

---

# End-to-end thumbnail of computing model and needs

Patrick R Brady

University of Wisconsin-Milwaukee

LIGO Scientific Collaboration

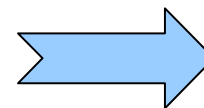
- The combination of LSC computational and data storage resources with grid-computing middleware to create a distributed gravitational-wave data analysis facility is the LIGO Data Grid.



- Compute centers at
  - » LIGO Hanford Observatory
  - » LIGO Livingston Observatory
  - » Tier-1: Caltech
  - » Tier-2: MIT, UWM & PSU
- Other clusters in Europe
  - » Birmingham, Cardiff and the Albert Einstein Institute (AEI)
- Grid Computing software
  - » e.g. Globus, GridFTP, and Condor and tools built from them

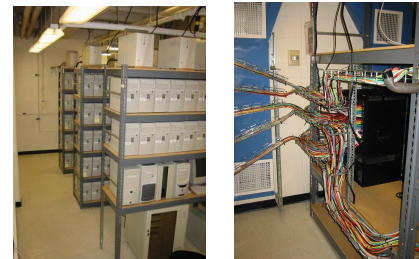
- Low latency analysis is needed if we want opportunity to provide alerts to astronomical community in the future
- Maximum scientific exploitation requires data analysis to proceed at same rate as data acquisition
- Requirements for flagship searches ( $\sim 2$  x fudge-factor)
  - » Stochastic = 1 unit (3 Ghz workstation day per day of data)
  - » Bursts = 50
  - » Compact binary inspiral = 600 (BNS), 300 (BBH), 6,000 (PBH) .....
  - » All sky pulsars = 1,000,000,000 (but can tolerate lower latency & .....

*LIGO's scientific pay-off is bounded  
by the ability to perform  
computations on the data.*



LIGO Data Grid

- Pre 2000:
  - » Commodity cluster computing shown to be ideally suited to LIGO data analysis needs in prototype analysis
  - » Trade study shows that clusters also provide best performance per dollar spent for LIGO data analysis
- 2000:
  - » Grid Physics Network (GriPhyN) funded via ITR program; LIGO is one of the founding experiments
  - » R&D program to prototype and develop grid-computing paradigm for data intensive experiments; LIGO portion funds development of LIGO Data Replicator
  - » UWM deploys Medusa cluster (funded by MRI) “a system for quick turnaround exploration, and development”



- 2001:

- » International Virtual Data Grid (iVDGL) funded via ITR program
- » Deployment of a Grid test bed for data intensive experiments
- » LIGO portion funds deployment of Tier 2 center at PSU and enhancement of storage capabilities at UWM

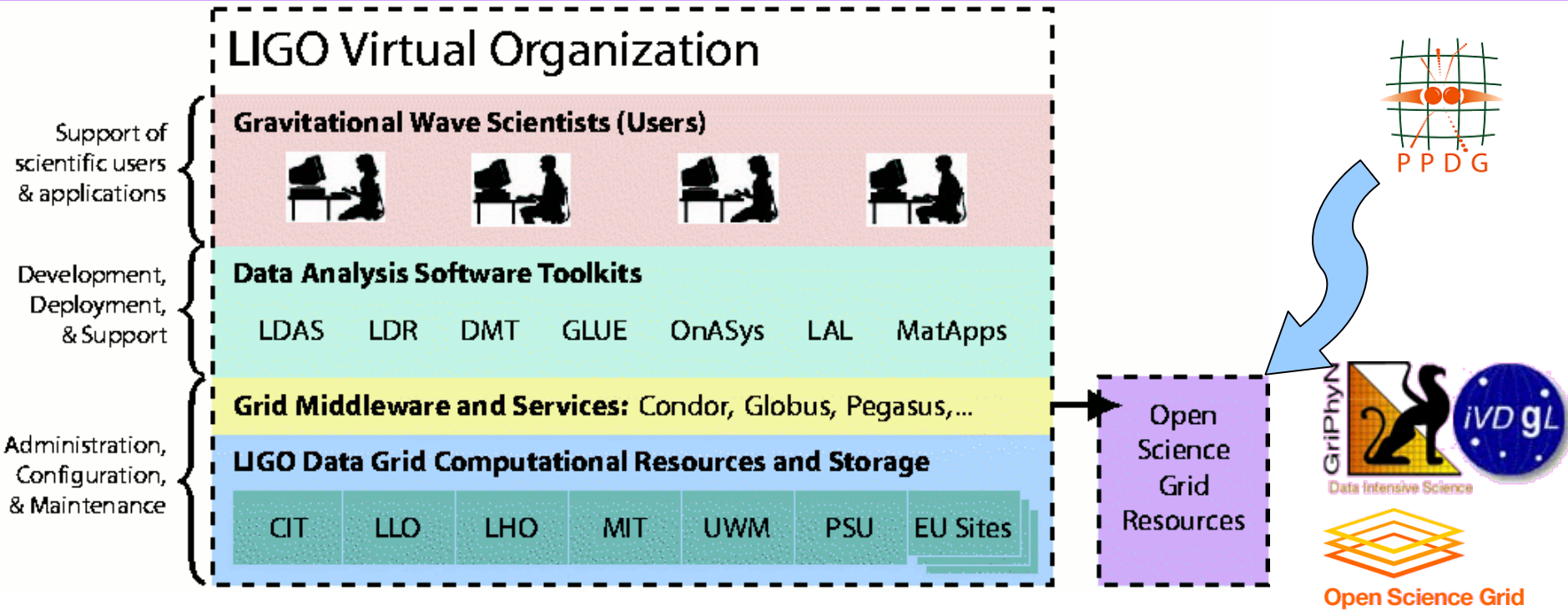


- 2003:

- » “Deploying the LIGO Data Grid; Grid-enabling the GW community” proposal by the LSC to transition from R&D to production deployment and use of the LIGO Data Grid.

- LIGO Data Grid now:

- » Consists of 2000 (US) and 1000 (EU) CPUs with total peak performance ~5 TFLOPS, 2 TB RAM, and 500 TB of distributed mass storage, in addition to 1.2 PB of tape storage at Caltech.
- » Provides dedicated computing support for data-intensive gravitational-wave research by 200 scientists of the LSC.



- **Cyberinfrastructure for the LIGO VO**

- » Hardware - administration, configuration, maintenance
- » Grid middleware & services - support, admin, configuration, maintenance
- » Core LIGO analysis software toolkits – support, enhance, release
- » Users - support

- Hardware and Operating System Maintenance
  - » Commodity hardware running Linux; track changes & enhancements
- Grid Middleware Administration
  - » Deploy LIGO Data Grid Server, configure Condor, LDR & other services.
- Data Distribution and Storage
  - » SAM-QFS, commodity storage on nodes
  - » LIGO Data Replicator to transfer data onto clusters before jobs are scheduled.
- User Support
  - » This is a big job because we have many inexperienced users who are prototyping analyses for the first time ever



- LIGO Certificate Authority
  - » Under development, needs long term personnel commitment
- Problem tracking and security
  - » Crude problem tracking in place, needs effort to make useful
- Virtual organization management service
  - » With 200 users, this is an essential service
- Metadata services
  - » Data catalogs, instrument quality information, resource information ..
  - » Need better resource monitoring
- Data Grid Server/Client bundles built on VDT
  - » Bundling of tools for users and admins on LIGO Data Grid



# LIGO Data Grid needs

- Need personnel committed multi-years to support, enhance and deploy this dedicated cyber-infrastructure
- System administration: 12 FTEs
- LIGO Data Grid Services: 8 FTEs
- Core analysis software: 14 FTEs
- Analysis profiling/help: 1 FTE

Fiscal Year ( Starts 1 Oct. of previous year)

Task	2007		2008		2009		2010		2011	
	Req.	Curr.	Req.	Curr.	Req.	Curr.	Req.	Curr.	Req.	Curr.
System Administration	3	9	5	7	5	7	5	7	5	7
LIGO Data Grid Services	4	4	7	1	7	1	7	1	7	1
Core Analysis Software	5	9	9	5	9	5	9	5	9	5
Analysis Profiling/Help	1	0	1	0	1	0	1	0	1	0
Totals	13	22	22	13	22	13	22	13	22	13
Total Need	35		35		35		35		35	

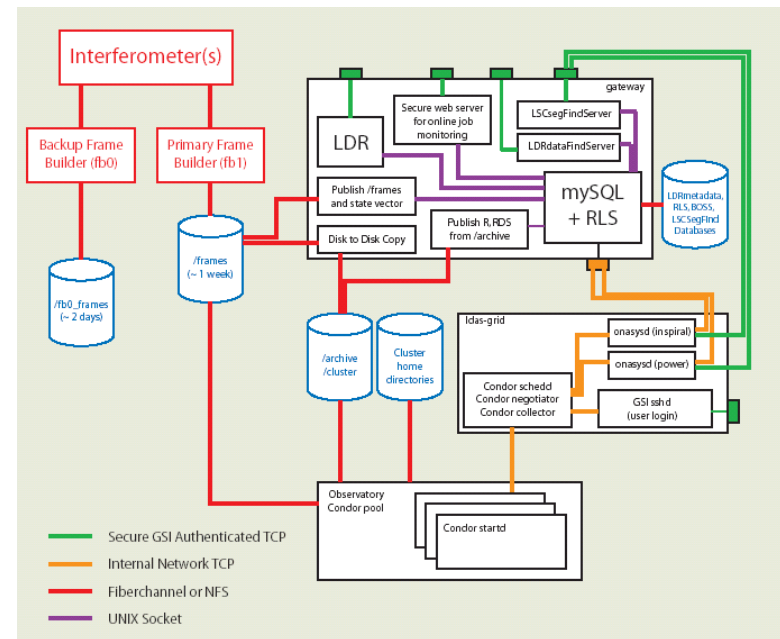
○ PIF Proposal to Fund Positions

○ LIGO Ops Funded Positions

# LIGO Grid LSC User Environment (GLUE)

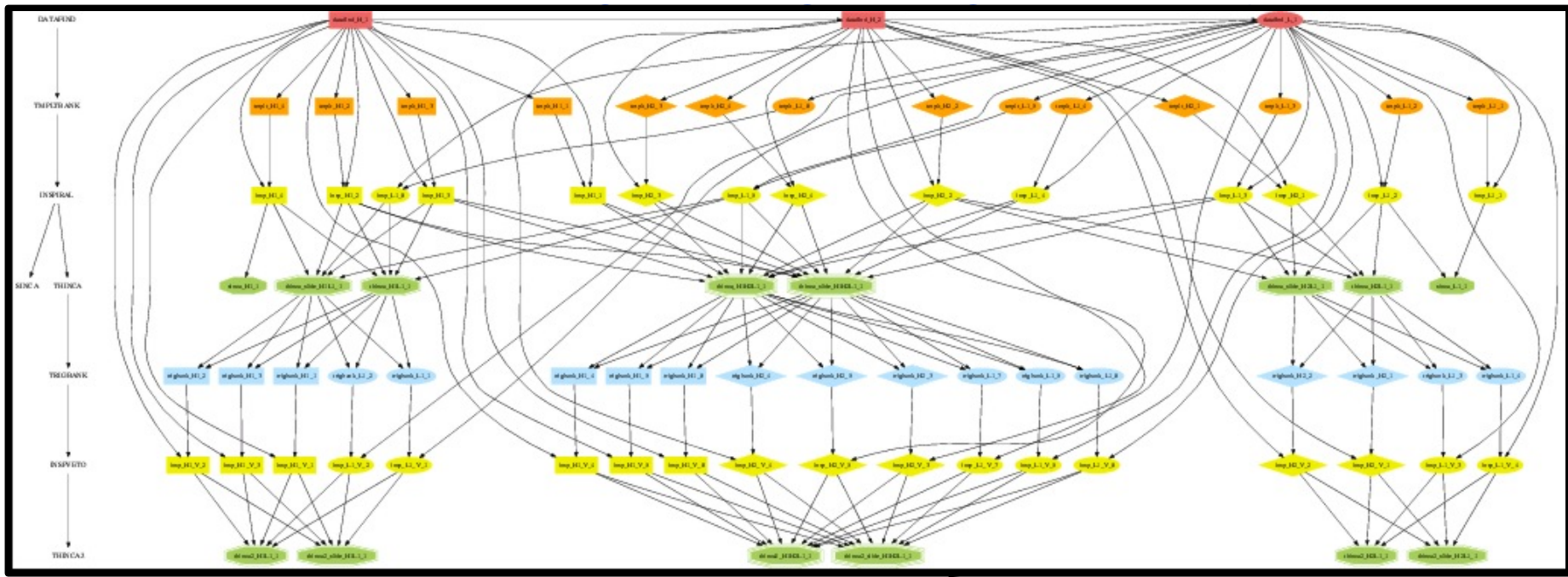


- Provides high-level infrastructure for running applications on the LIGO Data Grid
  - » provides an infrastructure to simplify the construction of workflows by treating data analysis applications as modules to be chained together.
  - » use of metadata (e.g. data quality information) allows complicated workflows to be easily constructed.
  - » contains certain LSC specific metadata clients and servers, such as data discovery tools.

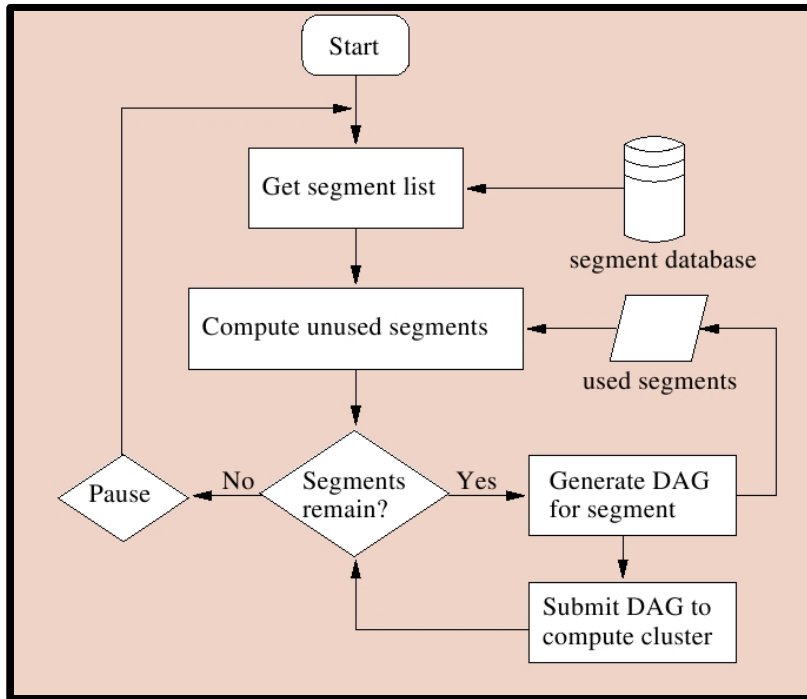


- Complicated workflows

- » to perform all steps to search data from four LSC detectors
- » workflow generation built on top of GLUE & LALApps analysis codes



Part of binary inspiral workflow  
full analysis workflows have over 10,000 nodes



- How it works

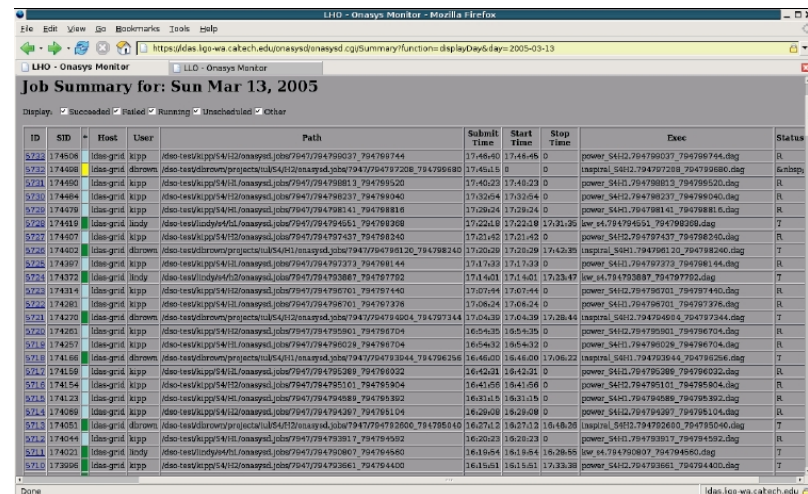
- » Identify data to analyze by query to GSI-authenticating instrument status & data quality service
- » Find data with LSCdataFind
- » Configure analysis pipeline using user defined pipeline construction tool
- » Execute on the grid

- ONline Analysis SYStem

- » Tools to automate real time analysis of GW data
- » Built on top of GLUE
- » Uses scientific data analysis pipelines from LSC users

- Built on top of Condor, GLUE, Globus
- Database of job information maintained to track progress through workflow
- Online monitoring via a web interface which queries job information metadata database.

- LDG is a lean effort; LIGO specific software is built on Condor & VDT ...
- ... but has relied on much volunteer effort which cannot be sustained .....



ID	STD	Host	User	Path	Submit Time	Start Time	Stop Time	Exec	Status
5732	174508	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947950037_794795744	17:46:40	17:46:46	0	power_S412.7947950037_794795744.dag	RL
5732	174498	lides-gnd	dbrown	ldao-tes/dbrown/project/ld54h12/onasysd.job/7947/7947952008_7947956880	17:46:15	0	0	nsprml_S401.7947952008_7947956880.dag	RL
5731	174450	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794795881_794795520	17:40:23	17:40:23	0	power_S411.794795881_794795520.dag	RL
5730	174464	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794798237_794799040	17:32:04	17:32:04	0	power_S412.794798237_794799040.dag	RL
5730	174479	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794798814_794798816	17:29:24	17:29:24	0	power_S4H1.794798814_794798816.dag	RL
5728	174187	lides-gnd	lindy	ldao-tes/lindy/h2/onasysd.job/7947/794794561_794799288	17:22:18	17:22:18	17:21:26	lsc_e4.794794561_794799288.dag	7
5727	174107	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794797373_794799340	17:21:42	17:21:42	0	power_S412.794797373_794799340.dag	RL
5726	174002	lides-gnd	dbrown	ldao-tes/dbrown/project/ld54h12/onasysd.job/7947/794798120_794798240	17:20:29	17:20:29	17:14:25	nsprml_S401.794798120_794798240.dag	7
5725	174397	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794797373_794798144	17:17:33	17:17:33	0	power_S412.794797373_794798144.dag	RL
5724	174372	lides-gnd	lindy	ldao-tes/lindy/h2/onasysd.job/7947/7947973887_794797792	17:14:01	17:14:01	17:33:47	lsc_e4.7947973887_794797792.dag	7
5724	174334	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947976701_794797440	17:07:44	17:07:44	0	power_S412.7947976701_794797440.dag	RL
5722	174261	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947976701_794797378	17:08:24	17:08:24	0	power_S411.7947976701_794797378.dag	RL
5721	174276	lides-gnd	dbrown	ldao-tes/dbrown/project/ld54h12/onasysd.job/7947/7947976004_794797344	17:04:39	17:04:39	17:28:41	nsprml_S401.7947976004_794797344.dag	7
5720	174202	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947976001_7947976704	16:54:35	16:54:35	0	power_S412.7947976001_7947976704.dag	RL
5719	174257	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947950238_794798704	16:54:32	16:54:32	0	power_S411.7947980238_794798704.dag	RL
5718	174160	lides-gnd	dbrown	ldao-tes/dbrown/project/ld54h12/onasysd.job/7947/7947973044_794798256	16:46:00	16:46:00	17:06:25	nsprml_S401.7947973044_794798256.dag	7
5717	174158	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794795269_794795902	16:42:31	16:42:31	0	power_S412.794795269_794795902.dag	RL
5716	174154	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794795101_794795904	16:41:55	16:41:55	0	power_S412.794795101_794795904.dag	RL
5715	174123	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794794989_794799392	16:31:15	16:31:15	0	power_S4H1.794794989_794799392.dag	RL
5714	174086	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947943497_794795104	16:29:08	16:29:08	0	power_S412.7947943497_794795104.dag	RL
5713	174051	lides-gnd	dbrown	ldao-tes/dbrown/project/ld54h12/onasysd.job/7947/7947976000_794798504	16:27:12	16:27:12	16:48:35	nsprml_S401.7947976000_794798504.dag	7
5712	174044	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/794797381_794794592	16:20:23	16:20:23	0	power_S411.794797381_794794592.dag	RL
5711	174021	lides-gnd	lindy	ldao-tes/lindy/h2/onasysd.job/7947/7947970807_794794580	16:19:54	16:19:54	16:28:55	lsc_e4.7947970807_794794580.dag	7
5710	173998	lides-gnd	krpp	ldao-tes/kpp54h12/onasysd.job/7947/7947973661_794794400	16:15:43	16:15:43	17:33:38	power_S412.7947973661_794794400.dag	7