



Enabling Grids for E-science

An Approach to Grid Interoperability using Ganga

Soonwook Hwang, KISTI

with YoonKee Lee, EunSung Kim, HeonYoung Yeom, Hangi Kim

OGF25/EGEE User Forum, Catania, March 2-6 2009

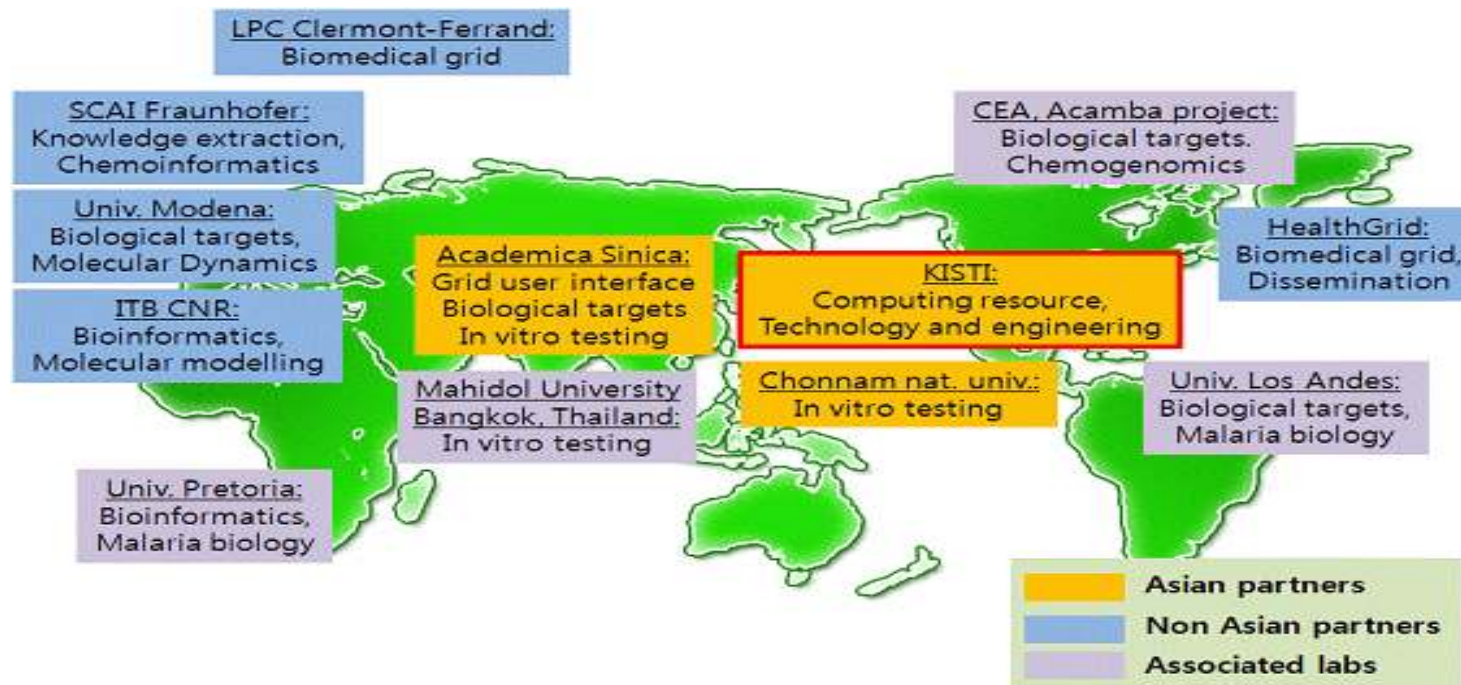
www.eu-egee.org



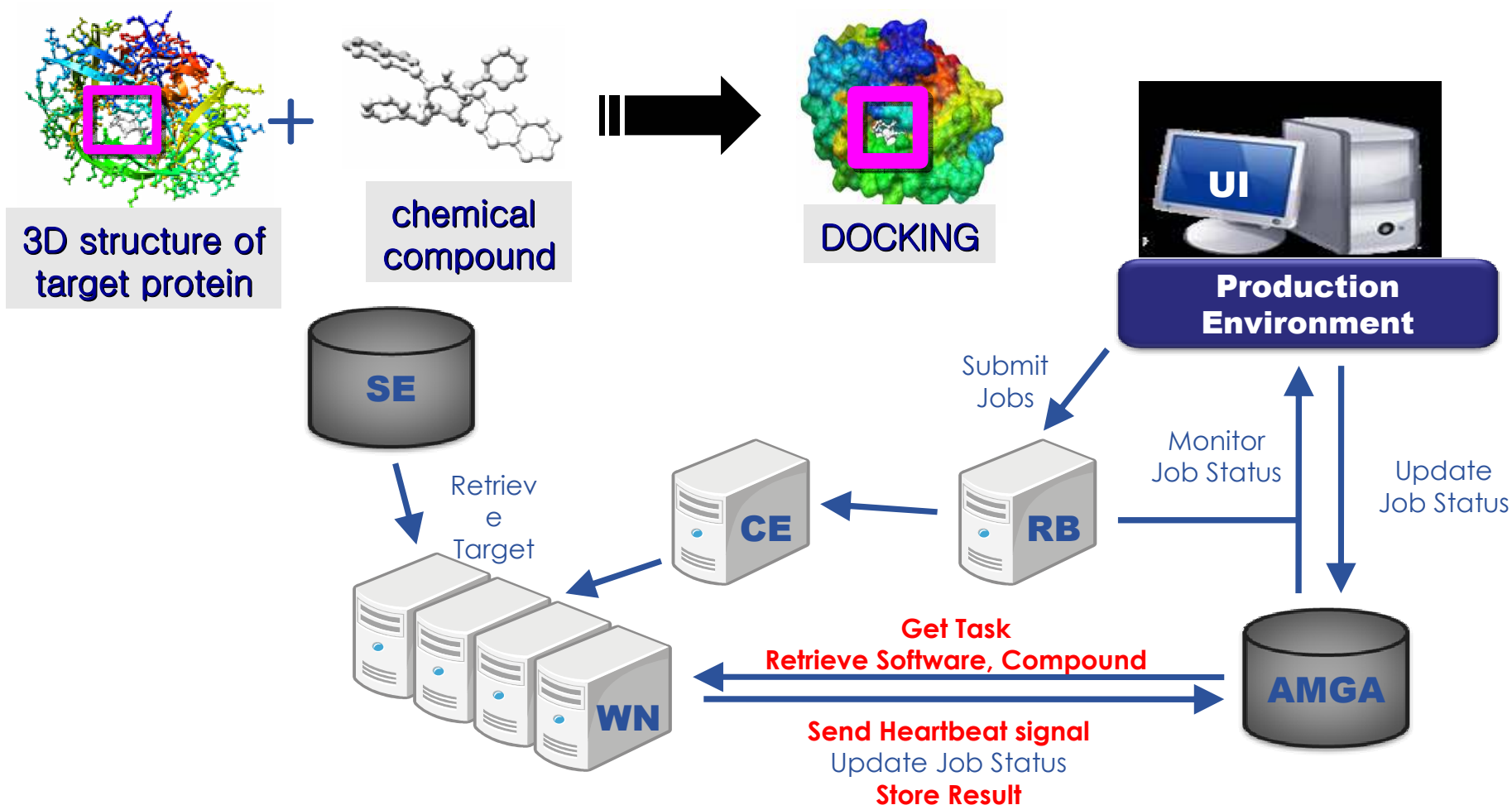
- **KISTI**
 - has been an official partner of EGEE project since EGEE-II, 2006
 - one of the founding members of PRAGMA
- **EGEE**
 - has Grid infrastructure based on **gLite** middleware services
- **PRAGMA**
 - has been operating its own Grid testbed based on **Globus** middleware services

- **Multiple Grid Infrastructures available**
 - EGEE (gLite), PRAGAM, OSG, TeraGrid (Globus), DAISA(Unicore), etc.
 - Users have to learn commands or tools for each grid infrastructure
 - Not easy to use resources across multiple Grids
- **Our Goal**
 - High-level Tools that enable the use of as many as resources possible from multiple grid infrastructures without having to know the details of each grid middleware
- **Our Approach**
 - Chose to use the GANGA as a high-level tool for job execution management exploiting multiple Grid infrastructures
 - Chose to use the WISDOM as a target application

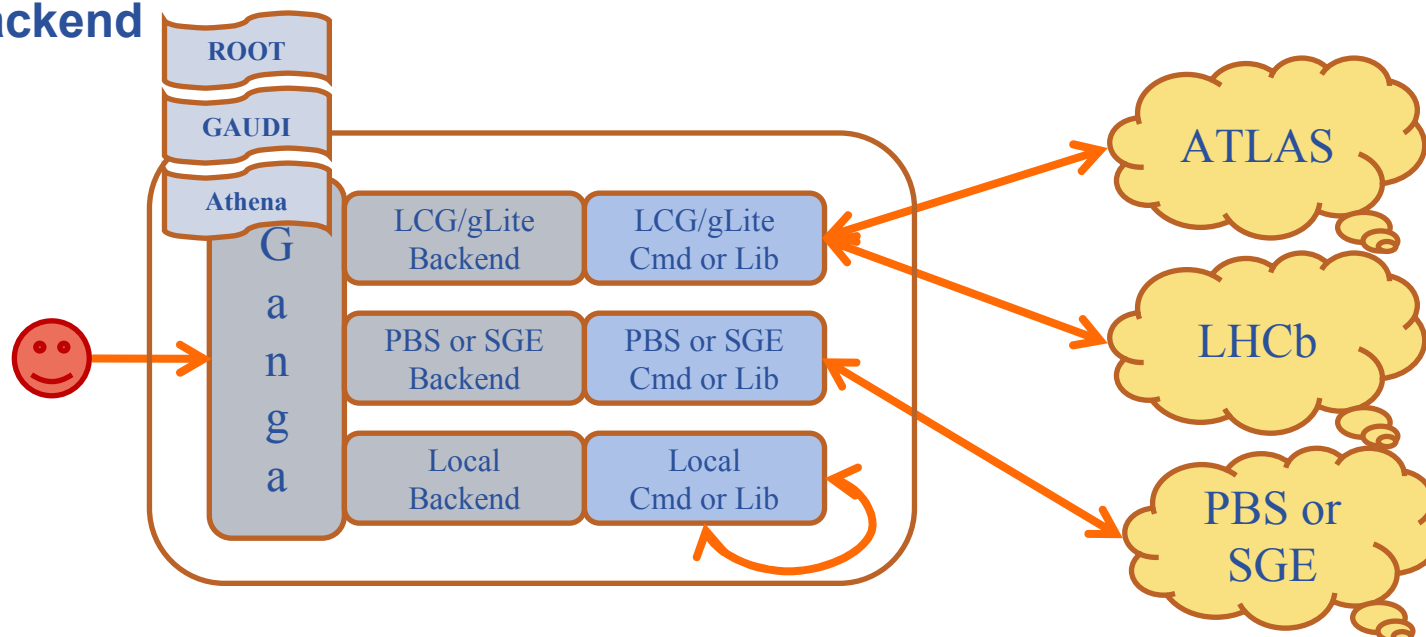
- International initiative to deploy large-scale in-silico docking on a **public grid infrastructure**
- An attempt to find potential drugs against neglected or emerging diseases
 - e.g., Malaria, Avian Flu



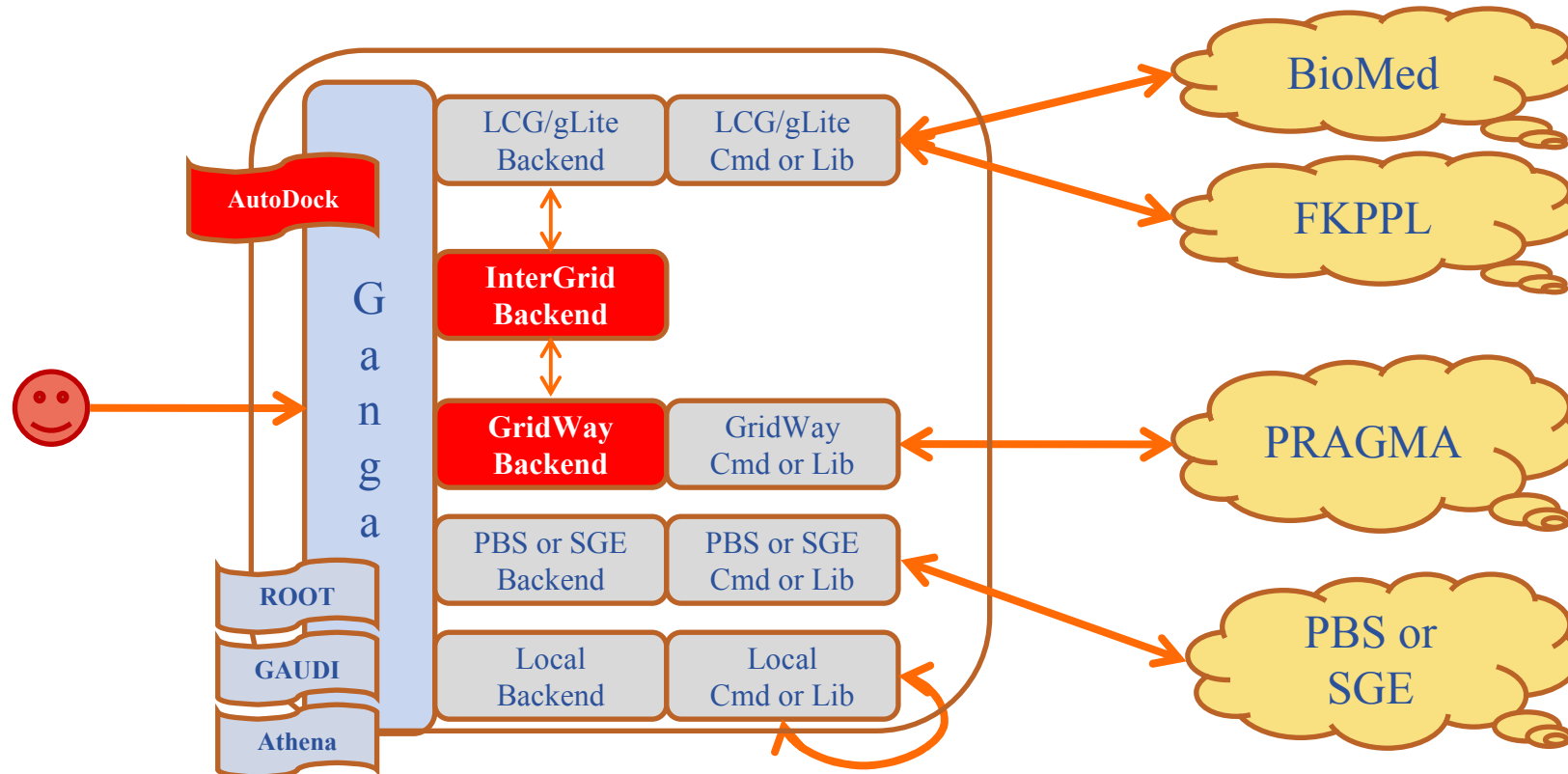
- Millions of independent docking jobs
 - Millions of chemical compounds
 - One or more targets



- **easy-to-use user interface for job submission and management**
 - Implemented in Python: Interactive shell or Script files
 - Use **python syntax** to submit a job
- **provide a plug-in mechanism for job submission on multiple computing backends (e.g., Local PC, Cluster and Grid) and applications (e.g., ROOT, GAUDI and Athena)**
 - easily extended and customized to meet the needs of different user communities
- **Currently, support only the LCG/gLite backend for Grid, not Globus backend**



- Have integrated new plug-ins into GANGA:
 - AutoDock applications
 - GridWay backend
 - InterGrid backend



- **AutoDock Application**
 - Customize configuration for running autodock applications
 - *A LigandSplitter module* has been developed to spread ligand-protein docking jobs on the grid
- **GridWay Backend**
 - Provides access to Globus resources
- **InterGrid Backend**
 - Provides access to both LCG/gLite and Globus resources via the LCG and Gridway backend, respectively
 - Intelligent resource selection
 - Based on load on Grid

- **Authentication**

- International Grid Trust Federation (IGTF) contribution to maintaining world trusted CAs makes it straightforward
- The voms-proxy-init works well to have access to both glite CE and globus gram services.

- The original shell script developed in the WISDOM project **fails to run** on Globus resources
 - The autodock program takes a file describing a 3D structure of target protein and a ligand file as its input files
 - The WISDOM script relies on the globus-url-copy command for copying the two files into the grid node
 - LCG: globus-url-copy – OK
 - Globus: globus-url-copy – Failed (No job delegation supported)

- **Our approach**

- We had to modify the original WISDOM shell script not to rely on the globus-url-copy for file copying
- The new shell script relies on the file staging feature of the Grid.
 - Put all the required files needed by the autodock code into
 - *inputSandbox* in case of LCG backend
 - *inputFile* in case of Gridway backend

It Worked !!!

- One AutoDock Job



```
horn@cloud:~/ganga_autodock

In [1]:!cat autodock_lcg.py
#!/usr/bin/python
j=Job()
j.application=Autodock()
j.application.exe="/bin/sh"
j.application.script="/home/horn/ganga_autodock/autodock.sh"
j.application.binary="/home/horn/ganga_autodock/autodock.tar.gz"
j.application.protein="/home/horn/ganga_autodock/1u2y.tar.gz"
j.application.ligand="/home/horn/ganga_autodock/ligands/9004736_1.pdbq"
j.application.parameter="/home/horn/ganga_autodock/dpf3gen.awk"
j.backend=LCG()
j.submit()
```

```

horn@cloud:~/ganga_autodock

In [2]:execfile("autodock_lcg.py")
Ganga.GPIDev.Lib.Job      : INFO      submitting job 62
Ganga.GPIDev.Adapters    : INFO      submitting job 62 to LCG backend
Ganga.Lib.LCG             : WARNING   The size of /home/horn/gangadir/workspace/horn/LocalAMGA/62
/input/_input_sandbox_62.tgz is larger than the sandbox limit (10000000 byte). Please wait while pre-staging ...
Ganga.Lib.LCG             : WARNING   The size of /home/horn/gangadir/workspace/horn/LocalAMGA/62
/input/_input_sandbox_62_master.tgz is larger than the sandbox limit (10000000 byte). Please wait while pre-staging ...
Ganga.GPIDev.Lib.Job      : INFO      job 62 status changed to "submitted"
    
```

```

horn@cloud:~/ganga_autodock

In [3]:jobs
Out[3]:
Job slice:  jobs (1 jobs)
-----
# fqid      status      name      subjobs      application      backend
backend.actualCE
# 62      completed      Autodock      LCG      kalkan1.ulakbim.gov.tr:2119/jobmanager-lcgpbs
Ganga.GPIDev.Lib.Job      : INFO      job 62 status changed to "running"
Ganga.GPIDev.Lib.Job      : INFO      job 62 status changed to "completing"
Ganga.GPIDev.Lib.Job      : INFO      job 62 status changed to "completed"
    
```

- Multiple AutoDock Jobs (*for* statement)

```
horn@cloud:~/ganga_autodock

In [1]:!cat autodock_gridway_10.py
#!/usr/bin/python
for x in range(10):
    j=Job()
    j.application=Autodock()
    j.application.exe="/bin/sh"
    j.application.script="/home/horn/ganga_autodock/autodock.sh"
    j.application.binary="/home/horn/ganga_autodock/autodock.tar.gz"
    j.application.protein="/home/horn/ganga_autodock/1u2y.tar.gz"
    j.application.ligand="/home/horn/ganga_autodock/ligands/9004736_"+str(x)+".pdbq"
    j.application.parameter="/home/horn/ganga_autodock/dpf3gen.awk"
    j.backend=Gridway()
    j.submit()
```

- Multiple AutoDock Jobs (*for statement*)

```

horn@cloud:~/ganga_autodock
[2]:execfile("autodock_gridway_10.py")
Ganga.GPIDev.Lib.Job      : INFO      submitting job 64
Ganga.GPIDev.Adapters     : INFO      submitting job 64 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 64 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 65
Ganga.GPIDev.Adapters     : INFO      submitting job 65 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 65 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 66
Ganga.GPIDev.Adapters     : INFO      submitting job 66 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 66 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 67
Ganga.GPIDev.Adapters     : INFO      submitting job 67 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 67 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 68
Ganga.GPIDev.Adapters     : INFO      submitting job 68 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 68 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 69
Ganga.GPIDev.Adapters     : INFO      submitting job 69 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 69 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 70
Ganga.GPIDev.Adapters     : INFO      submitting job 70 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 70 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 71
Ganga.GPIDev.Adapters     : INFO      submitting job 71 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 71 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 72
Ganga.GPIDev.Adapters     : INFO      submitting job 72 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 72 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      submitting job 73
Ganga.GPIDev.Adapters     : INFO      submitting job 73 to Gridway backend
Ganga.GPIDev.Lib.Job      : INFO      job 73 status changed to "submitted"
Ganga.GPIDev.Lib.Job      : INFO      job 64 status changed to "submitted"
  
```

- Multiple AutoDock Jobs (*for statement*)

```

horn@cloud:~/ganga_autodock
In [5]:jobs
Out[5]:
Job slice:  jobs (11 jobs)
-----
# fqid      status      name      subjobs      application      backend
backend.actualCE
# 62  completed      Autodock      LCG  kalkan1.ulakbim.gov.tr:2119/j
obmanager-lcgpbs
# 64  running      Autodock      Gridway      cloud.
kisti.re.kr/Fork
# 65  running      Autodock      Gridway      cloud.
kisti.re.kr/Fork
# 66  running      Autodock      Gridway      sun.
kisti.re.kr/Fork
# 67  running      Autodock      Gridway      sun.
kisti.re.kr/Fork
# 68  submitted      Autodock      Gridway
--
# 69  submitted      Autodock      Gridway
--
# 70  submitted      Autodock      Gridway
--
# 71  submitted      Autodock      Gridway
--
# 72  submitted      Autodock      Gridway
--
# 73  submitted      Autodock      Gridway
--

```


- Multiple AutoDock Jobs (*LigandSplitter*)

```

horn@cloud:~/ganga_autodock
In [1]:!cat autodock_inter_split.py
#!/usr/bin/python
j=Job()
j.application=Autodock()
j.application.exe="/bin/sh"
j.application.script="/home/horn/ganga_autodock/autodock.sh"
j.application.binary="/home/horn/ganga_autodock/autodock.tar.gz"
j.application.protein="/home/horn/ganga_autodock/1u2y.tar.gz"
#j.application.ligand="/home/horn/ganga_autodock/ligands/9004736_1.pdbq"
j.application.parameter="/home/horn/ganga_autodock/dpf3gen.awk"
j.splitter=LigandSplitter()
j.splitter.ligandDir="/home/horn/ganga_autodock/ligands/"
j.backend=InterGrid()
j.backend.targetBackends=["LCG","Gridway"]
j.submit()
    
```

- Multiple AutoDock Jobs (*LigandSplitter*)

```

horn@cloud:~/ganga_autodock
In [3]: execfile("autodock_inter_split.py")
Ganga.GPIDev.Lib.Job      : INFO      submitting job 86
Ganga.GPIDev.Adapters    : INFO      submitting job 86.0 to InterGrid backend
Ganga.Lib.InterGrid      : WARNING  LCG Utilization : 53.1438574519%
Ganga.Lib.InterGrid      : WARNING  Gridway Utilization : 100.0%
Ganga.Lib.InterGrid      : WARNING  Actually submitted to LCG
Ganga.Lib.LCG            : WARNING  The size of /home/horn/gangadir/workspace/horn/LocalAMGA/86
/0/input/_input_sandbox_86_0.tgz is larger than the sandbox limit (1000000 byte). Please wait while pre-
tagging ...
Ganga.GPIDev.Lib.Job      : INFO      job 86.0 status changed to "submitted"
Ganga.GPIDev.Adapters    : INFO      submitting job 86.1 to InterGrid backend
Ganga.Lib.InterGrid      : WARNING  LCG Utilization : 54.5194371572%
Ganga.Lib.InterGrid      : WARNING  Gridway Utilization : 100.0%
Ganga.Lib.InterGrid      : WARNING  Actually submitted to LCG
Ganga.Lib.LCG            : WARNING  The size of /home/horn/gangadir/workspace/horn/LocalAMGA/86
/1/input/_input_sandbox_86_1.tgz is larger than the sandbox limit (1000000 byte). Please wait while pre-
tagging ...
Ganga.GPIDev.Lib.Job      : INFO      job 86.1 status changed to "submitted"
Ganga.GPIDev.Adapters    : INFO      submitting job 86.2 to InterGrid backend
Ganga.Lib.InterGrid      : WARNING  LCG Utilization : 54.5194371572%
Ganga.Lib.InterGrid      : WARNING  Gridway Utilization : 25.0%
Ganga.Lib.InterGrid      : WARNING  Actually submitted to Gridway
Ganga.GPIDev.Lib.Job      : INFO      job 86.2 status changed to "submitted"
Ganga.GPIDev.Adapters    : INFO      submitting job 86.3 to InterGrid backend
Ganga.Lib.InterGrid      : WARNING  LCG Utilization : 54.5194371572%
Ganga.Lib.InterGrid      : WARNING  Gridway Utilization : 25.0%
Ganga.Lib.InterGrid      : WARNING  Actually submitted to Gridway
  
```

- Multiple AutoDock Jobs (*LigandSplitter*)

```

horn@cloud:~/ganga_autodock
In [5]:jobs(86).subjobs
Out[5]:
Job slice:  jobs(86).subjobs (10 jobs)
-----
# fqid      status      name  subjobs  application  backend
backend.actualCE
# 86.0  submitted  manager-lcgpbs-b  Autodock  InterGrid  epge1.ph.bham.ac.uk:2119/job
# 86.1  submitted  nager-lcgpbs-bio  Autodock  InterGrid  cs-grid1.bgu.ac.il:2119/jobma
# 86.2   running  kisti.re.kr/Fork  Autodock  InterGrid  cloud.
# 86.3   running  kisti.re.kr/Fork  Autodock  InterGrid  cloud.
# 86.4   running  kisti.re.kr/Fork  Autodock  InterGrid  sun.
# 86.5   running  kisti.re.kr/Fork  Autodock  InterGrid  sun.
# 86.6  submitted  --  Autodock  InterGrid
# 86.7  submitted  --  Autodock  InterGrid
# 86.8  submitted  --  Autodock  InterGrid
# 86.9  submitted  --  Autodock  InterGrid
  
```

- **Performance Evaluation**

		100 Jobs	1000 Jobs
FKPPL VO	Submission Time	8m 50s	79m 22s
	Success Percentage	100%	98.90%
Biomed VO	Submission Time	237m 39s	X
	Success Percentage	55%	X

- gLite supports bulk (faster) submission
 - Splitting jobs may be more efficient than submitting bunches of individual jobs



Enabling Grids for E-scienceE

Thank you for your attention

www.eu-egee.org

