LIGO Laboratory Operations

Plan for the Data Analysis & Computing Group

NSF Review of the LIGO Operations Proposal for FY2002 - 2006

LIGO Hanford Observatory

Hanford, Washington

26 February 2001

Albert Lazzarini

LIGO Laboratory Caltech

Outline

- Data analysis & computing group organization,functions
- Staffing
- Budgets

Data and Computing Group Roles

- Throughout Engineering & Science Runs, the Laboratory's Data Analysis & Computing Group fulfills the following roles:
  - » LIGO science, data analysis: scientific staff are actively engaged in the astrophysics searches
  - » Simulation & Modeling: detector support, data analysis
  - » Continuous management and movement of large volumes of data
  - » Maintaining pipeline analyses running, archive running
  - » Software maintenance/improvements/enhancements
  - » LSC support, visitors
  - » LIGO Laboratory-wide IT support

Data and Computing Group: Organization

- Three subgroups:
  - » LIGO Data Analysis System Group (LDAS)
  - » Modeling and Simulation Group
  - » General Computing Group (IT support)

Data & Computing Group Activities

Example LIGO Data Analysis System Group Activities

- Early upper limit searches
  - » Acquire, archive, process data from early engineering runs to set observational upper limits on gravitational wave source rates and strengths, etc.
  - » Complete software environment for pipeline analyses through mock data challenges (MDCs) jointly within LSC
- LIGO I Science Run astrophysical searches
  - » Data analysis will become a key "business" of the Laboratory across all sites
    - – Participation by LIGO scientists in astrophysical searches through the LSC, including use of multiple interferometer data streams (detector network)
    - – Database, archive use & maintenance
    - – Pipeline analysis SW and HW maintenance
    - – Continued R&D into algorithms, new technologies

Data & Computing Group Activities

Example LIGO Data Analysis System Group Activities

- LIGO Data Analysis System & Environment SW Support:
  - » LDAS code base is >600,000 lines of C++ and Tcl/Tk now -- will grow to ~1,000,000 lines by time of LIGO I Science Run.
  - » Frame data archive now includes ~ 5 TB of raw data.
    - – Will grow at 750 GB/day once science run starts.
    - – HPSS archive at Caltech will accommodate 500+TB
  - » Relational databases expected to grow to 1TB over course of science run
  - » Data are generated at remote sites, must be continually staged to main archived
  - » LDAS Documentation is entirely www based (http://www.ldas-sw.ligo.caltech.edu)

Data & Computing Group Activities

Example LIGO Data Analysis System Group Activities

- LIGO Data Analysis System & Environment HW Support:
  - » Disk farms
    - – 28TB of fibre-channel RAID 5 systems across 4 sites
    - – 50+TB of SCSI RAID 5 systems for data caching, mainly at Caltech near main archive
  - » 500+ TB HPSS installation at Caltech -- main archive
  - » Servers across all 4 LIGO sites:
    - – 15 enterprise class servers for database, data distribution
    - – 9 SMP Linux servers for data processing
  - » PC clusters across all 4 LIGO sites:
    - – 350+ PCs running Linux in 6 distinct clusters varying between 32 - 144 nodes.
  - » Homogenous controlled environment with few user accounts, little user services (e.g., no sendmail, no browsers, few connections to outside, etc.)

LIGO Data Analysis & Computing Group

LDAS FTE allocation for activities during operations

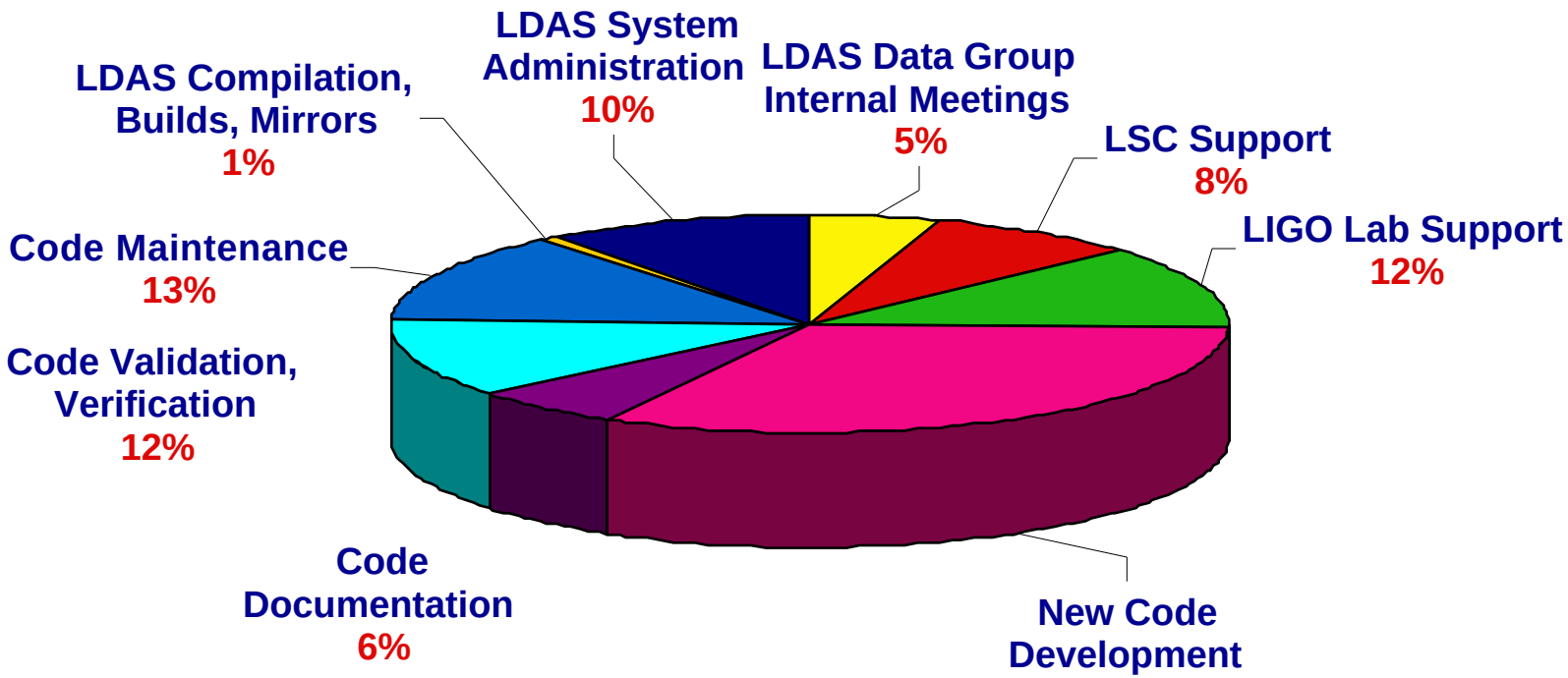
		LDAS Staffing Plan						
Personnel	FTE	LIGO Science Run; Algorithm Perf. Mon	LIGO Science Run; HW Perf. Mon	LIGO Science Run; Data Archival	LIGO Science Run; Data Analysis & Research	Operations: DB Mgmnt	Operations: Sys. Admin	SW Dev & Maint.
Graduate Student	2				2			
Postdoctoral Scholar	3	0.45	0.33	0.25	1.97			
Sci.	3.8	1.38	0.2	0.45	1.24			0.53
Sw. Eng	3.8	1.42						2.38
Contract SW Eng.	*1*							*1*
Contract DB Admin	*1*			*0.5*		*0.5*		
Contract HPSS Admin	*1*			*0.5*			*0.5*	
Contract LDAS-SysAdmin	*1*						*1*	
CIT Totals ->	**16.6**	**3.25**	**0.53**	**1.7**	**5.21**	**0.5**	**1.5**	**3.91**
Caltech Personnel Classifications	Students				2			
	Postdoctoral	0.45	0.33	0.25	1.97			
	Scientists	1.38	0.2	0.45	1.24			0.53
	SW Eng.	1.42						3.38
	SysAdmin			1		0.5	1.5	
LHO Sci*	1	0.2	0.2	0.2	0.2		0.2	
LLO Sci*	1	0.2	0.2	0.2	0.2		0.2	
LHO Eng..*	1			0.5		0.2	0.3	
LLO Eng..*	1			0.5		0.2	0.3	
LIGO Lab Totals ->	**20.6**	**3.65**	**0.93**	**3.1**	**5.61**	**0.9**	**2.5**	**3.91**

**** NOTE: LHO and LLO personnel are in the budgetary proposals from the respective sites***

LIGO-G010039-01-E

LDAS Operations - Professional Staff

Statistics derived from historical data



Breakdown of LDAS Workforce Activities

Data & Computing Group Activities

Example Modeling & Simulation Group Activities

- LSC science search participation
  - » Simulated data for data analysis, Monte Carlo simulation of detection efficiencies, etc.
- Commissioning support
  - » Provide realistic simulations to assist in improving interferometer performance up to and during science run
- Development of physics models
  - » Improvement of physical optics and mechanical systems models
- Code improvement & optimization
  - » Parallel computing, client-server model, etc
- Improvement of the User Interfaces
  - » Simplify the setting up of simulation sessions and viewing of results

Data & Computing Group Activities

Example Modeling & Simulation Group Activities

- LIGO End-to-End Simulation Environment SW Support:
  - » E2E consists of ~200k lines of C++
  - » Grouped into 80 primitive modules and 70 compound modules
  - » E2E documentation maintenance (http://www.ligo.caltech.edu/~e2e/)

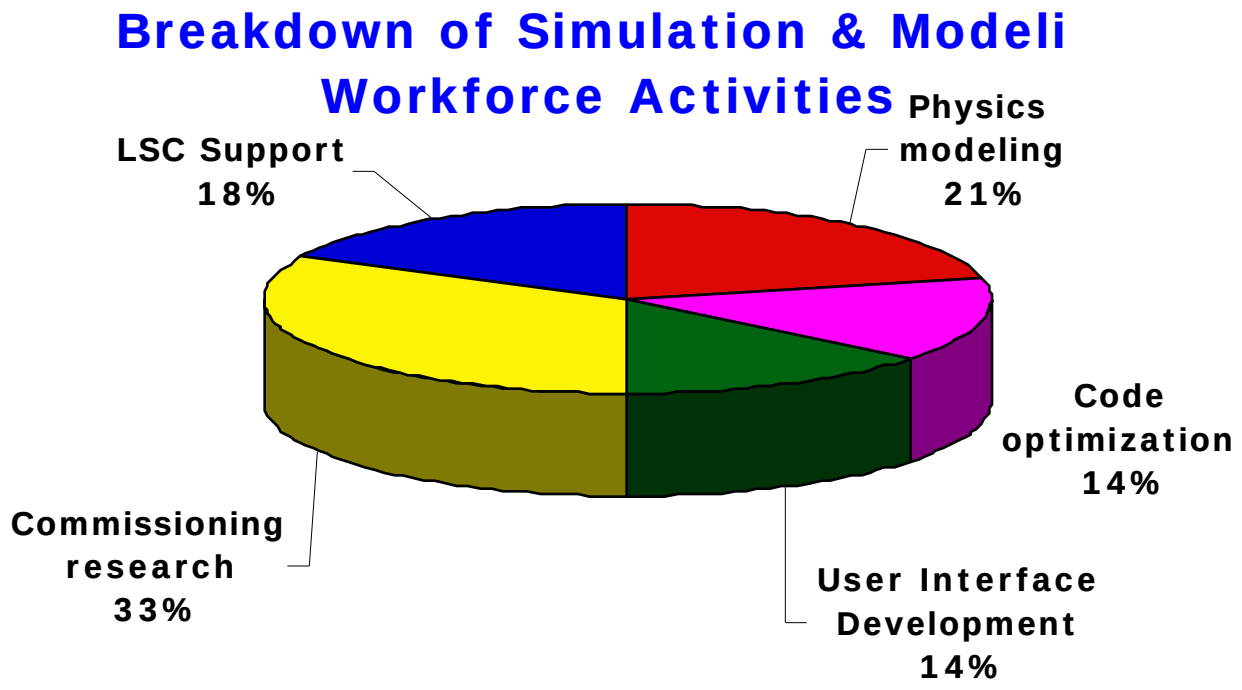
Data Analysis & Computing Group

Modeling & Simulation FTE activities during operations

		Modeling & Simulation Staffing Plan					
Personnel	FTE	Code environ. Development: Architecture	Code environ. Development: EM Fields, Optics	Code environ. Development: Mechanical systems	Code environ. Development: GUI Develop.	Detector Performance Modeling	Data Analysis Simulation
Graduate Student	2					1	1
Postdoctoral Scholar	1			0.5		0.5	
Sci.	2	0.25	0.75			0.75	0.25
Sw. Eng	1				1		
Contract SW Eng.	*1*	*0.75*		*0.25*			
CIT Totals ->	**7**	**1**	**0.75**	**0.75**	**1**	**2.25**	**1.25**
Caltech Personnel Classifications	Students					1	1
	Postdoctoral			0.5		0.5	
	Scientists	0.25	0.75			0.75	0.25
	SW Eng.	0.75		0.25	1		

Modeling and Simulation Operations

Estimates based on experience



Data & Computing Group Activities

Example General Computing Group Activities

- System Administration Activities & Support:
  - PC and PC server support: 75+ PCs, 5 NT servers
  - Unix systems support: 75+ Sun workstations
  - Unix Server support: 2 Enterprise servers, 6 smaller servers
  - Network Administration LSN (Caltech) and WAN (Laboratory-wide)
  - WEB administration for Laboratory and LSC functions
  - User and application support - 300+ user accounts backed up, maintained, Application licensing for unix servers, PCs, etc.
  - Meetings support (visitor workstations, DHCP, A/V equipment etc.)
  - Purchasing of all hardware, software for laboratory subgroups at Caltech
  - Document Control Center archive
  - Leadership and guidance to other Laboratory sites for IT support

Data Analysis & Computing Group

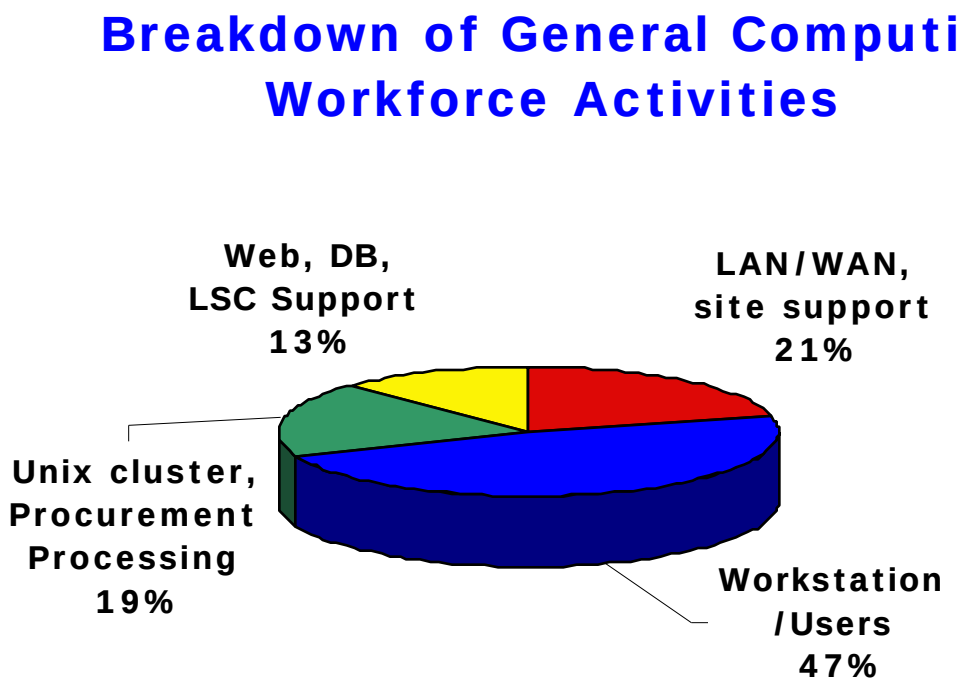
Principal General Computing activities during operations

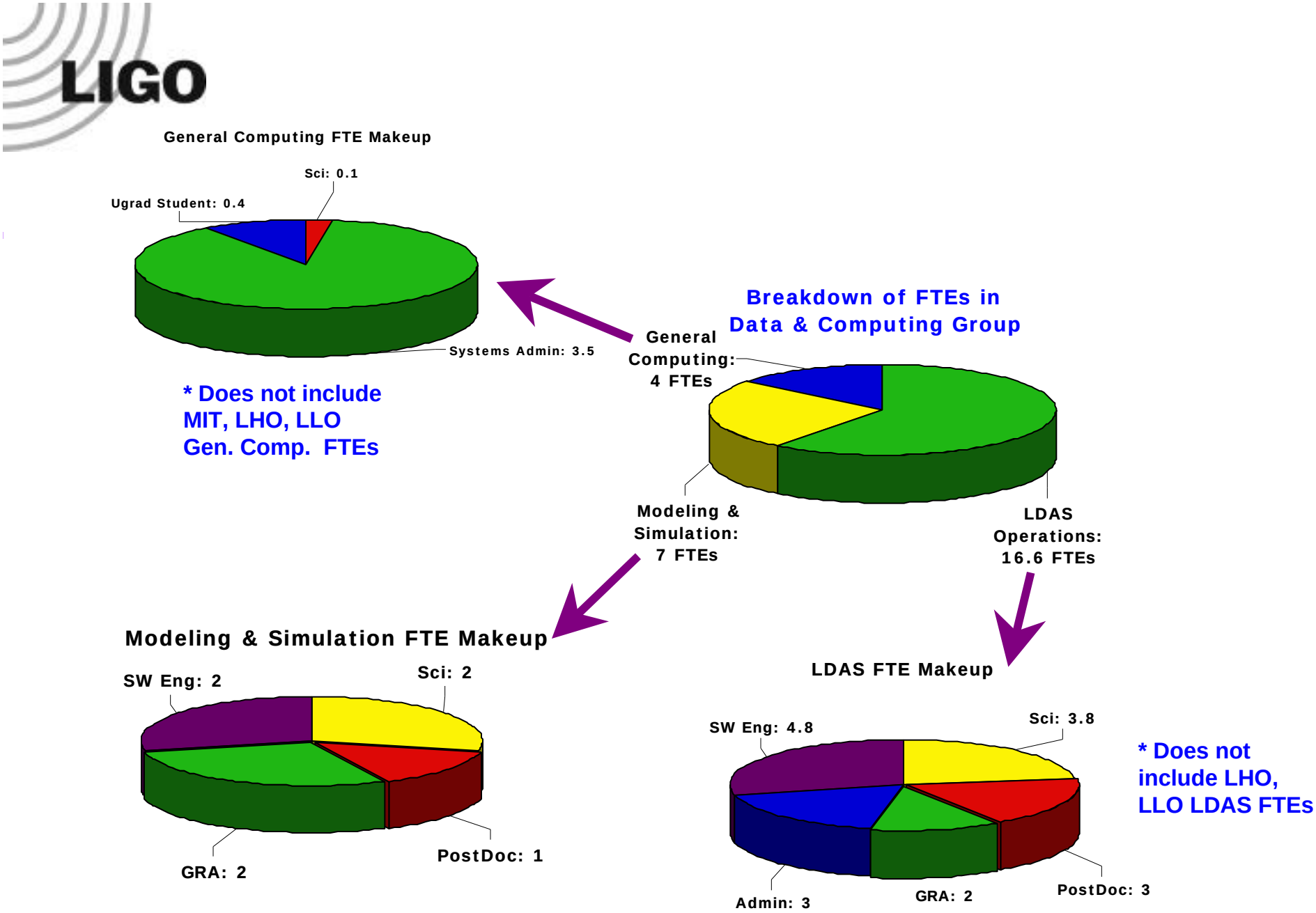
		General Computing Staffing Plan				
Personnel	FTE	Server & Data Systems	Desktop Workstations	LAN & WAN Administration	Telecommunications Administration	Web & DB Administration
Student Asst.	0.4		0.4			
www, DB Admin	0.5					0.5
Sci.	0.1			0.1		
Sr. Sys. Admin	1	0.2	0.15	0.3	0.35	
Asst. Sys. Admin	1	0.3	0.6	0.1		
Contract Sys. Admin.	*1*	*0.25*	*0.75*			
CIT Totals ->	**4**	**0.75**	**1.9**	**0.5**	**0.35**	**0.5**
Caltech Personnel Classifications	Students		0.4			
	DB Admin.					0.5
	Sci.			0.1		
	SysAdmin	0.5	0.75	0.4	0.35	
MIT Sys Admin*	1	0.3	0.6	0.1		
LHO Sys Admin*	1	0.15	0.5	0.1	0.13	0.12
LLO Sys Admin*	1	0.15	0.5	0.1	0.13	0.12
LIGO Lab Totals ->	**7**	**1.35**	**3.5**	**0.8**	**0.61**	**0.74**

**** NOTE: MIT, LHO, and LLO personnel are in the budgetary proposals from the respective sites***

General Computing Operations

Statistics derived from historical experience




General Computing FTE Makeup
Sci: 0.1
Ugrad Student: 0.4
Systems Admin: 3.5
* Does not include MIT, LHO, LLO Gen. Comp. FTEs
Breakdown of FTEs in Data & Computing Group
General Computing: 4 FTEs
Modeling & Simulation: 7 FTEs
LDAS Operations: 16.6 FTEs
Modeling & Simulation FTE Makeup
SW Eng: 2
Sci: 2
GRA: 2
PostDoc: 1
LDAS FTE Makeup
SW Eng: 4.8
Sci: 3.8
Admin: 3
GRA: 2
PostDoc: 3
* Does not include LHO, LLO LDAS FTEs

Increase for Full Operations

Budget Category	Increase	FY 2002	FY 2003	FY 2004	FY 2005	FY 2006
Basic Operations						
	* CDS Hardware Maintenance	513,800	502,434	517,507	533,032	549,023
	* LDAS Maintenance	1,378,728	1,378,728	1,322,235	1,303,163	1,303,163
	Outreach	249,848	257,343	265,063	273,015	281,206
	Site Operations	558,485	575,240	592,497	610,272	628,580
	* Telecommunications / Networking	540,500	542,200	542,200	539,500	539,500
	Staff for Site LSC Support	254,678	262,318	270,187	278,293	286,642
Basic Operations Totals		**3,496,039**	**3,518,263**	**3,509,689**	**3,537,275**	**3,588,114**
Operations Support of Advanced R&D						
	Seismic Development	506,300	434,574			
	Engineering Staff	920,868	948,494	976,949	1,006,257	1,036,445
	* Simulation & Modeling Staff	282,485	293,949	305,614	317,772	330,617
R&D Total		**1,709,652**	**1,677,017**	**1,282,562**	**1,324,029**	**1,367,062**
Grand Total		**5,205,691**	**5,195,280**	**4,792,252**	**4,861,304**	**4,955,176**

*** Need recognized by NSF Review Panel**

LIGO Data Analysis System
LDAS Hardware Investment

- LIGO Laboratory is establishing a distributed network of data analysis installations covering all sites
  - » Managed/maintained by LDAS group (not a general computing function)
  - » 10+ enterprise class SMP unix (sun) servers for data distribution
  - » 5+ PC linux SMP servers for pipeline data preprocessing
  - » 500TB robotic archive based on HPSS
  - » 30TB - 60TB RAID disk farms for data caching
  - » 5 separate linux PC clusters for parallel (MPI) pipeline analysis -- total of 350+ PCs
  - » Gigabit/fast ethernet switches, fibre channel switches
- Total plant investment will be ~$4M (with educational/GSA discounts, etc.)

Data & Computing Operations Budget

LDAS Hardware Maintenance Budgetary Estimate

From Procurement Plan C90755-A

Hardware Class	*Procurement Expense*	*Maintenance @ 30%/yr* *
HPSS	*$1428.0K*	*$428.4K*
Servers	*$272.0K*	*$81.6K*
Switches	*$163.0K*	*$48.9K*
HW RAID, Small Robots	*$735.0K*	*$220.5K*
	Subtotal for major hardware	*$779.4K*

Hardware Class	*Procurement Expense*	*Maintenance @ 15%/yr*
Workststions, small switches, disks	*$1387.0K*	*$208.1K*
Totals, per year->	*$3985.0K*	*$987.5K***

*** Major equipment obtained at deep (educational, GSA) discounts carries non discounted maintenance fees; normal maintenance of 15% of list price becomes ~30% of discounted purchase price for LIGO**

	Skew towards maintenance during first half: FY2002 - FY2004	*Skew towards replacement during second half: FY2005 - FY2006*
	Maintenance, per year	*Maintenance, per year*
	$779.4K	*$428.4K*
	Replacement Equipment	*Replacement Equipment*
	$208.1K	*$559.1K*
Total, per year	*$987.5K*	*$987.5K*

**** Net maintenance costs for LDAS equipment are 25% per year of purchase price**

Data & Computing Operations Budget

LDAS Hardware Maintenance Budgetary Estimate

- Maintenance and replacement budget allocated to each site in order to take advantage of different tax and overhead structures among LIGO Laboratory sites.
- Caltech+MIT: $866k (includes $295k OH)
- LHO : $314k (includes $59k OH)
- LLO: $199k (includes $37k OH)

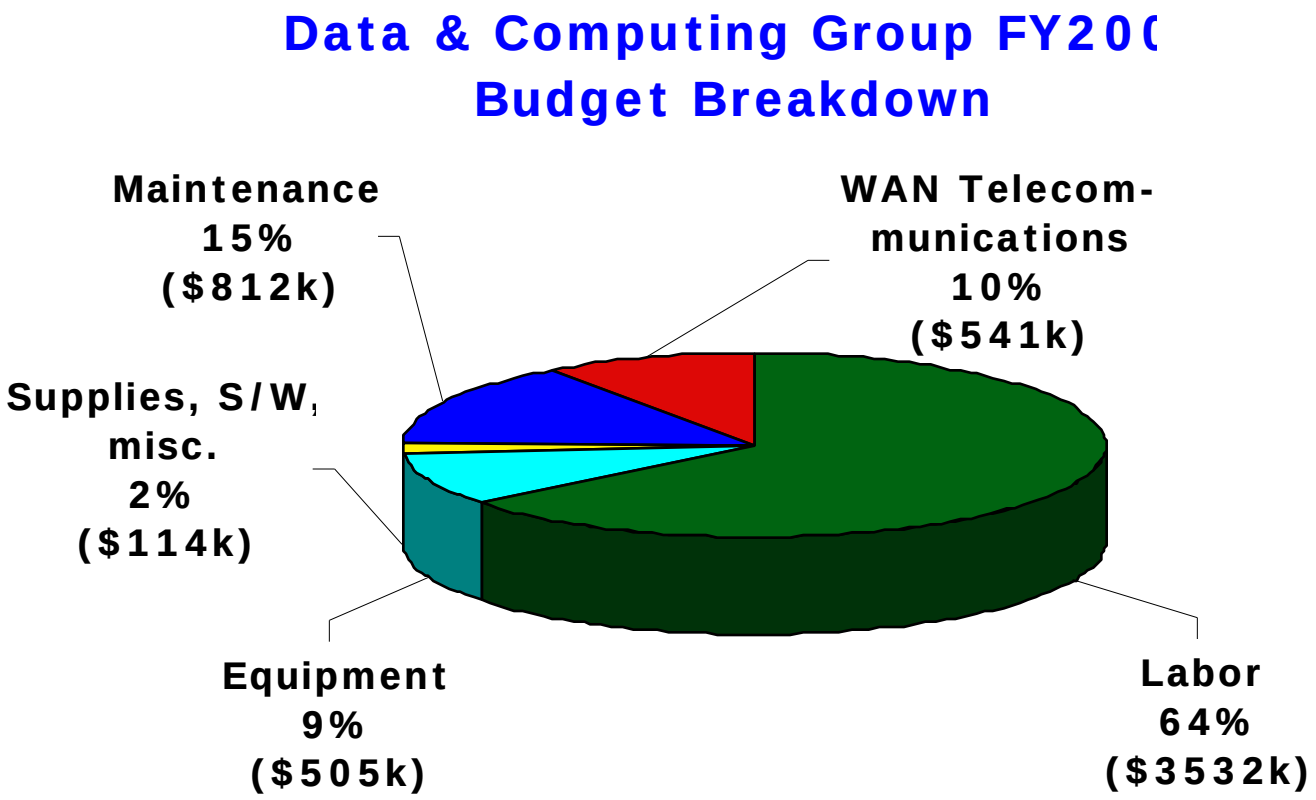
$1379k

LIGO WAN

- LIGO is working with DOE, PNNL, Internet2, and Abilene to establish a high bandwidth WAN (OC3) across the laboratory
  - » Caltech & MIT are already accessible
  - » LHO uses ESnet via PNNL for present internet access
    - – Established through an MOU between DOE and NSF in 1999
    - – Present MOU accommodates up to 4xT1 bandwidth
    - – Have begun move to OC3 via and ATM-ATM switch link between LHO and PNNL in Richland
    - – Estimated costs for OC3 operations from LHO-PNNL-Seattle: **$340k/year**
  - » LLO uses LAnet via LSU for present internet access
    - – Established through affiliation of LSU and LIGO at Livingston
    - – Present service accommodates up 1xT1 bandwidth
    - – Have had several meetings with State & BellSouth representatives to discuss move to OC3
    - – Estimated costs for OC3 operations from LLO-Baton Rouge: **$200k/year**
  - » Total estimated WAN operations costs for LHO + LLO: **$540k/year**

Data Analysis & Computing Budget

Breakdown



Summary

- LIGO Laboratory on threshold of transition to scientific observatory:
  - » Participation in astrophysical searches by the Laboratory scientists through the LSC
  - » Data analysis will become a key "business" of the Laboratory across all sites
    - – Database, archive use & maintenance
    - – Pipeline analysis SW and HW maintenance
    - – Continued R&D into algorithms, new technologies
  - » Collaboration within LSC requires continuing support by the Laboratory
  - » Simulation & modeling of interferometers needed to support detector understanding, data analysis
  - » LIGO WAN management crucial to daily operations

Plan to Reach Science Run

Detector & Data Analysis

- Jan to mid-March
  - ***LHO 2k, continued work on improving robustness of lock, some work on sensitivity***
  - ***LLO 4k, Lock single arm, recombined Michelson with Fabry-Perot (F-P) arms, Power Recycled Michelson (PRM)***
  - *LHO 4k, installation*
  - » HW: Procure Phase I, final RAID configurations, HPSS tape silo, small beowulf clusters for E6
  - » SW: Prepare LDAS release for E3
- March 9-12
  - ***E3 (engineering run): coincidence run between full 2km interferometer and recombined Michelson with F-P arms (possibly single arm) at LLO***
  - » SW/HW: Archive E3 data

Plan to Reach Science Run

Detector & Data Analysis

- mid-March to mid-May
  - ***LHO 4k, complete installation, lock mode cleaner***
  - ***LHO 2k, suspension sensor replacement, PRM studies***
  - ***LLO 4k, lock full interferometer, sensitivity/robustness***
  - » SW: Metadata Mock Data Challenge (Caltech/LHO)
  - » SW: Prepare LDAS release for E4
  - » HW: Install Phase I hardware
  - » HW: Benchmark HPSS on Sun hardware at Sun testbed facilities, Beaverton, OR
- May
  - ***E4 run: LLO 4 km only, operating in recombined mode (possibly recycling)***
  - » SW/HW: Archive E4 data
  - » SW: MPI inspiral search MDC (first of 4 MDCs tied to upper limits run)

Plan to Reach Science Run

Detector & Data Analysis

- May - June
  - ***LHO 2k, bring full interferometer back on-line, sensitivity studies***
  - ***LLO 4k, improve full interferometer lock, sensitivity studies***
  - ***LHO 4k, PRM locking (no arms yet)***
  - » SW: Prepare LDAS release for E5
  - » HW: Specify HPSS HW configuration for Phase II of procurement
- late June - early July
  - ***E5 LHO 2k and LLO 4k in full recycled configuration, LHO 4k in PRM mode***
  - » SW/HW: Archive E5 data
  - » SW: MDCs for remaining searches
- July - Sept
  - ***LLO 4 k suspension sensor replacement, bring back on-line***
  - ***LHO 2km sensitivity studies, 4k lock full interferometer***
  - » SW; Prepare LDAS release for E6
  - » SW: MDCs for remaining searches
  - » HW: Benchmark PCs for large beowulf procurement, Procure Phase II

Plan to Reach Science Run

Detector & Data Analysis

- late Sept
  - ***E6 triple coincidence run with all 3 interferometers in final optical configuration ("upper limit run")***
  - » SW/HW: Archive E6 data, on-site upper limit searches
- Oct – early 2002
  - *Improve sensitivity and reliability*
  - *Alternate diagnostic testing with engineering runs*
  - » HW: Specify HPSS drives, tapes, IDE/SCSI RADI 5 disk cache for data at Caltech; Procure Phase III
  - » SW: Data Archive MDC
- Jan - July 2002
  - » SW/HW: Prepare Release 1 of LDAS for Science run
  - » SW/HW: Integrated single interferometer running on-site
  - » SW/HW: multiple interferometer running off-site

LIGO/LSC Software Deployment Schedule

ID	Task Name
1	LDAS Software Releases
2	E2 release of LDAS
3	E3 release of LDAS
4	E4 release of LDAS
5	E5 release of LDAS
6	E6 release of LDAS
7	Release 1 of LDAS
8	LIGO/LSC Data Challenges
9	Data Conditioning Challenge
10	Metadata Data Challenge
11	MPI Pipeline Template Analysis Challenge
12	Data Archive Challenge
13	Science Inch Pebbles
14	Integrated Science Analysis (End to End 1 IFO) Chall
15	Integrated Science Analysis (End to End 3 IFO) Chall
16	LDAS Hardware
17	Caltech prototypes
24	LHO
31	LLO
38	CIT
54	MIT
58	Detector Milestones
59	Detector Commissioning
60	Detector Improvements/Noise Reduction
61	Engineering Runs
62	LIGO Science Milestones
63	Upper Limit Papers, Draft
64	First Science Run
65	Second Science Run
66	Third Science Run
67	First Science Paper, Draft