

# **State of LSC Data Analysis and Software**

**LSC Meeting**

**LIGO Hanford Observatory**

**November 11<sup>th</sup>, 2003**

**Kent Blackburn, Stuart Anderson, Albert Lazzarini**

# LIGO Data Analysis Hardware

- **LIGO's computer hardware installation complete!**
- Compute clusters:
  - LLO = 1X; LHO = 2X; CIT = 3X  $\Rightarrow$  420x2 total CPUs
    - “X” = 70 dual-2.66GHz P4-Xeon nodes, 2GB RAM, 200GB disk, & GigE
  - MIT = 112 single-2GHz P4 nodes with 512MB of RAM and FastEthernet
- Mass Storage:
  - Tape robotics with 4 STK tape drives for each Observatory(140TB)
  - 30TB disk cache & 10 STK tape drives for central archive (1.2PB)
    - 114 TBs of data currently archived (expect ~46TB more from S3)
- Servers:
  - Data (frame) servers upgraded to 8 CPU Sun servers just before S3
  - Database servers upgraded to dual-Intel servers just before S3
- Caltech cluster now running both LDAS & Grid software (e.g., condor)

# LIGO Data Analysis Software

- Preparation for S3 science run:
  - Pre-release of LDAS 0.8.0 software CVS tagged for RDS Mock Data Challenge
  - LDAS 0.8.0: new release of software for E10 run (*dry run*)
  - LDAS 0.9.0: new release of software for S3 run
    - Finished infrastructure requirements
    - Significant performance improvements
      - Infrastructure 2 – 3 times faster
      - Database access 10+ times faster
      - Some search codes that run under LDAS have been reworked to give O[10] speedup
    - Significantly more reliable software
- LDAS 1.0.0 software to be released shortly after S3 run ends
  - Plan is to fix any remaining issues identified by S3 run
- LDAS “2” software will be defined by LSC goals
  - Expect to include Grid technologies (ITR 2003)
  - Greater support role (recent successes with reworking some search codes)
  - Many paths to making software “better” available (faster, easier to use, etc.)
  - *Other flowers are blooming and more blossoms may appear*

# S3 Run Update

- Extremely smooth reduced data set frame generation
  - Rate of RDS as a multiple of real-time
    - LHO: L1 - 4.8x, L2 - 18x, L3 - 23x
    - LLO: L1 - 6x, L2 - 11x, L3 - 13x
    - CIT: L1 - 3.5x, L2 - 23x
  - Time between frame-file GPS time and reduction
    - LHO: L1 - 15 min, L2 - 30 min, L3 - 50 min
    - LLO: 5 min, L2 - 15 min, L3 - 35 min
  - FTP Transfer rates
    - LHO->CIT 4 MB/s, LLO->CIT 0.1 MB/s
  - Delay in L0->L1->L2->L3->CIT->MIT pipeline:
    - Total Level 3 RDS delays to MIT typically 2-3 hours!
- Data Pipelines should startup this week.

# ITR Proposal

- *Grant awarded to ITR2003 Proposal*
  - Proposal for 3 key tasks to complete deployment, support operation of Tier 2 Centers
    - Provide authenticated and secure access to LIGO data by collaboration members by developing a Tcl API to the Globus Package for use with existing infrastructure
    - Continued development & deployment of LIGO Data Replicator (LDR)
      - Efficient and secure mirroring of critical datasets across the LIGO Data Grid.
    - Port DMT software to a grid-based computing model
  - Proposal requests 6 FTEs for 4 years
  - 4 year award for \$3M
  - FTE allotment:
    - 1.5@CIT
    - 1@MIT
    - 1.75each @ UWM & PSU

# Super Computing 2003

- **Preparation for Supercomputing 2003 (“SC2003”)**
  - Scaled up from SC2002 to support large-area unbiased periodic source search (*Near term goal: ~1000 nodes*)
  - Joint effort with CS team at USC/ISI, and LSC CW working group -- Caltech, UWM, AEI
  - Targeted search of galactic center using pulsar demodulation code developed at AEI
    - Will use rich complement of “grid tools” to stage data, results, on/off distributed resources
    - Long list of cluster resources that will work together:
      - AEI, Birmingham, Caltech, Cardiff, LSU, USC/ISI, UTB, UWMadison, UWMilwaukee, possibly Teragrid, iVDGL Grid3 Testbed resources
- **THIS IS NO DEMO: intent is to achieve publishable results to be included in the next CW paper(s)**

# LSC Computing Update

## *New R&D Activities*

---

- LSC Computing Committee considering submitting proposal to next year's ITR2004
  - Last year for ITR proposals
  - Develop a proposal based on load balancing, scheduling of different jobs using different compute models on common distributed resources
    - Need identified within the grid development community; no tool exists other than “hand tuning”
- 2005 World Year of Physics Einstein Celebration
  - AIP approached LSC to develop a SETI@home - like screensaver to promote gravitational wave searches, links to Einstein
    - Plan is to team with AIP, Berkeley (SETI@home team) to port the all-sky pulsar search codes to this format
  - May propose to either ITR or Education and Interdisciplinary Research Program (EIR) of NSF