

Platforma wsparcia badań naukowych infrastrukturą informatyczną jest obecnie w wielu dziedzinach warunkiem koniecznym do rozwoju oraz współpracy, szczególnie tej międzynarodowej.

Superkomputery dla polskiej nauki

Kazimierz Wiatr

We współczesnym świecie praca bardzo wielu zespołów naukowych wymaga efektywnego wsparcia informatycznego, w tym poprzez udostępnienie zasobów obliczeniowych o wielkiej mocy. Technologie gridowe stanowią odpowiedź na wyzwania związane z problematyką jednolitego i sprawnego dostępu do zasobów za pomocą przyjaznych dla użytkownika mechanizmów wizualizacji i komunikacji. Tak jak sieć WWW umożliwiała wymianę oraz dostęp do informacji rozproszonych w Internecie, tak gridy łączą w jedną całość rozproszone zasoby obliczeniowe i repozytoria danych naukowych. Z tych ogólnych przesłanek powstał projekt, który ma zapewnić realizację potrzeb, w tym także poprzez współpracę naukową w Europejskiej Przestrzeni Badawczej ERA.

Ogólnopolski projekt *Polska Infrastruktura Informatycznego Wspomagania Nauki w Europejskiej Przestrzeni Badawczej* – PL-Grid tworzy gridową infrastrukturę obliczeniową na potrzeby środowisk naukowych. W jego ramach są budowane zasoby obliczeniowe o wielkich mocach oraz opracowywane i wdrażane narzędzia, pozwalające projektować i uruchamiać aplikacje naukowe na potężnych zasobach obliczeniowych z wykorzystaniem rozproszonych źródeł danych.

Projekt jest realizowany przez Konsorcjum PL-Grid, utworzone w styczniu 2007 roku, w skład którego wchodzi: Akademickie Centrum Komputerowe CYFRONET AGH w Krakowie – inicjator i koordynator Projektu, Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW w Warszawie, Instytut Chemii Bioorganicznej PAN – Poznańskie Centrum Superkomputerowo-Sieciowe, Wrocławskie Centrum Sieciowo-Superkomputerowe, Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej w Gdańsku (na fot. kierownictwo projektu na hali komputerowej).

Realizacja projektu jest możliwa dzięki połączeniu centrów komputerów dużej mocy niezwykle wydajną magistralą światłowodową akademickiej sieci komputerowej PIONIER – Polski Internet Optyczny. Akademicka sieć Pionier zapewnia także łączność o przepustowości 2x10Gb/s pomiędzy ośrodkami MAN, co powoduje, że zasoby obliczeniowe zgromadzone w centrach KDM są obecnie dostępne z każdego ośrodka akademickiego i badawczego w Polsce.

Metakomputer dla polskiej nauki

Naukowcy i zespoły badawcze z różnych dziedzin naukowych w Polsce mogą wykorzystywać infrastrukturę PL-Grid do obliczeń i symulacji wielkiej skali. W szczególności nowe możliwości obejmują:

- dostęp do klastrów dużej mocy obliczeniowej (docelowo 215 TeraFlops) i dużych pamięci dyskowych (docelowo 2500 TeraBytes),
- pomoc w zrozumieniu zagadnień związanych z uruchamianiem aplikacji naukowych na rozległych zasobach obliczeniowych,
- wsparcie technologiczne i informatyczne przy projektowaniu własnych aplikacji naukowych i ich wdrażaniu na infrastrukturze PL-Grid,
- pomoc techniczną przy adaptacji stosowanych obecnie narzędzi do działania w nowych warunkach,
- zaawansowane narzędzia do organizacji eksperymentów obliczeniowych, które będą dostosowywane do indywidualnych potrzeb.

Oferta projektu obejmuje także infrastrukturę do nauki i testowania, szkolenia, warsztaty i konferencje, a także materiały promocyjne dostępne na stronach internetowych projektu.

Dostęp do e-infrastruktury PL-Grid umożliwi naukowcom powiększenie skali obliczeń prowadzonych w ramach badań naukowych, co byłoby niemożliwe

do osiągnięcia przy wykorzystaniu pojedynczych komputerów, i przyczyni się do rozszerzenia współpracy naukowej w wymiarze międzynarodowym. Dostęp ten jest darmowy dla naukowców i wszystkich prowadzących działalność naukową, związaną z uczelnią lub instytutem naukowym w Polsce.

W ramach projektu PL-Grid zostało już uruchomione Centrum Operacyjne (koordynowane przez ACK CYFRONET AGH), mające za zadanie udostępnianie powstałej infrastruktury, wspieranie na bieżąco jej użytkowników oraz dbanie o doskonalenie jej jakości działania także po zakończeniu realizacji projektu. Opracowany został regulamin korzystania z infrastruktury PL-Grid oraz procedury uzyskania konta obliczeniowego. Działa także helpdesk – nowoczesny system wsparcia dla osób korzystających z zasobów projektu – obejmujący obsługę techniczną i organizację bieżącego wsparcia użytkowników przez ekspertów (obsługa tzw. *trouble tickets*). Helpdesk (helpdesk@plgrid.pl) jest istotnym elementem planu wdrożenia naukowców do pracy w nowym środowisku obliczeniowym, dzięki któremu będzie możliwe zbieranie i udostępnianie wiedzy przydatnej użytkownikom w takiej formie, aby pomagała ona efektywnie wykorzystywać środowisko obliczeniowe.

Wirtualne laboratorium już działa!

Realizowana w ramach projektu e-infrastruktura jest bazą, na której będą konstruowane specjalizowane systemy do różnych dziedzin nauki, obejmujące usługi i narzędzia zorientowane na wykorzystywane typy aplikacji. Oprogramowanie e-infrastruktury będzie obejmować:

- narzędzia użytkownika, takie jak: portale, systemy zarządzania i monitorowania aplikacji oraz wizualizacji wyników;
- biblioteki programistyczne;



Od lewej: prof. Jacek Kitowski, dr Marian Bubak, prof. Kazimierz Wiatr, Karol Krawentek, prof. Michał Turała

- system wirtualnych organizacji: podsystemy certyfikatów i kont, rozliczanie wykorzystania zasobów, podsystem bezpieczeństwa;
- system zarządzania danymi: katalogi metadanych, zarządzanie replikami, transfer plików;
- system zarządzania zasobami: zarządzanie zadaniami, monitorowanie aplikacji, usług gridowych i infrastruktury, zarządzanie licencjami, zarządzanie lokalnymi zasobami.

Przykładem systemu zintegrowanego z infrastrukturą PL-Grid jest Wirtualne Laboratorium. To zbiór komponentów, które – używane razem – tworzą rozproszone środowisko współpracy. Różnorodne, rozproszone geograficznie grupy naukowców mogą wykorzystywać różnorodne funkcje laboratorium do planowania i przeprowadzania eksperymentów numerycznych, jak również dzielić się wynikami swoich prac badawczych. Oferowane narzędzia do zaawansowanej organizacji eksperymentów mogą być również dostosowane do indywidualnych potrzeb grup badawczych.

Powstawanie gridowej infrastruktury obliczeniowej jest zgodne z aktualnymi działaniami Komisji Europejskiej, silnie wpierającej powstawanie i łączenie gridów narodowych, budowanych w ramach projektów. Rozwiązania zastosowane przy budowie infrastruktury PL-Grid będą od początku zintegrowane z podob-

nyimi platformami w świecie, a szczególnie w Europie, czemu służy powołana w ostatnim czasie europejska organizacja EGI.

Polska w europejskim Komitecie EGI-EB

Europejska Inicjatywa Gridowa (EGI) powstała w odpowiedzi na potrzeby i wymagania formułowane przez europejskie środowiska naukowe, dotyczące utworzenia trwałej infrastruktury gridowej w Europie. Misją EGI jest zapewnienie długoterminowego utrzymania europejskiej e-infrastruktury oraz koordynacja więzi i współpracy między Narodowymi Inicjatywami Gridowymi (NGI) w utrzymaniu produkcyjnej infrastruktury gridowej dla naukowców reprezentujących różnorodne dyscypliny naukowe.

W skład EGI wchodzi Narodowe Inicjatywy Gridowe, reprezentujące kraje, które podpisały Memorandum of Understanding (MoU), dotyczące współpracy z EGI. PL-Grid aktywnie reprezentuje Polskę w tworzącej się strukturze EGI. ACK Cyfronet AGH podpisał MoU w imieniu Konsorcjum PL-Grid, które zaangażowane będzie w realizację istotnych zadań, w tym także w nadzorowanie poprawnego funkcjonowania europejskiej infrastruktury gridowej.

3 lutego 2010, podczas posiedzenia Rady EGI w Amsterdamie, zostało wybra-

nych 6 członków Executive Board (EB) EGI – wśród nich znalazł się prof. Michał Turała (patrz fot.) z ACK Cyfronet AGH, reprezentant PL-Grid w Radzie EGI. Członkowie EB będą osobami wytyczającymi kierunki rozwoju i koordynującymi działania EGI. Ważnym zadaniem EB będzie także proponowanie nowych projektów EGI przeznaczonych do finansowania przez Komisję Europejską.

Uproszczony dostęp naukowców do mocy obliczeniowych

Ważnym zadaniem projektu jest także zapewnienie wygodnego dostępu do zasobów komputerowych dla zespołów badawczych. Obecnie Centrum Operacyjne PL-Grid wprowadza procedury, które mają zapewnić, że infrastruktura obliczeniowa programu PL-Grid – „zasoby na żądanie” – umożliwi naukowcom zgłaszanie żądań dotyczących użycia maszyn do różnych celów.

Standardowe procedury dostępu do mocy obliczeniowych we wszystkich pięciu polskich centrach Komputerów Dużej Mocy Obliczeniowej są związane z wypełnieniem przez zainteresowanego stosownego formularza, który następnie powinien zostać zaakceptowany przez właściwego pełnomocnika w danej uczelni lub jednostce naukowej. Tak przynaj-

mniej jest w krakowskim Cyfronecie, a w pozostałych centrach KDM zapewne jest to też podobnie zorganizowane. To spora mitręga i nieco archaiczne podejście, ale chcąc zadowolić wszystkich potrzebujących przy stosunkowo średniej podaży mocy obliczeniowej wielkich komputerów, trzeba wybierać.

Budowane duże i nowoczesne zasoby obliczeniowe dla polskiej nauki w ramach projektu PL-Grid znacznie poprawią bilans potrzeb i możliwości w tym zakresie. Dlatego kierownictwo projektu zdecydowało się na znacznie uproszczony dostęp do zasobów obliczeniowych. Wypełniony formularz elektroniczny będzie automatycznie weryfikowany na podstawie bazy OPI.

Więcej informacji o Projekcie PL-Grid, oferowanych usługach, o tym, jak zostać użytkownikiem PL-Gridu, a także o planowanych szkoleniach, seminariach i innych wydarzeniach organizowanych dla wszystkich zainteresowanych można znaleźć na stronie www.plgrid.pl.

Gridy dziedzinowe

Kierownictwo i wykonawcy projektu PL-Grid mają świadomość niezwykle zróżnicowanych potrzeb obliczeniowych poszczególnych grup użytkowników. W pierwszym rzędzie te zróżnicowane potrzeby wyrażają się w zapotrzebowaniu na całkowicie różne oprogramowanie specjalistyczne, dzięki któremu możliwe jest przetwarzanie danych uzyskanych w eksperymentach, a także danych symulacyjnych. Znacznie rzadziej pewne grupy obliczeń wymagają nieco innego sprzętu komputerowego, różnicowanego konfiguracją węzła oraz architekturą połączeń węzłów i oprogra-

owaniem zarządzającym pracą systemu wieloprocesorowego. Centra KDM mają świadomość takiego zróżnicowania i budując infrastrukturę obliczeniową starają się zapewnić pewne zróżnicowanie sprzętu. Kompletowane są komputery SMP o wspólnych zasobach obliczeniowych, komputery klastrowe, stacje graficzne, tzw. tłuste węzły w klastrach o dużych zasobach pamięciowych i kilku-kilkunastu korach procesorowych w jednym węźle itd. Monitorując popyt naukowców na poszczególne zasoby, centra KDM starają się prowadzić w miarę zrównoważony rozwój zasobów w poszczególnych grupach, chociaż ze względów finansowych nie zawsze jest to możliwe.

Jednak mamy także świadomość, że nie wszystko można kupić. Istnieje spory obszar potrzeb tak unikatowych, że potrzebna jest własna aktywność wsparcia informatycznego. Dotyczy to przede wszystkim tworzenia niepowtarzalnego oprogramowania. Z tego względu łączenie grup użytkowników o podobnych potrzebach informatycznych jest na pewno bardzo efektywne i w paru obszarach już przyniosło oczekiwane efekty. Dlatego kolejne plany Konsorcjum PL-Grid to budowa wsparcia gridów dziedzinowych. Wsparcie to rozumiemy jako organizowanie dedykowanych zasobów sprzętowych i oprogramowania, ale także innych działań organizacyjnych i informatycznych dla zorganizowanych, nawet odległych od siebie terytorialnie, grup użytkowników o podobnych potrzebach.

Zadania te są celem kolejnego, przygotowanego przez Konsorcjum PL-Grid, projektu *Dziedzinowo zorientowane usługi i zasoby infrastruktury PL-Grid do wspomaganie Polskiej Nauki w Europiej-*

skiej Przestrzeni Badawczej – PLGrid Plus.

Warto wymienić najważniejsze gridy dziedzinowe: biologia i nanotechnologie, chemia kwantowa i fizyka molekularna, fizyka wysokich energii, astronomia i astrofizyka, akustyka, problemy life science, zdrowie, bioinformatyka, ekologia, energetyka, materiały, badania synchrotronowe. Lista ta oczywiście nie jest zamknięta, a wymienione dziedziny posłużą do budowy instalacji pilotowej. Już po opracowaniu wniosku zgłosiła się do nas grupa metalurgów ze specjalistycznymi potrzebami dostępu do zasobów obliczeniowych. Jesteśmy przekonani, że pojawiają się następne grupy, a zorganizowana współpraca użytkowników infrastruktury z jej operatorami przyniesie dobre owoce.

Platforma wsparcia badań naukowych infrastrukturą informatyczną jest obecnie w wielu dziedzinach warunkiem koniecznym do rozwoju oraz współpracy, szczególnie tej międzynarodowej. Mamy nadzieję i przekonanie, że pracujemy na rzecz budowy dobrych podstaw eNauki – ważnego elementu społeczeństwa informacyjnego i gospodarki opartej na wiedzy. Projektując rozwój infrastruktury nauki warto pamiętać jak ważnym jej składnikiem jest infrastruktura informatyczna, szczególnie w zakresie sieci komputerowych, mocy obliczeniowych oraz bibliotek wirtualnych i repozytoriów danych.

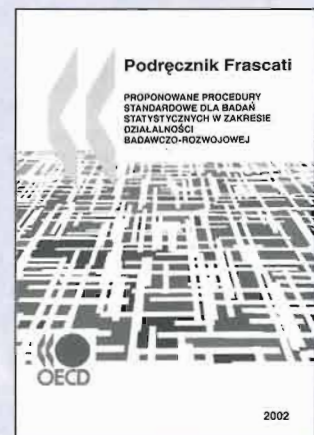
Prof. dr hab. inż. Kazimierz WIATR, profesor zwyczajny na Wydziale Elektrotechniki, Automatyki, Informatyki i Elektroniki AGH, dyrektor Akademickiego Centrum Komputerowego CYFRONET AGH, przewodniczący Rady Konsorcjum PIONIER – Polski Internet Optyczny.

Statystyka w B+R

Departament Strategii Ministerstwa Nauki i Szkolnictwa Wyższego opublikował polski przekład (dokonany przez Danutę Przepiórkowską) publikacji OECD *Podręcznik Frascati. Proponowane procedury standardowe do badań statystycznych w zakresie działalności badawczo-rozwojowej 2002*. Podręcznik jest nie tylko standardem do badań statystycznych sfery B+R w państwach członkowskich OECD. Dzięki inicjatywom podejmowanym przez OECD, UNESCO, Unię Europejską i inne organizacje regionalne stał się standardem w badaniach statystycznych sfery B+R. Co więcej, zawarte w nim definicje, taksono-

mie i rozróżnienia terminologiczne stały się podstawowym budulcem legislacji oraz programów i instrumentów polityki naukowej i innowacyjnej na całym świecie.

Zwracamy przy tym uwagę, że za zgodą OECD po tekście oryginalnym umieszczono dwie klasyfikacje, pochodzące z odrębnych dokumentów: *Zmieniająca klasyfikacja dziedzin nauki i techniki z: Revised Field of Science and Technology (FoS) Classification in the Frascati Manual OECD DSTI/EAS/NESTI(2006)19/FINAL* oraz *Cele społeczno-ekonomiczne z: Comparison between NABS 2007 and NABS 1992*, Eurostat, October 2008.



Podręcznik jest także dostępny w wersji elektronicznej: www.nauka.gov.pl/nauka/polityka-naukowa-panstwa/analizy-raporty-statystyki/artukul/podrecznik-frascati/.

jk