

IMPLEMENTASI DAN EVALUASI PENERAPAN GLOBUS TOOLKIT UNTUK APLIKASI GRID COMPUTING

Widodo Budiharto

Computer Science Department, School of Computer Science, Binus University
Jln. K. H. Syahdan No. 9 Palmerah Jakarta Barat 11480
widodo@widodo.com

ABSTRACT

Grid computing is a distributed computing technology that utilizes resources connected through a free computer network, yet coordinated with a specific mechanism. The development of grid computing infrastructure is not easy because it takes skill and experience in the installation and configuration of both Linux-based and open source program. In this study, the author built a grid computing infrastructure based on Debian 4, and used Globus Toolkit 4.1.2 on three computers. The WSRF technology was tried to run as an indication that the grid infrastructure has been successfully built. Based on some experiments in this study, grid computing can run well on the three computers with a user interface of web-based grid system using the UCLA Grid Portal. Overall the system runs fine, but it requires high experiences and comprehensions upon the Linux operating systems as well as computer networks in the installation process.

Keywords: *grid computing, Globus Toolkit, WSRF, UCLA Grid Portal*

ABSTRAK

Grid computing adalah teknologi komputasi terdistribusi yang memanfaatkan sumber daya yang terhubung melalui jaringan komputer secara bebas tapi terkoordinasi dengan mekanisme tertentu. Pembangunan infrastruktur Grid Computing tidaklah mudah karena membutuhkan kemampuan dan pengalaman di dalam instalasi dan konfigurasi program berbasis Linux dan Open source. Pada penelitian ini, penulis membangun infrastruktur Grid Computing berbasis Debian 4, dan menggunakan Globus Toolkit 4.1.2 pada tiga buah komputer. Teknologi WSRF dicoba untuk dijalankan sebagai indikasi bahwa infrastuktur Grid sudah sukses dibangun. Berdasarkan percobaan, grid computing dapat dijalankan dengan baik pada tiga komputer tersebut dengan user interface sistem grid berbasis web menggunakan UCLA Grid Portal. Secara keseluruhan system berjalan dengan baik, tetapi membutuhkan pengalaman dan pemahaman yang tinggi dalam bidang sistem operasi Linux dan jaringan komputer di dalam proses instalasinya.

Kata kunci: *grid computing, Globus Toolkit, WSRF, UCLA Grid Portal*

PENDAHULUAN

Pesatnya kebutuhan akan sumberdaya komputer skala besar, telah melahirkan berbagai macam teknologi komputer terkini, antara lain teknologi komputasi tersebar (*distributed computing*). *Grid computing* hadir untuk mengatasi keterbatasan sumberdaya komputer untuk melakukan proses komputasi skala besar. Penerapan superkomputer dalam beberapa hal sangat tidak realistis karena harganya yang sangat mahal. Dengan *grid computing*, user dapat menggunakan semua komputer yang tersedia dan terhubung pada jaringan komputer untuk membantu komputasi data skala besar. Untuk mewujudkan ini, dibutuhkan perangkat dan *software* sebagai pengendali aplikasi ini.

Banyak perangkat bantu dibuat guna mengembangkan aplikasi tersebar /parallel tersebut, salah satunya adalah MPI (*Message Passing Interface*) yang dijalankan pada infrastruktur *grid*. Pada paper ini, penulis mencoba dan mengevaluasi penerapan *Globus Toolkit 4.12* untuk aplikasi *Grid Computing*. Oleh karena itu, penulis ingin mencoba mengevaluasi perangkat tersebut sebagai sarana belajar. Penelitian ini sangat penting, karena penulis dapat melakukan berbagai penelitian di bidang *Grid Computing* jika infrastruktur ini telah berjalan.

Cluster dan *Grid computing* mengalami pertumbuhan pesat pada seluruh aplikasi seperti aplikasi akademis, ilmiah, teknik dan komersial. *Grid computing* adalah konsep membangun organisasi virtual atau *supercomputer* terdistribusi atau computer dengan *output* tinggi menggunakan sumber daya yang terhubung pada jaringan komputer di dunia (Lee et al., 2002). *Grid computing* merupakan teknologi baru di dalam ilmu komputer yang merupakan pengembangan baru dari *distributed networking*.

Menurut IBM, *grid computing* mengizinkan kita untuk menggabungkan server, storage system, dan jaringan ke dalam sebuah system besar sehingga kita dapat memberikan sebuah sistem dengan kekuatan besar ke sebuah tujuan komputasi spesifik.

Konsep *grid* dan teknologinya sendiri masih baru, pertama kali dikemukakan oleh Ian Foster dan Kesselman pada 1998 melalui bukunya *The Grid: Blueprint for a New Computing Infrastructure*. Sebelum ini, usaha untuk menggabungkan sumber daya terdistribusi pada area luas dikenal sebagai *metacomputing*. Pada tahun 1998, Ian Foster dan Carl Kesselman memberikan definisi awal pada bukunya yaitu: "A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high end computational capabilities". Pada tahun 2001, Ian Foster, Kesselman dan Tuecke mendefinisikan ulang *grid* sebagai "Coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations".

Foster (2002) lalu membuat sebuah checklist yang dapat digunakan untuk memahami secara tepat apakah yang dapat diidentifikasi sebagai *grid system*, dengan membagi menjadi tiga bagian. Bagian pertama, adalah ada sharing sumberdaya terkoordinasi dengan tanpa titik control tersentralisasi di dalam domain administrative berbeda. Jika hal tersebut tidak benar, sistem tersebut bukan termasuk *grid system*. Bagian kedua, jika menggunakan standar, open, protocol dan antarmuka umum. Bagian ketiga adalah mengecek apakah sistem tersebut menjalankan kualitas layanan *non trivial* yang lebih menguntungkan dibandingkan komponen individual. Layanan ini dapat diasosiasikan dengan *throughput*, *response time*, *meantime between failure*, *security* atau hal lainnya.

Tujuan utama dari sebuah *grid* adalah memvirtualisasi sumber daya untuk mengatasi masalah sains dan teknik. Sumber daya utama *Grid computing* didesain untuk memberikan akses antara lain: *computing/processing power*, *data storage/networked file system*, komunikasi dan *bandwidth*, dan *software* aplikasi.

Terdapat enam komponen utama dari *grid computing*: *security*, *user interface*, *workload management*, *scheduler*, *data management*, dan *resource management*.

Komputer pada sebuah sistem *grid* terhubung melakukan jaringan dan menjalankan aplikasi *grid*. Mereka juga menangani data yang penting dan sensitif, sehingga komponen *security* dari *Grid computing* adalah sangat penting. Komponen ini termasuk enkripsi, otentikasi dan otorisasi.

Arsitektur standar *Grid computing* pertama kali didefinisikan oleh *Open Grid Services Architecture* (OGSA), yang dibangun oleh para anggota *Global Grid Forum* (GGF), yang sekarang dikenal sebagai *Open Grid Forum* (OGF). OGSA mendefinisikan *grid services*, dan keseluruhan struktur dan layanan yang harus disediakan pada lingkungan *grid*.

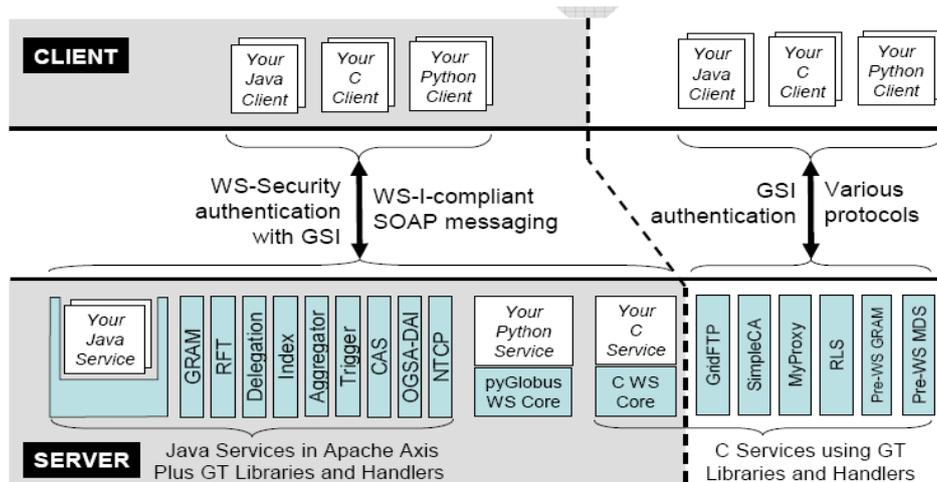
Dibangun pada standar *web services*, OGSA mendefinisikan sebuah *grid services* sebagai *web service* yang mematuhi perjanjian tertentu. Sebagai contoh, *grid services* didefinisikan pada format standar *web services definition language* (WSDL). Hal ini penting karena ia memberikan standar teknik untuk mengakses berbagai *grid service* menggunakan standar yang ada, seperti SOAP, XML, dan WS-*security*.

Menurut Maozhen Li and Mark Baker (2005), spesifikasi *grid* terus berkembang, group yang aktif pada organisasi seperti GGF dan OASIS terus mendefinisikan standar *grid* pada area: model aplikasi dan pemrograman, arsitektur, manajemen data, *security*, performa, *scheduling* dan *resource management*.

Untuk membangun *Grid*, tool yang dibutuhkan adalah: (1) komponen infrastruktur termasuk *file system*, *scheduler* dan *resource manager*, *messaging system*, aplikasi *security*, *Certificate Authorities*, dan mekanisme *file-transfer* seperti GridFTP; (2) *discovery services* untuk mendefinisikan dan memonitor topologi *grid* untuk *sharing* dan berkolaborasi. Untuk melakukan ini, diimplementasikan model yang terkenal seperti LDAP, DNS, *protocol* manajemen jaringan dan *indexing services*; (3) salah satu *benefit* dari sebuah *grid* adalah kemampuan memaksimalkan efisiensi. Ini diwujudkan dengan *scheduler* dan *load balance*. *Scheduler* memastikan job diselesaikan pada beberapa pertimbangan (*priority*, *deadline*, *urgency* dan lainnya), dan *load balancer* mendistribusikan tugas dan manajemen data pada sistem untuk mengurangi peluang *bottleneck*.

Open Grid Services Architecture (OGSA), dibuat oleh *Global Grid Forum* dengan tujuan mendefinisikan arsitektur yang umum, standar dan terbuka untuk aplikasi berbasis *grid* dan membutuhkan *stateful services*. Aplikasi *Grid* biasanya terdiri dari beberapa komponen, misalnya: (1) *VO management service* untuk mengatur *node* dan *user* manakah yang bagian dari tiap *organization*; (2) *resource discovery and management service*: untuk memungkinkan aplikasi pad agri dapat menemukan sumber daya dan mengaturnya; (3) *job management service* – *user* dapat mengirim *task* ke *grid*; (4) layanan lainnya seperti *security* dan *data management*

Globus Toolkit adalah sebuah *software toolkit* yang dibangun oleh The *Globus Alliance* (2002), yang dapat digunakan untuk membangun *grid system*. *Globus Toolkit 4* adalah tollkit open source yang diorganisasikan sebagai sekumpulan komponen *loosely coupled*. Komponen ini terdiri dari layanan, library pemrograman dan tool pengembangan yang didesain untuk membangun aplikasi berbasis *Grid*. Komponen di atas berjalan pada *platform Linux*. Berikut ini merupakan arsitektur dari *Globus Toolkit 4* (Gambar 1):

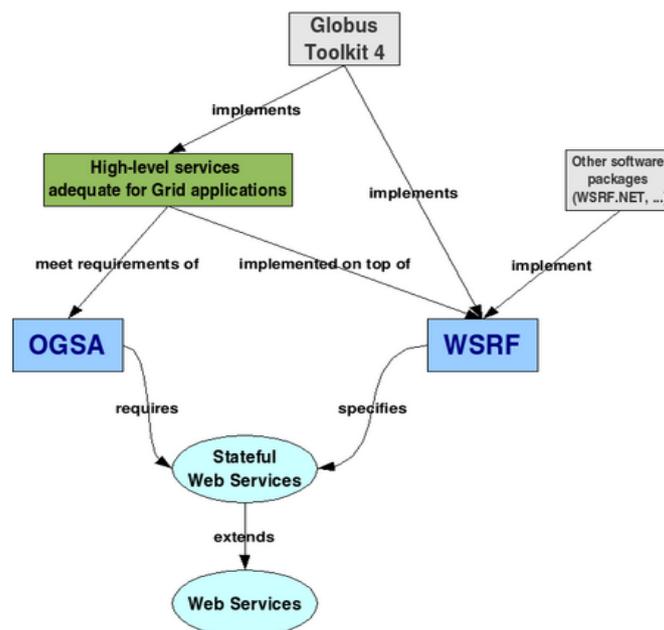


Gambar 1. Arsitektur GT 4 service (The Globus Alliance, 2002).

Komponen utama dari GT antara lain: GSI (*Grid Security Infrastructure*), *data management*, *execution management*, dan *information services*.

Salah satu fitur *Grid* yaitu *GridFTP* yang dapat menyediakan transfer data yang aman dan *reliable* di antara host *grid*. Protocol tersebut merupakan pengembangan dari *FTP* untuk menyediakan fitur tambahan seperti: (1) *Grid Security Infrastructure* (GSI) and Kerberos support yang memungkinkan kedua tipe autentikasi; (2) *transfer data third-party* mengizinkan sebuah *third party* mentransfer *file* di antara banyak *server*; (3) transfer data paralel menggunakan *TCP stream* akan meningkatkan *bandwidth* secara agregat.

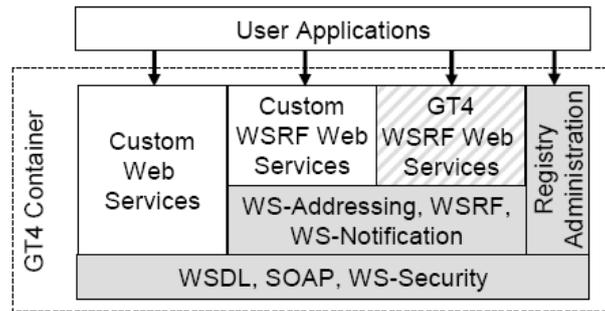
WSRF merupakan sebuah spesifikasi dari OASIS. WSRF menentukan bagaimana kita dapat membuat *web services* menjadi *stateful*. Hubungan antara OGSA dan WSRF adalah WSRF menyediakan *stateful services* yang dibutuhkan oleh OGSA (Gambar 2).



Gambar 2. Hubungan antara OGSA, GT4, WSRF dan *web services*.

Web service adalah komponen *software* yang dapat berinteraksi antar komputer dengan sebuah alamat jaringan yang dideskripsikan pada *Web Service Description Language* (Foster dan Kesselman, 1998). Interaksi *service* tersebut dideskripsikan menggunakan *Simple Object Access Protocol (SOAP)*. Spesifikasi WSRF mendefinisikan *framework* terbuka untuk pemodelan dan pengaksesan sumber stateful menggunakan *web services*.

Untuk membangun aplikasi *grid* yang tersimpan pada container GT4, dibutuhkan program java /C yang diprogram menggunakan *web services*, serta menggunakan *protocol Simple Object Access Protocol (SOAP)* seperti Gambar 3 berikut:



Gambar 3. Java GT 4 container.

METODE

Metode penelitian yang digunakan pada penelitian ini adalah studi literatur dari teknologi *Grid computing* yang berkembang, lalu melakukan eksperimen dengan instalasi *software Globus* pada tiga buah komputer bernama nodeA, NodeB dan NodeC. Terlebih dahulu komputer harus dihubungkan melalui jaringan dan harus dibuat *DNS server* dengan nama percobaan ui.ac.id.

Tabel 1 di bawah ini memuat spesifikasi *hardware* dan *software* yang digunakan.

Tabel 1
Spesifikasi *Hardware* dan *Software* yang Digunakan

No.	Nama	IP	Keterangan
1	NodeA • Laptop PIV 1 • Memory 1G	192.168.0.1	- Debian 4/Fedore Core - GT 4.1.2 -UCLA <i>Grid</i> Portal
2	NodeB • PC PIV • Memory 1 M	192.168.0.2	- Debian 4 -UCLA <i>Grid</i> Portal - GT 4.1.2
3	NodeC • Laptop PIV • Memory 512M	192.168.0.3	- Debian 4 - GT 4.1.2

Proses Instalasi dan Konfigurasi

Untuk membangun infrastruktur *Grid*, diperlukan paket yang lengkap, dengan *software* utamanya adalah: sistem operasi Debian 4.0 etch, Fedore Core 4, UCLA Grid Portal, dan GT 4.1.2. Setelah Linux Debian di-*instal* pada nodeB dan nodeC, paket tambahan yang perlu di-*instal* melalui *Synaptic Package Manager* adalah: *build-essential*, *openssh-server*, *autogen*, *mysql-server*, *automake*, *g++*, *gcc*, *zlib* *zlib-bin* *zlib-dev*, *xinetd*, *rsh* *rsh-server*, *bind*, *postgresql*, *libiodbc* dan *libiodbc2 dev*, dan *perl* *perl-base* *perl-modules* *perl-suid*.

HASIL DAN PEMBAHASAN

Untuk mencoba menjalankan GLOBUS yang sudah di-*instal*, berikan perintah *start* untuk menjalankan kontainer Globus:

```
root@nodec:~# chmod +x /etc/init.d/globus-4.1.2
root@Nodec:~# /etc/init.d/globus-4.1.2 start
Starting Globus container. PID: 29985
```

Untuk melihat *log* dari *container* tersebut, dapat digunakan perintah berikut:

```
root@nodec:~# cat /usr/local/globus-4.1.2/var/container.log
2005-11-15 08:48:00,886 ERROR service.ReliableFileTransferImpl [main,<init>:68]
Unable to setup database driver with pooling.A connection error has occurred:
FATAL: no pg_hba.conf entry for host "140.221.8.31", user "globus",
database "rftDatabase", SSL off

2005-11-15 08:48:02,183 WARN service.ReliableFileTransferHome [main,initialize:97]
All RFT requests will fail and all GRAM jobs that require file staging will fail.
A connection error has occurred: FATAL: no pg_hba.conf entry for host
"140.221.8.31", user "globus", database "rftDatabase", SSL off

Starting SOAP server at: https://192.168.0.3:8443/wsrf/services/
With the following services:

[1]: https://192.168.0.3:8443/wsrf/services/TriggerFactoryService
[2]: https://192.168.0.3:8443/wsrf/services/DelegationTestService
[3]: https://192.168.0.3:8443/wsrf/services/SecureCounterService
[4]: https://192.168.0.3:8443/wsrf/services/IndexServiceEntry
...
```

Pesan di atas menandakan server AXIS sudah berjalan, dan beberapa *web services* sudah aktif. Untuk mencoba menjalankan aplikasi WSRF pada klien, gunakan perintah:

```
bacon@nodec $:export JAVA_HOME /usr/java/jdk1.5.0_16
bacon@nodec$: export ANT_HOME /usr/local/apache-ant-1.6.5/
bacon@nodec$: export PATH $ANT_HOME/bin:$JAVA_HOME/bin:$PATH
bacon@nodec$: counter-client -s
https://nodec.ui.ac.ui.id:8443/wsrf/services/CounterService
Got notification with value: 3
Counter has value: 3
Got notification with value: 13
```

Dari tampilan program di atas, terlihat bahwa klien sudah dapat mengakses method *web services* bernama *CounterService* yang ada di *server*. Berdasarkan pengalaman, infrastruktur *grid* ini hanya berjalan jika semua paket pendukungnya di-*instal* lengkap dan benar. Oleh karena itu, perlu diperhatikan pada saat instalasi program di tiap komputer. Selanjutnya method *web services* inilah yang dapat digunakan oleh siapa saja untuk melakukan komputasi data berbasis *grid computing*.

PENUTUP

Dari hasil penelitian eksperimen ini, dapat diambil kesimpulan: (1) instalasi infrastruktur *Grid computing* membutuhkan konfigurasi DNS server yang benar; (2) *Certificate Authority* (CA) sangat diperlukan pada saat instalasi dan konfigurasi *Globus Toolkit 4.0*; (3) infrastruktur *Grid computing* dapat dicoba menggunakan tiga buah komputer sebagai sarana belajar teknologi tersebut; (4) sistem operasi Debian 4 direkomendasikan digunakan sebagai sistem operasi utama *Globus Toolkit 4.x*.

Selanjutnya, berikut saran-saran untuk penelitian lanjutan atau panduan bagi pemula yang ingin membangun *grid computing*: (1) perlu memahami secara mendalam instalasi dan administrasi Linux Debian dan Fedora; (2) menggali berbagai sumber baik buku dan internet tentang proses instalasi hingga tantangan pengembangan *Grid Computing*; (3) melakukan instalasi tambahan paket MPI untuk dapat menjalankan program paralel berbasis bahasa C/Fortran.

DAFTAR PUSTAKA

- Foster, Ian & Kesselman, Carl. (1998). *The Grid: Blueprint for a New Computing Infrastructure* (1st edition). San Francisco: Kaufmann Publishers.
- Foster, Ian. (2002). *What is the Grid? A Three Point Checklist*. Diakses 12 Januari 2012 dari <http://www.mcs.anl.gov/~itf/Articles/WhatIsTheGrid.pdf>.
- Lee, C., Buyya, R., and Roe, P. (2002). Special issue on cluster *computing* and the *grid*. *Journal of Future Generation Computing Systems* (FGCS), Elsevier Science, The Netherlands, 18 (4).
- Li, Maozhen & Baker, Mark. (2005). *The Grid Core Technologies*. England: Wiley.
- The Globus Alliance. (2002). *About the Globus Toolkit*. Diakses 12 Januari 2012 dari <http://www.globus.org/toolkit/>.