Computing Infrastructure for Gravitational Wave Data Analysis

Junwei Cao for the LIGO Scientific Collaboration

Tsinghua University / MIT LIGO Laboratory

The 3rd China-U.S. Roundtable on Scientific Data Cooperation

Qingdao City, China

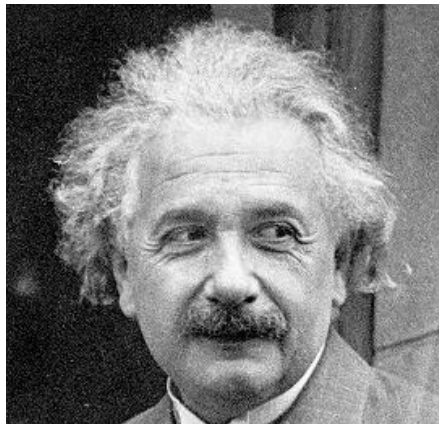
March 23 2009

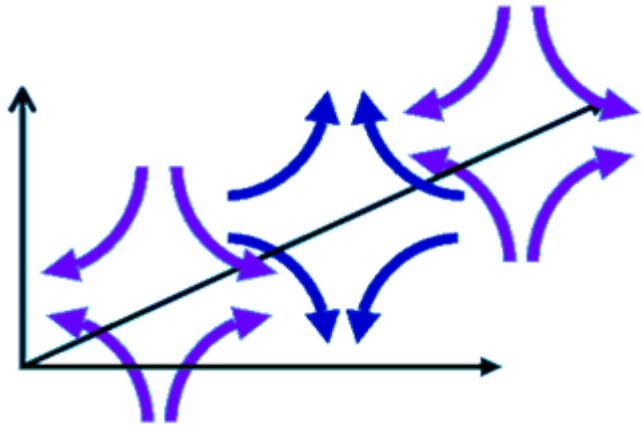
Outline

- Background introduction
  - » Gravitational wave detection
  - » LIGO and LIGO Scientific Collaboration
- LIGO data computing infrastructure
  - » The LIGO data archiving
  - » The Data monitoring system (online & offline)
  - » The LIGO data grid
  - » The LIGO data replication
  - » Grid enabling data monitoring
  - » The Open Science Grid
- Cyberinfrastructure perspective

Gravitational Waves

- Gravitational waves are prediction of Einstein's General Theory of Relativity.
- No direct detection of GWs is successful for nearly 100 years.
- A new window to observe the universe





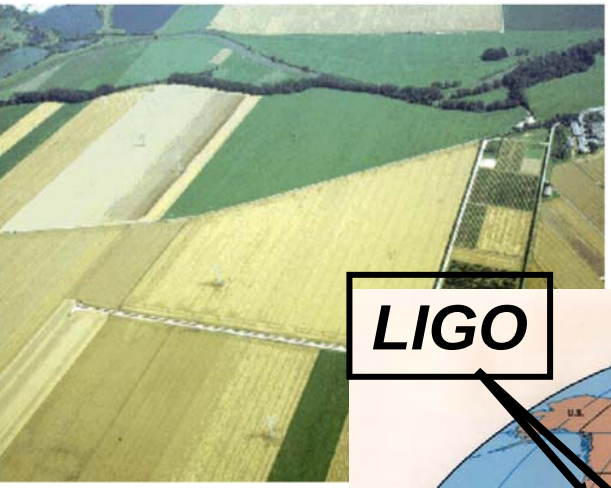
LIGO & LSC

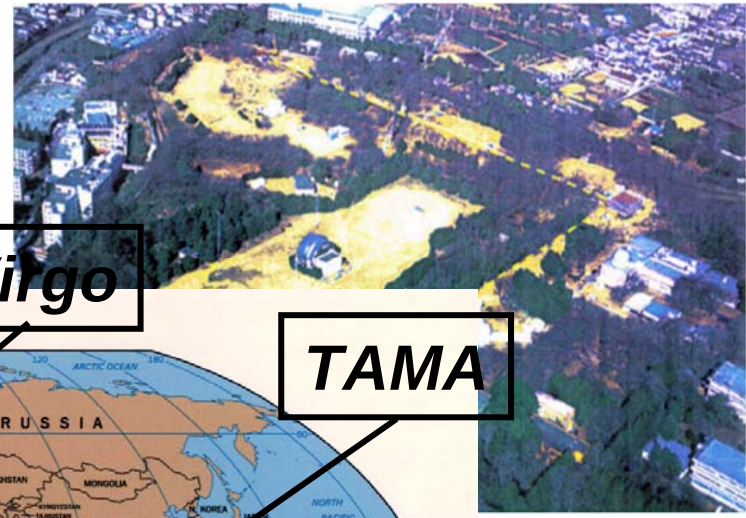
- LIGO: Laser Interferometer Gravitational wave Observatory (激光干涉引力波天文台)
- U.S. NSF flagship project with largest single investment; a joint effort of Caltech and MIT
- LSC: the LIGO Scientific Collaboration with over 500 scientists from over 60 institutes worldwide

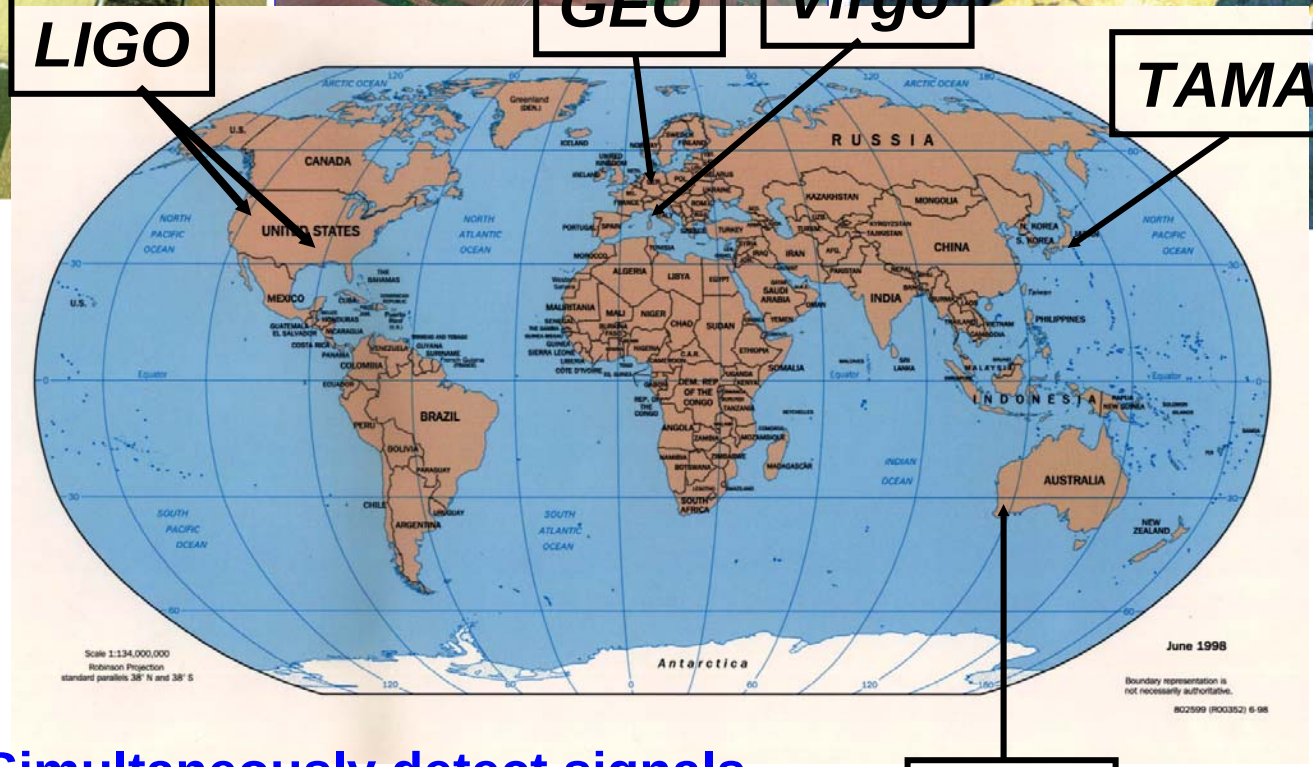
LIGO Hanford Observatory (H1, H2)

LIGO Livingston Observatory (L1)

LIGO International Network of Detectors



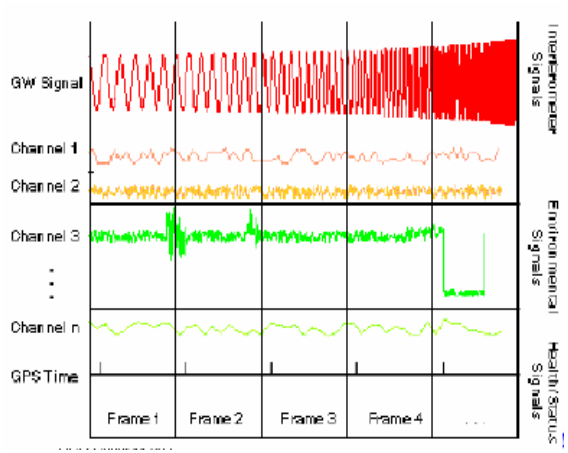




Simultaneously detect signals

LIGO Data Archiving

- Data are comprised of:
  - » Gravitational Wave channel (AS_Q)
  - » Physical Environment Monitors
  - » Internal Engineering Monitors
- Multiple data products beyond raw data
  - » RDS_R_L1, including both GW and environmental channels
  - » RDS_R_L3, including only AS_Q channel



L-RDS_R_L3-751658016-16.gwf

Site (H,L)

Data Type

GPS Start Time

Duration

Data Monitoring Toolkit (DMT)

SMP

lsmp

Server

Name Server

client

gui

lmsg

Single data stream

DMT Online Use Scenario – control-room type

DMT Monitors

MonServer

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

……

frameio

Stdout
Trigger files
Alarm files
Trend files
……

xml

html

xsil

event

trig

……

DMT Libraries

DMT Offline (DOL)

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

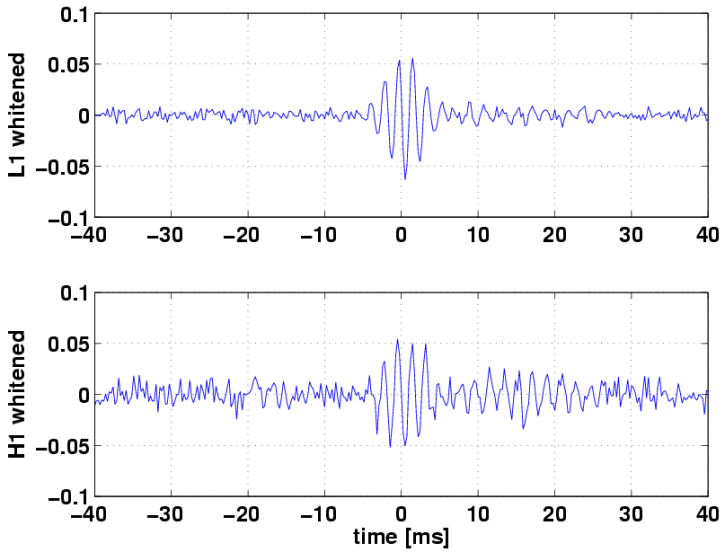
stride	16.0
channel_1	H1:LSC-AS_Q
channel_2	L1:LSC-AS_Q

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

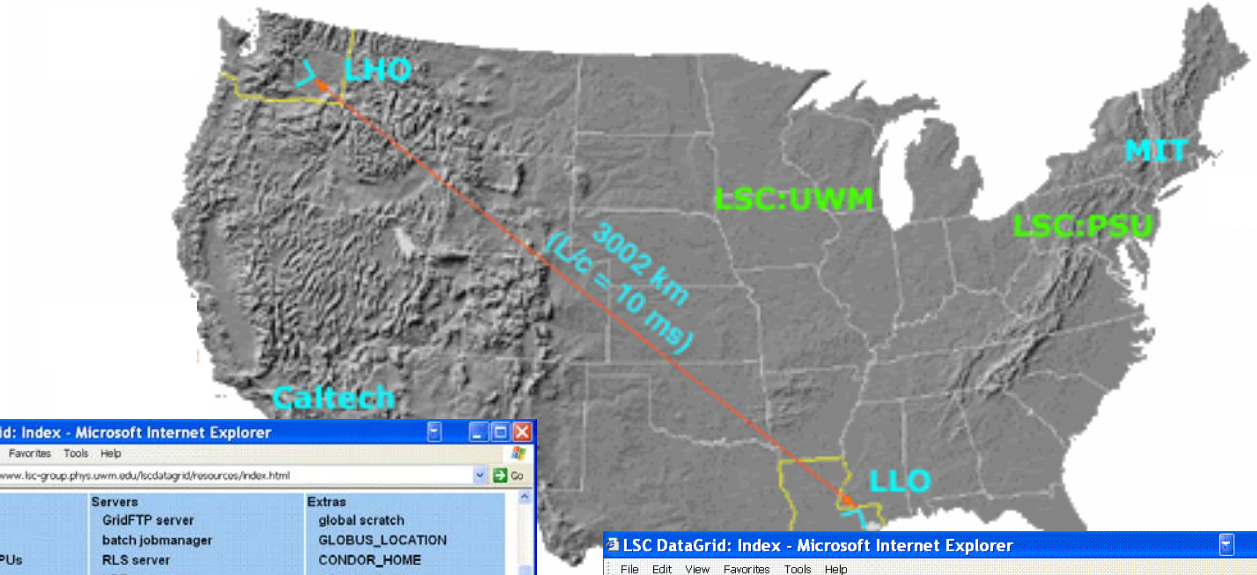
/data/node10/frame/S3/L3/LHO/H-RDS_R_L3-751658016-16.gwf
/data/node11/frame/S3/L3/LHO/H-RDS_R_L3-751658032-16.gwf
/data/node12/frame/S3/L3/LHO/H-RDS_R_L3-751658048-16.gwf
/data/node13/frame/S3/L3/LHO/H-RDS_R_L3-751658064-16.gwf
/data/node14/frame/S3/L3/LHO/H-RDS_R_L3-751658080-16.gwf
/data/node15/frame/S3/L3/LHO/H-RDS_R_L3-751658096-16.gwf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

The LSC Data Grid (LDG)



LSC DataGrid: Index - Microsoft Internet Explorer

File Edit View Favorites Tools Help

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

Basic Info	Servers	Extras
Site	GridFTP server	global scratch
Web Home	batch jobmanager	GLOBUS_LOCATION
Number of CPUs	RLS server	CONDOR_HOME
login node	LDR server	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 ldas-grid.ligo.caltech.edu	ldas-grid.ligo.caltech.edu:15000 ldas-grid.ligo.caltech.edu/jobmanager-condor rls://ldas-gridmon.ligo.caltech.edu:39281 ldas-cit.ligo.caltech.edu (for external use) ldas-gridmon.ligo.caltech.edu (for local Condor jobs)	The directory /dso-test/ is visible across all the nodes ? /usr1/lcldsk/ldg-2.0/globus /opt/ldg/condor-6.6.1 yes
LHO Web Home 280 ldas-grid.ligo-wa.caltech.edu	ldas.ligo-wa.caltech.edu (tier 1 transfers) ldas-grid.ligo-wa.caltech.edu:15000 (cluster users) ldas-grid.ligo-	The directory /dso-test/ is visible across all the nodes available /idcg/ldg/globus

LSC DataGrid: Index - Microsoft Internet Explorer

File Edit View Favorites Tools Help

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

LSC

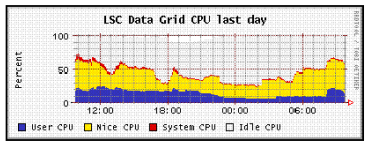
LSC Data Grid

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

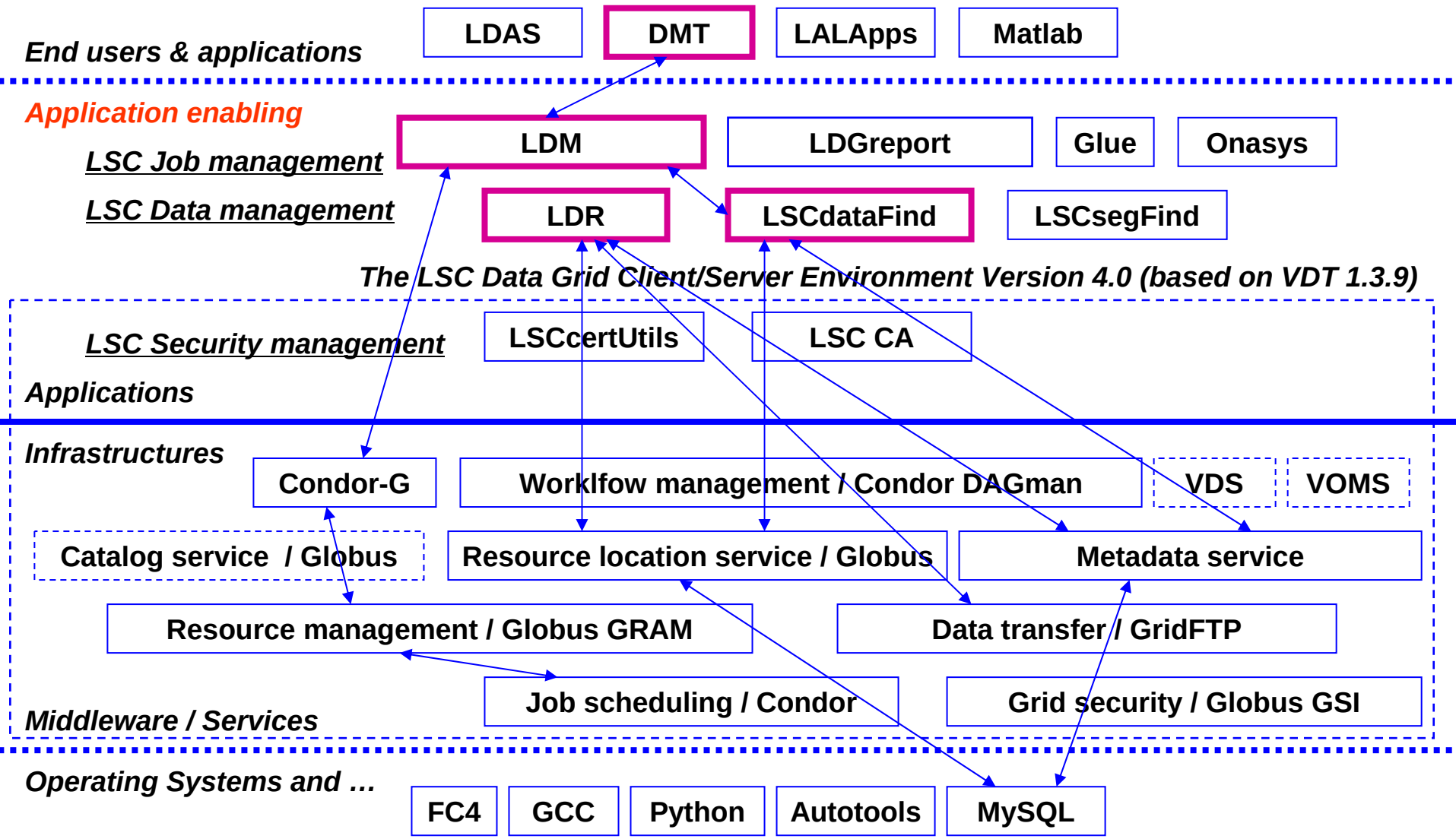
Getting started?

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.

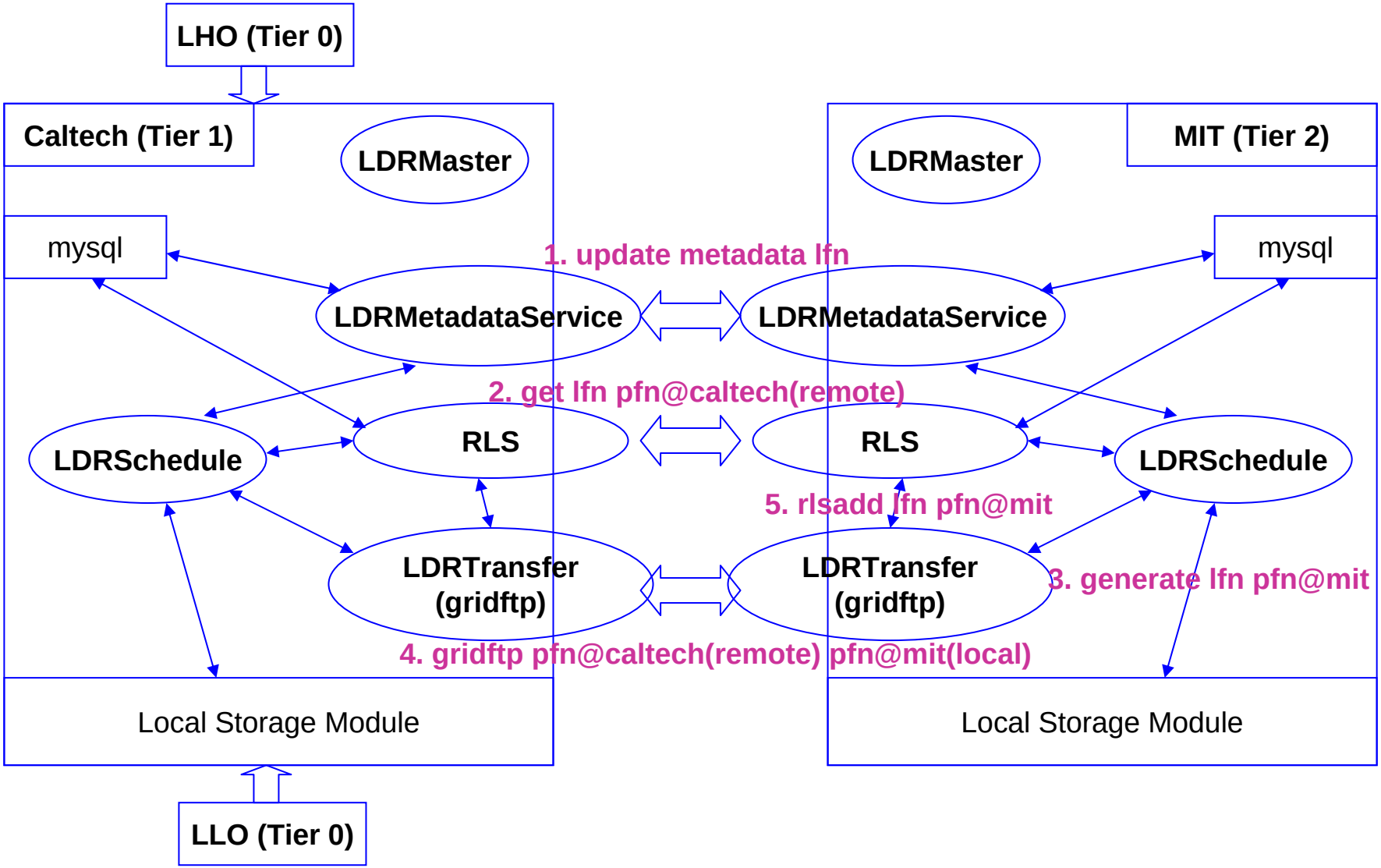
News

- **7/11/2005**: Version 3.5.0f 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.

The LDG Software Stack



LIGO Data Replication (LDR)



LDM: Grid Enabling DMT

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-G submission file

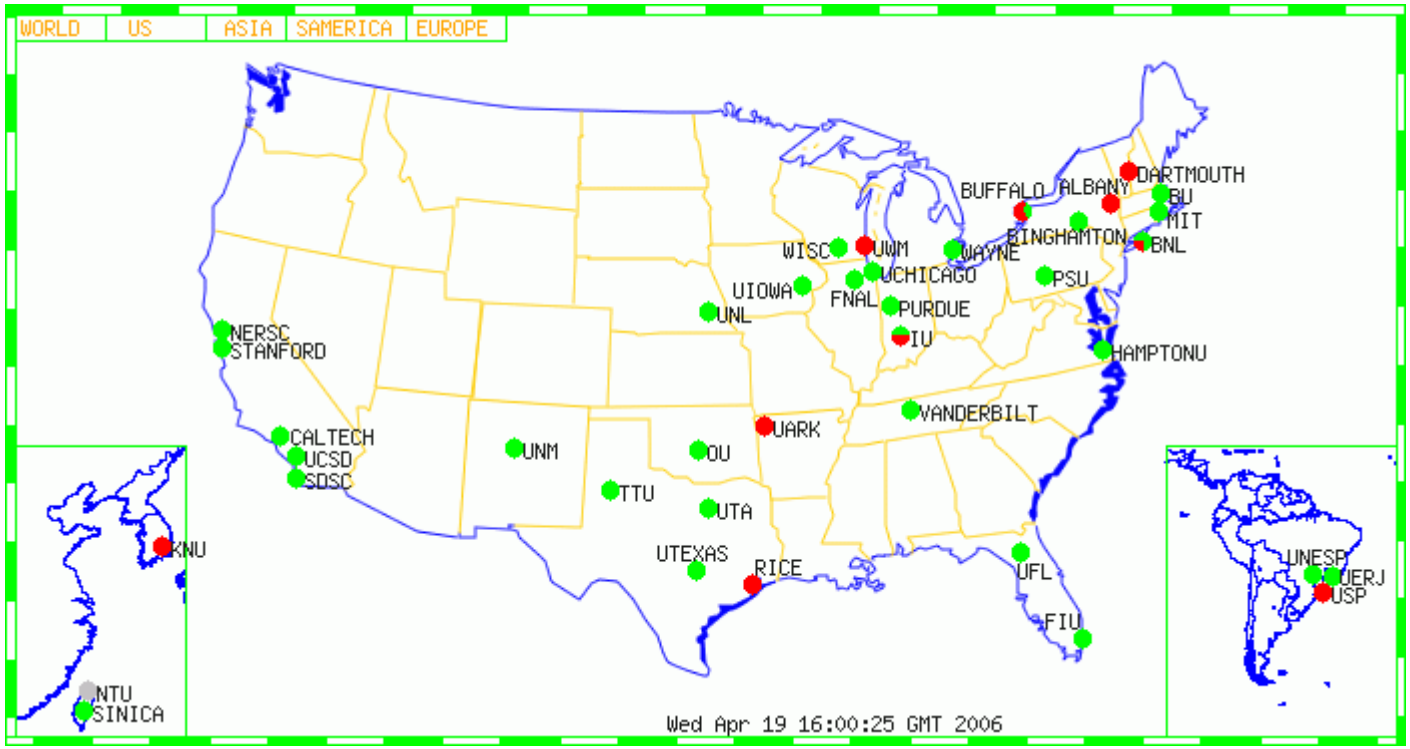
```
universe = globus
globusscheduler = ldas-grid.ligo.caltech.edu/jobmanager-condor
log = ldm_test_condor.log
output = ldm_test_condor.out
error = ldm_test_condor.err
should_transfer_files = YES
when_to_transfer_output = ON_EXIT
transfer_input_files = ldm_test_CIT_multilist.txt, ldm_test_CIT_filelist1.txt,
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

Open Science Grid (OSG)



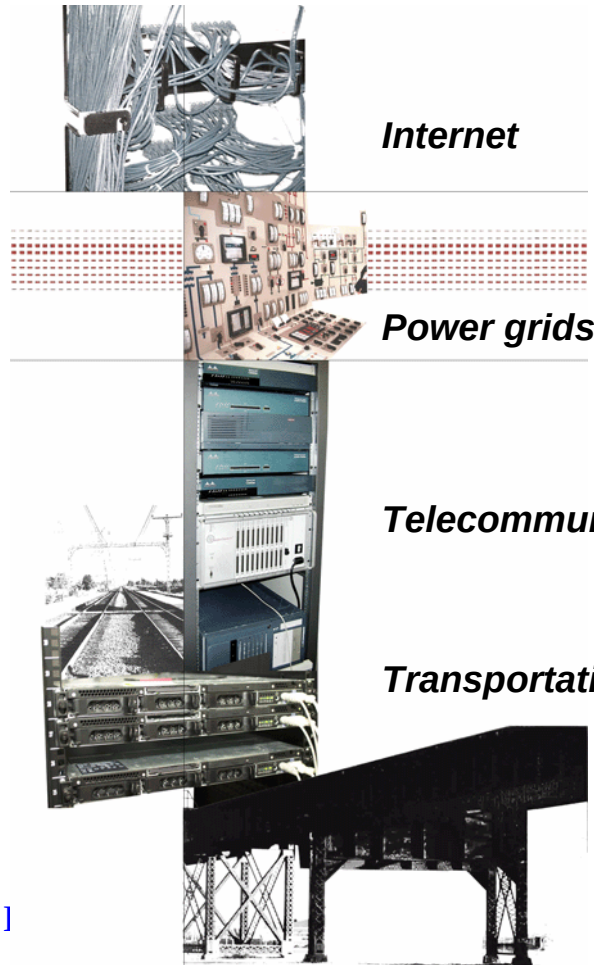
- Grid of Grids
- 20 thousand CPUs
- Petabytes of data storage

- High energy physics
- Bioinformatics
- Nanotechnology
- ……

Cyberinfrastructure Vision

- Providing high-end services – HPC, video conferencing, visualization, digital libraries, collaboration, …
- Infrastructuralizing high-end resources – supercomputers, data archives, telescopes, observatories, …
- More technical challenges: security, interaction, cross-VO, data transfer, fault tolerance, QoS …

Cyberinfrastructure – NSF's vision of 21st century's infrastructure



Technical Challenges

- Open vs. dedicated resources
- Applications are various
- Scientific applications are always evolving
- Cross-domain resource management and scheduling
- Cross-VO security (authentication and authorization) and trust management
- Dynamic VO construction
- Data availability
- Hybrid resource management
- QoS using virtual machines
- Too many to be exhausted …

Summary

- LIGO already heavily relies on a large-scale computing infrastructure
- Cyberinfrastructure implementation is still facing major technical challenges

References:

- http://www.ligo.caltech.edu
- http://www.ligo.org
- http://www.lsc-group.phys.uwm.edu/lscdatagrid/
- http://www.opensciencegrid.org
- http://www.nsf.gov/oci