Gravitational Wave Data Analysis

Software Implementation

Junwei Cao (曹军威) and Junwei Li (李俊伟)
Tsinghua University
Gravitational Wave Summer School
Kunming, China, July 2009
Computing Challenges

- Large amount of data
- Parallel and distributed processing
- Clusters and grids
  - Clusters: collection of workstations
  - Grids: collection of clusters/supercomputers
- Software becomes the key
- LIGO data analysis software working group (DASWG)
- Future computing infrastructures
The LSC Data Grid (LDG)
The LDG Software Stack

End users & applications
- LDAS
- DMT
- LALApps
- Matlab

Application enabling
- LDM
- LDGreport
- Glue
- Onasys

LSC Job management
- LDR
- LSCdataFind
- LSCsegFind

LSC Data management

LSC Security management
- LSCcertUtils
- LSC CA

The LSC Data Grid Client/Server Environment Version 4.0 (based on VDT 1.3.9)

Applications

Infrastructures
- Condor-G
- Workflow management / Condor DAGman
- VDS
- VOMS
- Catalog service / Globus
- Resource location service / Globus
- Metadata service
- Resource management / Globus GRAM
- Data transfer / GridFTP

Middleware / Services
- Job scheduling / Condor
- Grid security / Globus GSI

Operating Systems and ...
- FC4
- GCC
- Python
- Autotools
- MySQL
The LDG 4.0 Client/Server

- Written in Pacman 3
- Based on VDT 1.3.9
- Support LDG:Client and LDG:ClientPro
- Support multiple platforms: FC4, Solaris and Darwin at client and only FC4 at server
- Support both 32bit and 64bit machines
- Server includes client
- Online documentation for step-by-step installation

```
platformGE('Linux');
package('Client-Environment');
cd('vdt');
package('VDT_CACHE:Globus-Client');
package('VDT_CACHE:CA-Certificates');
package('VDT_CACHE:Condor');
package('VDT_CACHE:Fault-Tolerant-Shell');
package('VDT_CACHE:GSIOpenSSH');
package('VDT_CACHE:KX509');
package('VDT_CACHE:MyProxy');
package('VDT_CACHE:PyGlobus');
package('VDT_CACHE:PyGlobusURLCopy');
package('VDT_CACHE:UberFTP');
package('VDT_CACHE:EDG-Make-Gridmap');
package('VDT_CACHE:Globus-RLS-Client');
package('VDT_CACHE:VDS');
package('VDT_CACHE:VOMS-Client');
cd();
package('Client-FixSSH');
package('Client-RLS-Python-Client');
package('Client-Cert-Util');
package('Client-LSC-CA');
```

OR
```
platformGE('Sun');
package('SolarisPro');
OR
platformGE('MacOS');
package('Mac');
```
1. update metadata lfn
2. get lfn pfns@caltech(remote)
3. generate pfns@mit
4. gridftp pfns@caltech(remote) pfns@mit(local)
5. rlsadd lfn pfns@mit
Data Monitoring Toolkit (DMT)

DMT Online Use Scenario – control-room type

DMT Offline Use Scenario – standalone or grid enabled

Single data stream

Multiple data streams

DMT Monitors

MonServer

Stdout

Trigger files

Alarm files

Trend files

DMT Libraries

Name Server

Server

lsmp

lmsg

client

gui
An Example DMT Monitor

**filelist1.txt**
/data/node10/frame/S3/L3/LHO/H-RDS_R_L3-751658016-16.gwf
/data/node11/frame/S3/L3/LHO/H-RDS_R_L3-751658032-16.gwf
/data/node12/frame/S3/L3/LHO/H-RDS_R_L3-751658048-16.gwf
/data/node13/frame/S3/L3/LHO/H-RDS_R_L3-751658064-16.gwf
/data/node14/frame/S3/L3/LHO/H-RDS_R_L3-751658080-16.gwf
/data/node15/frame/S3/L3/LHO/H-RDS_R_L3-751658096-16.gwf
/data/node16/frame/S3/L3/LHO/H-RDS_R_L3-751658112-16.gwf

**filelist2.txt**
/data/node10/frame/S3/L3/LLO/L-RDS_R_L3-751658016-16.gwf
/data/node11/frame/S3/L3/LLO/L-RDS_R_L3-751658032-16.gwf
/data/node12/frame/S3/L3/LLO/L-RDS_R_L3-751658048-16.gwf
/data/node13/frame/S3/L3/LLO/L-RDS_R_L3-751658064-16.gwf
/data/node14/frame/S3/L3/LLO/L-RDS_R_L3-751658080-16.gwf
/data/node15/frame/S3/L3/LLO/L-RDS_R_L3-751658096-16.gwf
/data/node16/frame/S3/L3/LLO/L-RDS_R_L3-751658112-16.gwf

**multilist.txt**

```
[jcao@ldaspc1 rmon]$ export LD_LIBRARY_PATH=/opt/lscsoft/dol/lib
[jcao@ldaspc1 rmon]$ ./rmon -opt opt -inlists multilist.txt
```

__standalone run of rmon DMT offline monitor__

```
Processing multi list file: multilist.txt
Number of lists added: 2 Total data streams: 2
Processing frame list file: /home/jcao/rmon/filelist1.txt
Number of files added: 1188 Total frame files: 1188
Processing frame list file: /home/jcao/rmon/filelist2.txt
Number of files added: 1188 Total frame files: 1188
startgps=751658000 stride=16 r-statistic=-0.00251782
startgps=751658016 stride=16 r-statistic=-0.0122699
startgps=751658032 stride=16 r-statistic=0.0168868
```

\[ r_k = \frac{\sum_{i=1}^{n}(x_i-x)(y_{ik}-y)}{\sqrt{\sum_{i=1}^{n}(x_i-x)^2 \sum_{i=1}^{n}(y_{ik}-y)^2}} \]
The LDM Modules and Flowchart

LDM_CONFIG

[AGENT]
RESOURCES = @MIT@CIT@LHO@LLO
SITES = /home/jcao/ldm/etc/LDM_SITES
EXEC = /home/jcao/ldm/bin/ldm_exec_script
LOCATE = /home/jcao/ldm/bin/ldm_locate_script
PID = /home/jcao/ldm/var/ldm.pid
LOG = /home/jcao/ldm/var/ldm.log
LDG = /home/jcao/ldg-3.0/

LDM_MODULES

ldm_agent
ldm_submit
ldm_locate_script
ldm_exec_script
ldm_q
ldm_rm
LSCdataFind
condor_master
condor_submit

LDM_SITES

[MIT]
lsdatafindserver = ldas-gridmon.mit.edu
globusscheduler = ldas-grid.mit.edu/jobmanager-condor
environment = LD_LIBRARY_PATH=/dso-test/home/jcao/dol/lib

[CIT]
lsdatafindserver = ldas-gridmon.ligo.caltech.edu
globusscheduler = ldas-grid.ligo.caltech.edu/jobmanager-condor
environment = LD_LIBRARY_PATH=/dso-test/jcao/dol/lib

[LHO]
lsdatafindserver = ldas-gridmon.ligo-wa.caltech.edu
globusscheduler = ldas-grid.ligo-wa.caltech.edu/jobmanager-condor
environment = LD_LIBRARY_PATH=/data2/jcao/dol/lib

[LLO]
lsdatafindserver = ldas-gridmon.ligo-la.caltech.edu
globusscheduler = ldas-grid.ligo-la.caltech.edu/jobmanager-condor
environment = LD_LIBRARY_PATH=/data2/jcao/dol/lib

Other tools

Globus
Job Manager

Condor

Modules developed or deployed

Modules designed and underdeveloped
**Data Monitoring Environment (LDM)**

**Grid-enabled run of rmon DMT offline monitor using LDM**

```
[jcao@ldaspc1 ~]$ cd ldm
[jcao@ldaspc1 ldm]$ source setup.sh
[jcao@ldaspc1 ldm]$ cd ../rmon
[jcao@ldaspc1 rmon]$ ldm_agent
[jcao@ldaspc1 rmon]$ ldm_submit ldm.sub
Jobs test has been submitted.
[jcao@ldaspc1 rmon]$ more ldm_test_condor.out
Processing multi list file: ldm_test_CIT_multilist.txt
   Number of lists added: 2 Total data streams: 2
   ... startgps=751658000 stride=16 r-statistic=-0.00251782
   ...
```

**Automatically generated Condor-G submission file**

```
universe = globus
globus_schedulder = ldas-grid.ligo.caltech.edu/jobmanager-condor
log = ldm_test_condor.log
output = ldm_test_condor.out
error = ldm_test_condor.err
should_transfer_files = YES
when_to_transfer_output = ON_EXIT
transfer_input_files = ldm_test_CIT_multilist.txt, ldm_test_CIT_filelist1.txt,
                      ldm_test_CIT_filelist2.txt, /home/jcao/rmon/opt
arguments = -inlists ldm_test_CIT_multilist.txt -opt opt
environment = LD_LIBRARY_PATH=/dso-test/jcao/dol/lib
executable = /home/jcao/rmon/rmon
Queue
```

**LIGO friendly language.**

Users are interfaced with a LIGO friendly language.

**Users do not bother with technical details of LSC data grid services.**

Data are located and file lists are generated automatically.
Open Science Grid (OSG)

- Grid of Grids
- 20 thousand CPUs
- Petabytes of data storage

- High energy physics
- Bioinformatics
- Nanotechnology
- .......
Gravitational Wave Data Analysis

Thank You!

Junwei Cao
jcao@tsinghua.edu.cn
http://ligo.org.cn