

Zofia MOSURSKA, Jacek KITOWSKI,
Robert PAJĄK, Kazimierz WIATR
AKADEMICKIE CENTRUM KOMPUTEROWE CYFRONET AGH W KRAKOWIE

ARTYKUŁ TECHNICZNY



PL-Grid – Polska Narodowa Inicjatywa Gridowa dla naukowców w Polsce

1. Czym jest komputerowy Grid?

Grid to infrastruktura informatyczna, umożliwiająca gromadzenie i współdzielenie geograficznie rozproszonych zasobów – komputerów, oprogramowania, danych – należących do różnych organizacji. Celem tworzenia Gridu jest umożliwienie tym organizacjom wzajemnej wymiany np. mocy obliczeniowej i innych zasobów, z zachowaniem lokalnych polityk dostępu i bezpieczeństwa. W ten sposób tworzone są tzw. *Wirtualne Organizacje* (ang. *Virtual Organizations, VO*).

Grid można traktować jako uwspólnione, rozproszone geograficznie (wirtualnie) zasoby obliczeniowe, które są: heterogeniczne sprzętowo i programowo, dostępne dynamicznie, zarządzane przez różne organizacje (z odmienną polityką zarządzania zasobami), połączone heterogeniczną siecią (różne warstwy, protokoły komunikacyjne, topologie).

Połączone siecią zasoby obliczeniowe wielu organizacji tworzących Grid, stanowią warstwę fizyczną (sprzętową) środowiska gridowego. Ponad warstwą sprzętową istnieje warstwa programowa (ang. *middleware* – usługi, narzędzia), która pozwala na udostępnianie i współdzielenie zasobów oraz umożliwia rozliczanie członków VO z użycia zasobów.

Zasoby gridowe są udostępniane w ramach Wirtualnych Organizacji i przeznaczone do rozwiązywania złożonych problemów dużej skali (obliczenia równoległe, rozproszone). Cała skomplikowana infrastruktura sprzętowo-programowa jest tak ukryta przed użytkownikiem, aby mógł on w prosty sposób zlecić swoje zadanie i otrzymać wyniki.

2. Oferta projektu PL-Grid dla polskich środowisk naukowych

W Polsce, prace nad utworzeniem polskiej Narodowej Infrastruktury Gridowej (ang. *National Grid Initiative - NGI*) podjęte zostały już w 2007 roku poprzez ustanowienie Konsorcjum PL-Grid, w skład którego weszło pięć polskich centrów obliczeniowych (rys. 1). Obecnie Konsorcjum realizuje projekt PL-Grid (numer umowy: POIG.02.03.00-00-007/08-00).



Rys. 1. Partnerzy Konsorcjum PL-Grid

PL-Grid – Polska Infrastruktura Informatycznego Wspomagania Nauki w Europejskiej Przestrzeni Badawczej jest projektem współfinansowanym ze środków Europejskiego Funduszu Rozwo-

ju Regionalnego w ramach Programu Operacyjnego Innowacyjna Gospodarka w latach 2009-2011. Dzięki pozyskanym środkom finansowym, organizacje partnerskie PL-Grid mogą znacznie rozbudować swoje zasoby obliczeniowe i tworzyć struktury Polskiego Gridu Narodowego.

Celem Projektu jest dostarczanie polskiej społeczności naukowej usług informatycznych opartych na gridowych klastrach komputerowych, służących e-Nauce w różnych dziedzinach. Rozwiązania zastosowane przy budowaniu e-infrastruktury są od początku zintegrowane z podobnymi platformami w świecie, a w szczególności – z systemami gridowymi powstałymi w Europie w wyniku międzynarodowych projektów EGEE i DEISA. Utworzona gridowa infrastruktura obliczeniowa dla potrzeb polskich naukowców umożliwi w przyszłości szeroką współpracę międzynarodową. W ramach Projektu opracowywane i wdrażane są narzędzia, które pozwalają na projektowanie i uruchamianie aplikacji naukowych na potężnych zasobach obliczeniowych z wykorzystaniem rozproszonych źródeł danych.

Dzięki funduszom projektu PL-Grid zespoły badawcze reprezentujące różne dziedziny naukowe w Polsce, zainteresowane wykorzystaniem infrastruktury PL-Grid do obliczeń i symulacji wielkiej skali, mogą korzystać bezpłatnie z oferty Projektu, która obejmuje:

- dostęp do klastrów dużej mocy obliczeniowej i dużych pamięci dyskowych PL-Grid; do końca 2010 r. w skali kraju zostanie zainstalowany sprzęt obliczeniowy o mocy ok. 180 Tflops i zasoby pamięciowe o pojemności ok. 1900 TBytes (docelowo będzie 215 Tflops i 2500 TBytes),
- zaawansowane pakiety specjalistyczne, które umożliwiają przeprowadzanie obliczeń naukowych z dziedziny biologii, chemii kwantowej, fizyki, obliczeń numerycznych i symulacji,
- narzędzia i usługi, np. do organizacji eksperymentów obliczeniowych i wizualizacji wyników użytkowników, do zarządzania zasobami gridowymi i użytkownikami (zgrupowanymi w organizacje wirtualne),
- pomoc w zrozumieniu zagadnień związanych z uruchamianiem aplikacji naukowych na rozległych zasobach obliczeniowych (poprzez szkolenia i system helpdesk).

3. Innowacyjność w projekcie PL-Grid

Symulacje komputerowe, praca na wielkich zasobach danych, przy pomocy zaawansowanych technologii informatycznych i potrzeba ciągłej współpracy pomiędzy zespołami badawczymi, stają się nieodzownymi elementami procesu wypracowywania innowacji technologicznych w różnych dziedzinach nauki, a szczególnie w: chemii, biochemii, fizyce wysokich energii, astrofizyce, naukach biologicznych i medycznych, naukach o Ziemi, nanotechnologiach. Aby uczestniczyć w takich pracach, naukowcy potrzebują odpowiedniej infrastruktury informatycznej, która umożliwi integrację danych eksperymentalnych i wyników zaawansowanych symulacji komputerowych, prowadzonych zazwyczaj przez geograficznie rozproszone zespoły.

Utworzenie narodowej infrastruktury gridowej jest przedsięwzięciem o bardzo wysokim potencjale innowacyjności, wymienianym w dokumentach rządowych ustalających kierunki rozwoju polskiej infrastruktury informatycznej dla nauki oraz o dużym znaczeniu dla techniki i gospodarki.

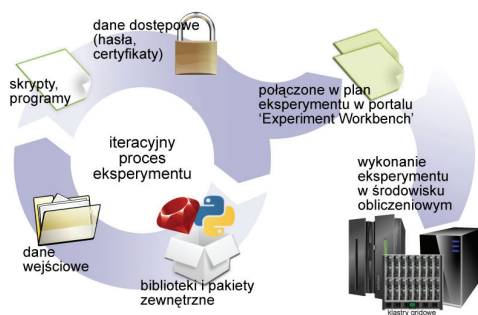
Aby zrealizować jeden z celów Programu Operacyjnego Innowacyjna Gospodarka – „Wzrost konkurencyjności polskiej nauki” – sektor nauki winien być wzmocniony poprzez zapewnienie odpowiedniej bazy badawczej. Taką bazą jest także infrastruktura gridowa, na której będą konstruowane specjalizowane systemy,

specyficzne dla różnych dziedzin nauki, obejmujące usługi i narzędzia zorientowane na wykorzystywane typy aplikacji.

Ustanowiony system gridowy musi umożliwić efektywne i innowacyjne wykorzystanie nowo utworzonej infrastruktury sprzętowej. Aby to było możliwe, zostały opracowane i wdrożone zasady efektywnego użycia infrastruktury przez użytkowników zgrupowanych w wirtualne organizacje oraz zestaw usług do zarządzania zasobami gridowymi na różnych poziomach.

Jednym z przykładów innowacyjnych narzędzi rozwijanych w ramach projektu PL-Grid, łatwo dostępnym poprzez przeglądarkę internetową, jest *GridSpace2* – nowatorska platforma Wirtualnego Laboratorium, która umożliwia naukowcom przeprowadzanie wirtualnych eksperymentów na zasobach gridowych (<https://wl.plgrid.pl/>).

W ramach Wirtualnego Laboratorium PL-Grid poszczególne aplikacje są określane mianem eksperymentów, co oznacza, że wykraczają one poza proste i powtarzalne uruchamianie zainstalowanych programów. Co więcej, mogą one wiązać się z eksperymentalnym lub eksploracyjnym programowaniem, kiedy aplikacja składa się z wielu kroków, które nie są z góry znane i często są wybierane *ad-hoc*, w oparciu o wyniki poprzednich kroków. Przykładowy cykl tworzenia eksperymentu wykonywanego w środowisku gridowym przedstawiony jest na rys. 2.



Rys. 2. Cykl życia eksperymentu

4. Oprogramowanie specjalistyczne

Na zasobach PL-Grid jest instalowane różnorodne oprogramowanie specjalistyczne, dotychczas dostępne tylko na komputerach dużej mocy o architekturze SMP w kilku polskich centrach obliczeniowych. Wkrótce planujemy udostępnić następujące pakiety specjalistyczne z różnych dziedzin nauki:

- biologia: AutoDock, BLAST, ClustalW2, CPMD, Gromacs, NAMD;
 - chemia kwantowa: ADF, CFOUR, Dalton, GAMESS, Gaussian, Molcas, Molpro, MOPAC, NWChem, OpenBabel, Siesta, TURBOMOLE;
 - fizyka: ANSYS Fluent, Meep;
 - obliczenia numeryczne i symulacje: Mathematica, MATLAB;
 - inne: Blender, POV-Ray.
- Obecnie, następujące oprogramowanie zostało już udostępnione w niektórych lokalnych ośrodkach Partnerów Projektu:
- w ACK Cyfronet AGH: ADF, ANSYS Fluent, Blender, CFOUR, CPMD, Dalton, GAMESS, Gaussian, Gromacs, Mathematica, MATLAB, NAMD, NWChem, POV-Ray, TURBOMOLE;
 - w PCSS: CFOUR, GAMESS, Gaussian, NAMD;
 - w CI TASK: ANSYS Fluent, Blender, CFOUR, GAMESS, Gaussian;
 - w WCSS: ANSYS Fluent, CPMD, Dalton, GAMESS, Gaussian, Gromacs, NWChem, TURBOMOLE.

5. Jak zostać użytkownikiem infrastruktury PL-Grid?

Użytkownikiem infrastruktury PL-Grid może zostać każda osoba prowadząca działalność naukową, związaną z jednostką na-

ukową w rozumieniu ustawy z dnia 8.10.2004 r. o zasadach finansowania nauki, jak również współpracownicy osoby prowadzącej działalność naukową (doktoranci, studenci, współpracownicy zagraniczni).

Osoby, które chciałyby korzystać z zasobów obliczeniowych projektu PL-Grid, powinny zapoznać się z Zasadami Funkcjonowania Infrastruktury PL-Grid oraz Regulaminem Użytkownika Infrastruktury PL-Grid. Dokumenty te są dostępne na stronie <http://www.plgrid.pl/projekt/dokumenty>.

Aby uzyskać dostęp do infrastruktury Projektu obejmującej usługi dostępne, zasoby obliczeniowe oraz specjalistyczne oprogramowanie, należy zarejestrować się w Portalu Użytkowników Infrastruktury PL-Grid wypełniając formularz on-line. Portal ten jest dostępny pod adresem URL: <http://konto.plgrid.pl>. W procesie rejestracji użytkownika dane weryfikowane są przez Centrum Operacyjne PL-Grid w oparciu o dane z bazy danych "Ludzie nauki" prowadzonej przez Ośrodek Przetwarzania Informacji (OPI).

Po zarejestrowaniu się w Portalu, użytkownik uzyskuje dostęp do zasobów obliczeniowych poprzez platformę gLite lub UNICORE, do danych na temat wykorzystywanych przez siebie zasobów oraz do systemu pomocy *helpdesk*. W Portalu użytkownik może także zarejestrować certyfikat, jeśli go już posiada. W przeciwnym przypadku może przejść przez procedurę uzyskiwania osobistego certyfikatu gridowego lub natychmiast uzyskać certyfikat ograniczony do zasobów w Polsce.

Usługi obliczeniowe w ramach infrastruktury PL-Grid są świadczone nieodpłatnie, ale użytkownicy są zobligowani do załączania informacji o jej wykorzystaniu w publikacjach powstałych w oparciu o te obliczenia.

6. Wsparcie użytkownika

W projekcie PL-Grid został uruchomiony helpdesk – nowoczesny system pomocy dla osób korzystających z zasobów Projektu – obejmujący obsługę techniczną i organizację bieżącego wsparcia użytkowników przez ekspertów (obsługa tzw. *trouble tickets*). Problemy i zapytania techniczne dotyczące działania infrastruktury lub aplikacji można zgłaszać na adres e-mail: helpdesk@plgrid.pl, ale użytkownicy posiadający konto w infrastrukturze PL-Grid mogą również korzystać z systemu helpdesk online, poprzez adres URL: <https://helpdesk.plgrid.pl>.

W ramach Projektu organizowane są regularne szkolenia w siedzibach Partnerów Projektu, a także szkolenia na zamówienie – dla dowolnej liczby uczestników i w uzgodnionym miejscu i czasie.

Ponadto, opracowano kilka podstawowych kursów na platformie e-learningowej Blackboard, pomocnych dla osób, które chciałyby skorzystać z zasobów i usług infrastruktury PL-Grid. Obecnie dostępne są już, dla użytkowników zarejestrowanych w Portalu Użytkownika PL-Grid, następujące kursy:

- ABC użytkownika PL-Grid
- Obsługa systemu kolejkowego PBS
- Wstęp do oprogramowania pośredniczącego gLite
- UNICORE
- Kurs dla użytkowników eksperymentu LHCB

Konto uprawniające do korzystania z tych kursów uzyskuje się po zarejestrowaniu się w Portalu Użytkownika Infrastruktury PL-Grid.

Więcej informacji o Projekcie, 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 oficjalnej stronie Projektu: <http://www.plgrid.pl/>.