LIGO Data & Computing Update

Albert Lazzarini

LIGO Laboratory at Caltech

NSF Annual Review of LIGO Laboratory

MIT, 24-26 October 2002

Cambridge, MA

Outline for this talk

- **Responses to Prior NSF Recommendations**

- **Simulation & Modeling Update**

- **Data Analysis Systems Update**

- **Grid Computing**

Responses to previous recommendations
- LIGO Laboratory IT Security-

- Requested and received copies of several DOE National Laboratory IT security plans to obtain insight, guidance
  - » SLAC
  - » FNAL

- Used Caltech Administration Audit Services to perform an internal audit of LIGO (Caltech) IT system
  - » Findings, recommendations folded in LIGO IT Plan

- Developed a LIGO Laboratory IT Security Plan to cover all four sites of the Laboratory
  - » Recently approved by LIGO Directors
  - » Much of plan is already in place

LDAS usernames/password Secure Web Server

All shared account have been removed!

Responses to previous recommendations
- Networking (Wide area network)-

- Lab-wide:
  - » LIGO requested and received sponsored access authorization to Inet2, Abilene consortium through Caltech participation.
  - » Authorizes LIGO to peer to Inet2 at Gigapop servers in Baton Rouge, Seattle for Observatories (Caltech, MIT already on Inet2)

- LHO
  - » NoaNet (WA State PUC infrastructure) and PNNL/DOE chosen for connection to Internet2
    - – Have the option to bypass PNNL/DOE if necessary
  - » Discussions with both organizations lasted for more than 1 year.
  - » Recently upgraded to a 10Mbps connection through PNNL and DOE
    - – Actual peak capability is 6Mbps -- superior to previous T1 (<1.55 Mbps)
  - » Discussion proceeding to get OC3 (or 100 Mbps) connection to Inet2.
    - – Technical details being addressed by NoaNet and PNNL
    - – Final logistics should be completed by December
  - » Installation completion estimated by April of 2003 -- mid S2
    - – *SUBJECT TO AVAILABILITY OF FUNDS*

LIGO-G020464-02-E

Responses to previous recommendations
- Networking (Wide area network)-

- LLO
  - » 2 x T1 lines were installed in 2001
  - » We have experienced difficulties in getting a further upgrade
    - Bell South pricing structure unacceptably high;
    - Third party carriers no longer viable options due to economic downturn.
    - RF option considered briefly, then dismissed;
      - Large costs and too many unknowns with microwave towers
  - » Local cable company may provide a solution -exploring this with LSU
- MIT
  - » MIT LIGO group has met with MIT IT services to discuss present situation
    - Performance is not as good as other LIGO sites
      - » CIT-to-MIT >> MIT-to-CIT
      - » 17 Mbps (~2MB/s) vs. <1Mbps (less than T1)
  - » Options are few - MIT is not likely to welcome on-campus carrier competition
    - DOE connection to ESnet at MIT via the NSF-DOE MOU
      - Depends on MIT's acceptance of DOE infrastructure expansion on campus
    - Third party commercial carriers not obvious
      - Cogent -"last mile problem"

Responses to previous recommendations
- Networking (Wide area network)-

- **Caltech**
  - » Current network access via CACR's High Performance Network to Calren2/Inet2
    - – 4 x OC12

  - » Campus-wide upgrade being implemented over the next 2 quarters.
    - – 3 x 10 Gbps as part of Teragrid (dedicated to Teragrid and network research)
    - – 1 Gbps for campus traffic
    - – LIGO, CACR, and Caltech Information Technology Services (ITS) are discussing options for LIGO's access to Inet2 after reconfiguration

  - » Current experience shows LIGO utilizes sustained bandwidths of 20MB/s over multiple 24-hour periods when migrating data from Tier 1 center at Caltech to Tier 2 center at UWM

Responses to previous recommendations
-High performance computing and mass storage-

- Made site visit to BNL (Star, Phenix, ATLAS) HPSS site;
  - » Met with RCF management, administrators
  - » Discussed RCF experience with then-emerging STK 9940 drive technology

- Made site visit to NCAR
  - » Met with Bernard O'Lear and NCAR IT group
  - » Compared our needs to NCAR needs, approaches -- little overlap
  - » Learned that significant HW resources needed to migrate existing data from older to next technology

- 3Q2001 - now: evaluation of Sun/LSC SAM-QFS technology
  - » Following positive experience with SAM-QFS at SDSC
  - » Discussed with IRIS Project (R. Braman, Dept. of Geophysics, U. Washington, Seattle)

- Supercomputing 2001, 2002 participation with GriPhyN collaborators

Responses to previous recommendations
- Database administration -

- LIGO has hired a scientist at Livingston Observatory as on-site data & computing liaison, data analysis scientist
- Dr. Yakushin has begun to assume Laboratory-wide role as DB expert
- Coordinates with Caltech group
  - » DB backups, migration, federation, mirroring
- Dr. Yakushin is growing into DB management role
  - » Preferable to professional IT database administrator
    - – Scientific background and experience

Simulation & Modeling for LIGO I
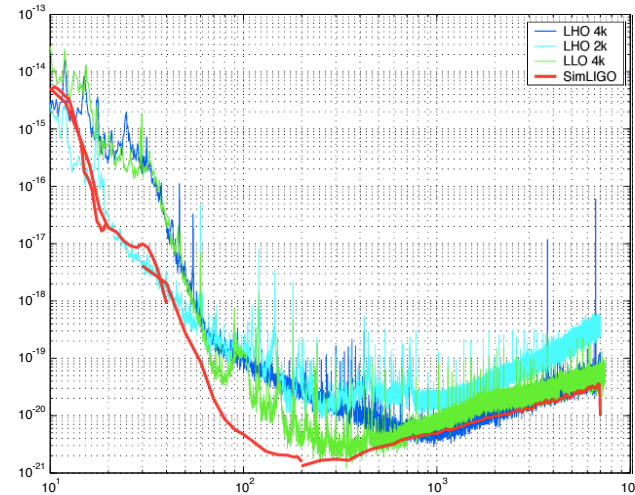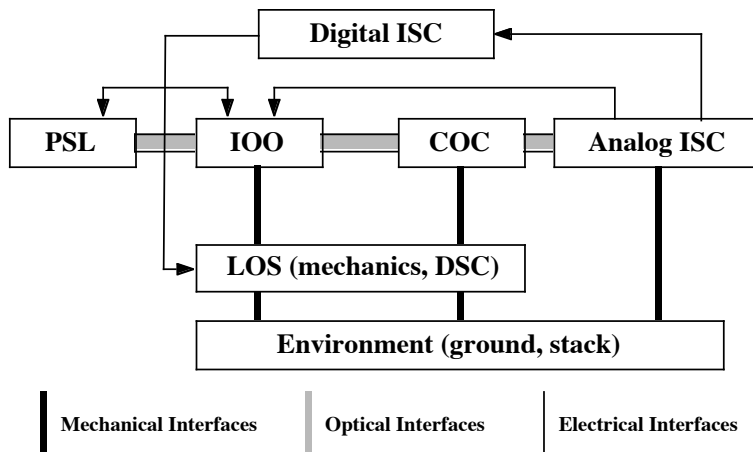
- *Hiro Yamamoto (Task Leader)*
  - Biplab Bhawal (Scientist)
  - Luca Matone (LHO postdoctoral scholar)
  - Matt Evans (Graduate Student, Staff Programmer/Scientist)
  - Bruce Sears (Staff Programmer)
  - Melody Araya (Contract Programmer)

Modeling & Simulation Activities Overview

E2E ("end-to-end")

- **General purpose GW interferometer simulation framework**
  - » Generic tool like matlab or mathematica
  - » Time domain simulation written in C++ (enables modeling non-linearities)
  - » Optics, mechanics, servo, ...
    - – time domain modal model : $TEM_{mn}(t)$
    - – single suspended 3 D.O.F. mass, modular mechanical modeling tool (MSE)
      - *Virgo Project collaboration, contribution*
    - – analog and digital controllers - ADC, DAC, digital filters, etc

- **LIGO I simulation packages built upon E2E**
  - » *Han2K*: rudimentary model, used for the lock acquisition design, studies:
    - – simple seismic noise, all analog servo, 1 D.O.F. mass motion, etc
    - – allows a fast simulation of LIGO I interferometers
    - – includes cavity response with misalignment
    - – thermal lensing interactions with lock acquisition system

LIGO-G020464-02-E

LIGO Modeling & Simulation Activities Overview

E2E ("end-to-end")

- LIGO I simulation packages built upon E2E (continued)
  - » *SimLIGO* : available for LIGO I commissioning, assist identification of noise sources
    - Modal beam representation
      - alignment, mode matching, thermal lensing
    - 3 dimensional mechanics
      - 6x6 D.O.F. stack transfer function
      - 3D optics with 4 local sensor/actuator pairs
    - Complete analog and digital electronics servo models with noise
      - Common mode feedback
      - Wavefront sensing
      - "Noise characterization of the LHO 4km IFO LSC/DSC electronics" study by P. Fritschel and R. Adhikari, 12-19 March 2002 incorporated into model
    - All major noise sources included:
      - seismic, thermal, sensing, laser frequency and intensity, electronics, mechanical

LIGO data vs. SimLIGO

Triple Strain Spectra - Thu Aug 15 2002

Legend:
- LHO 4k
- LHO 2k
- LLO 4k
- SimLIGO
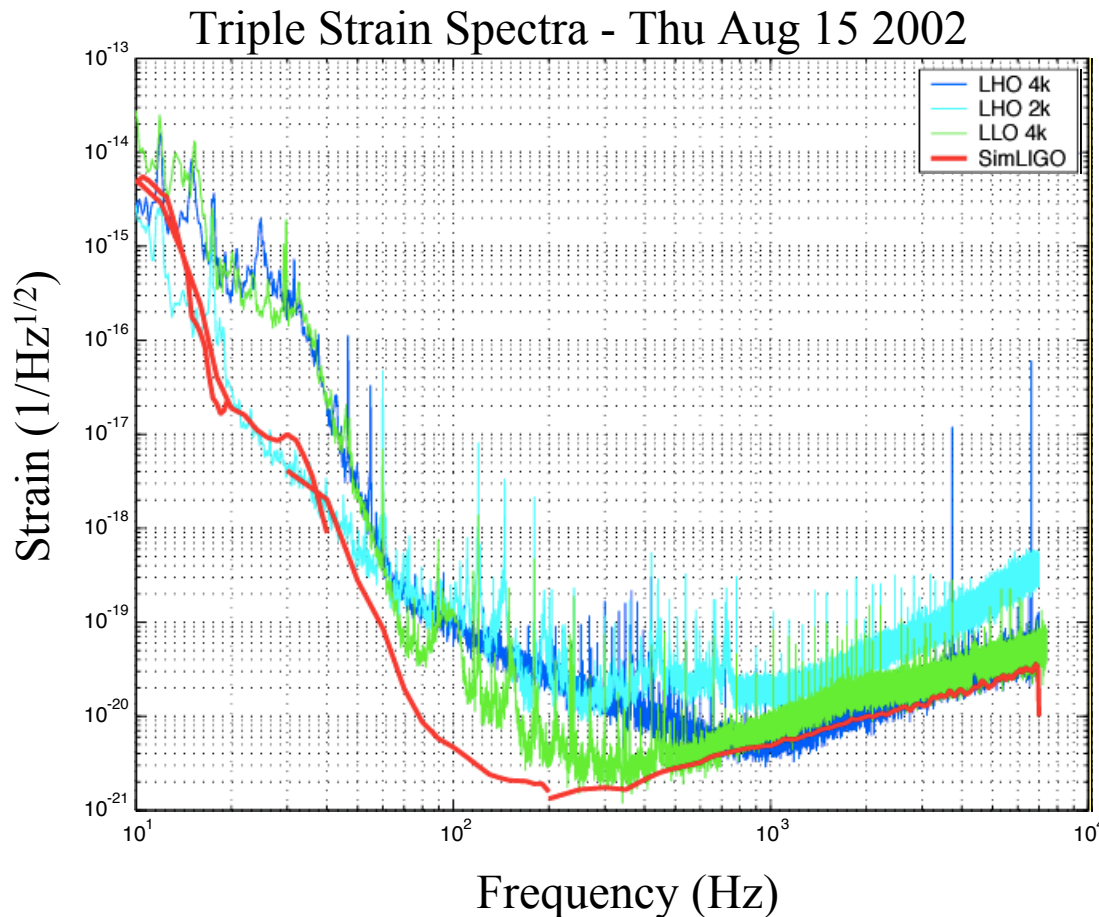
Simulation includes:
- Locked interferometer
- Mechanics
- Sensor-actuator, optical lever damping
- Servo electronics
- Signal extraction
- Simulated and measured noises
- …

▪ *Data provided by Rana Adhikari (MIT)*

E2E -- future plans

- ## Evaluation of the as-built LIGO performance
  - » Identify minimum achievable noise performance for the as-built design
  - » Study, quantify specific noise sources and issues:
    - – Bi-linear couplings, non-linearities, non stationarity, transients, lock instabilities, etc
  - » LIGO simulation package is ready
    - – Length control: complete
    - – Alignment control: complete

- ## Outreach -- E2E will be used for education
  - » Southeastern Louisiana University plans to use E2E to teach undergraduate physics
    - – Virtual optics laboratory

- ## Apply E2E to Adv. LIGO when appropriate

Data Analysis Systems
- Software -
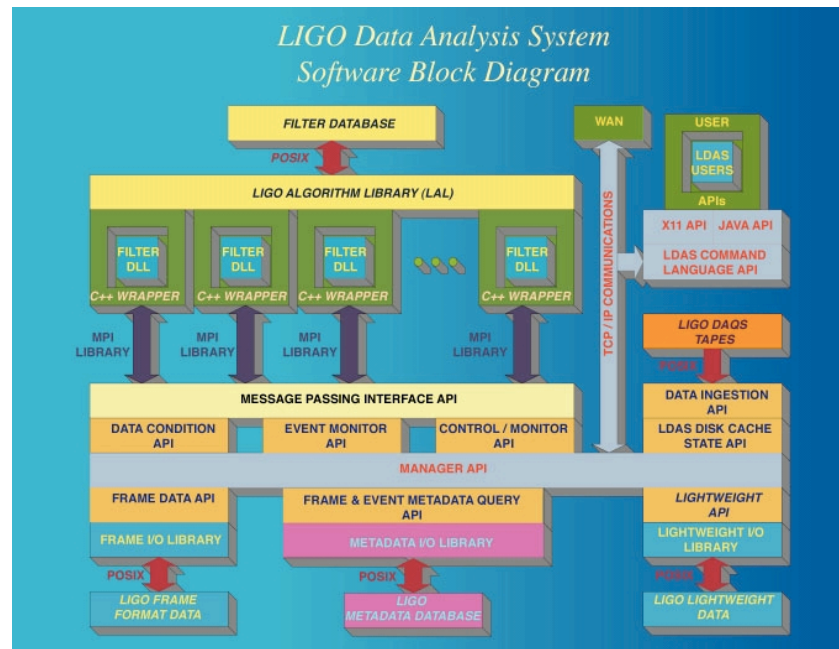
- **_J. Kent Blackburn (Task Leader)_**
  - Peter Shawhan (Scientist)
  - Philip Charlton (Postdoctoral Scholar)
  - Igor Yakushin (LLO Scientist)
  - Greg Mendell (LHO Scientist)
  - Edward Maros (Staff Programmer)
  - Philip Ehrens (Contract Programmer)
  - Masha Barnes (Contract Programmer)
  - Mary Lei (Contract Programmer)
  - Isaac Salzman (Contract Programmer)

LIGO-G020464-02-E

Data Analysis for LIGO Overview

- Pipeline Analysis Design encompasses two components:
  - » LDAS -- LIGO Data Analysis System
    - Pipeline analysis system for parallelized searches using linux clusters (beowulfs)
    - Data pre-processing, conditioning, reduction
    - Astrophysics source search codes running under parallel environment
    - Large data archive/data cache of raw frame data
    - Integrated with a relational database to store outputs of pipeline analysis

  - » DMT -- Data Monitoring Tool
    - DMT provides an environment for online analysis of interferometer and environmental data.
      - Provides operators with immediate feed-back on interferometer state and operating conditions
    - Record interferometer state summaries and special events (e.g. transients) for use in offline analysis.
    - Provides instrumental, environmental vetoes to science analysis

Astrophysical Search Pipeline
- example: burst group analysis -

LLO

Strain Data

Quality Check Data split

Burst Analysis Algorithms (DSO) Feature Extraction *LDAS*, DB (T, δT, SNR)

Aux Data (non GW)

...

Aux Data (non GW)

Glitch Analysis Algorithms (*DMT*) Feature Extraction *DB* (T, dT, SNR)

GW/Veto anticoincidence Event Analysis Tools

Simulated data Based on Astrophysical Source Knowledge

Sanity Checks

Quantify Upper Limit

Quantify efficiency

IFO-IFO Coincidence And Clustering

LHO

Strain Data

Quality Check Data split

Burst Analysis Algorithms (DSO) Feature Extraction *LDAS*, DB (T, δT, SNR)

Aux Data (non GW)

...

Aux Data (non GW)

Glitch Analysis Algorithms (*DMT*) Feature Extraction *DB* (T, dT, SNR)

GW/Veto anticoincidence Event Analysis Tools

LIGO-G020464-02-E

LDAS Performance
- E7, S1 -

- E7: 28 December 2001 - 14 January 2002
  - » LDAS up entire run: Approximately one job every 10 seconds (averaged).
    - – Approximately five rows (events) per second (averaged).
  - » Greater than 92% of jobs completed successfully
    - • LHO roughly 92%; LLO roughly 95%;
    - • Resubmission of few failed jobs

- S1: 23 August - 09 September 2002
  - » Averaged approximately one job every 5 seconds
    - – 2X E7 rate
  - » Averaged approximately five rows (events) inserted each second
    - – same as E7
  - » Total failure rate roughly a 1%-2%
    - – 99% of failures due to operator usage – not software

LIGO-G020464-02-E

DMT Summary, S1
- DMT -

- Both observatories ran with DMT

Loss Mechanism	LHO	# data frames	LLO	# data frames
Data Acq. Sys. reboot	8	~1750	1	~130
DMT reboot	1	~300	2	~750
Other (transmission errors, data not available, etc.)	-	~420	-	~20
Total unanalyzed frames	-	2470 (0.2%)	-	900 (0.1%)

LIGO-G020464-02-E

S1 Summary
- LDAS -

- All sites ran with LDAS release 0.4.0
  - » LDAS-LHO restarted once (database stalled)
  - » LDAS-LLO restarted once (power outage)
  - » LDAS-MIT restarted 6 times (don't know reasons?)
- CIT not yet configured to see full archive (frames)
- RDS frame generation began ~ 2nd day
- Data Pipeline searches at sites ~ week late
  - » August 29th at LHO
  - » August 30th at LLO
- *Very few jobs failed due to LDAS bugs*

LIGO-G020464-02-E

S1 Summary
- DMT -

- ## Current status:

  - 30+ monitors have been written for use under the DMT release 2.2.3

    - bandlimited channel rms, transient, glitches, earthquakes, wind, etc.

  - DMT is running 24/7 at LIGO observatories

    - Hanford:

      - 41 monitor processes to check data QA

      - 20 support processes

    - Livingston:

      - 27 monitor processes to check data QA

      - 15 support processes

LDAS Software To Do: Short List for S2

- Finish integration of new *frameCPP* based on Frame Specification Version 6 (previously labeled 5).

- Develop Federated Database solution between LDAS sites.

- Further improve performance of LDAS.

- Integrate SAM/QFS storage technology into *diskCacheAPI*.

- Improve metadata handling in *dataConditionAPI*.

- Implement pool of DB2 connections in *metaDataAPI*.

- Implement new actions, improved metadata, resolve remaining memory leaks in *dataConditionAPI*.

- Close out problem reports.

LIGO-G020464-02-E

Changes to Release Schedule
Prompted by LIGO Science Run Schedule

- **Accelerated release of 0.4.0 to support rescheduled S1**
- **Release 0.5.0 to support new data Frame 6 Specification**
- **Likely to have 2 additional β versions before 1.0.0**

LIGO-G020464-02-E
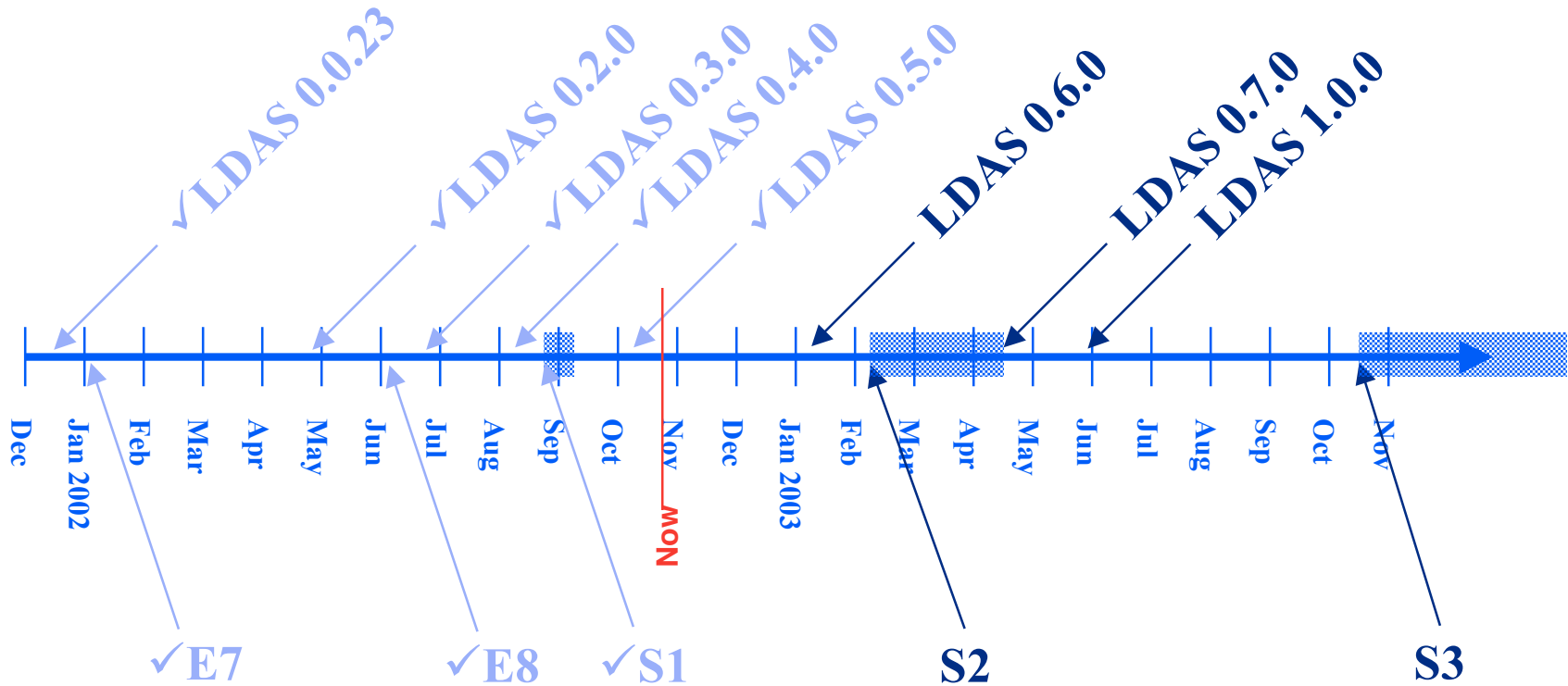
Data Analysis Systems
- Hardware -

STK@CIT

MPI Cluster@ LHO

T3 RAID @LHO

- ***Stuart Anderson (Task Leader)***
  - Greg Mendell (LHO Scientist)
  - TBD (LHO Admin)

  - Shannon Roddy (LLO Admin)
  - Igor Yakushin (LLO Scientist)

  - Dan Kozak (Contract Administrator)
  - Al Wilson (Contract Administrator)

Current LDAS
Hardware Configuration

LDAS Installation	SAN (TB)	IDE (TB)	CPU (Aggregate GHz)	Tape (TB)
LHO	10	2	139	2
LLO	5	2	107	2
CIT	3	16	34	92 (includes S2)
MIT	1	2	45	0
DEV	1	2	25	2
TEST	1	0	8	0

SAN -- Storage Area Network (disk farm)
IDE -- Intelligent/Integrated Drvie Electronics (commodity disks)
CPU -- Central Processing Unit s, beowulf cluster PC nodes
TB -- 10^{12} byte

LIGO Laboratory

SAM-QFS vs HPSS
Last major trade study, completed before S2:

- SAM-QFS - Sun Microsystems acquisition (LSC), provides alternative to HPSS for integrated mass storage (disk+tape)
  - » http://www.sun.com/products-n-solutions/hardware/docs/Software/Storage_Software/Sun_SAM-FS_and_Sun_SAM-QFS_Software/index.html
- SAM-QFS advantages
  - » Simplicity/reliability -- *Fewer FTEs*
  - » Media import/export
  - » License cost allows for use at observatories
  - » Disaster recovery (GNU TAR)
  - » Metadata performance (x1000)
  - » Single vendor solution (server, software and OEM storage)
  - » Reduced dependency on CACR
  - » 16 TB S1 data successfully stored and *retrieved* at 27MB/s
- HPSS advantages
  - » Integrated several man-years experience
  - » Free at Caltech
    - – cannot be used other Laboratory sites
  - » 54 TB E1 - S1 data successfully *stored* to date

Selected SAM-QFS for LIGO I

LIGO-G020464-02-E

Remaining hardware procurements S2 and beyond

- Additional Fibre Channel ports to connect disk and tapes at LHO and LLO.
  - Planned availability -- S2 (14 February 2003)
- STK9940B tape drives for archive system at Caltech.
  - Planned availability -- S2 (14 February 2003)
- Tape robotics at LHO and LLO (700 slots ea., 1 year on-line look-back & backup)
  - Planned availability -- S2 (14 February 2003)
- Large volume IDE RAID storage @ Caltech, in front of tape archive.
  - Build up disk farm as data volume grows over next several years
- S2 Beowulf clusters:
  - Get quotes, test performance for 32-bit Intel and AMD processors.
  - ~400 compute nodes
  - Procure 4Q02-very early 1Q03

- S3 Server configuration - operations upgrade
  - Data conditioning servers will be upgraded to Science Run configuration.
    - The desire to get to 64-bit platforms may push this out until 2-3Q03
  - Expand Sun V880 SPARC servers to full CPU configurations (4->8), 900 MHz CPUs

LIGO-G020464-02-E

Data Analysis
- Grid Computing -

GriPhyN: Grid Physics Network
iVDGL: International Virtual Data Grid Laboratory

Grid activities
GriPhyN & iVDGL

- LIGO/Caltech, UWM, PSU members of collaborations

- GriPhyN - CS + Applications research focused on:
  - » Virtual data in LIGO analysis
    - Use of transformation catalogs to keep track of previously requested, performed transformations
    - Determine if the data product is instantiated, if so where
    - Plan & execute data movements and computations required to support specific analyses
  - » LIGO data mirroring using gridtools
    - Automated, robust replication between Tier 1 & Tier 2 sites
  - » Developing an interface between the grid environment and the native LDAS environment
    - Extending LDAS into a grid-enabled environment -- longer term goal
  - » Grid security infrastructure
    - Provide data access only to collaboration members
    - Provide a level of security to LDAS computing

Grid activities
GriPhyN & iVDGL

- iVDGL will provide resources to deploy, support (in a limited fashion) the first 2 LIGO Tier 2 centers:

  » UWM

    – extension of existing MRI-funded hardware - provides out-year hardware renewal

  » PSU

    – greenfield Tier 2 center to be implemented by end of CY2002

  » *Limited FTE support to operate centers*

    – *Postdocs, students, very limited IT support*

Integration with GRID

- LIGO/Caltech is working closely with USC-ISI, UWM/NCSA

- Globus Toolkit installed on LDAS software servers

- LDAS integrated (with thin layer) onto the GRID

- Developed demonstration prototype application for SC 2001

- Will expand demonstration for SC 2002 this year
  - » New standalone wrapperAPI (remoteAPI)
  - » Pulsar search will be highlighted

LIGO-G020464-02-E

ITR 2003
Proposal to NSF for iVDGL Operations

- **ITR 2003 announcement of opportunity issued by NSF this summer**

- **LSC Computing Committee organizing a collaboration-wide effort to request funding to operate LSC Tier 2 centers for LIGO Science Run(s)**

  - » *ITR2000: GriPhyN ->C/S + Applications R&D, prototyping*

  - » *ITR2001: iVDGL-> Center buildup (2 for LSC), port, install GriPhyN deliverables to Tier 2 centers*

  - » ITR2003: request manpower to operate centers **to do the science**

    - Scientists, systems administration (~ 10 FTE across 2 LSC sites, Lab. sites)

    - Help desk, 7x24 operations support

LIGO-G020464-02-E