



LIGO Data Grid

Patrick R Brady
University of Wisconsin-Milwaukee

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



- 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

Why do we need the LIGO Data Grid

- 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
- Computers required for flagship searches
 - » 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.

- LIGO Data Grid (LDG) supports ~200 LSC scientists
 - » Demand for resources is growing rapidly as experience increases and more data become available

```

inspiral_pipe.ini (~/projects/lalapps/src/inspiral) - VIM
[datafind]
; use only level 1 rds data
type = RDS_R_L1

[data]
; data conditioning parameters common to tmplbank and inspiral
pad-data = 8
segment-length = 1048576
number-of-segments = 15
sample-rate = 4096
resample-filter = ldas
enable-high-pass = 100.0
high-pass-order = 8
high-pass-attenuation = 0.1
spectrum-type = median
    
```

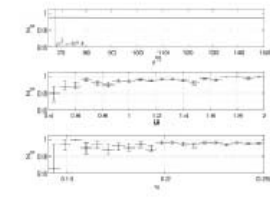
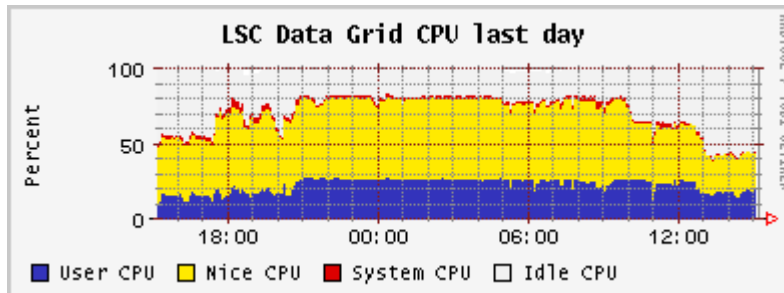
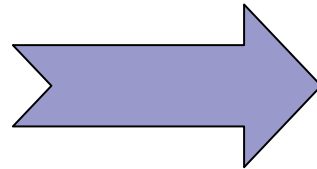


FIG. 5. The top panel shows the sensitivity in MW units \mathcal{N}_{100} of the search with a signal population as a function of the total mass M . The larger SNR observed in this analysis was $\mathcal{N}_{100} = 0.1$ meaning that the search is sensitive to a fraction $\mathcal{N}_{100}^{-2} = 0.01$ of the total. The middle panel shows \mathcal{N}_{10} as a function of total mass M . The search is sensitive to a fraction $\mathcal{N}_{10}^{-2} = 0.01$ of the total. The bottom panel shows \mathcal{N}_1 as a function of total mass M . The search is sensitive to a fraction $\mathcal{N}_1^{-2} = 0.01$ of the total. We note that the efficiency is a weak function of the total mass, as the amplitude of the inspiral signal is a function of the total mass. The efficiency of the search does not depend strongly upon \mathcal{M} .

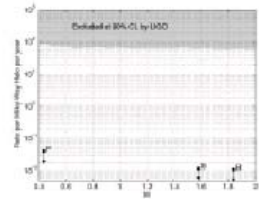


FIG. 6. The stacked region search rate was calculated at 80% confidence by the observational upper limit as PBH binary coalescence generated in this paper as a function of total mass $M = m_1 + m_2$ of the binary. The three points show the calculated rate using Eq (25) for the black hole masses $m_1 = 1.0 M_\odot$, $m_2 = 1.1 M_\odot$, and $m_1 = m_2 = 1.0 M_\odot$ of [9].

$$R_{\text{PBH}} = 81 \gamma^2 \mathcal{M}^2 \text{yr}^{-1}. \quad (4)$$

$\mathcal{N}_{10} = 0.01$ MW. The relative contribution to the error in the measured value of detection rate using our detection is defined in [2]. It is necessary that the systematics errors are also accounted in the observational upper limits, as the systematic errors due to differences between the simulated signal and the Monte Carlo simulation. In this analysis, we neglect errors due to the spatial distribution of the PBH binaries as studies show that the population is relatively uniform over the shape of the Milky Way halo. This is because the maximum range of all three detectors is greater than 50 kpc for PBH masses $\geq 0.1 M_\odot$. The systematic error is still affected the rate through the assumed SNR of the binary event. We note that the efficiency of the search depends only weakly on the SNR of the background signal due to the range of the search compared to the halo radius. The statistical errors in the Monte Carlo analysis also affect the error in Eq. The one found error due to systematic search and the calibration uncertainty is about 0.1×10^{-4} MW. The effect of spin was ignored both in the population and in the waveform used to detect inspiral signals. Estimates based on the work of Agostinelli [21] suggest that the maximum likelihood signal from spinning PBHs induces a bias that is at least a factor of 2 smaller than the population. To be conservative, however, we will use an upper limit only a factor of 2 higher. PBHs may still induce this error systematically in future analysis. Combining the errors in quadrature and assuming the detection efficiency of \mathcal{N}_{10} to be conservative, we obtain an observational upper limit on the rate of PBH binary coalescence with two

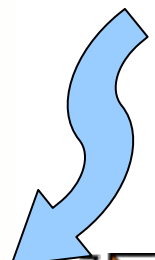
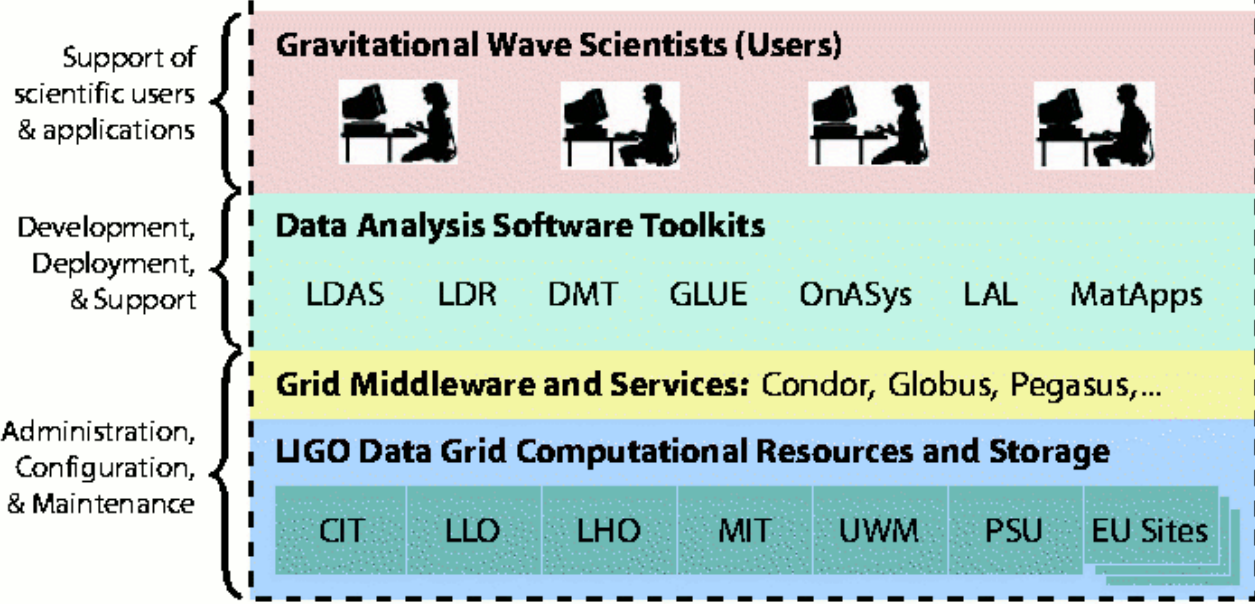
By considering statistical uncertainties of three-body PBH formation in the early universe, Iida et al. [15] obtain a probability distribution for the formation rate and coalescence rate of PBH binaries. This depends on the PBH mass in which we assume to be the MACHO mass. From this distribution, we may obtain an estimate of the rate of PBH coalescence at the present time, given by

$$R = 4 \times 10^{11} \left(\frac{M}{M_\odot} \right)^{-2} \left(\frac{h}{h_0} \right)^{-2} \text{yr}^{-1} \text{MW}^{-1} \quad (5)$$

where h is the MACHO mass and h_0 is the rate of the halo in MW units, which is obtained from assuming spherical symmetry. There is a small variation in the halo model used in the analysis of the coalescence results [25, 26]. The rate is set in Eq. (5) corresponds to model B of the MACHO Collaboration [25]. The rate including a barrier and PBH formation results, assume a 3-body halo size distribution, as also the rate estimate in Eq. (5). We use one from Fig. 2 that our calculation efficiency is not strongly dependent on the rate of the binary masses, m , and so we can neglect this parameter to obtain the rate as a function of total PBH mass \mathcal{M} , which was compared with the predicted rates from our remaining 14 different halo models. The analysis of 5.7 yrs of observation of 11.5 million stars in the LMC suggests a MACHO mass of $m = 0.29^{+0.02}_{-0.01}$ and a MACHO mass $\mathcal{M} = 10^{22} \pm 10^{23} M_\odot$. For halo model B [15]. Accounting all the MACHOs and PBHs, we obtain the rate estimate

- LIGO production computing tests grid concepts at both the high level of ideas & at the detailed level of tools.
- Example collaborative research with comp. sci.:
 - » LIGO runs a very large dedicated Condor pool with a complex usage model that stress tests Condor & identifies areas that need to be more robust – excellent collaboration with Condor Team.
 - » LIGO's environment has identified the need for Pegasus to better support rapidly changing executables & to better integrate internal & external resources to appear seamless to the user – ISI Group
 - » LIGO's needs led to improved scaling of the Globus Replica Location Service (RLS) to handle many millions of logical filenames, improved reliability of the server, and an improved API for faster publication of available data and queries – Globus Team

LIGO Virtual Organization



Open Science Grid Resources



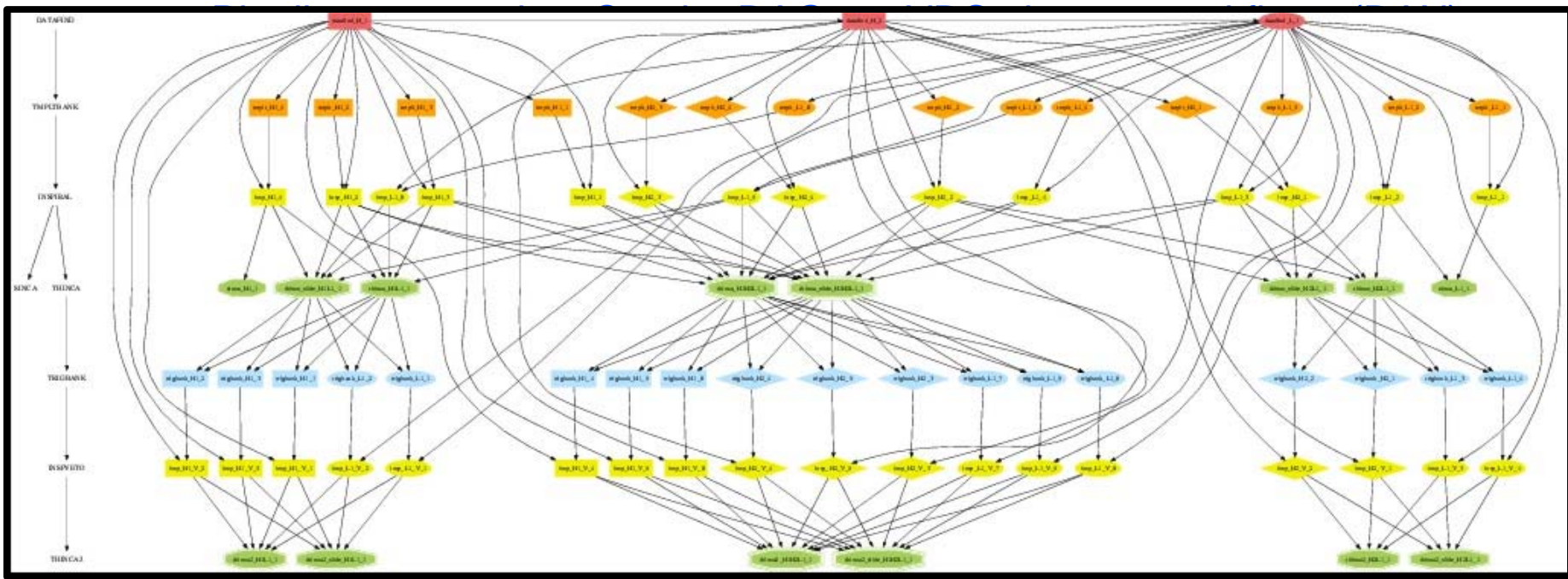
- Cyberinfrastructure for the LIGO VO
 - » Hardware - administration, configuration, maintenance
 - » Grid middleware & services - support, admin, configuration, maintenance
 - » Core LIGO analysis software toolkits – support, enhance, release
 - » Users - support

- 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

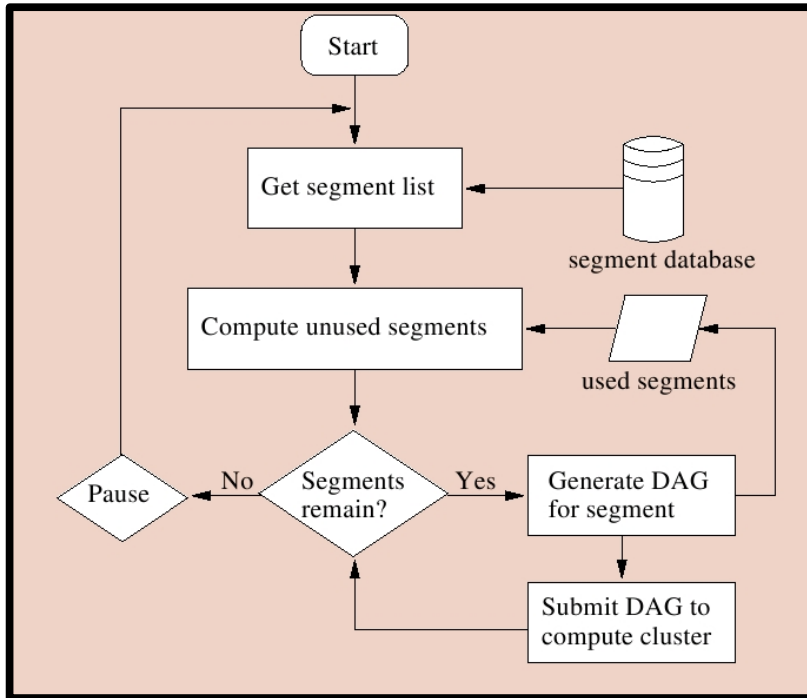
- LIGO Certificate Authority
 - » Under development, needs long term personnel commitment
- Problem tracking and security
 - » Crude problem tracking in place, needs effort to make useful
- Virtual organization management service
 - » With 200 users, this is an essential service
- Metadata services
 - » 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

- 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

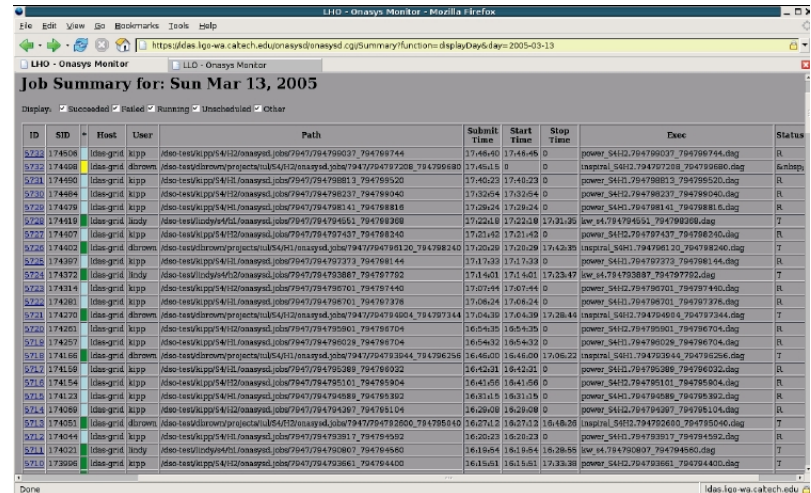
- » Identify data to analyze by query to GSI-authenticating instrument status & data quality service
- » Find data with LSCdataFind
- » Configure analysis pipeline using user defined pipeline construction tool
- » Execute on the grid

- 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



Job Summary for: Sun Mar 13, 2005

ID	STD	Host	User	Path	Submit Time	Start Time	Stop Time	Exec	Status
5232	174508	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947724792623_794769744	17:46:40	17:46:46	0	power_54H2.794795037_794769744.diag	R
5233	174408	lido-gnd	dlrcwm	ldao-tes/dlrcwm/project/ld54/f12/onasys1_job77947794797208_794799680	17:45:15	0	0	insprtl_54H2.794797228_794799680.diag	6:ntstp
5234	174450	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794798813_794795520	17:45:23	17:48:23	0	power_54H2.794798813_794795520.diag	R
5235	174484	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794798237_794799040	17:32:04	17:32:04	0	power_54H2.794798237_794799040.diag	R
5236	174478	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794798141_794798816	17:28:24	17:28:24	0	power_54H2.794798141_794798816.diag	R
5237	174510	lido-gnd	lidy	ldao-tes/lidy4/f2/onasys1_job779477947984651_794798308	17:22:18	17:22:18	0	lve_4.794798465_794798308.diag	7
5238	174407	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794797437_794798340	17:21:45	17:21:45	0	power_54H2.794797437_794798340.diag	R
5239	174025	lido-gnd	dlrcwm	ldao-tes/dlrcwm/project/ld54/f12/onasys1_job779477947979120_794798824	17:20:28	17:20:28	0	insprtl_54H1.794799120_794798824.diag	7
5240	174367	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794797273_794798144	17:11:33	17:11:33	0	power_54H2.794797273_794798144.diag	R
5241	174372	lido-gnd	lidy	ldao-tes/lidy4/f2/onasys1_job779477947973867_794797702	17:14:01	17:14:01	0	lve_4.794797387_794797702.diag	7
5242	174314	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job779477947976701_794797440	17:07:44	17:07:44	0	power_54H2.7947976701_794797440.diag	R
5243	174281	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job779477947976701_794797378	17:08:24	17:08:24	0	power_54H2.7947976701_794797378.diag	R
5244	174279	lido-gnd	dlrcwm	ldao-tes/dlrcwm/project/ld54/f12/onasys1_job779477947974604_794797344	17:04:10	17:04:30	17:28:44	insprtl_54H2.794797464_794797344.diag	7
5245	174257	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job779477947976704	16:54:35	16:54:35	0	power_54H2.7947976701_7947976704.diag	R
5246	174257	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794798028_794798704	16:54:32	16:54:32	0	power_54H2.794798028_794798704.diag	R
5247	174166	lido-gnd	dlrcwm	ldao-tes/dlrcwm/project/ld54/f12/onasys1_job7794779479793044_794798628	16:46:00	16:46:00	17:06:22	insprtl_54H1.7947993044_794798628.diag	7
5248	174158	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job7794779479795289_794798022	16:42:01	16:42:01	0	power_54H2.79479795289_794798022.diag	R
5249	174154	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job7794779479795101_794798504	16:41:58	16:41:58	0	power_54H2.79479795101_794798504.diag	R
5250	174123	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job7794779479794988_794798302	16:33:15	16:33:15	0	power_54H2.79479794988_794798302.diag	R
5251	174088	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job7794779479794267_794798104	16:28:08	16:28:08	0	power_54H2.79479794267_794798104.diag	R
5252	174025	lido-gnd	dlrcwm	ldao-tes/dlrcwm/project/ld54/f12/onasys1_job7794779479792600_794798540	16:27:12	16:27:12	18:14:38	insprtl_54H2.7947992600_794798540.diag	7
5253	174044	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job77947794797939817_794798992	16:20:23	16:20:23	0	power_54H2.794797939817_794798992.diag	R
5254	174022	lido-gnd	lidy	ldao-tes/lidy4/f2/onasys1_job7794779479790807_794797480	16:19:54	16:19:54	18:28:58	lve_4.79479790807_794797480.diag	7
5255	173906	lido-gnd	kjpp	ldao-tes/ldpp54/f12/onasys1_job7794779479793661_794797440	16:13:53	16:13:51	17:33:38	power_54H2.7947973661_794797440.diag	7

- The LDG has enabled 13 results papers on searches for gravitational waves from
 - » binary neutron star and black hole systems,
 - » isolated spinning neutron stars
 - » a gravitational stochastic background from the early universe
 - » gravitational-wave bursts from supernovae or other energetic events
- The LDG has also enabled more than 17 technical data analysis papers
- Six (6) students have received PhDs based on results obtained using these resources; Fifteen (15) more are in pipeline

- 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