

Web Services and PC Grid

ONG Guan Sin

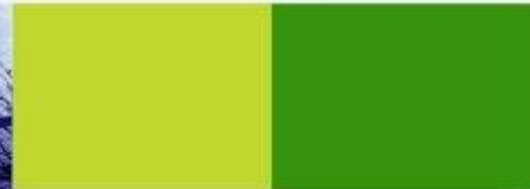
<GuanSin.Ong@scs.com.sg>

Grid Evangelist

Singapore Computer Systems Ltd

APBioNet/APAN

18 Jul 2006



Start

[Home](#) > [Platform](#) > [Device Groups](#)

User: [MPAdmin](#) | [Help](#) | [Logout](#)

Device Groups

+ Actions

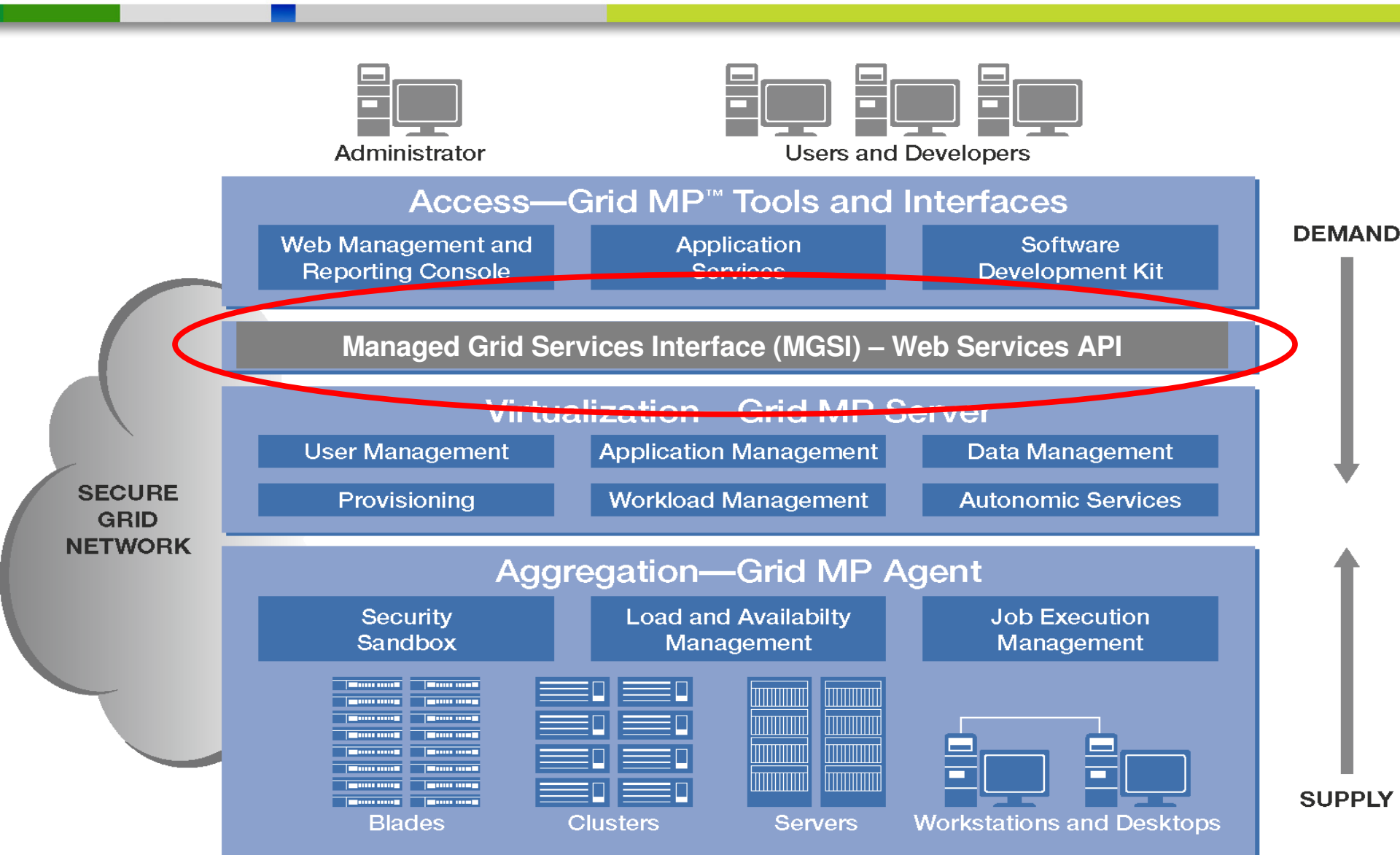
Device Groups

+ Filter/Search Options

Refresh ▾

Name	State	Device Counts				Total Whetstone	Average Whetstone	Average Memory
		Total	Connected	Disconnected	Dormant			
All Devices within all Device Groups	—	<u>1,327</u>	<u>552</u>	<u>17</u>	<u>758</u>	316,227 MFlops	238 MFlops	725 MB
BCH	Enabled	<u>32</u>	<u>20</u>	<u>0</u>	<u>12</u>	4,327 MFlops	135 MFlops	491 MB
BIZ	Enabled	<u>83</u>	<u>0</u>	<u>2</u>	<u>81</u>	515 MFlops	6 MFlops	444 MB
CCE	Enabled	<u>256</u>	<u>102</u>	<u>4</u>	<u>150</u>	75,949 MFlops	297 MFlops	501 MB
CCS	Enabled	<u>16</u>	<u>10</u>	<u>0</u>	<u>6</u>	570 MFlops	36 MFlops	680 MB
CELC	Enabled	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	0 MFlops	0 MFlops	0 MB
ChBE	Enabled	<u>39</u>	<u>35</u>	<u>0</u>	<u>4</u>	18,585 MFlops	477 MFlops	758 MB
CIT	Enabled	<u>33</u>	<u>2</u>	<u>2</u>	<u>29</u>	11,083 MFlops	336 MFlops	657 MB

UD Grid MP Architecture



```
guansin@segamat:~/Grid/NUS/Project/Apps/UD_HMMER_v5.1.01
File Edit View Terminal Tabs Help

      ADD JOB
      -----

Job name: Test

Type(hmmpfam/hmmsearch): hmmpfam

Do you want to use an existing HMM database?(y/n): y

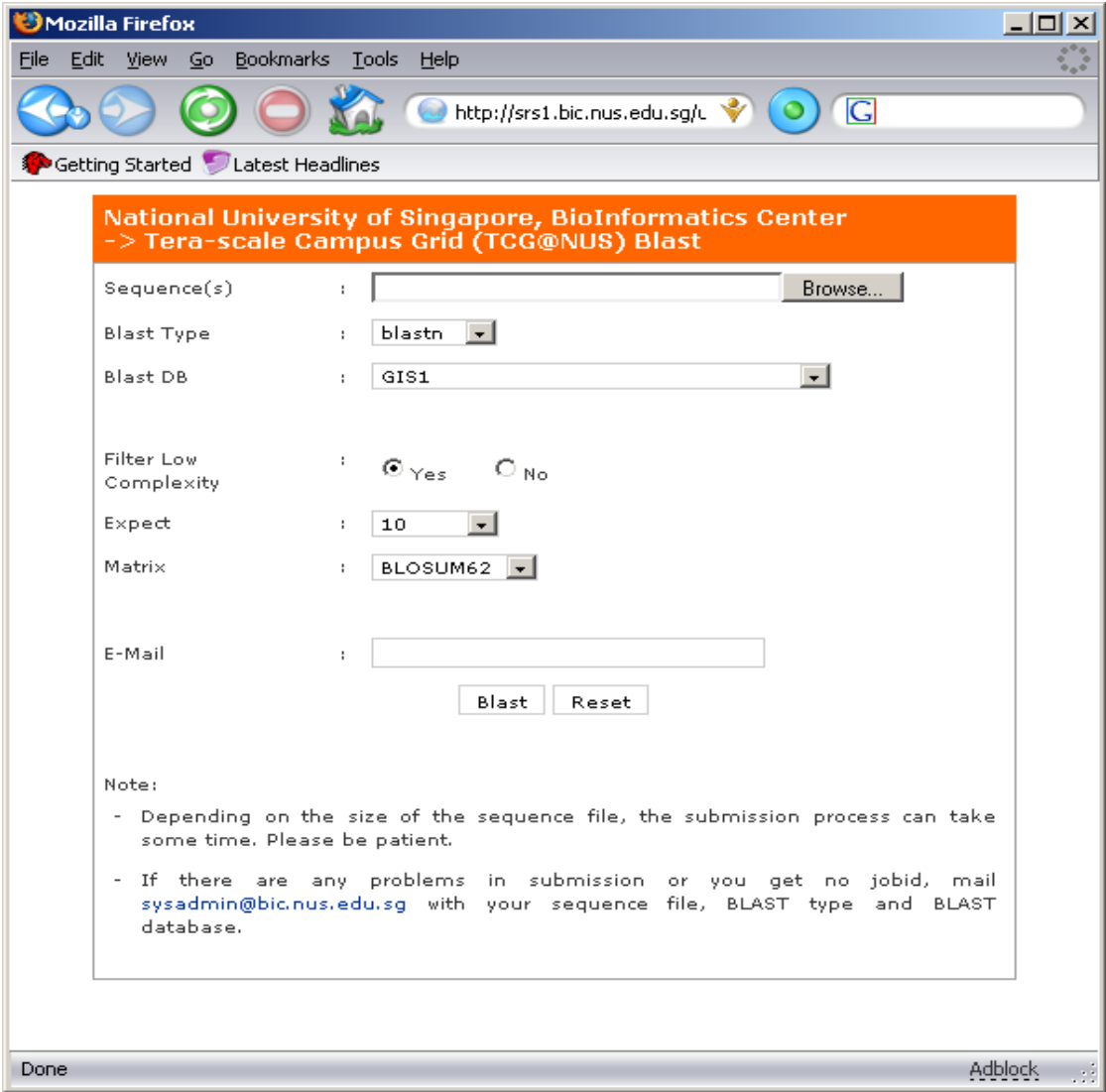
Please specify name of the existing HMM database: PFAM-20051221

Sequence File: ../HMMER-Bernett/proteins.fasta

Please provide command line parameters:
-n          : nucleic acid models/sequence (default protein)
-A <n>      : sets alignment output limit to <n> best domain alignments
-E <x>      : sets E value cutoff (globE) to <x>
-T <x>      : sets T bit threshold (globT) to <x>
-Z <n>      : sets Z (# seqs) for E-value calculation
--cpu <n>   : run <n> threads in parallel (if threaded)
--domE <x>  : sets domain Eval cutoff (2nd threshold) to <x>
--domT <x>  : sets domain T bit thresh (2nd threshold) to <x>
--forward   : use the full Forward() algorithm instead of Viterbi
--null2     : turn OFF the post hoc second null model
--xnu       : turn ON XNU filtering of target protein sequences

*****
NOTE: -n parameter is used only for hmmpfam tasks.
      It is used if model and sequence are nucleic acid and not protein
*****
Command Line: █
```

Simple Web Interface



Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://srs1.bic.nus.edu.sg/

Getting Started Latest Headlines

**National University of Singapore, Bioinformatics Center
-> Tera-scale Campus Grid (TCG@NUS) Blast**

Sequence(s) : Browse...

Blast Type :

Blast DB :

Filter Low Complexity : Yes No

Expect :

Matrix :

E-Mail :

Note:

- Depending on the size of the sequence file, the submission process can take some time. Please be patient.
- If there are any problems in submission or you get no jobid, mail sysadmin@bic.nus.edu.sg with your sequence file, BLAST type and BLAST database.

Done Adblock

▶ Application Service

- Is a job submission and result retrieval program which provides users with a simple interface for performing work on the Grid
- It is responsible for Splitting and Merging Application Data

▶ Features

- Control the Grid MP Sever with SOAP or XML-RPC communications
- SOAP and XML-RPC Communications are protected with SSL encryption
- SOAP and XML-RPC are language and platform independent with many publicly available toolkits and libraries.
- User interface can be command-line, web-based, GUI, etc.
- Can be written to run on various operating systems

▶ MP Grid Services Interface (MGSI)

- All Objects in the Grid MP platform can be controlled through the MGSI
- The Grid MP SDK provides a MGSI Clients in C++ and Java for Easy Application Migration to a Grid Environment

- Initial setup
 - **Login** to the Grid MP and get authentication token to perform all MGSI Calls
 - **Create an Application** object by determining which application you wish to work on, by querying for a list of applications or from information stored in a configuration file
 - **Create a Program** object. This is usually done by getting a list of Programs that belong to the Application
- Creating Data
 - **Create a DataSet** Object to hold All Data chunks to be processed
 - **Split** original input data into chunks and for each chunk -
 - Package data using **buildpkg**
 - **Upload** each Data to the Grid MP File Server
 - **Add** Data Information to a List
 - **Create all Data** Objects in List using a MGSI Call

- Creating Job
 - **Create a Job** Object. This Job will contain All Job Steps (usually there is only 1 Job Step per Job) to be processed
 - **Create Job Step** to contain new pieces of work to be processed
 - **Generate all Workunits** automatically by using a MGSI call that requires a list of Datasets to use per workunit
 - **Set Job Step State to Closed** so the Grid MP System knows that no more work will be added to this Job Step:
 - Get Job Step Structure to Update
 - Change Job Step State to Closed
 - Submit Changes to Job Step
 - **Output Job ID** to the User for use in the Post processing Step

Post-processing Program Flow

- Initial setup
 - **Login** to the Grid MP and get authentication token to perform all MGSI Calls
 - **Create an Application** object by determining which application you wish to work on, by querying for a list of applications or from information stored in a configuration file
 - **Create a Program** object. This is usually done by getting a list of Programs that belong to the Application

- Merging
 - **Get the Job Structure** from the Grid MP using the Job ID returned in the Preprocessing Step
 - **Get the Job Step** for the Job
 - **Get a list of the completed Workunits** belonging to the Job Step
 - For each Workunit:
 - Get the Result record for the Workunit
 - Download Result file from the Grid MP File Server
 - **Merge** all the Results that have been downloaded into the Final Result

Thank You

