

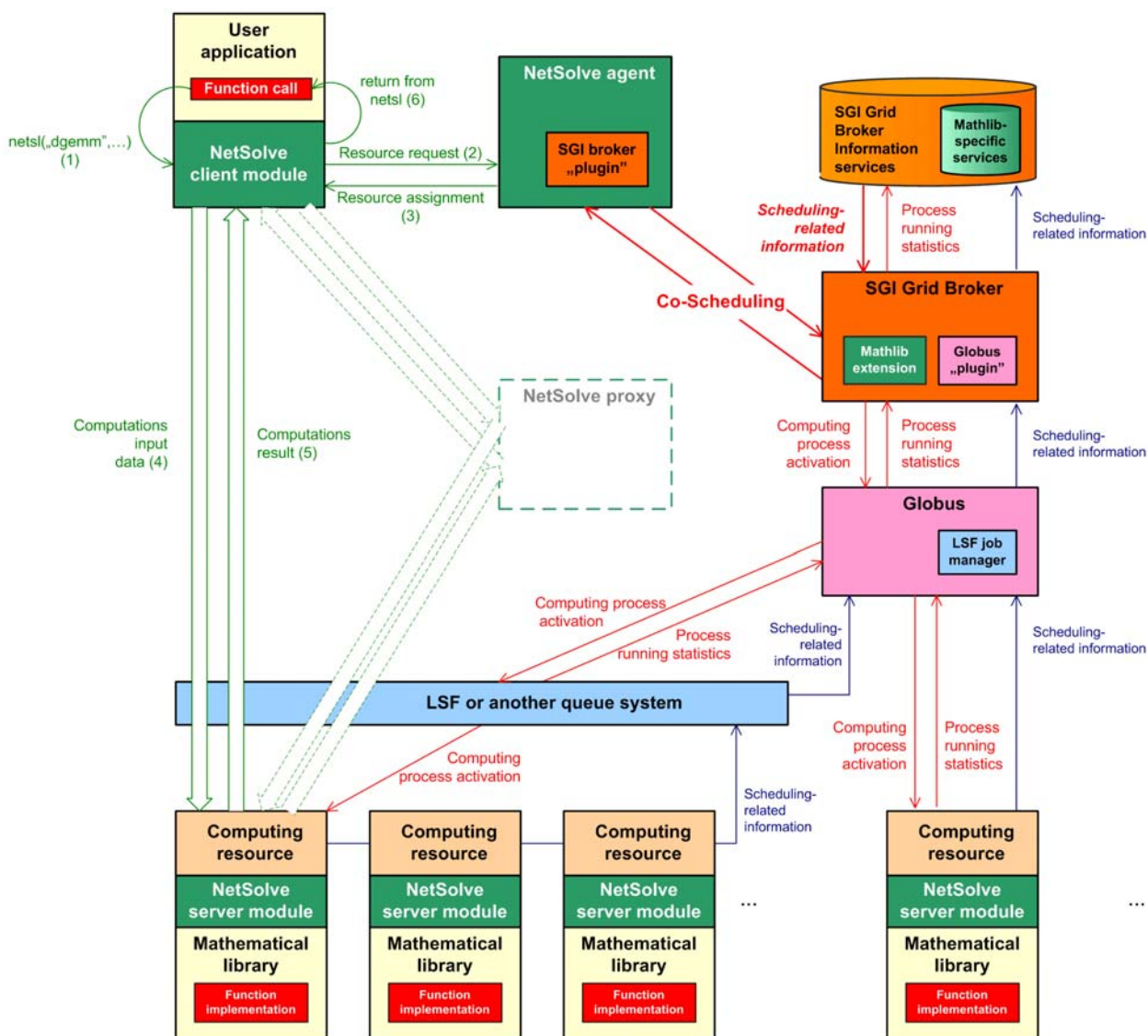
Architektura systemu udostępniania bibliotek matematycznych w środowisku klastra SGI Grid

Wynik dyskusji i ustaleń w ramach spotkań w dniach: 30 lipca i 4 września 2003 r.

Uczestnicy spotkań:

Maciej Brzeźniak (PCSS, WP 2.1. projektu)
Krzysztof Kurowski, Ariel Oleksiak (PCSS, WP 5 projektu)

Na spotkaniu projektu celowego w lipcu 2003 r. ustalono ostatecznie architekturę klastra SGI Grid. Uzgodniono, że klastr SGI będzie oparty o klastry LSF natomiast dostęp do poszczególnych klastrów LSF będzie odbywać się za pomocą systemu Globus. W oparciu o te ustalenia opracowano metodę włączenia systemu udostępniania bibliotek matematycznych do klastra SGI Grid. Poniższy rysunek przedstawia planowaną architekturę systemu udostępniania bibliotek matematycznych.



Rysunek 1 Architektura systemu udostępniania bibliotek matematycznych w środowisku klastra SGI

Na rysunku kolorem zielonym oznaczono elementy systemu NetSolve, na którym bazować będzie system udostępniania bibliotek matematycznych. Przepływy informacji, które nie zostały zmienione w stosunku do oryginalnego systemu NetSolve zaznaczone są na rysunku również kolorem zielonym. Przepływy informacji, które są zmienione w stosunku do oryginalnego systemu i odbywają się w momencie uruchamiania obliczeń systemu udostępniania bibliotek matematycznych zostały oznaczone kolorem czerwonym. Przepływy informacji, które odbywają się cyklicznie -

niezależnie od faktu, czy w danej chwili uruchamiane są obliczenia w systemie udostępniania bibliotek matematycznych - zaznaczone są kolorem niebieskim.

Moduł klienta NetSolve

Dla zachowania zgodności z klientami oryginalnego systemu NetSolve, moduł klienta nie będzie podlegał dodatkowym modyfikacjom (poza mechanizmami zwiększenia niezawodności i wydajności opracowywanymi przez ośrodek TASK). Podobnie, komunikacja pomiędzy klientem i agentem a także klientem i modułem serwerowym NetSolve pozostanie niezmienną.

Może zaistnieć potrzeba realizacji komunikacji klient-serwer NetSolve poprzez moduł Proxy NetSolve.

W przypadku braku możliwości bezpośredniego komunikowania się klienta systemu i modułu serwerowego (np. z powodu systemów firewall czy niemożności otwierania przejść w systemach firewall) komunikacja odbywać się będzie przez proxy systemu NetSolve (jest to opcja dostępna w systemie NetSolve). Komunikacja za pośrednictwem Proxy NetSolve wymagana będzie również w przypadku realizacji opcji przekazywania danych wejściowych i wyników obliczeń w postaci plików (szczegóły w dalszej części dokumentu). W takim wypadku konieczna będzie również modyfikacja modułu Proxy systemu NetSolve (by mógł on przekazywać dane wejściowe i wyniki do Brokera SGI).

Współpraca Agent - Broker SGI

Zmianom podlegnie moduł agenta NetSolve. Zostanie wyposażony we "wtyczkę" do Brokera klastra SGI (kolor pomarańczowy). Umożliwi to koordynację procesu szeregowania a także wymianę informacji, związanych z wydajnością obliczeń i stanem systemów obliczeniowych.

Szeregowanie:

Rozważane są dwie opcje współpracy Agent a Brokera SGI:

- **Agent szereguje wykonanie zadania** na podstawie dostępnych dla niego informacji (pobiera je m.in. z systemu informacyjnego Brokera); wybiera zasób i prosi Brokera o uruchomienie zadania na wybranym zasobie
- Agent "przekazuje" żądanie klienta o wykonanie obliczeń do Brokera SGI; żądanie zawiera opis zadania w odpowiednim formacie; **wybór zasobu odbywa się w brokerze**; broker musi realizować odpowiednią, specyficzną dla bibliotek matematycznych funkcjonalność potrzebną dla wybrania zasobu

Wszystkie funkcje Brokera klastra SGI specyficzne dla systemu udostępniania bibliotek matematycznych będzie realizować dodatkowy, wyróżniony element Brokera (moduł rozszerzający Brokera) lub będą one realizowane w Agencji.

Uruchamianie zadań:

Uruchamianie zadań na zasobach obliczeniowych (komputerach na których zainstalowane są biblioteki matematyczne) będzie się odbywać za pomocą systemu Globus. Dla zasobów, które pracują pod kontrolą systemów kolejkowych wykorzystywane będą tzw. job-manager'y systemu Globus dla poszczególnych systemów kolejkowych (np. dla systemu LSF). Dostęp do zasobów, które nie są kontrolowane przez system kolejkowy będzie odbywał się bezpośrednio z systemu Globus (z wykorzystaniem standardowego job manager'a).

Z kolei usługi systemu Globus będą eksploatowane przez Brokera SGI lub przez Agent a za pośrednictwem Brokera SGI (w zależności od przyjętego ostatecznie trybu szeregowania - w Agencji czy w Brokerze).

Systemy informacyjne

Serwisy informacyjne Brokera zostaną rozszerzone o informacje specyficzne, potrzebne do szeregowania wykonania funkcji matematycznych. Obejmą one m.in. historyczne informacje o poprzednich wywołaniach: czasie, zużytych zasobach, wartości poszczególnych argumentów i parametrów zadania itd. Ponieważ dane te są specyficzne, niepotrzebne do szeregowania innych niż wywołania funkcji matematycznych zadań w klastrze SGI, rozważane jest umieszczenie ich w oddzielnym serwisie informacyjnym. Mogłoby to być również korzystne dla wydajności operacji na tych danych.

Informacje te będą wykorzystywane przez Agent a lub Brokera w zależności od tego, gdzie ostatecznie odbywać się będzie szeregowanie. W przypadku, gdy szeregowanie odbywać się będzie w Agencji systemu udostępniania, poza danymi specyficznymi, będzie on musiał mieć dostęp również do danych ogólnych, m.in. dotyczących stanu/obciążenia zasobów zgromadzonych w serwisach informacyjnych Brokera.

Dostęp Agent a do danych w serwisach informacyjnych Brokera możliwy będzie na jeden z dwóch sposobów:

- bezpośrednio (w tej chwili preferowana),
- za pośrednictwem usług brokera dających dostęp do serwisów informacyjnych.

Pierwsze podejście pozwala na działanie Agent a niezależnie od dostępności i wydajności Brokera SGI. Wiąże się jednak nakładem pracy na oprogramowanie funkcji dostępu do systemu informacyjnego. Ostateczna decyzja zostanie podjęta w dalszym etapie prac po oszacowaniu korzyści dla wydajności i dokładnej ocenie potrzebnego nakładu prac.

Przebieg procesu uruchamiania

Z punktu widzenia klienta systemu udostępniania przebieg procesu uruchamiania będzie wyglądał podobnie do procesu uruchamiania w oryginalnym systemie NetSolve. Będzie obejmował następujące kroki:

1. Aplikacja klienta wywołuje funkcję modułu klienta systemu udostępniania, zlecając mu wyliczenie danej funkcji matematycznej.
2. Moduł klienta żąda od Agenta przydziału zasobu do obliczeń.
3. Wynik procesu szeregowania jest przesyłany do modułu klienta systemu udostępniania.
4. Moduł klienta NetSolve przesyła do aktywowanego modułu obliczeniowego (serwerowego) NetSolve dane wejściowe dla obliczeń.
5. Po dokonaniu obliczeń moduł serwerowy NetSolve przesyła wyniki do modułu klienta NetSolve.
6. Z kolei następuje wyjście z funkcji klienta NetSolve i przekazanie wyników obliczeń aplikacji.

Wewnątrz systemu udostępniania (z punktu widzenia Agenta, modułu serwerowego systemu oraz Brokera SGI) procesu uruchamiania obliczeń będzie miał następujący przebieg:

1. Agent w porozumieniu z Brokerem dokonuje uszeregowania żądania.
2. Na wybranym zasobie obliczeniowym uruchamiany jest serwerowy moduł obliczeniowy. Odbywa się to za pośrednictwem usług systemu Globus (patrz wyżej).
3. Aktywowany moduł obliczeniowy wykonuje obliczenia na podstawie danych dostarczonych przez moduł klienta.
4. Informacje dotyczące wywołania funkcji matematycznej przekazywane są przez moduł serwerowy do Brokera klastra SGI i umieszczane w systemie informacyjnym.

Przesyłanie danych wejściowych i wyników obliczeń.

Rozważana jest modyfikacja sposobu przekazywania danych wejściowych od modułu klienta do modułu serwerowego, który wykonuje obliczenia. W zmodyfikowanym podejściu argumenty dla obliczeń przesyłane byłyby za pośrednictwem Agenta systemu udostępniania (lub przez Proxy NetSolve) do Brokera SGI. Broker SGI kierowałby te dane do odpowiedniego pliku umieszczonego na maszynie obliczeniowej. Podobnie, wyniki obliczeń umieszczane by były w odpowiednim pliku wynikowym a następnie przez Broker SGI i Agenta lub Proxy NetSolve przekazywane do modułu klienta systemu udostępniania. Taka technika miałaby następujące zalety:

- jednolitość techniki uruchamiania zadań w klastrze SGI,
- pełna informacja brokera o zadaniu (np. rozmiar danych).

Jej stosowanie może się jednak wiązać ze spadkiem efektywności (dane przepływają przez Agenta lub Proxy NetSolve oraz Broker SGI a także zapisywane i odczytywane są z systemu plików maszyny obliczeniowej). Dlatego podjęcie ostatecznej decyzji o jej stosowaniu będzie poprzedzone dokładną analizą i testami wydajnościowymi.

Dane zbierane okresowo

Okresowo Broker SGI zbiera informacje o stanie systemów komputerowych pracujących w poszczególnych klastrach LSF, stanie klastrów a także maszyn, które nie są kontrolowane przez systemy kolejkowe itp. Zbierane informacje obejmują dane niezbędne dla działania Brokera klastra SGI a także informacje specyficzne, które są niezbędne dla prawidłowego szeregowania obliczeń w systemie udostępniania bibliotek matematycznych.

Informacje ogólne (po spotkaniu 2003.09.04):

Dostęp do zasobów obliczeniowych:

- Agent nie musi mieć bezpośredniego dostępu do zasobów, LSF czy systemu Globus, zamiast tego może korzystać z usług brokera i GRMS (nazwa w projekcie GridLAB); GRMS ma "przejsiówki" do Globusa, a przez job managery Globusa do systemów kolejkowych (LSF i inne).
- Uruchamianie obliczeń za pośrednictwem Brokera SGI, będzie możliwe po opracowaniu opisów zasobów, czyli wszystkich informacji, które są potrzebne do uruchomienia danego zasobu obliczeniowego przez Broker SGI (specyfikacja pliku wykonywalnego, bibliotek itd.). Trzeba również opracować metodę opisu konkretnych żądań, tak by Broker mógł je obsługiwać. Szczegóły dopracowane zostaną w terminie późniejszym.
- W klastrze SGI zdecydowano się wykorzystać Globus w wersji 2.2.4.

Problemy/ pytania otwarte:

- autoryzacja, technika delegacji uprawnień - rozważane są dwie opcje:
 1. klient w stosunku do Agenta, a następnie Agent występuje w imieniu jednego, tego samego użytkownika na wszystkich zasobach
 2. agent ma wiele certyfikatów dla różnych użytkowników, zarządza odwzorowania autoryzacji klientów NetSolve na autoryzację użytkowników klastra SGIuzgodnić z zadaniem dot. bezpieczeństwa (WP 7) a także z zadaniem WP 6 (zarządzanie użytkownikami)