a spustit, a která se stará o zabezpečení dat a jejich distribuci.

Z pohledu infrastruktury je Hadoop prostředí příkladem „platformy“, která je poskytována jako vyšší vrstva nad základem MetaCentra. Platforma využívá služby MetaCentra a transformuje je do prostředí vyladěného pro specifické využití, obvykle konkrétní organizaci dat a řízení jejich zpracování. V určité formě platformy již existují dlouhou dobu, například prostředí pro paralelní výpočty MPI či různé nástroje pro řízení zpracování dat - workflow. Dalším příkladem může být portál Galaxy, který je opět poskytován přímo MetaCentrem a pomocí kterého jsou uživatelé odstíněni od spouštění úloh na gridu pomocí příkazového řádku. Platformy jsou nejen přímo použitelné, ale zejména jsou prefabrikátem pro prostředí na míru - nabízíme spolupráci na rozšíření aktuálních platform o další funkcionalitu i pomoc při přípravě nových platform pro větší uživatelské komunity.

Různé cesty, společný cíl

Cílem MetaCentra je propojovat uživatele a jejich potřeby s dostupnými zdroji. Konceptně jde o různé cesty pro různé situace realizované uvnitř jednoho společného prostředí, nad univerzálním souborem zdrojů. Nástrojem je jednotná infrastruktura MetaCentra umožňující technické i organizační propojení různorodých zdrojů různých vlastníků a jejich transformaci do odpovídajících služeb. Přidělování zdrojů do různých „cest“ probíhá na základě potřeb uživatelů a zatížení infrastruktury a umožňuje optimální využití vložených investic.

Základní metodou přístupu k MetaCentru zůstává zadávání výpočetních úloh. Úlohy v high-throughput prostředí gridu jsou zpravidla dávkově spuštěné výpočetní aplikace, zatímco v prostředí cloudu se jedná o spuštění kompletního virtuálního stroje. Podle práce s daty uživatelé mohou zvolit prostředí sekvenčních nebo paralelních úloh v gridu nebo formu MapReduce výpočtů nad distribuovanými daty v prostředí Hadoop clusteru. Všechna prostředí poskytují přístup přes příkazovou řádku, pro specializované případy podporujeme i programové rozhraní a grafické rozhraní (Galaxy portál pro bioinformatiku, webové rozhraní pro cloud). Integrujeme zdroje institucí a vědeckých skupin i dalších e-infrastrukturních projektů, pomáháme budovat infrastrukturní zabezpečení konkrétních týmů pro výzkum a výuku či v rámci rozvojových projektů (například bioinformatickou infrastrukturu v rámci projektu ELIXIR). Různé cesty jsou součástí poslání MetaCentra, zejména ve formě spolupráce s uživateli na hledání cest vedoucích ke špičkovým výsledkům na jejich odborném poli.