

The LIGO Scientific Collaboration Data Grid Client/Server Environment

Junwei Cao

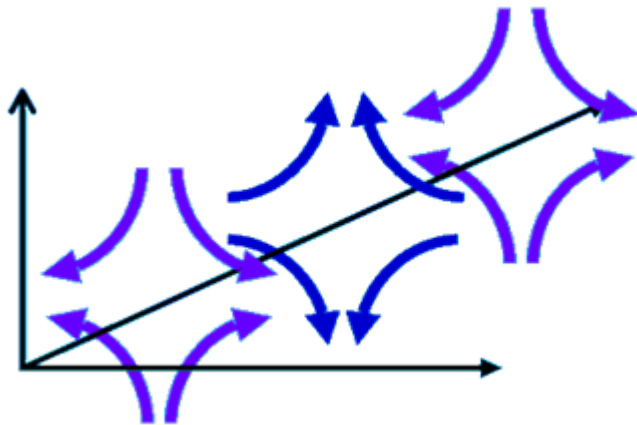
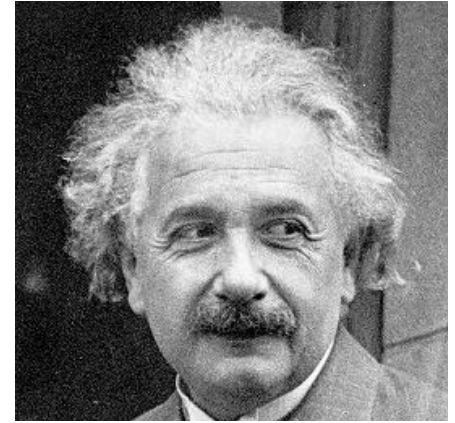
jcao@ligo.mit.edu

MIT LIGO Laboratory

For the LIGO Scientific Collaboration
GridWorld 2006, Washington D.C., U.S.A.

- Gravitational Waves
- LIGO
- LSC
- LIGO Data
- LSC Data Grid (LDG)
- The LDG Client/Server Environment
- Case Study I – LDM
- Case Study II – LDGreport
- The Open Science Grid (OSG)
- Summary

- Gravitational waves are prediction of Einstein's General Theory of Relativity
- They are emitted when matter moves or changes its configuration
- They distort space itself: stretching one direction and squeezing the perpendicular in the first half period and vice versa in the second half
- A new window to observe the universe



- LIGO: Laser Interferometer Gravitational wave Observatory
- Aiming at first direct detection of gravitational waves
- Led by Caltech and MIT, and Funded by NSF
- LSC: the LIGO Scientific Collaboration with over 500 scientists from over 40 institutes worldwide

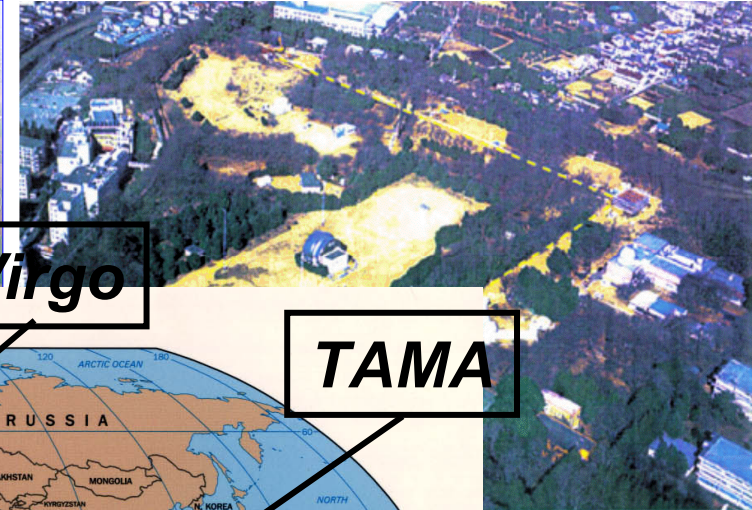
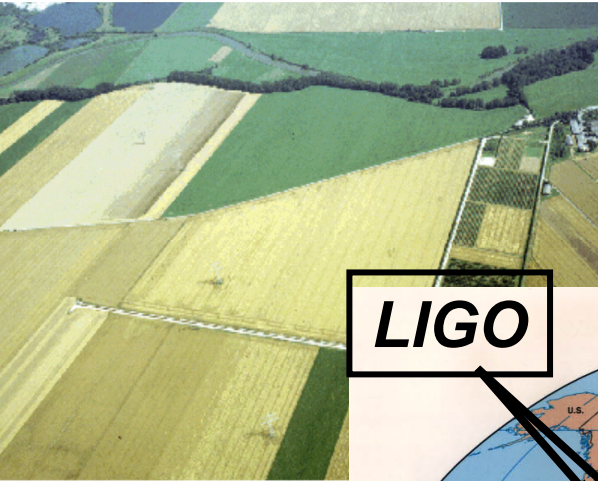


LIGO Hanford Observatory



LIGO Livingston Observatory

LIGO International Network of Detectors



LIGO

GEO

Virgo

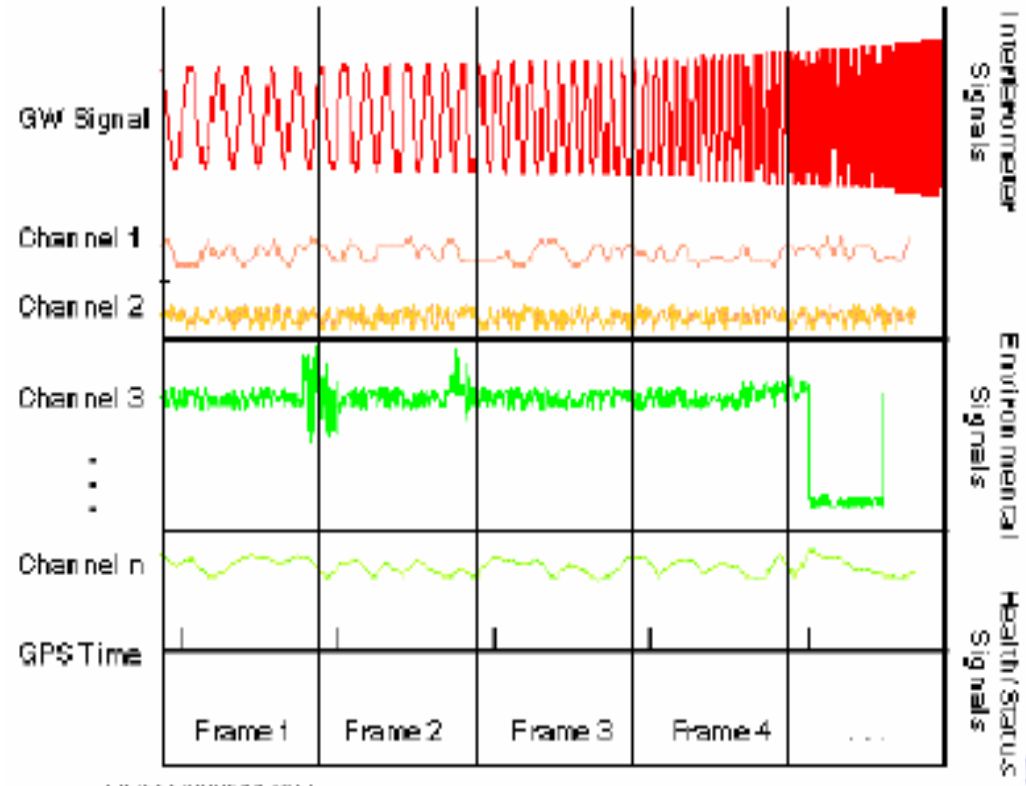
TAMA



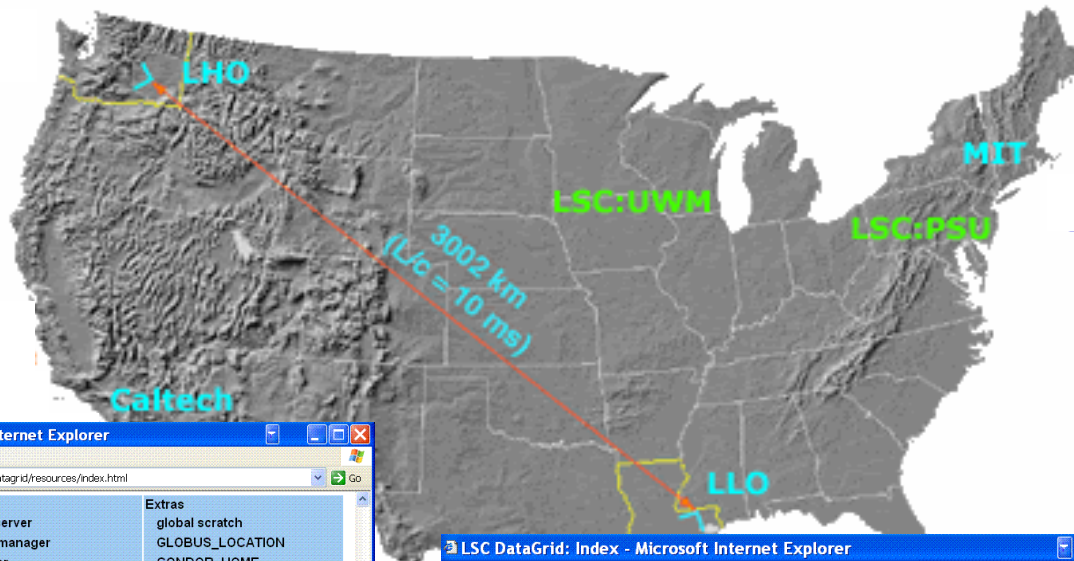
Simultaneously detect signal (within msec)

AIGO

- Continuous time series (1Hz, 128Hz ... 16kHz), all written in an internationally agreed data format: frames
- Data are comprised of:
 - » Gravitational Wave channel: 96 KB/s
 - » Physical Environment Monitors (seismometers, microphones etc): ~1.2 MB/s
 - » Internal Engineering Monitors (sensing, housekeeping, status etc): ~7.2 MB/s
- Multiple data products beyond raw data



1TB/day in 24/7 operation



LSC DataGrid: Index - Microsoft Internet Explorer

Address: http://www.lsc-group.phys.uwm.edu/lscdatagrid/resources/index.html

Basic Info	Servers	Extras
Site Web Home Number of CPUs login node	GridFTP server batch/jobmanager RLS server LDR server	global scratch GLOBUS_LOCATION CONDOR_HOME private subnet
AEI 359 morgane.aei.mpg.de	morgane.aei.mpg.de morgane.aei.mpg.de/jobmanager-condor not yet	/home/NOBACKUP ? /opt/ldg/globus /opt/ldg/condor yes
Birmingham 200 tsunami.sr.bham.ac.uk	tsunami.sr.bham.ac.uk tsunami.sr.bham.ac.uk/jobmanager-condor none	/raid/1/<login> ? /opt/ldg/globus /usr/local/condor yes
Cardiff 160 mini.astro.cf.ac.uk	mini.astro.cf.ac.uk mini.astro.cf.ac.uk/jobmanager-condor not yet	none ? /opt/ldg/globus /opt/ldg/condor yes
CIT Web Home 420 lidas-grid.ligo.caltech.edu	lidas-grid.ligo.caltech.edu:15000 lidas-grid.ligo.caltech.edu/jobmanager-condor rts://ldas-gridmon.ligo.caltech.edu:39281 lidas-cit.ligo.caltech.edu (for external use) lidas-gridmon.ligo.caltech.edu (for local Condor jobs)	The directory /dso-test/ is visible across all the nodes ? /usr/1/clidsk/ldg-2.0/globus /opt/ldg/condor-6.6.1 yes
LHO Web Home 280 lidas-grid.ligo-wa.caltech.edu	lidas.ligo-wa.caltech.edu (tier 1 transfers) lidas-grid.ligo-wa.caltech.edu:15000 (cluster users) lidas-grid.liao-	The directory /dso-test/ is visible across all the nodes available /ldcg/ldg/globus

LSC DataGrid: Index - Microsoft Internet Explorer

Address: http://www.lsc-group.phys.uwm.edu/lscdatagrid/

LSC Data Grid

DataGrid Details

- What is LSC DataGrid?
- Cluster Usage
- Monitoring
- Service details

User Manual

- How to get started
- Install Data Grid Client
- Getting Certificates
- Renewing Certificates
- Account Request
- Intro to Data Grid Tools
- Matlab Cluster Tips
- FAQ

Admin Manual

- Install DataGrid Server
- Get server

Welcome to the LSC DataGrid

The LSC DataGrid is the combination of LSC computational and data storage resources with so called "Grid Computing middleware" to create a coherent and uniform LIGO data analysis environment. The graph on the right shows the current CPU usage across the six active centers across the world

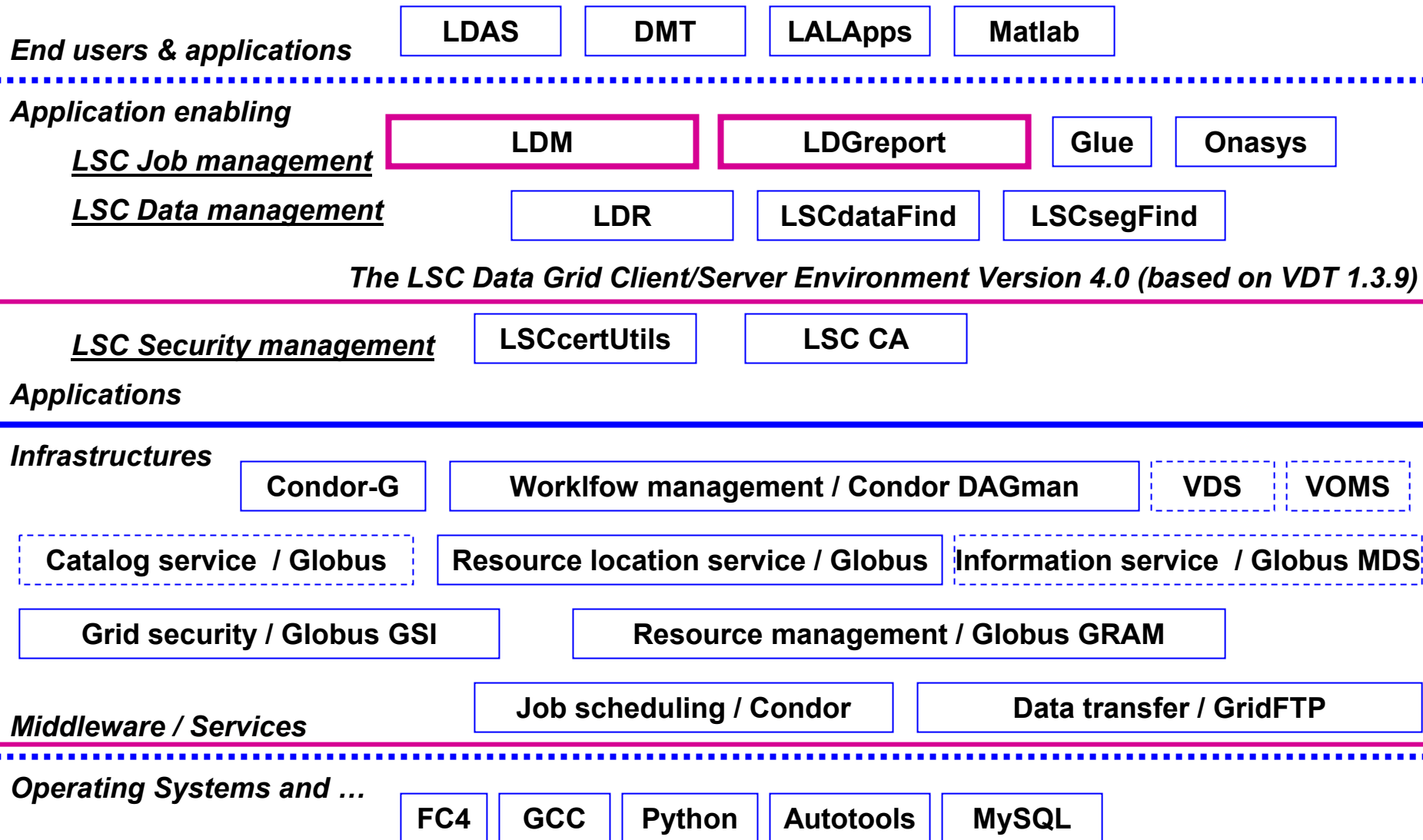
Getting started?

If you are new to the LSC DataGrid and need instructions for installing grid tools, getting a certificate, and requesting access to LSC resources then please see [Getting Started on the LSC DataGrid](#) or click on "Getting started" on the navigation bar on the left.

News

- **7/11/2005:** Version 3.5 Of the [LSC DataGrid Client](#) package and [LSC DataGrid Server](#) package with support for Fedora Core 3, Red Hat Enterprise Linux 3, and Solaris (only for client) are available.

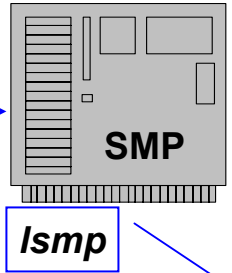
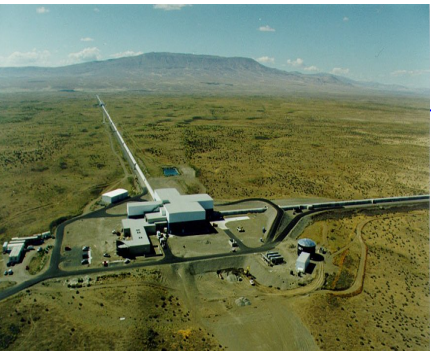




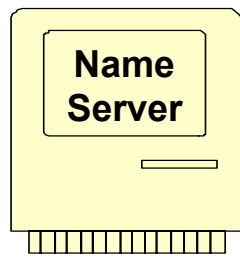
- Written in Pacman 3
- Based on VDT 1.3.9
- Support LDG:Client and LDG:ClientPro
- Support multiple platforms: FC4, Solaris and Darwin at client and only FC4 at server
- Support both 32bit and 64bit machines
- Server includes client
- Online documentation for step-by-step installation

```
.....
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' );
```

LIGO Data Monitoring Toolkit (DMT)



Server



client **gui**



Single data stream



DMT Online Use Scenario – *control-room type*

DMT Offline Use Scenario – *standalone or grid enabled*

Multiple data streams



```

/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
    
```

```

/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
    
```

- base**
- container**
- sigp**
- ezcalib**
- dmtenv**
-

Stdout
Trigger files
Alarm files
Trend files
.....

- xml**
- html**
- xsil**
- event**
- trig**
-

frameio

DMT Libraries

multilist.txt

filelist1.txt
filelist2.txt

rmon

$$r_k = \frac{\sum_i (x_i - \bar{x})(y_{i+k} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_{i+k} - \bar{y})^2}}$$

opt

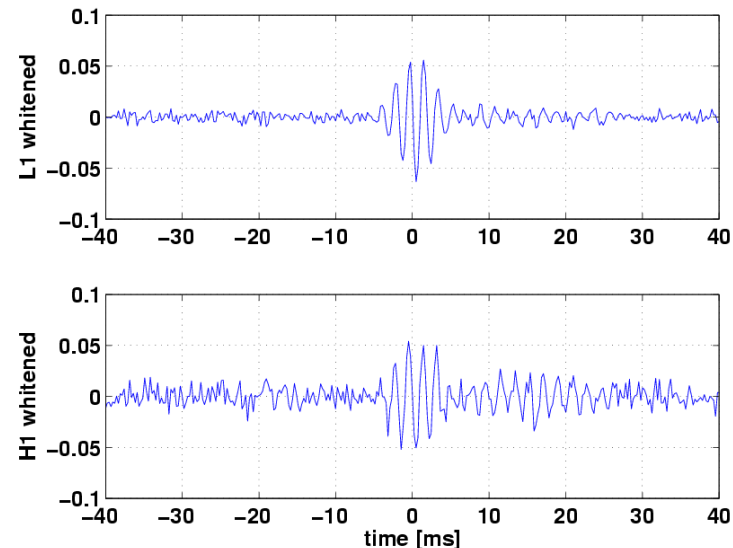
stride	16.0
channel_1	H1:LSC-AS_Q
channel_2	L1:LSC-AS_Q

filelist1.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
```

filelist2.txt

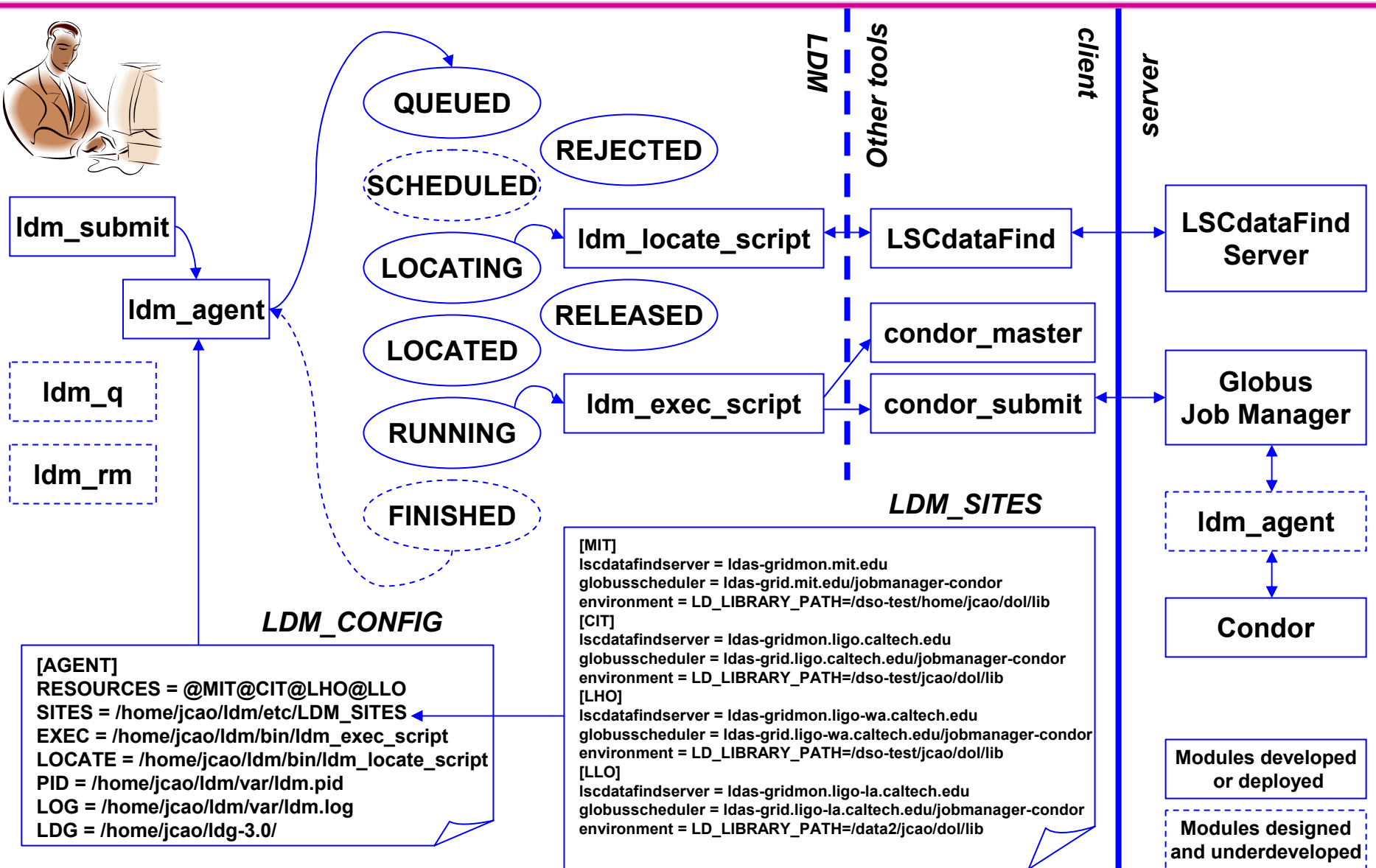
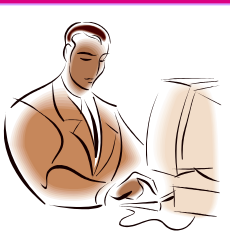
```
/data/node10/frame/S3/L3/LHO/H-RDS_R_L3-751658016-16.gwf
/data/node11/frame/S3/L3/LHO/H-RDS_R_L3-751658032-16.gwf
/data/node12/frame/S3/L3/LHO/H-RDS_R_L3-751658048-16.gwf
/data/node13/frame/S3/L3/LHO/H-RDS_R_L3-751658064-16.gwf
/data/node14/frame/S3/L3/LHO/H-RDS_R_L3-751658080-16.gwf
/data/node15/frame/S3/L3/LHO/H-RDS_R_L3-751658096-16.gwf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



standalone run of rmon DMT offline monitor

```
[jcao@ldaspc1 rmon]$ export LD_LIBRARY_PATH=/opt/lscsoft/dol/lib
[jcao@ldaspc1 rmon]$ ./rmon -opt opt -inlists multilist.txt
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
channel[1]=H1:LSC-AS_Q channel[2]=L1:LSC-AS_Q
startgps=751658000 stride=16 r-statistic=-0.00251782
startgps=751658016 stride=16 r-statistic=-0.0122699
startgps=751658032 stride=16 r-statistic=0.0168868
.....
```


LIGO The LDM Modules and Flowchart



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
Job 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
.....
```

ldm.sub

```
[job]
id = test
monitor = rmon
args = -opt opt
input = opt
[data]
observatory = @H@L
type = @RDS_R_L3@RDS_R_L3
start = 751658000
end = 751676993
```

automatically generated Condor submission file

```
universe = globus
globusscheduler = 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
```

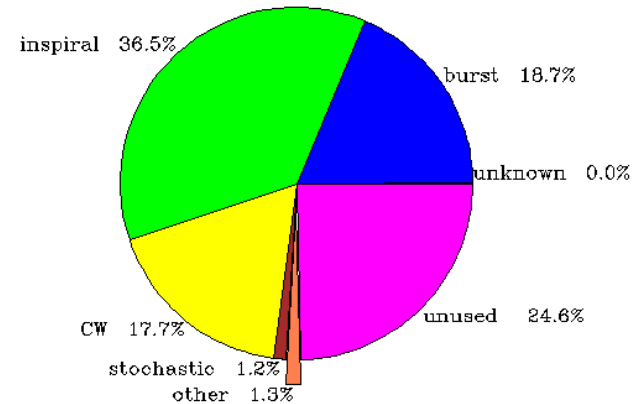
Users are interfaced with a LIGO friendly language.

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

Data are located and file lists are generated automatically

- A command line tool to generate LDG cluster usage statistics
- Integrated with LDG cluster and LSC user information
- Using globusrun to call remote Condor utilities and retrieve user-based usage information
- Automatic chart generation

CIT Total Usage (Available THz-hrs: 15466.7)

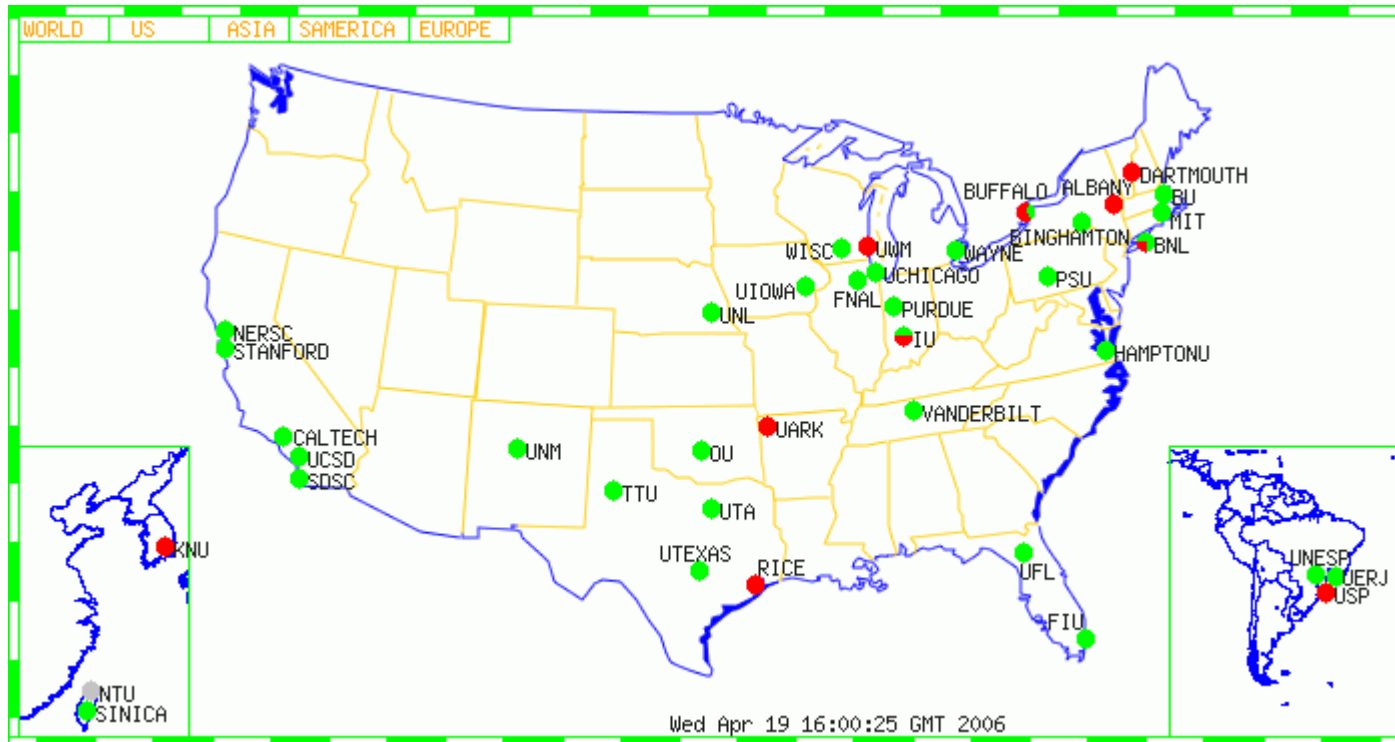


cluser.conf

```
[MIT]
address = ldas-grid.mit.edu
scheduler = condor
starttime = 1-1-2005,3-1-2006
endtime = 2-28-2006,present
nodenumber = 105,80
cpupernode = 1,2
cpuhz = 2,3
uptime = 0.962,0.962
[CIT]
.....
```

user.conf

```
#####
## LDG user information
## LSC working groups enums:
## 0 unknown
## 1 burst
## 2 inspiral
## 3 CW
## 4 stochastic
## 5 other(sysadmin/grid3/osg/...)
#####
[jcao]
name = Junwei Cao
affli = MIT
group = 1
[jzweizig]
name = John Zweizig
affli = CIT
group = 1
...
```

- Grid of Grids
- 20 thousand CPUs
- Petabytes of data storage
- High energy physics
- Bioinformatics
- Astroinformatics
-

The LSC Data Grid:

- A production level grid environment with professional maintenance and supports
- Over 10 grid sites in the USA and Europe; Thousands of CPUs and terabytes of data storage; Hundreds of users
- Producing real scientific computing results

Future work:

- Enabling LIGO applications on LDG/OSG
- LDG/OSG Compatibility

References:

- <http://www.ligo.mit.edu/lsc-itr2003>
- <http://www.ligo.mit.edu/ldg>
- <http://www.lsc-group.phys.uwm.edu/lscdatagrid/>
- <http://www.opensciencegrid.org>